IBM

Using zEnterprise for Smart Analytics

Volume 1 Assessment

- Evaluating the requirements of a Smarter Analytics workload
- Using the IBM Fit for Purpose methodology
- Deploying the workload on a zEnterprise system

Alex Louwe Kooijmans
Chris Hine
Colin Page
Concha Tourné
Eduardo C. Oliveira

Redbooks

ibm.com/redbooks

International Technical Support Organization

Using zEnterprise for Smart Analytics: Volume 1 Assessment

July 2013

SG24-8007-00

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

First Edition (July 2013)

This edition applies to the IBM zEnterprise system, IBM Cognos 10.1 Business Intelligence Server, SPSS Modeler 14.2, and IBM DB2.

Contents

Notices . . . vii
Trademarks . . . viii

Preface . . . ix
Authors . . . ix
Now you can become a published author, too! . . . xi
Comments welcome . . . xi
Stay connected to IBM Redbooks . . . xi

Chapter 1. Introduction and objectives . . . 1
1.1 Audience . . . 2
1.2 Objectives . . . 2
1.3 Overview of the book . . . 2
1.4 Some guidelines when reading the book . . . 2

Chapter 2. Current infrastructure . . . 3
2.1 Introduction to the Smarter Banking showcase . . . 4
2.1.1 Showcase objectives . . . 4
2.2 Showcase architecture . . . 5
2.3 IBM Banking Framework . . . 7
2.3.1 Core banking transformation . . . 7
2.3.2 Payments and securities . . . 8
2.3.3 Integrated risk management . . . 8
2.3.4 Customer care and insight . . . 9
2.4 Showcase infrastructure . . . 10
2.4.1 Hardware configuration . . . 11
2.4.2 Workload simulation . . . 12
2.4.3 Systems monitoring . . . 14
2.5 Product offerings for IBM Business Analytics Optimization . . . 16
2.6 Summary . . . 17

Chapter 3. Overview of the banking analytical workload . . . 19
3.1 Introduction to IBM business analytics . . . 20
3.2 IBM business analytics for banking . . . 20
3.3 IBM Customer Profitability Analytics blueprint . . . 21
3.3.1 Profitability calculator . . . 23
3.3.2 Reporting functionality . . . 24
3.3.3 Predictive analytics . . . 25
3.3.4 Summary of the blueprint . . . 26
3.4 Technical requirements . . . 27

Chapter 4. zEnterprise hybrid workload placement approach . . . 31
4.1 Overview of the approach . . . 32
4.1.1 Decision points . . . 32
4.2 Stage 1: Initiate project . . . 33
4.3 Stage 2: Perform zEnterprise business value assessment . . . 34
4.4 Stage 3: Assess new workload . . . 35
4.5 Stage 4: List solution architecture alternatives . . . 36
4.6 Stage 5: Prioritize solution architecture alternatives . . . 36

4.7 Stage 6: Sizing and capacity planning . . . 37
4.8 Stage 7: Execute total cost of ownership (TCO) analysis . . . 38
4.9 Stage 8: Create business case . . . 38

Chapter 5. Stage 1: Initiate project . . . 39
5.1 Overview . . . 40
5.1.1 Expectations regarding the outcome of the project . . . 40
5.1.2 Scope of the project . . . 40
5.1.3 Project organization . . . 40
5.1.4 Tools needed . . . 41
5.2 Inputs . . . 42
5.3 Outputs . . . 42
5.4 Tasks and decisions . . . 42
5.4.1 Identify and confirm the executive sponsor at the client organization . . . 43
5.4.2 Determine the key stakeholders and technical SMEs . . . 43
5.4.3 Establish the project organization . . . 43
5.5 Conclusions . . . 44

Chapter 6. Stage 2: Perform zEnterprise business value assessment . . . 45
6.1 Objectives . . . 46
6.2 Process overview . . . 46
6.3 Tools and accelerators . . . 47
6.4 Resources needed . . . 47
6.5 zEnterprise business value process steps . . . 48
6.5.1 Stage 1 . . . 48
6.5.2 Stage 2 . . . 51
6.5.3 Stage 3 . . . 52
6.6 Conclusions . . . 54

Chapter 7. Stage 3: Assess new workload . . . 55
7.1 Objectives . . . 56
7.2 Process overview . . . 56
7.3 Tools and accelerators . . . 57
7.4 Resources needed . . . 58
7.5 Detailed steps . . . 58
7.5.1 NFRs . . . 59
7.5.2 Component characterization . . . 60
7.5.3 Conclusion . . . 61

Chapter 8. Stage 4: List solution architecture alternatives . . . 63
8.1 Objectives . . . 64
8.2 Process overview . . . 64
8.3 Tools and accelerators . . . 64
8.4 Resources needed . . . 64
8.5 Detailed steps . . . 64
8.6 Rating platform options based on NFRs . . . 65
8.6.1 Data marts . . . 65
8.6.2 Cognos 10 BI . . . 66
8.6.3 Cognos TM1 . . . 67
8.6.4 SPSS Modeler Server . . . 68
8.7 Conclusions . . . 69

Chapter 9. Stage 5: Prioritize solution architecture alternatives . . . 71
9.1 Introduction . . . 72

9.2 Tools and accelerators 72
9.3 Resources needed 72
9.4 Detailed steps 72
9.4.1 Select the platform that is the best fit for each component 72
9.5 Conclusions 76

Chapter 10. Stage 6: Sizing and capacity planning 79
10.1 Overview 80
10.1.1 Sizing considerations and objectives 82
10.2 Process overview 82
10.2.1 Sizing an IT optimization (consolidation) environment 83
10.3 Tools and accelerators 85
10.4 Resources needed 85
10.5 Detailed steps 85
10.5.1 Sizing a new application 87
10.5.2 Total cost of ownership (or cost and value) 87
10.5.3 Capacity planning considerations 88
10.5.4 Cognos 10 BI sizing for Linux on System z: Input 90
10.5.5 Cognos 10 BI sizing on Linux on System z: Results 93
10.6 Conclusions 94

Chapter 11. Stage 7: Execute total cost of ownership (TCO) analysis 95
11.1 TCO versus CO 96
11.2 Objectives 97
11.3 Tools and accelerators 98
11.4 Resources needed 99
11.5 Detailed steps 99
11.5.1 Building the TCO model 102
11.6 Conclusions 105

Chapter 12. Stage 8: Create business case 107
12.1 Objectives 108
12.2 Inputs 108
12.3 Outputs 108
12.4 Tasks 108
12.4.1 Determine migration impact 109
12.4.2 Prepare business case 109
12.4.3 Present and discuss the business case 110
12.4.4 Review assumptions and variables 110
12.5 Tools and accelerators 110
12.6 Summary 110

Related publications 111
IBM Redbooks 111
Other publications 111
Online resources 111
Help from IBM 112

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.

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:

AIX®
BladeCenter®
CICS®
Cognos®
DataPower®
DB2 Connect™
DB2®
FICON®
GDPS®
Geographically Dispersed Parallel Sysplex™
HiperSockets™
HyperSwap®
IBM®
InfoSphere®
MVS™
OMEGAMON®
Parallel Sysplex®
Power Cube®
POWER7®
POWER®
RACF®
Rational®
Redbooks®
Redbooks (logo) ®
Resource Measurement Facility™
RMF™
Smarter Analytics™
Smarter Banking®
SPSS®
System x®
System z10®
System z®
Tivoli®
TM1®
WebSphere®
z/OS®
z/VM®
z10™
zEnterprise®

The following terms are trademarks of other companies:

Intel, Intel logo, Intel Inside logo, and Intel Centrino logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

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.

UNIX is a registered trademark of The Open Group in the United States and other countries.

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

Preface

This IBM® Redbooks® publication is the first of two volumes explaining the approach and implementation of a new hybrid workload integrated in the existing IBM Smarter Banking® showcase, which is hosted at IBM in Montpellier, France. The volumes contain the following information.

Volume 1 Assessment describes how to evaluate the requirements of a new IBM Smarter Analytics™ workload. It addresses the user, system resources, and data processing profiles to identify the optimal configuration using IBM methodologies, such as Fit for Purpose (F4P). Because the existing showcase is based on IBM zEnterprise®, the deployment options include IBM z/OS®, Linux on IBM System z®, IBM AIX® running on POWER7® blades within the zEnterprise BladeCenter® Extension (zBX), and Microsoft Windows 2008 Server running on Intel blades also within zBX.

Volume 2 Implementation describes the steps involved in deploying the Smarter Analytics workload in the showcase. With multiple components, including IBM Cognos® 10.1 Business Intelligence, IBM Cognos TM1®, Cognos Metrics Studio, IBM SPSS®, IBM DB2® for z/OS, and many application design tools, the workload spans multiple operating environments. Application clustering, setting up performance policies using Unified Resource Manager, and simulation test execution results are included.

This book is intended for an audience of professionals in an infrastructure architecture role. However, architects with a focus on business intelligence and analytics, as well as IT managers, will find value in reading this book. If you want to see this solution in action, contact your IBM representative or send an email directly to mopbcoe@fr.ibm.com.

Authors

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

Alex Louwe Kooijmans is a senior architect at the Financial Services Center of Excellence at IBM Systems & Technology Group. Prior to this position, he spent almost 10 years in the International Technical Support Organization leading IBM Redbooks projects, teaching workshops, and running technical events with a focus on using the IBM mainframe in new ways. Alex also worked as Client Technical Advisor to various banks in The Netherlands and performed various job roles in application development. His current focus is on modernizing core banking systems and the role of IBM's current mainframe technology.

Chris Hine is a System z IT Architect in the UK. He has 25 years of experience working in IBM in the mainframe field. He holds a PhD in Life Sciences from The University of Dundee, Scotland. His areas of expertise include System z architecture, data center consolidation, workload virtualization, and infrastructure testing programs.

Colin Page is the lead architect for the Smarter Banking showcase in IBM Montpellier, France, although he is originally from the UK. He is a Certified IT Specialist and System z Champion. He has over 20 years experience working on mainframe projects, mostly on client sites, specializing in DB2 for z/OS and now across all middleware and infrastructure components.

Concha Tourné is an System z IT Architect in Spain. She has more than 25 years of experience in IBM working in the mainframe field. She holds a degree in Physics from Universidad Complutense of Madrid, Spain. Her areas of expertise include System z architecture and workload optimization. She spent two years working on client activities in the IBM Products and Solutions Support Centre in Montpellier, France.

Eduardo C. Oliveira is an Executive IT Specialist in the United States. He has 27 years of experience on the System z architecture. He holds a Masters degree in Computer Science and Software Engineering from the University of Waterloo, Canada, and an undergraduate degree in Electrical Engineering from the State University of Campinas (UNICAMP), Brazil. His areas of expertise include sizing and capacity planning studies, total cost of ownership assessments, business analytics optimization (BAO), IBM smart analytics systems, IBM servers, z/VM®, and Linux. He has written extensively on server consolidation on Linux on System z, total cost of ownership, and capacity planning. His current position is in the Advanced Technical Skills (ATS) organization as part of the ATS Smarter Analytics Team.

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

Mike Ebbers
International Technical Support Organization, Poughkeepsie Center

Montgomery Bauman
IBM Certified IT Specialist, East Region, United States

Stephane Faure
IBM IT Architect, IBM Product and Solutions Support Center (PSSC), Montpellier, France

Fabrice Jarassat
IBM Certified IT Specialist, IBM Product and Solutions Support Center (PSSC), Montpellier, France

Manuel Jaén
IBM Executive Staff, Madrid, Spain

Yann Kindelberger
IBM European Design Center, Montpellier, France

Cedrine Madera
IBM Senior Information Architect, System z BI tiger team, Montpellier, France

Dennis Mosby
IBM Certified IT Specialist, Global Techline, United States

Robert Neidig
Program Director, zPO, United States

Adriana d'Aragona Oliveira
Independent editor and translator

Marie-Laure Pessoa Caucat
IBM System z New Technology Center, Montpellier, France

Nigel Williams
IBM Certified IT Specialist, IBM Product and Solutions Support Center (PSSC), Montpellier, France

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

- 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 and objectives

Within IBM's financial services division, several groups have come together to create a set of business analytics assets of specific use to the banking, insurance, and financial market industries. These assets are called *blueprints*.

In this IBM Redbooks publication, the first of two volumes, a blueprint designed to provide profitability information for both the client and commercial operational aspects of a retail bank is assessed for integration with an existing bank's information technology environment. The blueprint is known as the *Customer Profitability Analytics* (CPA) blueprint.

The CPA blueprint consists of several models developed using a common suite of IBM products. These are the principal components:

- IBM Cognos TM1 Server (9.5.0 or greater)
- IBM Cognos 10.1 Business Intelligence Server
- IBM Cognos 10.1 Metrics Server
- IBM DB2 Database 9.5 (Express-C or greater)
- IBM SPSS Modeler 14.2

The blueprint has been developed and made available to IBM sales support personnel to demonstrate to clients. The package contains sample data and is based on a Microsoft Windows 32-bit runtime environment. However, based on the findings of the assessment, we implemented the IBM CPA blueprint on several platforms within a zEnterprise system.

1.1 Audience

This book is primarily written for an audience of professionals in an infrastructure architecture role. However, architects with a focus on business intelligence and analytics, as well as IT managers, will find value in reading this book.

1.2 Objectives

The objectives of this book are two-fold:

- Show an end-to-end methodology that can be followed to make decisions about deploying a new or existing workload to zEnterprise.
- Demonstrate the value of placing an end-to-end business intelligence and analytics workload on zEnterprise and how to incorporate it into an existing banking IT environment.

1.3 Overview of the book

This book is organized as follows:

- Chapter 2, “Current infrastructure” on page 3 describes the current banking environment in place.
- The business intelligence and analytics workload in scope is an existing workload currently running on stand-alone Windows servers. Chapter 3, “Overview of the banking analytical workload” on page 19 describes the workload from a business and logical architecture perspective.
- Chapter 4, “zEnterprise hybrid workload placement approach” on page 31 provides an overview of the methodology used to get to the point of making a business decision on placement of a specific application or workload on zEnterprise.
- Chapter 5, “Stage 1: Initiate project” on page 39 through Chapter 12, “Stage 8: Create business case” on page 107 describe stages of the methodology in more detail. There is a separate chapter on each stage.

1.4 Some guidelines when reading the book

Here are some guidelines and thoughts while reading this book:

- The business and analytics workload is an example, and the information in this book can be applied to other workloads or applications as well.
- The methodology and its stages provide a logical and structured way to go through the process. We have given specific names to the stages, but each stage is a logical step in the process and cannot be skipped.
- We do not prescribe specific tools to be used in each stage, and you can use your own questionnaires, frameworks, and tools. However, IBM has a wealth of experience, tools, frameworks, and standard workshops to assist you in your decisions.

2

Current infrastructure

We started our project with an infrastructure with various banking solutions already in place. This infrastructure, which is also referred to as the “Smarter Banking showcase” has evolved over many years. This chapter describes the key business drivers for the Smarter Banking showcase and provides an overview of its architecture and current infrastructure.

2.1 Introduction to the Smarter Banking showcase

The IBM Smarter Banking showcase is a simulation of a real bank. A team of banking and IT infrastructure specialists working in the Banking Centre of Excellence at the IBM Montpellier location in France created and developed the showcase.

The banking systems that are used in the showcase run a mixed workload of real-world financial transactions including cash withdrawals, deposits, mortgages, and car loans. The COBOL core banking component runs on IBM CICS® and stores customer and account records in DB2 for z/OS.

The initial focus of the showcase team was to create a multi-channel core banking environment capable of running a representative banking workload. As the showcase has evolved, we adopted an approach based on service-oriented architecture (SOA) for reusing the core banking system. We also developed a comprehensive monitoring solution to provide a real-time status of the health of the IT infrastructure.

The fundamental components of multiple channel integration, and advanced systems management still form the basis of the showcase. However, the showcase also evolved in other directions. In particular, we added new scenarios based on products from Independent Software Vendors (ISVs). For example, the loan decision management processing is based on Chordiant Decision Manager from PegaSystems.

Several showcase components run on the zEnterprise Blade Extension (zBX).

2.1.1 Showcase objectives

The Smarter Banking showcase has the following objectives:

- Demonstrate a live banking operating environment that is based on the IBM Banking Framework.
- Share a vision of a modern, efficient, smarter bank that can cope with the IT challenges of today and position itself for the challenges of tomorrow.
- Show the linkage between infrastructure and business value to CIO and Line of Business (LoB) management through a set of scenarios that represent banking problem points and opportunity areas.
- Highlight the value of deploying hybrid workloads that span multiple platforms and architectures, on a zEnterprise infrastructure managed centrally with the Unified Resource Manager.

To achieve these objectives, we use a realistic mix of workloads and run these workloads at operational volumes representative of a typical European bank (Table 2-1).

Table 2-1 Simulated banking workload

Number of clients in database	Number of accounts in database	Transactions per second	Number of simulated clients
6 million	12 million	300 average 1,000 peak	7,400

For information about the breakdown of simulated banking operations, see Table 2-2 on page 13.

While the banking workload was running, we demonstrated the following proof points:

- The ability of the IT infrastructure to provide optimized customer service (fast and consistent response times) across multiple channels, while responding to varying peaks in branch, Internet, Point of Sale (POS), and ATM traffic
- Management and monitoring of physical blades, virtual servers, appliances, and workloads by using the Unified Resource Manager
- Key resiliency capabilities in planned and unplanned failure scenarios, while always retaining customer service
- How system capacity can be increased and removed to address peak workloads and immediate business opportunities
- Improved management of operational risk by defining a clear relationship between IT infrastructure and business services

2.2 Showcase architecture

We provide an overview of the existing showcase's operational and logical architecture. Figure 2-1 shows the multiple channel architecture and major components.

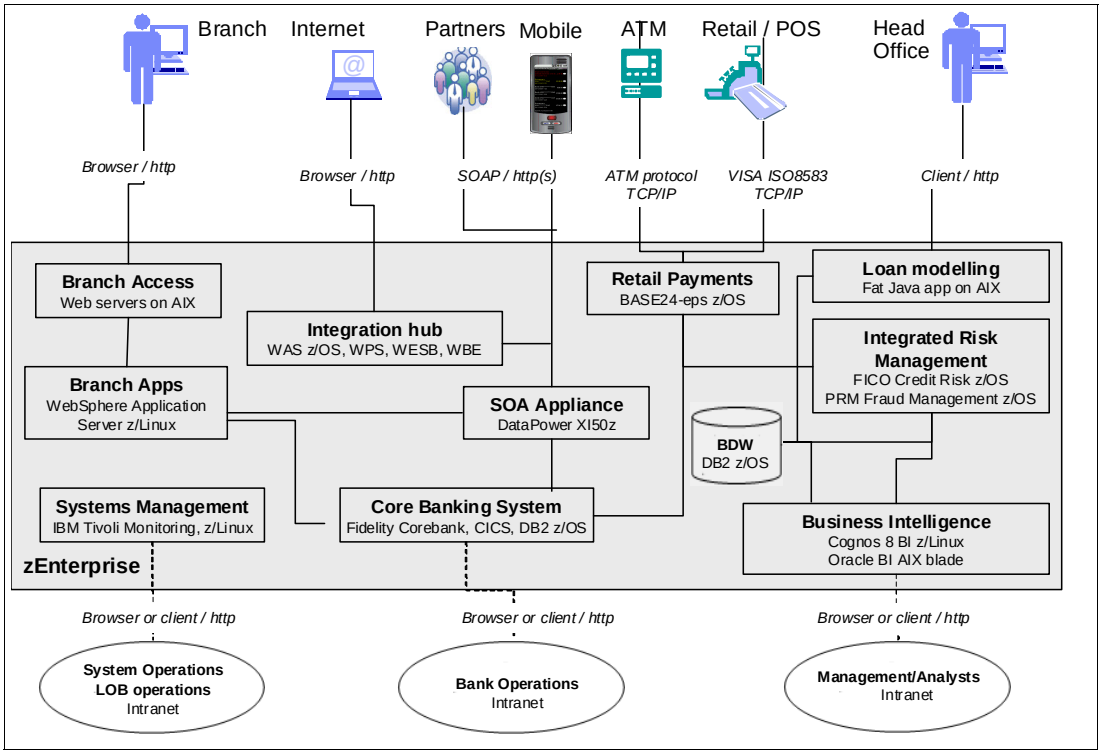


Figure 2-1 Architecture overview diagram

The following list shows the key IBM and ISV software components used in the showcase:

- Core banking system

  Fidelity Corebank V4.2 is a real-time retail banking application that is based on a physical implementation of the IBM Information Framework Financial Services Data Model (FSDM). The showcase implementation of Fidelity Corebank is based on CICS, COBOL, and DB2 running on System z. We store we store the 12 million accounts and the 6 million customers in the Fidelity Corebank application database.

The showcase uses a number of financial products available with Fidelity Corebank, including current accounts, savings accounts, term deposits, car loans, and mortgages. These represent a good selection of the core banking functionality normally found in a retail banking institution.

- Retail payments:
  - ACI Worldwide BASE24-eps V8.2 is used as a payment hub for the retail payments. BASE24-eps can handle the device protocols for most commercially available ATM and point of sale devices and for the standard interchanges, such as MasterCard and Visa. BASE24-eps handles the routing and authorization of these card transactions. In the Smarter Banking showcase implementation, BASE24-eps authorizes in real time against Fidelity Corebank. This authorization is made possible by using an External CICS Interface (EXCI) call from the C++ processes running in UNIX System Services on z/OS to CICS.
  - BASE24-eps runs natively on z/OS UNIX System Services (z/OS UNIX), is written in C++, and uses shared IBM WebSphere® MQ and DB2 structures in the coupling facility.
- Integrated risk management:
  - FICO TRIAD V8.0 is a risk calculation engine running in a COBOL batch environment on z/OS. FICO is used to generate the credit risk and probability of default scores for customers. This information is then stored in the Banking Data Warehouse (BDW), which is used as an analytical database.
  - ACI Worldwide Proactive Risk Manager (PRM) release 7.1 SP6 v8.2 provides fraud management capabilities for payment transactions. It works natively with BASE24-eps but it can also take feeds from any payment transaction source. PRM can calculate risk by using a choice of two basic techniques, either with neural network algorithms to detect suspicious patterns or with rules defined by the customer. The fraud detection processes can be invoked in real time to stop the transaction or in near real time so that analysts can block the card at a later time.

    PRM is written mostly in C++ and runs under z/OS UNIX using WebSphere MQ and DB2. A Java client runs in WebSphere Application Server.
- Business Intelligence:
  - Banking Data Warehouse (BDW), which is implemented using DB2 for z/OS. Initially, this database was built with the FICO interface to create a reporting solution around BASEL II[1]. Leveraging this customer information, we now use the same database for other purposes, such as operational statistics, operational transaction summaries, real-time general ledger positions, customer segmentation, and customer insight. We use near real-time feeds to update the BDW from our operational channels.
  - Business Intelligence (BI) is supported by a combination of Oracle BI, which provides an online dashboard reporting capability for BASEL II and customer segmentation, and Cognos 8 BI, which provides an Executive Dashboard on fraud and operational reporting for the payments infrastructure.

Note: The project outlined by this book is an enhancement to the existing BI implementation.

- Internet channel and integration hub

  WebSphere Application Server and IBM Business Process Manager run on z/OS and provide our central hub, which exposes the core functions of Fidelity Corebank as

[1] Basel II is the second of the Basel Accords, which are recommendations on banking laws and regulations issued by the Basel Committee on Banking Supervision.

services, and provides a standard way to create business processes. This provides key capabilities from WebSphere Process Server to support our enterprise integration and transaction process management requirements. We use the Service Component Architecture (SCA) to create mediation routines and Java Enterprise Edition applications, which are used by the Internet Banking channel and across other channels.

We have also deployed WebSphere Operational Decision Management as part of the integration hub so that business events can be collected in a central location for improved business process visibility and for making operational decisions.

- Branch servers:
  - WebSphere Application Server is also used on Linux on System z to provide a range of teller applications to the branch employees. Branch applications reuse the core banking functions by making service requests to the Core Banking System.
  - A WebSphere cluster running within a Linux on System z and z/VM environment allows us to consolidate the branch servers.
- Systems management

  IBM Tivoli® provides the systems management capability. The Tivoli Enterprise Portal, IBM OMEGAMON®, and Tivoli Enterprise Monitoring Agents monitor the banking infrastructure (see 2.4.3, “Systems monitoring” on page 14 for more information). These tools are typically used by operational analysts and systems programmers.

2.3 IBM Banking Framework

Within the Smarter Banking showcase, we talk about *proof points*, which are discrete scenarios that we use to illustrate ways to address the specific business problems faced by banks. We structure the proof points in relation to the four domains that constitute the IBM Banking Framework.

2.3.1 Core banking transformation

The initial proof points developed by the showcase team addressed some of the basic building blocks now found in the Core Banking Transformation framework:

- How to improve core banking process efficiency and reduce costs
- How to build flexible business processes
- How to efficiently reuse core banking functions

IT foundation transformation

The core banking platform runs on a simplified IT infrastructure, which uses the System z operating system and middleware capabilities to reduce operational cost and risk. The core banking system runs in a CICSplex for high availability and uses DB2 data sharing so that there is a single copy of customer and product data. IBM Rational® development tools are used to build business service components.

Core banking process agility

Through the use of WebSphere Process Server, we can quickly create business processes aligned to new business requirements. Business processes can access the core banking components as services. We have created a number of business processes that allow our multi-channel architecture to reuse the same core functionality and to provide business intelligence by updating the BDW in near real time.

Core banking application modernization

Fidelity Corebank is supplied with a comprehensive set of application programming interfaces (APIs) that allow different banking channels to access the core banking system. We have used these APIs to create a multi-channel architecture.

We have evolved toward a service-oriented architecture (SOA), reusing the APIs provided by Fidelity Corebank as our service building blocks. We have built composite applications using the Service Component Architecture (SCA). These composite applications provide additional functionality, but, at the same time, they reuse core system functions provided by Fidelity. We use an enterprise service bus (ESB) to provide intelligent routing and transformation. This flexible approach allows us to call other application functions provided by other ISVs to enrich our business processes without having to touch the Fidelity Corebank application.

2.3.2 Payments and securities

The showcase is focused on retail banking rather than wholesale banking. It does not address corporate services, derivative products, and the trading operations covered by the financial markets area. We do however deal with certain payment types, in particular card-based payments. Debit and credit card payments, either with the card present (ATM or POS) or not present (Internet or contactless) is a highly strategic area in retail banking, driven by opportunity and new regulations.

Retail payments

We implemented ACI Worldwide's BASE24-eps application to handle card payments and Proactive Risk Manager to monitor and alert on potential payment fraud. BASE24-eps authorizes each payment with Fidelity Corebank as the card issuer.

BASE24-eps is a secure, modern enterprise payment platform that can help a financial institution transform its payments strategy. In the showcase, we have enabled an ATM channel and the Visa interchange channel and we inject payments over both channels during a demonstration.

2.3.3 Integrated risk management

To address different aspects of risk, we implemented several proof points in the showcase.

Financial risk

We installed the FICO TRIAD product to provide credit risk and default probability scores on our loan portfolio.

Financial crimes

With the integration of ACI's Proactive Risk Management (PRM) with BASE24-eps, we can assess the risk of fraud to our enterprise from our card payment channels. PRM can receive feeds from any channel and, in time, our other channels will also be assessed here. PRM allows us to analyze transactions in real time or in near real time.

Fraud Analysts work the PRM queues to alert on potential fraud, and have the ability to stop a card if necessary. In real-time, PRM can block a transaction if it breaks specific rules.

We have exposed the metrics and statistics stored within the PRM DB2 database to a Cognos executive dashboard to give the banks executives a near real-time view of the current fraud exposure with the key performance indicators (KPIs) that are normally reported to the board.

Operational and IT risk

Using a variety of System z capabilities, including IBM Parallel Sysplex®, IBM Geographically Dispersed Parallel Sysplex™ (IBM GDPS®[2]) Hyperswap, System Automation, and IBM Tivoli Monitoring, we can monitor the service level agreements (SLAs) of our workload and react to planned and unplanned outages in order to maintain continuous availability.

Governance and compliance

The loan portfolios are extracted from the Fidelity database and stored in the BDW to address some of the BASEL II regulatory requirements. Oracle Business Intelligence provides dashboard analysis and reporting on BASEL II compliance.

Note: The BDW is part of the IBM Banking Industry Model, Information Framework (IFW).

The TRIAD product pulls data for our loan customers from the BDW and stores the resulting scores in this analytical database. We then use analytical tools to access the BDW to create the reports. The reports fundamentally address credit risk. However, through our analytical tools, we can also identify trends in the way that the bank has managed this risk, which then starts to address operational risk.

2.3.4 Customer care and insight

We previously discussed how we use the BDW for risk assessment. When creating an analytical warehouse, it is best to tackle the problem project by project. Have a clearly defined deliverable for each project and slowly build the warehouse, which is the approach that we took with the showcase BDW.

The BASEL II credit risk reporting was our first BDW project. We loaded customers (involved parties in the IFW model) and loan accounts (arrangements in the IFW model) to the database along with a number of more static reference tables, such as periods, rates, classifications, and locations. After we completed this project, we were able to reuse this data for other projects and slowly build more content, views, accumulations, and summaries into the database.

The creation of a BDW helped us to build a single view of a customer and to address some of the customer care and insight challenges described next.

Marketing process optimization

Using Oracle Business Intelligence and its marketing functions, we were able to perform customer segmentation based on the customer data that we stored in the BDW. We were able to define a pool of customers with similar characteristics to use, for example, to launch a new marketing campaign for a new product. In our case, we were able to leverage the credit risk scoring analysis to help the segmentation process.

Customer information optimization

The BDW does not just store static data. We have several near real-time feeds that provide extract, transform, and load (ETL) processes to load operational data into the warehouse.

We use WebSphere MQ as a technique to separate the synchronous unit of work for the real banking channel response to the customer or branch, from the asynchronous, near real-time update of the warehouse. This provides us with current customer data.

[2] Geographically Dispersed Parallel Sysplex (GDPS) is the ultimate Disaster Recovery and Continuous Availability solution for a System z multi-site enterprise.

Cognos 8 Business Intelligence (BI) is used to provide an executive dashboard showing fraud key performance indicators (KPIs). We used the same analytics server to also provide an operational dashboard for payments based on the near real-time feeds to the BDW.

Multi-channel transformation

We enabled multi-channel transformation through the reuse of the core banking services across all of the channels. New channels can be added with a rich functionality after the new channel access protocols are established.

2.4 Showcase infrastructure

The Smarter Banking showcase highlights some of the zEnterprise unique qualities that enable a smarter computing infrastructure, including service management, virtualization, business and information resiliency, and energy efficiency.

The showcase runs in the Montpellier Green Data Center and inherits the energy monitoring capabilities of this new innovative center. The IT infrastructure is based primarily on the System z196 and the zBX. With its increased capacity and number of available processor cores per server, and with reduced energy usage and floor space, the z196 can be a perfect fit for Green IT.

Figure 2-2 shows a simplified view of the operational model, illustrating the main operating environments used in the showcase.

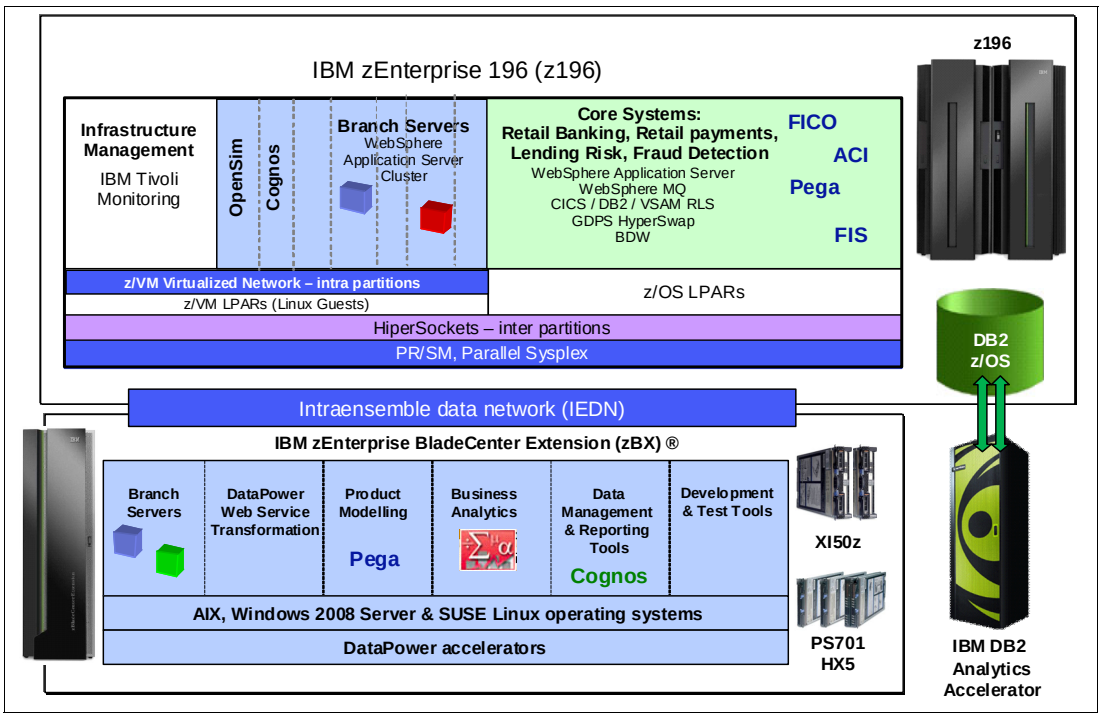


Figure 2-2 Operational model

The System z196 is shared with other projects and there are several logical partitions (LPARs) and sysplexes defined in this single machine. The production environment uses three z/OS LPARs. One LPAR runs IBM GDPS HyperSwap® software and the other two are the key application owning LPARs where the banking workloads run. The systems running

within the Parallel Sysplex access the same shared operational data, using the Coupling Facility (CF) to share resources.

IBM MVS™ Workload Manager (WLM) is used to manage the workloads that are injected into the demonstration environment and each request is classified and assigned a performance goal, based on channel and customer status. The performance goals are then monitored by using the IBM Tivoli Monitoring infrastructure.

Increasingly, we use Linux on System z environments hosted under z/VM. Initially, we established a WebSphere Network Deployment cluster of four application server nodes to represent our branch servers on Linux on System. We have since added some of the major components of IBM Tivoli Monitoring to Linux.

The zEnterprise Blade Extension (zBX) hosts several showcase components:

- IBM Power blades host IBM HTTP servers that are used for branch access, and the Pegasystems Visual Business Director server that is used for loan product forecasting and profitability simulation.
- IBM System x® blades host a range of business applications running on Linux and Microsoft Windows.
- IBM DataPower® XI50z blades perform security and data transformation for Web Services that are made available to the bank's business partners.

2.4.1 Hardware configuration

The hardware configuration including our two storage devices is shown in Figure 2-3.

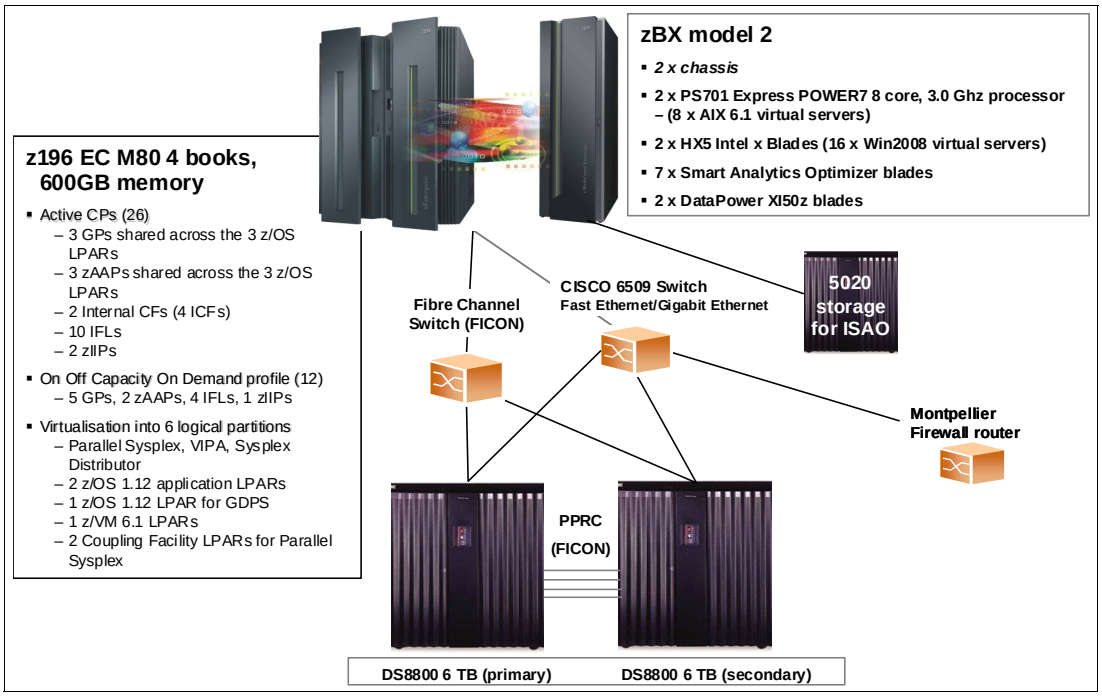


Figure 2-3 Hardware configuration

We use *Peer-to-Peer Remote Copy* (PPRC) between the two storage devices. PPRC mirrors primary disk updates to the secondary disk, so that an exact replica always exists in the event of an unplanned or planned outage, allowing us to swap to our secondary disk if necessary.

The main hardware configuration includes the following components:

- System z196 Model M80

  The processors of the z196 machine are shared across different LPARs, including two showcase environments:

  - Production sysplex (ZBPLEX) consisting of three z/OS 1.13 LPARs ZB01, ZB02, and ZB03
  - Development sysplex (BAPLEX) consisting of two z/OS 1.13 LPARs BA01 and BA02
- One z/VM 6.1 LPAR for development and one for production, each with the following Linux guests:
  - Six SUSE V10 Linux guests used for the WebSphere Network Deployment cluster (one HTTP server, one deployment manager, and four application servers) for the branch servers
  - One SUSE V10 Linux guest for the Cognos 8 BI server
  - One SUSE V10 Linux guest for the OpenSim 3D environment
  - Two SUSE V10 Linux guests for the Tivoli Enterprise Portal Server and IBM Tivoli Monitoring
- One zBX with two IBM Power7 blades, two IBM HX5 blades, and two DataPower XI50z blades
- One coupling facility LPAR for development and two for production
- Two DS8300 storage devices with 6.6 terabytes (TB) of data each
- Four IBM FICON® channels between the DASD devices
- One CISCO 6509 switch

The environment is designed to be resilient and highly available by using many of the technologies available natively with IBM System z:

- IBM Geographically Dispersed Parallel Sysplex (GDPS) HyperSwap for disk resiliency
- Parallel Sysplex for server resiliency
- Sysplex Distributor for workload distribution and high availability
- Virtualization to allow dynamic resource allocation and sharing of resources
- IBM HiperSockets™ and virtual local area network (VLAN) for virtualization of the network within the IBM System z10® machine
- Workload Manager (WLM) for transaction-based workload management on z/OS and machine priorities on z/VM

2.4.2 Workload simulation

Rational Performance Tester is used to inject the online workload into the banking system. The online workload is derived from research into actual banking workloads and is designed to be a reasonable representation of a typical day's online activity, with a transaction mix covering balance inquiries, statement requests, cash transactions, transfers, and check deposits.

Figure 2-4 on page 13 shows the exact transaction mix used in the demonstration, and the various channels that are simulated; branch, Internet, partner, retail payments, and ATM.

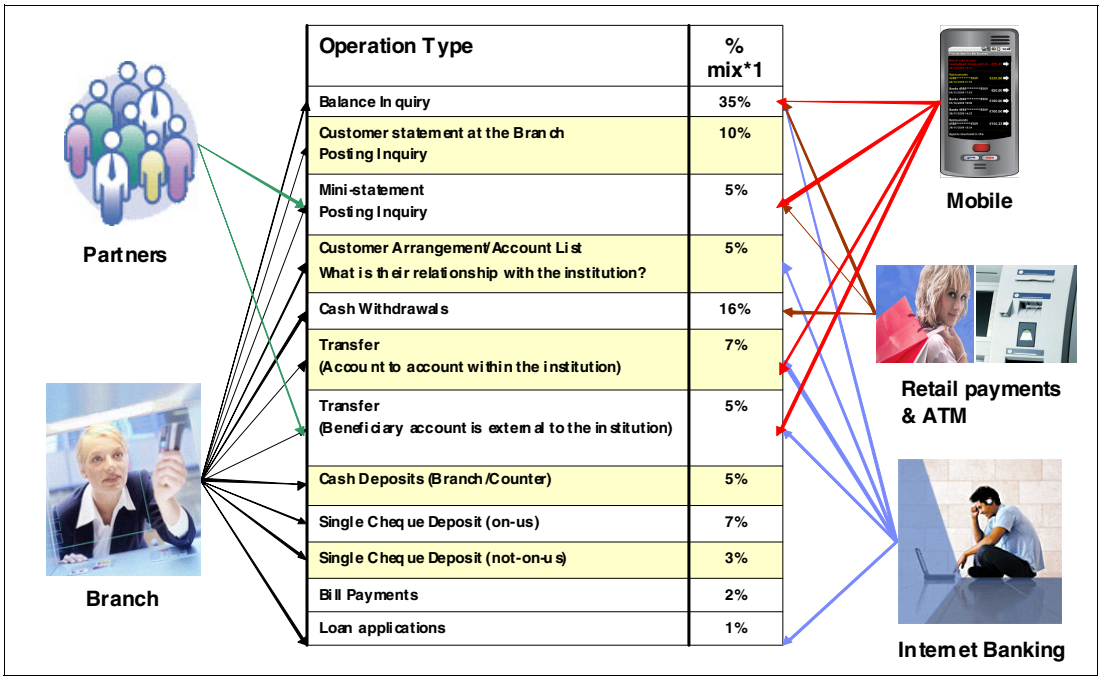


Figure 2-4 Workload simulation

All online transactions, whatever the channel, are processed by the core banking system (Fidelity Corebank), which runs in CICS. Each transaction has a unique transaction type as shown in Table 2-2.

Table 2-2 Simulated banking transactions

Transaction type	Description
BI	Balance Inquiry.
PIC	Posting Inquiry. Customer statement at the branch. Shows the last 20 transactions in detail.
PIA	Posting inquiry. Mini-statement at the ATM. Shows the last eight transactions in summary.
CAL	Customer arrangement list. Customer relationship with the bank.
CW	Cash withdrawal. ATM and counter.
TWF	Account to account transfer where both accounts are within the Smarter Banking financial institution.
TOF	Account to account transfer where the beneficiary account is in another financial institution.
CD	Cash deposit at the branch.
CDO	Check deposit on us (drawn on the Smarter Banking financial institution).
CDT	Check deposit on them (drawn on another financial institution).
BP	Bill payments.

We also inject business transactions manually, either by using the Smarter Banking web application or by using a real ATM or a virtual ATM in OpenSim. Transaction simulation shows the range of financial transactions that are simulated by the web application.

Figure 2-5 Transaction simulation

During a typical demonstration, we inject a workload in the range of 300 - 1,000 financial transactions per second.

2.4.3 Systems monitoring

The showcase uses many Tivoli System Management products and monitoring tools to track, monitor, and react to different infrastructure events that can occur in the course of a normal working day:

- IBM Tivoli Monitoring Services V6.2.3 is the foundation product and includes three components:
  - IBM Tivoli Enterprise Portal (TEP)
  - IBM Tivoli Enterprise Portal Server (TEPS)
  - IBM Tivoli Enterprise Monitoring Server (TEMS)

  Figure 2-6 on page 15 shows how we have deployed these components, specifying one IBM Tivoli Enterprise Monitoring Server as the hub to which all other IBM Tivoli Enterprise Monitoring Servers send monitored data. The Tivoli Enterprise Monitoring Agents (TEMAs) and OMEGAMON agents (for z/OS) send metrics from the individual subsystems and operating systems to the IBM Tivoli Enterprise Monitoring Server. The IBM Tivoli Enterprise Portal Server then extracts the data from the IBM Tivoli Enterprise Monitoring Server hub.

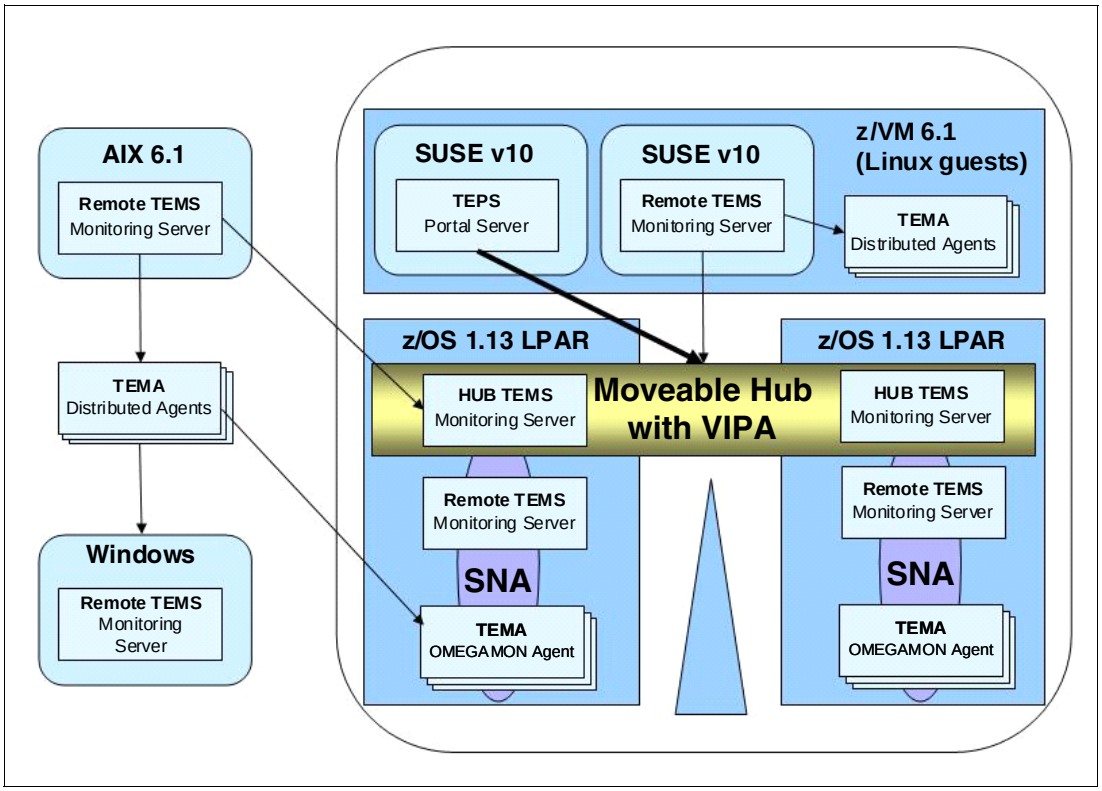


Figure 2-6 Systems monitoring infrastructure

- Figure 2-6 shows the moveable IBM Tivoli Enterprise Monitoring Server hub, which provides us with a more highly available solution. If the IBM Tivoli Enterprise Monitoring Server hub (a started task on z/OS) has a problem, Automatic Restart Manager (ARM) will restart another instance of the IBM Tivoli Enterprise Monitoring Server hub on another z/OS LPAR.
- The following list of OMEGAMON agents and IBM Tivoli Enterprise Monitoring Servers have been implemented as part of the monitoring solution:
  - IBM Tivoli OMEGAMON XE for CICS on z/OS V4.2.0
  - IBM Tivoli OMEGAMON XE on z/OS V4.2.0
  - IBM Tivoli OMEGAMON XE for DB2 V5.1.0
  - IBM Tivoli OMEGAMON XE for Messaging V7.0
  - IBM Tivoli Composite Application Manager for Application Diagnostics V7.1
  - IBM Tivoli Composite Application Manager for SOA V7.1
  - IBM Tivoli Composite Application Manager for DataPower V7.1
  - IBM Tivoli Enterprise Monitoring Agent for Windows V6.2.3
  - IBM Tivoli Enterprise Monitoring Agent for AIX V6.2.3
  - IBM Tivoli Enterprise Monitoring Agent for Linux V6.2.3

The monitoring infrastructure flags technical infrastructure events, such as a CICS failure or a network component failure, and sends alert events to the Tivoli Enterprise Portal. Figure 2-7 on page 16 shows how the Tivoli Enterprise Portal is used to provide a high-level view of the current health of the infrastructure.

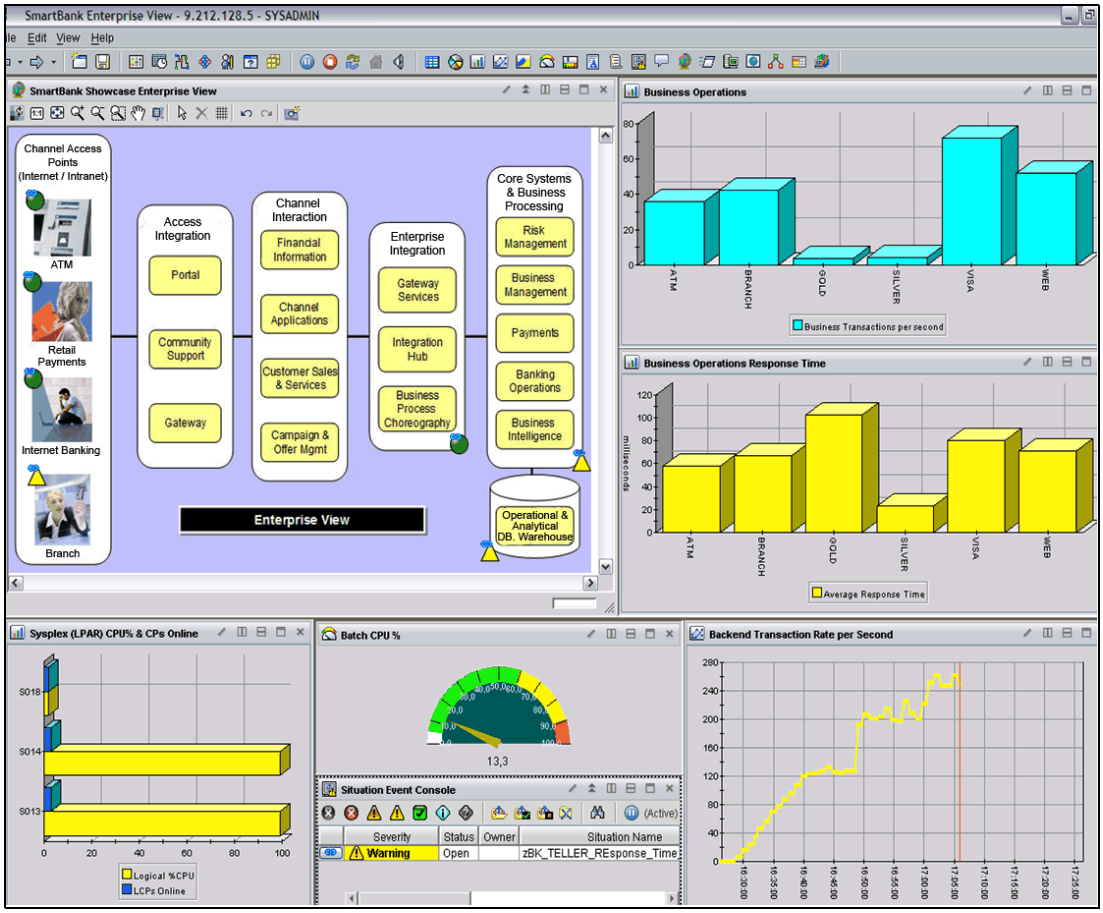


Figure 2-7 Tivoli Enterprise Portal

Figure 2-7 shows the following monitoring information:

- Channel status
- Transactions per second by channel
- Transaction response time by channel
- Transaction rates by time of day
- Batch CPU usage
- Infrastructure events
- Overall sysplex CPU usage

2.5 Product offerings for IBM Business Analytics Optimization

IBM offers the following products for *IBM Business Analytics Optimization* (BAO):

- zEnterprise Analytics System 9700/9710
- IBM DB2 Analytics Accelerator for z/OS
- IBM zEnterprise System
- Information Server

2.6 Summary

In this chapter, we provided an overview of the Smarter Banking showcase, its architecture, and current IT infrastructure, which is based on zEnterprise.

In the subsequent chapters of the book, we describe our new business analytics application and analyze how best to deploy the new workload to meet the associated nonfunctional requirements.

3

Overview of the banking analytical workload

This chapter describes the proposed workload that is to be assessed and deployed in the Smarter Banking showcase.

3.1 Introduction to IBM business analytics

IBM business analytics delivers complete, consistent, and accurate information that decision-makers trust to improve business performance. A comprehensive portfolio of business intelligence, predictive analytics, financial performance, strategy management, and analytical applications provides clear, immediate, and actionable insights into current performance and the ability to predict future outcomes.

IBM has developed a comprehensive approach to support the growing analytics demands. We refer to this as *IBM Business Analytics Optimization* (BAO).

The IBM BAO reference architecture describes the major foundational components, providing a common language to explain the architectural components within a framework that enables scope identification, roadmap definition, risk assessment, and gap assessment. See Figure 3-1.

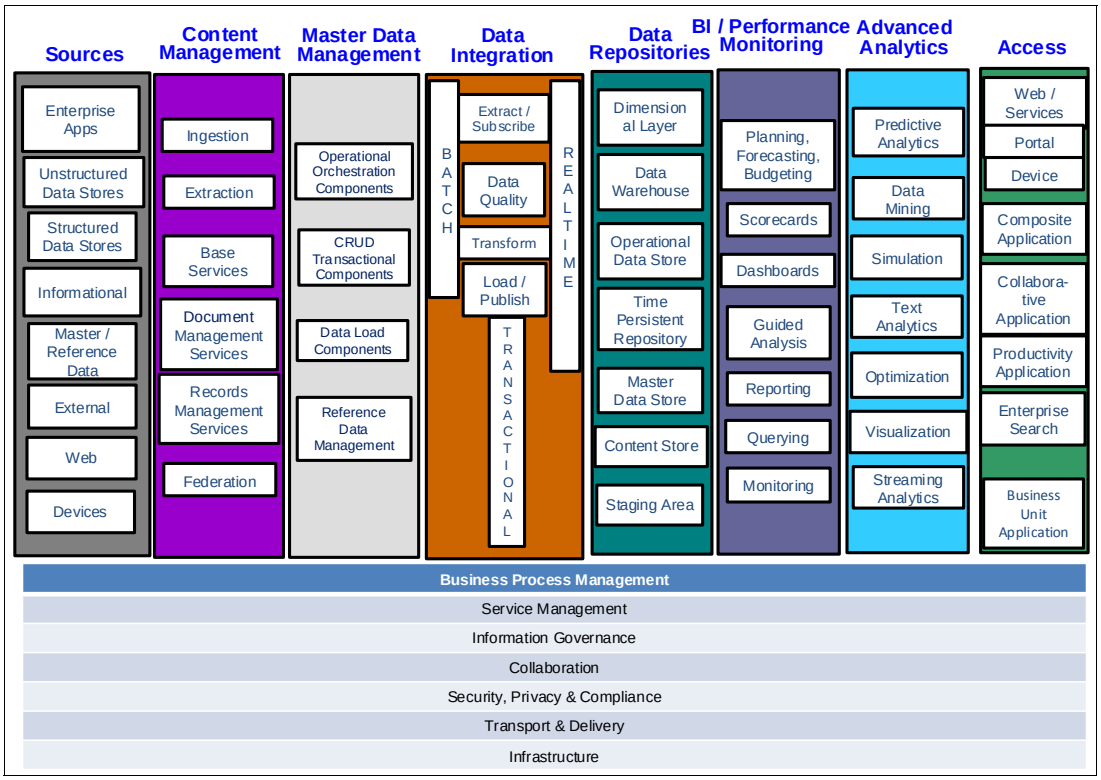


Figure 3-1 IBM BAO reference architecture

3.2 IBM business analytics for banking

Several groups within the IBM financial services division have collaborated to create a set of business analytics assets of specific use to the banking, insurance, and financial market industries. These assets are called *blueprints* and include, for example, sales reports, regulatory and compliance reporting, and detailed cost and value assessment reporting. For more information on the business analytics for banking blueprints, see this website:

http://www-01.ibm.com/software/analytics/banking/

These blueprints have been developed on a common framework mapped to the BAO reference architecture (see Figure 3-1 on page 20) to ensure consistent standards and integration across shared components. They are interchangeable in the same infrastructure and can be implemented en-masse or as individual projects. One of the key objectives of this IBM Redbooks project is to implement a common set of tools to support a range of financial services.

This book involves the assessment of a blueprint designed to provide profitability information for both the customer and commercial operational aspects of a retail bank. It is known as the *Customer Profitability Analytics* (CPA) blueprint and is described in detail in the following chapter.

Note: Although this solution is one of a number of blueprints developed for retail banking, the requirements, assessment, and deployment steps that we describe in these two books can be equally applied to any industry that requires a comprehensive business analytics solution.

3.3 IBM Customer Profitability Analytics blueprint

With the IBM Customer Profitability Analytics (CPA) solution, you can maximize the value of your customer base by understanding the interconnected factors that influence profitability through a single at-a-glance view.

The blueprint can provide answers to several business questions:

- Which geographies are profitable, how profitable are they, and why?
- Which products are profitable, how profitable are they, and why?
- Which customers are profitable, how profitable are they, and why? What will they do next and how do we anticipate their needs?
- How can we improve our customer retention, upselling, and account acquisition?
- Which business channel is more profitable?
- Which customer industry segments are profitable, how profitable are they, and why?
- Who are the Relationship Officers contributing to the bank's profitability and who are the ones needing more attention?

Figure 3-2 on page 22 shows an example of a CPA dashboard.

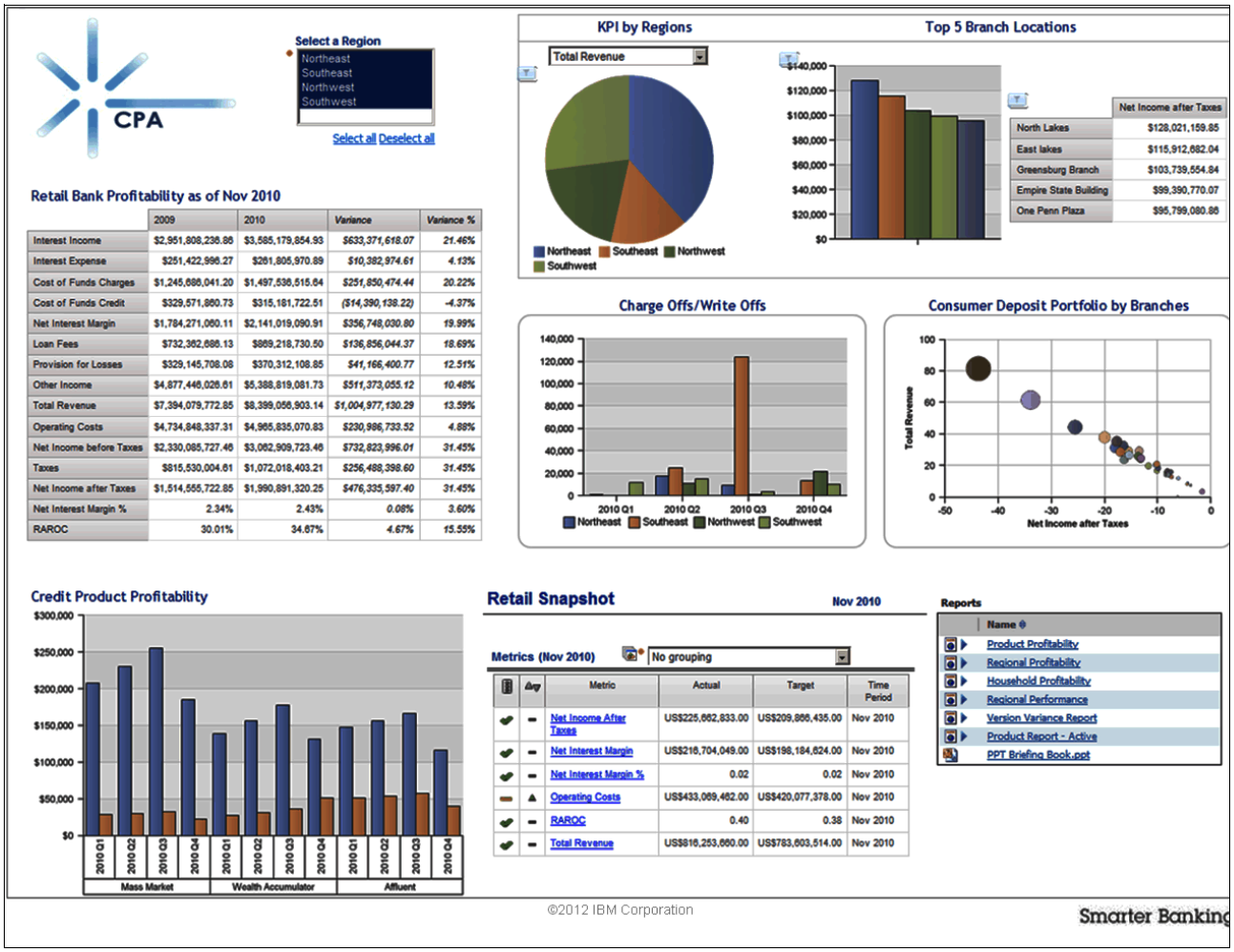


Figure 3-2 Customer Profitability Analytics retail dashboard

The blueprint offers these features:

- Comprehensive executive dashboards to highlight the current state of both customer (Figure 3-2) and commercial business areas (Figure 3-3 on page 23).
- Profit and loss statements at the individual account level, aggregated customer information including household data, personal, and business connections, line of business, products owned, demographics, and other profitability factors that create a 360-degree view of customers via a self-service portal.
- Predictive modeling so you know which products and services are most important to customers and will provide the biggest returns for the bank.
- With deep customer insights and the ability to predict customer lifetime value, you can make smarter decisions and execute strategies that will help increase the profitability of your entire customer base.

	2009	2010	Variance	Var %
Interest Income	$167,489,224	$204,251,814	$36,762,590	21.95%
Interest Expense	$16,029,783	$16,644,537	$614,755	3.84%
Cost of Funds Charges	$71,887,095	$87,267,564	$15,380,468	21.40%
Cost of Funds Credit	$21,716,113	$20,745,558	($970,555)	-4.47%
Net Interest Margin	$101,288,459	$121,085,271	$19,796,812	19.54%
Loan Fees	$42,475,484	$50,870,484	$8,395,000	19.76%
Provision for Losses	$21,676,728	$24,274,169	$2,597,441	11.98%
Other Income	$330,476,805	$369,087,113	$38,610,308	11.68%
Total Revenue	$474,240,748	$541,042,868	$66,802,121	14.09%
Operating Costs	$327,774,771	$329,658,338	$1,883,567	0.57%
Net Income before Taxes	$124,789,249	$187,110,361	$62,321,112	49.94%
Taxes	$43,676,237	$65,488,626	$21,812,389	49.94%
Net Income after Taxes	$81,113,012	$121,621,735	$40,508,723	49.94%
Net Interest Margin %	2.30%	2.37%	0.07%	3.04%
RAROC	24.51%	32.18%	7.67%	31.27%

	% of Exposure
Energy and Utilities	38.33%
Chemical and Petroleum	21.82%
Automotive	12.94%
Aerospace and Defense	11.28%
Consumer Products	8.49%
Electronics	5.46%
N/A	1.56%
Financial Services	0.09%
Education	0.03%

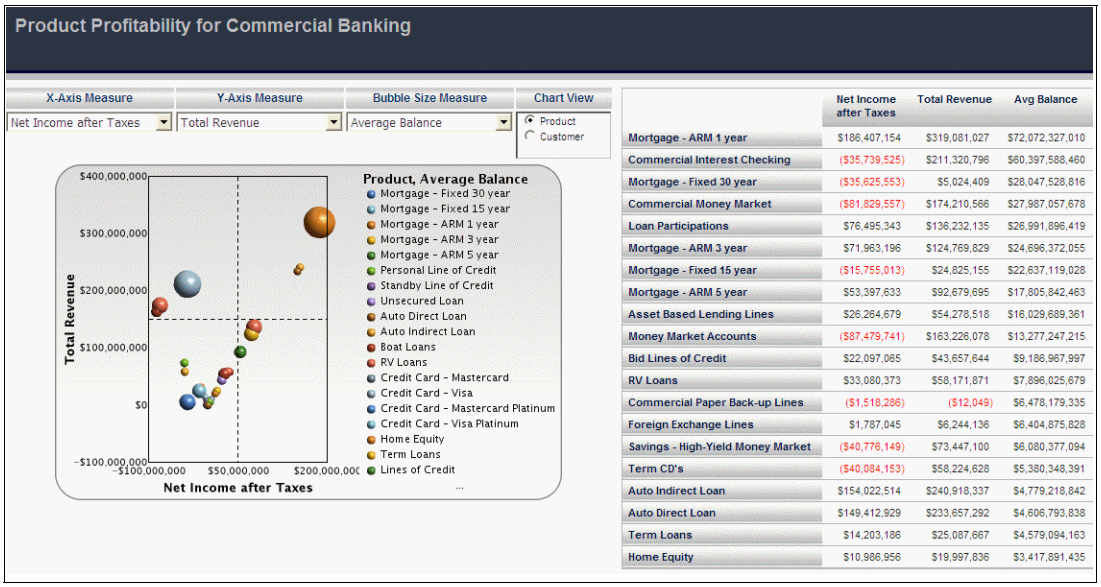
Figure 3-3 Customer Profitability Analytics commercial dashboard

The IBM Customer Profitability Analytics solution consists of three major components on an integrated software platform to provide the bank with credible, insightful, and actionable profitability information to drive better business outcomes.

3.3.1 Profitability calculator

The foundation of the solution is to provide the bank with a profitability calculator that can efficiently produce customer account profitability statements. The solution calculates and applies best practice management accounting theory to calculate measurements, such as cost of funds, risk-adjusted return on capital (RAROC), provision for losses, and non-interest expenses using both activity-based costing and allocation methodology. The solution also provides the calculation of many key performance ratios (RAROC, return on equity, expense ratio, loan yields, deposit rates, return on asset, and so on).

It is critical to perform these computations at the instrument or customer account level. Then, use the calculation engine to summarize these results in an integrated fashion across all of the key dimensions to support best practice profitability reporting and analytics. Cognos TM1 (Figure 3-4 on page 24) and Metrics Studio provide an advanced analytics design and delivery toolkit to deliver these calculations.



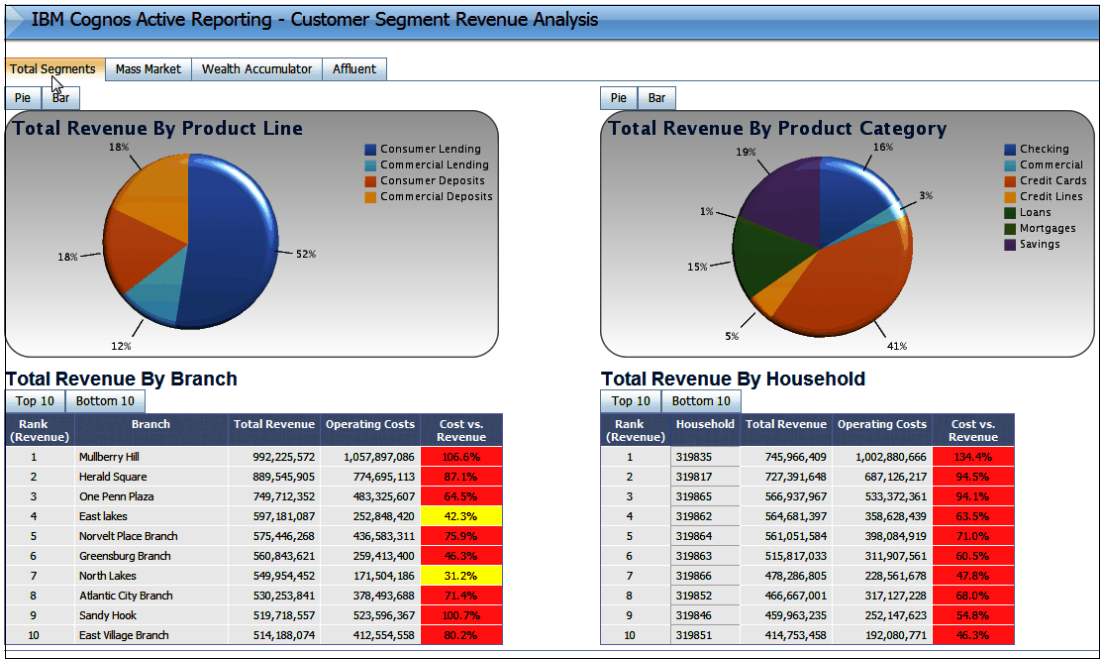
	Net Income after Taxes	Total Revenue	Avg Balance
Mortgage - ARM 1 year	$186,407,154	$319,081,027	$72,072,327,010
Commercial Interest Checking	($35,739,525)	$211,320,796	$60,397,588,460
Mortgage - Fixed 30 year	($35,625,553)	$5,024,409	$28,047,528,816
Commercial Money Market	($81,829,557)	$174,210,566	$27,987,057,678
Loan Participations	$76,495,343	$136,232,135	$26,991,896,419
Mortgage - ARM 3 year	$71,963,196	$124,769,829	$24,696,372,055
Mortgage - Fixed 15 year	($15,755,013)	$24,825,155	$22,637,119,028
Mortgage - ARM 5 year	$53,397,633	$92,679,695	$17,805,842,463
Asset Based Lending Lines	$26,264,679	$54,278,518	$16,029,689,361
Money Market Accounts	($87,479,741)	$163,226,078	$13,277,247,215
Bid Lines of Credit	$22,097,065	$43,657,644	$9,186,967,997
RV Loans	$33,080,373	$58,171,871	$7,896,025,679
Commercial Paper Back-up Lines	($1,518,286)	($12,049)	$6,478,179,335
Foreign Exchange Lines	$1,787,045	$6,244,136	$6,404,875,828
Savings - High-Yield Money Market	($40,776,149)	$73,447,100	$6,080,377,094
Term CD's	($40,084,153)	$58,224,628	$5,380,348,391
Auto Indirect Loan	$154,022,514	$240,918,337	$4,779,218,842
Auto Direct Loan	$149,412,929	$233,657,292	$4,606,793,838
Term Loans	$14,203,186	$25,087,667	$4,579,094,163
Home Equity	$10,986,956	$19,997,836	$3,417,891,435

Figure 3-4 Cognos TM1 analysis report

3.3.2 Reporting functionality

The second major component is a robust reporting capability that exposes the profitability information to the correct business user at the correct time to enhance business outcomes. The solution includes executive dashboards and scorecards for the bank's major lines of business, as well as financial/marketing analyst dashboards. The reporting capabilities include many functions that highlight key performance indicators (KPIs) that pertain to the business user's role.

The reporting capability is delivered in the IBM Cognos 10 technology software platform. This is a series of new business analytics innovations from IBM that changes how the bank will make decisions, allocate resources, predict and plan the future, and ultimately enhance its competitive advantage. IBM Cognos 10 revolutionizes how banking organizations use business intelligence. See Figure 3-5 on page 25.



Rank (Revenue)	Branch	Total Revenue	Operating Costs	Cost vs. Revenue
1	Mullberry Hill	992,225,572	1,057,897,086	106.6%
2	Herald Square	889,545,905	774,695,113	87.1%
3	One Penn Plaza	749,712,352	483,325,607	64.5%
4	East lakes	597,181,087	252,848,420	42.3%
5	Norvelt Place Branch	575,446,268	436,583,311	75.9%
6	Greensburg Branch	560,843,621	259,413,400	46.3%
7	North Lakes	549,954,452	171,504,186	31.2%
8	Atlantic City Branch	530,253,841	378,493,688	71.4%
9	Sandy Hook	519,718,557	523,596,367	100.7%
10	East Village Branch	514,188,074	412,554,558	80.2%

Rank (Revenue)	Household	Total Revenue	Operating Costs	Cost vs. Revenue
1	319835	745,966,409	1,002,880,666	134.4%
2	319817	727,391,648	687,126,217	94.5%
3	319865	566,937,967	533,372,361	94.1%
4	319862	564,681,397	358,628,439	63.5%
5	319864	561,051,584	398,084,919	71.0%
6	319863	515,817,033	311,907,561	60.5%
7	319866	478,286,805	228,561,678	47.8%
8	319852	466,667,001	317,127,228	68.0%
9	319846	459,963,235	252,147,623	54.8%
10	319851	414,753,458	192,080,771	46.3%

Figure 3-5 Cognos BI reporting: Customer segment revenue analysis

3.3.3 Predictive analytics

The third major component is the ability to apply predictive analytics to the historical profitability information and customer attribute data to predict what will happen next. The information you captured and calculated can now be analyzed by predictive models that help you understand and anticipate what customers want and will do next. These models use advanced analytics to uncover patterns in the data that help predict the likelihood of future events. For example, you can use predictive analytics to segment your customers based on profitability and then target each customer within a segment with the most relevant recommendation that is driven through predictive analytics to increase customer profitability and improve customer loyalty and satisfaction.

IBM SPSS predictive analytics technology provides a comprehensive tool to support this requirement. See Figure 3-6 on page 26.

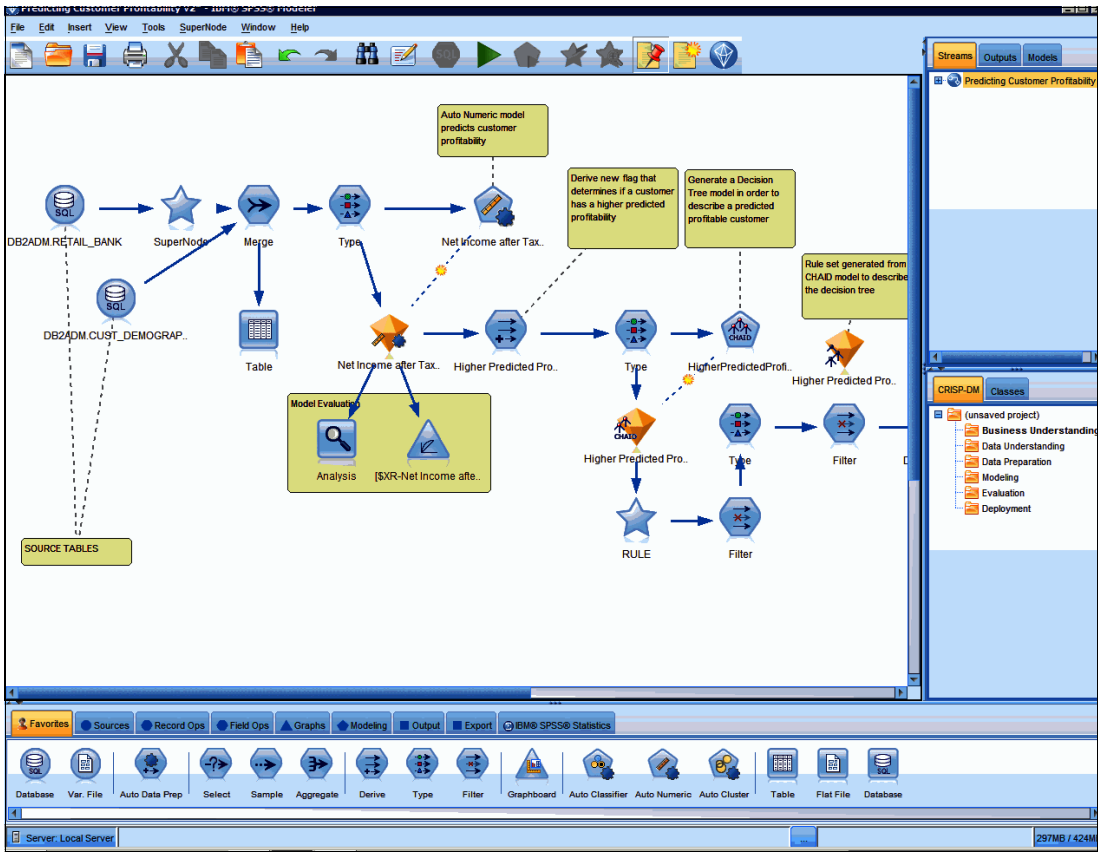


Figure 3-6 SPSS Modeler: Profitability model

3.3.4 Summary of the blueprint

The IBM Customer Profitability Analytics blueprint provides the Smarter Banking showcase with the following capabilities:

- Calculate, monitor and maximize the profitability of each customer
- Provide an integrated view of profitability across all key dimensions:
  - Product
  - Line of business/organization
  - Customer
  - Geography
  - Time
- Understand customer needs and propensity to buy new products leveraging predictive analytics
- Analyze real-time “What if?” profitability scenarios
- Increase customer lifetime value
- Perform more effective profitability forecasting and planning
- Understand the product profile of each customer relationship in its entirety

3.4 Technical requirements

The IBM CPA blueprint consists of a number of models developed using a common suite of IBM products. These are the principal components:

- IBM Cognos TM1 Server (9.5.0 or greater)
- IBM Cognos 10.1 Business Intelligence Server
- IBM Cognos 10.1 Metrics Server
- IBM DB2 Database 9.5 (Express-C or greater)
- IBM SPSS Modeler 14.2

The blueprint has been developed and made available to IBM sales support personnel to demonstrate to clients. The package contains sample data and is based on a Microsoft Windows 32-bit runtime environment (Figure 3-7).

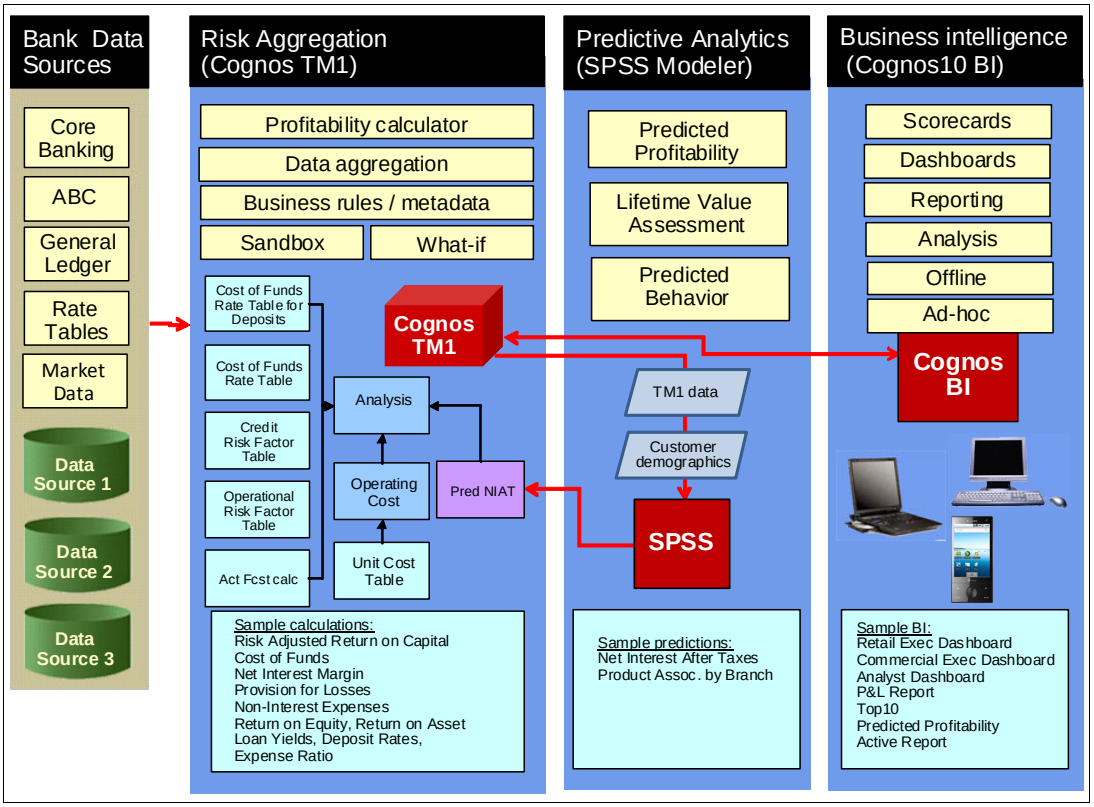


Figure 3-7 Customer Profitability Analytics blueprint: Logical architecture

Because the Smarter Banking showcase is a realistic representation of a mid-sized retail bank, the requirement is to implement this package in the zEnterprise based runtime environment and scaled to support the volumes of customer, product, and transaction data available.

This presents a number of key technical challenges:

- Users

  Support a variety of user roles: consumers, recipients, power users, and developers

The showcase simulates up to 8,000 concurrent users accessing multiple channels, of which the new “analytics” channel is expected to support up to 2,000 concurrent users of various role types.

- Transactions

  Support both batch and online report requests, many of which include active and online analytical processing (OLAP) reporting. Support dynamically executed reports based on drop-down selections (dimensions) or additional parameters

  The showcase simulates online transactions. However, there is also a suite of batch processing jobs that will be enhanced to support the data warehouse and reporting processing.

- Data integrity

  Ensure consistent point-in-time data analysis results, which can be set at regular intervals (daily, monthly, quarterly, or annually) or real-time data retrieval upon agreement

  All of the data in the showcase is stored in DB2 for z/OS. There are currently three principal subsystems for core banking, data warehousing, and messaging services.

- Scalability

  Support high volumes of data, which are used by the reporting services and accessed by a potentially high number of concurrent users

  The showcase contains 6,000,000 clients, 12,000,000 accounts, and more than 1,000,000,000 transaction records stored across multiple tables within the Core Banking subsystem. Over 400,000 customer, arrangement, and product records are stored in the data warehousing subsystem.

- Extensibility

  Support current reporting requirements and plan for future data, reporting, application, and user growth trends.

  This initial application is the first of many predefined business analytics solutions intended to be integrated within the showcase. Therefore, future business growth should be taken into consideration when assessing this workload.

- Security

  How will user authentication be managed using the different channels, for example, through branch, Internet, and mobile services?

  The showcase operates inside an internal authentication firewall that requires user validation before accessing the web services or showcase channels (mobiles, virtual world, or applications). Further authentication is required for access to back-office systems. Data security is protected by IBM RACF® and DB2 authorities.

- Availability:
  - Is the service continuously available 24x7?
  - While applying maintenance or service updates, what availability is provided?

> There are various availability requirements expected within the showcase to support reporting services and data access:
>
> - Continuous availability: While the Showcase users do not typically expect 24x7 coverage, the reporting services (web application access) and data retrieval are required during extended working hours.
> - High availability should provide a failover capability in the event of planned or unplanned outages.
> - Recovery Time and Recovery Point objectives are not applicable to the read-only service. However, in the event of a disaster, the system should be made available for critical business reporting within 30 minutes.

The remainder of this book describes the assessment of the blueprint against the business technical requirements outlined above.

4

zEnterprise hybrid workload placement approach

With the objective of producing a final business case for deploying a specific workload on zEnterprise, we developed a methodology. This methodology starts with any of the following events:

- Existence of certain problem points in a current workload or application
- A strategic initiative forcing a company to review its current implementation of workloads and applications
- A new application has been developed and the “best fit” runtime platform has to be chosen

The latter event applies to our situation. A new workload, Customer Profitability Analytics (CPA), has to be added to an existing IT infrastructure.

IBM has various methods, tools, and processes to make infrastructure decisions. With the methodology presented in this IBM Redbooks publication, we provide one approach that blends the best techniques of different approaches.

Our methodology encompasses eight stages:

- Stage 1: Initiate project
- Stage 2: Perform zEnterprise business value assessment
- Stage 3: Assess new workload
- Stage 4: List solution architecture alternatives
- Stage 5: Prioritize solution architecture alternatives
- Stage 6: Sizing and capacity planning
- Stage 7: Execute total cost of ownership (TCO) analysis
- Stage 8: Create business case

In the following sections, we discuss each stage briefly. More details about each stage can be found in the individual chapters in this book.

4.1 Overview of the approach

Figure 4-1 on page 33 shows the methodology at a high level. Later, we break down each stage into detailed activities, decisions, and input and output work products.

4.1.1 Decision points

There are three major decision points, but you may decide to use more decision points in between stages. Decision points are collaborative activities in which all stakeholders are involved. This approach uses the following major decision points:

- Possible zEnterprise value?

  This decision is made based on the outcome of Stage 2: Perform zEnterprise business value assessment. There are basically two possible decisions:

  - There is definitely no value in deploying the workloads or applications under investigation on the zEnterprise platform.
  - There is potential value in deploying the workloads or applications under investigation on the zEnterprise platform.

- Continue?

  Based on the solution architecture alternatives identified in Stage 4: List solution architecture alternatives and prioritized in Stage 5: Prioritize solution architecture alternatives, a decision is made to continue with the project. If the solution architecture alternatives are too intrusive or unrealistic, the decision might be to (temporarily) halt the project. In any case, an explicit continuation decision needs to be made to start with activities, such as sizing, capacity planning, and TCO analysis.

- Business case satisfactory?

  This decision might be made per solution architecture alternative or for all solution architecture alternatives combined. If none of the solution architecture alternatives results in a positive business case, the project needs to be halted and a decision needs to be made about whether to attempt the project.

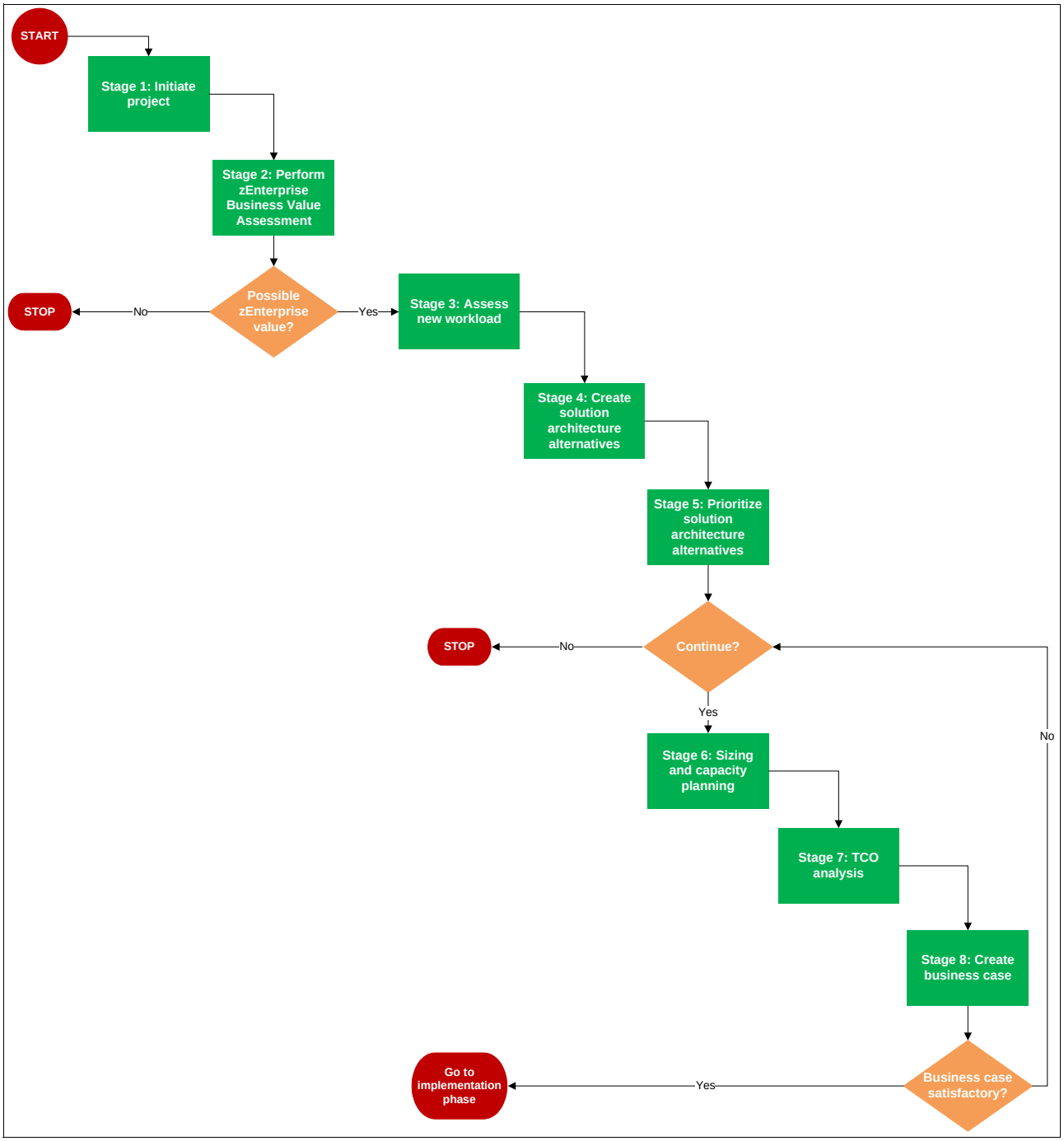


Figure 4-1 zEnterprise hybrid workload placement approach

4.2 Stage 1: Initiate project

As shown in Figure 4-2 on page 34, the first stage in the methodology after starting the project consists of activities and decisions to shape the project.

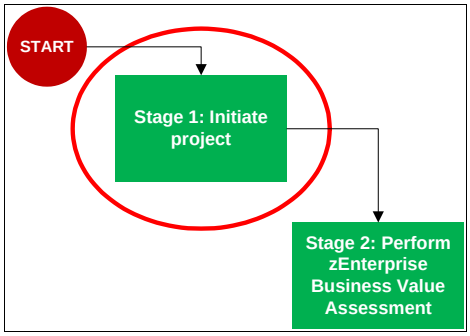


Figure 4-2 "Stage 1: Initiate project"

In this stage, it is important to define the following information:

- Expectations regarding the outcome of the project: It is important that everybody understand the background of the project and expectations of the stakeholders. For example, expectations could exist regarding documents and reports to be produced or to convince somebody of a certain solution.
- Scope of the project: The scope of a zEnterprise deployment project is typically set by specifying the applications or workloads to be included. A selection could be made based on various criteria, such as applications and workloads with similar problem points or applications and workloads belonging to a specific business unit or geography.
- Project organization: It is essential that all required stakeholders and subject matter experts (SMEs) are "onboard" and committed to perform their part during the project. Not everybody is required in all stages, though. Both IBM and the client will be represented in the project.
- Terms and conditions (T&Cs): It is important to agree on the T&Cs for the project. Is there any funding required and who provides the funding? Also, there needs to be agreement on interim milestones.
- Project plan: A project plan is needed, covering the different stages, their work products and milestones, and assumptions and dependencies.
- Tools needed: In some of the stages, we suggest certain tools. Some of these tools may be for IBM internal use only. The tools and who should run them should be defined up front.

This stage is discussed in more detail in Chapter 5, "Stage 1: Initiate project" on page 39.

4.3 Stage 2: Perform zEnterprise business value assessment

As shown in Figure 4-3 on page 35, the second stage in the methodology has the objective to determine whether the deployment or redeployment of a certain application or workload has any potential business value.

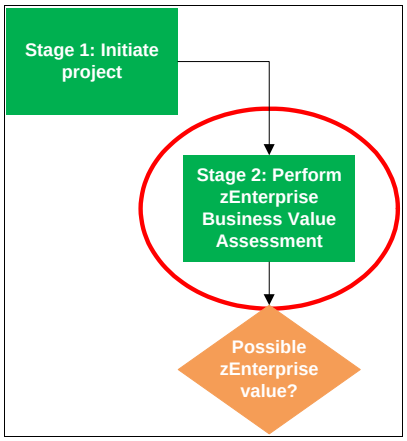


Figure 4-3 "Stage 2: Perform zEnterprise business value assessment"

Before spending valuable time and resources on detailed assessments and sizings, it makes sense to first perform a high-level study to find out whether there is potential value in deploying or redeploying an application or workload to zEnterprise. If the outcome of this stage is negative (no value can be expected), the project is either halted or restarted to investigate other platforms. If the outcome is positive, you can continue the project with the next stages.

This stage is discussed in detail in Chapter 6, "Stage 2: Perform zEnterprise business value assessment" on page 45.

4.4 Stage 3: Assess new workload

If the decision regarding potential zEnterprise business value based on the outcome of stage 2 is positive, stages 3 through 5 can be executed (Figure 4-4). These three stages are typically executed in one stream.

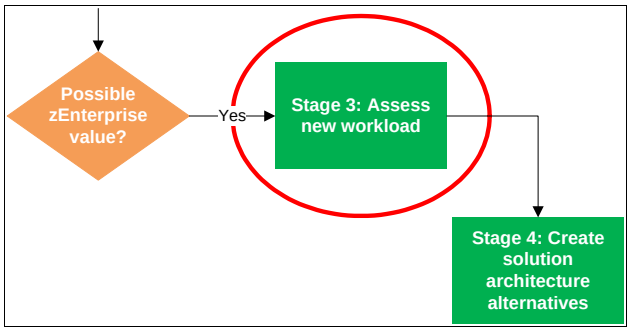


Figure 4-4 "Stage 3: Assess new workload"

4.4, "Stage 3: Assess new workload" on page 35 includes a variety of techniques, questionnaires, and workshops to obtain a good understanding of the applications and workloads in scope. This is necessary to be able to define candidate solution architectures in the next stage.

This stage is discussed in detail in Chapter 7, "Stage 3: Assess new workload" on page 55.

4.5 Stage 4: List solution architecture alternatives

This stage is a logical continuation of the previous stage and can only commence if a complete assessment has been made of all applications and workloads within scope. Based on the assessment results and knowledge of zEnterprise, several solution architecture alternatives are constructed, typically 2 - 4.

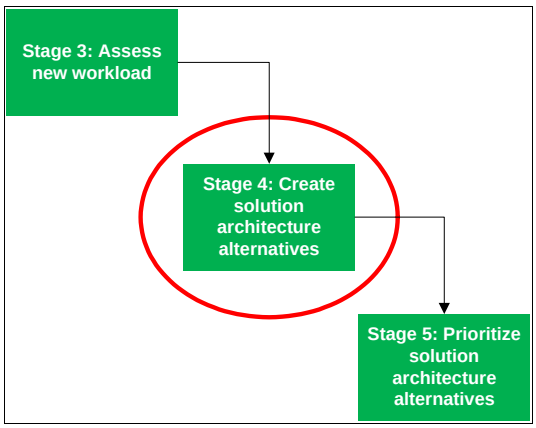


Figure 4-5 "Stage 4: List solution architecture alternatives"

This stage is discussed in detail in Chapter 8, "Stage 4: List solution architecture alternatives" on page 63.

4.6 Stage 5: Prioritize solution architecture alternatives

In this stage (Figure 4-6 on page 37), the solution architecture alternatives defined in the previous stage are prioritized. This prioritization is done based on qualitative criteria. The alternatives are matched with nonfunctional requirements (NFRs), service level agreements, and service level objectives.

Important: In this stage, no total cost of ownership (TCO) analysis has been performed yet. It is possible that the most favorable solution architecture from a qualitative point of view is not necessarily the most favorable solution architecture from a TCO point of view. In Stage 7: Execute TCO analysis, this evaluation is made with its advantages and disadvantages.

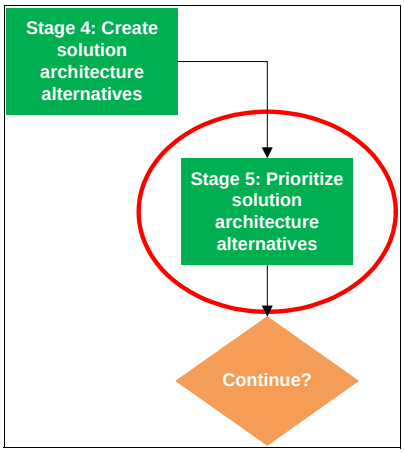


Figure 4-6 "Stage 5: Prioritize solution architecture alternatives"

This stage is discussed in detail in Chapter 9, "Stage 5: Prioritize solution architecture alternatives" on page 71.

4.7 Stage 6: Sizing and capacity planning

This stage can be executed for each solution architecture alternative or for all solution architectures all at once. The latter option allows you to make direct comparisons between the solution architecture alternatives with respect to sizing and capacity planning data. If this stage is executed per solution architecture alternative, the one that was ranked the highest in 4.6, "Stage 5: Prioritize solution architecture alternatives" on page 36 would be the first one to examine.

Figure 4-7 shows this stage in the context of its predecessor and successor.

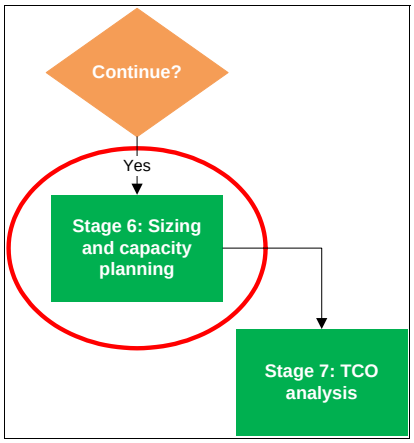


Figure 4-7 "Stage 6: Sizing and capacity planning"

This stage is discussed in detail in Chapter 10, "Stage 6: Sizing and capacity planning" on page 79.

4.8 Stage 7: Execute total cost of ownership (TCO) analysis

A proper TCO analysis (Figure 4-8) can only be performed if all metrics are available for the applications and workloads in scope. That's why this stage can only take place after the previous stage. Like the previous stage, this stage can either be executed per solution architecture alternative or on all solution architecture alternatives all at once. The latter will allow direct comparisons.

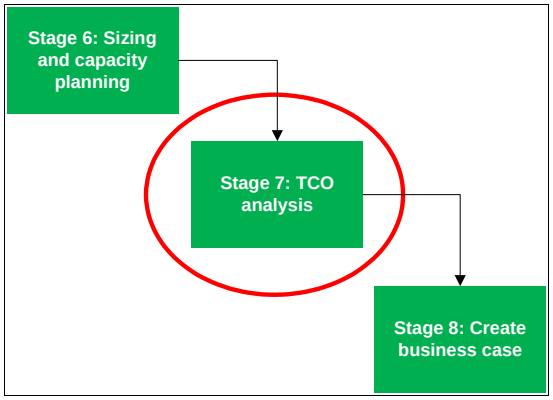


Figure 4-8 "Stage 7: Execute total cost of ownership (TCO) analysis"

At the end of this stage all information to make a business case should be available. This stage is discussed in detail in Chapter 11, "Stage 7: Execute total cost of ownership (TCO) analysis" on page 95.

4.9 Stage 8: Create business case

This stage (Figure 4-9) is the final step in the methodology. Based on TCO information and the earlier defined attributes of the solution architecture alternatives a business case is built. There are two options. Only one solution architecture alternative is justified and presented, or all solution architecture alternatives are evaluated, ranked, and presented in the context of both TCO and qualitative aspects.

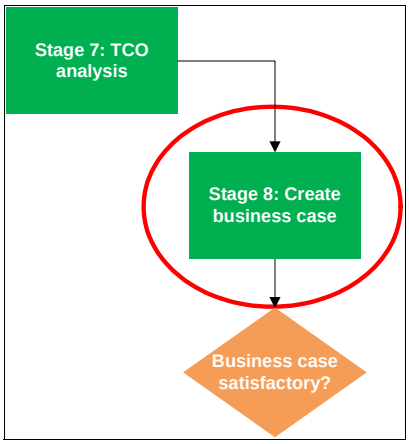


Figure 4-9 "Stage 8: Create business case"

This stage is discussed in detail in Chapter 12, "Stage 8: Create business case" on page 107.

5

Stage 1: Initiate project

The first stage, initiating the project, includes tasks such as formulating goals, objectives, and assumptions, as well as gathering requirements and project organization. This chapter covers the basics of project initiation, including tasks that experienced project managers instinctively perform.

5.1 Overview

This chapter describes a high-level view of the requirements to initiate a project to assess the placement of a workload on zEnterprise:

- Expectations regarding the outcome of the project
- Scope of the project
- Project organization
- Tools needed

5.1.1 Expectations regarding the outcome of the project

To initiate and effectively manage change in an organization, it is necessary to describe what is involved and obtain the sponsorship to make it happen. To achieve this level of sponsorship, it is necessary to be able to describe these areas:

- The business problems you are trying to address
- The value that the client can expect from this project
- When you expect the business value to flow into the organization

5.1.2 Scope of the project

Projects fail most frequently when the scope is not clearly understood or adequately defined. It is advisable to state the project scope and revisit it frequently to ensure that it continues to meet the business needs that the project sets out to address.

This book describes the use of a methodology that will assess the relevance of the *Customer Profitability Analytics* (CPA) business application workload for placement on the zEnterprise hybrid platform.

Later parts of this study will analyze the information and use it during the implementation phases of the project.

5.1.3 Project organization

The project is led by a project leader that reports to a senior executive sponsor. The project leader manages a team of subject matter experts (SMEs) from the business and operations departments. In support, the IBM account team can supply System z SMEs during the phases concerned with assessments and the analysis of workloads. Figure 5-1 on page 41 shows an example of a project organization. There are many possible variations, but essentially, the project needs executive stakeholders, architects, and SMEs involved, from both the client and IBM.

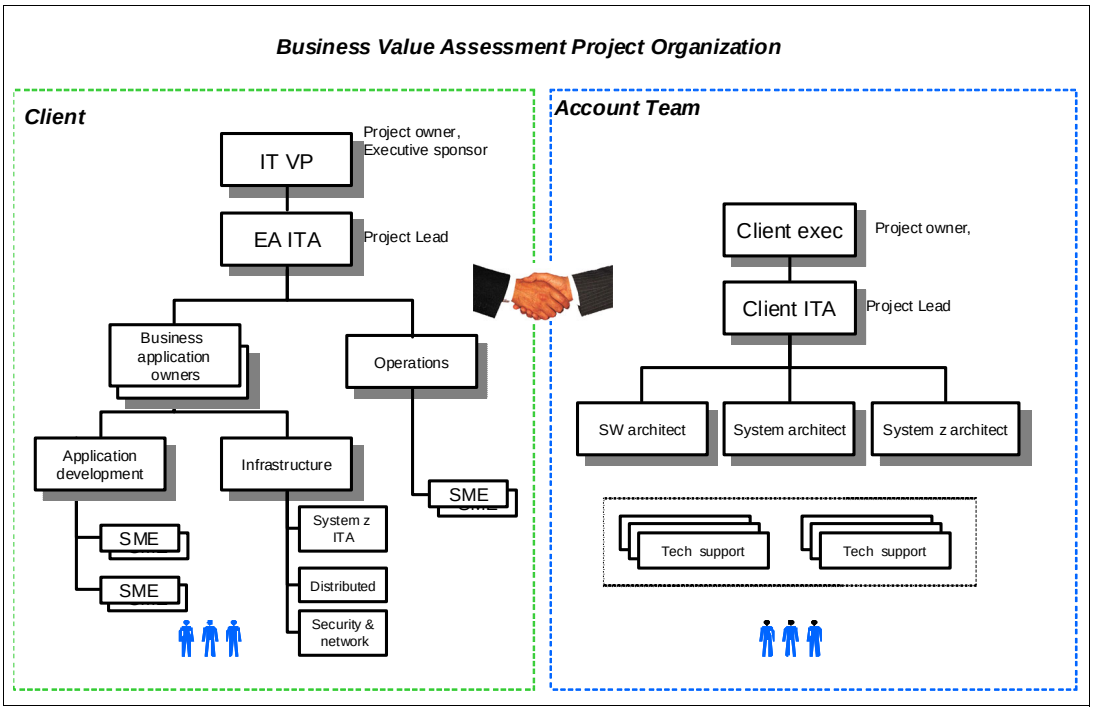


Figure 5-1 Project organization

5.1.4 Tools needed

Tools and methodologies are used to accelerate the rate of change delivered by a project. Several of these can be used during the workload assessment and placement activities described in this and subsequent books. This section describes the methodologies in general terms and explains when and how they should be applied. The following tools and methodologies have been used during the course of this project:

- A project management methodology

  During the assessment phase, the project management methodology should include the following milestones:

  - Initiate project:
    - Gain project executive sponsorship
    - Identify key stakeholders
    - Establish project organization
  - Perform zEnterprise business value assessment

    Confirm zEnterprise as a candidate
  - Assess workload
  - Create architectural solution alternatives
  - Prioritize architectural solution alternatives
  - Perform sizing and capacity planning
  - Execute total cost of ownership (TCO) analysis
  - Build business case
  - Select architectural solution

- zEnterprise business value assessment methodology determines if zEnterprise can add business value to the application workload.
- A fit for purpose methodology provides an assessment to verify that a platform delivers business value
- A sizing methodology determines the hardware requirements for defined workloads.
- A TCO methodology delivers a cost analysis for the solution as a whole.

5.2 Inputs

Workload placement in a hybrid environment is a challenge that is ultimately driven by a real business need. Business needs can be defined as requirements that will help a company achieve a strategic vision or goal.

The business need that defines and drives this project can be stated as: “CPA is a strategic business application that can be used to leverage business intelligence to better understand customers and their purchasing activity. Leveraging this business intelligence is perceived as providing a competitive advantage within the marketplace and will assist the company in gaining greater market share and achieving financial targets.”

After defining a business need, you must know to whom to articulate it within your organization. Identifying and exploring the need with the appropriate stakeholders allow this need to be more readily adopted. Knowing your organization’s structure at both a business and an IT level is paramount for success in initiating a project.

Having the appropriate relationship with a powerful sponsor who can resolve issues and problems is a prerequisite to success, especially where a hybrid platform infrastructure is concerned.

5.3 Outputs

The result of being able to articulate a business need to the appropriate sponsor with the backing of the technical community and business managers and being able to describe the changes that will solve this business need to provide a competitive advantage usually results in a formal project being initiated.

5.4 Tasks and decisions

The tasks that are involved in project initiation are shown in Figure 5-2 on page 43. We describe the following tasks:

- Gain executive sponsorship
- Determine key stakeholders and SMEs
- Establish the project organization

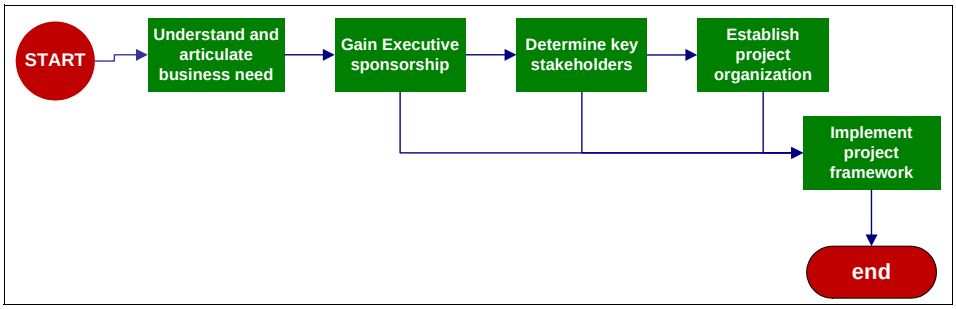


Figure 5-2 Initiate project process

5.4.1 Identify and confirm the executive sponsor at the client organization

The zEnterprise System brings a paradigm shift to the corporate IT environment. For a business value assessment to start and to lead to a successful conclusion, cross-departmental collaborative communication is essential.

To succeed, you have to identify the executive sponsor in the decision-making process capable of resolving inter-department issues. It is important to secure the approval at the start of the project to facilitate information sharing and client participation.

Ideal candidates are the CxO (for example, CTO or CIO) or the director or manager in charge of IT infrastructure.

5.4.2 Determine the key stakeholders and technical SMEs

With executive sponsorship in place, the key stakeholders and key SMEs can be engaged and the business value assessment project team created. Technical resources with business application knowledge and resources with infrastructure knowledge should be included in the project team. The following candidate roles and areas are suggested:

- Business application technical owners
- IT Architects:
  - Enterprise architecture
  - Application development and maintenance
  - Infrastructure (mainframe, distributed, and cross-platform)
  - Middleware architects
- Operations and IT infrastructure (data center and distributed)
- Security, network, and storage
- Operations managers, including capacity planners

5.4.3 Establish the project organization

The third activity is to put the complete project organization in place and apply the project management framework that has been adopted by the project.

When a rough project organization has been established, a kickoff meeting is held to define a formal set of goals and objectives for the project.

In the kickoff meeting, the following topics are discussed:

- Client's goals and objectives for the project (understand, share, and validate)
- Deliverables and outcomes
- Selection of business applications for analysis
- Additional SME or other team members that may be required
- Project time line

The project members will be finalized when the target business application has been selected.

5.5 Conclusions

With the formal initiation of an assessment project, resources can now look at the value of the platform through a selection process. The question to decide the platform on which to run the business application can now be addressed.

The CPA application uses data that resides on DB2 on the mainframe. As a consequence of its multi-platform characteristics, it was identified as a potential candidate for zEnterprise. Now, we start to look for the best place to host this business application.

The zEnterprise value assessment methodology allows us to determine whether zEnterprise can add value to the business for CPA as a platform. A generic approach to applying this methodology is described in the next chapter.

6

Stage 2: Perform zEnterprise business value assessment

The zEnterprise business value assessment (zBVA) is a methodology to determine whether the hybrid capabilities of the zEnterprise platform provide value in resolving identified problem points with a certain workload or group of workloads.

This stage is not mandatory and can be skipped if the choice to use zEnterprise for the workloads is already made. Or perhaps a workload is completely new, therefore, there are no known problem points[1], but just theoretical assumptions. Both these conditions were the case in our project, but we nevertheless include an explanation of how the zBVA can be used if the workload is not new and if the choice for zEnterprise has not yet been made.

For more information: You can also read this information, including more details, in *A Structured Approach to Assessing Heterogeneous Workload Deployment to the IBM zEnterprise System*, REDP-4709.

[1] Problem points are sometimes called “pain” points.

6.1 Objectives

The zEnterprise platform incorporates hybrid technology that spans mainframes, distributed systems, and speciality engines or appliances. This unique architecture requires a new approach to both business value and workload assessment. The zBVA provides a focused methodology to address these challenges by answering the question of which business value is brought by zEnterprise to an architectural challenge. Or, how do you decide what value zEnterprise can add to the placement of the new workload?

The goal of the zBVA is to develop consensus among all appropriate stakeholders about the business value of the zEnterprise System with relation to a new workload or target business application.

6.2 Process overview

The zBVA is applied early in the overall process to start thinking in terms of architectural end-to-end workload value, and to qualify workloads that look like a good fit for the zEnterprise platform (Figure 6-1).

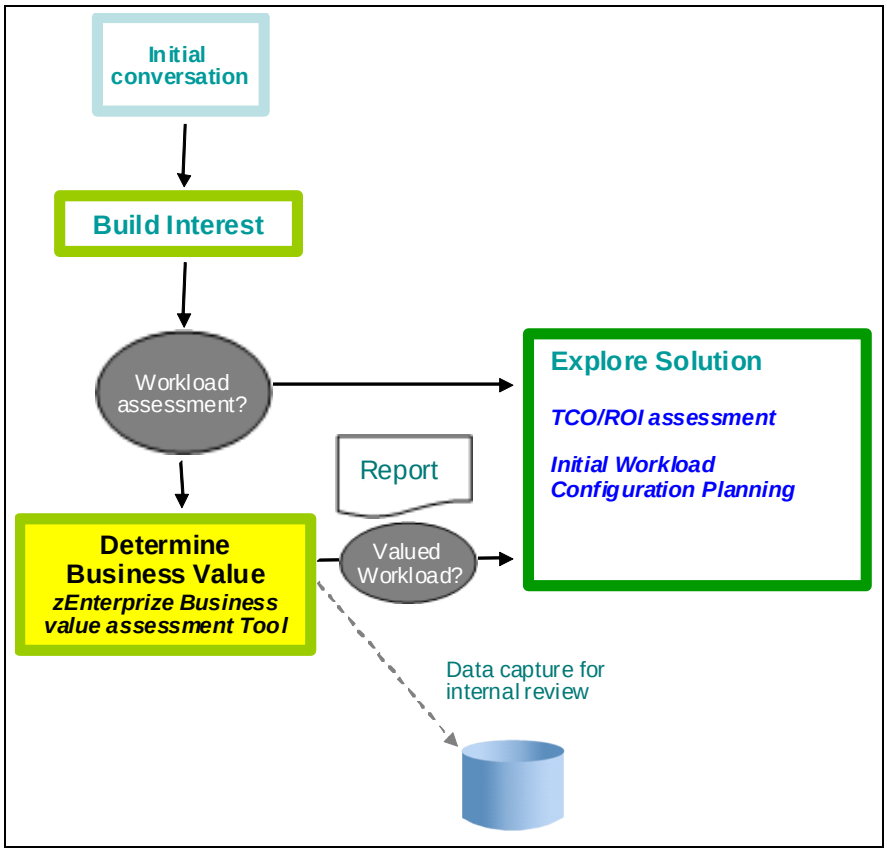


Figure 6-1 Positioning zBVA when considering workload placement

6.3 Tools and accelerators

zEnterprise business value assessment (zBVA) is a methodology that is delivered by questionnaire, review, and workshop approaches. This methodology gathers data used to calculate results in an assessment to use as input to the case for change.

The zBVA is applied to a selected business application and has several steps, which are shown in Figure 6-2.

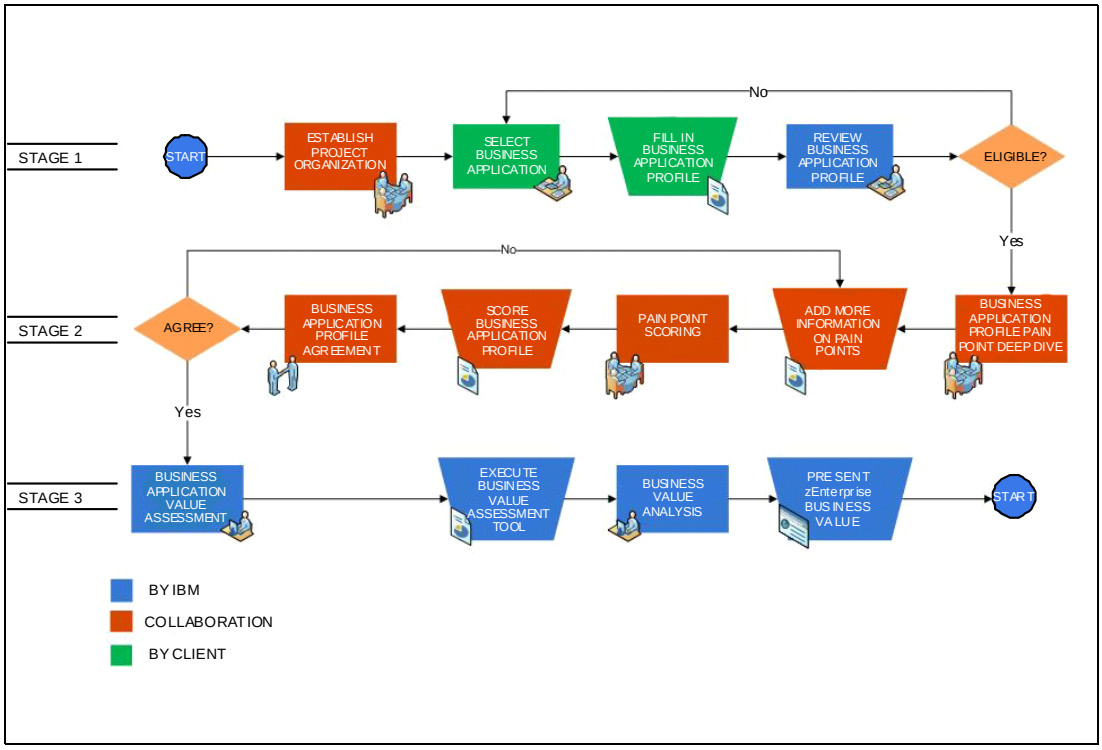


Figure 6-2 zBVA methodology

6.4 Resources needed

To perform a successful business value assessment, you need cross-departmental collaboration, including an executive sponsor, so that a consensus can be reached and supported by all those involved. Typically, a business value assessment involves people from the company or organization that owns the business application, along with IBM technical team members.

The following key stakeholders and technical SMEs are described in the methodology:

- Business application technical owners
- IT Architects:
  - Enterprise architect
  - Application development and maintenance
  - Infrastructure (mainframe, distributed, and cross-platform)
  - Middleware architects
- Operations and IT infrastructure (data center and distributed)

- Security, network, and storage
- Operations managers, including capacity planners

Technical managers and IT Architects will have different views on the correctness of a technical solution based on the priorities given to system characteristics. It is important that technical managers representing all potential areas of involvement are included. Their input is important in reaching a consensus, because the hybrid nature of workloads can span new platforms and challenge existing organizational structures.

A complete assessment must take into account all these views. The zBVA methodology provides a comprehensive approach to a placement of workloads on System z. Therefore, this assessment needs to take place early in the project before architectural decisions that might influence platform selection have been made.

6.5 zEnterprise business value process steps

The zBVA can be performed in six steps that can be grouped across three main stages:

- Stage 1: Establish the project organization, select the business application to be examined, and create a business application profile of the application.
- Stage 2: Describe in detail and score the problem points in the business application.
- Stage 3: Perform the actual value assessment, which analyzes the extent to which zEnterprise resolves the problem points that are defined in stage 2.

6.5.1 Stage 1

Stage 1 initiates the zBVA process. It creates the governance framework and gathers initial data that will be used to define the scope of the assessment. We briefly explain each step of stage 1.

Step 1: Establishing project organization

This activity must have an executive sponsor to ensure success. This might include people from multiple departments and organizations. For recommendations about establishing a project organization for this kind of project, see 5.1.3, “Project organization” on page 40.

Step 2: Selecting and documenting the candidate business application

The zBVA takes time to complete and can be run iteratively as new workloads are investigated. This approach involves resources that might be expensive or time-constrained. Therefore, consider these factors:

- A good workload to assess is one that has perceived problem points (sometimes called pain points). A better workload is one that is also representative of other business applications.
- Sometimes where applications already reside on a heterogeneous infrastructure, they can be selected based on potential engineering improvements.
- The business application is selected through a filtering approach. The criteria for filtering suggested by the zBVA methodology are shown in Figure 6-3 on page 49.

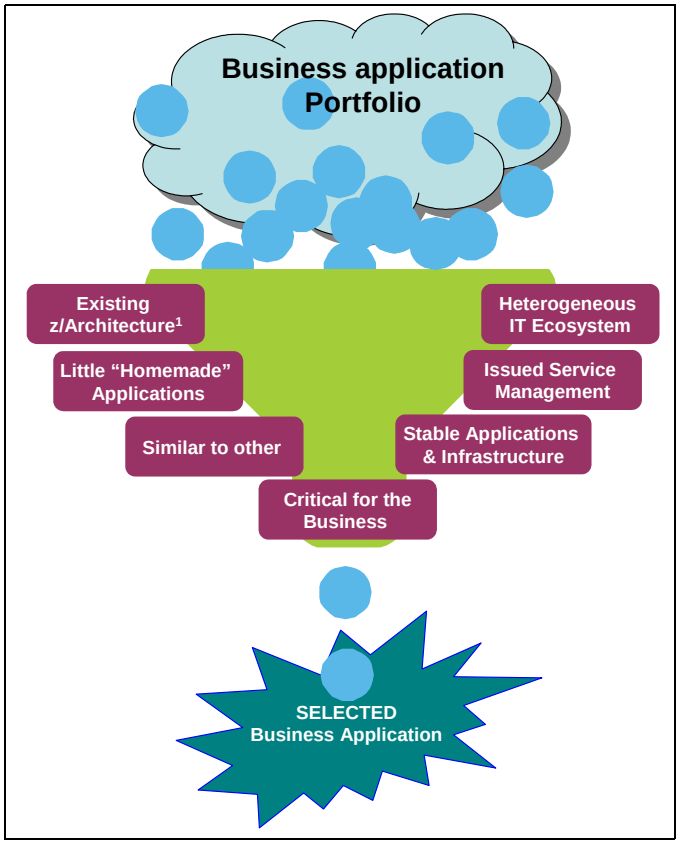


Figure 6-3 Business application candidate filtering

After an application or applications are selected, supporting documentation is collected to build a better understanding for the next steps in the process.

Completing and reviewing the business application profile

This step is an information-gathering exercise and describes the application profile in detail, with a strong focus on problem points. The minimum requirement is a set of infrastructure architectural documentation that can include the following information:

- Overview architecture
- Component architecture
- Operational architecture
- Security architecture

This documentation should have been collected in the previous step.

zBVA provides a methodology for identifying problem points in a business application. It does this during the analysis of business applications by following this process:

- Dividing a business application profile into domains that describe aspects of an IT organization (Figure 6-4 on page 50):
  - Security
  - Networks
  - Facilities
  - Performance
  - Operations

- Maintenance
- Support
- Business Intelligence (BI)
- Service-oriented architecture (SOA)

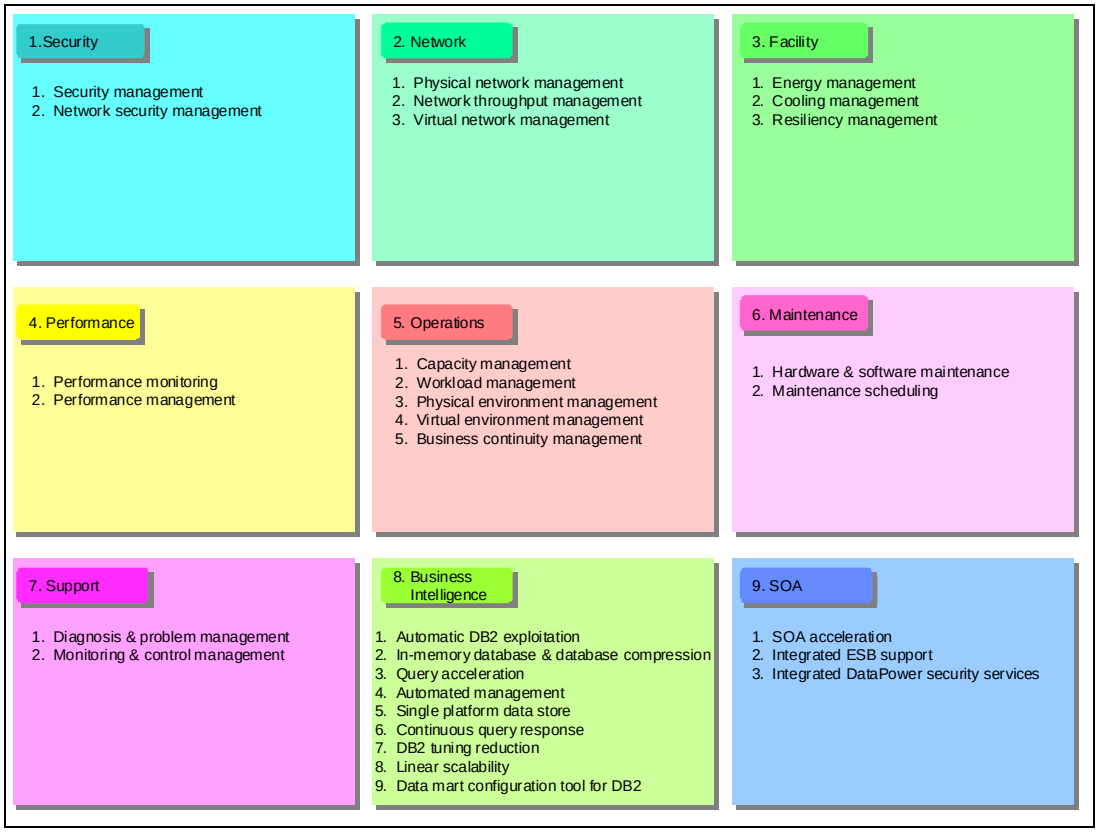


Figure 6-4 Services and domains in the business application profile

- Considering the functions performed in each domain as a service. For example, the performance domain may contain the following services:
  - Performance monitoring
  - Performance management
- Identifying the real problem points associated with each service in relation to each service. Performance management might have the following problem points:
  - Multiple tools required to manage all components of the application
  - Multiple skills required to manage across platforms
  - No holistic performance view of the entire end-to-end process

The benefits of using the zBVA methodology are highlighted when we consider specific scenarios during the analysis of the subject business application by relating this to business objectives. By using this approach, we can consider the following scenarios:

- Scalability
- Cost
- IT management
- Network with focus on security
- Business resilience

Within the zBVA, this is achieved by applying a weighting to a prescriptive set of questions within the assessment tool, which skews their importance. Other problem point scenarios can be designed and created by applying a weighting to a new set of questions. By using this approach, we can look at a business application and ask whether zEnterprise can add value to a specific area. For example, can *Customer Profitability Analytics* (CPA) scale to consume 1 terabyte (TB) of data daily and 2,500 concurrent users?

After we identify all the problem points, we are ready for the next step.

zBVA stage 1 applied to CPA

The CPA workload seemed to be a good candidate for a zBVA because of the following criteria:

- The application already uses a component that is currently resident on z.
- The application is supported by back-end and middleware components, IT systems, and IT components that run across multiple platforms and that are heterogeneous in nature.
- The application provides critical business information.
- The expected workload growth is very large.
- The workload will be both batch and online with peaks of activity.
- The application data is stored as operational data, in a data warehouse, and as data marts.
- The management of the business workload currently spans multiple management silos and is perceived as complex.

Documentation of the CPA workload was collected by the team. An overview of the workload is documented in Chapter 3, “Overview of the banking analytical workload” on page 19.

We did not create the business application profile based on problem points because CPA was a new workload without a current implementation.

6.5.2 Stage 2

The identification of eligible business applications and the documentation of the business application profile in stage 1 are used as starting points in stage 2 of the zBVA. This stage is concerned with the further exploration and analysis of pain points associated with the business application. If there are no problem points in the current business application, there is no compelling reason to act or change the existing architecture. Where a new workload is being assessed, problem points may be theoretical based on how nonfunctional requirements (NFRs) will be resolved by platform choices.

Problem points can be defined as inhibitors to future growth or factors that will have a negative influence on agreed to service-level objectives (SLOs) and SLAs. They are not necessarily showstoppers now, but may be in the future.

Stage 2 consists of the following activity.

Step 4: Performing the in-depth analysis and scoring of problem points

In this step, each problem point described in the previous step is analyzed and a ranking is produced that assigns an importance to it from the perspective of that service.

The following scale is used. You may deviate from this ranking and establish your own way of ranking.

Very high	Very high risk for meeting the business objectives. The impact of this situation causes direct economic losses to the company. These effects are uncontrolled and difficult to contain.
High	High risk for meeting the business objectives. The impact of this situation occurring causes indirect economic losses to the company. The effects are uncontrolled and difficult to contain.
Medium	Medium risk for meeting the business objectives. The impact of this situation can cause some disturbance. It is likely to get worse and become high or very high risk in the future. The effects are uncontrolled, but can be contained.
Low	Low risk for meeting the business objectives. The impact of this situation is minimal. The effects are controlled and can be contained.
Very low	Very low risk for meeting the business objectives. No impact occurs. The effects are controlled in all stages.

After all problem points are scored, we are ready to review these problem points in relation to the business application and the whole enterprise.

Applying zBVA stage 2 to CPA

CPA is a new workload, so the problem points that we considered are theoretical. Therefore, we could not complete this stage as part of our project.

6.5.3 Stage 3

With the completion of the analysis of the business application profile, the question about the value that zEnterprise brings to the architecture can now be addressed. This is achieved by using the zBVA tool that takes the profile information and generates a report. The team will then analyze the information and findings for the selected business application. Then, the team produces a report, which shows the value of zEnterprise for this business application, to present to the stakeholders.

The steps in stage 3 are now described.

Step 5: Assessing the business application value

zBVA requires the use of the zBVA tool and entering data gathered from the analysis of the business application profile into it. The use of the tool is restricted to IBM employees and qualified IBM Business Partners.

Information gathered in stage 2 is used here as reference data for input to the tool. The assessment asks to rate system characteristics as they relate to the business application under analysis. The problem point analysis draws attention to those system characteristics that affect the current or future state of the business application and introduces the data skew to reflect their importance in the decision-making process.

The subjective nature of many of these data points is removed as far as possible by a consistent rating scale but is open to change where required.

Interviews: The number of people interviewed also affects the analysis. Interviewing too few people or too many people may give average results that do not identify real concerns so care is needed in identifying the correct number of people.

Step 6: Analyzing the business value

The zBVA report is delivered as data points represented by a series of star chart graphics in the shape of a diamond divided into four coordinates. Each coordinate represents a business value area. These four business value areas identify the strengths and values of zEnterprise:

New capabilities The ability to enable new business offerings and better align IT to business processes and how the business position is improved in the near future

Quality of service The ability to improve some aspects of the NFRs of the existing and new business services

Time to value The ability to decrease the time needed for the IT business services to begin providing a positive impact on revenue and profit

Increased efficiency The ability to perform technological, organizational, and process improvements that lead to a cost reduction (such as a reduction of the labor force) or to a productivity improvement (such as the capability to perform more work with the same resources)

Applying zBVA stage 3 to CPA

The introduction of CPA into the Smarter Banking showcase represents a new workload. Problem points do not currently exist, so an extrapolation of the technical requirements was used to gather data as input to the zBVA.

When considering the result of the CPA workload analysis, the business value area in Figure 6-5 shows a significant area of green in the zEnterprise assessment, indicating that business value could be gained by incorporating zEnterprise in the CPA application workload architecture.

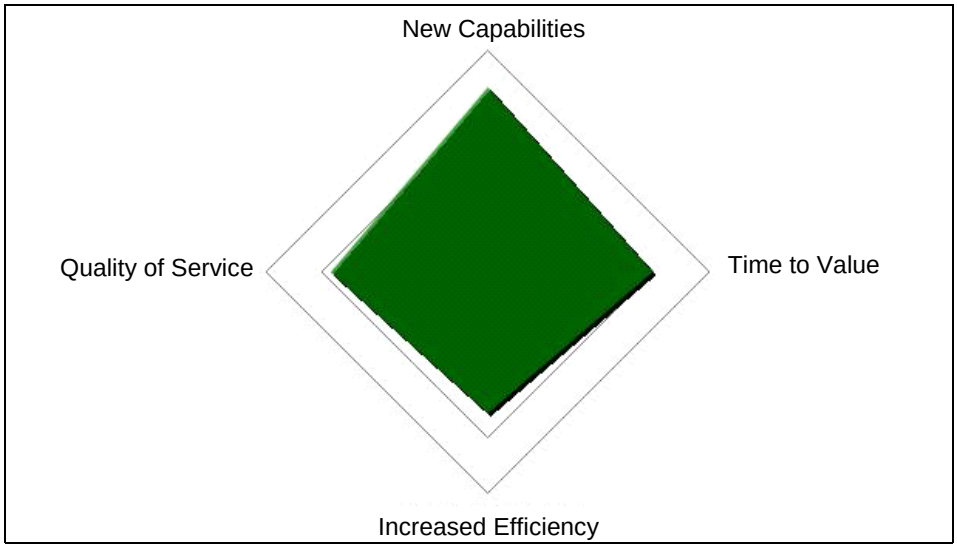


Figure 6-5 zEnterprise today: Comprehensive score for business value areas for CPA workload

6.6 Conclusions

The star charts produced in the various zBVA reports are a visual representation of the potential benefits for zEnterprise, recognized at an early stage in the assessment phase of a workload placement project.

Figure 6-5 on page 53 indicated that the architecture used to implement the CPA business application workload could derive real business benefits from zEnterprise.

The successful completion of this stage of an assessment project should see an agreement reached that further workload placement evaluation work will include zEnterprise in the candidate architectures of the solution.

7

Stage 3: Assess new workload

After we define the functions of *Customer Profitability Analytics* (CPA), we need to define on which platforms to deploy it. We use a three-stage process. First, we assess the different components of our workload following its nonfunctional requirements (NFRs). Then, we determine how the different available platforms comply with each NFR. Finally, from the list of available options, we select the best choice for each component.

Chapters Chapter 7, “Stage 3: Assess new workload” on page 55 to Chapter 9, “Stage 5: Prioritize solution architecture alternatives” on page 71 show the procedure we followed to select the best platform option for CPA. In this chapter, we determine which components of CPA need to be assessed and we define their NFRs.

7.1 Objectives

The global objective for this and the following two chapters is to select the best platform for our application. The platform selected will determine the quality of service (QoS) that can be expected from the application and the accepted way of measuring QoS is through the use of NFRs.

An NFR is not strictly related to business logic. Examples include security, subscription and profile provisioning, and usage data.

The objective of this chapter is to determine the NFRs for each CPA component as a first step in the process of selecting the platform for CPA. We will present the list of commonly accepted NFRs and then we will determine which NFRs apply to each component. The resulting matrix will be used in following chapters to select a platform, based on best fit, for each component.

Other considerations for selecting a platform include possible constraints, unsupported environments, and software requirements.

Note: We are assessing a new implementation of an existing application but, due to the limited time available, we are not considering the possibility of a re-engineering change. In different circumstances, especially when the assessment is performed on already-running applications, a re-engineering change might be considered because it might prove beneficial and lead to better QoS.

7.2 Process overview

Figure 7-1 on page 57 shows all CPA components along with core applications. New components are shadowed in the figure. We will only assess new components.

In this chapter, we first define the NFRs. Then, we determine which NFRs apply to each CPA component. In chapters Chapter 8, “Stage 4: List solution architecture alternatives” on page 63 and Chapter 9, “Stage 5: Prioritize solution architecture alternatives” on page 71, we build the list of possible architectures and then determine the one that best fits our requirements.

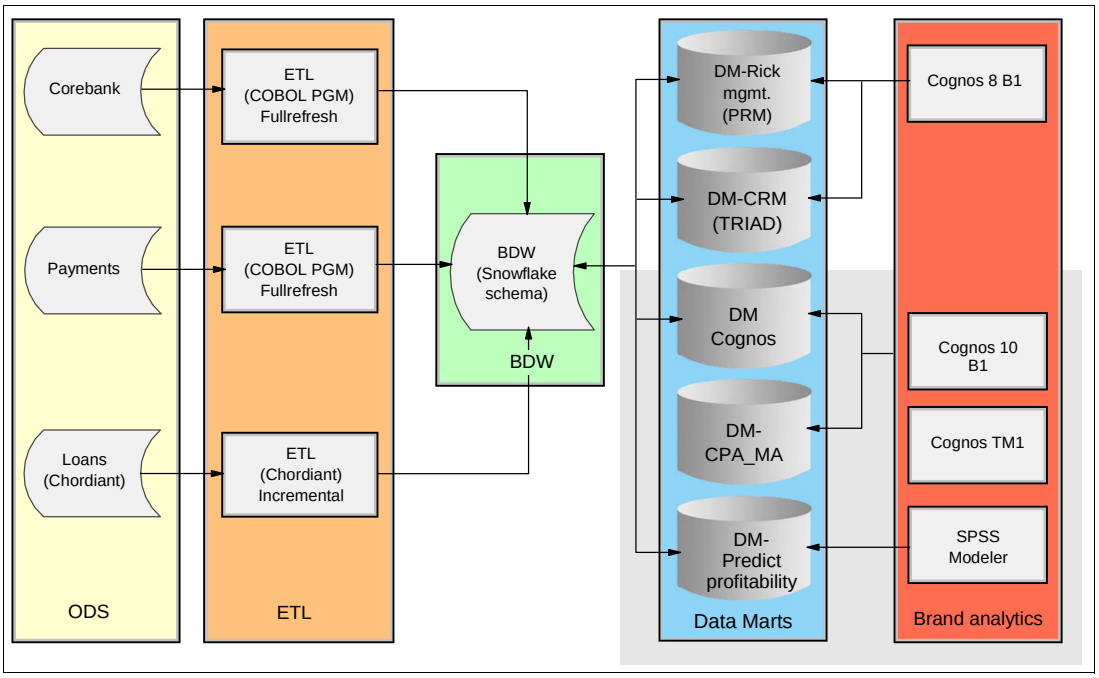


Figure 7-1 CPA components overview

7.3 Tools and accelerators

The tool we used in the assessment is the IBM internal tool used for mapping application requirements across platforms. The first activity in the tool is to build the list of applicable NFRs for one workload. We need to input one set of NFRs for each workload.

The main input to this stage is the list of CPA components. The new components in CPA are in two categories:

- Data marts

  Three data marts will be created:

  - Cognos BI Reports
  - Predict profitability
  - CPA-Metrics Studio (CPA-MS)

  There is one database that we left out of the assessment: the metrics store. Because IBM Cognos Metrics Server does not support DB2 for z/OS, this metrics store will reside in DB2 9.5 Enterprise Server Edition.

- Software components

  The following new components are identified:

  - Cognos 10 BI
    At the time of writing this book, some components of Cognos 10 BI were only available on specific platforms. For this reason, IBM Cognos Metrics Studio was not included in the assessment and was installed on an AIX platform.
  - Cognos TM1

    Uses Cognos Data Access Pack for access to data.

- Cognos 10 Framework Manager
- SPSS Modeler

  Is composed of SPSS Modeler Server and SPSS Modeler Client.

There are prerequisites for the software components:

- WebSphere Application Server Network Deployment (for Cognos 10 BI)
- IBM HTTP Server
- IBM DB2 Connect™ for Microsoft Windows (for SPSS Modeler Client)

These prerequisites will be placed along with the component that needs them. They are not a functional part of the application so they will not be part of the best fit exercise.

An extract, load, and transform (ETL) process that is capable of loading data from an existing data warehouse is needed. In our first stage, we decided to use DB2 utilities to unload data from the current data warehouse. The unloaded data will be moved to data mart environments and then loaded into the different data marts by using data mart utilities.

As a condition for the assessment, we decided to place all data marts on the same platform.

SPSS Modeler Client and Cognos 10 Framework Manager must be installed on a computer running the Microsoft Windows operating system. We decided that they will be deployed on zEnterprise BladeCenter Extension (zBX)-installed x86 blades running Microsoft Windows. They were left out of the assessment because we established in previous chapters that we will use the value provided by zBX and zEnterprise Unified Resource Manager.

This final list of components will be assessed:

- New data marts
- Cognos 10 BI
- Cognos TM1 (including Cognos Data Access Pack)
- SPSS Modeler Server

7.4 Resources needed

Assessing NFRs needs an excellent knowledge of the application to be assessed. The main resource needed is a team of users of the application. In our exercise, the team that authored this book acted as the users, defining a list of five NFRs for each component and prioritizing them.

7.5 Detailed steps

We used the process flow shown in Figure 7-2 on page 59.

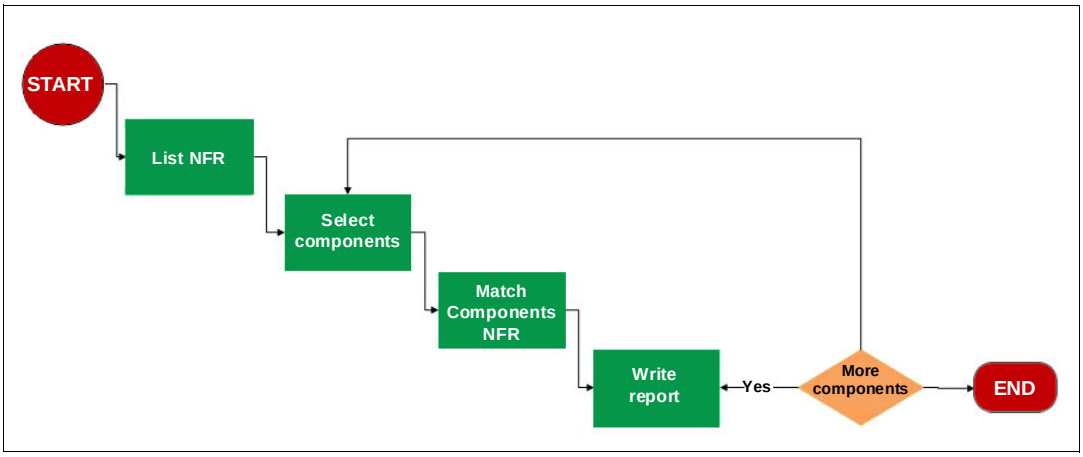


Figure 7-2 Steps for workload assessment

The following steps make up the process flow:

List NFRs We built a comprehensive list of NFRs and defined each of them.

Select component and match component to NFRs
We compared each component with the list of NFRs and decided which five NFRs are most important for the component.

Write report We documented the five NFRs in order of importance.

7.5.1 NFRs

The following list contains the NFRs used for the assessment:

Response time The application will be able to respond within its allotted time.

Disaster recovery In case of a disaster, the application will go back to operations within a maximum time frame and with minimal or no data loss.

Scalability Ability to handle growing amounts of work in a graceful manner or its ability to be enlarged to accommodate that growth.

Security Process that ensures the confidentiality, integrity, and availability of an enterprise's assets, information, data, and IT services.

Manageability Degree of ease to change, configure, and operate the platform through its lifecycle.

Skills Current state of an enterprise's personnel skills or ability to obtain skills to define, deploy, manage, and support a particular platform.

Data center constraints
Limitations on the physical infrastructure in the data center relating to space and energy constraints that influence or dictate IT solutions. Standards include the documentation of guidelines, reference architectures, and mandatory requirements that must be followed when deploying IT solutions.

Strategy Documentation (or verbalization) of an enterprise's principles and overall objectives.

Governance Structure, oversight, and management processes that specify a decision and accountability framework to encourage and enforce desirable behaviors in the use of IT.

Standards Established norms. Degree of respect to the norms.

Workload fit Process of evaluating a certain workload and mapping the demands against the proposed server deployment and underlying architectures.

Data proximity Relative distance of data placement to the application based on latency.

Efficiency Ability to best utilize IT resources to deliver services with the minimum consumption of time, money, people, or other resources.

Data replication Replication needs and options based on workload requirements.

High availability Attribute of a system to provide service during defined periods, at acceptable or agreed-upon levels, and that masks unplanned outages from users. It employs fault tolerance, automated failure detection, recovery, bypass reconfiguration, testing, and problem and change management.

Continuous availability Attribute of a system to deliver nondisruptive service to the user seven days a week, 24 hours a day (there are no planned or unplanned outages). It includes the ability to recover from a site disaster by switching computing to a second site.

7.5.2 Component characterization

For each component in 7.4, “Resources needed” on page 58, we now select the five most relevant NFRs from the list described in 7.5.1, “NFRs” on page 59.

SPSS Modeler Server

SPSS Modeler Server has the following unique characteristics:

- Optimized performance by leveraging proven parallel processing technology
- Increased flexibility with IBM SPSS Modeler Server Batch Mode to automate production tasks
- High efficiency with *Incremental Model Refresh* technology
- Additional in-database algorithms for IBM InfoSphere® Warehouse: Logistic Regression, Naive Bayes, Time Series, and Radial Basis Function (RBF)
- Extensive in-database modeling capabilities making the most of the existing IT infrastructure
- Use of unlimited number of CPUs and scoring capability
- Secure Sockets Layer (SSL) encryption and password protection ensuring that all sensitive data is fully protected

We want to insure the most important properties with SPSS Modeler Server:

- Fast and secure access to data marts
- Support for a growing number of users
- High availability
- Deployment in a server that is managed according to global policies

These properties led us to select the following NFRs from our list: security, data proximity, scalability, high availability, and manageability.

New data marts

Following the same methodology described for SPSS Modeler Server, we selected the following NFRs for data marts:

- Security: Data in our data marts is classified as confidential. Unauthorized access must be stopped.
- Response time: BI applications need to execute complex queries. Good response time from its database is critical.
- High availability: Data mart availability will contribute to the availability of CPA according to the SLA. Continuous availability is not needed.
- Scalability: Data marts must accommodate a growing number of users.
- Data proximity: Data marts will be populated with data that comes from the current data warehouse.

Cognos 10 BI

Following the same methodology described for SPSS Modeler Server, we selected the following NFRs for Cognos 10 BI:

- Security: Data in our data marts is classified as confidential. Unauthorized access must be stopped.
- Data proximity: Cognos 10 BI will not only read data from its data mart, but it will also populate data in it. Its proximity to the data mart will contribute to a better response time.
- High availability: CPA must respect SLA-defined availability. Continuous availability is not needed.
- Scalability: CPA must accommodate a growing number of users.
- Response time: Response time is very critical and will benefit from close proximity to the data.

Cognos TM1 and Cognos Data Access Pack

Following the same methodology described for SPSS Modeler Server, we selected the following NFRs for Cognos TM1:

- Security: Data in our data marts is classified as confidential. Unauthorized access must be stopped.
- High availability: CPA must respect the availability as defined in the SLA. Continuous availability is not needed.
- Scalability: CPA must accommodate a growing number of users.
- Response time: Response time is critical. Cognos TM1 loads its data in memory at start-up. This removes the need for data proximity.
- Manageability: Cognos TM1 will obtain its data from its directly attached devices but also from SPSS Modeler Server. It will benefit from a centrally managed policy along with SPSS Modeler Server.

7.5.3 Conclusion

Table 7-1 on page 62 summarizes the selected NFRs for each component.

Table 7-1 NFRs selected for each component

COMPONENTS				
	Cognos 10 BI	**Cognos TM1 and Cognos Data Access Pack**	**SPSS Modeler Server**	**Data marts**
Scalability	2	3	3	4
Response time	4	1	1	2
High availability	3	5		3
Continuous availability				
Disaster recovery				
Security	1	2	2	1
Skills				
Workload fit				
Data center constraints				
Manageability		4	4	
Standards				
Strategy				
Data proximity	5		5	5
Data replication				
Governance				
Efficiency				

8

Stage 4: List solution architecture alternatives

In this book, we introduced a new application, *Customer Profitability Analytics* (CPA). Its requirements, both functional and non-functional, were assessed. In view of all these needs, different architectural alternatives are available. This chapter discusses the available options.

8.1 Objectives

The objective of this chapter is to document system architecture alternatives that allow us to deploy our application.

In Chapter 7, “Stage 3: Assess new workload” on page 55, we selected the most important nonfunctional requirements (NFRs) for each component. This chapter qualifies the available platform options for each component in our application in view of these NFRs.

8.2 Process overview

We first determine the list of candidate platform options for our application. Then, we evaluate, as future users of the CPA application, how each platform performs for each of the selected NFRs. We repeat this process for each component of our application.

8.3 Tools and accelerators

The tool we used in the assessment is the IBM internal tool used for mapping application requirements across platforms. The second activity in the tool is to assess each platform following the list of applicable NFRs for one workload. This tool also allows us to assign a weight to each NFR for each component.

8.4 Resources needed

As in Chapter 7, “Stage 3: Assess new workload” on page 55, the assessment has to be performed by a team with an excellent knowledge of the application and its needs. The team writing this book acted as final users of the application and performed the evaluation.

In order to perform the assessment, we determined the list of platform options to be considered. We selected the following platforms as possible options:

- Stand-alone IBM POWER® servers
- zEnterprise BladeCenter Extension (zBX)-connected IBM POWER7 blades
- zBX-connected x86 blades running Linux
- zBX-connected x86 blades running Microsoft Windows
- Linux on System z
- z/OS

Not all components are available for all platforms, limiting the options. In each case, of course, we limited our assessment to supported platforms.

8.5 Detailed steps

We used the process flow that is shown in Figure 8-1 on page 65.

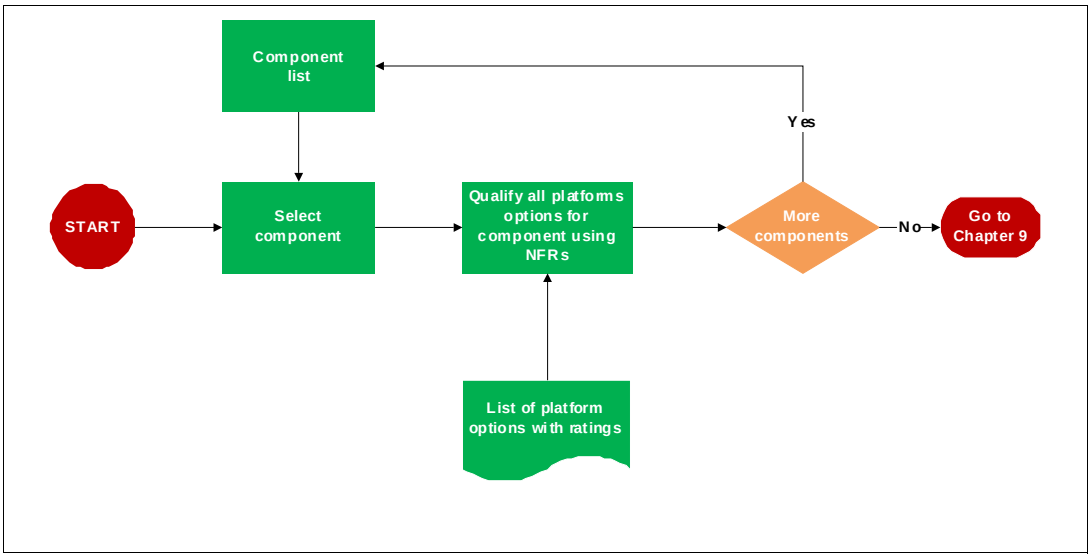


Figure 8-1 Detailed process flow for platform assessment

The process flow consists of these steps:

Select component We used the list of components with NFRs that we built in previous chapter.

Qualify all platform options for component using NFRs
For the component selected, we built the list of candidate platforms.

Rate platform options
Each platform received a rate regarding its level of compliance for each NFR for the component selected.

8.6 Rating platform options based on NFRs

The following findings are for each platform option.

8.6.1 Data marts

We decided to keep our data marts in DB2 databases on z/OS for the following reasons:

- DB2 offers performance, scalability, and reliability.
- DB2 is available on multiple platforms.
- Core banking applications already use DB2.
- The team in charge of CPA is already highly skilled.

There is one database that we left out of the assessment: the metrics store. Because IBM Cognos Metrics Server does not support DB2 for z/OS, the metrics store will reside in DB2 9.5 Enterprise Server Edition.

For our data marts, we considered the following environments:

- DB2 for Linux, UNIX, and Windows (LUW) running on a stand-alone POWER server
- DB2 for LUW on a zBX-connected POWER7 blade
- DB2 for z/OS
- DB2 for z/OS with IBM DB2 Analytics Accelerator

We selected the following NFRs in 7.5.2, “Component characterization” on page 60 for data marts:

- Security: Data in our data marts is classified as confidential. Unauthorized access must be stopped.
- Response time: Business Intelligence (BI) applications need to execute complex queries. Good response time from its database is critical.
- High availability: Data marts’ availability will contribute to availability of CPA according to the service level agreement (SLA). Continuous availability is not needed.
- Scalability: Data marts must accommodate a growing number of users.
- Data proximity: Data marts will be populated with data coming from our current data warehouse.

We qualified the different platform options as shown Figure 8-2. “10” means the highest possible rating and “1” means the lowest possible rating.

Non-Functional Requirements	DB2 Power	DB2 Blades	DB2 z/OS	DB2 IDAA
Security	6	7	10	10
Response Time	6	3	5	10
High Availability	7	6	10	10
Scalability	8	6	10	10
Data Proximity	6	7	10	10

Figure 8-2 NFRs for CPA data marts

Note: Values used here for response time and scalability represent our beliefs based on the circumstances of our workload. Any values used for response time, scalability, or other performance measures should be obtained in a benchmark with real applications, data, and environment.

Highlights from the assessment

We identified the following highlights from our assessment:

- The current data warehouse is located in DB2 for z/OS. It is configured in data sharing with two partitions.
- Configuring our data marts along with the current data warehouse will simplify the data population and administration for data marts.
- We considered the volume of data in direct access storage and its effect on each platform.
- Analysis of our queries show that they would benefit from the IBM DB2 Analytics Accelerator.

8.6.2 Cognos 10 BI

For Cognos 10 BI, we considered the following platform options:

- Stand-alone POWER servers
- zBX-connected POWER7 blades
- zBX-connected x86 blades running Linux
- Linux on System z

We selected these NFRs in 7.5.2, “Component characterization” on page 60 for Cognos 10 BI:

- Security: Data in our data marts is classified as confidential. Unauthorized access must be stopped.
- Data proximity: Cognos 10 BI will not only read data from its data mart; it will also populate data in it. Its proximity to the data mart will contribute to a better response time.
- High availability: CPA must respect SLA-defined availability. Continuous availability is not needed.
- Scalability: CPA must accommodate a growing number of users.
- Response time: Quick response time is critical and will also benefit from a proximity to the data.

For Cognos 10 BI, we qualified the different platforms options as shown Figure 8-3.A rating of “10” means the highest possible rating and “1” means the lowest possible rating.

Non-Functional Requirements	Power-SA	Power-zBX	Linux-zBX	Linux on z
Security	6	8	8	9
Data Proximity	4	5	5	8
High Availability	6	5	5	6
Scalability	8	7	6	8
Response Time	8	8	7	7

Figure 8-3 NFRs for CPA Cognos 10 BI

Note: Values used here for response time and scalability represent our beliefs based on the circumstances of our workload. Any values used for response time, scalability, or other performance measures should be obtained in a benchmark with real applications, data, and environment.

Highlights from the assessment

We did not find that continuous availability was needed.

8.6.3 Cognos TM1

For Cognos TM1, we considered the following platform options:

- Stand-alone POWER servers
- zBX-connected POWER7 blades
- zBX-connected x86 blades running Linux
- Microsoft Windows

We selected following NFRs in 7.5.2, “Component characterization” on page 60 for Cognos TM1:

- Security: Data in our data marts is classified as confidential. Unauthorized access must be stopped.
- High availability: CPA must respect SLA-defined availability. Continuous availability is not needed.
- Scalability: CPA must accommodate a growing number of users.

- Response time: Keeping it under the limits accorded in SLA is critical.
- Manageability: Cognos TM1 will obtain its data from its directly attached devices but also from SPSS Modeler Server. It will benefit from a centrally managed policy along with SPSS Modeler Server.

For Cognos TM1, we qualified the different platforms options as shown in Figure 8-4. A rating of “10” means the highest possible rating and “1” means the lowest possible rating.

Non-Functional Requirements	Power-SA	Power-zBX	Linux-zBX	Windows
Security	6	8	8	3
Manageability	4	8	8	4
High Availability	6	5	5	4
Scalability	8	7	6	5
Response Time	8	8	7	7

Figure 8-4 NFRs for CPA Cognos TM1

Note: Values used here for response time and scalability represent our beliefs based on the circumstances of our workload. Any values used for response time, scalability, or other performance measures should be obtained in a benchmark with real applications, data, and environment.

Highlights from the assessment

Cognos TM1 stores its data in its own files and loads it in memory at start-up. Because data is not read from the database during normal operations, data proximity does not have an effect on performance.

8.6.4 SPSS Modeler Server

For SPSS Modeler Server, we considered the following platform options:

- Stand-alone POWER servers
- zBX-connected POWER7 blades
- zBX-connected x86 blades running Linux
- Linux on System z

We selected the following NFRs in 7.5.2, “Component characterization” on page 60 for SPSS Modeler Server:

- Scalability: CPA must accommodate a growing number of users.
- Security: Data in our data marts is classified as confidential. Unauthorized access must be stopped.
- Data proximity: SPSS Modeler Server will not only read data from its data mart; it will also populate data in it. Its proximity to the data mart will contribute to a better response time.
- Manageability: SPSS Modeler Server receives and produces data from one data mart but it also produces data for Cognos TM1. It will benefit from a centrally managed policy along with Cognos TM1.
- Response time: Response time is critical and will benefit from close proximity to the data.

For SPSS Modeler Server, we qualified the different platform options as shown in Figure 8-5. A rating of “10” means the highest possible rating and “1” means the lowest possible rating.

Non-Functional Requirements	Power-SA	Power-zBX	Linux-zBX	Linux on z
Data Proximity	4	5	5	8
Manageability	4	8	8	10
Security	6	8	8	9
Response Time	8	8	7	7
Scalability	8	7	6	8

Figure 8-5 Platform assessment for SPSS Modeler Server

Note: Values used here for response time and scalability represent our beliefs based on the circumstances of our workload. Any values used for response time, scalability, or other performance measures should be obtained in a benchmark with real applications, data, and environment.

8.7 Conclusions

We built the list of possible platform options for CPA components, with the reasoning behind each assessment. In the next chapter, we determine which, among all possible platform options, is the best fit for each component.

9

Stage 5: Prioritize solution architecture alternatives

Chapter Chapter 8, “Stage 4: List solution architecture alternatives” on page 63 presented the various platform options to deploy the *Customer Profitability Analytics* (CPA) application. In this chapter, we decide which platform is the most appropriate platform for each component of our application.

9.1 Introduction

In Chapter 7, “Stage 3: Assess new workload” on page 55, we determined which NFRs were relevant for each component. In Chapter 8, “Stage 4: List solution architecture alternatives” on page 63, we assessed the value of different platform options for our CPA components. In this chapter, from the values obtained in assessments and with the help of an IBM internal tool, we determine the best fit platform for each CPA component.

9.2 Tools and accelerators

The tool we used in the assessment is the IBM internal tool used for mapping application requirements over platforms. After assessing application components and qualifying platforms for each component, this tool provides a graphical comparison with the value of each platform for each CPA component.

9.3 Resources needed

In order to achieve our objective in this chapter, we need only the values determined in previous chapters and the tool to run on them.

9.4 Detailed steps

We used the following process flow that is shown in Figure 9-1.

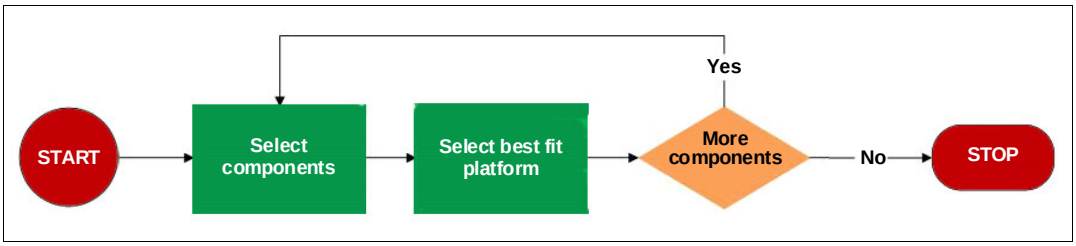


Figure 9-1 Steps for prioritizing architecture alternatives

In previous chapters, we established needs and rated platforms for each component of CPA. From these ratings, we can now prioritize solutions for each component.

9.4.1 Select the platform that is the best fit for each component

Note: Ratings used here for each component represent our beliefs based on the circumstances of our workload and might not be applicable to the same workload in different circumstances. Other workloads have to be assessed in order to obtain appropriate ratings.

New data marts

In “New data marts” on page 61, we selected the most significative NFRs for our data marts. The following list shows the final NFRs in order of importance:

1. Security
2. Response time
3. High availability
4. Scalability
5. Data proximity

In Chapter 8, “Stage 4: List solution architecture alternatives” on page 63, we rated four platform options for the new data marts:

- DB2 running on stand-alone POWER servers
- DB2 running on POWER7 blades within a zEnterprise BladeCenter Extension (zBX)
- DB2 for z/OS
- DB2 for z/OS and IBM DB2 Analytics Accelerator

Although DB2 running on stand-alone POWER servers rates highest among simple servers for response time, DB2 for z/OS rates highest for the other criteria. It rates highest for response time when attached with an IBM DB2 Analytics Accelerator.

Considering the selected NFRs, with platforms’ ratings weighed according to each NFR position in the list, Figure 9-2 shows the overall rating for each platform option.

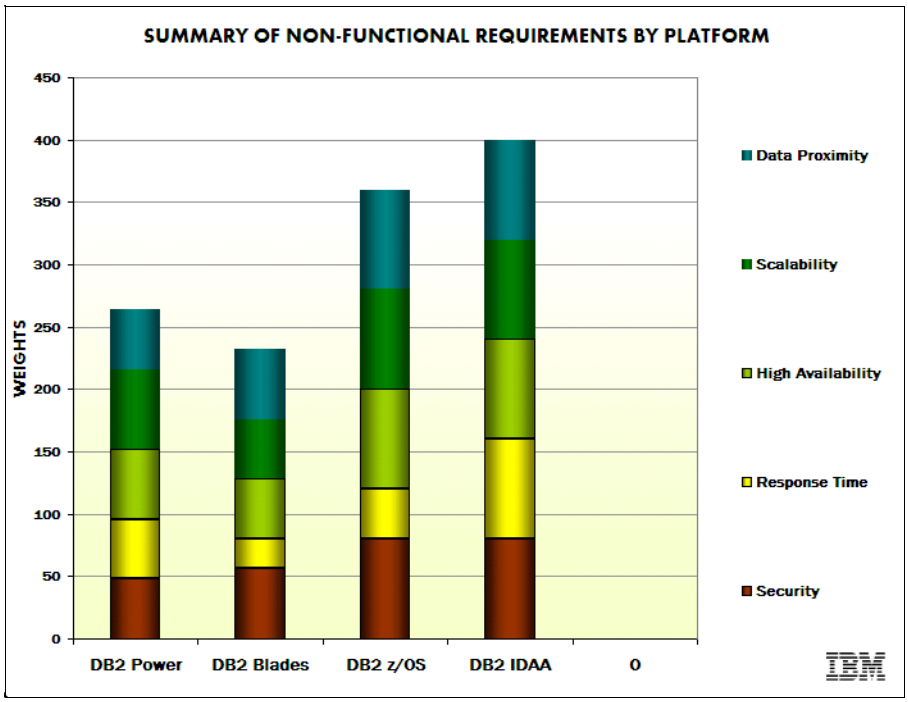


Figure 9-2 Overall ratings by platform option for CPA data marts

The graph shows that DB2 for z/OS with an IBM DB2 Analytics Accelerator has the highest overall rating for hosting CPA data marts.

Cognos 10 BI

In “Cognos 10 BI” on page 61, we selected the most significant NFRs for Cognos BI. The following list shows the final NFRs in order of importance:

1. Security
2. Scalability
3. High availability
4. Response time
5. Data proximity

In Chapter 8, “Stage 4: List solution architecture alternatives” on page 63, we rated four platform options for Cognos 10 BI:

- Stand-alone POWER servers
- zBX-connected POWER7 blades
- zBX-connected x86 blades running Linux
- Linux on System z

Linux on System z rates highest for security and data proximity. Stand-alone POWER servers rate highest in response time (along with POWER7 blades) and scalability (along with Linux on System z). Blade servers, either POWER7 or Intel, get intermediate ratings.

Considering the selected NFRs, with platforms’ ratings weighed according to each NFR position in the list, Figure 9-3 shows the following overall rating for each platform option.

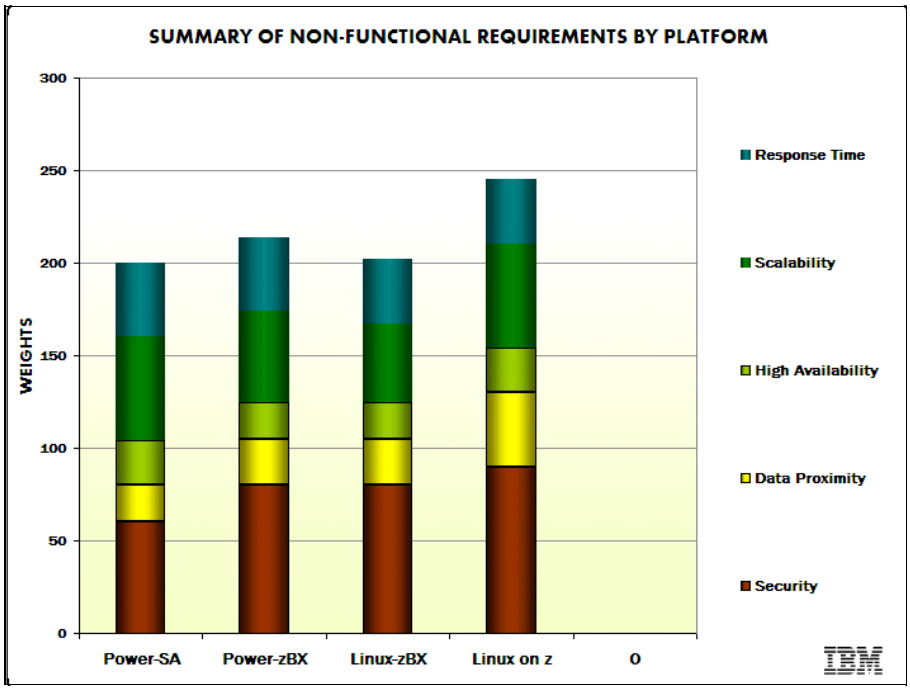


Figure 9-3 Overall ratings by platform option for Cognos 10 BI

Linux on System z gets the highest overall rating for Cognos 10 BI.

Cognos TM1

In “Cognos TM1 and Cognos Data Access Pack” on page 61, we selected the most significative NFRs for Cognos TM1. The following list shows the final NFRs in order of importance:

1. Response time
2. Security
3. Scalability
4. Manageability
5. High availability

In Chapter 8, “Stage 4: List solution architecture alternatives” on page 63, we rated four platform options for Cognos TM1:

- Stand-alone POWER servers
- zBX-connected POWER7 blades
- zBX-connected x86 blades running Linux
- zBX-connected x86 blades running Windows

Stand-alone POWER servers rate highest for scalability; Linux blades and POWER7 blades get highest ratings for security and manageability. Considering selected NFRs, with platforms’ ratings weighed according to each NFR position in the list, Figure 9-4 shows the following overall rating for each platform option.

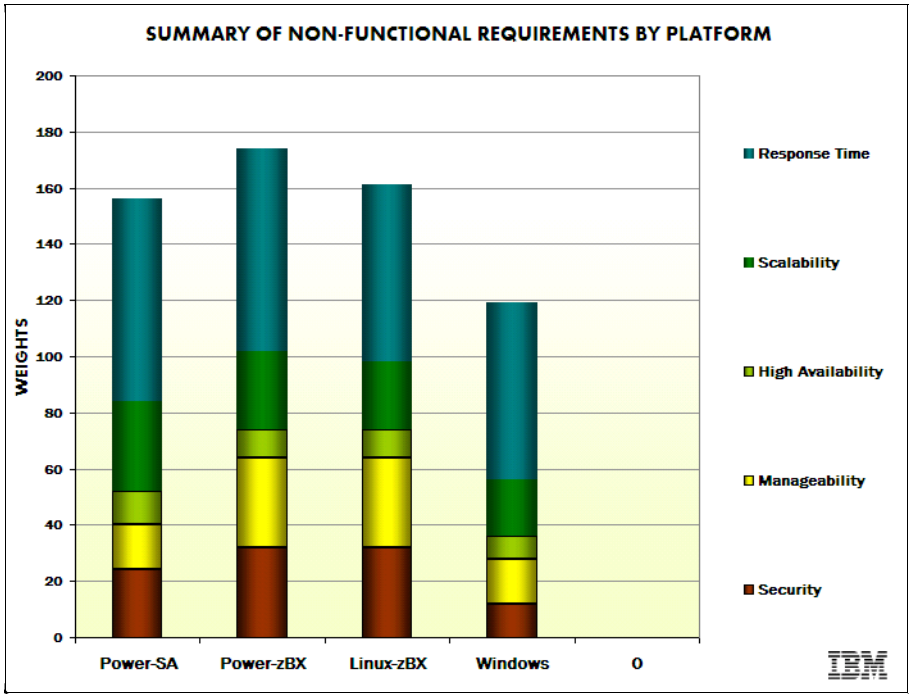


Figure 9-4 Overall ratings by platform option for Cognos TM1

AIX running on a zBX POWER7 blade gets the highest rating for Cognos TM1.

SPSS Modeler Server

In “SPSS Modeler Server” on page 60, we selected the most significative NFRs for this component. The following list shows the final list of NFRs in order of importance:

1. Response time
2. Security
3. Scalability

4. Manageability
5. Data proximity

In Chapter 8, “Stage 4: List solution architecture alternatives” on page 63, we rated four platforms for SPSS Modeler Server:

- Stand-alone POWER servers
- zBX-connected POWER7 blades
- zBX-connected x86 blades running Linux
- Linux on System z

Linux on System z rates highest for manageability, security, and data proximity. Stand-alone POWER servers rate highest in response time (along with POWER7 blades) and scalability (along with Linux on System z). Blades servers, either POWER7 or Intel, get intermediate ratings.

Considering selected NFRs, with platforms’ ratings weighed according to each NFR position in the list, Figure 9-5 shows the following overall rating for each platform.

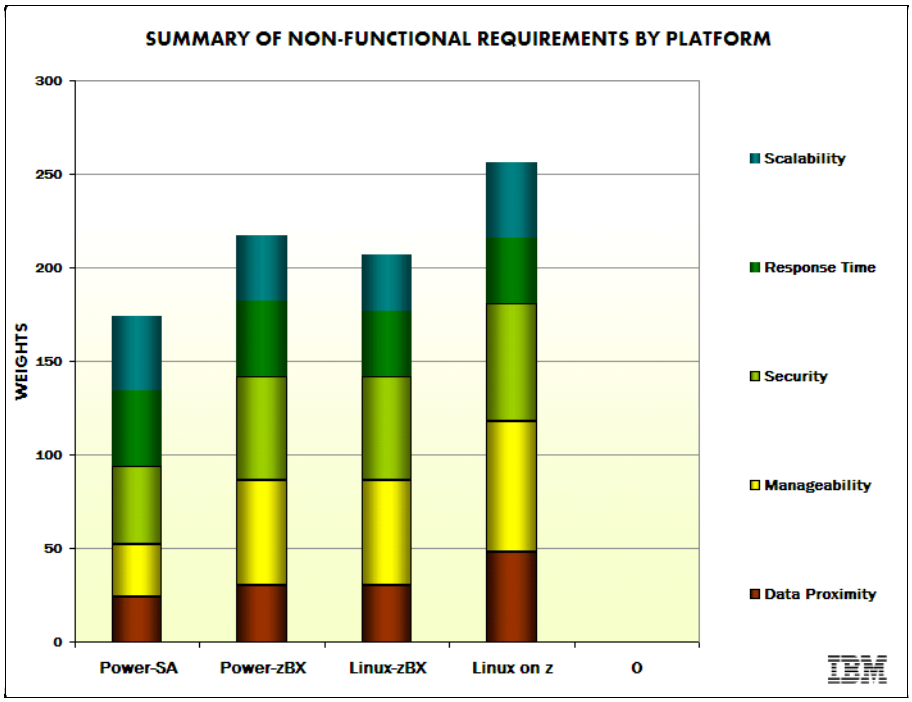


Figure 9-5 Overall ratings by platform option for SPSS Modeler Server

The graph shows that the most appropriate environment for SPSS Modeler Server is Linux on System z.

9.5 Conclusions

After performing the assessment for CPA, we determined the platforms that are the best fit for each component:

- New data marts:
  - Cognos BI reports: DB2 for z/OS with IBM DB2 Analytics Accelerator
  - Predict profitability: DB2 for z/OS with IBM DB2 Analytics Accelerator

  - CPA Metrics Studio: DB2 for z/OS with IBM DB2 Analytics Accelerator
- Software components:
  - Cognos 10 BI: Linux on System z
  - Cognos TM1 with Cognos Data Access Pack: zBX-connected POWER7 blade running AIX
  - SPSS Modeler Server: Linux on System z
- The following components were not included in this exercise:
  - Cognos 10 Framework Manager: zBX-connected x86 blade running Microsoft Windows
  - SPSS Modeler Client: zBX-connected x86 blade running Microsoft Windows

10

Stage 6: Sizing and capacity planning

In this chapter, we discuss relevant aspects of sizing and capacity planning on System z. We also focus on how these aspects relate to the previously described banking analytics environment.

It is our goal to help you understand the importance and complexity of sizing and capacity planning. Because the quality of the output depends directly on the quality of the input ("garbage in → garbage out"), we identify the characteristics of quality input. The ultimate goal is to enable you to provide quality input to the practitioners, enabling them to perform their duties by using the best available data, therefore producing quality reports.

10.1 Overview

Some process steps are required before the sizing or capacity planning stage. At this point, the architectural options have already been evaluated and defined. A proper sizing or capacity planning can only be executed after we know the components of the solution. Then, we can discuss how to transform these options into actual configurations. Sizing or capacity planning can be correctly accomplished by using a variety of tools and techniques that are based on our industry experience, knowledge of different server architectures, type of applications and workloads, current environment, and performance metrics.

Performance metrics are normally expressed in three different ways:

- If an application already exists and this application is being moved from a distributed platform to zEnterprise, performance metrics are normally related to the existing real CPU and real memory consumed in the execution of the task under analysis. In the scenario used in this book, the sizing study is associated with a consolidation or IT optimization scenario.
- If an application is new and zEnterprise will host it for the first time, the term “performance metrics” refers to the raw data specifications, such as the number of transactions per second or the size of the database. This is also a sizing study, but the application is sized from the ground up.
- If an application already runs on a System z or zEnterprise and the goal is to evaluate the impact of migrating this application from “z to z”, this scenario is related to a capacity planning study, and the term “performance metrics” refers to the specific System z parameters, such as IBM Resource Measurement Facility™ (RMF™) reports and System Management Facilities (SMF) data.

This chapter discusses sizing and capacity planning studies. The sizing has two subcategories, depending on the initial environment. We also elaborate on “z to z” capacity planning tools and techniques. We provide examples of the required steps to correctly size the infrastructure to support a smart analytics environment for banking, such as *Customer Profitability Analytics* (CPA).

Each approach previously discussed is adequate for a specific situation and each of these approaches also offer different levels of “precision.” We use the term “precision” loosely. There is no precision when you are converting applications to run on different platforms, or estimating requirements for a new application. Only in a capacity planning environment (from z to z) can you use the term “precision” to indicate the level of confidence in the results, which is commonly expressed as an interval defined by plus or minus a percentage value (example: 5500 MIPS +/- 10%).

What we are really emphasizing is that sizing is not an exact science nor a magical process. Actually, sizing is a very fine combination of art and science, where the practitioners use their skills and experience to process estimated or measured (preferred) performance data to produce performance projections about how specific workloads would run when migrated to, or first running on a specific System z or zEnterprise server. Capacity planning (z to z migration) also requires the same art and science, but considering its maturity level and the fact it handles growth within the same architecture, it is more biased to the science side.

The importance of the correct execution of sizing or capacity planning cannot be minimized since these are the techniques that will make the difference between having a good-performing system or a system presenting performance issues. Therefore, well-executed sizing or capacity planning is the key to preventing losses on the quality of service (QoS) defined by corporations, preventing the occurrence of critical situations, and enabling IBM to ensure high levels of user satisfaction.

Most of the advice given in this chapter specifically targets the System z architecture, a very flexible platform that offers alternatives in terms of hardware (HW) and software (SW), tuned for mixed workloads and with virtualization that has evolved for more than 45 years. However, the concepts presented are applicable to sizing and capacity planning in general.

Whether you are building a completely new system, working with existing systems, or planning capacity within the System z or zEnterprise families, you do not need to be an expert to leverage sizing and capacity planning in your environment. If you qualify for IBM support and are unfamiliar with the required techniques, help is available through the IBM Global Techline or from your local IBM technical resource, who can advise you on how to gather the required information to enable the experts to perform an appropriate sizing or capacity planning analysis.

The IBM Global Techline provides the required expertise in qualified pre-sales situations. It offers technical support in areas, such as IT Optimization Analysis, through cross-platform sizing, the sizing of new applications, and capacity planning studies. Through its capable pool of IT professionals, Techline provides the resources to analyze each situation on a case-by-case basis. The following examples are Techline work products:

- Detailed sizing or capacity planning report
- Total cost of ownership (TCO) comparison
- Hardware and software infrastructure configurations

How do sizing and capacity planning relate to the real world? The answer is simple: they are the technical solution that needs to be integrated into the business solution. The technical solution needs to be merged into other constraints, such as business requirements, and local factors such as budget, skills, and policies.

The technical solution does not constitute the absolute and only recommendation. It is important to note that any sizing results should be taken for what they really are: guidelines or a starting point. The person in charge of the infrastructure can and should override any sizing results to accommodate local business requirements. However, any identified risks must be assessed and mitigated.

Consider these helpful resources:

- *Considerations for Transitioning High Available Applications to System z*, SG24-7824:

  http://www.redbooks.ibm.com/abstracts/sg247824.html?Open
- *OS/390 MVS Parallel Sysplex Capacity Planning,* SG24-4680:

  http://www.redbooks.ibm.com/abstracts/sg244680.html
- *IBM zEnterprise System Technical Guide*, SG24-7833:

  http://www.redbooks.ibm.com/abstracts/sg247833.html
- *Linux on System z, an end-to-end view: From Sizing to Performance Analysis to Capacity Planning*:

  http://www-03.ibm.com/support/techdocs/atsmastr.nsf/84279f6ed9ffde6f86256ccf00653ad3/1c76a8c5aa63c3b28625723a0026816b/$FILE/Linux%20on%20System%20z,%20an%20end%20to%20end%20view.pdf
- IBM Global Techline Center of Excellence:

  http://w3-03.ibm.com/support/techline/global/index.html/

10.1.1 Sizing considerations and objectives

We assume the platform hosting decisions have already been made. We described how zEnterprise business value assessment (zBVA) and Fit for Purpose (F4P) methodologies were used to make these decisions for the CPA workload. This would be a very lengthy chapter if we describe the sizing of every workload component on every possible platform for the entire banking analytics implementation.

Therefore, this chapter focuses on just one of the components: Cognos 10 BI for System z Linux. It is our intent to use the sizing of this component as an example. The other components in our solution have a similar sizing process, so you will gain an understanding of the process by following the Cognos 10 BI example.

10.2 Process overview

A proper sizing is considered the starting point and the basis of an IT optimization project where workload and application relocation or consolidation will occur (see Figure 10-1), or when a new application is about to be brought up for the first time. You need to understand the requirements to move workloads from one platform to another, or to properly host a new application.

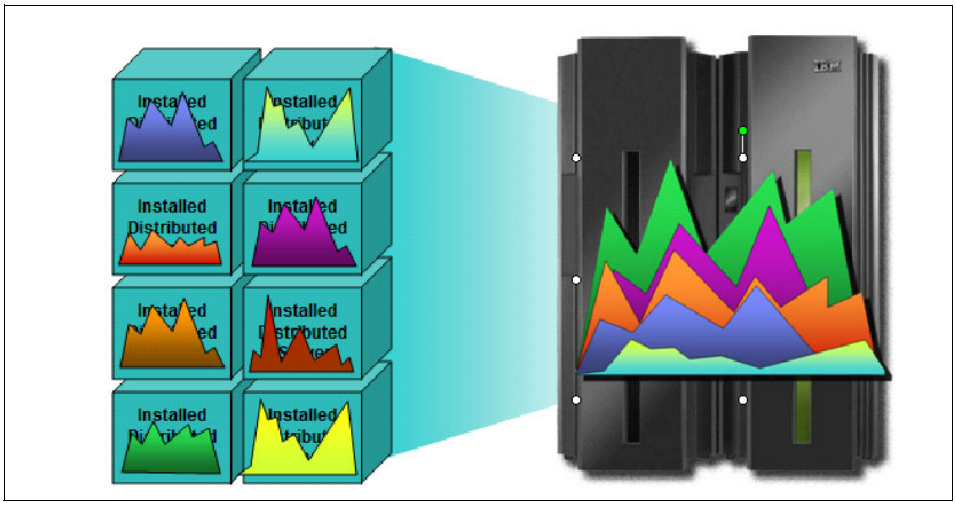


Figure 10-1 The z architecture can consolidate workloads from many servers

In the scope of this chapter, we will always consider the target platform as the zEnterprise, in its stand-alone form factor (z114, z196, or zEC12), or in its hybrid form factor, which includes these components:

- POWER7 and x86 blades
- IBM DB2 Analytics Accelerator for z/OS

A relocation or sizing discussion will always cover workloads or applications that are hosted in “non-z” platforms that are being moved to the “z” architecture. It also might include any new workload that is sized to run on the zEnterprise, on its “z” side or in its blades (zEnterprise BladeCenter Extension (zBX)) or analytics (IBM DB2 Analytics Accelerator) hybrid components.

10.2.1 Sizing an IT optimization (consolidation) environment

This is a relatively complex task, particularly when handling sizing across platforms built on top of very different architectures. However, provided the input data (hardware description, performance metrics, and application or workload type) is known and trustable, there are sizing methodologies that can produce practical results that can be used as the starting point to build a new solution hosted into the new architecture.

A proper sizing requires the best and most accurate data and description of the current environment. Although there are tools and methodologies that would handle estimated data by using statistical modeling to represent the existing environment, it is the opinion of the authors that nothing replaces the actual measured real data, such as CPU consumption metrics, memory consumption metrics, and I/O intensity metrics, collected from the existing servers. Real data will always better represent the environment that is being consolidated than estimated data.

A good description of the servers that are consolidated into zEnterprise includes the server make/model, number of chips/cores, clock frequency (GHz), memory utilization, CPU utilization, and a description of each workload or application running in each server. The number of intervals, duration of each interval, dates, and time stamps should be common to all servers that are being consolidated, or at least within a narrow window (quasi-simultaneous data collection). Variables, such as CPU and memory, are to be stacked in a single graph against time, enabling the assessment of the whole interval that is being considered in the study.

The data collected should represent what actually runs in production in these systems. This ensures that the future projection will be based on a true picture of today's environment. See Figure 10-2.

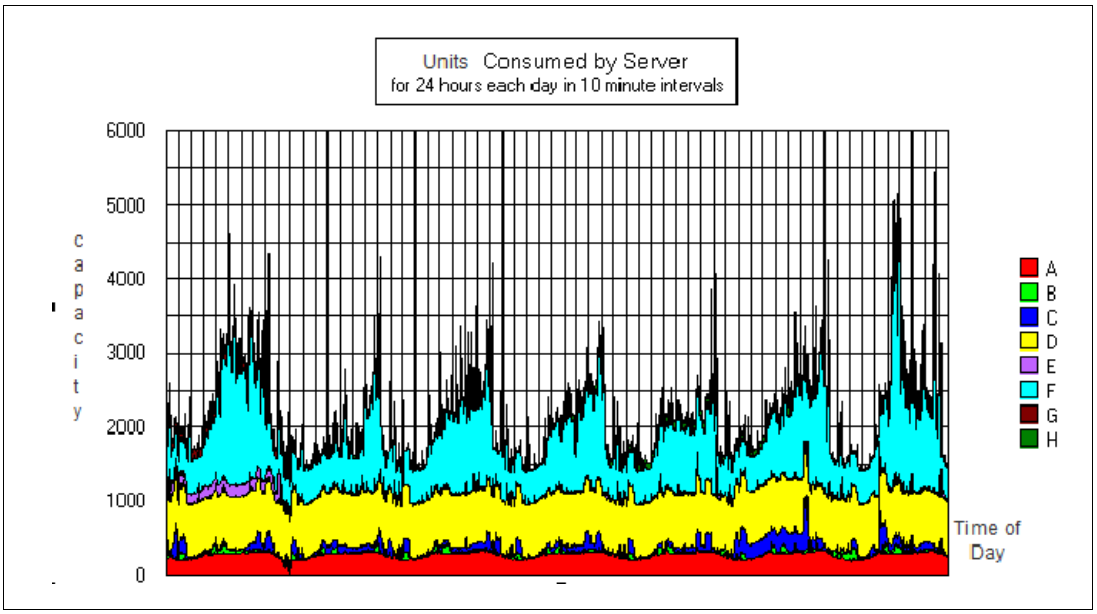


Figure 10-2 Example of a stacked graph representing a given interval for eight servers

One major exception to the real data policy: old data is not much better than no data. The data must be recent to properly represent the current consumed capacity. Planning a consolidation based on old data will almost certainly generate results that will not reflect the current capacity needs. Special care needs to be taken with the selection of the granularity of

the date/time, since employing very long intervals can lead to averages that may conceal important peaks and the true performance requirements of the target system.

In a consolidation scenario, sizing can be defined as the “translation process” from one platform to another. This is the science and the art that enable the visualization of two different architectures as a common base, allowing an “apples to apples” comparison across servers from diverse architectures.

First, you need to consider what current performance data is available for the servers that are being consolidated. If regular reports are being produced to track service level agreements (SLAs), possibly there is already a preferred tool to collect performance-related data. If there is no data collection standard in place, we will make recommendations of how the sizing input data can be gathered using freely available tools, or by using system-level commands.

A proper and reliable sizing is essential. If it results in significant underestimation of the required capacity, this is clearly going to cause problems when the new system enters into the production phase. If the results are overestimated, it will lead to a non-optimized solution.

Data collection

Data collection is an important step in any consolidation sizing effort. It involves gathering the data elements required to enable further analysis. As a starting point, the collected data needs to be adequately representative of the current system. Knowledge of how the system performs is useful, such as when peak loading periods occur and how they relate to any SLAs in place.

The most reliable way to perform consolidation sizing is to use measured real data as input. The sizing input requires, at minimum, the processor family (x86 or RISC), the number of chips, the number of cores, utilization, the clock frequency, and the application/workload characteristics.

Normally, this process is done for more than one server at a time. In this situation, the data collection needs to be performed in as synchronized a manner as possible or quasi-simultaneously, which means that you need to collect data from all servers being studied during the same time intervals and using the same time stamp format. The goal is to allow the servers’ data to be stacked into a single graph, where the overall utilization peak represents the peak of the entire interval for all servers combined.

It is important to note that the sum of the peaks is usually larger than the peak of the sum; therefore, by using the peak of the sum to represent the collective resource consumption, you can make use of the planning methods and techniques and produce the most efficient results.

Figure 10-3 represents an example of a sizing input sheet.

Input Sizing for Server Consolidation to Linux on System z

Machine Name	Vendor	Model/ Type	Total # Chips	Total # Cores	CPU Mhz	Utilization	Type of Workload	Comments
Sample: App Server 1	IBM	RS/6000 B50	1	2	375	27%	Web Application Server: With substantial JAVA application	peak utilization is during database re-org
Sample: Prod East Coast	Dell	PowerEdge 2950	2	4	1600	23%	Web Serving: HTTP Server,general usage	

Figure 10-3 Example of a data collection spreadsheet

The data collection must be done within a period of time that represents the typical production business processes, normally when critical workloads are running. Invariably, this will involve consideration of the peak loading periods since when a system is configured to perform adequately during expected peaks, the system should be able to perform well on non-peak periods. The period of time chosen to represent the entire interval being considered should ideally show the most important high-level resources (processor, memory, I/O, and network) being exercised at the typical levels seen when the system is the busiest.

An alternate, more aggressive technique replaces the overall peak by the *90th percentile*. To help you understand this approach, the 90th percentile value for a set of values states that at least ninety percent (90%) of the values in the set are less than or equal to this value. In summary, the top 10% of the available samples is disregarded.

Data collection for sizing is conducted with the aid of appropriate tools. Very often, basic data collection tools are already part of the system and well known to sizing practitioners and capacity planners. Although often not as powerful or easy to use, operating systems offer their own standard monitoring tools.

10.3 Tools and accelerators

IBM has tools and methodologies to deal with sizing and capacity planning. However, these are not the only methods and tools that can be used. Clients often have their own methods to size and plan capacity. Use the method with which you are most familiar. Since sizing and capacity planning is a complex business, we also suggest that a qualified practitioner be engaged.

IBM offers to the public a capacity planning (z to z) tool named zPCR. Information and a download link can be found at the following URL:

http://www-03.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/PRS1381

10.4 Resources needed

Most of the time, a new application sizing starts with a customer interview. This interview can be conducted in person or with a sizing questionnaire. The objective of the interview is to gather the required input information that is necessary to produce the sizing results. The information provided by the customer is just the starting point. After the data is collected, you need to interpret the requirements and then apply these requirements as input in one of the sizing methods. A sizing method might include a manual or automated processing of the input data. When the data is processed in an automated manner, that implies that the methodology is implemented into a sizing tool. In addition to the questionnaire or interview to collect input data and a method or a tool, you need the required skills to understand what to do with the information to produce the sizing. If the person does not have the required skills, they must seek technical support from an experienced sizing practitioner.

10.5 Detailed steps

Figure 10-4 on page 86 illustrates the sizing and capacity planning process flow. It starts with the need to size a specific application, represented by the red circle “START”. The next step is to understand whether the application being sized is already running on a specific platform or whether it is a new application being deployed.

If this is a new application, you need to understand the application and determine which requirements need to be fulfilled by the application, such as the number of transactions per second, transaction complexity, and number of users. Raw data like this is a normal input requirement to the sizing of a new application from the ground up. After the requirements are understood and collected, you need to determine whether a sizing process is already established and whether there are methods and tools in place to perform the sizing. If these conditions are met, the sizing can be done by a capable technical person familiar with the process and tools. If the sizing process is not yet available to a technical subject matter expert (SME), the SME needs to secure support from the application developer and seek guidance in how to proceed. In either situation, the SME applies skills and experience to produce the sizing report, specifying the characteristics of the infrastructure capable of hosting the application or workload.

If the application already exists (on an existing server), the question is whether the current server is a z architecture-based mainframe or not. If the server is zEnterprise architecture-based, the follow-on question is whether the architecture is pure z (System z) or hybrid (System z and zBx, for instance). The flow is very similar for both situations, except that, when the hybrid component is present, we then need to involve an SME who is knowledgeable of the hybrid platform. The z side is handled by standard “z to z” capacity planning.

If the existing server is not a server that is based on the z architecture, the sizing is actually an IT optimization/consolidation exercise. The consolidation sizing process involves the performance data collection, the review of this data, and the use of the SME skills to produce a sizing report laying out the target hardware characteristics to host the application or workload. See Figure 10-4 for details.

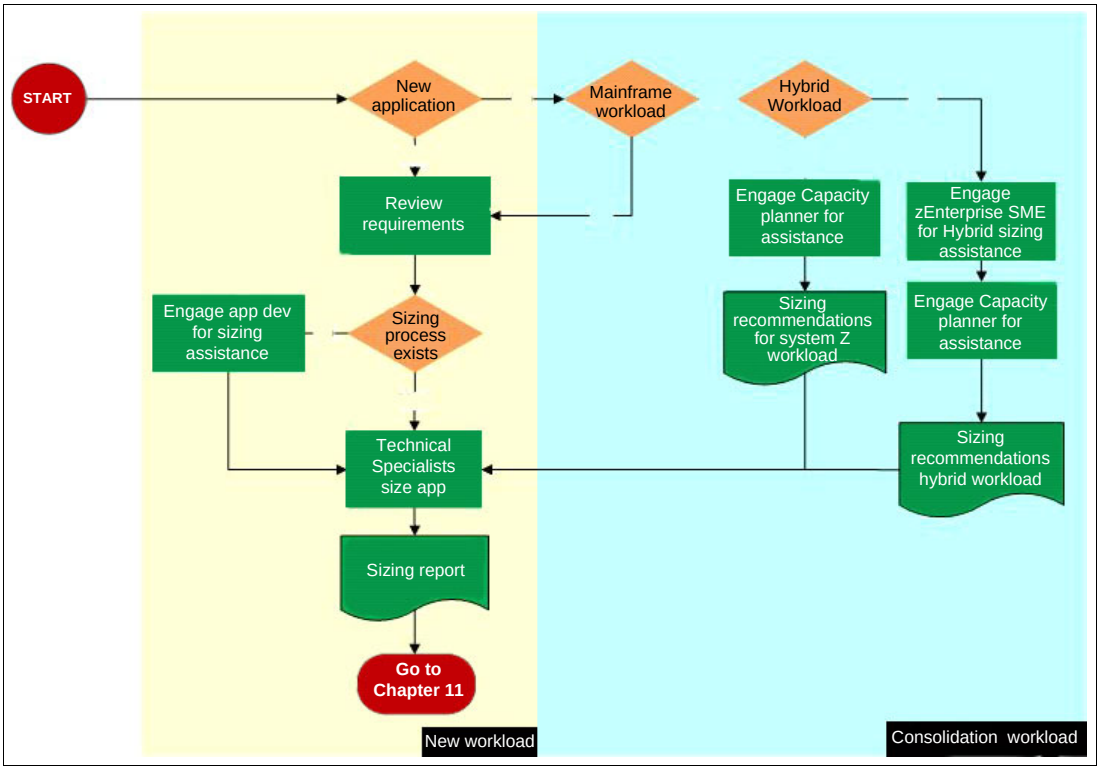


Figure 10-4 Sizing and capacity planning process flow

10.5.1 Sizing a new application

The sizing of a new application requires a different approach, since there is no actual data to be viewed because the application does not yet run on any platform. A new application sizing might fall into one of the following scenarios:

- The new application sizing process is already established. There is an existing, matured methodology, and an existing sizing tool is available to the SME to apply the required skills to produce a sizing report. However, the sizing process might not include the target platform, subdividing this scenario in two versions:
  - The existing sizing process is available for the target platform.
  - The existing sizing process is not available for the target platform. In this scenario, you need to size the application to any of the existing platforms supported by the sizing process and then apply consolidation sizing techniques to estimate the application requirement in the target platform.
- The application is immature and there is no established sizing process. You need to seek advice from the application developers to be able to produce or validate a sizing report.

Typically, the sizing of a new application requires the client or the technical resource working with the client to answer a questionnaire. This questionnaire is required to capture information about the application, such as the intended number of users, database size, number of transactions per second, or transaction complexity.

The data collected is then sent to an SME who will work with development to estimate the infrastructure requirements.

10.5.2 Total cost of ownership (or cost and value)

One important factor to mention is that the sizing process, although required, often is not all you need to effectively prove the value of the System z platform. A total cost of ownership (TCO), which is also known as “cost and value”, analysis is an effective mechanism to demonstrate the value of the z platform. It elevates the conversation from the technical aspects to the financial vocabulary, which is better understood by the C*xx* executives, mostly the CIOs and CFOs.

It is crucial to demonstrate that, for instance, a proposed consolidation of workloads into the z platform would reduce or keep costs manageable within an expected range. TCO normally factors costs, not only in terms of infrastructure acquisition costs, but also in terms of software licenses, footprint, maintenance costs, power and cooling costs, and costs of administration, all spread along a specific period of time, normally three to five years.

The chance of having the value of zEnterprise recognized by the clients’ high-level executives moves the decision-making process from the passionate IT side of the house (local view) to the more strategic view provided by the C*xx*s. This enterprise view offered by the C*xx*s is more likely to enable decision-making based on a more complete analysis, where not only are the technical aspects considered, but also the business component.

There are different methodologies to establish the System z and zEnterprise cost and value (TCO). In chapter Chapter 11, “Stage 7: Execute total cost of ownership (TCO) analysis” on page 95, we discuss the basic common guidelines to help you understand the process.

Figure 10-5 illustrates the TCO concept.

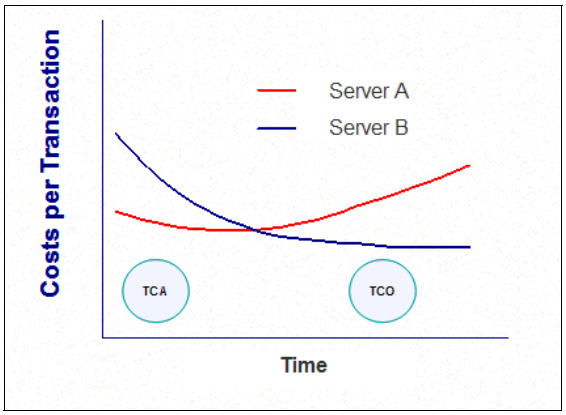


Figure 10-5 A typical TCO factors the operational costs for 3 - 5 years

10.5.3 Capacity planning considerations

Capacity planning is the methodology to verify and analyze the hardware requirements needed to accommodate the growth of workloads that are already running on or sized for the z architecture.

The capacity planning process takes into consideration the workload performance while it is running on a System z or zEnterprise. Then, the process enables you to forecast and create projections by inserting a specific growth rate into the equation. After the analysis is completed, it will produce a recommendation defining the proper System z or zEnterprise machine model that is capable of absorbing the specified growth within the z platform. This projection can be made to estimate a model within the same z family (for example, zEnterprise), as well as across generations (such as System z9s or z10s to zEnterprise).

Typical situations that require the use of the capacity planning methodology can be found when an existing z architecture-based server is running out of capacity. The existing workloads are consistently reaching 100% utilization and there is a need to deploy new workloads or to improve and maintain SLAs. For example, assume a client already has an IBM System z or zEnterprise server and has seen an increase in CPU usage. Among other things (aside from an atypical spike), it indicates the current machine does not have enough capacity to process the inflated workload within the agreed-to SLAs. There is then a need to add more hardware to bring the SLAs back to an acceptable level. The following questions arise from this situation. How many additional processors are needed? Should the client replace the existing machine with a newer z generation?

Another possible scenario is a client that wants to add a new workload to the existing z architecture-based server. How much extra processor consumption can be expected? Will the new requirements be able to fit into the existing server?

Capacity planning provides the methodology and tools to handle such situations, because it helps you to understand and define the requirements to ensure adequate processing capabilities.

Ultimately, capacity planning helps you to verify and analyze a specific system configuration and respective performance and then project the expected growth into the future. It also helps assess the effect of this growth into the current or proposed new z architecture-based environment.

This approach is important because it enables the view of the current environment in the future, as though it were a time machine. Given that the entire system will grow over time, the idea is to devise a system that anticipates and supports that growth. It is also important to correlate the future needs of a system with *any* planned changes, such as changes in the underlying business areas. For example, there are times when a contraction might be required. When capacity planning is well executed, it will provide the means to an optimized system performance, which in turn enables cost management. See Figure 10-6.

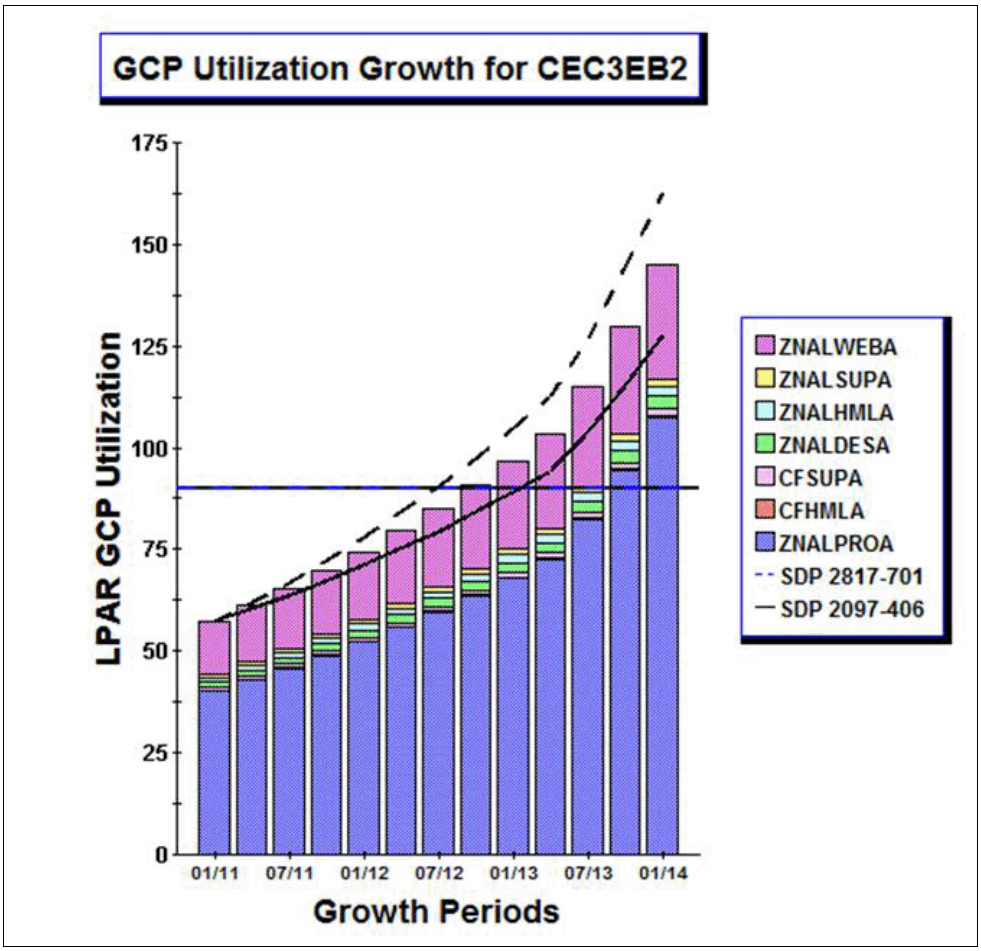


Figure 10-6 Example of a capacity planning estimation

Although a capacity planning study normally covers more than one year, it is advisable to recheck the results in regular intervals, at least after each year. In one year, many things might change and these changes can invalidate the multi-year performance prediction previously executed.

An optional but important element of the capacity planning process (also applicable to the sizing studies) is the verification of how close the real performance is when compared to the projected performance. In a comparison situation, it is important to check whether the initial assumptions that drove the initial study are still true. You need to be alert to detect whether new workloads that were not part of the initial study have been added, whether the properly sized hardware has been deployed, or whether the SLAs have been changed. This emphasizes the importance of monitoring the system on a regular basis.

Whenever possible, verify actual results against projected results; that is, undertake the classic “before and after” comparison. The goal is to show that the base capacity planning (when from one z architecture-based server to another z architecture-based server) or a specific consolidation sizing projection is allowing the system to run within the expected levels of performance.

Tuning can also be part of the process. It targets the optimization of the system to enable it to run as efficiently as possible. If tuning is required, it is helpful to modify only a single parameter at a time. After the change, the consequences of this specific change need to be evaluated. After the consequences of the first targeted individual tuning are fully understood in terms of its impact to the system, then (and only then) you can move to change the next parameter. Iterate until no more parameters are left to be changed. “Bulk” (simultaneous) changes are not helpful because they do not enable you to understand the cause and effect of these changes. Unraveling the interdependent effects of multiple changes can be difficult and sometimes impossible.

The goal of understanding the “before and after” situations, whether by capacity planning (z to z) or through sizing (consolidation or new workload), is to better understand the results, validate the input, better forecast future growth, and create a larger database of real cases to be fed back to development. This process enables each of us to participate in the enhancement of existing tools and methodologies.

We have described many versions of sizing and capacity planning processes. Keep in mind that from this point on, when we refer to sizing, that will be a reference to the process that handles sizing of a new application. The sizing description that follows will be an *example* of a scenario, similar to the Banking Smart Analytics scenario we are implementing.

The example that follows is only to illustrate the process. The input numbers used and the results produced do not map against the real implementation described in other chapters of this book, nor in other related volumes.

10.5.4 Cognos 10 BI sizing for Linux on System z: Input

The process to size Cognos 10 BI for Linux on System z is well established. When targeting the System z architecture, it is a two-step process that we explain later in this section. It is recommended that you seek IBM’s support for help with these tasks.

The sizing guidelines are based on IBM internal testing and real-life experience. Successful sizing guidelines are always developed in collaboration between a qualified client and IBM Cognos representatives. By working together, we ensure the proper interpretation of the information collected throughout the client interview process and the correct application of Cognos 10 sizing methodology.

Sizing input data

You need to gather the following types of information to enable the Cognos 10 BI for Linux on System z sizing:

- User roles
- User concurrency
- Use of portlets and Business Insight Consumer dashboards (BIC)
- Solution access patterns
- Application complexity
- Query complexity
- Output complexity
- Percentage of overall BI content for heterogeneous query
- Percentage of overall BI content for master/detail query

- Percentage of overall BI content using personal data merge
- Percentage of overall BI content Dimensionally Modeled Relational (DMR) data query
- Percentage of overall BI content for statistics
- Batch report processing requirements
- IBM Power Cube® building requirements
- Technology preference
- Deployment option

As you can see, there are many areas required to be part of the input to produce the proper Cognos 10 BI sizing. Some of these entries are only the top header of many sub-items. Describing all the entries here is beyond the scope of this book. However, we chose a few areas to illustrate the level of the input requirement.

User roles

The idea behind asking the client to define user roles is to provide the required granularity to enable you to differentiate ratings of concurrent users based on different user roles. Examples of user roles are illustrated in Figure 10-7.

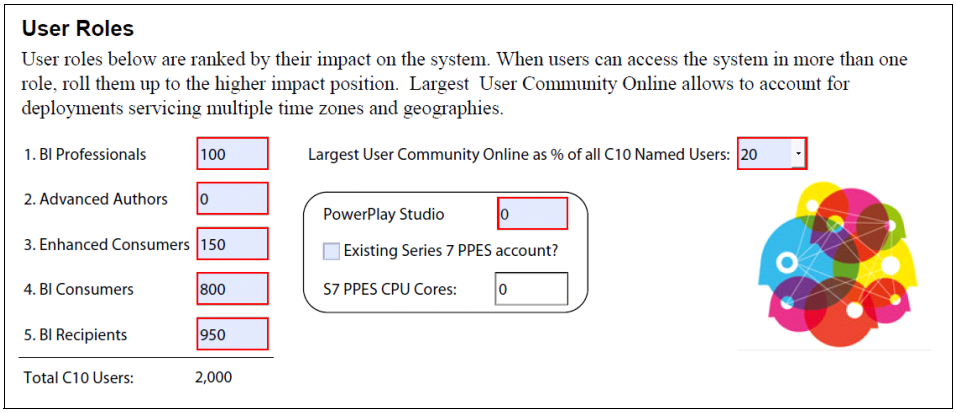


Figure 10-7 Types of Cognos 10 BI users

Application complexity

No two client applications are exactly alike. To produce a proper sizing, you have to provide a simplified measure of projected or existing application complexity, described in a number of areas. In each area, application complexity is expressed through four categories: simple, intermediate, extended, and highly extended. Figure 10-8 on page 92 illustrates an example of query complexity parameters.

Query Complexity

The following criteria represent complexity levels in this area.

Please specify projected query complexity. When applicable, percentages on each side (interactive and batch) have to add up to 100%.

Interactive		Batch	
Simple %	40	Simple %	0
Intermediate %	50	Intermediate %	70
Extended %	10	Extended %	20
Highly Extended %	0	Highly Extended %	10
Total %	**100**	**Total %**	**100**

☐ Use Defaults ☐ Use Defaults

Figure 10-8 An example of how to classify a query complexity

Batch report processing requirements

One of the areas that affects the final sizing results is the size of the required batch window. The batch processing window is typically the period that begins when report data becomes available (that is, the database load is complete) and ends when users expect to receive reports based on that data, regardless of whether there is a dedicated window or if the batch processing is time-shared with an interactive access.

Burst reports are different from simple batch execution. With *burst processing*, a database query brings back data for all report recipients at once. Cognos 10 processes this master data set and produces personalized output versions according to the defined burst keys. If there is a need to process a batch job in two hours rather than 12 hours, the hardware requirements will be much larger for the former job. Figure 10-9 illustrates examples of input batch parameters.

Processing window for batch reports (Hours):	6 ☐ Dedicated?
Number of reports to process per batch window (excluding burst):	5
Number of burst reports to process per batch window:	15
Average number of recipients per burst report:	50
Average number of pages per burst report output:	15
Average number of versions retained per batch report:	1
Average number of versions retained per burst report[1]:	1

Figure 10-9 Parameters defining our Cognos 10 BI application batch requirements

Figure 10-10 is a summary of all the assumptions we used to create the application profile.

Assumptions

BI Concurrent Users:	5.35	PP Studio Concurrent Users:	0
Batch Processing:	15 burst reports to 50 users each and 5 batch reports		
OLAP Builder Required:	No		
Dedicated Batch Window:	No, 6 hours	Dedicated Window for OLAP:	No, 0 hours
Inter. Query Complexity:	40% Simple, 50% Intermediate, 10% Extended, 0% Highly Extended		
Inter. Output Complexity:	40% Simple, 50% Intermediate, 10% Extended, 0% Highly Extended		
Batch Query Complexity:	0% Simple, 70% Intermediate, 20% Extended, 10% Highly Extended		
Batch Output Complexity:	0% Simple, 60% Intermediate, 35% Extended, 5% Highly Extended		
Local Processing:	20% Heterog., 30% MD, 40% Personal Data, 70% DMR, 0% Statistics		
Personal Data Size:	500-10000 rows or 3MB	Statistics Data Size:	Not Applicable
Deployment Platform:	System Z IFL, Linux		
C10 Graphics Engine:	AVS	Dynamic Query Mode:	No

Figure 10-10 Summary of the assumptions used to size phase 1

10.5.5 Cognos 10 BI sizing on Linux on System z: Results

Note: This sizing is for illustration purposes only, based on a hypothetical application profile. The results are *not* linked to the implementation.

After considering the input to each sizing parameter that is included in the interview form, we came up with the initial sizing results. One of the characteristics of the Cognos BI 10 sizing tool is that it provides the results in two phases: phase 1 expresses the results targeting the x86 platform and phase 2 converts these results to the Linux on System z architecture platform. The initial results provided by phase 1 are listed in Figure 10-11.

C8 Component	CPU Cores	RAM (GB)	Virtual Servers
Web Gateway	0.1	6.1	1
Report Service	4.18	22.36	2
Content Manager Service	0.7	8	1

Figure 10-11 Phase 1 results are expressed using the IBM x3690 as a base

These results are then converted to millions of instructions per second (MIPS) and the number of necessary Integrated Facilities for Linux (IFLs)[1]. Also, the memory requirements are converted to the z architecture. The results of the phase 2 sizing process are included in Table 10-1 on page 94.

[1] Specialized processors known as Integrated Facility for Linux

Table 10-1 Phase 2 sizing process

	MIPS needed	IFLs needed[a]
Web Gateway	19	0.02
Report service	2603	2.08
Content Manager service	336	0.27
Totals	2959	2.36

a. At 1259 MIPS per IFL

These results need to be recalculated if z/VM is used. In our scenario, the Linux virtual machine is hosted under z/VM. Therefore, we need to use the appropriate tools to calculate the resources that will be required by the virtualization layer. In our example, the total required resources are three z196 IFLs.

10.6 Conclusions

Sizing and capacity planning require detailed input, proper methodologies and tools, and qualified technical SMEs. The goal is to provide the technical guidance to define the required hardware that can run a new application, run an existing consolidated application, or in the case of capacity planning, project growth within the same platform.

During our Linux on System z Cognos 10 BI sizing example, we demonstrated some of the characteristics and laid out the assumptions that are required to complete this application sizing.

We described the two-phased process and elaborated on the requirements to run it virtualized with z/VM.

In the end, our results indicate that three z196 IFLs are enough to run our Cognos 10 BI on Linux on System z (in our hypothetical example). Leveraging the System z and zEnterprise scalability will be important to accommodate future growth.

After the application is running on the z architecture, it is strongly recommended that capacity planning is done. At that time, the architectural conversion will be over and the capacity planning process will assess the requirements from z to z. This is the best scenario to produce results that are more accurate.

In summary, the strategy is to size the application and use the output of this sizing (including all phases and z/VM) as a guideline to start the initial deployment on the z platform. Then, after you are using the target architecture, use the capacity planning capabilities to help plan the move of the system from pilot to production.

11

Stage 7: Execute total cost of ownership (TCO) analysis

In every business decision, cost is an important factor. How much does it cost? How much will you save by making a certain decision? Which of the available solutions will meet the service-level agreements (SLAs), and, at the same time, be the most economical?

These questions are part of the *repertoire* of any decision maker. However, anyone who is to provide an answer to these questions needs to understand the meaning of the word “cost”.

Let us start from a generic definition of the word cost and then work it down to the IT environment. One of many possible definitions is that cost is the amount of a specific resource you have to offer to acquire an item or to receive a service.

Let us say you want to buy a car. The cost is the amount of resources (in this case currency) you must have available to acquire the car, license the car, tag the car, insure the car, provide fuel to be able to run the car, store the car when not in use, or fix the car when needed.

In other words, if you approach the business of buying a car solely based on the cost of acquiring the car, with no resources to pay for the expenses inherent to the car operation, the car will not be able to run. If broken, it will not be able to be fixed, and there will be no place for the car to be stored. If there is no tag, the car cannot be driven, and so on.

Why are ideas that seem so obvious when put into the context of buying a car so easily forgotten when the business is buying a server? Are these tasks so different?

The authors propose that they are not, and that is the notion behind this chapter. We discuss, under the cost perspective, what is involved in the business of acquiring a server, running it, and maintaining it.

11.1 TCO versus CO

There can be a tremendous difference between the total cost vs. just the cost of acquisition. Back to our car purchase example, let us look into what is associated in deciding which car to buy. Let us say we have two options, a brand new small car and a used luxury car. There are many nonfunctional requirements (NFRs) that could steer the decision maker one way or the other, such as the pleasure to drive a bigger car, or the desire to drive a brand new car. The approach to deal with these NFRs related to the car buying process is similar to the approach to deal with the NFRs previously described in Chapter 7, “Stage 3: Assess new workload” on page 55.

Let us assume that, in our example, we ran the assessment to compare these two cars against NFRs, such as maximum speed and maneuverability, and the assessment result positioned these two cars very closely to each other (they both ranked similarly against the NFRs).

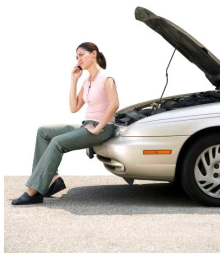
In this situation, the total cost of owning and running the chosen car is even more important. If the cars had ranked far apart with a clear *winner*, the choice is simpler, assuming you have unlimited resources. You would be willing to choose the car that ranked highest, therefore better fulfilling the NFRs. Since the resources are unlimited, the purchaser would be willing to pay more to secure the higher level of satisfaction offered by the car that best fits the purchaser needs.

But the reality is that no one has unlimited resources and a choice should be made between the two cars that, hypothetically, equally fulfill the NFRs.

Figure 11-1 shows some of the areas to consider in the car buying decision and how they correlate to the decision of buying a server.

Many Factors Affect Choice

Would you purchase a family car solely on one factor?



Car	Server Platform
Purchase price	Purchase price
Gas mileage, cost of repairs, insurance cost	Cost of operation, power consumption, floor space
Reliability	Reliability
Safety, maneuverability, visibility, vendor service	Availability, disaster recovery, vendor service
Storage capacity, number of seats, towing capacity	Scalability, throughput
Horsepower	Chip performance
Dash board layout Steering wheel location	Instrumentation and skills
Handling, comfort, features	Manageability
Looks, styling, size	Peer and industry recognition

Figure 11-1 Considerations when buying a car and an IT server

Factors affecting the decision in buying a car are not that different from the considerations required to purchase a server.

To make our case more interesting, let us assume that the initial purchase cost is the same for both cars, fifteen thousand units in the local currency (LU 15K). The key question is not changed. Which car should the purchaser choose? The purchaser could choose either one of these cars because they closely satisfy the NFRs at almost the same level.

That brings us to the concepts of total cost of acquisition (TCA) and total cost of ownership (TCO, also known as cost and value). The same cost of LU15K reflects only the cost related to the acquisition of either car. Since one car is used, bigger, and more luxurious, and the other one is new, smaller, and basic, there are differences in the cost of ownership that should be factored along the car's useful life. Figure 11-2 illustrates a few areas involved in the long-term costs.

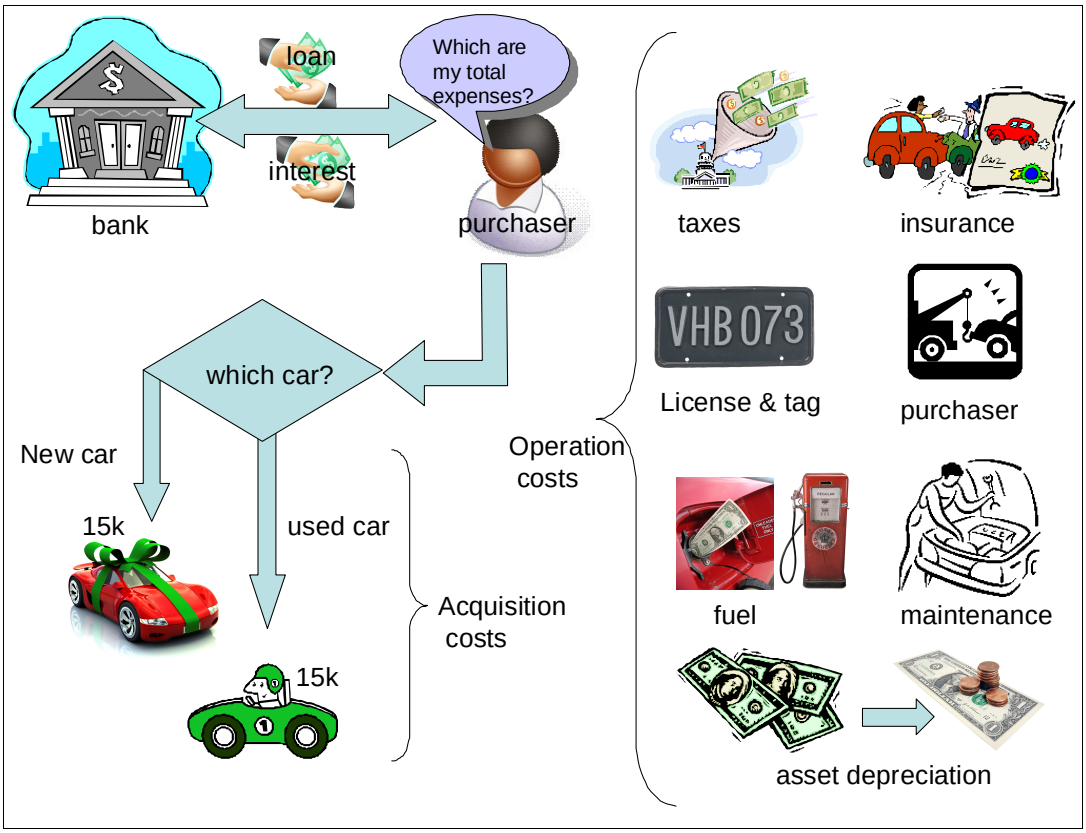


Figure 11-2 Total cost is the acquisition plus the operation costs

The purchaser in this scenario faces a situation where the acquisition costs are the same (TCA), but the total cost of ownership is likely different since the cars belong to different categories.

The TCA reflects the tip of the iceberg, and only a thoughtful TCO (cost and value) analysis will provide the proper information to the buyer. This enables an informed decision based on the total costs involved, and not just the costs of acquisition.

11.2 Objectives

This chapter presents some key TCO concepts and important considerations when you conduct a TCO study, using the car purchase experience as an example and then discussing similar concepts applied to the IT environment.

Figure 11-3 illustrates the TCO process flow. It started as a sub-product of an infrastructure solution architectural choice, where different platforms were ranked closely together and the need for differentiation triggered a TCO analysis. The next step is to define the TCO scope, in which many TCO considerations are included into the analysis. Since TCA is always a component of the TCO itself, the total acquisition costs for the infrastructure being analyzed need to be known. After that, the TCO study is performed, a report is generated, and then you can iteratively refine the process and include other relevant components. This continues until no more components are left to add nor are new iterations for refinement required.

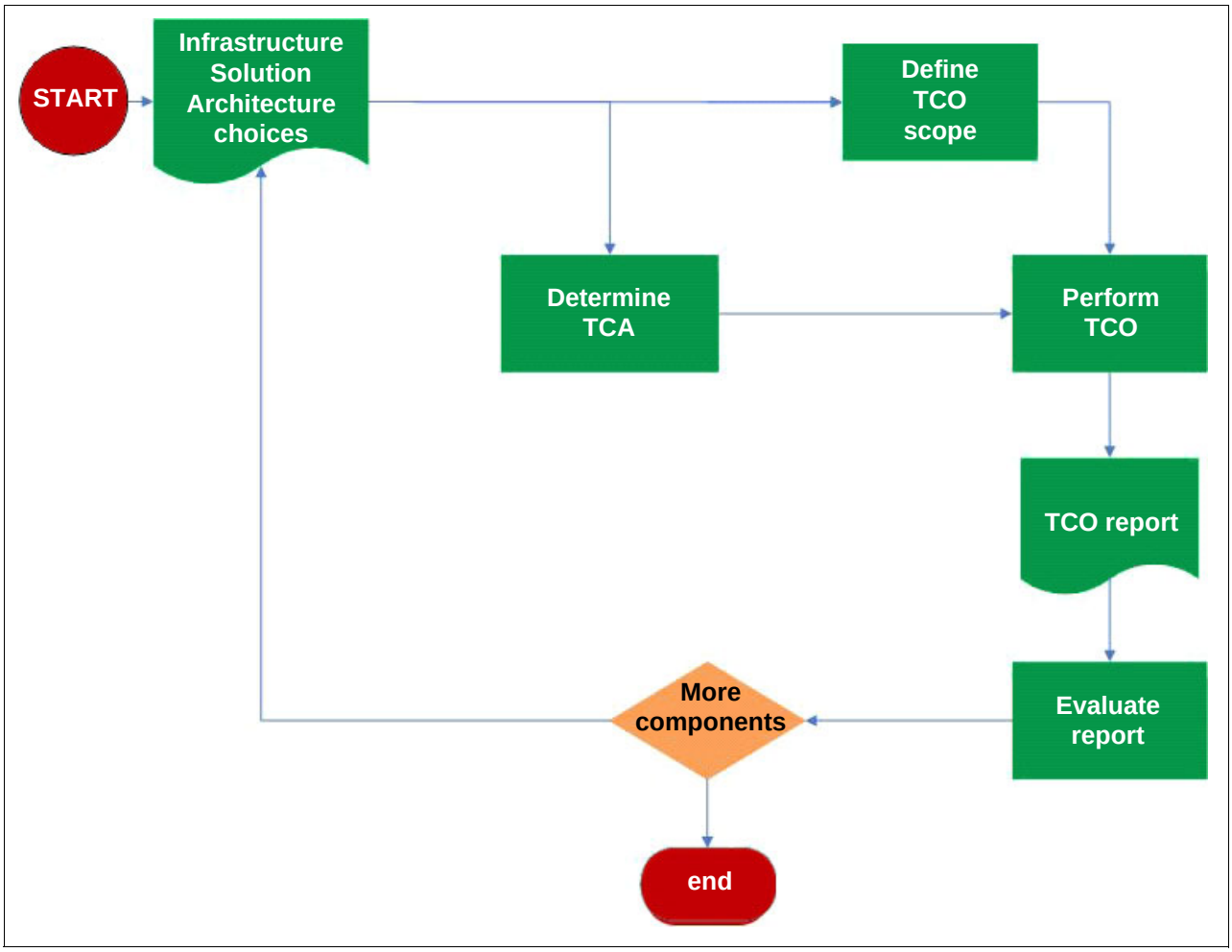


Figure 11-3 TCO process flow

11.3 Tools and accelerators

There are many TCO tools available in the market, so we leave it up to you to choose the tool that best fits your needs. The important component here is not the tool itself, but the process used to calculate the TCO. A good TCO tool is transparent to the user. The user can see the input being used and modify it according to the user's local factors, for instance, local electricity costs, local space costs, or local administrative costs. The tool should not be a "black box" where the user has access to just a few input parameters and the bulk of the assumptions are unknown.

However, a tool is not mandatory. If you apply the TCO concepts, a simple spreadsheet can be used to represent the cost components being integrated into the study, and produce the resulting tables and graphs.

The most important thing is to ensure that the TCO study, its input, and its output provide useful results to the client. The client needs to see that the study is unbiased and executed with the client's best interests in mind. If the client can see the numbers being used and can understand the results, the TCO objective is accomplished to provide the client with another component to enable the client's decision making.

11.4 Resources needed

A TCO assessment requires the following resources:

- A specific TCO methodology
- An optional tool implementing this methodology or a manual approach to perform the equivalent function
- A skilled TCO practitioner

Although, at first glance, a TCO assessment might seem simple enough to be done by any individual, the unique requirement of merging the proper technical, business, and financial skills is a determinant factor to narrow down the field to qualified professionals.

Any TCO should be preceded by a sizing or capacity planning study, as described in Chapter 10, "Stage 6: Sizing and capacity planning" on page 79. You need to know the target hardware to produce hardware acquisition cost, software license cost, and maintenance cost.

11.5 Detailed steps

Chapter 7, "Stage 3: Assess new workload" on page 55 described the process of selecting an Infrastructure Solution Architecture. We use the result of that process with the sole intent to provide us with an example for TCO discussion.

In some situations, the summary of NFR results is enough to enable the decision maker to act. However, in many situations, the cost plays a role in the selection process.

The decision maker should carefully evaluate the output of the infrastructure selection process. That might or might not be enough for the final decision, but if a decision is not reached, additional input, such as the output of a TCO analysis, might be considered.

In the scope of this chapter, our example assumes the decision maker required the TCO assessment to provide additional input for the final decision.

To provide a scenario, Figure 11-4 on page 100 shows the results of the assessment of the Cognos 10 BI component in Chapter 7, "Stage 3: Assess new workload" on page 55.

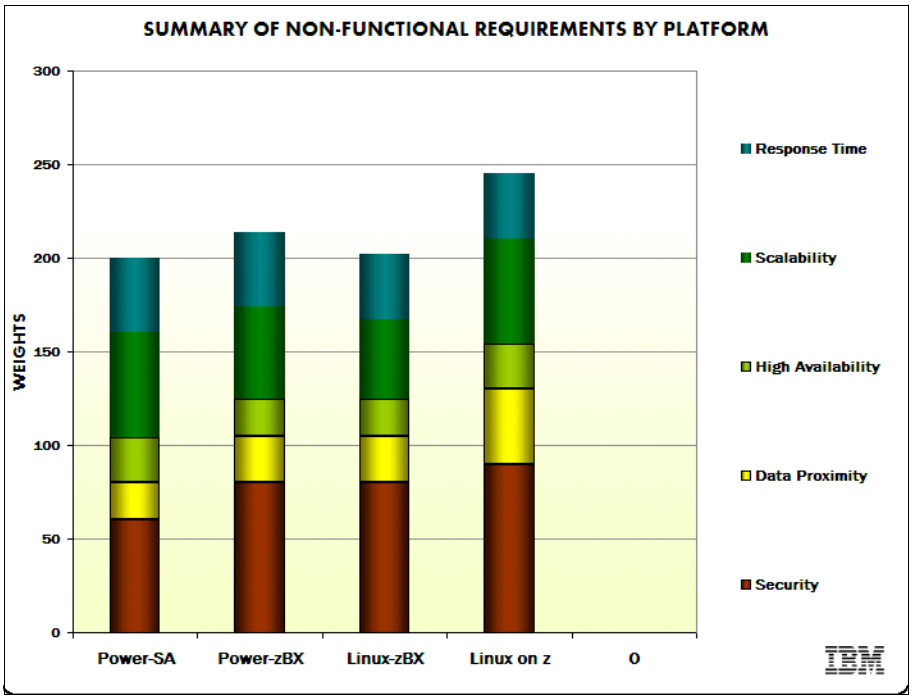


Figure 11-4 The NFRs are best met by Linux on System z

As you can see, based on the Infrastructure Solution Architecture assessment previously completed, the Linux on System z option is the solution that best fulfills the NFRs. It is also obvious that there is no clear second place solution, since all three remaining architectures are ranked the same or very close. In theory, in most instances, there would be no need to proceed with a TCO study because the “Fit for Purpose” results obtained for Linux on System z far exceed the other options. The NFRs are better satisfied with Linux on System z, making it the “best fit” platform for the application.

However, just as an example, we proceed with the TCO discussion as though we need it to enable the decision maker to best choose an option.

There are many ways and methodologies to enable a TCO assessment. In this chapter, we use the method that is most familiar to the writers. It does not mean this is the only method or necessarily the best method.

Let us define the scope of our TCO example. Figure 11-5 on page 101 gives us examples of areas to consider when creating a TCO assessment.

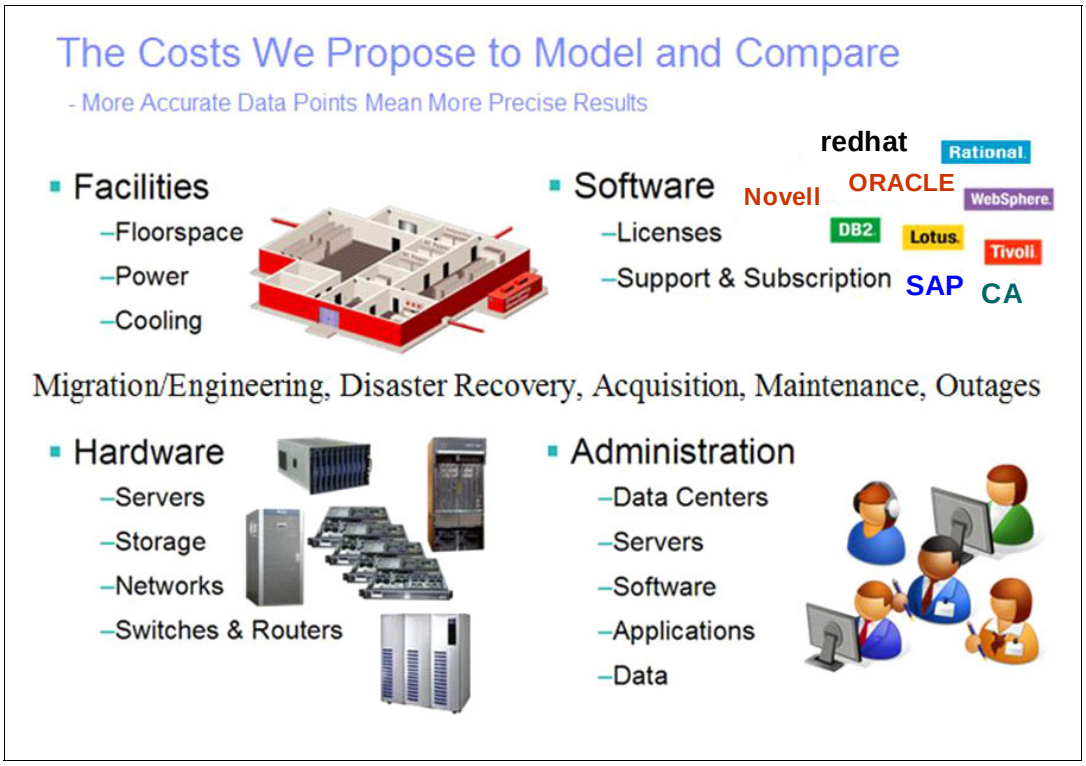


Figure 11-5 TCO data points

The methodology that we chose has a good tool to implement it. The tool is spreadsheet-based and transparent to the client. Prices are included as defaults, but any price can be overridden by the user, making the tool very flexible. The same is true for all the required input. The Cost Analysis categories included in the Right-Fitting Applications into Consolidated Environments (RACE) methodology are listed in Figure 11-6.

Cost Analysis Categories

1. Power
2. Floorspace
3. Facilities
4. Migration
5. Engineering
6. Server Acquisition
7. Server Maintenance
8. Connectivity Acquisition
9. Connectivity Maintenance
10. Disk Acquisition
11. Disk Maintenance
12. Software Licenses
13. Software Maintenance
14. Network Bandwidth
15. Systems Administration
16. Disaster Recovery Equipment Acquisition
17. Disaster Recovery Equipment Operation
18. Cost of Downtime
19. Cost of “Solution Editions”

Figure 11-6 Right-Fitting Applications into Consolidated Environments cost-analysis categories

Some of these parameters need to be calculated on the side and then entered into the tool, such as engineering or migration costs. But costs, such as server acquisition, software acquisition, and services and support (S&S) are included in the tool and default prices are provided. The TCO practitioner can and should override any default numbers that do not reflect the client's environment. It is good practice to work together with the client when validating or overriding the input, which will not only produce more accurate results, but also bring more credibility to the results that are obtained.

The tool we chose for our example allows you to select only a subset of the listed TCO categories, narrowing down the scope of the analysis. For instance, if the cost of downtime is not important to a specific client, it can be easily excluded from the analysis by simply clearing the related box.

11.5.1 Building the TCO model

The TCO model has to contain the platforms being assessed and the related pricing structure being considered in the context of Figure 11-6 on page 101, simplified to exclude the areas not being considered.

In this example, we are trying to differentiate among the four platforms that are being considered to host the target workload: Power, zEnterprise BladeCenter Extension (POWER7), zBX (Linux), and Linux on System z. The workload being considered is Cognos BI 10. We are comparing the cost of a single application, but normally the comparison is done for the cost of running multiple applications, in order to maximize the virtualization characteristics of a specific server.

The primary input to the TCO model is the total cost of acquisition (TCA). To be able to assess this cost (including hardware and software), you need to have the environment configured correctly. You need to create a required sizing or capacity plan before you can configure the environment correctly. After the TCA cost is known, you can start the comparison, but not before considering which other TCO areas to include in the analysis. A simple TCO analysis considers at least these areas:

- TCA costs (including server, software, network, and storage costs)
- Recurring costs (such as hardware and software maintenance)
- Basic installation-related costs (such as power, cooling, and real estate)

Figure 11-7 on page 103 illustrates the process of defining a candidate server being considered in a TCO study. Note the server type definition, virtualization mechanism, and other factors.

To-be machine is identified on the Targets sheet

Target Server Setup and Processing C

Directory

Case 2 System Z Long Name

Virtual Domain:	**Domain 1**
	ServerName
Server Name	ServerName
zServer Family	z196
Water Cooled Option	No
DC Power Option	No
Part of URM Ensemble (Manage/Automate/None)	None
Hypervisor Selection	z/VM
Operating System Selection	zLinux

Change defaults as necessary

System z Server(s) and LPAR(s) Quick&Dirty Configurator

Required MIPS	48430
Reduction in Required MIPS due to Re-Architecting Application	0%
Net Required MIPS	48430
Hypervisor Overhead	7.0%
Override (from VMPlanner)	
Hypervisor Overhead Effect	7.0%
Overhead Effect - MIPS	3390

Figure 11-7 Example of input fields for one platform

Figure 11-8 shows the output of an example TCO study. Many TCO areas are included in the table representing the results. The tool generates one table per platform being considered.

RACEv Summation Sheet Filename: C:\$USER\$RACE\[RACEv_3.1 20100901.xls]Summation

Directory

ase 2

	Manufacturer/Vendor or Other Descriptor		Num. Servers	1	Num. Chips	0	
	Server/Model or Other Descriptor				Num. Cores	8	
n/Off	**Long Case Identifier**	**Long Name**					
witch		**1st Year**	**2nd Year**	**3rd Year**	**4th Year**	**5th Year**	**Total**
1	Solution Edition Cost	703,354	0	0	0	0	703,354
1	Power	7,231	7,231	7,231	7,231	7,231	36,154
1	Floor Space Analysis	46,800	46,800	46,800	46,800	46,800	234,000
1	Facilities Analysis	0	0	0	0	0	0
1	Migration and Engineering/Development Costs	0	0	0	0	0	0
1	Server Acquisition Analysis	90,000	0	0	0	0	90,000
1	Connectivity Acquisition Analysis	0	0	0	0	0	0
1	Disk Acquisition Analysis	101,363	0	0	0	0	101,363
1	Annual Server Maintenance Cost Analysis	0	0	0	140,064	140,064	280,128
1	Annual Connectivity Maintenance Cost Analysis	1,295	1,295	1,295	1,295	1,295	6,477
1	Annual Disk Storage Maintenance Cost Analysis	5,068	5,068	5,068	5,068	5,068	25,341
1	Software License Cost Analysis	0	0	0	0	0	0
1	Annual Software Maintenance Cost Analysis	198,800	198,800	198,800	243,111	243,111	1,082,622
1	Annual Enterprise Network Bandwidth Costs	7,680	7,680	7,680	7,680	7,680	38,400
1	Annual Sysadmin Cost Analysis	172,025	172,025	172,025	172,025	172,025	860,125
1	Disaster Recovery Equipment Acquisition	0	0	0	0	0	0
1	Total Annual Cost of DR Equipment	0	0	0	0	0	0
1	Annual Cost of Downtime Time	0	0	0	0	0	0
	Annual	1,333,615	438,899	438,899	623,274	623,274	3,457,962
	Cumulative	1,333,615	1,772,514	2,211,414	2,834,688	3,457,962	
	Startup Costs	894,716					894,716
	Operating Costs	438,899	438,899	438,899	623,274	623,274	2,563,246

Figure 11-8 Example of a TCO study output for a specific platform

Figure 11-9 is a graphical representation in which each line shows one server platform. This example is one of the rare cases where the platform that is represented by case 1 has lower initial (TCA) costs than the platform that is represented by case 0. The costs stay lower until the end of the fifth year of the TCO study. The more common graphical representation is similar to Figure 10-5 on page 88, where one platform has a higher TCA than the other, but the final TCO costs are lower.

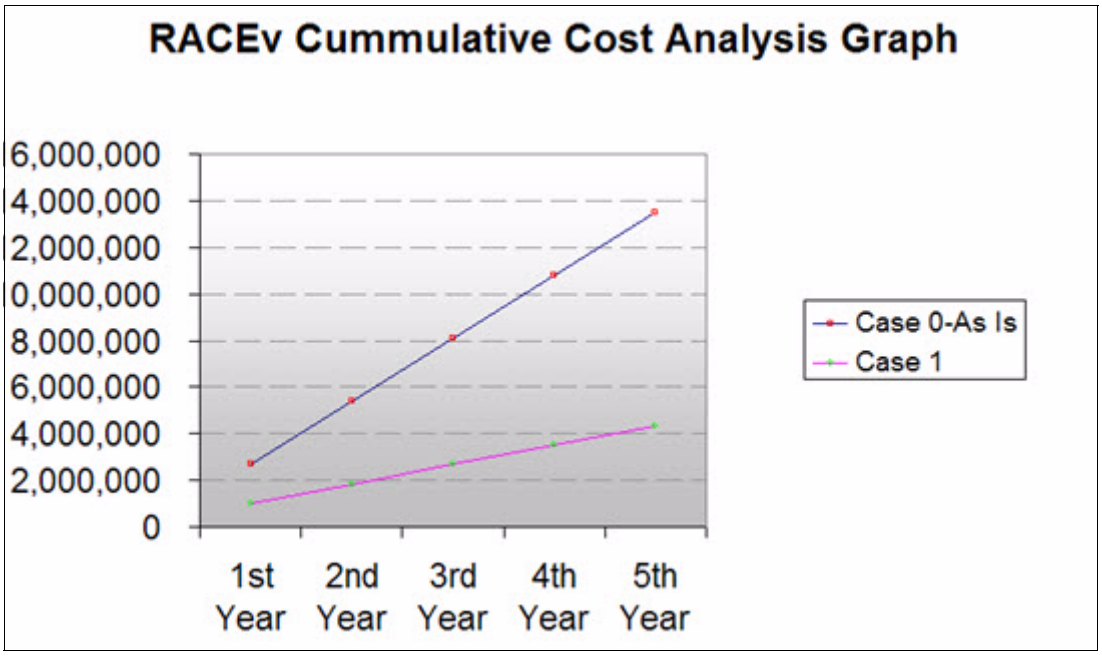


Figure 11-9 TCO numbers plotted for a five-year study

Figure 11-10 shows the cost breakdown, pinpointing the largest cost contributors. This is useful information to enable a more detailed analysis and guide the decision maker. By knowing the areas that are producing the higher costs, you can look again into the NFRs and see whether areas that were deemed required in an earlier assessment are still looked at the same way after the costs are known.

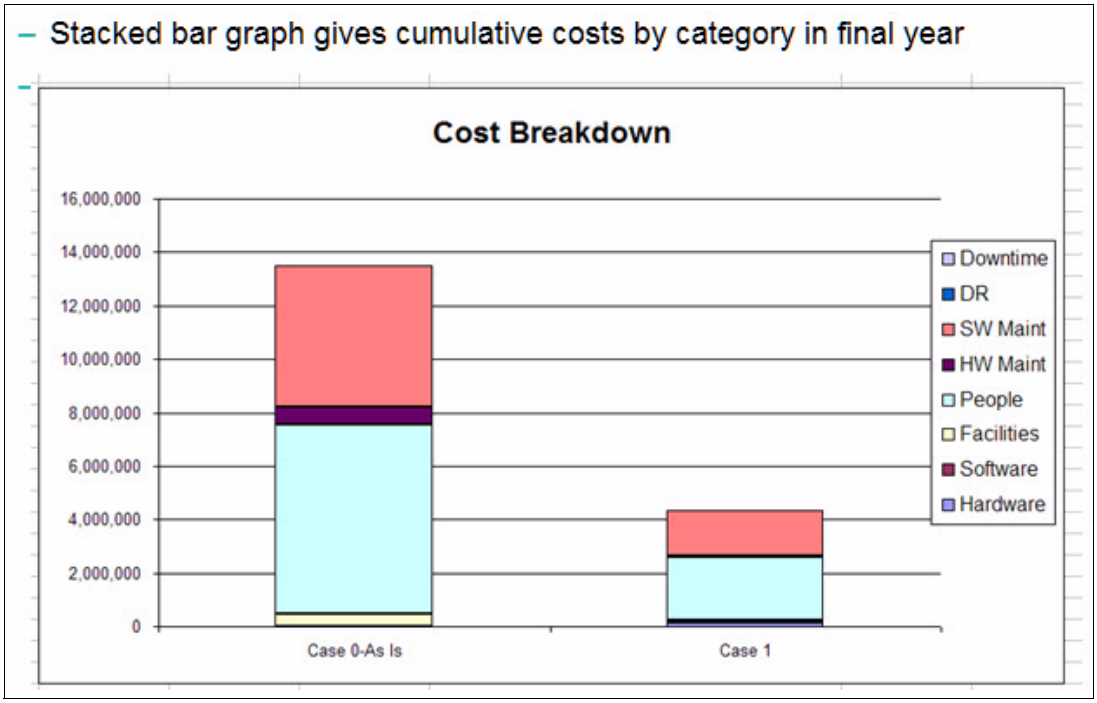


Figure 11-10 TCO breakdown per category

11.6 Conclusions

The intent of this chapter was not to present a full-blown TCO study, but to introduce you to the basic TCO drivers and concept. Not all situations will require a TCO study, although many will. The ability to express the value of any platform in economic terms is always a plus and extremely important to the C*xx* (CIO, CFO, or CMO) office. Nowadays, decision makers are not purely constrained to the IT side of any company. A cross-enterprise view offered by the C*xx* decision makers is not solely based on functionality or localized needs, but on dollars and “sense” and long-term strategic goals.

To be able to speak the same language as these decision makers, you need to use the TCO analysis process. That would enrich any proposal by combining not just the technical components but also the business component into the mix.

Many methodologies and tools are available, but the bottom line is that the best TCO model is the simplest model that addresses the client’s concerns. There is no value in complicating any TCO model by adding unnecessary areas to it.

Also, we discussed that the TCO results need to be validated by the client. The client needs to see their own operational costs on the final results. This implies that the default costs provided by any tool or used by any methodology should be customized, not only for the client’s specific environment, but also for regions and countries.

12

Stage 8: Create business case

After all the previous stages have been completed, and all data has been collected and analyzed, there must be enough information available to make a business decision on whether to proceed to the implementation phase. There are only a few key elements that should determine the decision:

- The projected total cost of ownership (TCO) to operate the new solution
- The cost and risk of migration, in the case of an existing workload
- The extent to which the business requirements are met or superseded and problem points are eliminated

In certain cases, other elements may be part of the equation, such as the total cost of acquisition (TCA) or strategic directions or factors from outside, such as compliance to new regulations. In this chapter, we discuss these aspects in more detail.

12.1 Objectives

The objective of this stage is to be able to make the business case and a decision on whether to proceed with implementation of the application or workload on zEnterprise.

12.2 Inputs

This stage has these inputs:

- TCO analysis from Chapter 11, “Stage 7: Execute total cost of ownership (TCO) analysis” on page 95
- Assessed workload from Chapter 7, “Stage 3: Assess new workload” on page 55
- Prioritized infrastructure solution architecture from Chapter 9, “Stage 5: Prioritize solution architecture alternatives” on page 71

12.3 Outputs

The output of this stage is a documented and informed decision made on the business case to move forward.

12.4 Tasks

Figure 12-1 on page 108 shows the tasks and decisions in this stage.

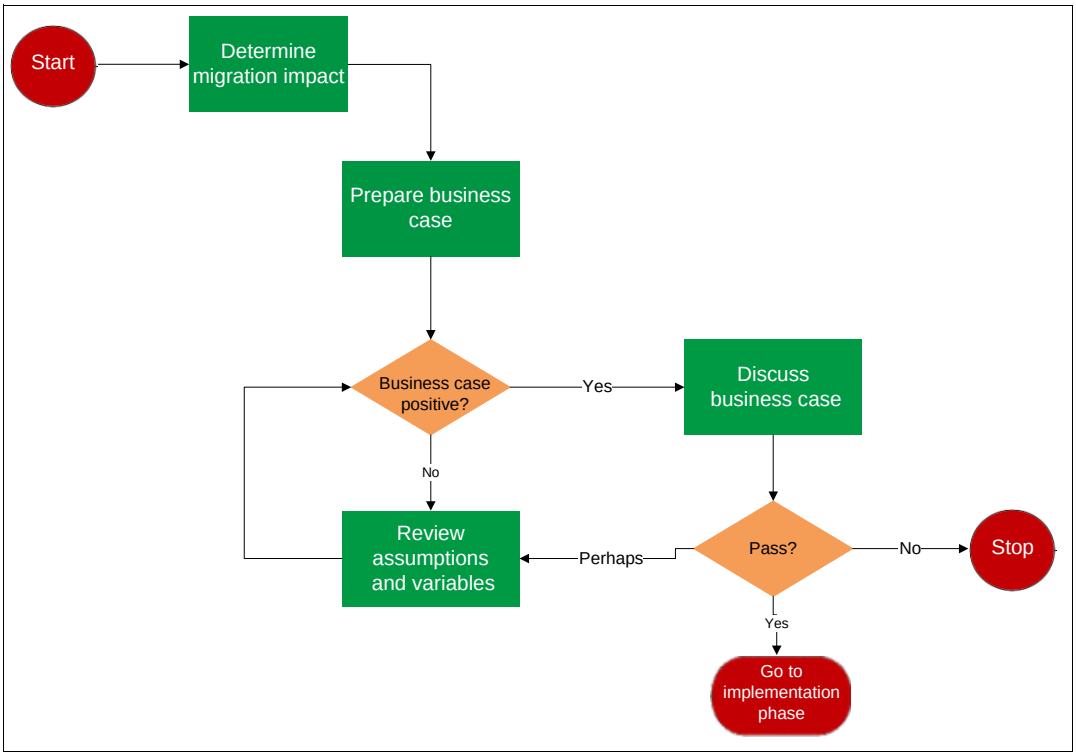


Figure 12-1 Chapter 12, “Stage 8: Create business case” on page 107: Tasks

12.4.1 Determine migration impact

This task only applies to the situation where the workload is an existing workload currently running on another platform. In that case, an assessment will need to be made of the migration impact in terms of cost and risk. The assessment will need to include these components:

- A high-level migration plan, with timelines, dependencies, and necessary resources
- A cost estimate of the migration
- A risk assessment

The migration impact and cost can vary greatly. On one extreme, the migration can be costly, take a long time, and be high risk. On the other extreme, it can be low risk and quick. The risk of the migration increases because of factors, such as the use of platform-specific technologies, such as programming languages and middleware. The risk also increases because of dependencies of the application or workload in scope on other workloads and applications.

For an elaborate discussion of migration scenarios, see *Considerations for Transitioning Highly Available Applications to System z*, SG24-7824.

How this task applied to our Customer Profitability Analytics (CPA) workload

Our *Customer Profitability Analytics* (CPA) workload is a newly designed workload to address specific business needs, so there was no migration or transition task involved. See Chapter 3, “Overview of the banking analytical workload” on page 19 for the business value of this workload.

12.4.2 Prepare business case

The business case must be well-documented and contain the following information:

- An executive summary: This is typically not more than one page with a summary of the business benefits of deployment on zEnterprise and a cost overview, including TCO and migration cost.
- In the case of an existing workload or application, a migration plan that includes this information:
  - Timelines
  - Resources needed
  - Dependencies
  - Cost of the migration
  - Cutover scenarios and fallback plans
- A projection of the total cost of operation of the solution on zEnterprise. See Chapter 11, “Stage 7: Execute total cost of ownership (TCO) analysis” on page 95 for a discussion of how TCO is calculated. A good projection window is five years, but fewer or more can be chosen, as well.
- Return on investment (ROI), putting the cost of migration in perspective to the savings achieved.

- A description of how the business benefits from deploying the solution on zEnterprise. Use these hints for describing these benefits:
  - Describe how the problem points found in Chapter 4, “zEnterprise hybrid workload placement approach” on page 31 are eliminated or significantly reduced with the new solution.
  - Describe other advantages, not currently experienced as problem points, that may come with the new solution as quality enhancements, such as more agility, better time to market, or cloud-readiness.

12.4.3 Present and discuss the business case

After the business is documented, it must be presented and discussed. The team involved must consist of the following roles:

- Sponsors
- Subject matter experts (SMEs) in the following areas:
  - Finance, with knowledge of TCO calculations, depreciation schemes, financing methods and offerings, licensing, and so on
  - Solution architecture
- Decision-making executives

The outcome of this step can be one of these alternatives:

- “Yes”, meaning arrangements can be made to proceed to the implementation stage.
- “No”, resulting in stopping, holding, restarting, or delaying the project with other parameters.
- “Perhaps”, making it worthwhile to review the entire business case and all of its input again to see how it can be optimized. See “Review assumptions and variables”.

12.4.4 Review assumptions and variables

Perhaps the business case does not pass the first time. It may provide insufficient benefits to move forward. In that case, it is worthwhile to review all parameters and variables again and investigate how the business case can be further optimized. Especially, from a financial perspective, a lot of assumptions are typically made, such as depreciation schedules and acquisition costs of equipment. Sometimes, even external factors, such as estimates of future currency conversion rates and energy prices, could influence the business case.

12.5 Tools and accelerators

There are no specific tools used for this step, but it is key to have a good writer on board and ensure that the business case looks professional.

12.6 Summary

We hope this description of our Smart Analytics assessment methodology, including IBM’s Fit for Purpose (F4P) activities, is helpful. The resulting zEnterprise system is described in *Using zEnterprise for Smart Analytics: Volume 2 Implementation*, SG24-8008.

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

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.

- *Using zEnterprise for Smart Analytics: Volume 2 Implementation*, SG24-8008
- *IBM zEnterprise Unified Resource Manager,* SG24-7921
- *Using Rational Performance Tester Version 7,* SG24-7391
- *IBM zEnterprise 196 Technical Guide,* SG24-7833
- *z/VM and Linux on IBM System z: The Virtualization Cookbook for SLES 10 SP2*, SG24-7493
- *Considerations for Transitioning High Available Applications to System z*, SG24-7824
- *OS/390 MVS Parallel Sysplex Capacity Planning,* SG24-4680
- *HiperSockets Implementation Guide*, SG24-6816

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:

- *Coordinator of Processes Service Developer's Guide*
- *OMEGAMON XE for Mainframe Networks Planning Configurations*, SC27-2384
- *zEnterprise System Ensemble Performance Management Guide,* GC27-2607

Online resources

These websites are also relevant as further information sources:

- IBM Global Techline Center of Excellence

  `http://w3-03.ibm.com/support/techline/global/index.html/`
- SPSS

  `http://www.ibm.com/software/analytics/spss/`
- zEnterprise

http://www.ibm.com/systems/z/hardware/zenterprise/zbx.html/

- IBM DB2 Analytics Accelerator

  http://www.ibm.com/software/data/db2/z/OS/analytics-accelerator/

- Rational Performance Tester

  http://www.ibm.com/software/awdtools/tester/performance/

- *Linux on System z, an end-to-end view: From Sizing to Performance Analysis to Capacity Planning*

  http://www-03.ibm.com/support/techdocs/atsmastr.nsf/84279f6ed9ffde6f86256ccf00653ad3/1c76a8c5aa63c3b28625723a0026816b/$FILE/Linux%20on%20System%20z,%20an%20end%20to%20end%20view.pdf

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

IBM Redbooks

Using zEnterprise for Smart Analytics: Volume 1 Assessment

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

Using zEnterprise for Smart Analytics

Volume 1 Assessment

Evaluating the requirements of a Smarter Analytics workload

Using the IBM Fit for Purpose methodology

Deploying the workload on a zEnterprise system

This IBM Redbooks publication is the first of two volumes explaining the approach and implementation of a new hybrid workload integrated in the existing IBM Smarter Banking showcase, which is hosted at IBM in Montpellier, France. The volumes contain the following information.

Volume 1 Assessment describes how to evaluate the requirements of a new IBM Smarter Analytics workload. It addresses the user, system resources, and data processing profiles to identify the optimal configuration using IBM methodologies, such as Fit for Purpose (F4P). Because the existing showcase is based on IBM zEnterprise, the deployment options include IBM z/OS, Linux on IBM System z, IBM AIX running on POWER7 blades within the zEnterprise BladeCenter Extension (zBX), and Microsoft Windows 2008 Server running on Intel blades also within zBX.

Volume 2 Implementation describes the steps involved in deploying the Smarter Analytics workload in the showcase. With multiple components, including IBM Cognos 10.1 Business Intelligence, IBM Cognos TM1, Cognos Metrics Studio, IBM SPSS, IBM DB2 for z/OS, and many application design tools, the workload spans multiple operating environments. Application clustering, setting up performance policies using Unified Resource Manager, and simulation test execution results are included.

This book is intended for an audience of professionals in an infrastructure architecture role. However, architects with a focus on business intelligence and analytics, as well as IT managers, will find value in reading this book. If you want to see this solution in action, contact your IBM representative or send an email directly to mopbcoe@fr.ibm.com.

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

SG24-8007-00 ISBN 0738438251