

IBM Cloudant: The Do-More NoSQL Data Layer

IBM Redbooks Solution Guide

Cloudant represents a strategic acquisition by IBM® that extends the company’s Big Data and Analytics portfolio to include a fully managed, NoSQL cloud service. Cloudant simplifies the development cycle for creators of fast-growing web and mobile applications, by alleviating the burdens of mundane database administration tasks. Developers are then able to focus on building the next generation of systems of engagement - social and mobile applications - without losing time, money, or sleep managing their database infrastructure and growth. Critically, Cloudant is an enterprise-ready service that supports this infrastructure with guaranteed performance and availability.

Built atop a CouchDB-based NoSQL data layer, Cloudant’s fully managed database-as-a-service (DBaaS) enables applications and their developers to be more agile. As a part of its data layer, clients have access to multi-master replication and mobile device synchronization capabilities for occasionally connected devices. Applications can take advantage of Cloudant’s advanced real-time indexing for ad hoc full text search via Apache Lucene, online analytics via MapReduce, and advanced geospatial querying. Mobile applications can use a durable replication protocol for offline sync and global data distribution, as well as a geo-load balancing capability to ensure cross-data center availability and optimal performance. Cloudant’s RESTful web-based API, flexible schema, and capacity to scale massively are what empower clients to deliver applications to market faster in a cost-effective, DBA-free service model. This IBM Redbooks® Solution Guide describes the IBM Cloudant features. Figure 1 shows the Cloudant Do More Data Layer.

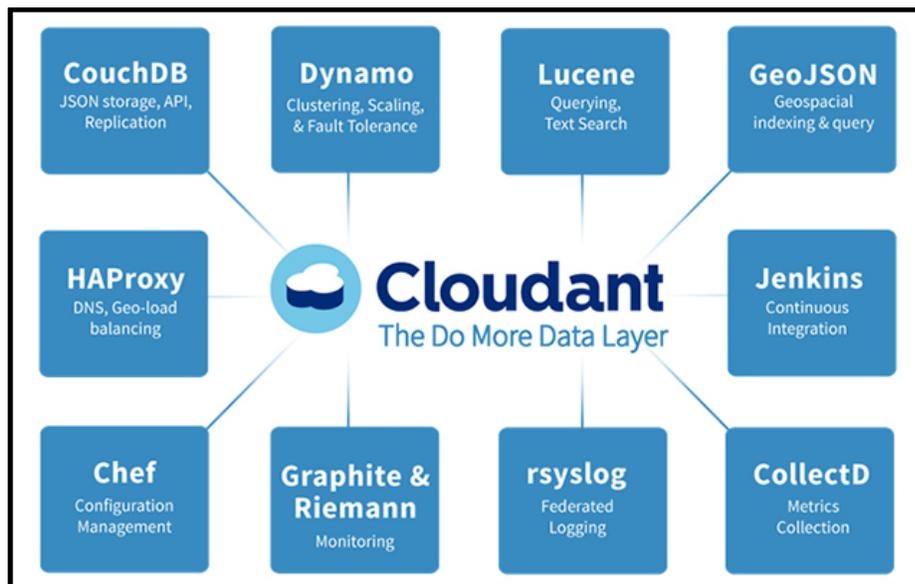


Figure 1. Cloudant delivers the most flexible, scalable, and always available solution for developers of big mobile and “Internet of Things” applications, via a fully managed database-as-a-service

Did you know?

The NoSQL data management market is burgeoning: forecasts expect the value of the market to grow to USD \$14 billion between 2014 and 2018. Additionally, over 50% of NoSQL solutions use JavaScript Object Notation (JSON)-based document data stores (including Cloudant).

Business value

It is clear that the momentum behind big data, mobile, social, Internet of Things (IoT), and cloud initiatives is transforming the modern IT profession. We are already seeing an explosion of new products and applications developed for these platforms, and the expectation is that these products will experience ever-increasing consumption by established and emerging markets. Agility and elasticity are key in the mobile application environment, where being able to rapidly scale up performance is necessary to accommodate fluctuations in usage and load on your infrastructure. In the mobile space, the demand on these systems can change on a day-to-day basis, and it is necessary for developers to have a solution in place that allows them to scale up without having to scale their internal resources, as well. Web and mobile application databases - and their administrators - must be prepared to face issues of cost, performance, scalability, availability, and security head-on; all of which carry uncertainty and risk.

IBM Cloudant eliminates this complexity by enabling developers to focus on building next-generation applications without the need to manage their database infrastructure or growth. A fully managed cloud data layer service, Cloudant offers its clients the high availability, scalability, simplicity, and performance that modern web and mobile applications demand. The scalability of a fully managed cloud DBaaS solution simplifies the application development cycle and offers Cloudant clients greater agility for launching new products or responding to an ever-changing market: build more, grow more, and sleep more.

Cloudant provides the following benefits:

- Offers a NoSQL data layer, delivered as a fully managed service. Liberates developers from the cost, complexity, and risk of do-it-yourself data layer solutions.
- Monitored and managed 24x7 by Cloudant's big data and database administration experts.
- Uses self-describing JSON "document" storage schemas to allow for flexible and agile application development.
- Mobile device and web replication and synchronization support for offline and occasionally connected devices.
- Built using a master-master (also known as "master-less") clustering framework that can span multiple racks, data centers, cloud providers, or devices.
- High availability and enhanced performance for customer applications that require data to be local to the user. Supplied by global data distribution and geo-load balancing technologies.
- Delivers real-time indexing for online analytics, ad hoc full-text search via integrated Apache Lucene, and advanced geospatial querying.
- Supplies a RESTful API for ease of access and compatibility with developers that live and work on the modern web.
- Based on open standards including: Apache CouchDB, Apache Lucene, GeoJSON, and others.

Solution overview

The data layer solution offered by IBM Cloudant delivers a fully managed cloud service that is always on, fast, and scalable. Cloudant provides all the database administration for their clients' applications, providing a fast-growing and scalable framework on which clients are able to focus purely on development of their next generation of applications.

IBM Cloudant provides database solutions tailored to address the following scenarios for its clients:

- Inadequate database performance is currently (or has the potential in the future) to hamper user base or business growth.
- Unreliable service availability has negatively affected user experience or resulted in lost revenue opportunities.
- Access required to application features and data on sometimes offline (mobile) devices, where network connectivity is poor or unavailable.
- Performance of advanced analytics on customer data and application metrics needed.
- Storage solutions need to use “variable” or multi-structure JSON data for maximum schema flexibility.
- No in-house database administration solutions; company does not want to hire DBAs.

Developers of mobile and web applications can host their business on a global network of service providers, including IBM SoftLayer, Rackspace, Microsoft Azure, and Amazon Web Services. Regardless of the service provider, Cloudant's data layer ensures that a company's underlying services are fully supported by a scalable and flexible NoSQL solution. Full-text search, advanced analytics technologies, and mobile data replication and synchronization further extend how clients interact with and use their data. For this reason, Cloudant typically targets verticals in the areas of online gaming, mobile development, marketing analytics, software-as-a-service (SaaS) companies, online education providers, social media, networking sites, and data analytics firms. Figure 2 shows how IBM Cloudant's fully managed DBaaS solution fits into the database market.

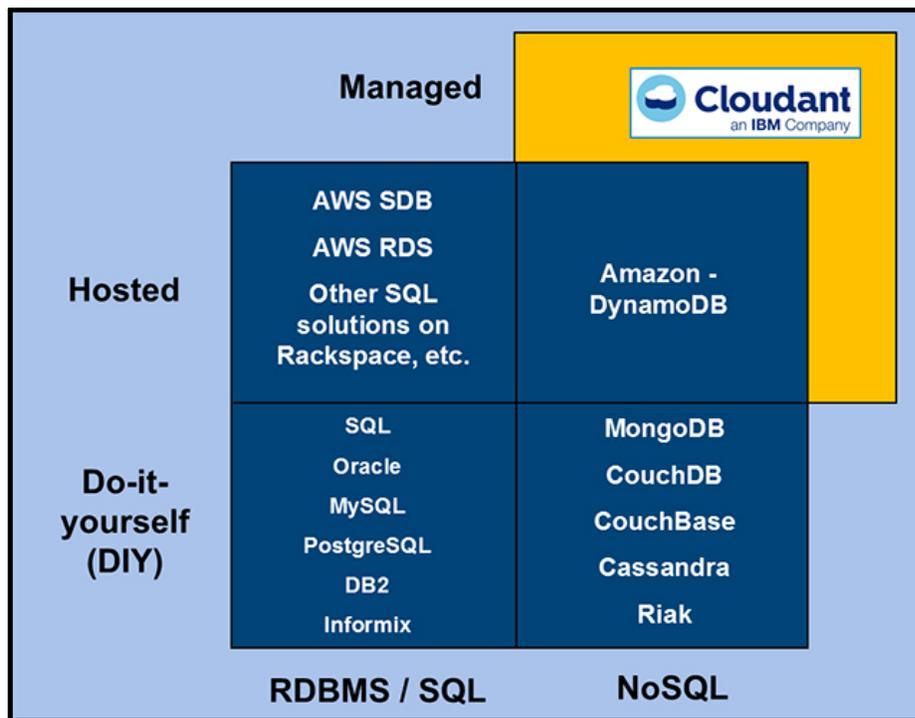


Figure 2. How IBM Cloudant's fully managed DBaaS solution fits into the database market

At the time of this writing, the Cloudant service is hosted in over 35 data centers around the world. IBM Cloudant allows for total flexibility in document and database design - as well as geo-location choices - to ensure maximum control and security over customer data. Cloudant is able to scale out these deployments up to millions of databases; furthermore, you can instantiate individual databases to isolate data on an individual database level. This combination of scaled development and deployment across geospatial locations - as well as partitioning of data across individual databases - enables the client to isolate and tightly control how data is persisted in the network. Figure 3 illustrates the IBM Cloudant NoSQL database, services, and API layer view.

Solution architecture

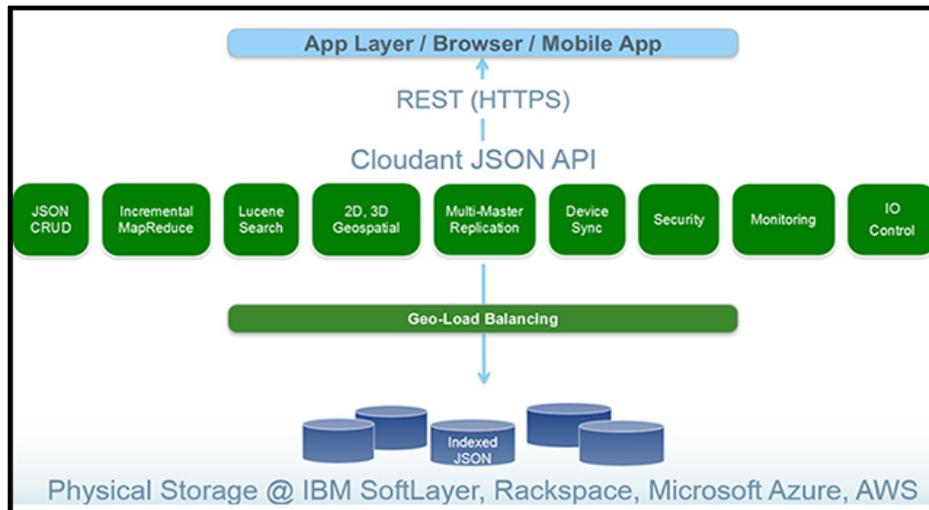


Figure 3. The IBM Cloudant NoSQL database, services, and API layer view

Notice in Figure 3 the fully integrated capabilities that are inherent in the Cloudant API and available without the need for third-party integrations. IBM Cloudant was designed to reduce the complexity of tasks and services that developers otherwise are required to manage themselves: complete synchronization and geo-load balancing features are tunable within Cloudant's toolsets, and its replication API is consistent across both Cloudant and CouchDB. No additional services or components between your device and storage endpoints are required to take advantage of these features. The advantage to you - being able to source database synchronization activity to only particular subsets of your data reduces network load and increases performance via targeted, geospatial-specific synchronization tasks.

One of IBM Cloudant's key API differentiators, which set it in a class apart from competitors, is a feature-laden JSON API, including document data stores, primary indexing, MapReduce-built secondary indexes, and full-text search. JavaScript Object Notation (JSON) is a lightweight data interchange format that has become the de facto data interchange format on the web because of its language independence and self-describing data structures. Data representation and structure can vary from document to document.

This schema flexibility allows you to describe all the aspects of data (in any formatting that you might encounter); moreover, JSON allows you to avoid the use of NULL values - such as you find in relational databases. Consequently, IBM Cloudant can be described as a "flexible schema" approach to data storage: this is not meant to imply that there is no schema, rather that the schema varies across subsets of documents and their data. No database downtime or table locks are required to alter a single document's schema, and because of this Cloudant is aptly suited for scenarios where database schema flexibility is key. Cloudant comes embedded with a variety of real-time indexing options to query your data. Secondary indexes built via MapReduce, also known as *views*, are ideal for searching for secondary keys or ranges of keys, and for doing heavy online analytics. Search indexes built using Apache Lucene are excellent for performing ad hoc or full text search; additionally, Cloudant's indexing supports search facets and groups, as well as search by both distance and bounding box.

Finally, advanced geospatial indexing allows for querying against complex structures, such as polygons and calculating advanced relations, such as overlap or intersection. All of these features are accessed by a RESTful web-based API, which is natural and intuitive to programmers familiar with developing for the web. Figure 4 depicts that the sole concern for clients of a database-as-a-service (DBaaS) solution is the design and development of their application. IBM Cloudant guarantees availability, eliminates risk, and ensures that service is able to scale out as clients (and their applications) grow.

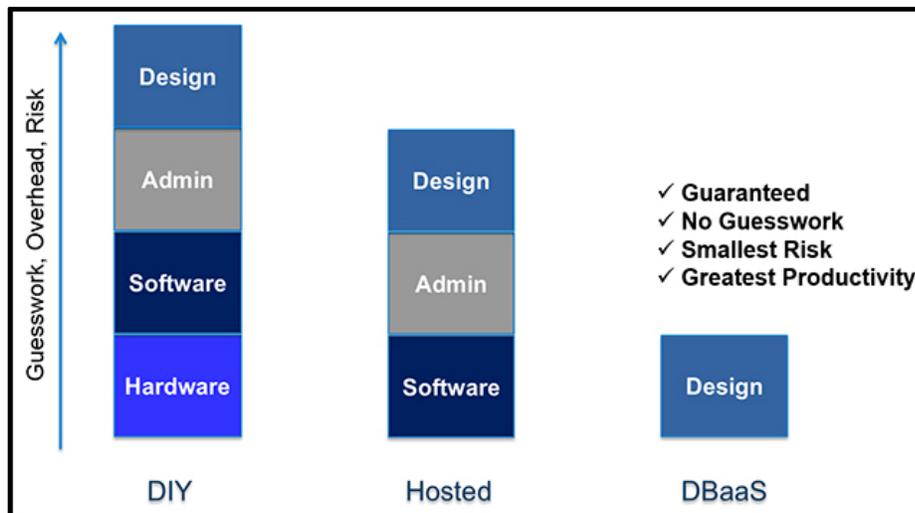


Figure 4. The sole concern for customers of a database-as-a-service (DBaaS) solution is the design and development of their application. IBM Cloudant guarantees availability, eliminates risk, and ensures that service is able to scale out as clients (and their applications) grow

For modern web and mobile applications, the speed of deployment onto a database is critical: databases need to adequately support usage requirements, scale (and downsize) rapidly, and provide high availability. IBM Cloudant's distinguishing feature is the delivery of these necessary services and rich portfolio of proprietary tools via a cloud-distributed, database-as-a-service (DBaaS) solution. Unlike do-it-yourself (DIY) solutions, which ask developers to handle everything from provisioning the hardware to database administration at the top of the stack, DBaaS handles cloud database provisioning, management, and scaling as a paid service to the client. The client receives guaranteed availability and reliability of their business, hardware provisioning that can grow elastically as required, and a rapid time to value with the greatest mitigation of risk. Hosted services simply cannot claim to offer the same degree of comprehensive services: hosted cloud solutions provision hardware and instantiate an image, then turn the keys over to you - the developer. Only a fully managed service, such as IBM Cloudant, can liberate developers from the burdens of database administration and allow clients to focus their energy on what really matters: building the next generation of web and mobile applications for their customers.

Usage scenarios

Consider the scenario of a small start-up developer for mobile and web-browser games with a problem: the runaway success of their newest mobile game. This company is of modest size and equally modest budget, with a game that is named a “Featured App” in the App Store the day of the launch. The unexpected success of the application required the studio to rapidly scale out their service in order to accommodate the increasing demand for the application. At present, this development studio and their application are at a standstill; the load on their inadequate database architecture is so great that the mobile game has become unusable. Despite their best efforts to prepare for scaling - including a soft launch of the game - no one on the team has ever faced such an onslaught of users. Negative customer reviews are piling up and the company’s ability to conduct business is gridlocked. Without experts in database administration on hand to support the product’s demand, the financial repercussions of this ongoing service outage might prove disastrous.

The design studio had five criteria that needed to be satisfied before they are able to commit to a solution:

1. The improved database back end needs to scale massively and elastically (up and down) in response to fluctuating demand on the App Store.
2. It needs to be available nonstop in order to not interrupt the delivery of entertainment to their users around the world.
3. They need to be up and running on it fast, while there was still a chance to capitalize on the initial popularity of the game.
4. The solution needs to be managed - hiring DBAs did not make sense for the company’s long-term objectives of developing better games for their customers.
5. The solution requires improved tools and techniques for data management over the messy and frustrating relational database management systems (RDBMS) that were used previously.

Enter IBM Cloudant, which delivered the robust scalability needed by the game developer - who was able to migrate to Cloudant within just a few days and without hiring a DBA. Cloudant addressed challenges of availability, synchronization, and geography. Millions of new users were able to interact with the online game’s world, without requiring the studio to hire (or have on staff previously) DBAs to administer the solution. Cloudant also provided monitoring of user activity and monetization analytics: tools that allowed the developer to track how customers were purchasing through in-application markets and ecosystems, as well as trace usage and application exposure. Key decision makers in this process were the company’s chief technology officer, as well as the lead application development team - for whom a fully managed Database-as-a-Service, such as Cloudant, was able to alleviate key pain points and drive new business.

As you might expect, IBM Cloudant is as equally appealing to enterprise developers as it is for small start-up companies. Nearly any system that needs to elastically scale concurrent access to data or manage multi-structured data can benefit from Cloudant. Several examples of enterprise success stories are described:

- **Elastic scaling.** A major consumer financial services company created a customer-facing web and mobile app for storing and sharing personal financial data. It was intended to serve as a digital safety deposit box. Providing this to customers generated better self-service and brand loyalty; however, the company did not have the experience or know-how needed to scale a system to serve up to 20 million users. They chose Cloudant over other offerings due to Cloudant's superior service, performance, and security.
- **Messy data.** One of the world's largest pharmaceutical companies uses the Cloudant DBaaS to stage and transform clinical trial data for a large data warehousing and analytics project they operate. Cloudant's ability to handle the wide variety of clinical studies data (via self-describing JSON indexing) - and ability to index it incrementally as new data is loaded - reduced the time needed for data processing from 18 hours (in Oracle) to just a few minutes. It also eliminated the expense of Oracle's hardware and software overhead.
- **Internet of Things.** A fitness metrics company collects data from Internet-enabled fitness devices (including mobile phones) to collect information about product usage and workout information. Users can subsequently tap into and monitor their fitness metrics online. It also collects product "health" readings to determine whether the devices collecting this data might require maintenance. The client relies on Cloudant to handle the large volume of data being concurrently collected and read by its products and users.
- **Social learning.** A publicly held developer of desktop language learning software wants to deliver their software as an online service. Additionally, they want to enable language learners to connect and communicate with each other, in order to practice their newly learned languages with other users. They use Cloudant to handle the large scale-up of course material and user data, including connections between users, states of conversations, full-text indexing, searches of curriculum and correspondence information, and more. Cloudant provided the scalability and eliminated the need to use separate databases for structured data, graph (connections) data, and full text.

Integration

The IBM enterprise-ready Big Data and Analytics portfolio enables clients to address the full spectrum of challenges across areas of mobile, social, big data, and the cloud. Cloudbant extends these capabilities by providing another leading solution to an already market-leading portfolio. Table 1 describes IBM product integration points with Cloudbant.

Table 1. IBM product integration points with Cloudbant

IBM product	Integration points
IBM BigInsights™	<ul style="list-style-type: none">● BigInsights is the IBM Hadoop platform.● Cloudbant is complementary to BigInsights: BigInsights handles analytics and Cloudbant handles transactional data.● The tight integration between Watson Foundations product architectures allows data from Cloudbant to be pushed into BigInsights for analytics.
IBM DB2® BLU	<ul style="list-style-type: none">● DB2 BLU is the IBM in-memory, high performance, relational database system (RDBMS) for analytics.● Data from Cloudbant can be loaded into DB2 BLU directly, or ingested via BigInsights, by using the interoperability of the Watson Foundations framework.
IBM InfoSphere® Information Server	<ul style="list-style-type: none">● Information Server is the IBM data integration platform.● Information Server is complementary to Cloudbant: Information Server can deliver trusted data from the enterprise to Cloudbant.● Clients who use Information Server as part of the enterprise data warehouse (EDW) and Analytics landscape will find it simple to import data from Cloudbant into that infrastructure.
IBM Worklight®	<ul style="list-style-type: none">● Worklight is the IBM platform for extending clients' businesses to mobile platforms.● IBM Worklight enables the development of HTML5, JavaScript, and native mobile applications on the front end, and integration with enterprise-scale data applications and services on the back end.● There is a Worklight adapter available for Cloudbant to facilitate service-based access to various components of Cloudbant's data layer.
IBM Bluemix	<ul style="list-style-type: none">● Bluemix is the IBM cloud-based delivery system for composable services.● Bluemix provides a marketplace where developers can rapidly provision, experiment, build, and test applications from a catalog of IBM and IBM partner-built services.● IBM Cloudbant will be one of many components deployable from Bluemix's catalog of composable services.

Supported platforms

The IBM Cloudant Database-as-a-Service (DBaaS) solution enables you to buy into a guaranteed data management service level agreement (SLA), rather than locking you into a database technology. Considerations must be made for a client's storage, throughput, latency, up-time, data access, and support requirements. Pricing and service tiers - offered without lock-in - are detailed in the following section.

Ordering information

You can try Cloudant at no charge at <https://cloudant.com/sign-up/>. Table 2 explains the available service tiers.

Table 2. Service tiers

Service tier	Features	Pricing
Enterprise (Dedicated) DBaaS, Single-tenant Cluster	<ul style="list-style-type: none">● Scalability: elastic scaling; handles billions of transactions per day.● Rich, NoSQL Database as a Service (DBaaS) API.● Guaranteed database performance and up-time.● Dedicated DBaaS cluster hardware.● Over 35 cloud hosting locations on IBM SoftLayer, Rackspace, AWS, or Azure.● Bare-metal performance on IBM SoftLayer or Rackspace.	Elastic: based on cluster size (number of server nodes in operation).
Multi-tenant Cluster	Usage is measured against three metrics: <ul style="list-style-type: none">● Data volume (in GBs per month).● "Heavy" API requests (including PUTs, POSTs, and DELETes). API requests that read or write multiple JSON docs in bulk are considered one API call per request.● "Light" API requests (including GETs and HEADs).	<ul style="list-style-type: none">● \$ 1.00 USD per GB/month● \$ 0.015 per 100 "heavy" requests● \$ 0.015 per 500 "light" requests● No charge ever if your monthly usage is under \$ 5.00.
Gold Support	<ul style="list-style-type: none">● Systems monitoring: APIs collect and report on system metrics (read load versus write load, disk saturation, cache utilization, and CPU core utilization).● Cluster resizing: Cloudant manages the expansion or downscaling of dedicated clusters, the reconfiguration of hardware, and the rebalancing of data - as required.● Continuous enhancement: code deployment (fixes, optimizations, and new features) across the data layer on a biweekly basis.	\$ 500 USD per month on Multi-tenant tier (included for no charge with Enterprise tier).

Related information

For more information about IBM Cloudant, see the following sources:

- IBM Cloudant website <https://cloudant.com>
- IBM Redpaper publication - IBM Cloudant: Database as a Service Fundamentals, REDP-5126-00 <http://www.redbooks.ibm.com/Redbooks.nsf/RedbookAbstracts/redp5126.html?Open>

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 give 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 Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites 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. 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.

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 measurement may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

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.

© Copyright International Business Machines Corporation 2014-2015. All rights reserved.

Note to U.S. Government Users Restricted Rights -- Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

This document was created or updated on March 26, 2015.

Send us your comments in one of the following ways:

- Use the online **Contact us** review form found at:
ibm.com/redbooks
- Send your comments in an e-mail 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 U.S.A.

This document is available online at <http://www.ibm.com/redbooks/abstracts/tips1187.html> .

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™
DB2®
IBM®
InfoSphere®
Redbooks®
Redbooks logo®
Worklight®

The following terms are trademarks of other companies:

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

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

Worklight is trademark or registered trademark of Worklight, an IBM Company.

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