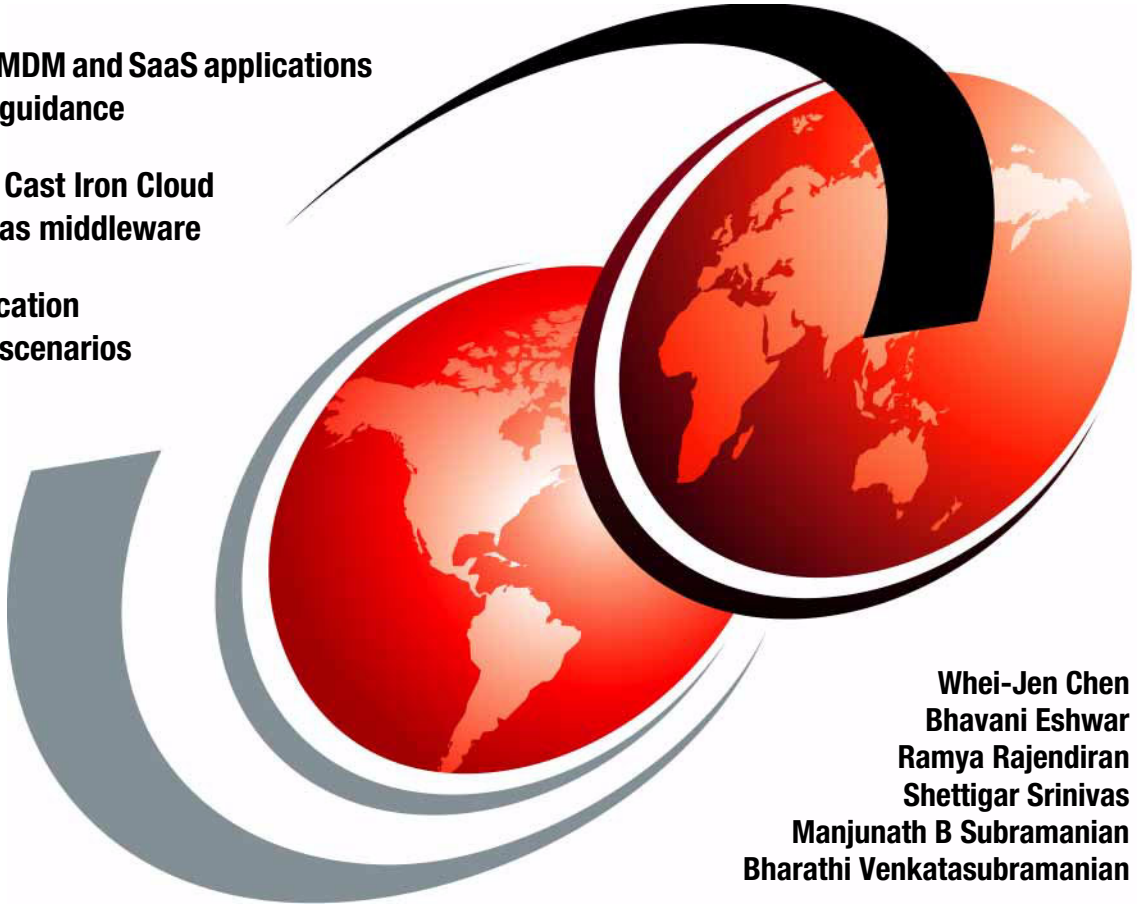
IBM

Master Data Management for SaaS Applications

- InfoSphere MDM and SaaS applications integration guidance
- WebSphere Cast Iron Cloud integration as middleware
- Cloud application integration scenarios



Whei-Jen Chen
Bhavani Eshwar
Ramya Rajendiran
Shettigar Srinivas
Manjunath B Subramanian
Bharathi Venkatasubramanian

Redbooks

ibm.com/redbooks

International Technical Support Organization

Master Data Management for SaaS Applications

October 2014

SG24-8220-00

Note: Before using this information and the product it supports, read the information in “Notices” on page vii.

First Edition (October 2014)

This edition applies to IBM InfoSphere Master Data Management Version 11.3, IBM InfoSphere Server Version 11.3, WebSphere Cast Iron Cloud Integration Version 7.0.

Contents

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:
This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. You may copy, modify, and distribute these sample programs in any form without payment to IBM for the purposes of developing, using, marketing, or distributing application programs conforming to IBM's application programming interfaces.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. These and other IBM trademarked terms are marked on their first occurrence in this information with the appropriate symbol (® or ™), indicating US registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at http://www.ibm.com/legal/copytrade.shtml

The following terms are trademarks of the International Business Machines Corporation in the United States, other countries, or both:

BigInsights™
Bluemix™
Cast Iron®
Cognos®
DataStage®
DB2®
developerWorks®
IBM Watson™
IBM®
InfoSphere®
QualityStage®
Redbooks®
Redbooks (logo) ®
SPSS®
Watson™
WebSphere®

The following terms are trademarks of other companies:

Linux is a trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Java, and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

Other company, product, or service names may be trademarks or service marks of others.

Preface

Enterprises today understand the value of employing a master data management (MDM) solution for managing and governing mission critical information assets. chief data officers and chief information officers drive MDM initiatives with IBM® InfoSphere® Master Data Management to improve business results and operational efficiencies, which can help to lower costs and to reduce the risk of using untrusted master information in business process. Cloud computing introduces new considerations where enterprise IT architectures are extended beyond the corporate networks into the cloud.

Many enterprises are now adopting turnkey business applications offered as *software as a service* (SaaS) solutions, such as customer relationship management (CRM), payroll processing, human resource management, and many more. However, in the context of MDM solutions, many organizations perceive risks in having these solutions deployed on the cloud. In some cases, organization are concerned with the legal restrictions of deploying solutions on the cloud, whereas in other cases organizations have policies and strategies in force that limit solution deployment on the cloud.

Immaterial of what all the cases might be, industry trends point to a prediction that many “extended enterprises” will keep MDM solutions on premises and will want its integrations with SaaS applications, specifically customer and asset domains. This trend puts a key focus on an important component in the solution construct, that is, the cloud integration middleware and how it fits with hybrid cloud architectures that span on premises and cloud services. As this trend pans out, the on-premises MDM solution integration with SaaS applications will be the key pain point for the “extended enterprise.”

This IBM Redbooks® publication provides guidance to chief data officers, chief information officers, MDM practitioners, integration architects, and others who are interested in the integration of IBM InfoSphere Master Data Management with SaaS applications. This book lays the background on how mastering and governance needs for SaaS applications is quite similar to what on-premises business applications would need. It draws the perspective for serving the on-premises application and the SaaS application with the same MDM hub. This book describes how IBM WebSphere® Cast Iron® Cloud Integration can serve as the “de-facto” cloud integration middleware to integrate the on-premises InfoSphere Master Data Management systems with any SaaS application by using Saleforce.com integration as an example. This book also covers aspects of handling bulk operations with IBM InfoSphere Information Server.

After reading this book, you will have a good understanding about the considerations for on-premises InfoSphere Master Data Management integration with SaaS applications in general and Salesforce.com in particular.

The MDM practitioners and integration architects will understand the deployable integrations patterns and, in general, will be able to effectively contribute to delivering strategies that involve building solutions in this area. Additionally, SaaS vendors and customers looking to build or implement SaaS solutions that might require trusted master information will be able to use this compilation to ensure that the right architecture is put together and adhered to as a set of standard integrations patterns with all the core building blocks is essential for the longevity of a solution in this space.

Authors

This book was produced by a team of specialists from around the world working at the International Technical Support Organization, San Jose Center.

Whei-Jen Chen is a Project Leader at the International Technical Support Organization, San Jose Center. She has extensive experience in application development, database design and modeling, and IBM DB2® system administration. Whei-Jen is an IBM Certified Solutions Expert in database administration and application development and an IBM Certified IT Specialist.

Bhavani Eshwar is a Senior Technical Staff Member in the R&D Team of IBM InfoSphere Master Data Management. He has worldwide responsibilities for technical oversight on MDM Architecture, MDM Integrations, and MDM Cloud Offerings. He works out of the India Software Labs of IBM India Pvt. Ltd. and has additional responsibilities for Talent Development, Geo Expansion, University Relations, and Intellectual Property Development as a Senior Technical Leader. He has been working in the area of information integration, data governance, and MDM in IBM Information and Analytics Group for about 10 years. He has around 19 years of IT industry experience. Bhavani is a member of the IBM Academy of Technology and he is also a member of various world wide architecture boards in IBM overseeing the technical developments in Products. He holds a few patents in the technology areas he works in.

Ramya Rajendiran is a Software Developer at IBM India Software Labs in Bangalore, India. She has a Bachelors degree in Computer Science Engineering and Masters degree in Management of Technology. She has close to 8 years of industry experience and has held technical lead roles with IBM WebSphere Commerce, IBM Supply Chain, IBM WebSphere Adapters, and IBM WebSphere Cast Iron Cloud integration. She has also authored with a team *WebSphere Commerce Line-Of-Business Tooling Customization*, SG24-7619.

Shettigar Srinivas is an Senior Developer working on Integration Portfolio in IBM InfoSphere Master Data Management in India Software Labs, Bangalore. He is currently working on the integration between Salesforce.com and InfoSphere Master Data Management. Before Integration Portfolio, Srini worked on traditional Master Data Management development. His areas of expertise are MDM, master data policy management, and integration around MDM. He holds a bachelor degree in electrical and electronics.

Manjunath B Subramanian works as a Product Manager for IBM InfoSphere Master Data Management. Manjunath has more than 12 years of experience in the field of MDM and has worked in a variety of client facing roles during this time. In his current capacity, he is responsible for defining the road map and strategy for InfoSphere Master Data Management installation, upgrade, configuration, and SaaS integration.

Bharathi Venkatasubramanian is a Senior Software Engineer at India Software Labs. She started her career with the WebSphere Adapters team in 2006 where she was responsible for the development and support of adapters like FTP, Enterprise Content Management, and Oracle E-Business Suite. She is now a lead developer for the WebSphere Cast Iron Cloud Integration team, specializing in Cast Iron Connectors and Cast Iron Studio. She has published many articles related to integration on IBM developerWorks®. She holds a Bachelors degree in Computer Science from BITS Pilani India.

Acknowledgement

Thanks to the following people for their contributions to this project:

Rajath Ramesh
Jeffrey I. Dare
Raghunathan Krishnamurthy
Software Group, IBM India Pvt Ltd

Jon Case
Scott Schumacher
Software Group, IBM USA

Joseph Tsang
Software Group, IBM Canada

Now you can become a published author, too!

Here's an opportunity to spotlight your skills, grow your career, and become a published author—all at the same time! Join an ITSO residency project and help write a book in your area of expertise, while honing your experience using leading-edge technologies. Your efforts will help to increase product acceptance and customer satisfaction, as you expand your network of technical contacts and relationships. Residencies run from two to six weeks in length, and you can participate either in person or as a remote resident working from your home base.

Find out more about the residency program, browse the residency index, and apply online at:

ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us!

We want our books to be as helpful as possible. Send us your comments about this book or other IBM Redbooks publications in one of the following ways:

- Use the online **Contact us** review Redbooks form found at:

  ibm.com/redbooks

- Send your comments in an email to:

  redbooks@us.ibm.com

- Mail your comments to:

  IBM Corporation, International Technical Support Organization
  Dept. HYTD Mail Station P099
  2455 South Road
  Poughkeepsie, NY 12601-5400

Stay connected to IBM Redbooks publications

- Find us on Facebook:

  http://www.facebook.com/IBMRedbooks

- Follow us on Twitter:

  http://twitter.com/ibmredbooks

- Look for us on LinkedIn:

  http://www.linkedin.com/groups?home=&gid=2130806

- Explore new Redbooks publications, residencies, and workshops with the IBM Redbooks weekly newsletter:

  https://www.redbooks.ibm.com/Redbooks.nsf/subscribe?OpenForm

- Stay current on recent Redbooks publications with RSS Feeds:

  http://www.redbooks.ibm.com/rss.html

1

Introduction

This chapter discusses the business drivers for integrating an on-premises IBM InfoSphere Master Data Management (InfoSphere MDM) system with a software as a service (SaaS) application, such as Salesforce.com. The chapter also describes the business challenges and technical challenges that must be overcome, along with the capability considerations that must be enabled for this integration to be successful. It includes the following topics:

- Master data management and SaaS applications
- The evolution of SaaS applications
- SaaS applications require data mastering and governance similar to on-premises applications
- Challenges in overcoming business and operational issues from within the SaaS applications using a native MDM or internal MDM
- Other integration challenges
- Setting the stage for MDM integration with SaaS applications

1.1 Master data management and SaaS applications

With data growing rapidly and the increasing number of ways for managing this data, master data management (MDM) initiatives are continuing to grow in importance to the chief data officer or the chief information officer. Delivering a trusted, accurate, and complete view of critical data assets to business users and applications improves business results and operational efficiencies and also can help to lower costs and reduce risks. Using MDM, information governance can be enforced at an organization level, where business and IT can have a conversation and agree on important topics, such as data quality, which otherwise might not have been possible. Furthermore, business applications can be shielded from disruptive forces, such as big data and social data, by allowing MDM to build the *golden profile* of the customer and deliver this profile in context and on demand to users and applications.

As a general principle, MDM promotes *horizontal thinking*, where users from different line of business (LOBs) think about actions that can create an enterprise level impact, thereby benefitting the overall business. These factors and several other factors have contributed to the rapid growth in the worldwide MDM market. Thus, companies continue to invest in MDM technologies and solutions to create better business results and to improve cross-LOB functions.

SaaS applications, such as Salesforce.com, are business applications that are provisioned on the cloud and delivered as a service. These applications are typically hosted by independent software providers (ISP) or application service providers (ASP) and are normally based on a multi-tenant model. Installation and system management of the software is completely transparent to the user, because the provider manages all these aspects. As a result, the initial setup cost and provisioning cost is low. The software licensing is based on a subscription basis where the usage fee is typically charged either monthly or annually.

The industry has seen a rapid growth in the adoption of SaaS applications in the last few years. A leading-industry analyst reported recently that by 2015, more than 50% of organizations will have a SaaS strategy defined as part of their IT roadmap and application providers will support this strategy by enabling a majority of their software solution to be cloud ready.

1.2 The evolution of SaaS applications

The evolution of SaaS applications, and cloud adoption in general, have given birth to the notion of an *extended enterprise* that organizations now have to deal

with. Data and applications can now be *hybrid* because they can originate both inside and outside the enterprise. Business critical data assets that were traditionally created within the enterprise can now originate from upstream SaaS applications, and this can create the need for additional integrations with back-end systems. Though this scenario is not a new use case, and Salesforce customer relationship management (CRM) is a prime example of a SaaS application that companies have integrated with back-end applications. The difference now is the massive growth in the number of applications that are moving into the SaaS space, creating the need for additional integrations with enterprise applications.

Another aspect worth noting is that, by design, SaaS applications are relatively easier to obtain compared to the on-premises applications. SaaS applications can also be provisioned fairly quickly because clients do not have to procure the hardware or software or complete installation and configuration tasks. They can simply pay for the required subscription costs for the SaaS application, and the SaaS environment can be provisioned within minutes. Departmental heads and other LOB users find this model encouraging because it reduces the dependency on IT. They can get started on projects quickly and economically. As a result, organizations are seeing a rapid rise in the growth of SaaS applications and solutions that are sponsored by the business user community. Although this approach brings agility to the business, it also creates additional data silos in the extended part of the enterprise, which then need to be monitored and governed.

Hybrid or extended enterprises will become a reality in the near future because of the benefits that they provide to a business. The challenge here is to give business users a seamless experience when accessing applications and data, irrespective of where the applications or data reside. This process is similar to accessing the Internet and intranet sites that we are exposed to today. The only difference is that in the case of SaaS and on-premises applications, there must be a seamless data integration between the Internet site (that is, the SaaS application) and the intranet site (the on-premises application) or vice versa. Additional requirements in terms of application integration and process integration might apply in the form of workflows or business process management modules.

1.2.1 Business and technical impact

Organizations gearing up to embrace the SaaS revolution might experience business and technical impacts in following areas:

- *Application integration*: SaaS integration with on-premises applications will become a key requirement and will open important conversations on data security. Organizations will be required to implement the SaaS integration in such a way that it does not compromise corporate IT security guidelines.

- *Data integration*: Organization will need to integrate critical data assets, such as master data and reference data, between SaaS applications and on-premises applications. Implementing a seamless governance model to ensure data quality and polices are kept in check will be key.
- *Process integration*: End-to-end process integration for customer relationship management, support and service delivery, across all touch points will form a important factor. Business users and data stewards should be able to participate in workflows and other business processes that include data, rules, and processes that span both on-premises applications and SaaS applications.

1.2.2 Other considerations for SaaS applications in the context of master data management

Master data is the information infrastructure that connects applications, both inside and outside the enterprise (SaaS). So, when organizations employ one or more SaaS applications as part of their business strategy, it is important to understand how master data dependencies and extensions are performed in such situations. Consider the following questions when determining the optimal business outcome in these situations:

- How will data be mastered in an extended enterprise when the authoring is distributed? (That is, consider whether creation and maintenance of master data happens both inside and outside the enterprise.)
- How will data governance be implemented so that it covers the core enterprise and the extended enterprise (SaaS)? How will workflow and remediation tasks be implemented using the data from on-premises applications and SaaS applications?
- What will be the recommended strategy for integrating master data between SaaS applications (such as Salesforce.com) and an on-premises application (such as InfoSphere MDM) so that a 360-degree view of the customer can be provisioned for any or all customer interaction points? How can the cost of integration be optimized when there are multiple SaaS applications (both now and in the future)?
- How will risks related to data security be handled when integrating an on-premises MDM with an external SaaS application?
- How will a user enrich the master data in the SaaS application by drawing information from other sources (both on-premises and other SaaS applications)?

- How will a user view a consolidated, 360-degree of master data profiled at an organization level?
- How will data quality be measured and tracked? How can SaaS applications use the existing knowledge base and assets already built in the on-premises applications (for example, data quality rules)?

1.3 SaaS applications require data mastering and governance similar to on-premises applications

In the context of CRM applications, such as Salesforce.com, sales and marketing users depend on good quality customer information so that they can search for and enrich that information to support their daily business use requirements. So the key point here is that customer information is continuously added or changed and then searched not just by the users in the CRM application but also from the other business applications. When there are duplicates, they should be reconciled, merged, or linked with the other records. Business users and IT must have an agreement on the data quality levels that they will maintain to ensure smooth functioning of the business. Whenever these agreements are violated by poor data quality levels, policy violations are automatically triggered, and the appropriate governance body members are notified. Remediation is carried out in a workflow.

In the context of SaaS applications, creating master data, governance challenges, and remediation tasks should be managed properly. Duplicates should be controlled at the source by the implementation of the "search-before-create" policy, and the search technology should be powerful enough to handle all kinds of variations, thereby preventing duplicate creation at the source.

1.3.1 Managing duplicate records

When authoring is distributed, different users create data on multiple applications. This duplication increases the likelihood of error and poor data quality. Performing a large number of data stewardship and reconciliation tasks can create a negative offset on productivity and is best avoided. Poor data quality affects sales and marketing performance in several ways. In the context of data quality governance, ensuring that business and IT are in alignment with respect to data quality metrics, rules, and policies is critical for keeping the bottom line costs in check and for ensuring high productivity for business users.

In the case of CRM, research has shown that there is direct correlation between good data quality and good CRM performance. CRM users on Salesforce.com

face a common problem with other source applications in the form of duplicate record creation. A big contributor to this problem is limited built-in search capabilities that do not find duplicates when customer records have typographical errors, misspellings, and variations of name or addresses. Many CRM applications, such as Salesforce.com, imply a simple search and index mechanism that requires fairly deterministic search criteria and that operates only on its own records. Thus, it has no way to bring in other sources of data. For example, if a sales representative needs to enter information about a contact containing the term “Saint,” a search for existing records returns only exact matches to the supplied search terms. The sales representative has to search for all variances of the term (for example, “St,” “St.,” and “Saint”) to find all of the possible matches. This effort would take multiple searches to find the one correct and accurate record, although most users expect the search to be more like a Google search, where all relevant results are returned. As a result, the sales representative finds few or no matches, and the likelihood that the sales representative will create a duplicate record increases. The limited search and index capability also creates a scenario where sales representatives searching records to prepare for a customer meeting or call might view only part of the customer’s account and customer interaction history based on the particular version of the customer name used during the search. As a result, customer expectations might be missed due to a lack of complete information.

This drawback might also cause the sales user to create new records when in fact the record already exists. Duplicate records can frustrate users and can compromise the use of the CRM application where the data is perceived to not be trustworthy. Even if the CRM has good fuzzy search capabilities, there might be instances when the customer record resides in different source applications. So a user would ideally want a capability to do federated search that searches across applications but that still has fuzzy or probabilistic search capabilities that negates variations. Such requirements are seldom met by CRM applications and, thus, control on duplicate records is generally found lacking. CRMs such as Salesforce.com depend on an InfoSphere MDM solution to assist them with such requirements.

This same issue also manifests itself when large data updates are required, such as when the marketing department needs to load the results of a lead generation or marketing campaign. In this case, marketing administrators must manually import the data into Salesforce.com, check as much as possible for duplicates, and try to avoid creating a whole set of new ones. This process is a time-consuming exercise that significantly delays the time it takes to update data and the ability to gain the potential benefits from the campaign while also using up valuable marketing resources on menial tasks.

This lack of a complete, enriched picture of a contact and their account history has some unfortunate consequences. For example, when a sales representative

visits a customer without complete, accurate knowledge about other calls made and the overall account history, the customer is made to feel undervalued. It also creates the perception of a supplier who is not aware of what is going on. This same lack of information might also lead to poor management and governance. Poor quality information can result in failing to meet regulatory requirements in areas such as customer care, and will also reduce management's ability to manage and control operational activities and performance.

1.3.2 Adherence to data quality standards and governance policies

When authoring remains distributed, it is important that users are guided by the application to capture data to the acceptable standards laid down by the organization. For example, a company might have data quality standards in the way they capture the details of a customer. Name, address, telephone number, phone number, fax, email address, and so on might require certain standardizations and normalizations as part of the company policy. Therefore, when implementing a SaaS application that serves as a source application for accounts and contacts, such as Salesforce.com, it is imperative that the same data quality rules are followed here for consistency.

In such cases, companies have the following choices for implementation:

- Implement the data quality policies (again) in Salesforce.com and replicate or rebuild all the rules.
- Use the knowledge base, data quality rules, and standards already incorporated in the enterprise InfoSphere MDM system, and use these rules during the account creation process in Salesforce.com.

The second option is the better choice, because all policies and rules are enforced by the master data application. In addition, similar to the other applications in the enterprise, Salesforce.com looks up to InfoSphere MDM for guidance on data quality standards and policies. There is no need to replicate these rules in every application (including Salesforce.com), thereby enabling consistency across the board. Those rules that are specific to Salesforce.com can be incorporated in Salesforce.com itself.

The other advantage in following this methodology is that it enables an organization to measure the data quality metrics and measures and allows the organization to compare these measures with the standard set by InfoSphere MDM. For example, if the data governance committee set benchmarks to capture completeness of records in terms of a percentage and if there are indicators for the lowest acceptable thresholds as part of the company governance policy, allowing source applications such as Salesforce.com to look up InfoSphere MDM for data quality standards enables you to measure, analyze, and report the

deviations or adherence, whichever the case might be, as part of the monthly, quarterly, and annual data governance reporting.

1.4 Challenges in overcoming business and operational issues from within the SaaS applications using a native MDM or internal MDM

MDM is more about *managing* the data than the data itself. So the depth of capabilities that help manage the data is important in determining the quality of the information. Industry-leading MDM offerings, such as InfoSphere MDM, have best of breed capabilities natively built in to support such. Some of these key capabilities are of paramount importance in context to SaaS applications, such as Salesforce.com:

- Real-time probabilistic search and match to drive use cases, such as search before create
- Robust business process manage (BPM) and workflow engine for remediation and stewardship
- Policy enforcement and monitoring for data governance inside and outside the enterprise
- Ability to analyze, measure, and perform data quality procedures
- Real-time and near real-time integration to meet transactional requirements and batch integration

Some SaaS applications might not have these capabilities or might lack the required depth to support life cycle management of accounts, contacts, and other critical information.

1.4.1 Incomplete information

To a user on Salesforce or a another SaaS application, getting access to a 360-degree view of customer information that is complete and up to date in all aspects is important. Today, CRM applications are dependent on MDM solutions to help them get access to this 360-degree view. Let us analyze a few use cases and understand why this is the case.

Customers today interact with an organization using several touch points and channels, as opposed to previously when there were only a few dominant channels, such as face to face, email, phone, or fax. In addition, customers expect organizations to have a seamless integration between these touch points

so that any update they make on one touch point is reflected on all the others. This expectation is forcing organizations to deliver an omni-channel experience to customers where they channel all client interactions to a single system, usually an MDM solution, and have this system as a point of a reference for all other channels (such as mobile, web, kiosk, and so on).

SaaS applications cannot do this channeling on their own because it requires them to do multi channel integration to gain insight. Add to this challenge disruptive forces, such as social media and mobile apps, and you realize that CRM applications are not designed to handle this level of complexity. And if there is more than one CRM (SaaS or on-premise) implemented within the enterprise, it will require the data to be consolidated, matched, and linked for single-view generation and consumption just within the CRM boundary space.

All these use cases and problem statements can be well handled by MDM, and CRM applications usually rely on an MDM system for information completeness and accuracy.

1.4.2 Consistent de-duplication strategy and results

Duplication of records is a common phenomenon observed in most applications. Applications typically get around this problem by applying *fuzzy logic* or probabilistic matching, as part of the “search-before-create” implementation.

However, just like how all duplicates are not created equal, the accuracy of all de-duplication engines is not the same either. Basically, the following types of problems must be managed:

- *Match accuracy*: What one application considers a duplicate might not be considered as a duplicate in another application.

  Probabilistic matching engines show variance in accuracy and results, because they all are coded differently and because they employ different techniques. The accuracy of the results depends on the depth of the matching capability and not all applications have the same depth. So consistency in matching and de-duplication is generally important.

- *Duplicates at source level versus duplicates at the enterprise level*: A record not considered a duplicate in one source application might already be present in another source application.

  So, during record creation, the first source application should ideally have the ability to check if the record exists anywhere in the enterprise. If it exists, and if it is determined that both records are referring to the same customer, records should be matched and linked to one entity record or physical record so that lifecycle management can be done at this entity level. This process ensures a 360-degree view of customer information at an enterprise level.

1.4.3 Data augmentation through business hierarchies

Different applications use different hierarchies to classify and categorize customer information. In most cases, a mapping between hierarchies exists, and these mappings are used by downstream business intelligence applications for aggregations and reporting. For example, one application can classify customers based on geography, and another application can segregate customers based on segmentation. A customer can be mapped to both hierarchies, and it is often useful for the business or a business intelligence (BI) application to determine the catchment concentration of their segmented customers so that they can do targeted marketing, which can be more effective.

It is perfectly normal that, with changing business requirements, these hierarchies undergo changes where additional nodes and levels are added. This change creates the need for adding and updating the mappings. So lifecycle management of hierarchies and relationship is another important area that must be addressed.

A SaaS application has challenges in maintaining business hierarchies on its own given this complexity, whereas an MDM solution, such as InfoSphere MDM, can help manage the different hierarchies and their mappings in one location. Further, an MDM solution can help manage the lifecycle of customer classification hierarchies by applying governance frameworks whenever there are changes to these hierarchies and also by ensuring that the updated definitions are synchronized with what downstream systems need.

1.4.4 Redundant enforcement mechanisms that hinder enterprise-wide governance initiatives

Master information is a shared entity that is owned at an “organization” level rather than at an “application” level. The definition, policies, and rules related to how customer information or other master information is to be captured and maintained is defined by an Information Governance Council that is set up at the highest levels of the organization. This committee usually is comprised of the department heads from both business and IT and is usually led the chief data officer or the chief information officer. In addition, the Information Governance Council also ensures that monthly or quarterly data quality standards are met and that business and IT are in agreement and have committed to meeting these standards. Some organizations treat the Information Governance Council so seriously that they incorporate a percentage of the salary pay for committee members depending on how well compliance was achieved in the previous month or quarter.

MDM solutions, such as InfoSphere MDM, drive governance in the organization. When the business has to integrate with one or more SaaS applications, it can be a challenge to embed or inject the governance policies and data quality rules in every application. Therefore, each SaaS application should look to MDM solutions, especially InfoSphere MDM, to help understand whether the information being captured is in-line with the standards and governance policies as agreed by the organization.

1.5 Other integration challenges

This section describes other integration challenges that must be addressed when integrating an on-premises MDM system with an external SaaS application.

1.5.1 Cross border

Cloud vendors typically have several data centers in different countries, and SaaS applications vendors host their applications in one ore more of these data centers. Organizations looking to integrate SaaS applications should carefully analyze information privacy acts, laws, and regulations that apply when personal information about customers are stored, maintained, or managed outside the country where the customer resides.

Certain countries and regions, such as the European Union, have strict regulations about storing personal data in countries outside the union. There are published laws and regulations that state that citizen information should not be stored or maintained outside the country of residence. Thus, in theory, the transfer of personal information from the European Union to the US is prohibited when equivalent privacy protection is not in place in the US. American companies that would work with EU data must comply with the Safe Harbour framework.

Even in the US, there are laws about information exchange and privacy. The Health Insurance Portability and Accountability Act (HIPAA) act, for example, talks about patient's privacy rights when dealing with how and when personal health information can be shared.

Companies that implement SaaS applications, especially those related to CRM and other customer facing initiatives, need to be aware of these regulations and should implement solutions accordingly. One plausible way to handle this challenge is to implement an MDM solution, such as InfoSphere MDM, and use the MDM solution to power the external SaaS application. That is, the data always resides in the "country of residence" MDM solution SaaS application, but it queries for all create, read, update, and delete (CRUD) services. The SaaS

application is implemented on a “no data persistence” model, and it invokes the services from an MDM solution when dealing with use cases that are related to customer data creation, search, and update.

1.5.2 Data integration

Based on what we have discussed so far, we have established the premise that master data integration between on-premises enterprise applications, such as InfoSphere MDM, and external SaaS applications, such as Salesforce.com, is an important capability. You can use the following methods to enable data integration in an extended enterprise:

- Real-time data integration
- Near real-time data integration
- Bulk data integration

The real-time and near real-time data integration capabilities enable use cases that require real-time transaction services invocation between SaaS applications and an on-premises MDM solution. Examples for such use cases include:

- “Search-before-create”
- View account details, including contact information, interaction history
- Publish account details from SaaS applications to the on-premises MDM system
- Create, update, or delete actions
- Event notification and auditing
- 360-degree view of account information, organization hierarchies, and relationships

Capabilities related to bulk data integration and movement support use cases where bulk load of data or mass updates have to be performed. Examples for such use cases include:

- Initial load of data that involves extraction of data from an on-premises application, such as InfoSphere MDM, and loading of this data into a SaaS application, such as Salesforce.com
- Mass de-duplication of marketing leads generated from a campaign program
- Bulk update of data in SaaS application, which must be propagated to on-premises MDM system

1.5.3 Security

Security concerns and their associated risks continue to be the biggest inhibitor for companies when it comes to cloud computing. Although security challenges in cloud computing realms can be a vast topic for discussion and can include several sub topics, such as application security, user security, platform security, and so on, we limit the scope of discussion in this book to *data* security.

Challenges related to data compromise, data hacking, and data loss can pose a serious security concern whenever data needs to be integrated between on-premises applications and SaaS applications. For example, a major concern arises when organizations implement point-to-point integrations using web services or Representational State Transfer (REST), which need special inbound access to be allowed in the company's firewall. This point-to-point integration is done to allow the external SaaS applications to make service calls with the on-premises applications and for integrating data. This type of access exposes a major security flaw that puts the company's entire IT infrastructure at a wide open compromise and risk. So organizations need to evaluate and implement integration solutions that offer secure access to their on-premises application layer and data.

The other challenge related to data security is data theft. Due to lack of control on what goes on in the data center where the SaaS applications are hosted, it opens a security concern where someone (in the data center) can easily steal sensitive customer information related to a customer, SSNs, and other critical information. So companies are required to review the security policy of the vendor to ensure that the possibility of data theft occurring is virtually non existent.

Several software vendors today offer specialized software solutions that allow a seamless, secure integration between on-premises applications (such as InfoSphere MDM) and SaaS application (such as Salesforce.com). These solutions have made significant progress in recent years in the area related to data security and have alleviated the security concerns to a large extent. For example, cloud integration solutions, such as IBM WebSphere Cast Iron Cloud Integration, offers a capability that performs SSH tunneling in the firewall for all service interactions. This function ensures that only trusted SaaS applications with valid security tokens and certificates can access the company's application and data.

1.5.4 Business process integration

Process integration is an integral part of master data creation and maintenance procedures because customer data creation can be on a distributed authoring

topology. For example, there might be several source applications in the enterprise where customer information is created, but this customer information needs to funnel down to the MDM systems for the ultimate system of record and for the enterprise 360-degree view. With time, whenever these customer records undergo updates or remediation, it is best done in a workflow process that has the required process orchestration, data steward task assignments and notifications, and so on.

When integrating an external SaaS source application, such as Salesforce.com, where customer information or account information is also created, it is important to ensure that the external application can effectively participate in the workflow processes as defined by the business. This requires good service integration between the MDM application, the workflow engine, and the SaaS application to ensure that data in the System of Record is clean and up to date.

The following approaches are effective for implementing business process integration with Salesforce.com using service integration:

- Implement a custom point-to-point integration between the MDM system, the workflow engine, and Salesforce.
- Use a cloud integration product, such as WebSphere Cast Iron Cloud Integration, for service integration and orchestration.

When integrating, use a cloud integration offering as the extended enterprise connector and enterprise application integration (EAI). This approach is secure and cost efficient (zero coding approach). In addition, it allows you to support future integrations by mapping the MDM services once but by also integrating these services with any SaaS application anytime in the future.

1.6 Setting the stage for MDM integration with SaaS applications

Cloud computing (including SaaS applications) can bring about significant innovation to both business and IT operations. However, it can also bring disruptive changes that must be addressed well for the innovation to be effective. For example, so far, the bulk of the master data originated from within the enterprise and was fed to the business applications that were also within the enterprise. However, this process is changing rapidly with the emergence of SaaS applications and hybrid topologies, which have given rise to a new generation of integration requirements. Further, we are also seeing a shift in the way that master data is created and integrated when compared to the traditional methods. For example, the bulk of the core master information that typically originated inside the enterprise is now originating from outside the enterprise, in

the form of big data, social data, and so on. The majority of the business applications (such as CRM applications, eCommerce applications, and so on) that resided within the enterprise and that either generated or consumed master information, are now residing outside the enterprise in the form of SaaS applications. This is slowly but surely shifting the integration epicenter from "enterprise only" applications to that of an "extended enterprise" that encapsulates enterprise applications and SaaS applications in one integrated operating environment.

When contemplating MDM integration with SaaS applications, consider carefully the following technical requirements and solution aspects:

- *Security*: Anytime a business critical application, such as an MDM solution, is wired to an application that practically resides outside the company's network, data security automatically becomes an important requirement. The integration component that facilitates the integration should not require any "hot wiring" that bypasses the fundamental IT security guidelines of the company.
- *Master Data Governance* (both inside and outside the enterprise): Master information can potentially originate in an SaaS application. A good example of this is Salesforce.com, where customer accounts and contacts get created. The ability for the MDM system to drive and enforce master data governance in these applications assumes an important role. Otherwise, issues related to substandard information can pose serious challenges that can negatively impact the business operations throughout the enterprise.
- *Integration*: Given the rapid rise in the emergence of SaaS applications, organization must ensure that their integration platform has the capability to quickly "wire and weld" any SaaS application and their on-premises MDM solution. This is critical to stay agile in an ever-changing business environment and also to meet competitive threats. In addition, it is also important that a multi dimensional integration approach is implemented that facilitates integration of data, application, and processes between the MDM system and SaaS applications.

1.6.1 High-level solution construct and logical architecture

IBM InfoSphere MDM and IBM WebSphere Cast Iron Cloud Integration are well positioned to meet the requirements described in 1.3, "SaaS applications require data mastering and governance similar to on-premises applications" on page 5, 1.4, "Challenges in overcoming business and operational issues from within the SaaS applications using a native MDM or internal MDM" on page 8, and 1.5, "Other integration challenges" on page 11.

1.6.2 IBM InfoSphere MDM

IBM InfoSphere MDM creates trusted views and elevates the effectiveness of your most important business processes and applications. It can help you to improve business results, lower costs, reduce risk, and enable strategic agility. InfoSphere MDM supports multiple MDM styles and domains to work together to provide a comprehensive set of MDM capabilities (as illustrated in Figure 1-1).

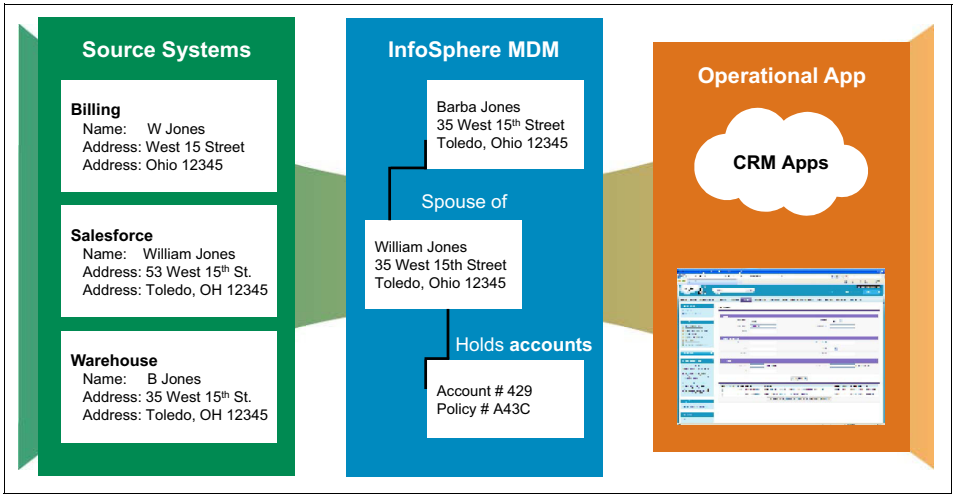


Figure 1-1 InfoSphere MDM provisioning a single view of customer

Some of the key capabilities of InfoSphere MDM include:

- Virtual, physical, or hybrid master repository
- Workflow capabilities to implement policies and processes for data stewardship and data governance
- Pre-built and extensible data models and business services
- Highly accurate probabilistic matching and search
- Policy management to author, monitor and enforce data quality
- Solutions framework to enable and enhance applications and processes with master data
- Data stewardship to inspect and resolve data quality issues
- Relationship and hierarchy management to manage household and B2B relationships
- Integration with multiple batch and real-time options

1.6.3 WebSphere Cast Iron Cloud Integration

IBM WebSphere Cast Iron Cloud Integration products can enable you to help integrate cloud and on-premises applications in days, to help reduce integration costs, and to help optimize resources and productivity in SaaS and cloud models. They provide a graphical configuration approach—rather than custom coding, on-demand tooling, or traditional middleware—to help you integrate applications quickly and simply. They use pre-configured templates based on common integration scenarios to accelerate integration.

WebSphere Cast Iron Cloud Integration products provide several capabilities for near real-time integration, data cleansing and migration, data synchronization and connectivity, and workflow and transformation that enable you to orchestrate integration processes across multiple applications. Mash up capabilities enable you to integrate information from disparate sources and to display it using the native user interface of a cloud application. WebSphere Cast Iron Cloud Integration products also support mobile applications by harnessing data and processes from other parts of the enterprise (as illustrated in Figure 1-2).

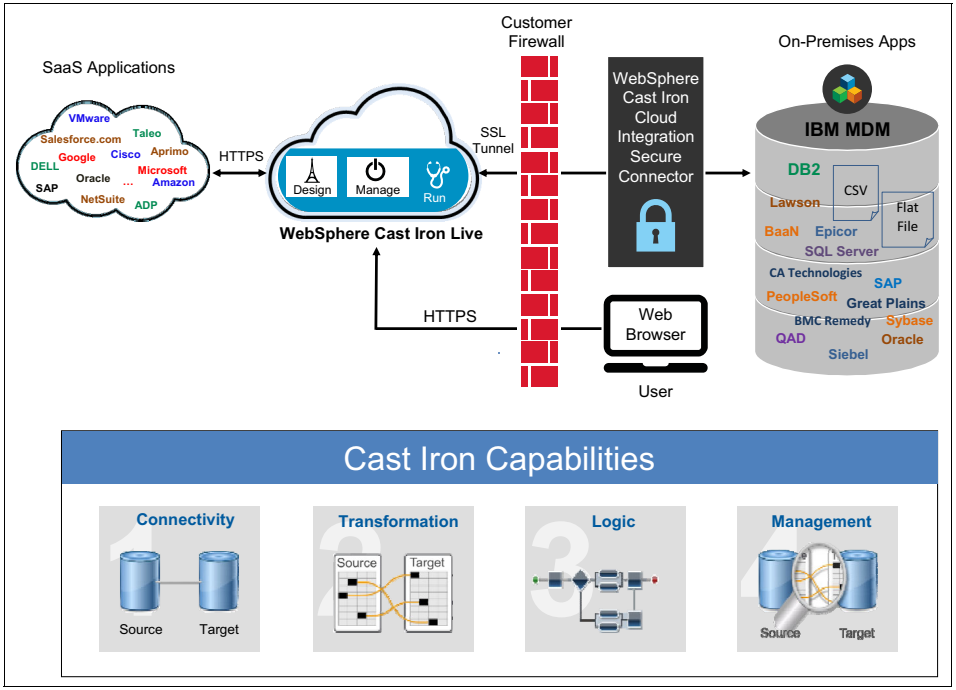


Figure 1-2 WebSphere Cast Iron Cloud Integration capabilities

The deployment options include:

- *WebSphere Cast Iron Express* is cloud-based software that helps you integrate information from SaaS applications, such as Salesforce.com, with other data sources.
- *WebSphere Cast Iron Hypervisor Edition* is a virtual appliance that can be installed on existing servers through virtualization technology.
- *WebSphere Cast Iron Live* is a multi-tenant, cloud-based platform for integrating cloud and on-premises applications and enterprise systems in a hybrid environment.

1.6.4 Logical architecture of the InfoSphere MDM and WebSphere Cast Iron Cloud Integration

Figure 1-3 on page 19 shows the logical integration architecture and positioning of InfoSphere MDM and WebSphere Cast Iron Live. InfoSphere MDM, as a mission critical application, is usually set up behind the firewall and in a network that has no inbound or outbound access. WebSphere Cast Iron Live resides outside the firewall and helps broker the calls between the MDM system and external SaaS applications. The WebSphere Cast Iron Cloud Integration Secure Connector ensures that communications between the MDM system and the SaaS applications are secured.

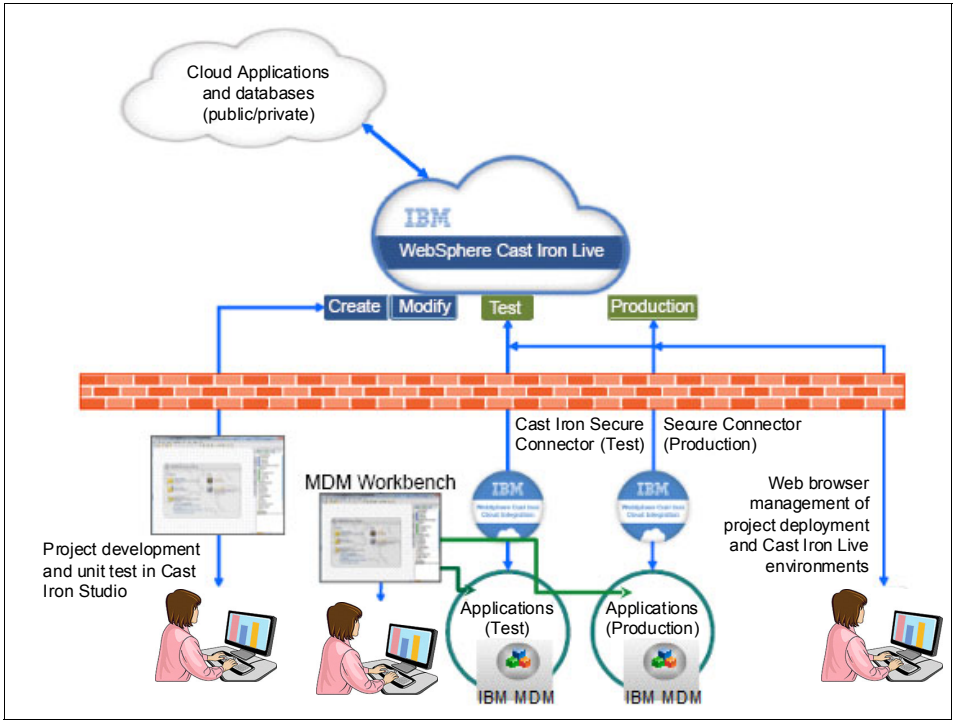


Figure 1-3 Logical architecture

2

MDM solution readiness to participate in integrations with SaaS applications

This chapter describes various integration strategies, relevant standards, and a other characteristics that on-premises applications generally use. It also describes integrations and master data management (MDM) solutions typically participates in. This chapter discusses the requirements for integrating on-premises MDM solutions with SaaS applications. It wraps up by outlining different facets of solution definition when it comes to integrating on-premises MDM solutions with SaaS applications.

This chapter includes the following topics:

- Integration strategies employed regularly by on-premises applications
- Integrations with IBM InfoSphere Master Data Management solutions
- General requirements for integrations of on-premises MDM solutions with SaaS applications
- Aspects of solution definition

2.1 Integration strategies employed regularly by on-premises applications

Enterprise applications deployed on premises are integrated for various reasons. The key reason is to draw synergies on the capabilities offered by different applications that have been purchased or built by these enterprises. Historically, point-to-point integration came in being with applications building independent connectors to support inter-application connectivity and processing hand off. Over time, it was seen that complex integration scenarios were not sustainable with this approach had high maintainability and ownership costs.

Point-to-point integration approaches have given way to *enterprise application integration* (EAI) architectures. EAI architectures predominantly governed how applications deployed on premises interacted with other, the standards they employed, and other non-functional considerations. Enterprises collaborating with other enterprises also employed integration practices under the umbrella of EAI to enable business-to-business (B2B) integration. EAI architectures helped enterprises enable the integration of people, processes, and information at departmental levels and was easily extended across enterprise boundaries.

Early EAI solutions used a sort of *hub and spoke* model to centralize integrations and typically dealt with message brokering from a centrally deployed application. These solutions used to have adapters, data transformation capabilities, orchestration capabilities, stewardship interfaces, and other complex functions to provide a single integration solution for the enterprises. EAI solutions paved the way for better consideration to terms such as loose coupling, made good strides in helping standards around reliable messaging, laid the foundation for connector architecture standards and a few others, however. During this time that EAI was evolving, the B2B integrations saw large companies mandating methods of receiving business information technology with the widespread adoption of electronic data interchange (EDI) standard.

However, given the solution was single point of inter-application messaging and vendor specific nuances were managed in the adapters, this construct fell under its own weight because it was not able to support transactionality and became a single point of failure with no proper standardized failover techniques.

The enterprise service bus (ESB) was an attempt to resolve hub and spoke related issues. ESBs did resolve some of the issues around the EAI solution but also brought to the surface service-oriented architecture (SOA). ESBs served well in real-time asynchronous integrations; however, real-time synchronous integrations needed better capabilities in integration service composability, reliability, transaction processing, and many others.

You can find more resource materials about SOA at:

http://www-01.ibm.com/software/solutions/soa/

The following summary provides a roundup of what has been written about SOA by various IBM experts:

- SOA provides a disciplined approach to integration that is essential for the enterprise integration architecture. SOA laid down the path for building modular composite applications, laying down high importance on reusable service components using open and available standards. The notion was loosely coupling was taken a step further and component architectures become more powerful as they provided the ability for components to act as service consumers and providers by exposing interfaces in a standard format across a distributed topology.

  This improved flexibility of IT systems and provided the following benefits:

  - SOA based integration approaches used open standards to deliver integration assets as services, helped reuse of existing assets and avoided proprietary vendor lock-in issues.
  - SOA provided a standard to represent and interact with integration components thus providing flexibility to reuse assets and allowing multiple business processes to seamlessly interact with each other.
  - SOA shifted the focus of integration solution assemblies rather than implementation details of each function which laid the foundations for Composite applications which ultimately support building a “composable” business.

- SOA was partially fueled by the prevalence of the World Wide Web and the HTTP protocol, which in turn helped the creation of remote procedure call (RPC) style and SOAP style web services, Although RPC style web services were more command-oriented, SOAP style was XML-document-oriented. Because of the rigid nature of SOAP-style web services, Representational State Transfer (REST) is an architectural style that can be used as a web service and has gained momentum over the years because it does not have a lock on to the XML payload. Talking about payloads, over time, REST style services with the “schema-less” JSON format has gained prominence. Enterprises continue to use REST style services and web services in the current implementation, and the cloud computing era has brought in web APIs into the landscape of the applications composer.

In summary, enterprises typically use reliable messaging or real-time asynchronous integrations, services for real-time synchronous integrations, and some ETL solution for bulk data operations.

2.1.1 Role of open standards

The role of open standards throughout the history of application integration has been evolving. Right from EDI all the way to SOA, the IT industry has made significant progress in technology and architecture standards. While the focus of technology standards has been on middleware products to enable interoperability and portability, SOA architecture standards have evolved to provide guidance to practitioners as they design and deploy SOA solutions. These techniques are agreed upon with industry validation, and the methodologies are vendor neutral with standard terminologies and taxonomies.

Integrating on-premises applications, in this case an MDM solution and SaaS applications, needs a good understanding of standards in play for cloud computing, regular standards for on-premises enterprise applications, and the standards applicable to the middleware technologies in use.

The Open Group

The Open Group is the custodian of the standards around SOA and primarily deals with following standards and guides:

- The Open Group Service Integration Maturity Model (OSIMM)
- Service-Oriented Architecture Ontology
- SOA Governance Framework
- SOA Reference Architecture

The Open Group as also covered Service-Oriented Cloud Computing Infrastructure (SOCCI) Framework and more recently, SOA and Open Platform 3.0, enabling Big Data, which deals with big data processing, based on SOA.

SOCCI does cover aspects of Service-Oriented Infrastructure (SOI) and SOA reference architecture (RA). This standards document also covers the aspects of the services supply chain. However, the coverage of hybrid IT architectures and how on-premises services providers and cloud solutions such as SaaS applications work with bidirectional integration does not seem to be covered holistically.

For more information about Open Group, see:

http://www.opengroup.org/standards/soa

Cloud Computing Reference Architecture

IBM hosts the Cloud Computing Reference Architecture (CCRA) at the following developerWorks wiki:

http://ibm.co/1BRM1fF

IBM also made an Open Group submission of the CCRA at the following website. A login ID for Open Group site is required to see this document.

http://bit.ly/ZR2sOk

Figure 2-1 represents the schematic diagram of CCRA because it is a representations of the aggregate experience from hundreds of cloud client engagements and IBM-hosted cloud implementations. It provides prescriptive guidance about how to build infrastructure as a service (IaaS), platform as a service (PaaS), SaaS, and service provider clouds using IBM technologies.

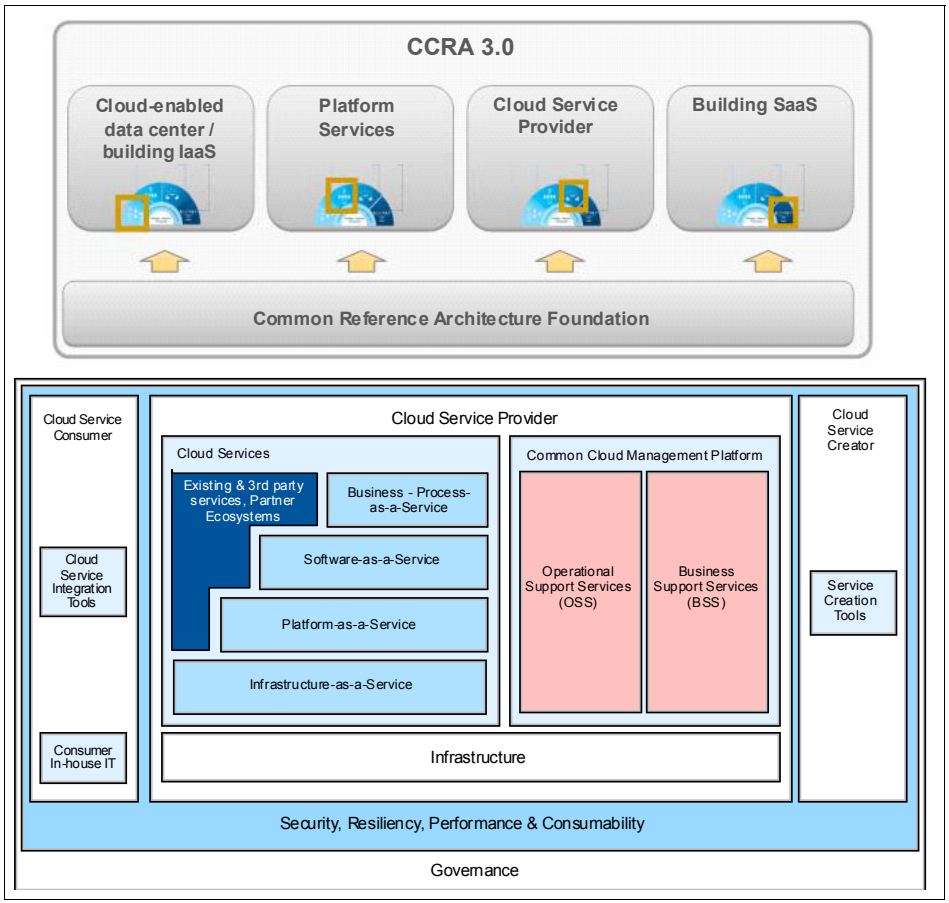


Figure 2-1 IBM general cloud reference architecture

Other parties have created cloud reference architectures with varying levels of quality and detail, for example:

- VMware vCloud Architecture Toolkit (vCAT)

  http://www.vmware.com/cloud-computing/cloud-architecture/vcat-toolkit3.html

- VMware vCloud Reference Architecture

  http://www.vmware.com/cloud-computing/cloud-architecture/vcat-toolkit.html

- HP CloudSystem Reference Architecture

  http://h18006.www1.hp.com/storage/pdfs/4AA3-4548ENW.pdf

- Cisco Virtualized Multiservice Data Center Reference Architecture

  http://www.cisco.com/c/en/us/solutions/collateral/enterprise/data-center-designs-cloud-computing/solution_overview_c22-714480.html

- Rackspace Open Cloud Reference Architecture

  http://www.rackspace.com/knowledge_center/article/rackspace-open-cloud-reference-architecture

- Amazon AWS Reference Architecture

  http://aws.amazon.com/architecture/

- Microsoft TechNet - Private Cloud Reference Architecture

  http://social.technet.microsoft.com/wiki/contents/articles/3819.reference-architecture-for-private-cloud.aspx

- Oracle Cloud Reference Architecture

  http://www.oracle.com/technetwork/architect/archday-rws-2013-krishnaswamy-1966514.pdf

A broad view on cloud-related reference architectures is important because it helps evolve the solution architectures and parallel and because it helps enterprises upgrade the policies around service level agreements (SLAs), security, interaction contracts with partners, and many other areas. Standards are updated from time to time, and users, practitioners, and decision makers should keep themselves updated, especially if the focus is on integrating on-premises applications with SaaS applications. It worth to see some trends at this point on how the cloud computing is playing on the topology decisions of the enterprise.

Figure 2-2 illustrates how on-premises resources and cloud-hosted resources are jointly used by the “extended enterprise.”

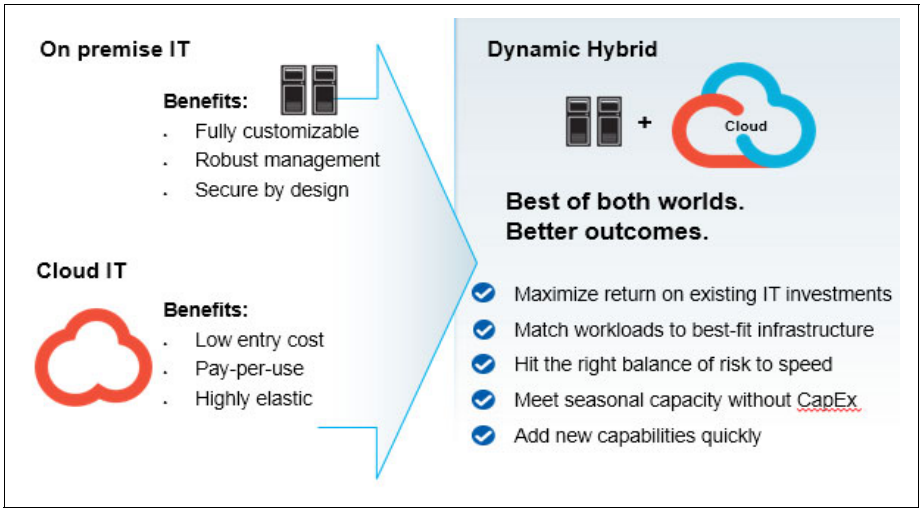


Figure 2-2 Dynamic hybrid topologies

The future will be a dynamic hybrid because it combines the best of both worlds and allows for better outcomes. For companies with large IT investments, the hybrid approach allows you to maximize its return on existing IT investments by transforming IT into self-service private clouds. The hybrid approach also allows enterprises to keep some workloads on a best fit infrastructure and move other workloads to public cloud for better economics. The hybrid approach lets you keep critical, sensitive data in house, while innovating on the public domain so that you have the correct balance of risk to speed. For example, you can connect ERP to public CRM to enhance client experience. The hybrid approach also allows enterprises to extend current IT with private clouds with public capabilities to meet seasonal capacity requirements (for example, holiday shopping), without the need for additional capital expenditure. Finally, the hybrid approach allows enterprises to set up development environments quickly for innovation, to add new capabilities, such as analytics or mobile SaaS, on the public domain quickly, and to compose more competitive business services and processes.

Typical facets of application integration

Applications are integrated to draw synergies, and typically this integration is done because applications have specializations. They either must interact with other applications, get data from other applications, or pass data to other applications to participate in this synergy. Integration of two applications does indeed have many facets. Integration begins with the business reasons on why the applications need to be integrated. The case needs to be compelling enough

to make the investment in doing the actual integration and putting the plan into production or operational use. Then, the considerations delve into functional aspects on what each application does and what kind on integration is done between the applications.

Keep in mind the following considerations:

- Does one application do some work and just post data to the other applications and vice versa? [data integration patterns]
- Does one application need interfaces of the other application natively invoked in the other? [SOA patterns]
- Is there a higher-level business process that needs these two applications to participate in? [business process manage (BPM) patterns]

After getting past the functional considerations, consider next the following technical specifications:

- What existing end points are provided by each application and are they worth using?
- What new end points are needed?
- What deployment topologies do each of these application follow?
- Do we need to synchronize data between these applications before the actual integration kicks in?
- Do these applications need a common user registry and authentication services?
- Do we need to have the authorization for these applications stored in a common repository?

After the technical considerations, the non-functional integration aspects, such as the SLAs for integrations with various types of end points, performance, high availability, failover and others, come into play.

Even though the discussion in this section might seem narrow and fairly indicative, they are some of the key facets of any application integration task.

2.1.2 Integration strategies for consumers and services providers outside the enterprise firewall

Enterprises today understand the value proposition of cloud platforms. Multi-platform architectures have been adopted rapidly in recent years as IT architectures span premises and host environments (sometime referred to as hybrid clouds). Giving the rise of *extended enterprises* whose IT footprint extend

into the clouds, enterprises wants to also extend on-premises application functions to hosted applications that serve their operational needs.

IBM cloud reference architecture addresses the following major roles in any cloud computing environment (as illustrated in Figure 2-3):

- Cloud service provider
- Cloud service creator
- Cloud service consumer

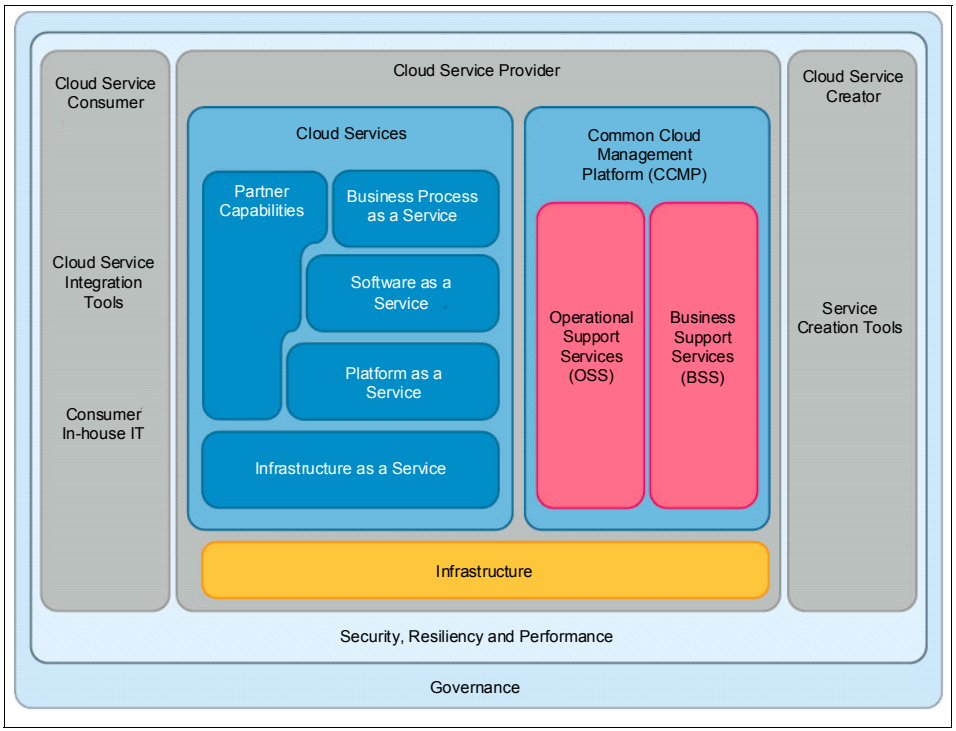


Figure 2-3 IBM cloud reference architecture core

In the content of on-premises MDM solution integrating with SaaS applications, the key area of focus is around how in-house IT consumers work with cloud services and how cloud services use services hosted in house, that is on premises. Enterprises do not expose ports on the Internet for interactions with mission-critical applications because of various security considerations. This is where the use of a cloud integration middleware becomes a requirement. The characteristics for the cloud integration middleware are as follows:

- Integrate applications from cloud-to-cloud and cloud-to-on premises primarily
- Must provide a rapid and easy-to-use method of setting up the integrations

- Should be available in multiple forms factors to align with the adoption cycle of cloud integration and on-premises priorities
- Secure connectivity support allowing the secure transfer of data from behind the firewall
- Single unified monitoring solution for multiple environments
- Broad set of connectivity options

2.2 Integrations with IBM InfoSphere Master Data Management solutions

IBM InfoSphere Master Data Management (InfoSphere MDM) solutions typically function as a repository that can centralize and manage an organization's key information assets, such as customer, product, supplier, and more. This solution provides the centralization that is the provenance of a single view of customers and products that results in other systems needing to integrate to contribute or use it. Potentially, these integrations result in better services, improved customer satisfaction, and improved relationships with partners and suppliers.

The topics in this section discuss the relevance of key integrations of InfoSphere MDM solutions.

2.2.1 Integrations with source systems

Every InfoSphere MDM implementation integrates with a set of source systems to bring data into the MDM solution in many different ways, for example:

- Source systems provide a one-time initial feed to the MDM solution in preparation for retirement.
- Sources systems provide a one-time initial feed for the go live of the MDM solution and continue to provide delta feeds.
- Source systems possibly post only events to the MDM solution when MDM solutions are built to "pull" data.
- There are write backs from the MDM solution back to the source systems.
- Source systems provide a one time initial feed and then stop owning the attributes of the master data entities and let the MDM solution manage them while they transition into becoming consumers of the MDM solution.
- Hybrid cloud topologies and the notion of extended enterprises brings in hosted or SaaS applications into considerations of being a source.

2.2.2 Integrations with business applications

Business applications might have many roles around integration in InfoSphere MDM projects. Integration of InfoSphere MDM solutions with business applications have the following key considerations around roles:

- Business applications acting as a source
- Business applications acting as a consumer

The facets of business applications acting as a source were outlined in 2.2.1, “Integrations with source systems” on page 30.

Business applications acting as a consumer are the ones who use the *trusted single views* provided by InfoSphere MDM solutions to begin with. They also use other aspects of governance and search capabilities. For example, business applications managing customer on-boarding solutions could use MDM services for customer creation processes, data validations, and so on. The MDM services could also provides probabilistic search services so that the customer on-boarding process would stop duplication at the grass root process itself.

Business applications today, as its being discussed, are not limited to deployments in the enterprise owned data centers. Businesses now understand the value of cloud computing, and the adoption of usage of SaaS applications has been ever growing for solutions needed for operations such as CRM, order management, human resource management, employee education, and so on. These kind of applications fall into the key considerations around the roles outlined.

In the context of on-premises MDM solutions integrating with SaaS applications, SaaS applications acting as a source and SaaS applications acting as a consumer are certainly key considerations.

2.2.3 Integrations with warehouses and other analytical environments

MDM integration with warehouses and other analytical environments is of paramount importance as these endpoint systems do need trusted master dimensions for trusted business insights. Analytics solutions spectrum is quite broad with the progressions in technology, computing environments and the form factors of computing available.

Traditionally speaking, MDM solutions feed data into warehouses and data marts for analytics to be done by business intelligence tools such as IBM Cognos®; this is what we can call *classic analytics solutions*. There are ETL tools in play that participate in data distribution strategies which ultimately ensure that the

warehouses and the data marts have the MDM data for enabling business intelligent (BI) tools to perform analytics. In some cases, the insights received from the classic analytics solutions can provide feeds back to MDM. For example, insights could be hierarchies for customer segmentation.

If we start looking further, considerations for the new class of analytical platforms is gaining prominence. There are as follows:

- *Cognitive analytics solutions*: These are solutions driven by machine learning and other cognitive elements that learn *dynamically* and provide insights. IBM Watson™ System is one example.
- *Prescriptive analytics solutions*: These are analytics solutions which deliver prescriptive outputs targeted towards *best outcomes*. IBM Next Best Action Signature Solution can be considered as one example.
- *Predictive analytics solutions*: These solutions are focused on “What could Happen?” type questions and are targeted towards helping enterprises make smarter decisions. IBM SPSS® Predictive Analytics software is a market leading solution in this area.
- *Descriptive Analytics*: These solutions are focussed on “What Has Happened?” type questions and are targeted towards enterprises doing compliance and investigational analytics.
- *Exploratory Analytics*: Analytical Solutions supporting exploration and discovery are gaining prominence. These solutions are more directed towards situational analytics focusing on “What Do You Have?” In fact, “exploration and discovery” is the primary use case for IBM BigInsights™ Platforms.

Looking still further, we are now seeing that lot of these analytics solutions that require feeds from MDM solutions that are now hosted on the cloud. The spectrum gets still broader with many of these cloud analytics providers coming in with X dollar per hour pricing all the way to niche specialized analytics capabilities that are relevant to enterprises.

In the context of on-premises MDM solutions integrating with SaaS applications, cloud-based analytics solution offering acting as a consumer is certainly a key consideration.

2.3 General requirements for integrations of on-premises MDM solutions with SaaS applications

Before we can completely understand the requirements for integrating on-premises MDM solutions with SaaS applications, it is worth understanding the different ways in which governance capabilities are provided to SaaS

applications. The following approaches are relevant for powering SaaS applications with governance capabilities:

- Embedded or dedicated governance capabilities
- Federated governance capabilities
- Integrated governance capabilities

There are SaaS applications that provide their own governance solutions that are focused only on the governance aspects of the SaaS application. They are not really focused on tying in to the governance strategies that the subscribing enterprise uses. The governance capabilities can be considered as the SaaS application features and are probably of less value to enterprise. It can be safely assumed that this is siloed governance.

Federated governance capabilities have not really come out in the SaaS applications so far. This approach of federated governance in SaaS applications is a class of governance capabilities where the SaaS application has skeletal governance capabilities, which is core to its functions. In addition, there are end points to tie into the governance strategies and solutions that the subscribing enterprise might be using so that a *federated* approach to avoid siloed governance is used. Currently, implementations use customization to employ this federated approach.

SaaS applications that depend on the subscriber to bring their own governance capabilities can use the approach of an *integrated* governance.

SaaS applications that are likely to use a federated governance and integrated governance fall into the purview of considerations for on-premises MDM solutions integrated with SaaS applications.

2.3.1 Integrations with SaaS applications acting as a source

An *extended enterprise* needs SaaS applications to be a source for MDM solutions. Consider the following key requirements for this type of environment:

- The MDM model must have mappings or new attributes created based on the amount of attribution the MDM solution has to manage for the SaaS application.
- The MDM matching algorithms must be calibrated to work with the model and data coming in from the SaaS application.
- Data cleansing and standardization rules employed in the MDM solutions are implicated with the characteristics of the data coming in from the SaaS applications.
- Initial load procedures must be defined to on board the data from the SaaS application to the MDM solution on premises.

- SaaS applications typically do not allow huge data extraction transactions. So any ETL tool employed in building the initial load jobs need to be able to handle some aspects of the *soft batching*, which is required based on the limits of the SaaS applications.
- Some SaaS applications provide export files, which must be securely moved to the on-premises systems to initiate load procedures.

2.3.2 SaaS applications consuming MDM services and MDM callouts to SaaS applications

This is one of the key topics that is associated with the day-to-day operations of an enterprise that wants to power its SaaS applications with the MDM capabilities deployed on premises. Beyond the considerations of MDM model implications, matching algorithms, cleansing and standardization, APIs or services are next in line.

The initial requirements are to ensure that the required services from the MDM solution are created or reused to cover all the interactions that are required for which the SaaS application is calling into the MDM solution. It preferable to have required coarse-grained APIs and services so that the “chatter” is reduced and well-qualified endpoint services are available. Another consideration is *reliable messaging* for certain integration scenarios. In this regard, the message format compliance as per the specs provided by the SaaS application is one key consideration.

Likewise, for the MDM solution, there should be proper APIs or services available or coded in the SaaS application.

The area of integration is where the *cloud integration middleware* fits in to broker the communications and interactions to and from the MDM solution and the SaaS application.

As discussed earlier, another important aspect for the cloud integration middleware is security, that is the need for secure connectivity which takes care of the secure transfer of data from behind the firewall onto SaaS applications and vice versa.

2.3.3 Bulk data operations between MDM solutions and SaaS applications

Bulk data operations are key considerations for pre-production setup of any solution that involves integrations that cater to data movement. In the case of MDM solutions integration with SaaS applications, initial data loads, batch loads,

and incremental data loads are areas where the capabilities of bulk data operations are required.

Consider the following cases that characterize how initial loads are expected to work:

- *Case 1*, where existing MDM solutions on promises and a new SaaS application is subscribed to

  In this example, the initial load scenarios are mainly directed towards the MDM solution onboarding all the data of the key master data entities and their relationships into the SaaS application.

- *Case 2*, where existing SaaS application and an MDM solution is being deployed

  In this example, the initial load scenarios are mainly directed towards the SaaS solution being treated as a sources. All of its relevant data is on boarded in the MDM solution for further treatment, which, in turn, might also result in bulk write-backs to the SaaS application.

- *Case 3*, where a SaaS application and an MDM solution is being deployed together

  In this example, the initial load scenarios can be tricky and the preferred practice is to first get the MDM solution deployed to a level of pre-production readiness by ensuring that it has covered all the on premises sources of data. Then, follow the guidance as described for Case 1.

  The considerations for batch load scenarios and incremental load scenarios are quite similar. The only difference is that the payload is not as large as initial loads and sometimes. There is a need for the high performance reconciliation routines to ensure that the data synchronization has indeed succeeded and that the exceptions encountered have the required remediation processes deployed.

In general, the bulks data operations are supported by ETL tools that have a broad set of connectors for SaaS applications and on-premises sources alike, which have scalable, high performance capabilities for data integration and data quality. The ETL tools should also support the exchange of business information such as terms, application metadata, taxonomies, and so on.

2.3.4 Preparations required for MDM solutions to integrate with SaaS applications

The cases discussed in 2.3.3, “Bulk data operations between MDM solutions and SaaS applications” on page 34 are used for the discussion in this section:

- *Case 1*, existing MDM solutions in premises and a new SaaS application is subscribed to
- *Case 2*, existing SaaS application and an MDM solution is deployed
- *Case 3*, SaaS application and an MDM solution is deployed together

The preparations will differ based on these cases too. However, one of the key points to understand is that the core business driver for integrating MDM solutions is information governance for the extended enterprise. Thus, the first set of preparations have to be focused on nuances of information governance. IBM Unified Data Governance Process provides some good guidelines about the information governance polices to be adopted.

Figure 2-4 shows the process flow for laying down a typical governance process in an enterprise.

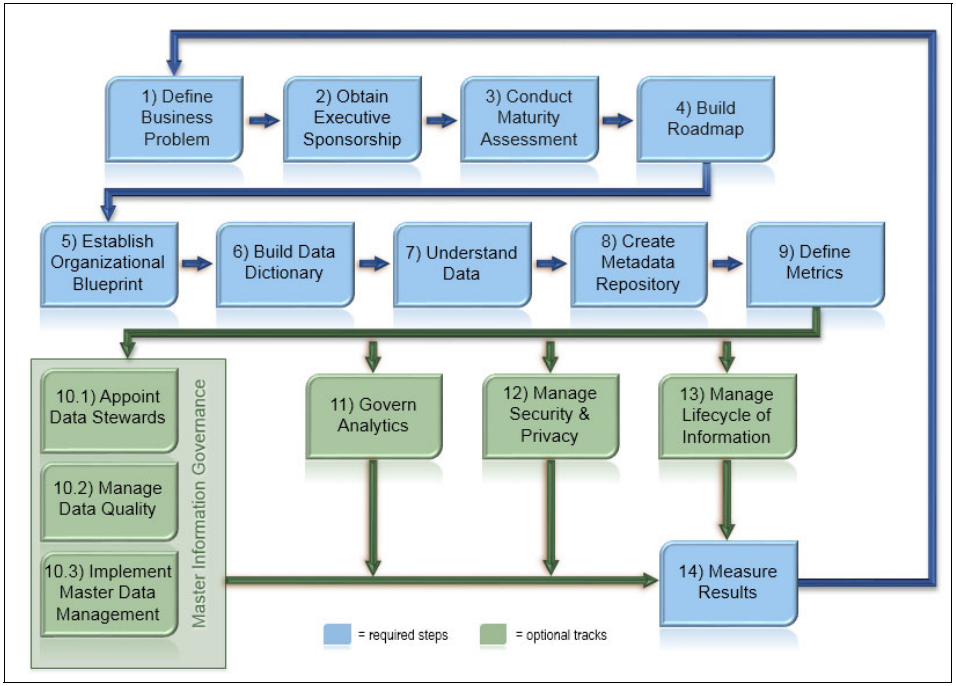


Figure 2-4 IBM Unified Information Governance process

The next area to be prepared is the understanding of the CCRA, which helps provide details around the Cloud Adoption Patterns, Cloud Enabled Data Centers, aspects about Cloud Service Provider, Business Solutions on Cloud, and finally Securing Cloud Solutions.

Figure 2-5 shows the advantages of cloud adoption patterns.

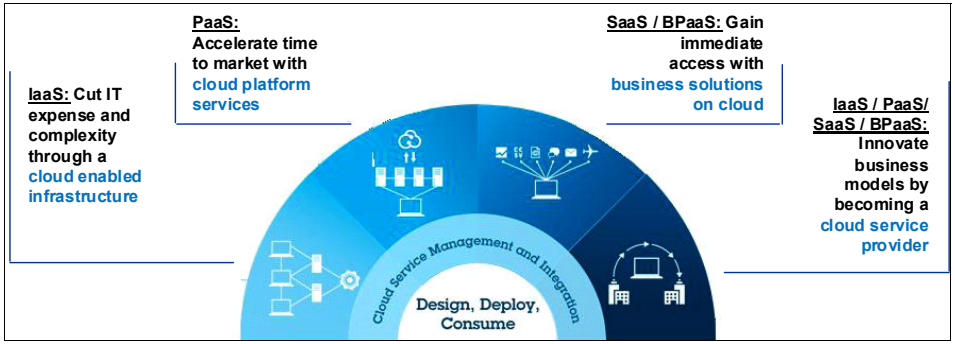


Figure 2-5 High-level advantages of adoptions of different cloud offerings

It is also good to review artifacts for background purposes, such as *Patterns For Information Management*, ISBN: 978-0-13-315550-1.

Figure 2-6 shows a layout of key patterns of information management.

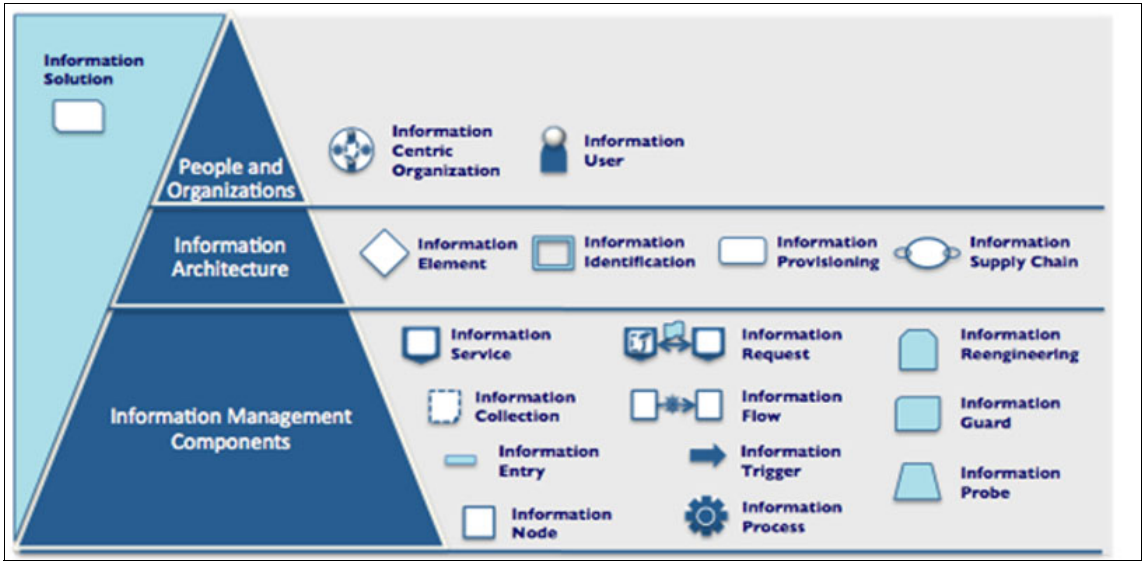


Figure 2-6 Key Patterns of information management

One of the key aspects of knowing the patterns of information management is understanding the big data reference architecture. A good perspective of this is provided by *Smarter Analytics: Information Architecture for a New Era of Computing*, REDP-5012.

Figure 2-7 shows the overall layout of the big data reference architecture, where all data sources participate in the information supply chain in their journey to the information delivery zone.

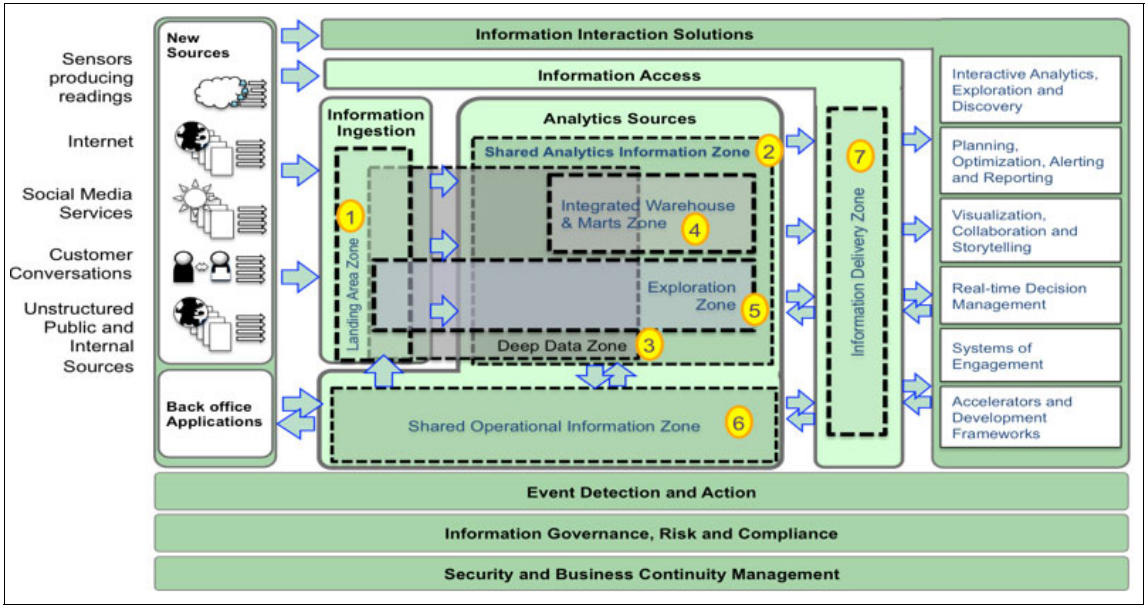


Figure 2-7 Big data reference architecture

Other preparatory requirements are the coverage for logical data mappings of the entities so that early view on design lineage is established.

Other actions to be undertaken are to review the security guidelines in play in the enterprise and to check whether there are any updates required as though this integration is the first cloud-related project being undertaken.

2.4 Aspects of solution definition

This section covers various considerations for defining the solution based on the background provided in 2.3, “General requirements for integrations of on-premises MDM solutions with SaaS applications” on page 32.

2.4.1 High-level requirements

There are various topics that have discussed the requirements for the solution of getting on-premises MDM solutions integrated with SaaS applications. This

section attempts to summarize all of that and draw a perspective on an indicative and crisp list of capabilities that must be provided by the solution.

The following is a simple list of key functional requirements:

- SaaS applications should be able to use the probabilistic search offered by the MDM solution to drive “embedded” de-duplication techniques.
- The MDM solution must have awareness of the master data entities in the SaaS application and should be able to enrich them with the “enterprise wide” awareness of this master entity.
- The MDM solution and the SaaS application should be able to jointly manage relationships and hierarchies where in relationships are shared across the two and hierarchies are rationalized wherever applicable.
- The solution should have relevant components to manage all the scenarios for bulk data operations.
- Each application should be modified or customized to handle events from either side or other common events affecting both these solutions.
- Cloud integration middleware needs to include the following capabilities:
  - Integrate applications from cloud-to-cloud and cloud-to-on-premises primarily.
  - Must provide a rapid and easy to use method of setting up the integrations.
  - Should be available in multiple form factors to align with the adoption cycle of cloud integration and on premises priorities.
  - Secure connectivity support to allow the secure transfer of data from behind the firewall.
  - Provide a single, unified monitoring solution for multiple environments.
  - Provide broad set of connectivity options.
  - Provide integration reliability and guaranteed delivery.
- Cloud information integration middleware (hybrid cloud ETL tool) needs to include the following capabilities:
  - Secure connectivity to SaaS applications.
  - Should work with the API provided by the SaaS application seamlessly to perform required operations.
  - Should support SaaS application bulk load and bulk extract limitation.
  - Provide automatic generation of SaaS application query language semantics and ability to support the same on job stages.

- Understand metadata of the SaaS application understood for data mapping and transformations.
- Should be able to scale linearly within the boundaries of the SaaS application cloud operating environments (resource limitations.

2.4.2 Security requirements

Security requirements are covered in this dedicated section because of this issue is of paramount importance. The security considerations typically fall into the following categories:

- Data security
- Services-related security

Security remains the top inhibitor to broad-scale cloud adoption. In the case of on-premises MDM solutions integrating with SaaS applications, security considerations typically start from the endpoints provided by the MDM solution.

Data *visibility considerations* are of primary consideration. Most MDM solutions do have some form of a “Rules of Visibility” implementation that covers most of the data visibility requirements. Next in line is *secure connectivity*, that is the connectivity between applications behind the enterprise DMZ and the cloud application. The focus here is on the secure connectivity that is provided by the cloud integration middleware without the need for VPNs or opening up of inbound ports for SaaS applications.

There are other security considerations that are SaaS application specific that are typically encapsulated in the connector provided by the cloud integration middleware. That said, the cloud integration middleware might not be protected against other security threats, such as denial-of-service attacks (DoS). Thus, steps must be taken to wrap the cloud integration middleware with what has already been done to ward off these threats.

Keep in mind the following additional low-level considerations for the cloud integration middleware:

- Advanced logging features
- Strong access control lists (ACL) support
- Granular audit capabilities
- P800-131a and other National Institute of Standard and Technology (NIST) compliance norms

- Service layer security, for example Security Assertion Markup Language (SAML), OAuth 2.0 support[1]
- Encryption and decryption, for example Pretty Good Privacy (PGP), Advanced Encryption Standard (AES), and some others

The same applies to ETL tools that support hybrid cloud topologies in an extended enterprise.

2.4.3 Different solution approaches

Multiple solution approaches are available to solve issues with having on-premises MDM solutions integrated with SaaS applications. For our discussion, we focus on InfoSphere MDM as the MDM solution. The cloud integration middleware of choice for many real-time integrations with SaaS applications is IBM WebSphere Cast Iron Cloud Integration. For ETL tools that support hybrid cloud topologies in an extended enterprise, drawing a context on bulk operations between InfoSphere MDM solutions and SaaS applications, IBM InfoSphere Information Server is the ETL tool of choice.

Figure 2-8 illustrates a solution architecture that is based on the following software components (where the numbers correspond to the numbers in the illustration):

1. The MDM solution that exposes REST or web services has messaging support and uses ETL tool for bulk operations.
2. Every enterprise typically uses a Reverse proxy solution to hide IPs of all servers in the DMZ.
3. Every enterprise has a firewall that protects its data centers.
4. Every interface that the MDM solution uses for integrating with a SaaS application ultimately should connect only to a secure connectivity component.
5. The cloud integration middleware can be of multiple form factors. It can be a virtual applicancy, a physical applicancy, a hosted SaaS offering, or capabilities in a PaaS platform. Depending on the security requirements, any one of the form factors can be used. The suggested option is to use the hosted SaaS offering form factor by tying into co-joined integration platform as a service (iPaaS) initiatives.
6. A set of generic coarse grained services, messaging, or ETL jobs should be served by the MDM solution in preparation to serve multiple SaaS

[1] OAuth is an open standard for authorization. For more information, see: http://oauth.net/

applications that the enterprise might subscribe to, thus extending governance capabilities to all of them consistently.

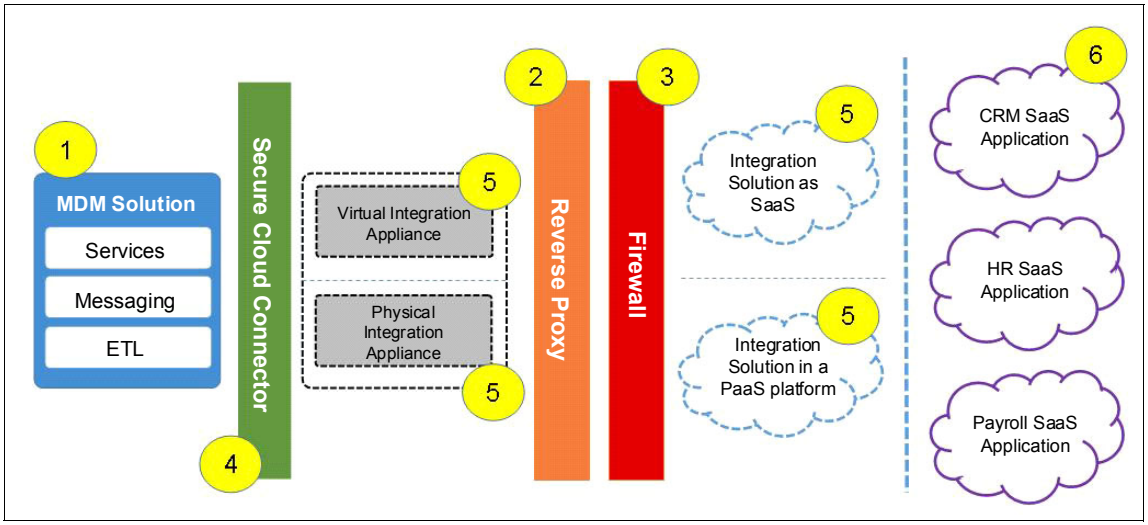


Figure 2-8 Tentative solution architecture layout

The remaining chapters in this book discuss details about how InfoSphere MDM can be integrated with SaaS applications in a generic fashion.

Note: To obtain generic perspectives about the IBM WebSphere Cast Iron Cloud Integration, see *Connect Cloud and On-premise Applications Using IBM WebSphere Cast Iron Integration*, REDP-4674.

3

Middleware considerations for integration

Software as a service (SaaS) applications provide companies the ability to innovate without a huge investment in the IT infrastructure and resources. Nevertheless, organizations usually have numerous existing investments in on-premises software. Furthermore, as the number of SaaS applications grows within an organization, so does the need for integration among SaaS applications and on-premises applications.

Given the persistent time and budget constraints, companies are not inclined to extend the use of traditional integration approaches to hybrid integration scenarios. Thus, an integration solution can provide services to bridge the gap between the existing on-premises systems and new cloud applications, platform, and infrastructure by providing a rapid and easy-to-use method of setting up the integrations. The integration challenges span data, access, and process.

This chapter discusses the following topics:

- Data integration between on-premises MDM systems and SaaS applications
- Access integration between on-premises MDM systems and SaaS application
- Process integration between on-premises MDM systems and SaaS applications

3.1 Data integration between on-premises MDM systems and SaaS applications

With the explosion of data spanning across multiple application systems, data integration has become complex. Organizations have a need to connect to multiple source applications and a need to retrieve and to manipulate data before passing the data to the target application.

This section describes the capabilities of IBM WebSphere Cast Iron Cloud Integration and IBM InfoSphere Information Server, which play a pivotal role in integrating data between the on-premises master data management (MDM) applications and SaaS applications.

3.1.1 Introduction to WebSphere Cast Iron Cloud Integration capabilities

Cloud computing has become a business evolution that impacts different business functions, such as manufacturing, finance, marketing, sales, human capital management, and so on. As each of these functions move to the cloud, there is a need for migration and integration to close the gap between existing on-premises applications and the newer cloud applications.

When companies must integrate in-house applications and SaaS applications, old tools and ad hoc methods do not stand up to the challenges in orchestration, security, data transformation, and more. IT organizations might assume that cloud application integration can be done by traditional integration methods, such as extract, transform, load (ETL), external electronic data interchange (EDI), or writing custom code for each of the applications. Solving integration problems with these traditional approaches is not beneficial in the long run because they need skilled programmers, they are time consuming, and they can be expensive. The poor fit between traditional integration approaches and the requirements of cloud computing has created the need for a new type of integration platform.

IBM WebSphere Cast Iron Cloud Integration helps companies integrate existing on-premises systems with cloud applications, cloud applications to other cloud applications, and on-premises applications to other on-premises applications. By following a “no-coding” approach, WebSphere Cast Iron Cloud Integration allows you to focus on the business requirements and removes the daunting complexity of integration. It has everything needed to support integrations in a hybrid world in a single integration platform, such as data migration and process integration. By using WebSphere Cast Iron Cloud Integration, IBM clients can expect fast and flexible SaaS and cloud application integration in the space of days that provides lower cost and higher return on investment in cloud and SaaS models.

Figure 3-1 depicts IBM offerings of cloud application integration with WebSphere Cast Iron Cloud Integration.

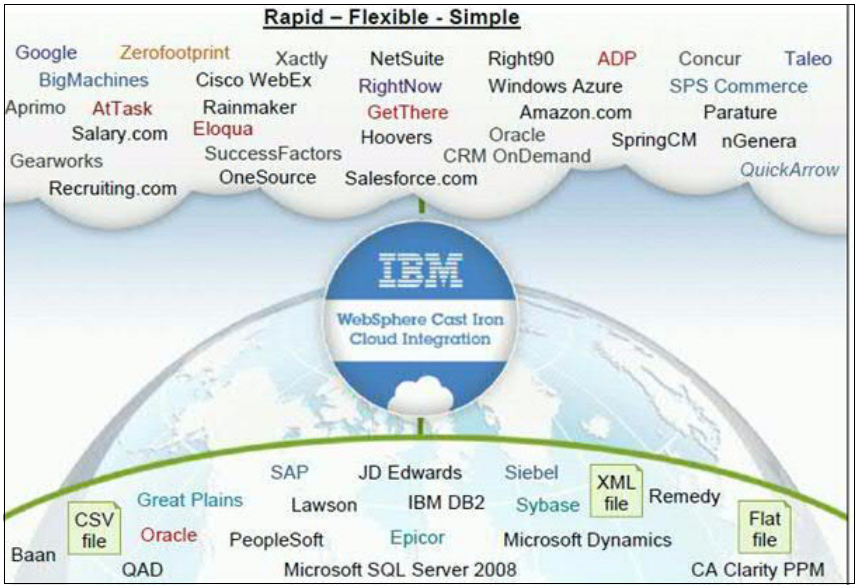


Figure 3-1 WebSphere Cast Iron Cloud Integration

Here is an example of how WebSphere Cast Iron Cloud Integration is used by a healthcare company, Heathcare Company A.

Heathcare Company A is a rapidly growing healthcare company that specializes in the management of emergency room facilities. Heathcare Company A must integrate customer master, project, and resource emergency room data with systems further downstream, where the care is delivered outside of the emergency room. The data must move between PeopleSoft and Salesforce.com applications and repositories kept in MS SQL Server and other data sources that are accessible through web services. Both batch and real-time integrations were needed to improve data quality and reliability and to allow for a continuing cycle of business process improvement. In addition, the company chose integration provided through one solution for both real time and batch integration. The initial project was completed in a few days by junior personnel and then was extended to solve new problems. The IT staff found WebSphere Cast Iron Cloud Integration complemented existing policies for disaster recovery. The flexibility of these cloud integration products enabled Heathcare Company A to easily integrate their hybrid world of cloud and on-premises applications.

Features in WebSphere Cast Iron Cloud Integration

WebSphere Cast Iron Cloud Integration has the following capabilities:

- Universal connectivity

  WebSphere Cast Iron Cloud Integration provides reliable, ready-for-use connectivity options to a wide range of on-premises and cloud applications. All the applications are exposed as endpoints and have a similar kind of user interface to work with. WebSphere Cast Iron Cloud Integration contains many built-in connectors for applications (for example SAP, Salesforce.com, and Netsuite), databases, and web services that make connecting to these endpoints straightforward.

- Transformation and workflows

  WebSphere Cast Iron Cloud Integration enables you to transform data from source to target applications using a graphical interface with a drag-and-drop approach. It has built-in functions to transform data from the source format to the format that is required by the target system and to build relevant mapping rules. You can create workflows that govern the integration process and define business rules without any coding. By using a graphical interface, WebSphere Cast Iron Cloud Integration can build these workflows quickly and easily.

  WebSphere Cast Iron Cloud Integration provides an approach to integrating applications without programming. You can build integration flows in WebSphere Cast Iron Studio, which is a graphical development environment that is installed to a personal computer. With WebSphere Cast Iron Studio, you can create an integration project that contains one or more *orchestrations*. Each orchestration is built with activities that define the data flow.

  A project contains all of the information that is required for the orchestrations to run, including any file schemas, Web Services Description Language (WSDL) files, and functions.

Figure 3-2 shows a sample orchestration in WebSphere Cast Iron Studio.

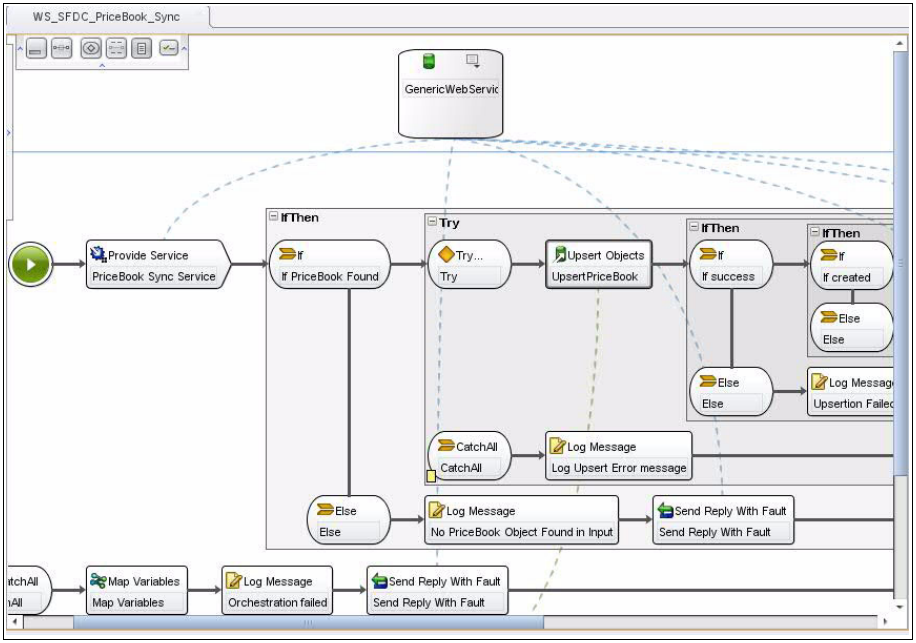


Figure 3-2 Sample orchestration in WebSphere Cast Iron Studio

- Management

  WebSphere Cast Iron Live provides a web-based console to administer and monitor projects. WebSphere Cast Iron Live can monitor real-time integration activity, handle exceptions, and provide proactive alerting for data and connectivity errors. You can use WebSphere Cast Iron Live to monitor the projects that are running by providing granular job-related information. WebSphere Cast Iron Cloud Integration contains useful tools to generate, visualize, and export logs.

Figure 3-3 shows a sample Web management console of WebSphere Cast Iron Live.

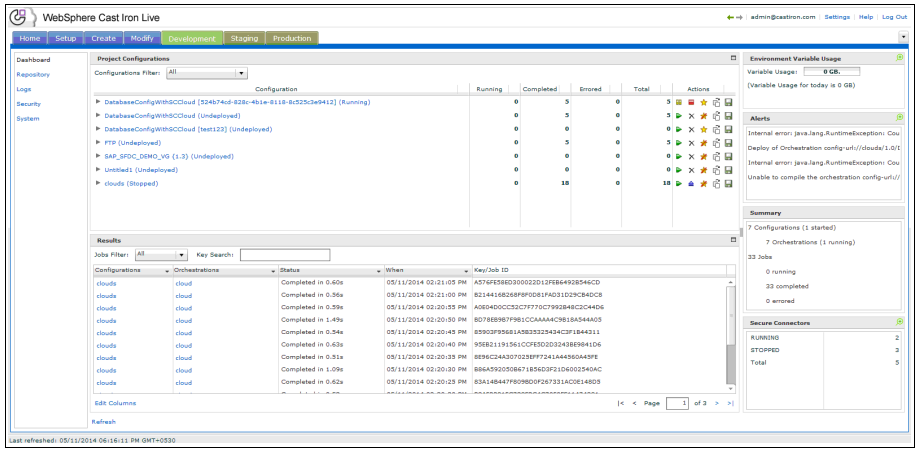


Figure 3-3 Web management console of WebSphere Cast Iron Live

WebSphere Cast Iron Live pushes log events to the system and job logs. Log events generated by orchestrations (*job instances*) go through job filters that can block events from being logged. The system log determines the level of logs it will accept through an additional filter. Although a log event can pass the job filter and be sent to the system log, the event can still be blocked. In addition to job instance log events, WebSphere Cast Iron Live also produces log events that are not related to orchestrations, such as security, hardware, and network events. These log events are pushed to the system log.

WebSphere Cast Iron Live can generate alert notifications allowing users to monitor the system for certain events. Email and SNMP notifications are available in the appliance model whereas only email is available in the Live model.

3.1.2 Deployment options in WebSphere Cast Iron Cloud Integration

WebSphere Cast Iron Cloud Integration has the following implementation models:

- *An appliance model*: The appliance can be either physical hardware or a virtual machine, where the integration appliance is installed on premises, behind the firewall.
- *A SaaS model* (Cast Iron Live): IBM WebSphere Cast Iron Live is a multi-tenant, cloud-based platform for integrating cloud and on-premises applications and enterprise systems in a hybrid environment. It enables you to

configure, run, and manage integration in the cloud without any infrastructure footprint.

WebSphere Cast Iron Studio is a development tool provided by WebSphere Cast Iron Cloud Integration for creating projects and orchestrations. When using the appliance model for deployment, WebSphere Cast Iron Studio is a local application that must be installed and run on a Windows operating system. For WebSphere Cast Iron Live, WebSphere Cast Iron Studio is a dynamically downloaded Java application that is started from the modify tab inside WebSphere Cast Iron Live.

WebSphere Cast Iron Cloud Integration appliance

IBM WebSphere DataPower Cast Iron Appliance XH40 is a self-contained, physical appliance that provides what is needed to connect cloud and on-premises applications. It can be installed and managed within a local data center. WebSphere DataPower Cast Iron Appliance XH40 can connect to the network with Ethernet.

IBM WebSphere Cast Iron Hypervisor Edition is a virtual appliance that can be installed on existing servers through virtualization technology. Virtual appliances improve the utilization of hardware and enable faster response to demands for newly deployed systems. It can help to reduce the costs of both hardware and software operation and maintenance.

On-premises appliance normally resides behind the firewall.

WebSphere Cast Iron Live

WebSphere Cast Iron Live provides development environment and runtime functions as a physical or virtual integration appliance that are similar to WebSphere Cast Iron Studio through a cloud-based service. In WebSphere Cast Iron Live, you publish a project to an environment, such as a development, test, and production environment. The environment contains the functions of an on-premises integration appliance as shown in Figure 3-4.

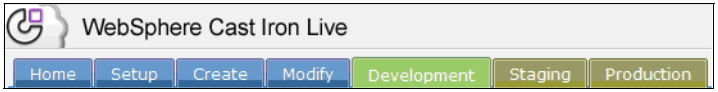


Figure 3-4 Environments in WebSphere Cast Iron Live

You can build integration flows in WebSphere Cast Iron Live and create an integration project that contains one or more orchestrations (Figure 3-5). Each orchestration is built with one or more activities.

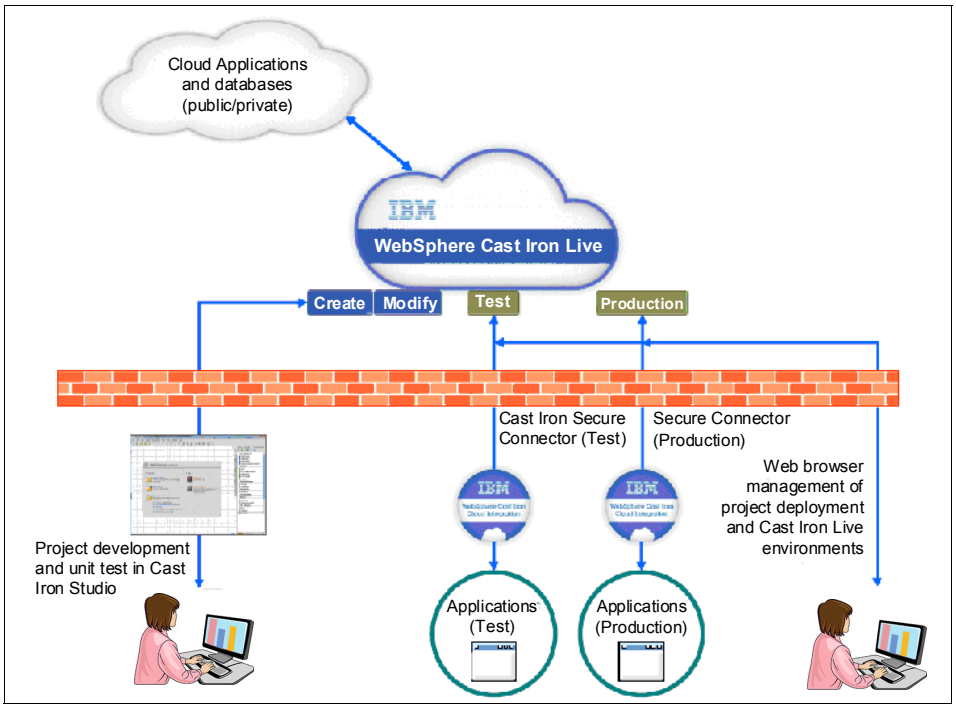


Figure 3-5 Building integrations using WebSphere Cast Iron Live

The multi-tenant WebSphere Cast Iron Live service includes the following components that allow you to design, run, and manage integrations all in the cloud:

- A clustered runtime engine that runs the integrations and that has built-in fault-tolerance and recovery mechanisms
- A multi-tenant highly available system to store the designed integrations
- A load balancer to intelligently manage the loads throughout the various runtime engines
- Highly available file systems to store and manage logs that are related to integrations

3.1.3 IBM Information Server

IBM InfoSphere Information Server is a market-leading data integration platform, which includes a family of products, to help you understand, cleanse, transform, and deliver trusted information to your critical business initiatives, such as big data, MDM, and point-of-impact analytics. InfoSphere Information Server provides the ability to flexibly meet unique information integration requirements, ranging from data integration, to data quality, to data governance, and more.

InfoSphere Information Server Packs enable enterprise applications to benefit from multiple capabilities of InfoSphere Information Server, such as support for complex transformations, automated data profiling, data quality, and integrated metadata management. InfoSphere Information Server Packs help your organization integrate data and create consistent, trustworthy information.

InfoSphere Information Server Pack for Salesforce.com enables repeatable, reusable data integration between Salesforce.com and other enterprise data repositories.

InfoSphere Information Server Pack for Salesforce.com includes the following features:

- Secure connectivity to Salesforce.com from InfoSphere Information Server through the Internet.
- Easy user interface-based, application-level access to Salesforce.com metadata through the Force.com application programming interface (API).
- Support for Salesforce.com asynchronous bulk load API to enable high volume batch access and delivery of Salesforce.com CRM data.
- Automatic generation of Salesforce Object Query Language (SOQL) syntax and the ability to use custom SOQL.
- Discover, load, and extract Salesforce.com business objects from InfoSphere Information Server.
- Help to reduce costs and increase productivity of ETL development teams by eliminating the need to learn Salesforce.com APIs.
- Enhancement for the ROI benefits of hosted Salesforce.com CRM and Salesforce.com automation (SFA) deployments.
- Ability to clean and transform Salesforce.com data in the cloud and integrate it with on-premises applications, including ERP and MDM systems.
- Increased visibility of back-office data to customer-facing employees, enabling them to make more intelligent and trusted decisions to transform the business.

3.2 Access integration between on-premises MDM systems and SaaS application

Although SaaS applications provide greater flexibility and lower costs, enterprises have identified access integration as a key obstacle to adopting SaaS applications. Security remains a key concern for users. A cloud integration solution must be able to authenticate and authorize access to resources, encrypt data, and comply to regulations. With the growing number of SaaS applications, mobile apps, and other services that must access enterprise data, a secure connection is required from the cloud system to the enterprise system. Ready-for-use connectivity to various applications is in great demand.

This section discusses how WebSphere Cast Iron Cloud Integration can help to solve issues related to access integration.

3.2.1 Security

WebSphere Cast Iron Cloud Integration is designed to provide a secure, scalable, and robust means to accomplish data integration within and between the cloud and the enterprise. This section describes the security features specific to *endpoints*. For more information about the complete list of security features offered by WebSphere Cast Iron Cloud Integration, see *Getting Started with IBM WebSphere Cast Iron Cloud Integration*, SG24-8004.

With WebSphere Cast Iron Cloud Integration, you can use a variety of secure communication protocols to communicate with endpoints such as applications and databases:

- HTTPS (HTTP over SSL): Supports bilateral authentication, privacy, and integrity
- Secure web services using SOAP/HTTP over SSL
- Secure FTP(FTP over SSH) and FTPS (FTP over SSL or implicit FTPS)
- Secure databases (SSL)
- OAuth 2.0 authentication: Connect to a resource server endpoint through OAuth 2.0 authenticated access

Secure connector

Applications that are deployed on premises can be integrated in a secure way by using WebSphere Cast Iron Cloud Integration Secure Connector. The WebSphere Cast Iron Cloud Integration Secure Connector is a light-weight piece of software that resides behind the firewall and connects databases, enterprise

applications, and message queues behind the firewall in a secure manner with the cloud.

The secure connector can help you overcome the limitations caused by a firewall between the on-premises applications and the orchestration run time. The secure connector works for both the outbound and inbound traffic. An outbound scenario is where accessing and publishing information to an endpoint behind firewall is required. An inbound example can be an orchestration that has a starter activity of HTTP receiving request where the HTTP listener can resides behind the firewall.

The WebSphere Cast Iron Live Secure Connector allows orchestrations in Cast Iron Live to connect to endpoints behind a firewall. A separate Secure Connector is configured for each environment and is installed to a computer that runs Windows or Linux (both 32-bit and 64-bit installers are available) operating systems behind the firewall.

Secure Connector includes the following security features:

- Secure Connector always initiates communication with WebSphere Cast Iron Live, and that communication is validated before the Secure Connector attempts further processing. It runs only activities marked remote behind firewall and not the entire project.
- Communication between Secure Connector and WebSphere Cast Iron Live is based on the standard SSL 128-bit encryption over HTTPS through port 443. When Secure Connector starts, it undergoes the SSL/TLS handshake, authenticates through standard X.509 certificates, and establishes a TLS-encrypted tunnel if all connections are successful.

  When the TLS connection is established successfully, Secure Connector sends a request to WebSphere Cast Iron Live for authentication. Based on information that is provided by the Secure Connector, including a private key, InfoSphere Cast Iron Live ensures that only the correct Secure Connector is granted access to a particular environment of a tenant.

After Secure Connector is installed and configured on the local system, its status can be monitored in web management console.

For more information about using Secure Connector on WebSphere Cast Iron Live, see the WebSphere Cast Iron Cloud Integration information Center at:

http://pic.dhe.ibm.com/infocenter/wci/v7r0m0/index.jsp?topic=/com.ibm.wci.live.doc/Secure_Connector/toc_using_secure_connectors.html

3.2.2 WebSphere Cast Iron Cloud Integration connectors

WebSphere Cast Iron Cloud Integration comes with built-in connectivity to many clouds, packaged, and proprietary on-premises applications, including ERP, CRM, databases, web services, and flat files. It is self-contained and includes everything needed to complete integrations in one place. The progressive integration approach makes no distinction between local and remote applications because they establish connectivity to the endpoints by using native application protocols. The advantage to this approach is that no additional adapters are required and that there is nothing to install or change at the endpoints.

The Connector Development Kit provides a method for building connectors to endpoints that are not already included in the product, thus reducing the need to manually build these connections over and over. Connectors can be built for every application that either exposes a web service or can be accessed by REST APIs.

You can build an orchestration from a number of activities which are listed by category in the WebSphere Cast Iron Studio Activities tab as shown in Figure 3-6 on page 55. Activities exist for all connectors, and there are activities for handling workflow logic, for example, error handling with Try/catch blocks and transformations.

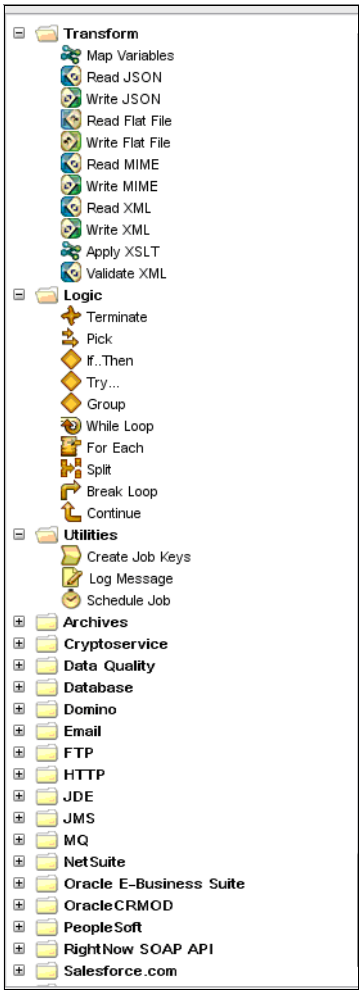


Figure 3-6 Activities in WebSphere Cast Iron Studio

The connectors are referred to as endpoints in WebSphere Cast Iron. All the endpoints have an endpoint panel where access information to the application must be entered. The properties include URL, user name, password, port number, and son on. If the endpoint is running behind a firewall, you have an option to specify the secure connector name. The endpoint configuration panel also has a Test Connection feature that helps developers to verify whether the information provided is sufficient and valid to connect to the desired application.

Figure 3-7 shows the endpoint panel of Salesforce.com connector.

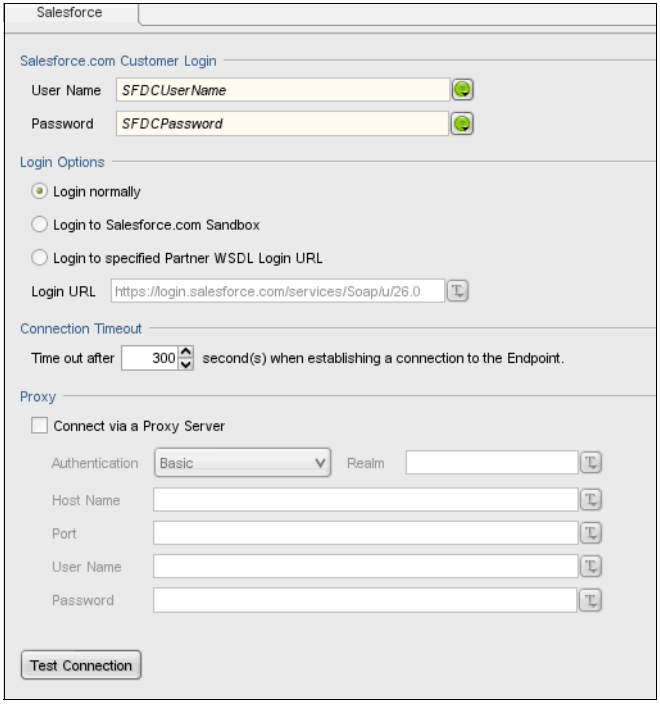


Figure 3-7 Endpoint panel of Salesforce.com connector

Each endpoint exposes different number of activities, depending on the application. The endpoints enable bidirectional communication between WebSphere Cast Iron and the target application. Basically, there are the following types of activities:

- An *outbound* activity accepts input, processes that input, and provides an output.
- A *starter* activity either listens to an external event or polls an external source periodically and eventually provides an output.

Some endpoints also have a *query* activity, which provide user-defined queries that can be run against the application and which provide search results that can be used for processing.

You use an activity by dragging it into the workspace. The configuration pane contains the checklist and items that are configured for each activity. When you select an activity, the configuration pane is displayed in the lower section of the studio. Activities for endpoints include a checklist of items and the input and

output mapping requirements. Logic and transformation activities list items that are required for the configuration of that activity.

Figure 3-8 shows the checklist panel for the Salesforce.com Upsert Objects activity.

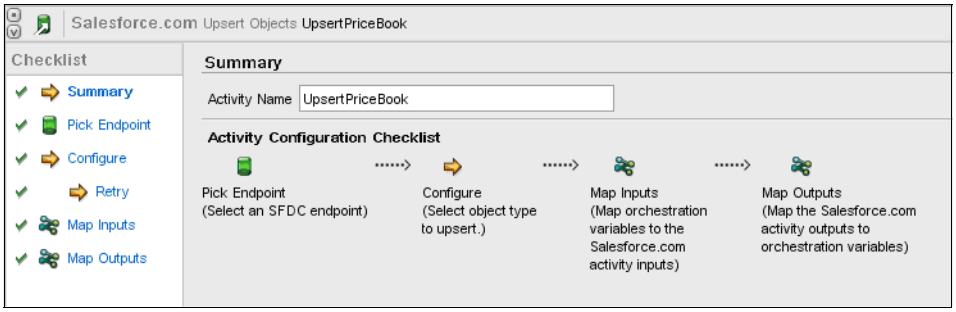


Figure 3-8 Checklist for Salesforce.com Upsert Objects activity

The activities for an endpoint also include a checklist of items for mapping input and output values, except for those activities that only have either inputs or outputs. In a mapping, the map inputs define the data that is sent to the activity. The map outputs determine the data that can be used by other activities in the orchestration. Mapping is easy to configure using the configuration panes as shown in Figure 3-9.

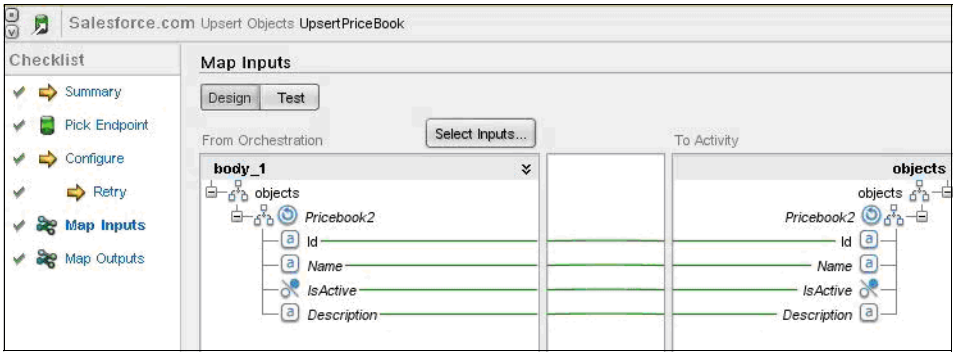


Figure 3-9 Map input panel for Upsert Objects activity

WebSphere Cast Iron Studio provides the Verify tool for unit test to verify that orchestrations are functioning as expected. WebSphere Cast Iron Studio provides the ability to validate the orchestrations before publishing them. After starting an orchestration job, the orchestration waits for the starter activity to initiate. You cannot verify or publish a project if there are incomplete configurations.

Only one job instance can run for each orchestration in the Verify toolbox tab at a time. However, after one job instance completes, another job instance for the same orchestration can be started. Trace data for each job instance for the orchestration displays in the Verify toolbox tab, as shown in Figure 3-10.

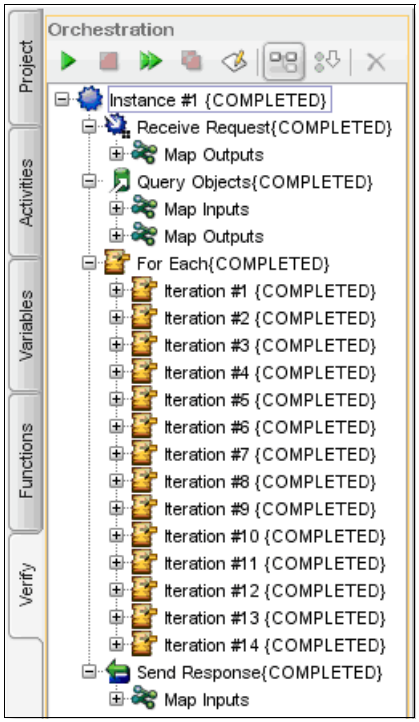


Figure 3-10 Verify

3.3 Process integration between on-premises MDM systems and SaaS applications

Cloud computing is about flexibility, speed, and cost reduction, yet the practical realities dictate that these advantages are often muted by the complexities of integration. Business process management and setting procedures for the smooth exchange of data inside and outside the enterprise is needed. Companies need a jumpstart in building their processes so that precious time is saved.

Reusability is a key factor in efficiency and for the ease of use of any product. In WebSphere Cast Iron, you can reuse orchestrations by storing them as template integration projects.

3.3.1 Data flow orchestration using TIPs

Although WebSphere Cast Iron Studio provides the capability to build complex integrations from the beginning, WebSphere Cast Iron Cloud Integration also provides Template Integration Processes (TIPs) that encapsulate a specific integration use case between specific endpoints and include preferred practices. A common repository that includes all of the available TIPs are published and is available for download by any customer or IBM Business Partner. These TIPs are authored by internal or external integration experts who have expertise with the endpoints for which they authored the integrations.

WebSphere Cast Iron Cloud Integration includes hundreds of reusable TIPs that can be searched within the repository. This online library contains templates for all of the most common cloud integration scenarios. These templates provide a question-and-answer-based wizard that walks users through a common integration scenario. For example, a TIP might request all the information needed to transform an opportunity in a CRM system to an order in an ERP system. Through this simplified interface, IBM customers and partners can create their own wizard-driven reusable templates with the point-and-click TIP Development Kit and make them instantly available to the entire WebSphere Cast Iron user community.

TIPs provide the following key benefits:

- Save costs because the majority of the developer effort to build integration is already codified in the template. The vast majority of the remaining effort involves using a wizard to configure the endpoint and apply any special customizations, such as custom fields or objects mapping and special business logic.
- Readily referenceable preferred practices because each certified TIP is carefully authored and provides a method to reference and learn from other users and WebSphere Cast Iron experts.
- Flexibility to adapt to changes because the TIPs and TIP configuration wizard allow a user to modify integration when business needs change and additional fields or new business rules must be accommodated. The wizard automatically provides options for a user to select custom fields, change maps, or change business rules.

Figure 3-11 shows a sample TIP Configuration Wizard.

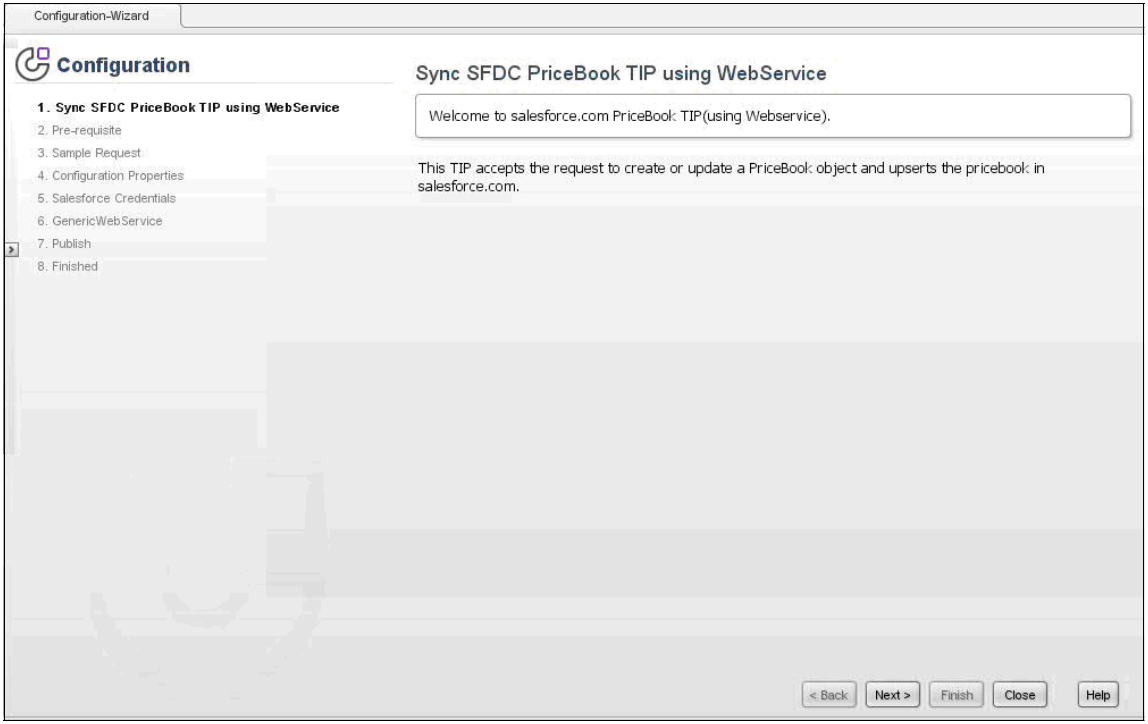


Figure 3-11 TIP Configuration Wizard

3.3.2 BPM based Stewardship

IBM InfoSphere Master Data Management (InfoSphere MDM) provides a variety of tools for data stewards, business analyst, and business users to manage sustainable governance of information quality, to master the complete lifecycle of information, to search for entities, and to update them. One such tool is IBM Stewardship Center.

Data stewards and managers use the tools and processes provided by IBM Stewardship Center in conjunction with the capabilities provided by IBM Business Process Manager (BPM) to perform data stewardship, and stewardship management.

The capabilities provided with IBM Stewardship Center are divided into the following categories:

- Data stewardship capabilities
  - Reviewing and updating the information in the records, including relationships and hierarchies
  - Resolving tasks, such as suspected duplicate tasks
  - Data steward managers can assign tasks to individual team members or groups
- Performance metrics capabilities
  - Monitoring how many tasks are overdue, at risk, or on track, which you can view by task type or by team
  - Monitoring the task turnover rate
  - Monitoring the list of open tasks for the current week and the upcoming week

By using IBM Stewardship Center and BPM, organizations can build custom workflows that can help integrate data maintenance processes between InfoSphere MDM and an external SaaS application. This method ensures that customer data and information is managed well through its lifecycle and also helps users on external SaaS applications to obtain a 360-degree view of customer information, leading to better decision making.

By implementing and achieving process level integration between InfoSphere MDM and external SaaS applications using BPM, organizations can drive and enforce governance mechanisms and policies, both inside and outside the enterprise.

4

Solution architecture

This chapter provides certain deeper perspectives about the applicable integration patterns applicable to the integration of master data management (MDM) and SaaS applications. This chapter also covers the overview of the suggested solution architecture and how the applicable integration patterns are applied.

The chapter uses the example of the integration of InfoSphere Master Data Management (InfoSphere MDM) with Salesforce.com by using IBM WebSphere Cast Iron Cloud Integration as the middleware of choice. It explains different *integration flows*, such as real-time synchronization, near real-time synchronization, and batch mode.

This chapter discusses the following topics specifically:

- Integration patterns applicable for integration of MDM systems and SaaS applications
- Solution architecture for integrating MDM systems and SaaS applications
- Considerations for bulk load operations

4.1 Integration patterns applicable for integration of MDM systems and SaaS applications

Integration patterns are pertinent in designing an integration solution architecture for integrating two or more applications or solutions. It is important to begin with good background and with tested standards and integrations patterns so that the solution architectures are based on mature approaches that are proven in other related or similar scenarios.

Most of the requirements for the integration of InfoSphere MDM with SaaS applications can be aggregated into the following main areas to define the approaches:

- Real-time synchronous integration
- Real-time asynchronous integration or near real-time integration
- Approaches for bulk operations

This top-level view of the approaches needs helps in isolating the patterns of integration that are applicable for building the solution architecture. Based on these approaches, the following integration patterns are key to building the solution architecture:

- Data integration patterns
- Access integration patterns
- Process integration patterns

As mentioned earlier, the allegiance to patterns is important because a good solution architecture is always a usage derivation of well-defined patterns.

For more information about integration patterns, see the following resources:

- Enterprise Integration Patterns

  http://www.eaipatterns.com/toc.html

- Integration Patterns and Practices from Salesforce.com

  http://www.salesforce.com/us/developer/docs/integration_patterns/integration_patterns_and_practices.pdf

4.1.1 Data integration patterns

In the context of MDM systems and SaaS application integration, data integration patterns typically deal with nuances of data movement between the two solutions. These patterns can either be initial load from the MDM system to the SaaS application or vice versa. It can also deal with patterns that hand the incremental data load from the MDM system to the SaaS application or vice

versa. Finally, these patterns can also include the use of services or reliable messaging by addressing data movement only between the MDM system and the SaaS application in real time or near real time.

Data integration patterns cover other areas such as information flow orchestration, data transformations, events handling, and other triggers that require data movement.

For InfoSphere MDM integration with SaaS applications, the primary data integration patterns to work with are as follows:

- Initial data load
- Batch and incremental data load
- Runtime data integration

Patterns for initial data load

The initial data load patterns must cover the following main data integration flows:

- Initial load of master records from InfoSphere MDM to one or more target SaaS applications
- Initial load of data from one or more SaaS application into InfoSphere MDM

Figure 4-1 shows the overall layout of the pre-production setup for initial loads in the context of InfoSphere MDM integration with SaaS applications.

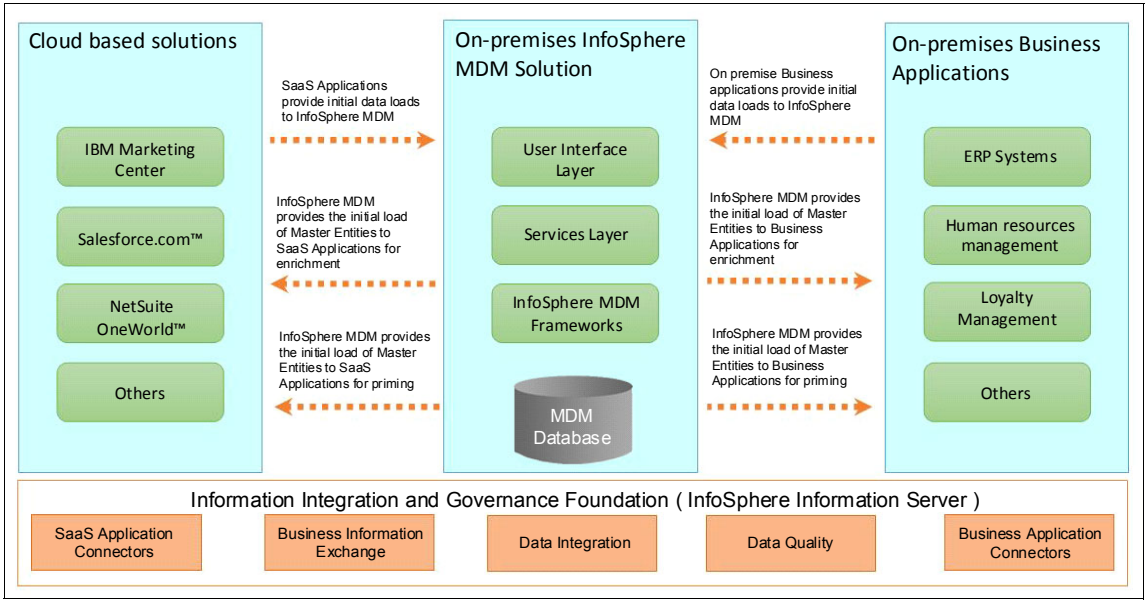


Figure 4-1 Pre-production setup for initial loads

Initial load of master records from InfoSphere MDM to one or more target SaaS applications

Given that InfoSphere MDM has the trusted, authorized, 360-degree view of master data, enterprises want to make use of InfoSphere MDM as a trusted source of information in the SaaS application they subscribe to.

An enterprise scenario used here is that the company has an existing InfoSphere MDM solution deployed and that they are moving forward to cloud initiatives to subscribe to SaaS applications, such as CRM system using Salesforce.com. This enterprise wants to use the “single view of customer” from the in-house InfoSphere MDM solution to power the Salesforce.com master entities, such as Account, Contacts, and so on.

In this type of scenario, there should be options provided to consistently load and prime Salesforce.com with the InfoSphere MDM mastered golden records when integrating Salesforce.com with InfoSphere MDM initially.

The SaaS solution also must be treated as a source after the solution implementation goes live.

Initial load of data from one or more SaaS applications into InfoSphere MDM

This flow should be considered when an enterprise is dealing with a scenario wherein they have an existing set of SaaS applications that they have subscribed to but where the master entities in these SaaS applications are not governed by their on-premises InfoSphere MDM deployment. The other alternative scenario can be that the enterprise does not have an existing MDM solutions and are now deploying an InfoSphere MDM solution that is expected to govern the on-premises business applications and SaaS applications. In both scenarios, the cloud-based SaaS application should be treated as a source system to InfoSphere MDM from the beginning.

There should be options provided to consistently and efficient load the data from the SaaS applications to the on-premises InfoSphere MDM system when integrating these two systems.

InfoSphere MDM can enrich its golden records with the data from the SaaS applications and provide data governance for the SaaS applications.

For the implementation, the enterprises that subscribe to Saleforce.com CRM can load the customer data from Saleforce.com CRM into the in-house InfoSphere MDM system. The data is treated as one of the sources for master data and uses the InfoSphere MDM capabilities such as de-duplication, cleansing, standardization, and so on for the customer data.

Special scenarios for initial load—SaaS application migration

There is a potential future scenario that is not in the scope of this compilation; however, we outline it here for awareness.

SaaS applications marketplace is going to see fierce competition as adoptions increase. There might be scenarios associated with what can be called “SaaS churn,” much like how a customer might switch from one Telecom provider to another. Enterprises might want to switch from one SaaS application in a particular business space to another. The reasons can be multiple, more capabilities, cheaper price, better performance, and so on. An enterprise can have an InfoSphere MDM solution running on-premises but might want to switch the SaaS CRM solution provider. In this case, the enterprise can expect a solution to migrate all the data, rules, and other implementation artifacts from one SaaS CRM provider to another while not incorporating many changes into the MDM implementation.

InfoSphere Information Server and other tools are capable of handling these kind of scenarios. However, given that this scenario is not prevalent, external artifacts in this area are a future consideration.

Patterns for batch and incremental data load

The incremental and batch loading can be used for the following types of business needs:

- Real-time data integration is not required and data can be moved in batches during off-peak hours.
- A large marketing campaign is initiated that triggers a one-time, large amount of data generated, and the company can move bulk data in batches.
- Data reconciliation is required in the event of job failures. When intelligent handling of job failures and restarts have to be incorporated in the data integration procedures that are deployed, you can use incremental data load to catch the missed data when job failed.
- Mergers and acquisitions when partial initial loads are required for the new sources.

InfoSphere MDM and the SaaS application are usually part of the mission or operations critical systems. So consider system availability when designing batch load strategies.

The patterns for batch and incremental data load also cover the following main data integration flows:

- Batch and incremental data load from InfoSphere MDM to one or more SaaS application
- Batch and incremental data load from one or more SaaS applications to InfoSphere MDM

The payload of data moved is lesser than the amount of payload to be dealt with during initial loads.

Batch and incremental data load from InfoSphere MDM to one or more SaaS applications

This data integration flow primarily deals with keeping the data in sync between the on-premises InfoSphere MDM system and one or more SaaS applications that the enterprise uses. In this data integration flow, data from InfoSphere MDM is synchronized into the SaaS application periodically based on a schedule set up for the data movement. This process ensures that the SaaS application has the up-to-date data from InfoSphere MDM at the end of the data movement schedule.

For handling this flow, a scalable and high performing method to identity the delta data in InfoSphere MDM from the last successful synchronization operation is required that in turn prepares the payload to be transformed and dispatched to the target SaaS application. The data load method must also be able to read all of the superset data required by multiple target SaaS applications. The superset data is read once only from the InfoSphere MDM system, transformed as required, and then dispatched to the target SaaS applications.

Batch and incremental data load from one or more SaaS applications to InfoSphere MDM

This data flow primarily deals with one or more SaaS applications, regarded as a source by the InfoSphere MDM system, providing periodic updates to an InfoSphere MDM system. This data flow can also result in updates in the InfoSphere MDM system that in turn might have to post data back to one or more SaaS applications. The data posted back can be the match or link information, standardized attributes, and so on. The feeds coming in from the SaaS applications, the return actions resulting in data flows back to the SaaS application, cannot be coordinated; therefore, proper reconciliation functions are also required. The procedures for updating the InfoSphere MDM system with the incoming data should be well optimized as well, along with proper policies for exception management and other considerations.

Patterns for runtime data integration

Runtime data integration flows typically deal with real-time data integration requirements. An InfoSphere MDM system is the authoritative source for master data, and it provides services that can be consumed by business applications, ETL tools, SaaS applications, and many other consumers in the enterprise.

The data of the InfoSphere MDM system is part of many business transactions and the system provides services for consuming applications to control the creation, management, quality, and access of master data. For example, as part of a process to add a new customer, a business system invokes an InfoSphere MDM service in real time to validate whether this customer is a new customer or an existing customer within the enterprise.

The InfoSphere MDM system can be customized to implement certain governance policies wherein updates done to critical data elements in the source systems and must be approved by an MDM data steward.

Terminology: The term *runtime* here refers to the running data integration transactions whenever is needed. The terms *runtime* and *real time* are used interchangeably further in this document.

The patterns for runtime data integration cover two main flows with the following additional considerations:

- InfoSphere MDM to SaaS application runtime integration
- SaaS application to InfoSphere MDM runtime integration
- Characteristics of 2-way integration between InfoSphere MDM and SaaS applications

InfoSphere MDM to SaaS application runtime integration

In this data flow, any create, read, update, and delete (CRUD) operation that causes change on the master golden record in the InfoSphere MDM system must be synchronized with the relevant target entities in the SaaS application. Likewise, any business rule which is run as a resultant of a governance policy in the InfoSphere MDM system might require some real-time data movement back to the SaaS application. This can be attributes being updated from another source, which in turn enriches a target entity in a SaaS application.

SaaS application to InfoSphere MDM runtime integration

In this data flow, any new records created, updated, and deleted in the SaaS application must be synchronized with the on-premises InfoSphere MDM system in real time so that the InfoSphere MDM system is up to date with all the source system updates. The consuming system of the InfoSphere MDM system might

have to be alerted if there are changes to the golden master record. Some of the other considerations could be that the SaaS application might want to use certain services provided by the InfoSphere MDM system such as the probabilistic search services, enterprise search services (provided by InfoSphere MDM integration with Watson Explorer), scoring services, and data cleansing services. For more details, see *Building 360-Degree Information Applications*, SG24-8133.

4.1.2 Access integration patterns

Access integration patterns are a category of Integration patterns that deal with certain security requirements when two applications are integration with each other. Here the focus is on how an InfoSphere MDM system and SaaS application integration is implemented to support the security requirements and data authorizations. The InfoSphere MDM system can reside in on-promises infrastructure, on a public cloud, or on a private cloud within the DMZ of the organization.

Based on where the InfoSphere MDM system is hosted, the following access integration patterns are important considerations:

- Access management from on-premises applications to SaaS applications
- Access management among applications hosted on cloud

Access management from on-premises applications to SaaS applications

In this pattern, InfoSphere MDM is hosted on premises within the DMZ of the enterprise and is integrated with SaaS applications. WebSphere Cast Iron Cloud Integration is used as the middleware for this integration. It provides connectivity between the InfoSphere MDM and the SaaS applications in the following types of deployments:

- An integration model with physical and virtual appliance

  The connections between the on-premises InfoSphere MDM and SaaS applications are through WebSphere Cast Iron Cloud Integration appliance (physical or virtual) that resides behind the firewall but not within the DMZ. In this type of integration, the access control from SaaS application to InfoSphere MDM is through web service endpoints. The security, data flow, and logging are all handled by the integration appliance, WebSphere Cast Iron Cloud Integration.

- Integration as a service using WebSphere Cast Iron Live

  The connections between the on-premise InfoSphere MDM and SaaS applications are through IBM WebSphere Cast Iron Live multi-tenant service provider. WebSphere Cast Iron Live makes use of WebSphere Cast Iron

Cloud Integration Secure Connector to connect to the on-premises InfoSphere MDM. WebSphere Cast Iron Cloud Integration Secure Connector is installed on premises within the firewall. WebSphere Cast Iron Cloud Integration Secure Connector builds a secure channel between WebSphere Cast Iron Live and the local network and helps to bridge secure communication between the Internet and intranet.

Figure 4-2 shows the schematic diagram of WebSphere Cast Iron Cloud Integration.

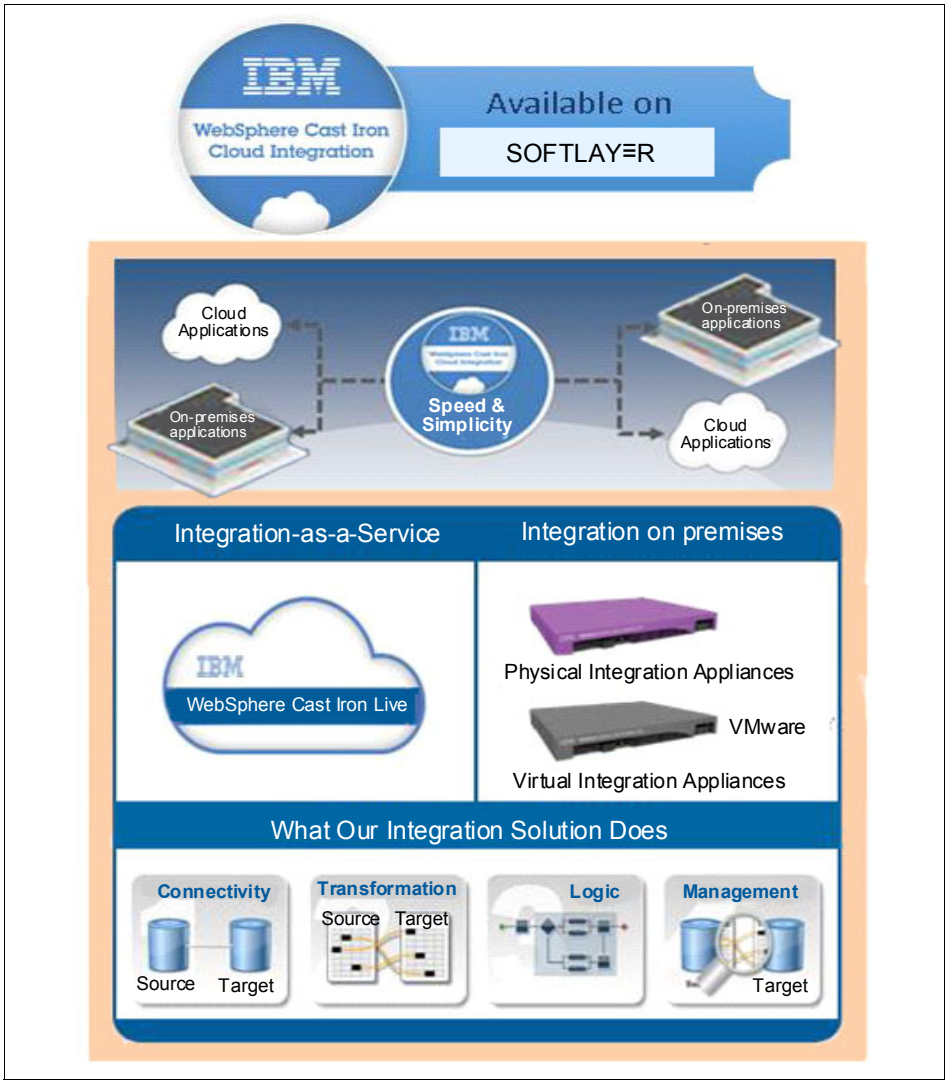


Figure 4-2 The schematic diagram of WebSphere Cast Iron Cloud Integration

Access management among applications hosted on cloud

In this pattern, InfoSphere MDM is hosted on cloud such as IBM Softlayer, Amazon EC2, or Rackspace. The security for InfoSphere MDM is governed by the cloud platform provider and the application server. In this pattern, InfoSphere MDM endpoints are available on the IaaS provider virtual machine but might still be secured for general access on public IPs.

The application-level security is provided by InfoSphere MDM.

Security considerations and capabilities

For InfoSphere MDM and SaaS integrations, the security is provided at different levels based on how and where InfoSphere MDM is hosted. In the on-premises scenario, InfoSphere MDM is integrated using WebSphere Cast Iron Cloud Integration as the middleware. The security is configurable at InfoSphere MDM, at WebSphere Cast Iron Cloud Integration, and at SaaS application.

The SaaS application usually provides access control to its objects. Take Saleforce.com as an example. The Salesforce.com objects are logically mapped to the InfoSphere MDM data model business objects. In Salesforce.com, users, groups, and authorization rules can be defined to provide specific access rights to the objects.

WebSphere Cast Iron Cloud Integration also has user and groups for managing the projects. This security does not deal at data access level but more at managing the orchestration projects and the authorization to run the services provided by WebSphere Cast Iron Cloud Integration. WebSphere Cast Iron Cloud Integration exposes the InfoSphere MDM services as provider services. WebSphere Cast Iron Cloud Integration can define who can InfoSphere MDM provider services through users and groups.

The next section covers the authentication and authorizations for InfoSphere MDM.

When InfoSphere MDM hosted on external clouds, the security is configurable at the following layers:

- Specific to InfoSphere MDM application: InfoSphere MDM security
- Specific to cloud integration middleware: WebSphere Cast Iron Cloud Integration security
- Specific to SaaS application: SaaS application security

If InfoSphere MDM is hosted on external cloud, the WebSphere Cast Iron Cloud Integration is an optional component because there is no traditional firewall. Thus, WebSphere Cast Iron Cloud Integration becomes an optional component.

The external public cloud that provisions InfoSphere MDM has additional authorization guidelines about user access to InfoSphere MDM.

InfoSphere MDM security

InfoSphere MDM addresses the authentication, authorization, and audit aspects of security for MDM. At a high level, authentication, authorization, and audit cover the establishment of a consumer's identity and access rights to services as well as providing audit trail of who did what respectively.

InfoSphere MDM security covers the requirements around how MDM entity data can be created, accessed, governed, modified, or deleted. Almost all the InfoSphere MDM editions depend on a user registry in the application container, that is, WebSphere Application Server. WebSphere Application Server in turn can delegate the user authentication to any directory sever or any other supported user security solution.

The authorizations are specific to the MDM editions but in general meet the goals of secured and trusted interactions with the system from all entry points.

If we take the example of InfoSphere MDM Standard Edition (InfoSphere MDM SE) and its integration with Salesforce.com, a set of services generated by the InfoSphere MDM enterprise service-oriented architecture (eSOA) toolkit form the primary set of APIs for use in the integration.

Authorization: When the eSOA Toolkit of InfoSphere MDM SE is used to develop web services, the credentials are stored in the `madclient.properties` property file. The WebSphere Cast Iron Cloud Integration orchestration calls the InfoSphere MDM service using the credentials specified in `madclient.properties`.

InfoSphere MDM treats the request as trusted request and provides the response based on the authorization configuration configured for this particular user. Any further authorization must be handled at the security configuration at Saleforce.com.

To know more about the InfoSphere MDM security, see the InfoSphere MDM Information Center at:

`http://ibm.co/1nev2CX`

WebSphere Cast Iron Cloud Integration security

WebSphere Cast Iron Cloud Integration offerings have different form factors such as physical appliance, virtual appliance, and a SaaS offering called WebSphere Cast Iron Live. Each of the form factors of the WebSphere Cast Iron Cloud Integration provides security at different layers as applicable.

All form factors have users and groups that control the configuration and deployment of WebSphere Cast Iron Service orchestration projects.

WebSphere Cast Iron Live, that is the SaaS form factor, has one additional component to manage secure connectivity—WebSphere Cast Iron Cloud Integration Secure Connector. Secure Connector also adheres to various industry specific security standards for connectivity and data transfer and is deployed behind the enterprise firewall. For the physical and the virtual appliance form factor, the secure connectivity capabilities are built in.

To understand more about WebSphere Cast Iron Live security and Secure Connector, see *Getting Started with IBM WebSphere Cast Iron Cloud Integration*, SG24-8004.

4.1.3 Process integration patterns

This section describes the aspects of process integration patterns and outlines a set of processes, in the context of InfoSphere MDM Standard Edition integration with Salesforce.com, where this type of pattern comes into play.

Process integration patterns primarily deal with scenarios where more than one application or solution comes together to run a business flow. It can entail runtime data movement alongside the execution of certain business logic as well, such as validation rules, governance policies, and so on. The aspects of process integration patterns can be understood well and in context with the “search before create” use case, where InfoSphere MDM SE participates in an account creation process in Salesforce.com.

“Search before create” use case with an InfoSphere MDM powered search

SaaS applications typically are not powered with strong capabilities to avoid de-duplication at the root processes. If we take the example of Salesforce.com, its de-duplication capabilities during creation of a Contact Object cannot do complex probabilistic matching. It does not have a view of the enterprises business applications that might have this contact details already, which is where a “search before create” subprocess powered by InfoSphere MDM SE helps meet the business goal of de-duplication at entry point for master data.

InfoSphere MDM SE search uses the InfoSphere MDM probabilistic matching capability. InfoSphere MDM SE search enhances the Salesforce.com search so that you can find variants of account or contact names with typographical errors, and account or contact names that sound similar. By finding the existing customer records, the point of entry can be governed efficiently and accurately, ultimately avoiding the act of creating duplicate records in Salesforce.com.

This capability also helps enterprises locate data that is contained not only within the Salesforce.com system, but also within other systems in the enterprise. The breadth of this data gives Salesforce.com users an enterprise wide, 360-degree view of customers. The enhanced search and resulting view can improve productivity with access to customer data and account contacts that is necessary in day-to-day activities, such as preparing for customer meetings or other sales activities.

The accounts and contacts in Salesforce.com and InfoSphere MDM Server must be synchronized to ensure that no latest information is missed when data is accessed. Any account and contact that are created, updated, and deleted must be synchronized with the on-premises InfoSphere MDM system.

To delve a littler deeper into this use case, the Salesforce.com Account or Contact creation process powered by InfoSphere entails the introduction of the following types of subprocesses:

- “Search before create”
- Subprocesses to synchronize records from Salesforce.com to InfoSphere MDM when add, update, and delete of records happens in Salesforce.com
- Synchronize Salesforce.com updates to MDM in batch (optional)

“Search before create” subprocess

A Salesforce.com user wants to create an opportunity for an account. Before creating an account, the user wants to make sure that the account does not exist in Saleforce.com, and the “Search before create” subprocess has to be initiated.

Figure 4-3 outlines the Account creation process augmented by the InfoSphere MDM Powered probabilistic search capabilities.

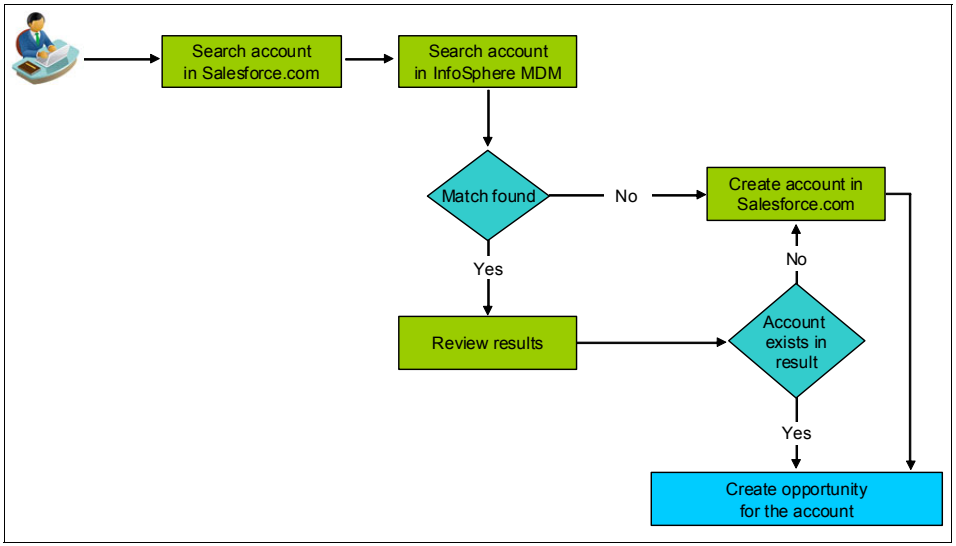


Figure 4-3 Integration flow using the "search before create" subprocess

The process is as follows:

1. Salesforce.com user starts the account search with an InfoSphere MDM Powered Search window.
2. The customizations supporting the InfoSphere MDM powered Search window in Salesforce.com call out to APIs in WebSphere Cast Iron serving the probabilistic search API of InfoSphere MDM to search for these account details.
3. If no results are found, the user creates a new account and later creates the opportunity for the newly created account.
4. If InfoSphere MDM returns matching accounts, the user reviews and validates the accounts.
   - If the user finds out that accounts listed does not match for the user is intending to create, the a new account creation process is initiated and later a new opportunity is created for the newly created account.
   - If the user indeed finds the account intended to be created, the new Opportunity is created for the existing account.

The newly created account in Saleforce.com must be synchronized with InfoSphere MDM.

Subprocesses to synchronize records from Salesforce.com to InfoSphere MDM

Given that an InfoSphere MDM solution has to govern the master entities in Salesforce.com, any add, update, or delete of accounts or contacts in Salesforce.com must be synchronized with the on-premises InfoSphere MDM system. This subprocess can be run in real time or near real time or can be pooled to be run as batch processes on a given schedule.

Real-time subprocess invocation

The real-time process typically operates within the same transaction in all the systems and is not advised for various reasons, such as performance and availability.

Neal real-time subprocess invocation

Salesforce.com provides the pooling API to periodically poll for any changes to objects in Saleforce.com. In a scenario where other cloud applications are used and where API polling API is not provided, the changes can be published to an external queue and WebSphere Cast Iron Cloud Integration orchestration can start the same process to read from the queue.

Figure 4-4 illustrates the example process flow for data synchronization on add, update, and delete of objects in Salesforce.com.

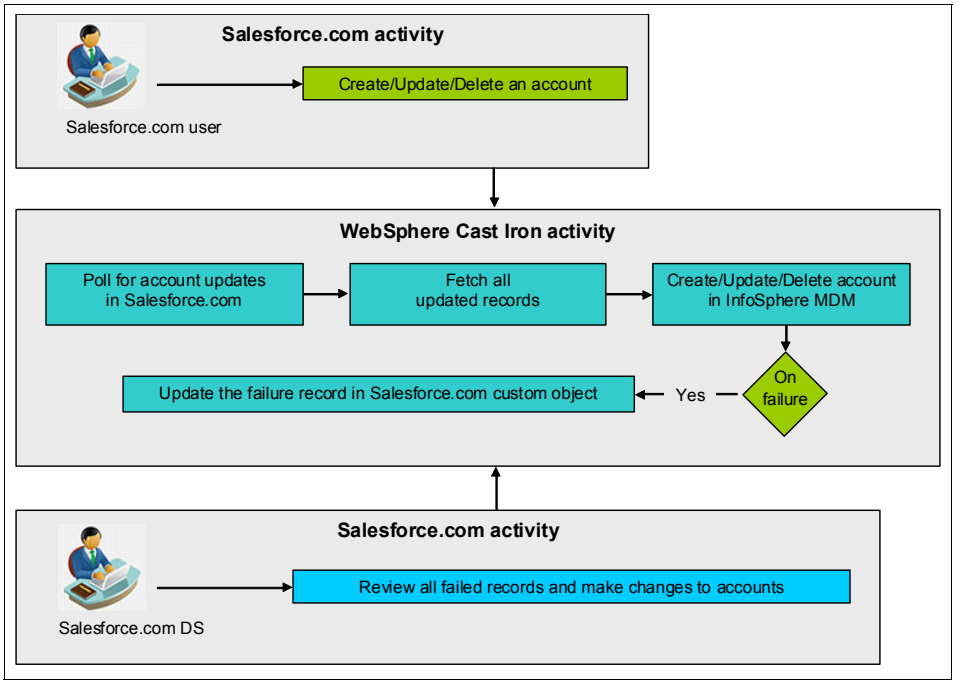


Figure 4-4 Process flow for data synchronization on add, update, and delete of objects in Salesforce.com

For Salesforce.com, this near real-time subprocess invocation for data synchronization has the following steps:

1. Salesforce.com user performs create, update, or delete operations in Salesforce.
2. WebSphere Cast Iron Cloud Integration orchestration job uses the Salesforce.com polling API to check for any add, update, or delete of objects in Salesforce.com.
3. The job fetches all the recently modified objects in Salesforce.com.
4. Based on the action (create, update, or delete) type, the corresponding business transaction is called in InfoSphere MDM to synchronize the changes from Salesforce.com to InfoSphere MDM.

   On failure of the transaction, the object details and failure messages is logged to a table or custom object in Salesforce.com. A custom object is just another object created in Saleforce.com to capture failure details for error handling.

Later, the data steward in Saleforce.com can check the error logs in the custom object and takes appropriate action to rectify the failure and synchronize the data back to the InfoSphere MDM system.

Invocation as a batch subprocess

With this batch invocation approach, the data in synchronized in batch either daily, weekly, or monthly based on the business requirements. This approach is typically use when the volume of updates on Salesforce.com is heavy. For the batch synchronization, IBM InfoSphere Information Server Sales Force Pack can be used to read the data in bulk from Saleforce.com and then synchronized into InfoSphere MDM.

Figure 4-5 describes the process flow the batch updates.

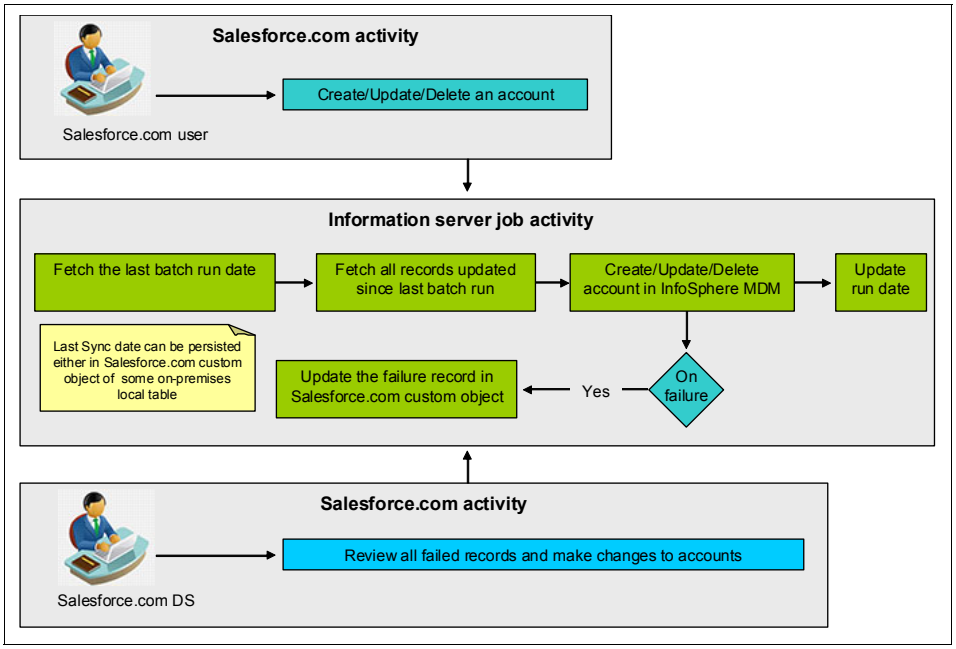


Figure 4-5 illustrates the batch data synchronization process

Batch process updates include following steps.

1. A Salesforce.com user creates, updates, or deletes objects in Saleforce.com.
2. A InfoSphere Information Server batch load job fetches the last updated date and time of previous batch job run.
3. A InfoSphere Information Server batch load job fetches all the objects that were modified after the previous batch job run.

4. Based on the action type (create, update, or delete), the corresponding business transaction is called in InfoSphere MDM to synchronize the changes from Salesforce.com to InfoSphere MDM.
5. On failure of the transaction, the object details and failure messages is logged to a table or custom object in Salesforce.com.
6. The current job run date and time are persisted for next job run. The date and time can be persisted in a Saleforce.com custom object or any tables within the enterprise.
7. Later, the data steward in Saleforce.com can check the error logs in the custom object and takes appropriate action to rectify the failure and synchronize the data back to the InfoSphere MDM system.

4.2 Solution architecture for integrating MDM systems and SaaS applications

This section describes the suggested solution architecture for integration of InfoSphere MDM with SaaS applications. The solution architecture is outlined using the integration of InfoSphere MDM SE with Salesforce.com.

Figure 4-6 depicts the architecture for three type of integration patterns between Salesforce.com and the on-premises InfoSphere MDM server.

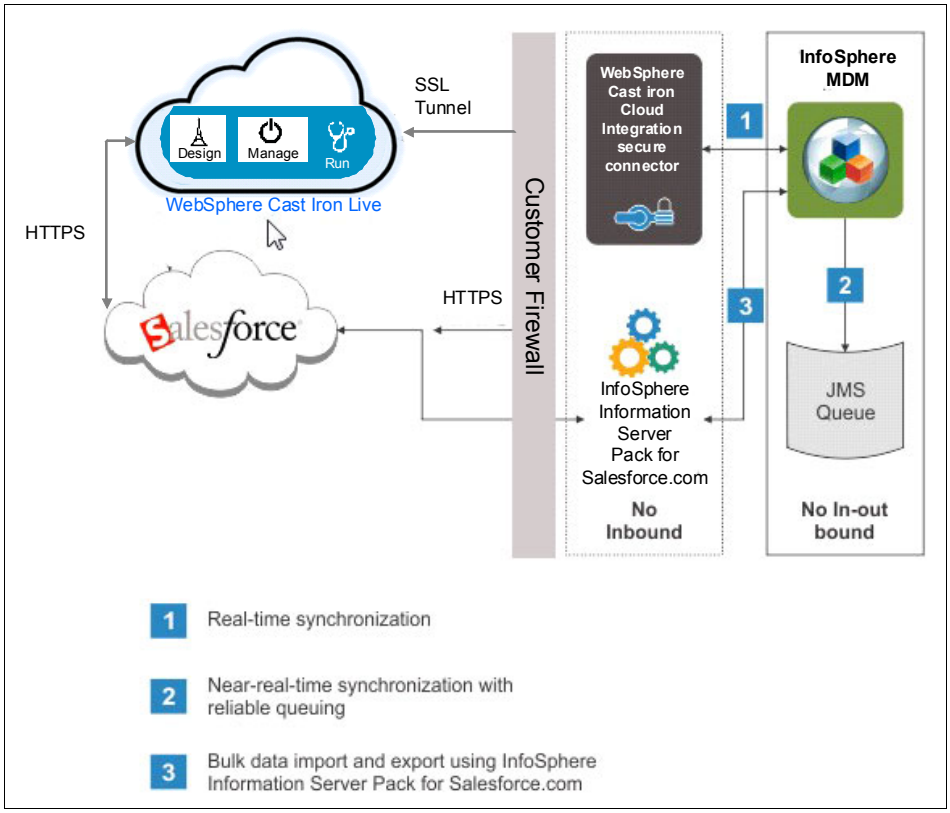


Figure 4-6 Overall solution architecture for InfoSphere MDM SE and Salesforce.com integration

Figure 4-6 is a semantic layout of the solution architecture:

- Network: Within the DMZ of the enterprise, the network is generally further classified as follows:
  - Red network with both inbound and outbound access
  - Network with only outbound access
  - Network with no inbound and outbound access
- Deployments: InfoSphere MDM SE is typically hosted in network without inbound and outbound access. WebSphere Cast Iron Secure Connector and IBM InfoSphere Information Server with Salesforce.com Service Pack are hosted in network with outbound access.

- Integrations: Runtime integration between Saleforce.com and InfoSphere MDM happens through the Cast Iron Live and standard queue infrastructure in WebSphere Application Server in which InfoSphere MDM SE is deployed. IBM InfoSphere Information Server Service Pack for Salesforce.com is used for bulk import and export of data from InfoSphere MDM to Salesforce.com and vice versa.

  Figure 4-6 also outlines the three types of integration patterns between Salesforce.com and the on-premises InfoSphere MDM Standard Edition.

- Real-time integration

  WebSphere Cast Iron Live is used to integrate the on-premises InfoSphere MDM SE with Salesforce.com. The Salesforce.com CRM application makes a call to InfoSphere MDM SE business web services using a WebSphere Cast Iron Cloud Integration web services provider that is deployed on the Cast Iron Live cloud. Similarly, InfoSphere MDM SE can perform real-time call Salesforce.com services through WebSphere Cast Iron Cloud Integration.

- Near real-time asynchronous integration

  Asynchronous integration is typically used to synchronize Salesforce.com CRM data with InfoSphere MDM SE and vice versa. WebSphere Cast Iron Live is used as the interface between Saleforce.com and the on-premises InfoSphere MDM system. Salesforce.com on its own does not maintain any queue such as JMS or IBM WebSphere MQ. The queue lifecycle has to be managed out side Salesforce.com. This can be done either using WebSphere Message Broker or Java Messaging Service (JMS).

- Bulk import and export

  Bulk import and export of InfoSphere MDM SE and Saleforce.com objects is accomplished using InfoSphere Information Server Pack for Salesforce.com(IIS-SF Pack). IIS-SF Pack uses Salesforce.com bulk API to import and export the data. This feature is also used to import and export data from InfoSphere MDM SE. When exporting the data from InfoSphere MDM SE, the MDM Connector stage in the InfoSphere Information Server IBM DataStage® component (InfoSphere DataStage) can be used to read data using the configured the composite view from InfoSphere MDM SE. The output of the InfoSphere Information Server MDM Connector stage is loaded to Salesforce.com using Information Server Service Pack. In bulk import, IIS-SF Pack reads bulk data from Salesforce.com and the output is transformed to the InfoSphere MDM SE input format using InfoSphere DataStage jobs. The transformed data is then loaded into InfoSphere MDM SE using the MDM Connector Stage in InfoSphere DataStage configured to run the `MemPut` API.

See the following link for more details about the MDM Connector stage:

http://www-01.ibm.com/support/knowledgecenter/SSWSR9_11.3.0/com.ibm.swg.im.mdmhs.mdmstages.doc/topics/c_mdmconnectorstage.html?lang=en

4.3 Considerations for bulk load operations

This section describes the requirements for bulk load operations and introduces the capabilities provided by IBM InfoSphere Information Server for bulk operations. In 4.2, “Solution architecture for integrating MDM systems and SaaS applications” on page 80, the description of “bulk import and export” provides a cursory view of bulk load operations. In this section, a deeper description of bulk data operations between the on-premises InfoSphere MDM SE and Salesforce.com is provided by using a InfoSphere DataStage job as a backdrop. To begin with, a deep dive into the key components is provided first, followed by a quick snapshot of the requirements.

4.3.1 IBM InfoSphere Information Server Pack for Salesforce.com

IBM InfoSphere Information Server Pack for Salesforce.com is an add-on offering for IBM InfoSphere Information Server that can be used in InfoSphere DataStage or InfoSphere QualityStage® jobs to help with connectivity and data integration with Salesforce.com.

The key business value of this add-on includes:

- Provide a secure, scalable, and performing method for high-volume batch loads of Salesforce.com data.
- Provide secure methods to synchronize Salesforce.com data with on-premises applications and other consumers.
- Enable on-premises analytical applications to use data in Saleforce.com.

For more information about IBM InfoSphere Information Server Pack for Salesforce.com, see:

http://www-03.ibm.com/software/products/en/infosphere-information-server-pack-salesforce

4.3.2 IBM InfoSphere DataStage

InfoSphere DataStage provides MDM APIs and utilities for the ETL operations for customer data. Installed with InfoSphere DataStage, MDM Connector stage is

one of the options to use the APIs to retrieve (read) and insert or update (write) MDM data.

Similar to other connectors in InfoSphere Information Server, the MDM Connector stage is a polymorphic stage that allows you to run interactions with the MDM instance for reading and writing data. The MDM Connector stage uses the `MemGet` API for data extraction jobs and the `MemPut` API for data ingestion into InfoSphere MDM system. The combination of MDM Connector stage and Information Server pack for Salesforce.com can be used to bulk read, cleanse, transform, and load data from InfoSphere MDM and Salesforce.com or to extract data from Salesforce.com.

For more information about MDM Connector Stage, see:

`http://ibm.co/1vPFHV9`

4.3.3 Initial extract, and bulk and incremental load requirements

This section describes the requirements of the initial extract, bulk and incremental load.

The following scenarios illustrate where the initial extracts are required:

- An enterprise is already subscribed to Salesforce.com CRM and wants to utilize the benefit of InfoSphere MDM. For this, they would like to load Salesforce.com records as an additional source into InfoSphere MDM Standard Edition. A job is required for extracting data from Salesforce.com and loading the extracted data into InfoSphere MDM while meeting performance and remediation requirements.
- An enterprise has InfoSphere MDM Standard Edition deployed which is the system of record for the master entities, now the enterprise would like to use this trusted source of single views in the Salesforce.com CRM. In this scenario, an ETL job that does a full extract from InfoSphere MDM and load the extracted data into the Saleforce.com entities is required. This job must meet the performance and remediation needs.

The following scenarios illustrate where the bulk and incremental loads are required:

- When the solution goes into production, InfoSphere MDM will get continuous or periodic updates from multiple source systems. To refresh the single view of an MDM entity that the Salesforce.com might have used earlier, a fresh synchronization is required.
- When the system is in production, there is also a possibility that some attributes are updated in Salesforce.com. Given that Salesforce.com will be

respected as a source by the InfoSphere MDM system, the Salesforce.com updates must be synchronized back to InfoSphere MDM.

Data synchronized: Bulk and incremental load might not be required if runtime integration can take care of keeping data synchronized between InfoSphere MDM SE and Salesforce.com. Runtime integrations take care of keeping the data synchronized by tracking the create, update, or delete events in either of the applications and then using these events to synchronize data either in real time (service calls) or near real time by using messaging.

4.3.4 Sample export job

IBM InfoSphere Information Server Pack for Salesforce.com is the key component for bulk operations between InfoSphere MDM and Salesforce.com. IBM InfoSphere Information Server Pack for Salesforce.com must be installed in the IBM InfoSphere Information Server instance. After installation, the Salesforce.com connector appears in the palette of the IBM InfoSphere DataStage and QualityStage Designer Client.

The MDM Connector stage, available with IBM Information Server V11.3, serves the purpose of reading and writing data to InfoSphere MDM Standard Edition.

This section includes a sample export job with InfoSphere Information Server Pack for Salesforce.com and MDM Connector stage. Figure 4-7 depicts the export job flow of InfoSphere MDM entities to Salesforce.com.

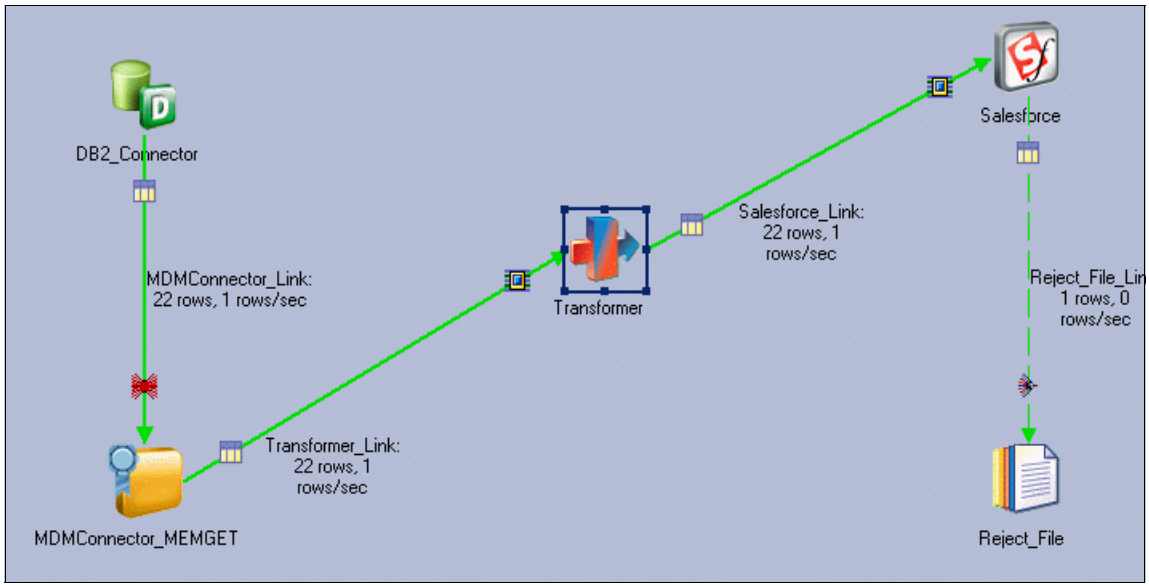


Figure 4-7 Job flow for bulk export of InfoSphere MDM entities to Salesforce.com

The job flow is outlined as follows:

- The DB Connector stage reads all the keys (entity IDs) based on a predefined criteria of the records that must be extracted for export.
- The MDM Connector stage, with `MemGet` API, takes each key provided by the DB Connector stage and reads the MDM entity records.
- The Transformer stage, configured with mappings of InfoSphere MDM attributes to Saleforce.com Contact and Account objects attributes, does the data preparation and makes the payload ready for the write to Saleforce.com.
- The Pack for Saleforce.com stage writes the “prepared” payload of MDM records to Saleforce.com Contact and Account objects using secured transport layer streams of HTTPS protocol.
- The Reject stage logs the rejected record. The remediation routines can be run to check and reload the rejected records.

For bulk imports, the MDM Connector stage that is configured to run with the `MemPut` API can be used to read data from Salesforce.com.

5

A sample InfoSphere MDM and SaaS application integration scenario

It is important for organizations to avoid data duplication at the point of entry to their information systems. Doing so reduces data maintenance, data inconsistencies, and storage requirements. You can avoid data duplication by providing your employees with a real-time search utility that can be used to check for duplication of records at the point of entry. This capability, when incorporated into daily processes, enables your organization to focus on core customer initiatives and to prevent data duplication.

This chapter describes a real-world integration solution scenario for the real-time search capability with IBM InfoSphere Master Data Management (InfoSphere MDM) and Salesforce.com. This scenario demonstrates how InfoSphere MDM, integrated with IBM WebSphere Cast Iron Cloud Integration, provides organizations with a real-time search capability. It includes the following topics:

- Business issue
- Solution architecture

5.1 Business issue

Container Company A is a plastic container manufacturing company that makes and sells plastic kitchen ware. The marketing model of Container Company A is primarily through resellers but also has direct marketing channels. Container Company A and their resellers have stores across the country in most cities. Often, Container Company A and their resellers have overlapping presence in the same city. Each reseller maintains its own record system for product inventory and customer accounts. The systems of records are maintained using disparate products, platforms, and technologies for different locations.

One of the challenges that Container Company A faced was managing its customer accounts and product inventory across all the resellers and direct marketing shops in an unified manner. This problem was fixed when Container Company A deployed InfoSphere MDM on their premises to manage the records and to provide a single consolidated view of these records.

Container Company A also started using Salesforce.com to enable its door-to-door sales consultants to easily create and track customer opportunities and leads. The problem faced in such a setup is the overlap of customer data and opportunities across its disparate systems. Sales consultants from the direct and indirect marketing channels can create duplicate customer leads or place duplicate orders.

Duplicate records can create the following issues:

- Difficulties in lead conversion leading to loss of productivity
- Difficulties in order tracking leading to lost or wrong orders
- Difficulties in after sales customer service leading to low customer satisfaction
- Difficulties in calculation for sales forecast and inventory maintenance
- Difficulties in bonus allocation to sales consultants

Figure 5-1 illustrates the business challenges that the container Company A faced.

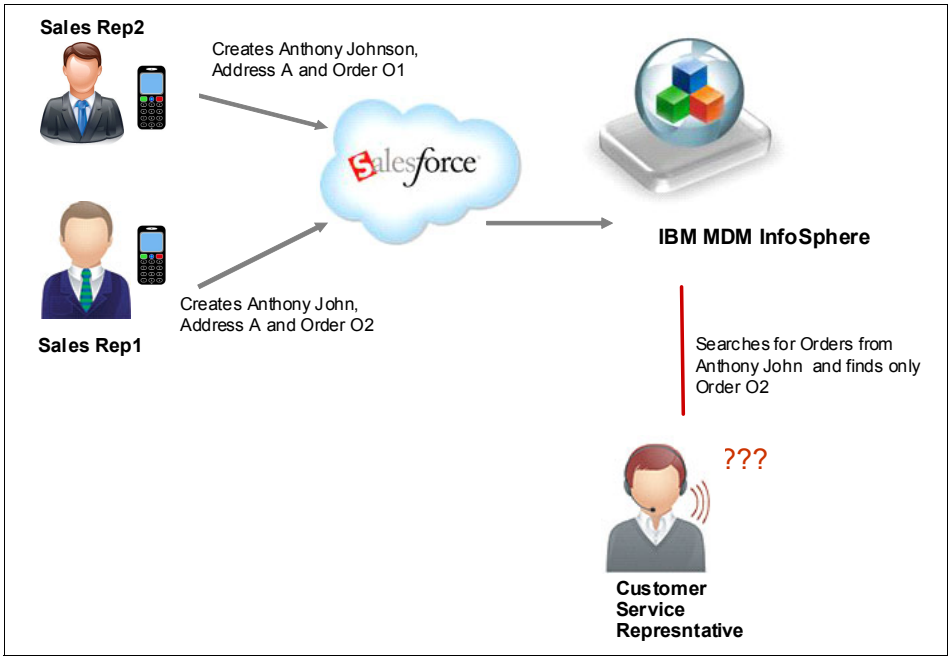


Figure 5-1 Business issues

5.1.1 Solution for the business issues

The solution to the business problem of Container Company A is to eliminate duplicate records at the point of entry by providing the data validation capability.

InfoSphere MDM has a built-in probabilistic search functionality that can be used to search customer accounts from other information system, including Salesforce.com, giving a 360-degree view of the customer data. This function finds accounts with exact match and also variants based on the search parameters. In addition, it accommodates typographical errors and fetches records that are similar. This function is best suited for real-time search where inputs are similar but not always an exact match. InfoSphere MDM exposes the search functionality using web services SOAP API.

Salesforce.com allows users to build user experience on demand from the Force.com platform. The VisualForce framework can be used to develop the custom user interfaces (UI) on Force.com, eliminating the need for external software tools. This solution use VisualForce to extend the Salesforce.com UI to build custom pages for customer record search.

The customer service representative or sales consultants of Container Company A can use the custom UI to search customer account information across all the information systems in the infrastructure. This utility is then incorporated into the process for sales consultants to create new leads. The request is routed to InfoSphere MDM to invoke its probabilistic search. The SOAP API calls contain the search results that are then displayed in the Salesforce.com custom UI extension. Figure 5-2 shows the solution outline.

With the search utility included in Salesforce.com application, the sales consultants can use this real-time utility to search for existing customer or order records. If the search reveals that a similar customer record does not already exist, then the consultant can create a new one from Salesforce.com.

The solution can also be used by the customer service representatives to find the correct customer record during after-sales support. By eliminating duplicate records, this solution provides accurate customer records view to support business needs and increasing productivity.

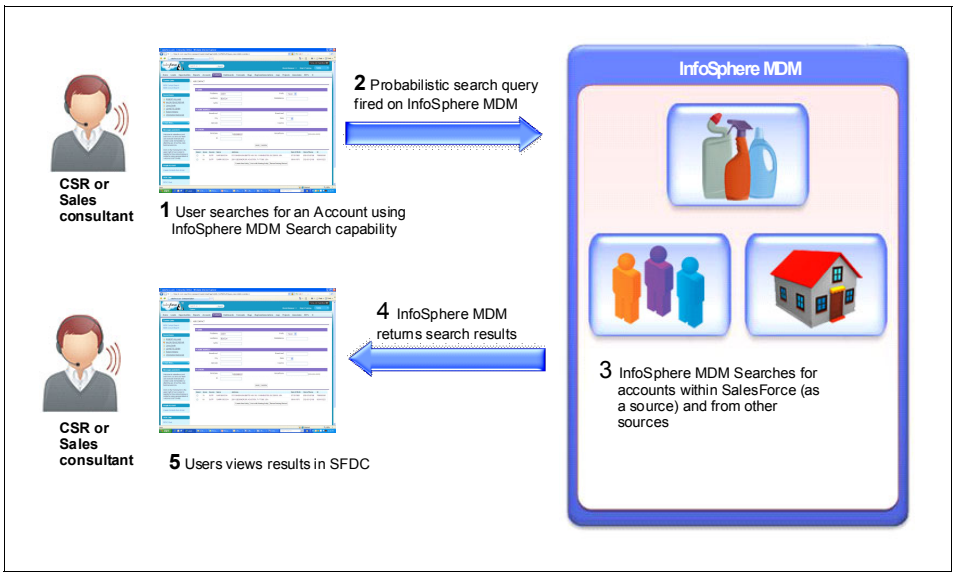


Figure 5-2 Solution outline

5.1.2 Relevance of the solution to Container Company A

This solution capability enables Container Company A to optimize the following client-focus initiatives:

- Reducing duplicate customer and lead records by identifying duplication at the point of entry.
- Finding correct customer information faster and providing a holistic and complete view of customer data.
- Enabling business clients to enrich the customer data hosted on their Salesforce.com application with collective knowledge from internal and external data sources.
- Improving productivity for sales consultants and customer service representatives in their day-to-day activities.
- Improving customer satisfaction through after-sales support.
- Enabling better planning of sales forecast and sales incentives.
- Enabling better inventory planning.

5.2 Solution architecture

Figure 5-3 illustrates a high-level solution architecture with Salesforce.com as the front-end application, WebSphere Cast Iron Live as the middleware, and InfoSphere MDM as the back-end systems that provide services for managing master customer records.

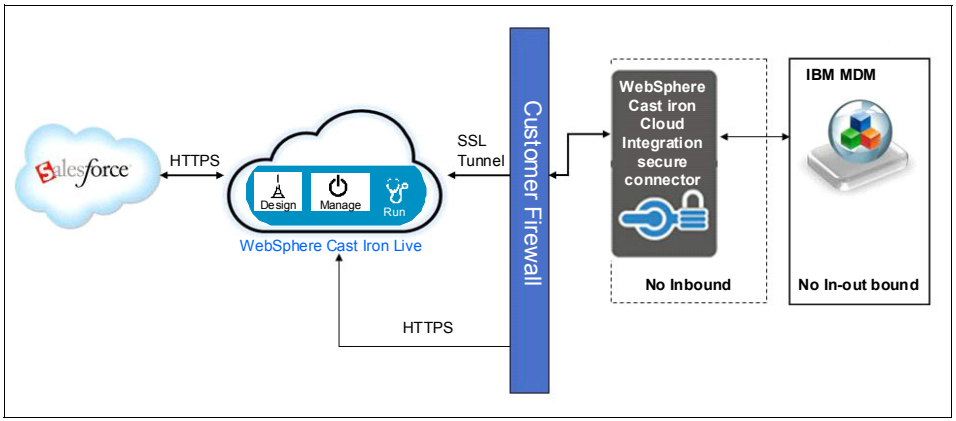


Figure 5-3 Solution architecture

5.2.1 Salesforce.com customization

A custom UI extension on Salesforce.com application elicits the search parameters from the user and invoke the WebSphere Cast Iron Live using a web services call. The communication between Salesforce.com application and WebSphere Cast Iron Live happens on HTTPS secure channel. WebSphere Cast Iron Live calls the correct web services on MDM InfoSphere and returns the results back to Salesforce.com.

5.2.2 IBM WebSphere Cast Iron Live as the middle ware

IBM WebSphere Cast Iron Cloud Integration products enable the user to integrate between cloud application and on-premises application quickly and efficiently. WebSphere Cast Iron Live is a multi-tenant, cloud-based platform for hybrid cloud and on-premises integration. WebSphere Cast Iron Live provides an built-in Designer to helps users create projects and orchestrations to depict business flows.

The orchestration created with WebSphere Cast Iron Designer accepts the incoming request from the Salesforce.com UI. The orchestration routes the web services call to InfoSphere MDM through Secure Connector. Secure Connector resides on the user's premises and provides the required security to the on-premises infrastructure.

Secure Connector is installed behind the firewall in the DMZ of the user's infrastructure. Secure Connector is the only component that communicates with the on-premises InfoSphere MDM servers. WebSphere Cast Iron Live does not talk directly to InfoSphere MDM, which is not part of the DMZ. This communication is achieved by a mechanism where the secure connector always establishes a secure tunnel to the WebSphere Cast Iron Live gateway and not vice versa. This configuration provides an additional layer of security. After Secure Connector establishes the connection and handshakes with the WebSphere Cast Iron Live gateway, the project containing the orchestrations is deployed on WebSphere Cast Iron Live. The activity in the project that talks to MDM InfoSphere web services runs as a remote activity on Secure Connector.

The orchestration that talks to the InfoSphere MDM web services provides a proxy type of setup for InfoSphere MDM web services. The orchestration contains a provider service that acts as a web services receiver. The request is then mapped to invoke the correct web services on InfoSphere MDM. Similarly, the response is mapped back to the WebSphere Cast Iron provider service and to the Salesforce.com application.

For more details about the secure connector and developing the orchestration, see Chapter 6, "Putting it all together" on page 95.

5.2.3 MDM InfoSphere web services

InfoSphere MDM provides a set of services as web services endpoints. Probabilistic matching service is also available as search web service. Using these web services you can search for accounts, contacts, customers, and other entities. These web services can be invoked using standard SOAP calls. In this solution capability, the MEMSEARCH web service API is invoked from WebSphere Cast Iron Live orchestration.

5.2.4 Prerequisite for this solution capability

This solution requires the following prerequisites:

- Bulk data movement from Salesforce.com to InfoSphere MDM

  Every integration between InfoSphere MDM and any application entails a bulk data transfer to the InfoSphere MDM servers. The reasons are discussed in Chapter 2, “MDM solution readiness to participate in integrations with SaaS applications” on page 21. After this initial data movement is complete, the probabilistic search can be used to search information on both the InfoSphere servers and Salesforce.com. This enriches the search functionality and gives a 360-degree view of the customer data. How a bulk data movement between Salesforce.com to InfoSphere MDM servers can be achieved is described in Chapter 4, “Solution architecture” on page 63.

- Near real-time synchronization of data between Salesforce.com and InfoSphere MDM server

  When customer records are created on Salesforce.com, the data should be updated on the InfoSphere MDM servers as soon as possible to provide real-time search results. The longer the gap of data synchronization between two systems, the higher the probability of duplicate records.

Note: The prerequisite for this solution is to make sure that the account information between Salesforce.com and InfoSphere MDM servers are nearly synchronized.

6

Putting it all together

This chapter provides detailed steps to achieve the integration scenario between InfoSphere Master Data Management (InfoSphere MDM) and Salesforce.com as described in Chapter 5, “A sample InfoSphere MDM and SaaS application integration scenario” on page 87.

This chapter includes the following topics:

- Pre-requisites
- Installing the WebSphere Cast Iron Live Secure Connector
- Configuring the project in WebSphere Cast Iron Studio
- Loading sample data
- Testing the scenario
- Running the sample

6.1 Pre-requisites

Before you begin to implement the InfoSphere MDM and Salesforce.com integration, the requirements shown in Table 6-1 must be met.

Table 6-1 Integration prerequisites

Requirement	Description
IBM InfoSphere MDM 11.3	Install InfoSphere MDM Version 11.3.
Salesforce.com account	You must create a Salesforce.com account before you begin the integration.
Force.com IDE	Force.com is an integrated development environment provided by Salesforce.com. You must complete the initial property configuration before the integration project is deployed.
InfoSphere MDM and Salesforce.com integration sample	You must download the sample asset. The download instructions for the sample is provided in 6.1.2, “Downloading integration sample”.
IBM WebSphere Cast Iron Studio 7.0.0.1	WebSphere Cast Iron Studio is used to design the orchestrations for the InfoSphere MDM and Salesforce.com integration.
IBM WebSphere Cast Iron Cloud Integrator 7.0.0.1 Subscription license	You must obtain a valid subscription for WebSphere Cast Iron Cloud Integration Version 6.3.

6.1.1 Downloading WebSphere Cast Iron Studio

You can download WebSphere Cast Iron Studio from the IBM Support Fix Central by choosing **WebSphere Cast Iron Cloud Integration** from the product selector drop-down menu.

6.1.2 Downloading integration sample

To download the InfoSphere MDM and Salesforce.com integration sample from the Samples and Assets site for InfoSphere MDM, complete these steps:

1. Log in to the Samples and Assets site for InfoSphere MDM, and download **IBM InfoSphere Master Data Management IWM Samples 11.3**.
2. Extract the files, and go to the `mdmsamples\mdm_integration` directory.

3. Copy the `com.ibm.mdm.integration.sfdc` directory to a location of your choice.

6.1.3 Creating and deploying a project in Force.com IDE

The Force.com integrated development environment (IDE) is a client application for creating, modifying, and deploying Force.com applications. It allows developers to code, compile, test, package, and deploy applications from the IDE.

A project must be created in the Force.com IDE application. This project contains the InfoSphere MDM and Salesforce.com integration project artifacts which will later be deployed to Salesforce.com. For more information about creating and deploying a project in Force.com IDE, see the *Creating and deploying a project in Force.com IDE* section in InfoSphere MDM V11.3 Knowledge Center at:

`http://www.ibm.com/support/knowledgecenter/SSWSR9_11.3.0/com.ibm.swg.im.mdmhs.sdfcintegration.doc/topics/c_mdmandsfdc_integration.html`

6.1.4 Setting the Salesforce.com configuration properties

The Force.com project for InfoSphere MDM and Salesforce.com integration contains a static configuration Apex class that has all the configuration properties such as Cast Iron endpoints, timeout, and so on defined. The user must update this file based on his environment before deploying the project onto Salesforce.com. For details about the configuration properties, see the *Setting the Salesforce.com configuration properties* section in InfoSphere MDM V11.3 Knowledge Center at:

`http://www.ibm.com/support/knowledgecenter/SSWSR9_11.3.0/com.ibm.swg.im.mdmhs.sdfcintegration.doc/topics/c_mdmandsfdc_integration.html`

6.2 Installing the WebSphere Cast Iron Live Secure Connector

WebSphere Cast Iron Live Secure Connector provides bidirectional data transfer between the endpoints located behind the firewall and WebSphere Cast Iron Live.

To enable a secure connector, complete the following tasks:

- Create the secure connector in Cast Iron Live
- Download, install and start the Secure Connector

A secure connector is defined for each cloud environment for security reasons. If you have multiple environments (development, test, and production), you must configure and install one secure connector per environment. To create the secure connector configuration in WebSphere Cast Iron Live, administrator privileges are required.

6.2.1 Creating a secure connector in WebSphere Cast Iron Live

Complete the following steps to create a secure connector in WebSphere Cast Iron Live:

1. Log in to `https://eval-cloud2.castiron.com/login` as an administrator.
2. Click the environment tab where the secure connector configuration must be created. Figure 6-1 shows an example of a Development environment.

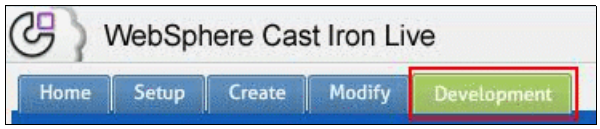


Figure 6-1 WebSphere Cast Iron Live panel when logged in as tenant administrator

3. On the Development tab, click **System** → **Secure Connectors**.

4. Click **New Secure Connector** (Figure 6-2).

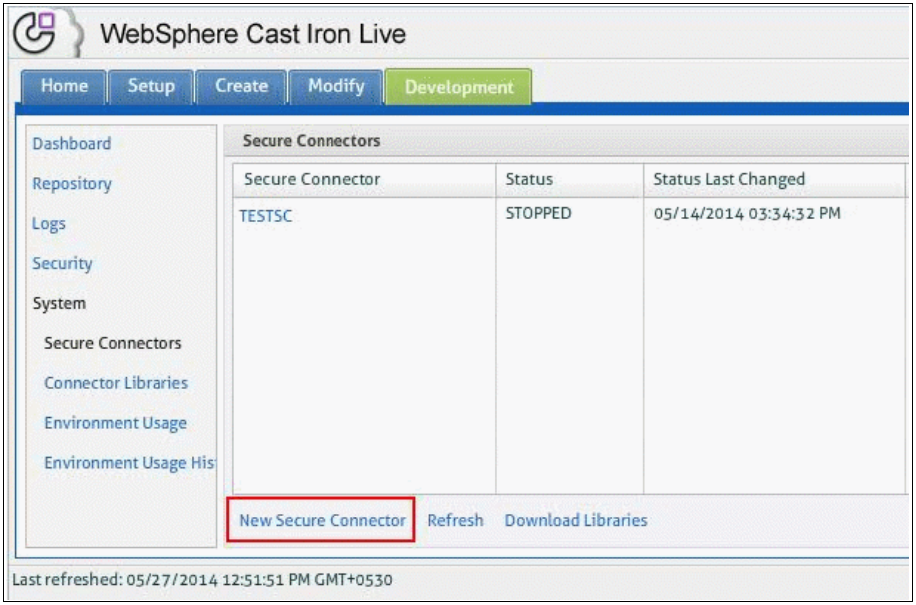


Figure 6-2 WebSphere Cast Iron Live Secure Connector

5. The Create Secure Connector dialog box opens. Enter a name for the secure connector, and click **Save** as shown in Figure 6-3.

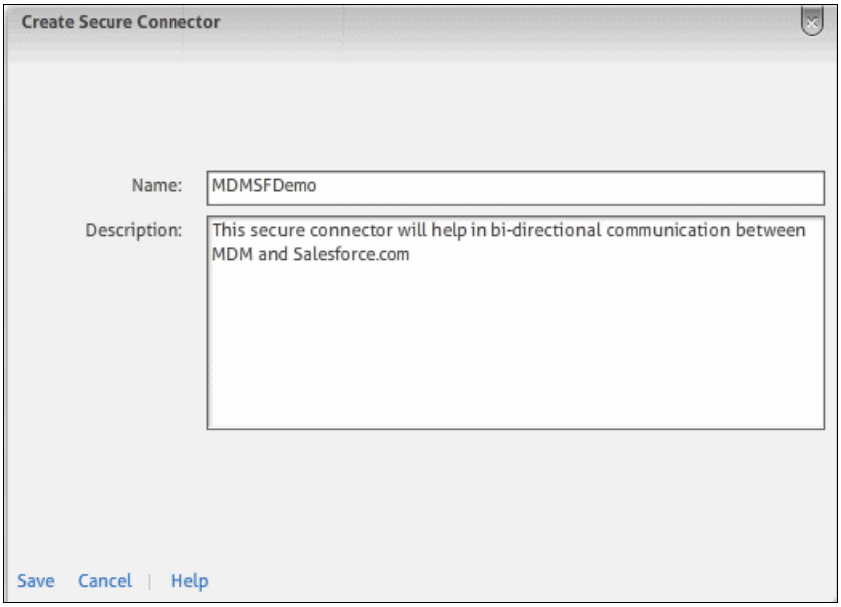


Figure 6-3 Create Secure Connector dialog box

6. Download the Secure Connector configuration file and the installer:
   a. From the Secure Connectors page, click **MDMSFDemo Secure** Connector.
   b. Click **Download Installer** (Figure 6-4 on page 101) and as save the file to your local system.

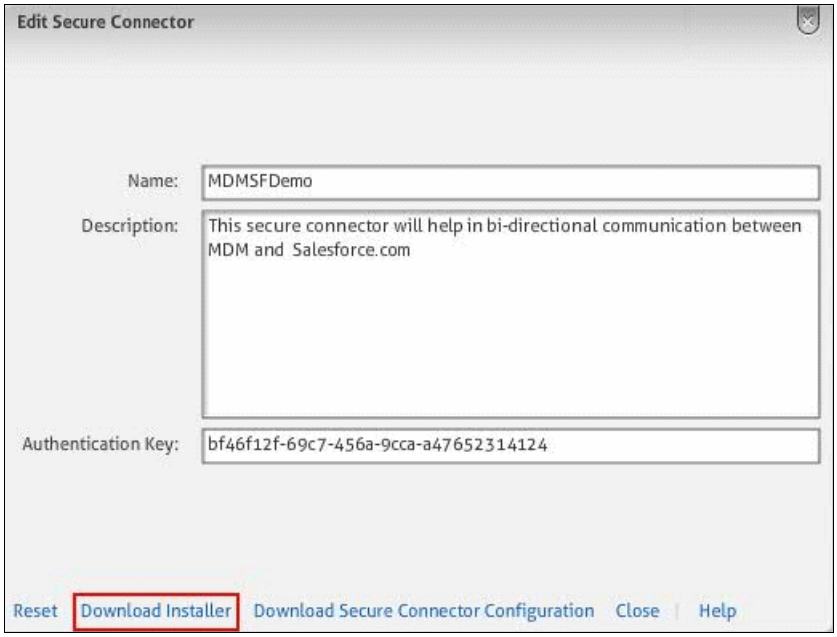


Figure 6-4 Edit Secure Connector

c. Choose the target OS for the installer.

The installer is available for Windows and Linux OS. For Windows OS, the file name is `windows-secure-connector-installer.exe` as shown in Figure 6-5. Save this file to your local system.

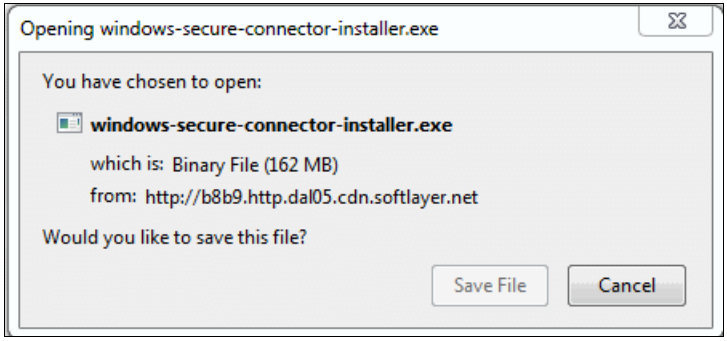


Figure 6-5 Secure connector installer

d. Click **Download Secure Connector Configuration** (Figure 6-6).

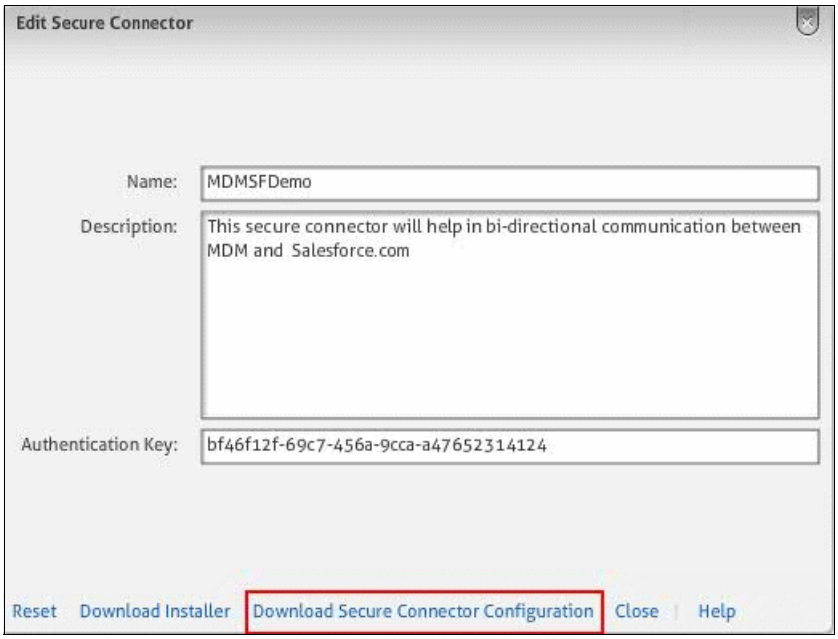


Figure 6-6 Edit Secure Connector

e. Save the file to your local system

6.2.2 Installing Secure Connector on Windows system

Complete the following steps to install Secure Connector on Windows system:

1. Click the installer file, `windows-secure-connector-installer.exe`, for Secure Connector to start the installation.

> **Verifying the publisher:** A window might open to inform you that the publisher cannot be verified and to prompt you regarding whether you want to continue to run the software. Click **Yes** to continue.

2. In the welcome panel, click **Next** to continue (Figure 6-7).

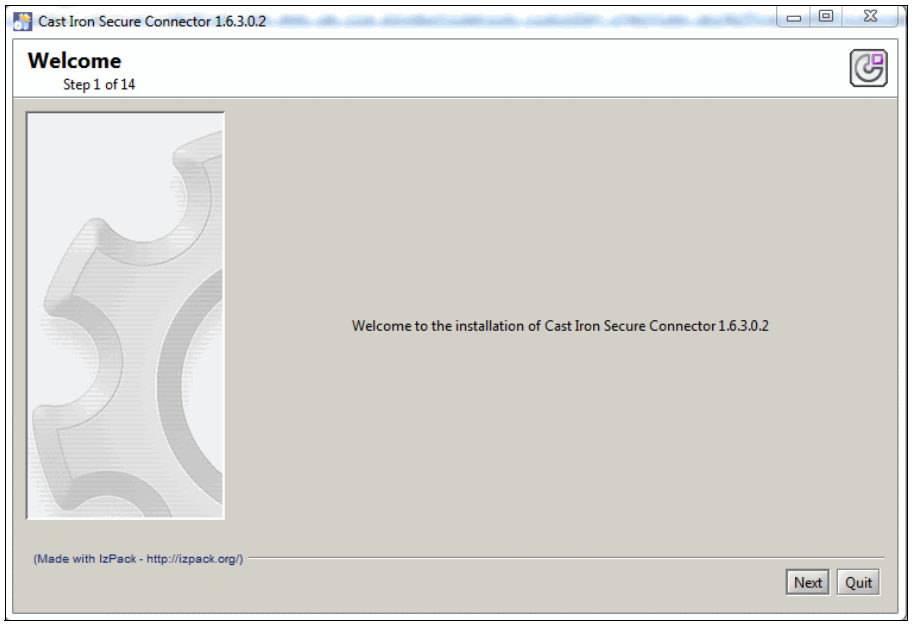


Figure 6-7 Cast Iron Secure Connector Installer, Welcome panel

3. Read and accept the license agreement.
4. Specify the installation path for Secure Connector.
5. If the target directory specified is not available, it is created, and the user is informed about it. Click **OK** (Figure 6-8).

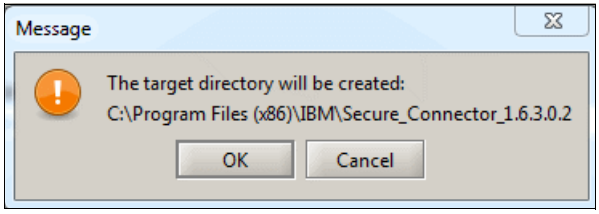


Figure 6-8 Message for creating target directory path

6. Choose to create shortcuts or go with the default settings (Figure 6-9).

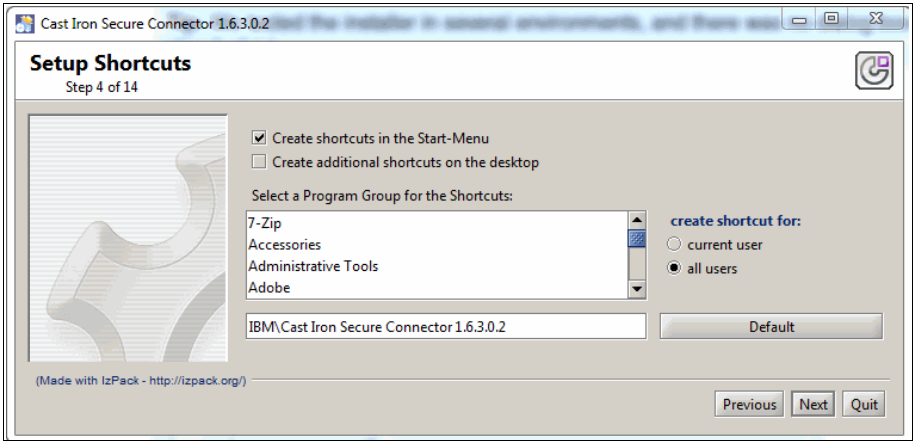


Figure 6-9 Cast Iron Secure Connector, Setup Shortcuts panel

7. The installer starts, and it takes a couple of minutes to complete the installation as shown in Figure 6-10.

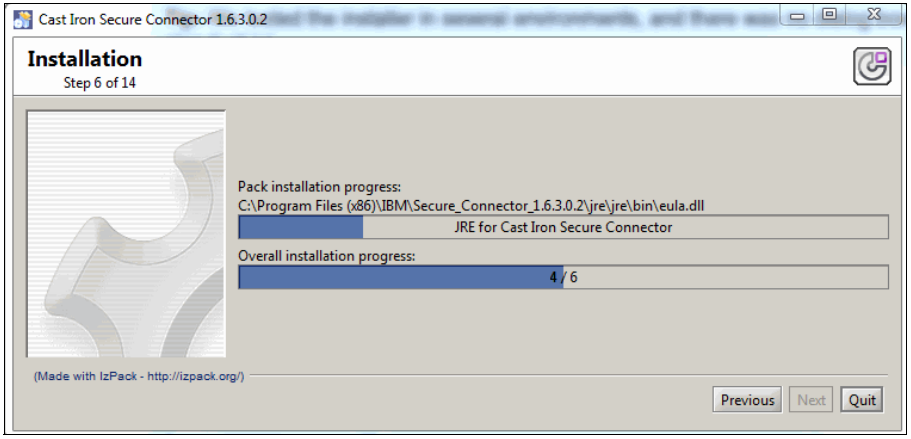


Figure 6-10 Cast Iron Secure Connector, Installation panel

8. After the installation is complete, the progress bar shows the status, click **Next** to continue (Figure 6-11).

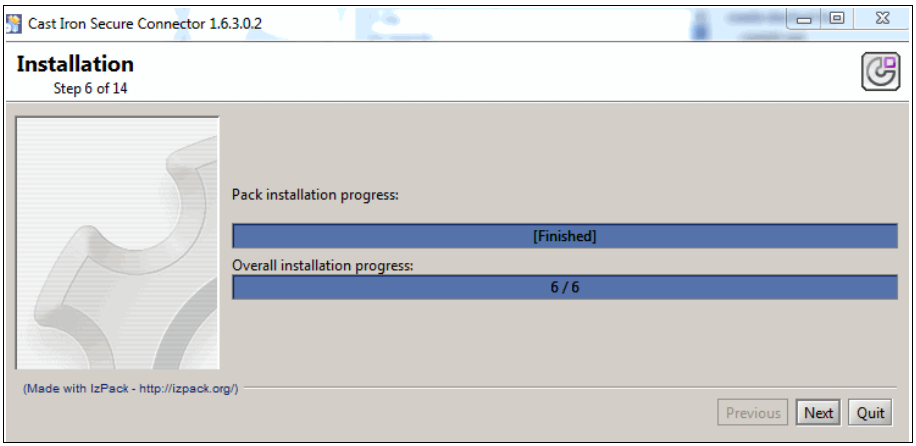


Figure 6-11 Cast Iron Secure Connector, Installation finished

9. To load the Secure Connector configuration file, click **Previous**.
10. Browse to the location where the Secure Connector configuration file is downloaded and choose the Secure Connector configuration file. Click **Next** when finished (Figure 6-12).

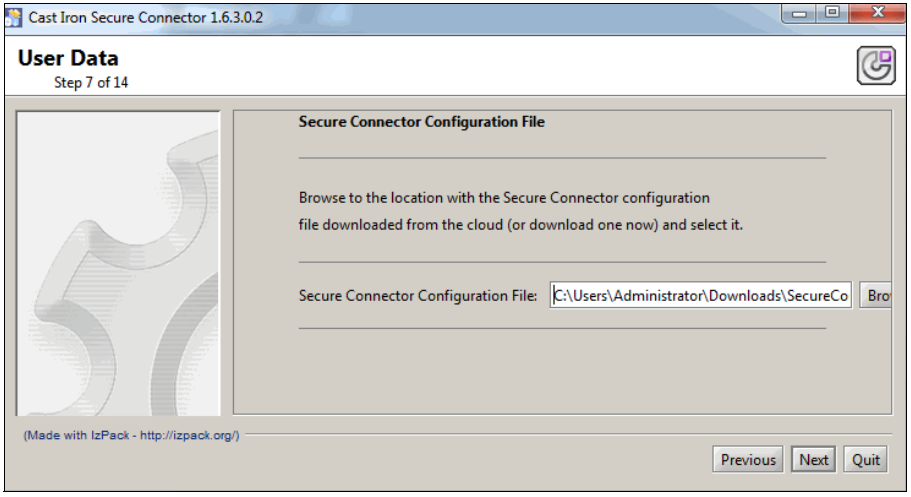


Figure 6-12 Cast Iron Secure Connector installation path to configuration file

11. The wizard reads the configuration file and displays the data (Figure 6-13), click **Next**.

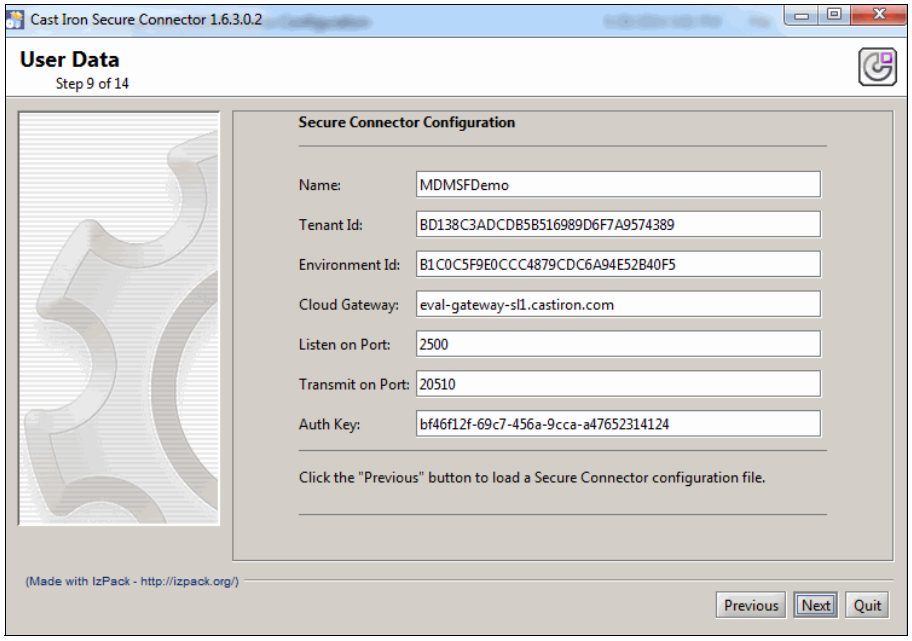


Figure 6-13 Cast Iron Secure Connector installation user data display

12. If you have a need to connect to the Internet through a proxy, enter the proxy server details in this step. Otherwise, click **Next**.

13. If you want to run Secure Connector as a Windows service, specify the details as requested. This scenario starts the service manually as shown in Figure 6-14.

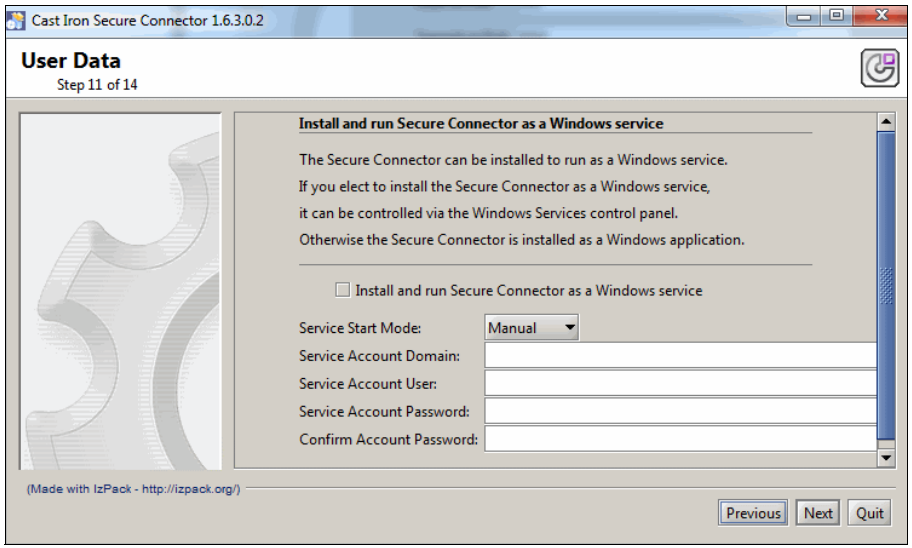


Figure 6-14 Cast Iron Secure Connector installation Windows service configuration

14. The installation is completed, click **Done** to close the wizard (Figure 6-15).

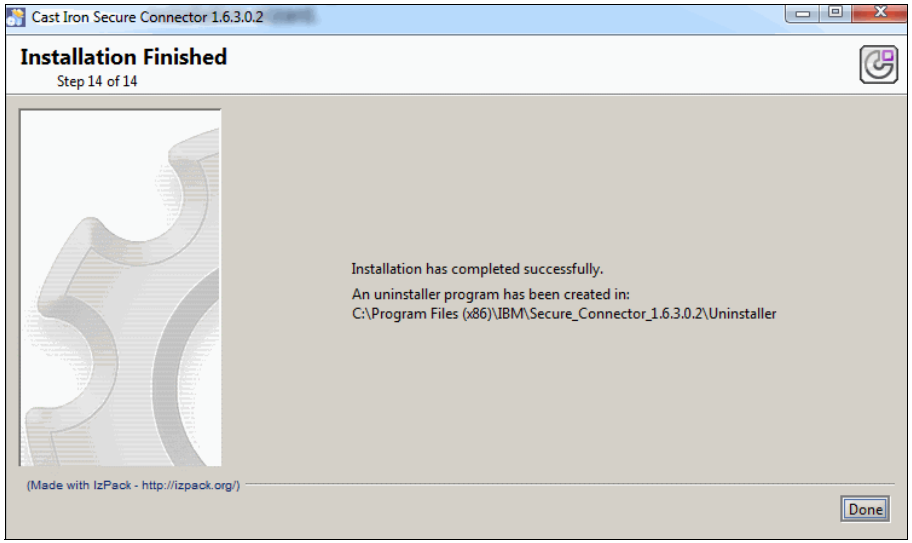


Figure 6-15 Cast Iron Secure Connector installation finished

6.3 Configuring the project in WebSphere Cast Iron Studio

The downloaded sample contains the WebSphere Cast Iron Cloud Integration project for the scenario. The project must be imported into WebSphere Cast Iron Studio and modified before deployment to the WebSphere Cast Iron Cloud Integration.

6.3.1 Downloading the Template Integration Process

Complete the following steps to download the template from the Template Integration Processes (TIPs):

1. Start the installed WebSphere Cast Iron Studio.
2. Click menu **Solutions** → **Search for TIPs**.
3. You can log in to the TIP repository using the default credentials:

   ```
   Username: TIPSreadonly
   password: IBM4you!
   ```

4. In the search window, select **Filter**. Enter the source endpoint and target endpoint, and click **Query**.
5. The available TIPs display. Select **T0310-IBM-MDMToSFDC-SearchBeforeCreate MDM Salesforce.com** and click **Download**.

For more details about TIPs, see 3.3.1, “Data flow orchestration using TIPs” on page 59.

6.3.2 Configuring the project

`MDM-Salesforce_Search_Before_Create` is WebSphere Cast Iron Cloud Integration orchestration project that makes InfoSphere MDM-SE eSOA web services available as provider web service. This project is provided as TIP in WebSphere Cast Iron Cloud Integration.

Use the following steps to configure the `MDM-Salesforce_Search_Before_Create` project:

1. After downloading the TIP, the configuration wizard to customize the TIP for this scenario opens. The configuration wizard opens with the Introduction panel, which describes the scenario, as shown in Figure 6-16.

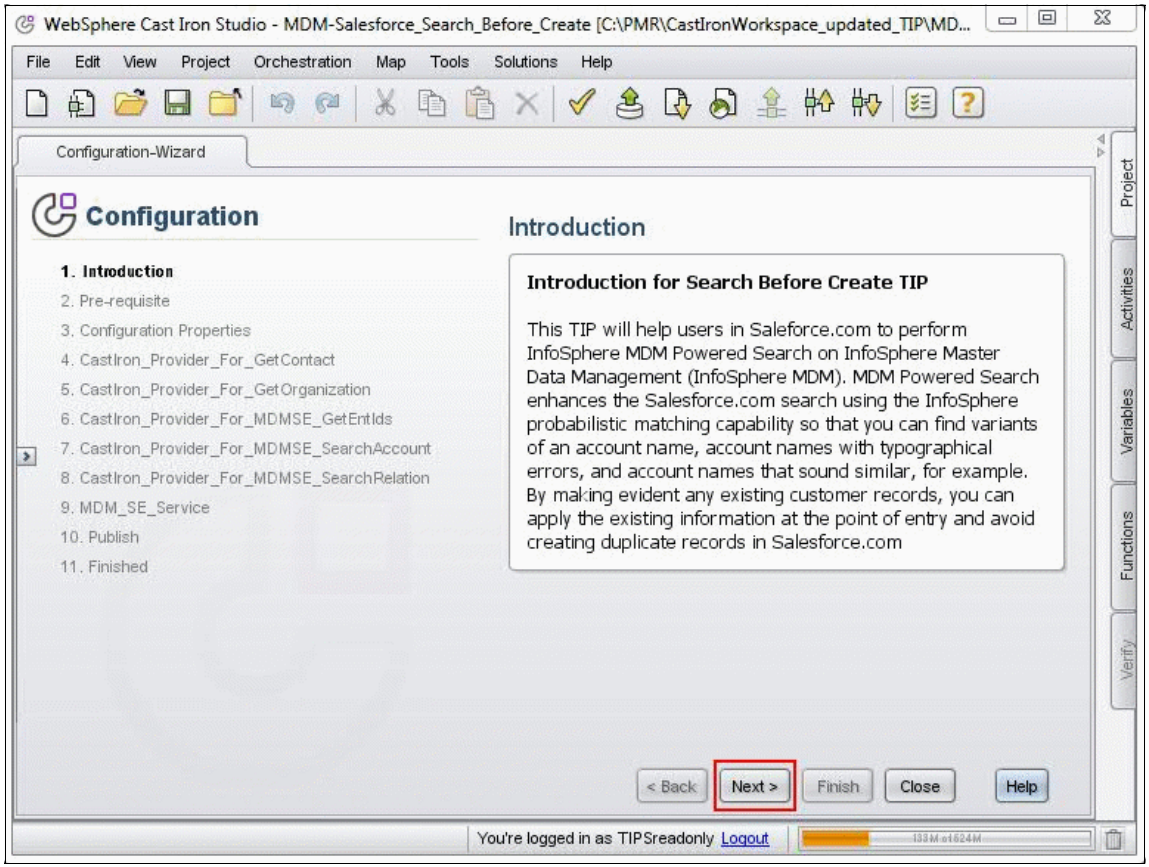


Figure 6-16 TIP Configuration Wizard, Introduction panel

2. Review the prerequisites that are required for running the scenario. Make sure that all the prerequisites are satisfied. Click **Next**.
3. The configuration properties for the project display. Change the HostForMDMWebService property to point to your MDM eSOA web service. For example:

   `http://mdmhostname:WC_defaulthost/madclient/services/mdm`

Enter the name of the Secure Connector that is configured and installed in your system as shown in Figure 6-17.

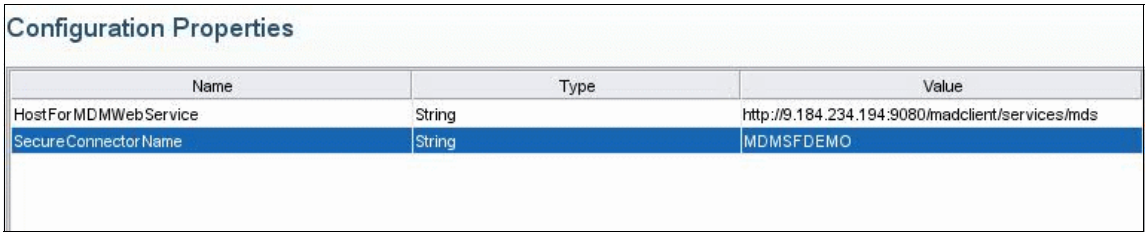


Figure 6-17 TIP Configuration Wizard, Configuration Properties panel

Click **Next** to move to the next step in the wizard.

4. From step 4 to step 8 of the wizard, the following provider services are displayed. Proceed with the default values.
   - CastIron_Provider_For_GetContact
   - CastIron_Provider_For_GetOrganization
   - CastIron_Provider_For_MDMSE_GetEntIds
   - CastIron_Provider_For_MDMSE_SearchAccount
   - CastIron_Provider_For_MDMSE_SearchRelation
5. In step 9 of the wizard, the web service MDM_SE_Service configuration displays. Click **Next** (Figure 6-18).

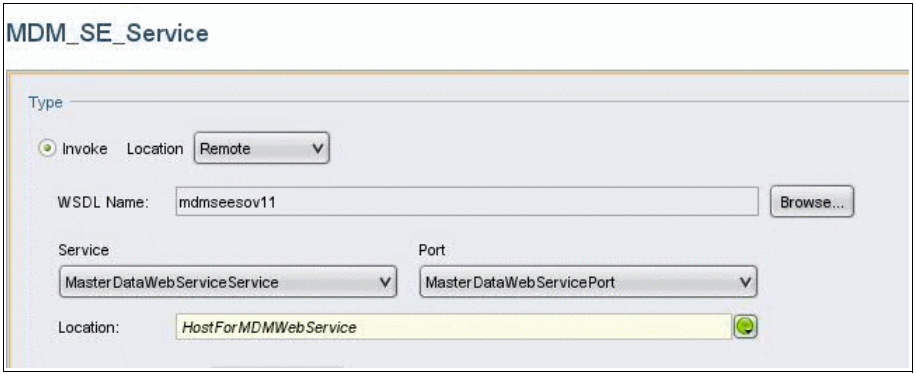


Figure 6-18 TIP Configuration Wizard, web service endpoint configuration

6. Step 10 of the wizard helps to publish the project to the WebSphere Cast Iron Live. By clicking the link in the panel, the Save Project window shows. Click **Yes** (Figure 6-19).

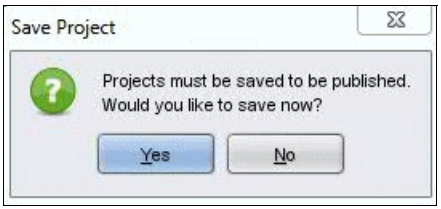


Figure 6-19 Save Project dialog

7. In the Publish Project panel, Enter the IP address of the WebSphere Cast Iron Live service and your tenant credentials to publish the project (Figure 6-20).

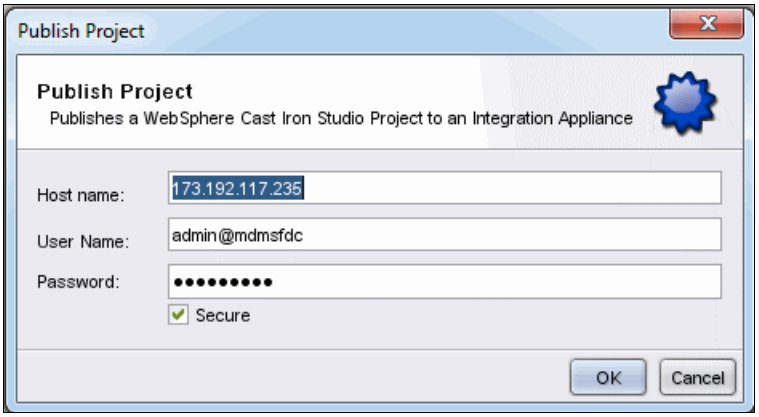


Figure 6-20 Publish project

Figure 6-21 shows the successful message when the project is published successfully.

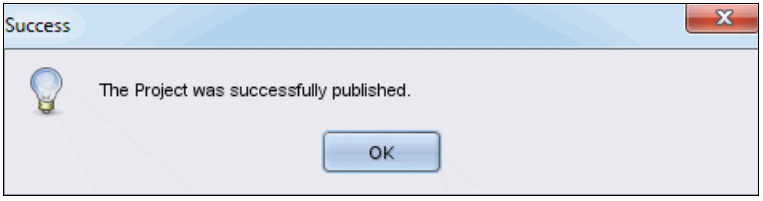


Figure 6-21 Success message after project is successfully published

8. You can log in to the web management console of WebSphere Cast Iron Live to check if the project is published, as shown in Figure 6-22.

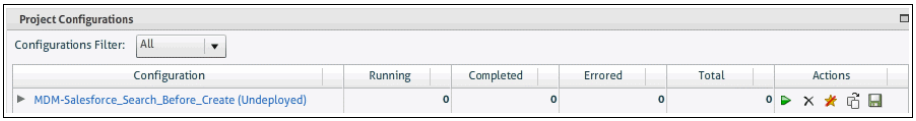


Figure 6-22 Project Configuration from WebSphere Cast Iron Live

6.4 Loading sample data

The scenario requires that data exists in both Salesforce.com and the InfoSphere MDM. For demonstrations or testing, you can use SOAP requests to create and load sample test data to the InfoSphere MDM server.

The user can load enterprise data and test the application. The downloaded sample asset provides a sample data set for testing purpose. For more details about loading the sample data, see *Loading sample data for the MDM Powered Search scenario* in InfoSphere MDM V11.3 Knowledge Center at:

http://www.ibm.com/support/knowledgecenter/SSWSR9_11.3.0/com.ibm.swg.im.mdmhs.sdfcintegration.doc/topics/c_mdmandsfdc_integration.html

6.5 Testing the scenario

This section provides details steps about how to run the project. After completing the configuration and loading sample data, you can test the sample scenario with the following steps:

1. Starting Secure Connector.
2. Starting the WebSphere Cast Iron project.
3. Running the sample.

6.5.1 Starting Secure Connector

When Secure Connector is installed on the Windows system as a Windows application instead of a Windows service, you can start Secure Connector from the Windows start menu by going to **Start Menu** → **All Programs** → **IBM** → **Cast Iron Secure Connector*<version number>*** → **Start Secure Connector**.

Figure 6-23 shows the message in the command prompt when Secure Connector is started.

Figure 6-23 Cast Iron Cloud Integration Secure Connector startup logs

6.5.2 Starting the WebSphere Cast Iron project

Log in to the WebSphere Cast Iron Live web management console as the tenant administrator. In the Development tab, you see the project `MDM-Salesforce_Search_Before_Create` in the Undeployed state.

To start the `MDM-Salesforce_Search_Before_Create` project, click the run icon (). The project moves to the Running state as shown in Figure 6-24.

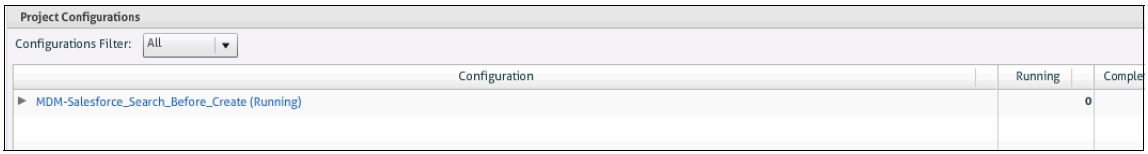


Figure 6-24 Orchestration running in Cast Iron Live

6.5.3 Running the sample

You can test the InfoSphere MDM and Salesforce.com integration application from the front-end application of Salesforce.com. Follow these instructions to run the sample:

1. Open a browser, enter `www.salesforce.com`, and log in using the Salesforce.com account that was used to deploy the Force.com project as shown in Figure 6-25.

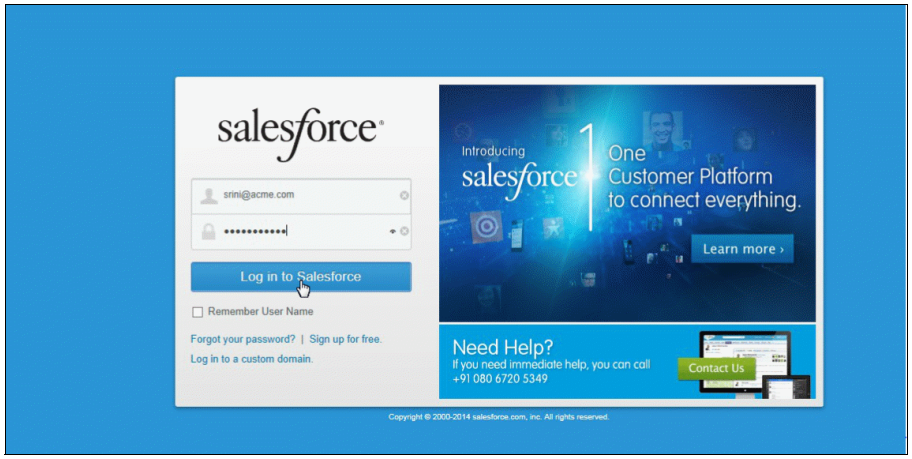


Figure 6-25 Salesforce.com login page

2. Salesforce.com now includes a IBM InfoSphere MDM tab (shown in Figure 6-26). Click **IBM InfoSphere MDM**.

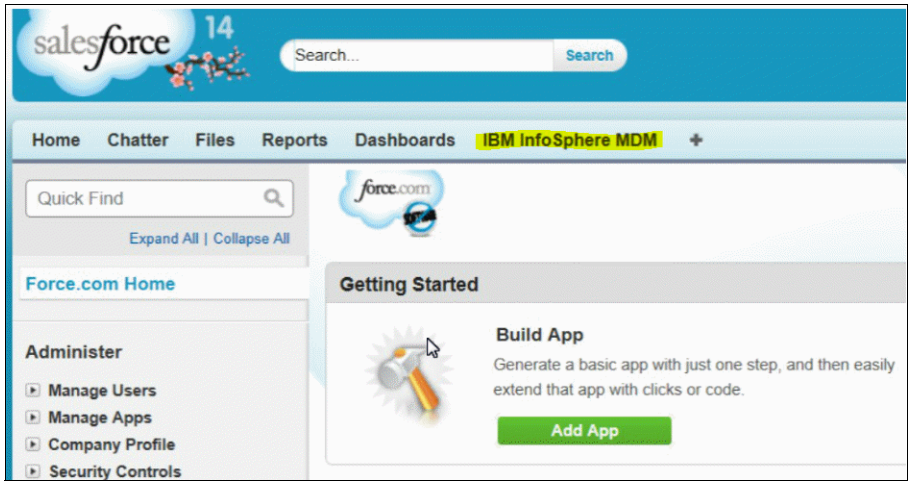


Figure 6-26 InfoSphere MDM tab in Saleforce.com

3. There are three capabilities provided as sample asset. We focus on InfoSphere MDM Powered search. Click the **Go to InfoSphere MDM Powered Search** link (Figure 6-27).

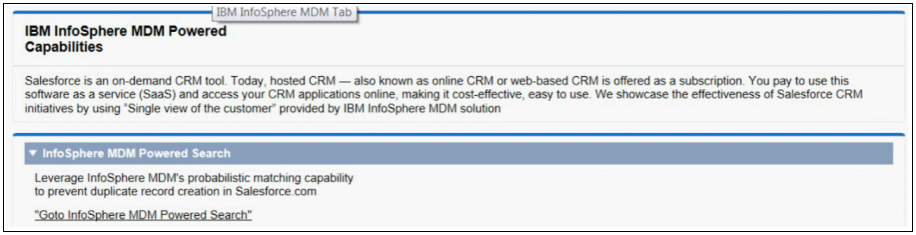


Figure 6-27 InfoSphere MDM Powered Capabilities window in Salesforce.com

4. Enter the account name in Search For Account in MDM and click **Go!**. If the sample data is loaded for testing, sample data loads account “UTI Limited”, you can enter the same account for search as shown in Figure 6-28.

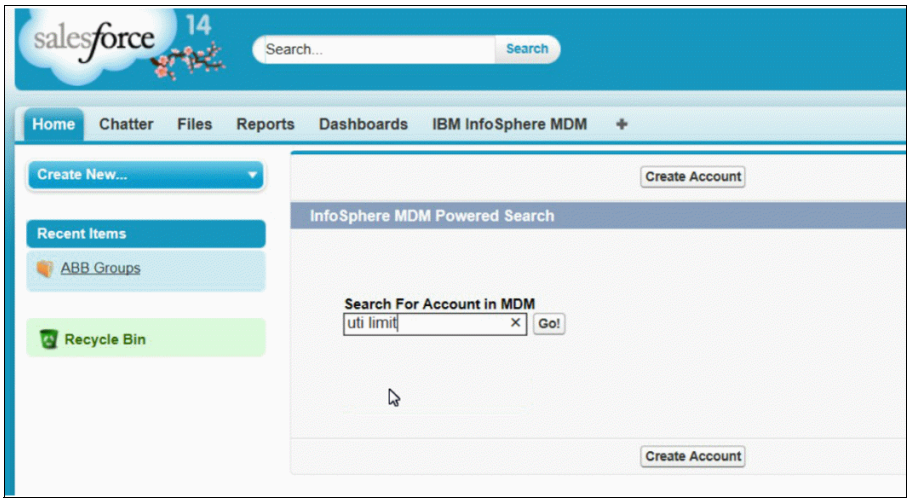


Figure 6-28 InfoSphere MDM Powered search window in Salesforce.com

The solution search for the account in InfoSphere MDM across all the source loaded in InfoSphere MDM and displays the result in tables as shown in Figure 6-29. You can click the account name in the result to view additional details of the account.

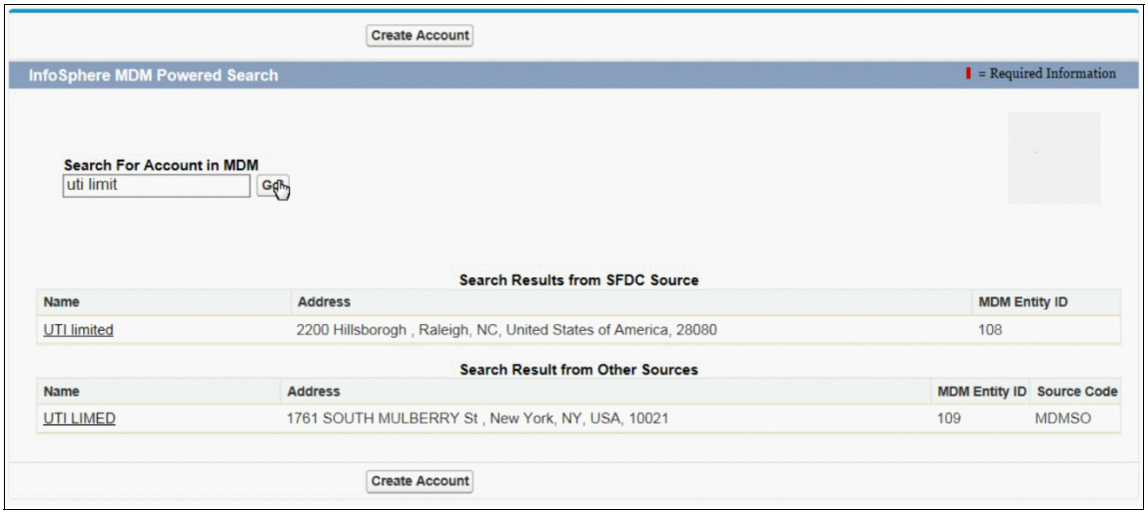


Figure 6-29 InfoSphere MDM Powered search results in Salesforce.com

When you closely observe the result, the user has wrongly typed the account name as *uti limit*. The probabilistic matching engine caught the variations and found two accounts that closely matched the search criteria. The search result found the record from Salesforce.com source and also fetched the matching account from all other sources within the enterprise.

7

Handling integrations with other SaaS applications

This chapter describes two potential integration scenarios of IBM InfoSphere Master Data Management (InfoSphere MDM) with Netsuite and Workday. It includes the following topics:

- InfoSphere MDM and Netsuite CRM integration
- InfoSphere MDM and Workday
- IBM Bluemix and services

7.1 InfoSphere MDM and Netsuite CRM integration

Netsuite Enrolled is a cloud-based business management software that includes components for enterprise resource planning (ERP), customer relationship management (CRM), and e-commerce. Netsuite provides its services in the software as a service (SaaS) model. Netsuite hosts and maintains the infrastructure, and users can access the service by paying a subscription fee with a simple browser and internet connection. Netsuite provides various applications for data process and integrations with exiting systems that are on the cloud or residing on the customer's premises.

This section briefly discusses Netsuite CRM and two process integration patterns with InfoSphere MDM.

7.1.1 InfoSphere MDM powered search

Netsuite Suite Script provides application-level scripting framework and tools for customers to customize the user interfaces, business logic, even-driven management, and debugging capabilities. It is based on JavaScript.

You can use the InfoSphere MDM powered search with Netsuite similar to the InfoSphere MDM powered search with Salesforce.com discussed in Chapter 5, "A sample InfoSphere MDM and SaaS application integration scenario" on page 87.

The following is an high-level integration overview of InfoSphere MDM and Netsuite:

- Netsuite Suite Script can be used to develop custom user interfaces for the search functionality.
- The search function invokes the InfoSphere MDM web services through the IBM WebSphere Cast Iron Cloud Integration orchestrations.
  - Secure connector must be installed on the customer's on-premises systems.
  - Orchestration with a provider service must be developed using WebSphere Cast Iron Live Designer.
  - The orchestration then invokes the InfoSphere MDM web services through Secure Connector and serves back the results.
- InfoSphere MDM powered probabilistic search helps with real-time scenarios where search parameters do not exactly match the records. This solution capability allows customers to reduce duplication at the point of entry to their information systems.

7.1.2 Real-time synchronization of records in Netsuite CRM to InfoSphere MDM

IBM WebSphere Cast Iron Live comes with a built-in connector for Netsuite. This connector provides a simplified interface over the Netsuite objects, fields, and operations to speed up the integration patterns built. You can use IBM WebSphere Cast Iron Live Designer to graphically design the orchestration to use the Netsuite connector.

You can use Netsuite Transaction Search records to obtain a list of updated objects after a certain date (last modified date should be the input). The WebSphere Cast Iron Cloud Integration connector for Netsuite has a Search Records activity that can be used to search for new accounts created. Different from the WebSphere Cast Iron Cloud Integration connector for Salesforce.com, Netsuite connector does not have starter activities to automatically retrieve the list of updated objects. Therefore, you can use the “Schedule Job” activity in the Designer to trigger the Search Record activity of the Netsuite connector at suitable time intervals.

When the list of new records is obtained from Netsuite, you can use the Invoke Request activity in the web services connector to start the appropriate web services on InfoSphere MDM to create a similar record.

Figure 7-1 shows a WebSphere Cast Iron Cloud Integration orchestration that synchronize updated objects between Netsuite and InfoSphere MDM.

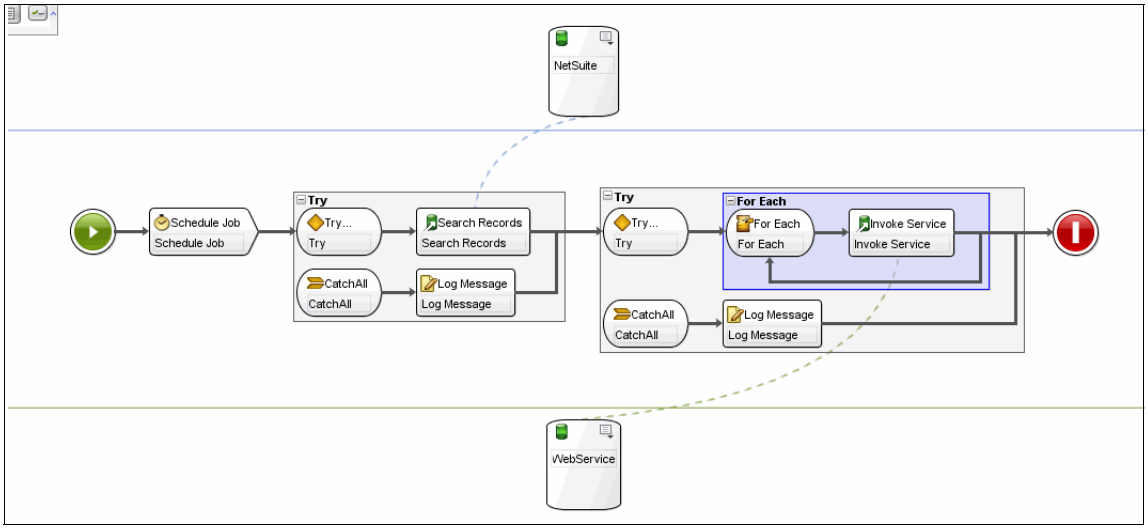


Figure 7-1 IBM WebSphere Cast Iron Cloud Integration orchestration

7.2 InfoSphere MDM and Workday

Workday is a cloud-based human capital management and financial management software provided by Workday Inc. Workday provides 18 services that are exposed as SOAP applications:

- Absence_Management
- Benefits_Administration
- Cast_Management, Compensation
- External_Integrations
- Financial_Management
- Human_Resources
- Identity_Management
- Integrations
- Payroll
- Payroll_Interface
- Performance_Management
- Professional_Services_Automation
- Recruiting, Resources_Management
- Revenue_Management
- Staffing and Talent

IBM WebSphere Cast Iron Live provides a connector for all of the 18 Workday services. The Workday connector of WebSphere Cast Iron Live also supports custom reports, which is the preferred way of integrating with Workday. The Workday connector provides a convenient abstraction over the underlying web service protocol and formats and enables integration to Workday through simple drag and drop operation.

Orchestrations can be built on IBM WebSphere Cast Iron Designer using the Workday connector and the web services connector to talk to InfoSphere MDM web services to suit customer's business needs.

7.3 IBM Bluemix and services

IBM Bluemix™ is a platform as a service (PaaS) offering that enables developers to quickly build applications to satisfy business needs with a composition of Bluemix services. IBM Bluemix additionally provides various SaaS offerings to enable clients to fully realize their business needs. Developers can deploy their applications on IBM hosted infrastructure that runs on IBM SoftLayer. IBM Bluemix also enables clients to manage these applications with necessary tools and utilities.

For more information about IBM Bluemix, see this website:

https://www.bluemix.net/

IBM WebSphere Cast Iron Live service is available under the Cloud Integration add-on on IBM Bluemix. From Bluemix, customers can use the cloud integration service (available under Add-Ons) to create an InfoSphere Cast Iron Live evaluation tenant or log in to the existing InfoSphere Cast Iron Live production account.

Related publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this book.

IBM Redbooks publications

The following IBM Redbooks publications provide additional information about the topic in this document. Note that some publications referenced in this list might be available in softcopy only.

- *Getting Started with IBM WebSphere Cast Iron Cloud Integration*, SG24-8004
- *Hybrid Cloud Integration and Monitoring with IBM WebSphere Cast Iron*, SG24-8016
- *IBM WebSphere Cast Iron Introduction and Technical Overview*, REDP-4840
- *Connect Cloud and On-premise Applications Using IBM WebSphere Cast Iron Integration*, REDP-4674
- *Integrating ERP and CRM Applications with IBM WebSphere Cast Iron*, TIPS0961
- *Integrating Two Cloud Services with IBM WebSphere Cast Iron Live*, TIPS0962
- *Smarter Analytics: Information Architecture for a New Era of Computing*, REDP-5012

You can search for, view, download or order these documents and other Redbooks, Redpapers, Web Docs, draft and additional materials, at the following website:

ibm.com/redbooks

Other publications

These publications are also relevant as further information sources:

- *Patterns For Information Management*, ISBN: 978-0-13-315550-1

Online resources

These websites are also relevant as further information sources:

- InfoSphere Master Data Management

  http://www-03.ibm.com/software/products/en/infosphere-master-data-management
- IBM WebSphere Cast Iron Cloud integration

  http://www-03.ibm.com/software/products/en/castiron-cloud-integration
- WebSphere Cast Iron Express

  http://www.ibm.com/software/products/us/en/castiron-exp
- WebSphere Cast Iron Hypervisor Edition

  http://www.ibm.com/software/products/us/en/cast-iron-hypervisor
- WebSphere Cast Iron Live

  http://www.ibm.com/software/products/us/en/cast-iron-live
- WebSphere Cast Iron Knowledge Center

  http://pic.dhe.ibm.com/infocenter/wci/v7r0m0/index.jsp?topic=%2Fcom.ibm.wci.live.doc%2FSecure_Connector%2Ftoc_using_secure_connectors.html
- InfoSphere Information Server

  http://www-03.ibm.com/software/products/en/infosphere-information-server
- SOA

  http://www-01.ibm.com/software/solutions/soa/
- SP800-131a

  http://csrc.nist.gov/publications/nistpubs/800-131A/sp800-131A.pdf
- National Institute of Standard and Technology

  http://www.nist.gov/

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

IBM Redbooks

Master Data Management for SaaS Applications

(0.2”spine)
0.17”<->0.473”
90<->249 pages

Master Data Management for SaaS Applications

InfoSphere MDM and SaaS applications integration guidance

WebSphere Cast Iron Cloud integration as middleware

Cloud application integration scenarios

Enterprises today understand the value of employing a master data management (MDM) solution for managing and governing mission critical information assets. chief data officers and chief information officers drive MDM initiatives with IBM InfoSphere Master Data Management to improve business results and operational efficiencies, which can help to lower costs and to reduce the risk of using untrusted master information in business process. Cloud computing introduces new considerations where enterprise IT architectures are extended beyond the corporate networks into the cloud.

This IBM Redbooks publication provides guidance to chief data officers, chief information officers, MDM practitioners, integration architects, and others who are interested in the integration of IBM InfoSphere Master Data Management with SaaS applications. This book lays the background on how mastering and governance needs for SaaS applications is quite similar to what on-premises business applications would need. It draws the perspective for serving the on-premises application and the SaaS application with the same MDM hub. This book describes how IBM WebSphere Cast Iron Cloud Integration can serve as the “de-facto” cloud integration middleware to integrate the on-premises InfoSphere Master Data Management systems with any SaaS application by using Saleforce.com integration as an example. This book also covers aspects of handling bulk operations with IBM InfoSphere Information Server.

INTERNATIONAL TECHNICAL SUPPORT ORGANIZATION

BUILDING TECHNICAL INFORMATION BASED ON PRACTICAL EXPERIENCE

IBM Redbooks are developed by the IBM International Technical Support Organization. Experts from IBM, Customers and Partners from around the world create timely technical information based on realistic scenarios. Specific recommendations are provided to help you implement IT solutions more effectively in your environment.

For more information:
ibm.com/redbooks

SG24-8220-00 ISBN 0738440043