

Compliments of **IBM**

IBM Limited Edition

Big Data Analytics Infrastructure

FOR
DUMMIES[®]
A Wiley Brand

Learn

- How Big Data Analytics is critical for competitive advantage
- Why infrastructure matters for Big Data Analytics
- The considerations for choosing a Big Data Analytics infrastructure

Barry Schoenborn



Big Data Analytics Infrastructure

FOR
DUMMIES®
A Wiley Brand

IBM Limited Edition

by Barry Schoenborn

FOR
DUMMIES®
A Wiley Brand

Big Data Analytics Infrastructure For Dummies®, IBM Limited Edition

Published by
John Wiley & Sons, Inc.
111 River St.
Hoboken, NJ 07030-5774
www.wiley.com

Copyright © 2014 by John Wiley & Sons, Inc.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the Publisher. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permissions>.

Trademarks: Wiley, For Dummies, the Dummies Man logo, The Dummies Way, Dummies.com, Making Everything Easier, and related trade dress are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates in the United States and other countries, and may not be used without written permission. IBM and the IBM logo are registered trademarks of International Business Machines Corporation. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc., is not associated with any product or vendor mentioned in this book.

LIMIT OF LIABILITY/DISCLAIMER OF WARRANTY: THE PUBLISHER AND THE AUTHOR MAKE NO REPRESENTATIONS OR WARRANTIES WITH RESPECT TO THE ACCURACY OR COMPLETENESS OF THE CONTENTS OF THIS WORK AND SPECIFICALLY DISCLAIM ALL WARRANTIES, INCLUDING WITHOUT LIMITATION WARRANTIES OF FITNESS FOR A PARTICULAR PURPOSE. NO WARRANTY MAY BE CREATED OR EXTENDED BY SALES OR PROMOTIONAL MATERIALS. THE ADVICE AND STRATEGIES CONTAINED HEREIN MAY NOT BE SUITABLE FOR EVERY SITUATION. THIS WORK IS SOLD WITH THE UNDERSTANDING THAT THE PUBLISHER IS NOT ENGAGED IN RENDERING LEGAL, ACCOUNTING, OR OTHER PROFESSIONAL SERVICES. IF PROFESSIONAL ASSISTANCE IS REQUIRED, THE SERVICES OF A COMPETENT PROFESSIONAL PERSON SHOULD BE SOUGHT. NEITHER THE PUBLISHER NOR THE AUTHOR SHALL BE LIABLE FOR DAMAGES ARISING HEREFROM. THE FACT THAT AN ORGANIZATION OR WEBSITE IS REFERRED TO IN THIS WORK AS A CITATION AND/OR A POTENTIAL SOURCE OF FURTHER INFORMATION DOES NOT MEAN THAT THE AUTHOR OR THE PUBLISHER ENDORSES THE INFORMATION THE ORGANIZATION OR WEBSITE MAY PROVIDE OR RECOMMENDATIONS IT MAY MAKE. FURTHER, READERS SHOULD BE AWARE THAT INTERNET WEBSITES LISTED IN THIS WORK MAY HAVE CHANGED OR DISAPPEARED BETWEEN WHEN THIS WORK WAS WRITTEN AND WHEN IT IS READ.

For general information on our other products and services, or how to create a custom *For Dummies* book for your business or organization, please contact our Business Development Department in the U.S. at 877-409-4177, contact info@dummies.biz, or visit www.wiley.com/go/custompub. For information about licensing the *For Dummies* brand for products or services, contact BrandedRights&Licenses@Wiley.com.

ISBN: 978-1-118-92136-4 (pbk); ISBN: 978-1-118-92311-5 (ebk)

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

Publisher's Acknowledgments

We're proud of this book and of the people who worked on it. For details on how to create a custom *For Dummies* book for your business or organization, contact info@dummies.biz or visit www.wiley.com/go/custompub. For details on licensing the *For Dummies* brand for products or services, contact BrandedRights&Licenses@Wiley.com.

Some of the people who helped bring this book to market include the following:

Project Editor: Carrie A. Johnson
Acquisitions Editor: Connie Santisteban
Editorial Manager: Rev Mengle
Business Development Representative:
Sue Blessing

Custom Publishing Project Specialist:
Michael Sullivan
Production Coordinator: Melissa Cossell
Special Help: Rick Perret and
Herbert Schultz

Table of Contents

Introduction	1
About This Book	2
Icons Used in This Book.....	2
Beyond the Book.....	3
Chapter 1: Getting to Know Big Data and Analytics . . .	5
Looking at Big Data and Analytics	6
Volume	6
Velocity	7
Variety.....	7
Examining Types of Analytics	8
Understanding the Business Need for BD&A	8
Competitive advantage	9
Enhanced ROI.....	10
Improved customer experience.....	10
Introducing BD&A Infrastructure.....	10
Promoting success	11
An approach you can follow.....	12
Chapter 2: Looking at Infrastructure for Big Data and Analytics	15
Looking at a BD&A Hardware Infrastructure.....	15
Customer business goals.....	16
Needs analysis.....	16
Introducing the Infrastructure Components of BD&A	16
Scalability of infrastructure.....	17
Parallel processing	17
Low latency resources	18
Data optimization	18
Understanding the Aspects of a Well-Architected Infrastructure.....	20
Speed/performance	20
Availability.....	20
Access	21
Chapter 3: Looking at IBM Infrastructure Choices . . .	23
IBM Servers.....	23
IBM Power Systems	24
IBM System z	25

IBM PureSystems	25
IBM System x	26
IBM Storage.....	26
Looking at storage basics	27
IBM FlashSystem.....	27
IBM System Software	29
GPFS.....	29
Platform Computing	29

Chapter 4: Solving Big Problems with Powerful Solutions.31

Energy and Utilities.....	31
Business Need	32
Solution	32
Fashion	32
Travel.....	33
Finding the solution.....	34
Looking at the benefits.....	34
Responding to customer behavior	34
Healthcare.....	35
Insurance.....	35
Overview	36
Solution	36
Travel and Transportation	36
Business need	37
Solution	37

Chapter 5: Ten Questions LoB and IT Execs Should Ask Each Other39

IT to LoB: Are You Aware of the Benefits of BD&A Infrastructure?	40
IT to LoB: Are You Aware of BD&A's Return On Investment?	40
LoB to IT: How Can I Benefit from Bringing the Back Office to the Front Office?.....	41
LoB to IT: Do I Need to Be On the Cloud?	41
LoB to IT: Should I Worry about Obsolescence?	41
IT to LoB: Where and When Would Performance Issues Occur?.....	42
IT to LoB: Are There QoS or SLA Concerns?	42
IT to LoB: What Security Requirements Exist?	42
IT to LoB: Does the BD&A Application Require Mission-Critical Resiliency?	42
IT to LoB: How Will the BD&A Application Expand?	43

Introduction



The current form of analytics frequently involves an end-user using a spreadsheet application. The trouble is that information is often slow coming in and may not contain all relevant data, and the result is unacceptable delays in getting actionable information. That's okay, but it probably won't get you where you want to go.

With Big Data and Analytics (BD&A), you get a new approach to gathering, processing, and understanding information. You may hear the term a lot, but many people still don't fully know what it is. Very broadly, BD&A is a combined hardware/software architecture that gathers vast quantities of disparate data for fast analysis. The analysis produces meaningful information. The result is that you can make faster and better-informed business decisions. The benefit is competitive advantage and (hopefully) increased profitability for your operation.

And don't forget infrastructure. There can be no doubt that infrastructure matters. Take, for example, the many structural failures throughout history: The Tacoma Narrows Bridge ("Galloping Gertie") in 1940 had poor girder design that caused the bridge to buckle and fail under moderate winds. Then there is the Leaning Tower of Pisa. Its builders failed to address infrastructure, and that's why the tower leans.

Fortunately, in the world of information technology, it's relatively easy to plan and implement appropriate infrastructure for BD&A. BD&A solutions produce positive, valuable outcomes. They need to address three things: securing access to all relevant data to get the best insights; making analytics a part of everyday operations and business transactions; and ensuring all employees can benefit from these insights generated.

About This Book

BD&A has several components: hardware, software, and services. In this book, I emphasize hardware infrastructure — processing, storage, systems software, and internal networks. The reason for talking about infrastructure is that it's the foundation of a healthy BD&A solution. You can't build anything worthwhile without a good infrastructure.

If you're reading this book, I assume that you've heard about BD&A but want to understand a little more about what it is, how it works, and how your company can benefit from it. This book is meant for executives — decision makers — and senior IT practitioners who want to put the concepts in front of them.

The book talks about companies of different sizes in different industries. Almost any business can benefit from a BD&A implementation. Some people think that BD&A is only for bigger companies, but smaller companies become big companies by finding and exploiting the kind of competitive edge that BD&A can provide.

Icons Used in This Book

You'll find several icons in the margins of this book. Here's what they mean.



A tip is a suggestion or a recommendation. It usually points out a quick and easy way to get things done or provides a handy piece of extra information.



A warning alerts you to conditions that require extra care and thinking. For example, you don't want to omit critical steps in evaluating your needs and planning your implementation.



Anything that has a Remember icon is something that you want to keep in mind.



Technical Stuff contains information that's interesting and useful but not vital to understanding the topic. Info here may include a brief history of a principle, the earliest practitioners, or the origin of a word. It also showcases technical points. You can either read these or skip over them.

Beyond the Book

You can find additional information about BD&A (and about IBM's approach to it) by visiting the following websites:

- ✓ **General information:** www.ibm.com/big-data
- ✓ **IBM BD&A Infrastructure:** [www.ibm.com/
big-data-infrastructure](http://www.ibm.com/big-data-infrastructure)

Chapter 1

Getting to Know Big Data and Analytics

.....

In This Chapter

- ▶ Getting a perspective on Big Data and Analytics
 - ▶ Seeing the business need for Big Data and Analytics
 - ▶ Introducing Big Data and Analytics infrastructure
-

Whether you're a Line of Business (LoB) executive or a senior IT manager, you've likely been exposed to a great number of "new" new things. Analytics isn't new; analysis has been a component of planning for millennia. What has always hampered civilization's enterprises was insufficient data and slow speed of getting it — but the need for intelligence and planning goes back a long way.

What *is* relatively new is Big Data. It's not an exaggeration to say that data is now one of the world's most valuable resources. And there's a lot of it. For example, consider the Internet. During the Christmas 2013 shopping season, Amazon sold 426 items *per second* — each transaction yielding a wealth of data. But big companies like Amazon aren't the only ones getting benefits from Big Data and Analytics (BD&A). Smaller companies without a big Internet presence have plenty of useful data to analyze, too. There's gold in BD&A if you're willing and able to mine for it. Check out Chapter 4 for examples of smaller companies making a big impact.

Looking at Big Data and Analytics

Executives often say, “If only I had known *this* (substitute your own situation) and if only I had known it *sooner*, I would have made a smarter decision.” That’s the problem BD&A addresses. It’s the mechanism that lets you know more, and sooner, about a lot of data. And, as a bonus, you can get information you need from *disparate* and *unstructured* data.



Remember the three Vs of BD&A: volume, velocity, and variety.

Volume

The first attribute of Big Data is *volume*. Big Data projects tend to imply terabytes to petabytes of information. However, some smaller industries and organizations are likely to deal with mere gigabytes or terabytes of data.

Nevertheless, smaller organizations may still require intense and complex analytics, based on less data. For example, a manufacturer may have only a few retail outlets, but getting and storing reliable sales data still is important. The same manufacturer may have a simple supply chain, again requiring acquisition and storage of reliable supplier data. Plus, data is stored plenty of other places that don’t include your traditional computing system or your hard. Together this data helps to paint a picture of what’s happening in the organization, and what may happen in the future.



Terabytes and petabytes, frankly, aren’t a big deal anymore. At the low end, a 1 terabyte external disk drive fits in a shirt pocket. A 4-terabyte external drive has dimensions of 5.5” x 1.9” x 6.7”. At the high end of the spectrum, ebay.com stores almost 90 petabytes of data, which requires a good deal of storage hardware.

Velocity

The second attribute of Big Data is *velocity* — the speed at which information arrives, is analyzed, and is delivered. The velocity of data moving through the systems of an organization varies from batch integration and loading of data at predetermined intervals to real-time streaming of data. The former can be seen in traditional data warehousing. The latter is in the world of technologies such as complex event processing (CEP), rules engines, text analytics, inferencing, and machine learning.



Real-time or right-time isn't just for air traffic controllers. It's about gathering data and insight as you run your business. The moment can happen when a customer calls and you capture notes from the conversation. It can be click streams from your website, hourly sales data, or anything that relates to the speed at which you operate your business.

Variety

The third attribute of Big Data is *variety*. In the past, enterprises had only to deal with a manageable number of data sources. Times have changed. Today's business environment includes not only more data but also more *types* of data than ever before. *Disparate data* is data from a variety of data sources and in a variety of formats, and is a major challenge that business analytics and Big Data projects must contend with.



You can realize two kinds of cost-saving efficiency enhancements:

- ✓ **Operational efficiency:** These gains are measured by reduced costs to get the same or better results. With BD&A, this is due to more efficient methods of data integration, management, analysis, and delivery.
- ✓ **Business processes:** These enhancements are measured by the impact of new, better ways of conducting business, including improvements to commercial transactions, sustainable management of communities, and appropriate distribution of social, healthcare, and educational services.

Examining Types of Analytics

After you get the data, you need to perform analytics to extract its value. There are four principal types of analytics:

- ✓ **Descriptive:** Uses basic statistics or cool visualization to characterize a set of data. The results may show averages, totals, frequencies, and perhaps a causal relationship. The vast majority of analytics done today falls in this category.
- ✓ **Predictive:** Helps you see what the future may hold. You use statistical models to forecast a condition such as revenue, profits, market-share, or operational outcome. Predictive analytics is based on a modeled relationship between a set of independent variables. Companies deploy this type of analytics in different ways, but it's most often used for planning.
- ✓ **Prescriptive:** Takes predictive analytics (see the previous bullet item) to a new level by optimizing the best outcome of a future prediction. It takes into account new inputs or constraints that are specific to a given situation. It may include the next best action of changing your credit or pricing policies or upselling/cross-selling strategies.
- ✓ **Cognitive:** Uses techniques and a high-performance infrastructure to extract relationship from diverse sets of data.

Understanding the Business Need for BD&A



An excellent technique for beating the competition and prospering is reducing the time gap between awareness of a condition (the trigger event) and action on it. This is called *time to insight*. Table 1-1 is a simple comparison of traditional time to insight versus BD&A time to insight.

What remains after review of the information is making a decision and taking action.

Table 1-1 Traditional versus BD&A Time to Insight

<i>Timeline</i>	<i>Days (without BD&A)</i>	<i>Days (with BD&A)</i>
Event	0	0
Time to get data	14	1
Time to analyze	7	1
Time to insight	21	2

For convenience, the times in the table are expressed in days. IBM can cite three much faster real-world examples:

- ✓ A global consumer packaged goods company reduced real-time reporting from 6 minutes to 736 microseconds.
- ✓ A manufacturing company reduced customer reporting from days to seconds.
- ✓ A financial company reduced cross-sell model calculations from 45 minutes to 5 seconds.

BD&A software, optimized on a well-tuned BD&A infrastructure, provides three results of improved time to insight: competitive advantage, enhanced return on investment (ROI), and improved customer experience.

Competitive advantage

Competitive advantage occurs when an enterprise acquires or develops attributes or processes that allow it to outperform its competitors. It's easy to define, but more difficult to get. Different executives have their favorite items that give competitive advantage. Broadly, they include one or more of the following:

- ✓ Acquire, grow, and retain customers
- ✓ Create new business models
- ✓ Transform financial processes
- ✓ Optimize operations and reduce fraud
- ✓ Improve IT
- ✓ Manage risk

Enhanced ROI

A universal business goal is increasing ROI — also known as “getting more bang for the buck.” There are many ways to measure ROI, including

- ✔ More profit through increased sales
- ✔ More profit through stable sales with less excess inventory
- ✔ More profit derived from reduced operating costs

BD&A helps enhance ROI by putting more well-analyzed data in the hands of decision makers at a fast rate.

Improved customer experience

With products becoming commoditized, price differentiation is no longer sustainable for competitive advantage. Look to improved customer experience to get an edge. *Customer experience* is the sum of all experiences a customer has with a supplier of goods and/or services, over the duration of their relationship. You could also call this customer satisfaction, and to a large degree you can manage and improve it.

In the past, brand loyalty could get you through, but that’s not true anymore. That’s because your customer base’s choices have widened, largely through the Internet.

You can use BD&A to learn what customers want and (therefore) enhance their attaining it. You have to deliver on customer wants and needs. Customer experience improves by your knowing what attracts customers (products, interactions, purchasing, warranties, “rewards,” and so forth).

Introducing BD&A Infrastructure

Broadly speaking, *infrastructure* is a combination of hardware, software, and services. The foundation is processing and storage hardware because the software has to live somewhere. The combination ends with services, as the infrastructure must stay robust, supportive, and scalable.

But first, here's a difficult reality: It's a certainty that your current systems can't keep pace with the growth of data and neither can your IT budget. The volume of data grows fast, and the parts needed to manage it grow increasingly complex.



To be a top-performing organization, you need to take a different approach to architecture, tools, and practices. To outperform in your industry, try the following:

- ✓ Use analytics to improve the core competitiveness of your business.
- ✓ Make speed a differentiator to reduce delay in decisions and business processes.
- ✓ Monetize the data itself, creating new products and services.
- ✓ Be more right, more often. A BD&A approach creates more context — and more confidence.

Promoting success

The key items needed for success are managing risk, promoting agility, acting strategically, and practicing forward thinking. These are covered in this section.

How to manage risk

To manage risk, you need to

- ✓ Guard against poor decision making by sourcing the right data. Source it at the right time, and integrate it.
- ✓ Protect against security and privacy risks, both internal and external.
- ✓ Get the risk-opportunity equation right through proactive management of your data. That includes regulatory risk and organizational risk.

How to create IT agility

To create IT agility

- ✓ Build agility into your operation through optimized systems and the cloud. Doing so should relieve pressure on your IT infrastructure.

- ✓ Adopt a new approach to the onslaught of data. There are techniques for moving the analytics closer to the data and analyzing in motion.
- ✓ Eliminate hidden costs by not starting piecemeal. Instead, build from an orchestrated roadmap.

How to act strategically

Develop a curiosity-driven and evidence-inspired workforce. It's possible to infuse analytics into everything employees touch.



Follow these tips:

- ✓ Start with people. Encourage discovery and insight.
- ✓ Put analytics into key business processes.
- ✓ Deploy the full range of analytics: prescriptive, predictive, and cognitive. See the section “Examining Types of Analytics” in this chapter.

How to adopt a forward-thinking approach

Forward thinking is an important element to introduce into corporate culture. Proactive is better than retroactive. A solid BD&A infrastructure helps that happen. Check out these ideas:

- ✓ Use thinking approaches that will maximize impact while balancing risk.
- ✓ Develop a shared set of trusted information to view the business as a whole.
- ✓ Set the tone around privacy and policies for protecting data.
- ✓ Enable risk-aware decision making and model exposure and understand variability of outcomes.

An approach you can follow

There's a three-step procedure you can follow to get your infrastructure in shape.

Master plan

Build against a master plan. Plan for all types of data and all types of analytics, with an efficient IT Infrastructure as a given. Collaborate. Cultivate new partnerships and roles. IT and LoB executives must join to understand and develop the master plan.

Infrastructure plan

Build an infrastructure designed to enable new levels of insights derived from exploiting all relevant data. The platform should be fluent in all forms of data and analytics: transactional data, Hadoop data, and so forth.

Build a highly flexible, scalable IT infrastructure, tuned for today's big data environment and designed to capitalize on the integration of social, mobile, and cloud technologies.

New approach

Take a new approach in thinking about the infrastructure for BD&A:

- ✓ Enable shared, security-rich access to trustworthy information while addressing ever-increasing data volumes, greater variety, and increased scale.
- ✓ Build intelligence into operational transactions through fast analytics and optimization of the solution stack for different analytics workloads.
- ✓ Maximize right-time availability and insights to address more users and more concurrency, changing demand and resiliency at the point of impact.

The questions you must ask are

- ✓ Is my infrastructure architected for today's challenges?
- ✓ Is it efficient and agile enough to support the transition ahead?
- ✓ Will it meet the rising expectations of my customers, employees, and business partners?
- ✓ Am I shifting my investment from maintaining my IT infrastructure to pursuing new projects?

If you were unable to answer this last question with confidence, you aren't alone. IBM research shows that just one in five IBM clients are able to answer this question affirmatively. But the CIOs who run highly efficient data centers allow them to allocate 50 percent or more of their IT budget to new projects. What could *you* accomplish if you could devote 50 percent of your IT budget to innovation? What new opportunities could you take advantage of? Don't get mired in the status quo while your competitor is out driving this type of innovation.

Chapter 2

Looking at Infrastructure for Big Data and Analytics

In This Chapter

- ▶ Taking a look at hardware infrastructure
- ▶ Examining infrastructure components
- ▶ Looking at three aspects of a well-architected infrastructure

Infrastructure for Big Data and Analytics (BD&A) is the combination of components that produce an architecture that delivers business outcomes. While developing this architecture *isn't* rocket science, it *is* computer science.

You can't build a house without a foundation, right? It's the same with BD&A. The finest software in the world won't run without optimal processing power and storage capacity as a foundation. So in developing a BD&A solution, you start with the foundation, which is the infrastructure. Infrastructure must be an early part of your planning. Infrastructure matters. Don't leave it for last.

Looking at a BD&A Hardware Infrastructure

A BD&A hardware architecture is largely synonymous with a hardware and systems software infrastructure. Granted, the architecture is a hardware/software combination, but it's the hardware and systems software infrastructure that's the foundation. It refers to those items that deliver desired amounts of data at desired speeds. A BD&A architecture is based on customer business goals and mathematical analysis of needs.

Customer business goals

The executive knows best. There are often unmet needs for fast acquisition and total storage of data that she can sense. It's rare to hear a customer say, "My computers run really fast, and I have all the storage I need."

Even if that were true, the executive must stay constantly alert to the possibilities of increasing revenue, reducing expenses, and competing more effectively. So chances are that the infrastructure that meets business goals today won't meet them tomorrow.

Needs analysis

An old IT saying is that "it proceeds from need." After the customer declares her business goals, mathematical tools address the infrastructure that meets those goals. For example:

- ✓ If processing is slow, replacements or additions exist in practically every family of computing hardware, from blade servers to mainframes.
- ✓ If storage is inadequate, storage additions exist in all sizes, from simple disk collections to massive disk arrays. Storage capacity planning is relatively easy to do, and executives can find options for both large one-time upgrades and gradual buildouts.
- ✓ If moving data is slow, you have numerous network options with predictable increases in speed (reductions in latency).
- ✓ If data can't easily reach the company, new, innovative ways (particularly mobile apps and the cloud) get the information to the data center.

Introducing the Infrastructure Components of BD&A

This section addresses four core infrastructure capabilities critical for optimal BD&A infrastructure: scalability, parallel processing, low-latency resources, and data optimization.

Scalability of infrastructure

Your BD&A needs *will* change. Presumably, the reason is that your needs grow as your enterprise grows. To accommodate this growth, the infrastructure must feature scalability and resilience.

Scalability, simply put, is the ability of a system to process a growing amount of work. In a well-planned BD&A architecture, scalability is a planned-in component and should pose little difficulty for an enterprise. Well-thought-out scalability makes it relatively easy to add processor and storage power.

Hardware requirements are derived from numbers about your current data utilization, your predicted data growth, complexity of the analytics and underlying algorithms, and your desires and decisions about business goals. Some analyses help predict the hardware that meets the goals.

Parallel processing

There's parallelism in processor design as in system level software. Of course, you need an intelligently designed traffic cop that understands the workloads so data and processing instructions can be intelligently threaded through the hardware layer — from processor to memory to storage. Greater core and thread densities can help boost performance in data and processing as more users begin to scale.

The same can be said for systems level software and for Big Data in particular. File system parallelism is coming of age. Big Data may need to be accessed in multiple ways from multiple places and at multiple rates of speed.

Take a simple example from back in the pre-computing days. Would you want a filing cabinet that has one long ten-foot drawer or ten one-foot drawers to store all your information? What if five people needed to access information at the same time?

The infrastructure design challenge is in determining where and how to store data for the best performance as data scales. High-performance parallel file systems spread the data around intelligently so it can be quickly accessed and updated at high speed.

Low latency resources

High speed (often referred to as *low latency*) is becoming a real differentiator for companies that are leveraging analytics. Low latency is pure speed — speed by design. One of the marvels of modern computing is the tremendous increases made in speed, which embraces CPU speeds, DRAM speeds, disk speeds, and transmission lines (think optical fiber cable). Together, hardware systems are orders of magnitude speedier than their predecessors.

In computing, speed is also architected in the system software. For example, flash memory is very fast, and the price is coming down. It provides real benefits in accelerating BD&A. Also, new forms of resource management software enable scale-up/scale-out architectures to perform at lightning speed. Speed can also be delivered when you optimize the hardware and software together, either at the middleware or application level.

You can be assured that infrastructure latency is capable of meeting your business goals, both simple and elaborate — if you have the right hardware infrastructure in place.

Data optimization

In the world of BD&A, data management takes on a whole new meaning. What data do I keep? Do I need to store it in a different location? What are the costs of dealing with the data? How quickly and when do I need it? All good questions that require new capabilities.

Developing the right data architecture enables data independence and provides a simple, resilient, and agile environment to support BD&A.



Big data isn't just about "social media" Hadoop data. It's about *all* your data and how internal and external data sources provide you with the best insights. Consider the following:

- ✓ More data = better context = better confidence
- ✓ Not all data needs to be stored on expensive media
- ✓ BD&A is dynamic — and your storage needs have to self-tune with your ongoing requirements

Network infrastructure

A *network* is a combination of points or *nodes* and the lines that connect them. In the world of IT, the definition holds up, but has several variants. The variations depend largely on the perspective of the people discussing the network — and they're all legitimate. In all cases, they contribute to the BD&A architecture.

The variants include the following:

- ✓ **Operating departments:** In a business or government environment, *network* likely refers to the connection between desktop computers and processors. That's why it's called a Local Area Network (LAN).
- ✓ **Data center:** *Network* often refers to the system of redundant full crosspoint switches that move data between storage and servers, and vice versa.
- ✓ **Failover:** There's much more to disaster recovery (DR) than local backups or offsite backups. Backups aren't all that helpful when a flood or earthquake hits. Sophisticated companies have robust connections between regional data centers. Should disaster strike, processing will "fail over" to another data center over a Wide Area Network.
- ✓ **Branch offices:** Many businesses have a central data center and branch offices, but banking surely offers the most dramatic example. Bank of America has 5,600 banking centers and 16,200 automated teller machines. That implies an enormous network.
- ✓ **Mergers:** Bank mergers are a special source of concern. There are security issues, compatibility issues, and generally no (or few) branch closings. IBM assisted Bilbao Bizkaia Kutxa (BBK) when it acquired CajaSur. Both institutions were about the same size. That implies an enormous network.
- ✓ **Internet:** Large Internet retailers have robust data center networks, but *also* get information from and send information to many users connected over the Internet.

Some important capabilities to consider in the design of your data architecture include

- ✓ Data virtualization so you can focus less on the data and more on the workload
- ✓ Automated data tiering and compression so you can store more and use more for less money

- ✓ Analytics in a flash; speed up to 12 times faster and more can be achieved by using the latest innovations in flash storage systems
- ✓ Platforms that promote diversity — diversity of workloads and data types and data demands that can be tuned and managed automatically

Understanding the Aspects of a Well-Architected Infrastructure

The following sections describe the three aspects of a well-architected infrastructure: speed, availability, and access.

Speed/performance

BD&A provides insightful answers, but you need the horsepower to get the job done. The infrastructure must have the power to meet current needs and grow to meet expanding needs. An application may have a key requirement to meet certain thresholds. For example, for a customer service analytics application, you can't wait two minutes to get an answer; the answer has to be available in seconds.

Throughput often consists of a series of analytical steps output to a real-time transaction process. Understanding how data and insights flow and how consistently they flow *at speed* is an important essential in planning.

You need to balance cost versus speed and the value of business outcomes. In addition, also consider future growth. A well-architected solution should achieve the best balance among all requirements.

Availability

Resilience is the ability of a system to absorb or avoid damage without suffering complete failure. A well-planned BD&A architecture allows for processor, storage, and facility failures (with protection via “failover” mechanisms), software failures, and at the least a strong backup and restore capability.

Ensuring availability begins with evaluating from where the data comes and to where the insights flow. If insights are important to your real-time operations and processes, resilience (and disaster recovery) is vital. Solutions may range from simple rebuilds of a failed disk to simultaneous asynchronous data replication at a remote site.

Evaluation of insights includes what ones are mission critical, what ones are critical, and what ones are non-critical. Understanding and differentiating insights is important for putting resiliency where it's most needed. The more supportable and manageable a system is, the easier it is to maintain availability.

Access

Understanding data access is the basis for three kinds of efficiency:

- ✓ **Data locality:** It's vital to know where the data is coming from, where it needs to go, and whether there's any pre-processing required.
- ✓ **Data life cycle management:** Be aware of how often data need to be accessed, how much data there is, and the cost of maintaining it. There may be governance issues, and data may need to be retained to keep models fresh.
- ✓ **Data deduplication:** There may be some excellent leverage in storage, achieved by eliminating the same data stored in multiple places.

Chapter 3

Looking at IBM Infrastructure Choices

In This Chapter

- ▶ Seeing a world of infrastructure options
- ▶ Examining the storage universe

A powerful infrastructure goes a long way toward improving success in implementing Big Data and Analytics (BD&A). The alternative is failure, which usually isn't an option for enterprises. Chapter 2 discusses infrastructure capabilities in a general way, but this chapter gives you a look at the IBM families of server, storage, and system software products.

Many products from various sources appear at first to meet the BD&A challenge, but that's not the whole story. You're better served by using products that have superior features designed into them. So, in addition to brute force processing/storage power, look for products with intuitive management that require a minimum of manual intervention and result in reduced administrator time.

IBM Servers

You may want to become familiar with key IBM product line names. They're the core of a BD&A infrastructure and fall into the worlds of processing and storage.



In all cases, infrastructure needs for BD&A meet six critical business needs. Examples are acquire, grow, and retain customers; transform financial management processes; optimize operations and reduce fraud; manage risk; improve IT economics;

and create new business models. For each entry point, certain infrastructure design points (speed, access, availability) are more important than others. Certain design points are better enabled by specific hardware and software infrastructure capabilities and architectures.

Some deployments are more aligned with *scale-out* server solutions, while others are used with *scale-up* solutions:

- ✓ *Scaling out* (scaling horizontally) means adding more nodes to a system, such as adding a new computer to a distributed software application. Hundreds of small computers may be configured in a cluster to obtain aggregate computing power that often exceeds that of computers based on a single traditional processor.
- ✓ *Scaling up* (scaling vertically) means adding resources to a single node in a system. It typically involves adding CPUs or memory to a single computer. Such vertical scaling of existing systems also enables them to use virtualization technology more effectively, as it provides more resources for the hosted set of operating system and application modules to share.



Choose a scale-in, scale-up, or scale-out server infrastructure to support the complexity and breadth of analytic workloads. IBM provides an integrated, high-performance infrastructure including core server, storage, networking, and systems software technology. You can accelerate the flow of data and insights, provide shared and highly secure access to all types of data where it resides, and significantly improve right-time availability. The core systems of the BD&A infrastructure are covered in this section.

IBM Power Systems

IBM Power Systems are highly virtualized, based on IBM next-generation POWER8 multicore processors. Designed around the world's first open server ecosystem, they bring together the computing power, memory bandwidth, and I/O in ways that are optimized for innovative data hungry applications. Power System's support Linux open standards, as well as the industry leading AIX UNIX. The platform enables choice in design and deployment to meet a wide range of data center requirements.

There are multiple deployment styles for business analytics, including server nodes for business analytics grids and scalable server nodes to analyze data stored in data warehouses. The proximity of data stored near (or *on*) these systems enhances overall throughput, as do fast server I/O links connecting to external storage and network switches.



These capabilities have made Power Systems the platform that has driven innovative BD&A solutions, such as the IBM Watson solution that's now commercially available for clients looking to take their analytics to the next level by leveraging cognitive computing.

IBM System z

IBM's mainframe is System z. Much of the world's transactional and financial data still reside on mainframe systems, and the System z BD&A solutions are centered on the concept of moving analytics processing closer to the data, integrating and managing the full variety, velocity, and volume of data that comprises advanced analytics business solutions. This approach relies on the structured data already residing on System z but also incorporates feeds from unstructured and streaming data sources. IBM System z mainframes deliver the ability to easily handle massive data volumes while also scaling quickly and cost-effectively as needs arise — efficiently protecting and governing data and delivering leading-edge computing performance.

IBM System z is highly virtualized through logical partitioning (LPARs) and supports several operating systems, including System z z/OS environment and Linux distributions running on System z Integrated Facility for Linux (IFL) specialty processors. This solution can dramatically reduce downtime (by over 99 percent) and provides the highest level of support in mission critical applications. The built-in support for Linux opens System z to a wide variety of business analytics use cases, supporting workload consolidation and linking to other servers (for example, IBM System x blades or Power Systems blades).

IBM PureSystems

IBM PureSystems is a series of integrated systems that allow choice and flexibility in server deployments, including both x86 and non-x86 server nodes. The nodes are connected by

a high-speed fabric and built-in switches. These integrated systems have built-in security features. Flexibility in server deployments means that you can add nodes later on as demand for computing resources grow. A single console manages shared system resources.



Three major areas of focus in the IBM PureScale design are as follows:

- ✓ **Scale-in systems design:** The system's overall design automates integrated server, storage, and networking resources for both physical and virtual environments.
- ✓ **Patterns of expertise:** The systems embed software tools that automate configuration, upgrades, and application requirements.
- ✓ **Cloud-ready integration:** Designed to be deployed to support private self-service cloud infrastructure and software. That enables flexible scale-up and provisioning of workloads.

IBM System x

IBM System x is IBM's x86 server product line that supports all x86 server-based software for virtualization, Big Data, and business analytics. Built on the IBM X6 technology, these systems are shipped in a variety of form factors, including rack-based and blades, as well as integrated high density storage and are the building blocks for an x86 server-based infrastructure for BD&A. The latest X6 technology provides new resiliency features for those high-availability in an x86 platform. They run Linux distributions (Red Hat RHEL and SUSE SLES) and Microsoft Windows server software, along with associated hypervisors (for example, VMware vSphere, Microsoft Hyper-V, Xen, and KVM).

IBM Storage

IBM has a portfolio of storage capabilities for BD&A deployments. In this section, you take a look at those and get the lowdown on storage basics.

Looking at storage basics

IBM storage solutions are designed to save data center real estate through the use of dense storage and power efficiency. Essential storage technologies are built in and tightly integrated. Essentially, they have the following features:

- ✓ Self-optimizing
- ✓ Cloud-agile
- ✓ Reduced space requirements
- ✓ Intuitive management
- ✓ No manual intervention required
- ✓ Reduced administrator time

IBM storage solutions use techniques such as virtualization, deduplication, data compression, automated tiering, and cataloging. Take a look at the nearby sidebar “A brief glossary of terms” for more information.

Like servers, storage can be scaled up and scaled out. Check out the section “IBM Servers” earlier in the chapter for more information on scaling up and scaling out.

IBM FlashSystem

Huge amounts of data are generated by sensors and computers in addition to data created by people. It’s hard to keep up with data growth. And then, without high-performance storage, it can be impossible to extract meaningful insights from rapidly accumulating data in areas such as customer satisfaction, operational efficiency, financial processes, risk, fraud, and compliance management.

The latest innovation from a BD&A perspective is the use of flash technology for storage. You can boost business agility and insight with flash storage for analytics. That’s where IBM FlashSystem comes in.

IBM FlashSystem is fast, so it unlocks storage performance to give analytics applications the speed they need. But speed is just one component of effective analytics applications.

Storage must also meet three key requirements in order to effectively support analytics workloads:

- ✔ Compelling data economics
- ✔ Enterprise resiliency and
- ✔ Easy infrastructure integration

FlashSystem delivers advanced technologies in each of these areas, so organizations can unleash the power of their data for critical business insights.

A brief glossary of terms

When you talk about infrastructure, you can acquaint several useful terms:

- ✔ **Virtualization:** Creates a virtual (rather than actual) version of something. This goes back to the days of creating mainframe partitions, so multiple jobs could run simultaneously on the same machine. Hardware virtualization is creating a “virtual machine” that acts like a real computer with an operating system. Hardware virtualization is a way of improving efficiency.
- ✔ **Deduplication:** Eliminates duplicate copies of the same data. You get two benefits from data deduplication: a single version of the truth and excellent leverage in storage.
- ✔ **Data compression:** Encodes information using fewer bits than the original representation. That’s a space saver. Compression helps reduce resource usage, such as data storage space or transmission capacity.
- ✔ **Automated tiering:** Automates the movement of data across different types of storage devices and media (*tiering*). The most important data goes to flash or cached disks. Less important data goes to less expensive, slower devices, such as optical disks and tape drives.
- ✔ **Cataloging:** Comprises views and tables that contain information about all the DB2 objects in the database — sometimes referred to as *metadata*. As the database changes, the catalog changes. You can leverage this in many ways, including report generation, automation of required database changes, and checking compliance with database naming conventions.

IBM System Software

IBM provides two key system software offerings designed for BD&A: General Parallel File System (GPFS) and Platform Computing.

GPFS

IBM GPFS is a high-performance, cost-effective storage methodology. GPFS started out as a clustered file system and evolved into much more. Today it's a full-featured set of file management tools, including advanced storage virtualization, integrated high availability, automated tiered storage management, and performance to effectively manage very large quantities of file-based data.

GPFS is designed to support a wide variety of application workloads and has been proven extremely effective in very large, demanding environments. GPFS is installed in clusters supporting BD&A, gene sequencing, digital media, and scalable file serving. All indications are that the future of BD&A includes bringing *more* unstructured and file-based data into the application.



A GPFS *cluster* can be a single server, two servers providing a high-availability platform supporting a database application, or thousands of servers used for applications such as the modeling of weather patterns, or supporting the needs of advanced analytics leveraging file-based data. GPFS was designed to support high-performance workloads and has been proven effective for a variety of BD&A applications.

Platform Computing

When you have big data problems to solve, you need to maximize the potential of your computing power and the supporting infrastructure to accelerate your applications at scale, extract insight from your data, and make better decisions faster. IBM Platform Computing provides that by pooling your computing resources, managing them efficiently across multiple groups, and getting the most out of your IT investment.



For those looking to accelerate Hadoop performance, IBM Platform Computing's Symphony delivers peak performance as measured by Terasort and UCB SWIM results. Symphony is a low-latency grid management platform for SOA and distributed applications. It supports diverse compute and data intensive applications.

Chapter 4

Solving Big Problems with Powerful Solutions

In This Chapter

- ▶ Seeing how businesses deploy infrastructure
 - ▶ Learning about problems addressed and improved outcomes achieved
-

IBM has deployed infrastructures for clients in a range of industries around the world. In this chapter, you see several areas that are affected by infrastructure and how it drives new revenue (or increases revenue), services customers better, and reduces the costs of IT.

Energy and Utilities

Oncor Electric Delivery Company, LLC, is an electric distribution and transmission company in Texas with more than three million customers and a service area covering nearly 117,000 square miles. Oncor began its advanced meter program (digital smart meters) in 2008 to collect data on usage, load, and availability. Oncor recently completed the deployment of 3.25 million meters serving the customers of north and central Texas. Last year, customers in Oncor's service territory showed during the company's biggest energy saver contest that by using the information from Oncor's advanced meter, they could reduce their electric usage and bills by 25 percent or more.

Business Need

The Oncor smart meter deployment creates immense volumes of data, including an average of 2,900 records a month for *each* of its 3+ million customers. The company needed a more efficient and cost-effective alternative to its existing storage setup.

To accomplish its goals Oncor, partnered with IBM and IBM Business Partner Ecologic Analytics, along with Advance Metering System provider and IBM Business Partner Landis + Gyr, to implement a Meter Data Management System (MDMS) and deploy smart meters across its grid — meters that would collect energy usage and event data automatically.

The meters provided consumers the data needed to become active participants in their power consumption management and enabled individual households to reduce electrical consumption by an estimated 5 to 10 percent.

Solution

Oncor deployed IBM Power Servers and System z with IBM ProtecTIER.

Fashion

Bernard Chaus (Chaus) is an American clothing company that retails consumer products. The clothing line is made to enhance the dynamic woman's personal style. The company is involved with wholesale distribution and services. Chaus found its sales and merchandising tracking efforts were weighted toward time-consuming report creation and data issues. That, combined with the increasingly competitive nature of the business, prevented Chaus from effectively analyzing and addressing its selling opportunities.

Chaus enlisted SKYPAD from IBM Premier Business Partner Sky IT Group to help analyze and address selling opportunities. SKYPAD is delivered via a Web portal and integrates weekly electronic data interchange (EDI) and spreadsheet feeds from each retailer selling various Chaus products.



EDI is a particular set of standards for computer-to-computer exchange of information.

By using SKYPAD, Chaus has the ability to track daily which clothing items are selling (and which aren't) at the retailer level. This unique insight is achieved by merging information from retail outlets around the world augmented by Chaus Enterprise Resource Planning (ERP) data.

This solution for Bernard Chaus runs on IBM infrastructure hosted by Sky I.T. Group. It offered Bernard Chaus the analytics capabilities it needed to make sense of sales data and generate business insight. The hosted infrastructure behind these IT capabilities included an IBM POWER server, which Sky I.T. Group uses to run the SKYPAD application (using data collected from the ERP and retail point of sale systems) and IBM DB2 software running on IBM BladeCenter servers to power the SKYPAD database.

Henry Thervil, president of sales for Vince Camuto, Bernard Chaus commented that “The sheer fact that we have IBM hardware and software with SKYPAD has given us the analytics capabilities to gain a big advantage over our competitors.”

Travel

Go Voyages, an online travel company in Paris, France, offers a full range of services including accommodations, cruises, short- and long-term vacation packages, and insurance. Go Voyages has 450 employees with websites in English, French, German, Italian, and Portuguese.

The company needed the ability to manage annual web-based business growth of up to 30 percent with intensive behavioral analytics and was looking for a solution. Part of the Go Voyages' business model is to analyze customer behavior based on response to specific travel offers and discounts and the volume of data threatened to overwhelm the existing systems. Go Voyages wanted to create a solution that would solve the immediate challenges and scale continuously to deal with future growth.

Finding the solution

Go Voyages chose IBM Power and System x servers for both their front office and back office systems. The IBM solution allows Go Voyages customers to compare prices from their travel providers and to propose the best price to online customers. For example, during the air travel disruptions caused by the Iceland volcanic ash clouds, Go Voyages was able to find flights for stranded clients when many other travel companies could not.

Looking at the benefits

Being able to connect to different global reservation systems means that Go Voyages can offer a far greater choice to its consumers, and the IBM Power servers provide quick cache services for more than two million daily unique web searches. Even with 700 query parameters, the new IBM solution is able to process customer requests twice as quickly as in the past.

The insight generated from behavioral analysis is used by the Go Voyages marketing team to develop better marketing promotions. This helps ensure that more customers complete their ticket bookings, generating increased sales. As mentioned, sales are rising by 30 percent annual — 25 percent more than the industry average.

Responding to customer behavior

Go Voyages has deployed a business intelligence solution, built on IBM hardware and software technologies, to create and support a comprehensive data storage environment for the company's travel data and payment records, and enables customer data to be integrated and analyzed using the software's predictive modeling capability.

For example, customers who want to purchase an airline ticket from Paris to New York may be discouraged by the high price. Using behavioral analytics, based on the IBM solution, Go Voyages prompts the customer with alternative routes

with a short layover — and with substantial savings. The analysis allows Go Voyages to tailor the response to flight inquiries and to increase the probability of ticket sales.

Healthcare

Park Nicollet Health Services provides healthcare services through its network of 25 clinics, six urgent care sites, and one hospital. From its headquarters in St. Louis Park, Minnesota, the organization provides inpatient and outpatient services, emergency care, pharmacy services, and health and wellness programs. Although many other healthcare organizations operate hospitals as separate entities from physicians, which can inhibit information sharing, Park Nicollet set out more than two decades ago to create an integrated care network designed to encourage collaboration.

Park Nicollet wanted to improve the quality of care for its patients by implementing a new electronic medical record system and new servers capable of supporting advanced analytics. So the company deployed an Epic system supported by IBM Power Systems.



The IBM solution achieved the following:

- ✓ Employed advanced analytics and reduced the heart-failure mortality rate of patients by about 20 percent
- ✓ Reduced rack-space requirements by more than 75 percent compared to previous servers
- ✓ Reduced failover time from more than two hours to less than two minutes

Insurance

The Swiss Re Group, headquartered in Zurich, is a leading worldwide wholesale provider of reinsurance, insurance, and other insurance-based forms of risk transfer. Its global client base consists of insurance companies, mid-to-large-sized corporations, and public sector clients. From standard products to tailor-made coverage across all lines of business, Swiss Re deploys its capital strength, expertise, and innovation power to enable the risk-taking on which enterprises and progress in society depends.

Overview

With 2.5 billion transactions and records dating back nearly 30 years, Swiss Re needed a solution with the flexibility and capacity to manage its growing data workloads. The company wanted to acquire insights from multiple international locations of claims information from numerous insurance companies. These insights would increase the understanding of risk, the agility in identifying profitable segments, and the ability to make decisive decisions quickly by internal business users. The company needed a system that would also scale to meet future data and analytics needs.

Solution

Swiss Re chose to deploy IBM zEnterprise System with IBM DB2 for z/OS to perform data analysis and reporting from a central location, with IBM DB2 Analytics Accelerator (IDAA), powered by IBM Netezza technology, to deliver faster responses to individual analytic queries.



With the solution in place, the company benefitted from

- ✓ Increased satisfaction among internal business users by speeding report generation by 70 percent through faster query response times with IBM DB2 Analytics Accelerator for z/OS
- ✓ Off-loaded workloads, reducing processing costs
- ✓ Transparently implemented deployment without program modification, enabling fast and cost-effective installation

Travel and Transportation

Stuttgarter Straßenbahnen AG (SSB) provides passenger transport services in the greater Stuttgart region of Germany. The company is owned by the State Capital Stuttgart and acts as the region's main public transport operator. SSB manages an extensive network of light rail lines, bus routes, and a rack railway, carrying a total of more than 171 million passengers annually. In all, the company operates approximately 450 buses and trains, and employs more than 2,900 people.

In addition to its public transport duties, SSB acts as an IT service provider for other public transport companies and medium-sized companies from various industries. Its IT service business specializes in SAP ERP hosting, and the company also provides general data center services and runs online shops, marketplaces, and business-to-business portals for its clients.

Business need

Slow response times or lack of application availability could have a huge impact on SSB's ability to properly coordinate its activity and deliver safe, efficient urban transport. It could also cause significant disruption to its clients' operations, leading to lost revenues if clients chose to switch hosting providers. The company wanted to ensure fast access to enterprise data, in order to provide its own executives and its clients with the information they needed to perform timely, accurate analysis and reporting and guide better decision-making.

Solution

SSB chose IBM Power Systems servers, running IBM AIX, to deliver the performance it needed for its business-critical systems, including an extensive SAP ERP application landscape. To manage growing quantities of data and provide fast access to vital information, the company extended its existing storage environment, based on IBM System Storage SAN Volume Controller and high-performance IBM FlashSystem technology.



The solution allowed

- ✓ Quicker payouts with optimized commission calculations in SAP Revenue Management and Contract Accounting (SAP RM-CA), reducing runtime from five hours to just 40 minutes, a time savings of more than 86 percent
- ✓ HR staff new opportunities to work better with data by accelerating complex payroll processing from six hours to six minutes, a time saving of more than 98 percent

IBM FlashSystem storage is expected to deliver a full return on investment within five years and offers a total cost of operation that is 50 percent less than traditional storage solutions.

Chapter 5

Ten Questions LoB and IT Execs Should Ask Each Other

In This Chapter

- ▶ Looking at common BD&A questions
- ▶ Providing additional info for the LoB executive
- ▶ Preparing the IT manager for talking with the LoB executive

Big Data and Analytics (BD&A) offers numerous advantages for an enterprise. However, different Line of Business (LoB) executives have different business processes and are likely to have different perceptions about the value of BD&A and BD&A infrastructure:

- ✔ To a senior executive responsible for sales and marketing (whether it's the light truck division or domestic airline ticket sales), BD&A can be a nimble trend spotter that builds sales and enhances the client experience.
- ✔ To a senior executive responsible for national or world-wide finance, BD&A provides a rapid view of what is and what may be financially beneficial for the firm.
- ✔ LoB executives are cost conscious, and when there's a proposed improvement in infrastructure that costs money, the IT executive has some explaining to do.
- ✔ As an IT person, you may need to do some selling to get buy-in. The core principles for these internal sales are the same as for client sales — emphasize benefits, not features, and put costs in proportion.

This chapter gives you ten questions LoB and IT executives should be asking each other (and also be prepared to answer).

IT to LoB: Are You Aware of the Benefits of BD&A Infrastructure?

Use case after use case shows the benefits of BD&A. Those cases address many industries, and your operation is likely very similar to one of them. The IT department and IBM (along with partners, sometimes) can analyze operational problems. There are metrics that will predict the benefits the enterprise is likely to see. Take a look at Chapter 4 to see five use cases from different industries.

IT to LoB: Are You Aware of BD&A's Return On Investment?

In relative terms, BD&A infrastructure is a bargain. The infrastructure deployment for the company is scaled to the company. When the company does a thorough cost/benefit analysis, it can factor in increased productivity (and hence lower overall labor costs), manufacturing efficiencies, marketing timing efficiencies, and so on. It's a game of chess, where you need to see the *whole* board. For example, wouldn't the company save money if it made only products that customers wanted and none that they didn't want?

Change in IT is *continuous* change, and a flexible, agile infrastructure minimizes longer-term changes and impacts. Overall, that can improve return on investment (ROI). Besides, in absolute terms, most change has produced lower costs. Consider Moore's Law, predicting regular increases in processing power with lower costs. The same is true of disk storage. A terabyte of storage was once obtained with 30 hard drive costing \$400.00 each — or \$12,000.00. These days that same terabyte costs about \$230.00. For more information on enhanced ROI, see Chapter 1.

LoB to IT: How Can I Benefit from Bringing the Back Office to the Front Office?

Back office operations have traditionally been thought of as non-customer-facing operations. But in fact, certain operations, such as procurement, shipping, or billing, affect customer satisfaction. And increased customer satisfaction can produce a competitive advantage. There is back office data, which, when aggregated and available for analysis, can create useful marketing information for those in the front office.

LoB to IT: Do I Need to Be On the Cloud?

The cloud has its advantages, and IBM has use cases that support that. Just the same, you can achieve business goals with a robust BD&A infrastructure and no cloud. Chapter 2 mentions that considering the cloud should be part of needs analysis. Chapter 3 discusses IBM systems that are “cloud-ready.” Check out those chapters for more information.



No matter what you do, make sure you and your IT team understand the infrastructure implications early.

LoB to IT: Should I Worry about Obsolescence?

Not at all. Business intelligence (BI) and predictive analysis have always been components of business. The concept of BD&A is the next step in logical evolution from slow, scattered “little data” to a big picture vantage point. Particular items go obsolete, but constant reinvention replaces them — sometimes at a lower cost than the items they replace. That’s the way a flexible, agile infrastructure works.

IT to LoB: Where and When Would Performance Issues Occur?

Given the data that must be gathered and analyzed and the speed that's required, think about where you may have performance issues. These could pertain to the amount of data required, the number of people or processes using the outputs, and where the data resides. For example, a performance issue may be the need for a sales/service agent to have up-to-date information *in seconds* after a customer query. Having the data could magnify upsell/cross-sell opportunities.

IT to LoB: Are There QoS or SLA Concerns?

An application (or the workload it must process) could run the risk of falling short of contracted Quality of Service (QoS) or Service Level Agreement (SLA) goals. This situation can be prevented by deploying a robust BD&A architecture. Refer to Chapter 2 for a brief discussion of speed/performance.

IT to LoB: What Security Requirements Exist?

Is the data exposed to a security risk? Are the outputs or the delivery of insights also exposed? The company usually wants a solution that provides security for all these. Chapter 1 gives you four key items that promote success in managing risk.

IT to LOB: Does the BD&A Application Require Mission-Critical Resiliency?

Consider the impact of a failed system. What would happen if the application stopped working for ten seconds, ten minutes, or ten hours? See Chapter 2 for more on resiliency and failover.

IT to LoB: How Will the BD&A Application Expand?

A well-architected BD&A application expands gracefully to meet future needs. Also, it should be able to leverage other data and business processes in the future. See Chapter 3 for more info on IBM storage. “Future needs” embraces the growth of an application to manage increased volumes and faster speeds in a primary application. The LoB executive should consider other company operations, probably with new data sources and different users, that would benefit from expansion of the BD&A application.

Deploy an infrastructure that meets your business challenges

Big Data Analytics is a critical tool to gain competitive advantage and to attract and retain happy customers. But simply deploying a software package isn't enough. Infrastructure matters for Big Data Analytics. This friendly book explains the value of infrastructure and how to choose what's right for your business.

- *Big Data Analytics is a game-changer — your competitive advantage depends on it*
- *Infrastructure matters for Big Data Analytics — don't leave it for last in your planning process*
- *IBM offers a broad portfolio of solutions — see what meets your infrastructure needs*
- *Big Data Analytics is deployed cross-industry — learn how companies have succeeded with the right infrastructure*



Open the book and find:

- How Big Data Analytics is changing business
- Why infrastructure matters
- How to choose the right infrastructure

Go to [Dummies.com](https://dummies.com)

for videos, step-by-step examples,
how-to articles, or to shop!