Brought to you by:

Cloud Analytics

for dummies®
A Wiley Brand

Learn which data is best for analysis

Pair big data with cloud analytics

Perform cloud analytics across ecosystems

Liz Kao

2nd Informatica Special Edition

About Informatica

Informatica is the only Enterprise Cloud Data Management leader that accelerates data-driven digital transformation. Informatica enables companies to unleash the power of data to fuel innovation, become more agile, and realize new growth opportunities, resulting in intelligent market disruptions. With over 7,000 customers worldwide, Informatica is the trusted leader in Enterprise Cloud Data Management. For more information, call +1 650-385-5000 (1-800-653-3871 in the U.S.), or visit www.informatica.com.

Cloud Analytics

2nd Informatica Special Edition

by Liz Kao

for dummies®
A Wiley Brand

Cloud Analytics For Dummies®, 2nd Informatica Special Edition

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

ISBN 978-1-119-45952-1

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

Project Editors: Carrie A. Johnson, Martin V. Minner

Editorial Manager: Rev Mengle

Senior Acquisitions Editor: Amy Fandrei

Business Development Representative: Karen Hattan

Production Editor: Magesh Elangovan

Special Help from Informatica: Clive Bearman

Contents at a Glance

Table of Contents

Introduction

Over the last few years, new buzzwords have cropped up in conference rooms across the world as a result of the oceans of data that new and improved technologies can capture. *Cloud analytics. Big data. Hadoop.* These words are often used to broadly define solutions for unlocking the secrets that monumental amounts of data can hold. You've either been in those meetings (heck, you probably used some of those words yourself without fully understanding their meanings), or you have some new work cut out for you as a direct result of those meetings. But really — what do all those phrases actually mean?

About This Book

This book gives you a quick overview that helps you become one of the smartest people in the room on the topic. Both business users and budding data scientists get an overview of the world of cloud analytics.

Find out why you should care about the evolving world of cloud analytics and the foundational concepts that every person should know. Along the way, discover the types of solutions offered within the analytics world with some example vendors, and understand the guidelines that steer you in the right direction as you navigate approaches to take for your company.

Icons Used in This Book

For Dummies uses helpful icons to point out information. You will find the following icons in this book.


TIP

The Tip icon calls out advice and shortcuts that you can use while learning more about cloud analytics.

The Technical Stuff icon marks information of a more technical nature that can be skipped over (unless you really want to delve into those details).

The Warning icon marks important information that, when heeded, may save you a lot of headaches.

Beyond the Book

This book can help you discover more about cloud analytics, but if you want resources beyond what this book offers, here's some insight for you:

- **State of Cloud Analytics Report:** `www.informatica.com/stateofcloudanalytics`
- ***Salesforce Integration For Dummies,* Informatica Special Edition:** `http://infa.media/sfdcbook`
- **Cloud Data Warehousing With Amazon Redshift:** `www.informatica.com/AWSWorkbook`
- **Free 30-day trial of Informatica Cloud:** `http://infa.media/freetrial`
- **Informatica Cloud iPaaS Platform:** `http://infa.media/ipaas`
- **Develop a Governed Self-Service BI Strategy:** `www.informatica.com/Tableau`

IN THIS CHAPTER

- Learning about traditional business analytics
- Making sense of all that information
- Finding out what's in it for your company

Chapter 1
Comprehending Cloud Analytics

Before I explain different ways you can use cloud analytics to benefit your business, I want to make sure all readers are on the same page regarding context and background. In this chapter, I tell you about the origins of cloud analytics, where the movement is today, and why you should care.

Reviewing Traditional Business Analytics

Looking back at the beginning of this century, companies that wanted to capture and analyze data about their customers had to spend a lot of money on expensive software, hardware, and people who specialized in getting business users the insights they wanted. These requests for data analysis would come from the business and take a wide-ranging amount of time to fulfill. It would be a minor miracle if a report was delivered in less than a week, since the specialized data folks were unquestionably

always understaffed. This realm was labeled "Business Intelligence" (BI for short, pronounced *bee-eye*) and is a fancy way of describing the world of getting metrics to business people so they could make smart, educated decisions to further benefit the business.

Business analysts, also often with technical backgrounds, worked on collecting business information and storing it within a *data warehouse* (a central repository of data that's been integrated from a number of sources). How the data was stored in the data warehouse (DW) took a lot of careful planning, because multiple technical people running reports need to query the DW for answers to business queries in a way that didn't slow down the entire system. (Few things are more frustrating than relying on a different, usually understaffed group to get you what you need, only to find out that the system it designed can't get you what you want. There have been more than a few times where one has had to wait for what seemed like an eternity just to hear, "Oh sorry, we aren't capturing that one metric, so we can't really get you the information you need.")

These teams of people and their skills didn't come cheaply. Only large enterprise customers with big warehouses of data to analyze could afford to hire these folks to configure and tweak the very specialized and not-so-sexy DW user interface. Think of large clothing retailers with national brick-and-mortar stores that need to analyze purchasing trends in stores across the country. Or a consumer packaged goods company with thousands of product lines that wants to find ways to better cross-sell or up-sell its cereal lines with its snack bar lines. This type of reporting is also described as "operational reporting," where people are tracking widgets (usually inventory), looking at historical purchasing trends and other customer activity to better understand which customer is more likely to buy what product line and when.

Although investing in a BI solution was historically an expensive endeavor, the growing acceptance of cloud-based applications (such as Salesforce.com) and advances in digital and mobile technologies have accelerated developments in the world of BI.

Identifying Current Business Analytics Trends

Salesforce.com's success as a cloud-based business application (where it takes care of the database, software upgrades, security, and backup of data all for the price of an ongoing subscription) opened the floodgates for other entrepreneurs to introduce cloud-based apps for other business processes. As more users adopted the cloud, the central source of data also shifted online.

Similar to how Salesforce.com's hosted CRM application offers "just what you need" functionality in a simpler user interface to businesses at an affordable price (compared with the traditional software options), cloud analytics vendors also offer a similar value proposition. Users can now pay for what they need and add more functionality later as the need arises. At the same time the business application sector saw a dramatic increase in cloud-based offerings, consumers also started adopting cloud apps (like Facebook), which suddenly made global interaction that much easier.

Innovation in the digital arena suddenly allowed vast amounts of music and photographs to be compressed (digitally) into a computer without requiring users to lug around awkward external hard drives. Of course, the introduction of the iPhone escorted in a new era of mobile adoption that added to the tons of data now being generated.

This confluence of factors is creating a wealth of data to store and analyze, all at a velocity never before experienced. Business analysts don't just have to look at data surrounding inventory trends; they now have data around usage that isn't immediately related to the inventory in question but could be correlated. For instance, do the number of "likes" on a fast-food retailer's Facebook page correlate to a growth in sales? What about the number of "pins" on a clothing retailer's Pinterest page?

Traditional BI never foresaw these types or volumes of data, so it can be a struggle to handle them. Add in a new generation of potential BI users who've grown up in this new digital age, and they're expecting the business tools they use to help them

easily process and understand all this information. Cloud analytic vendors (like Amazon Redshift, Salesforce Analytics Cloud and Tableau) have been built from the ground up ready to input these new streams, volumes, and types of data.

Spotting Where Your Business Fits In

Cloud analytics is compelling because smaller non-behemoth-size companies that never before could have afforded a BI solution can now gather and analyze all the information flowing around them. Not only are there cost savings to cloud-delivered applications but also time savings. These apps are easier to navigate and don't require a lot of setup to get up and running.

Advances in data visualization from companies like Tableau or Salesforce's Analytics Cloud mean more and more new people within a company can act as business analysts. These sleek tools have been created with design principles targeted toward empowering the business user — similar to how the iPhone's or iPad's interface is somehow easily comprehensible by young and old alike and not just the tech gadget fans. In some cases, business users themselves are empowered to build their own reports. This also means there is a whole new generation of users out there learning about BI or cloud analytics (of which you might be one).

As shown in Figure 1-1, the newest chapter in cloud analytics empowers business users to access and report on information that used to only be available from IT or specialized business analysts.

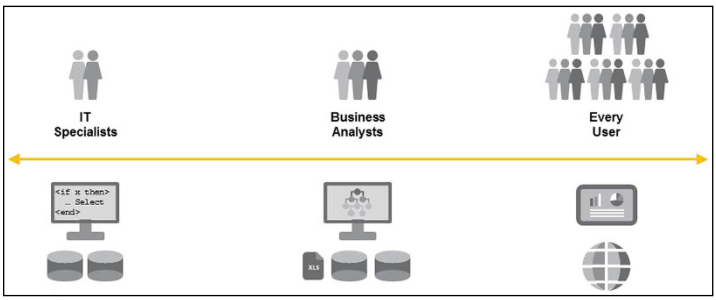

FIGURE 1-1: Every business user is now a business analyst.

Business users can also receive results in dramatically less time than before, which is driving competitive advantages to businesses that are nimble enough to look at the data and make smart, quick, strategic decisions.

You may already be a part of this new generation. Maybe you're a businessperson just learning what this is all about. Or maybe you're a technical person in IT who has a lot of traditional software implementation experience in-house, and you're wondering if this world of cloud analytics is as mature as you need it to be. Or maybe you're someone with a foot in both worlds. You're on the business side of the house (such as a sales operations team), but you're a logical, analytical thinker and have taken it upon yourself to be the report creator for the head of sales.

The explosion of users who now have access to analyze data has also introduced a variety of innovative ways in which cloud analytics can be used to solve business challenges that used to be within reach of only the biggest enterprise companies out there. In Chapter 2, I examine examples of business challenges that are being tackled today with cloud analytics.

IN THIS CHAPTER

- Reviewing technology drivers to cloud analytics
- Understanding business motivators
- Learning why some companies went with cloud BI

Chapter 2
Discovering New Use Cases for Cloud Analytics

Because a lot more data is being generated now than ever before, and technological advancements have made options available to more and more businesses, what can you do about it? In this chapter, I discuss common drivers that lead both technical and business people to adopt cloud analytics. Whether you're in management or you're a "boots on the ground" individual contributor, it's in your best interest to start thinking about ways your company can benefit from cloud analytics before you're left behind. I also provide a couple real-life examples of companies that chose to embrace the cloud and why.

Identifying Drivers for Technology Users

As processing power has advanced over time, it makes sense for technical leaders with an existing enterprise data warehouse to understand the benefits of using cloud-based data warehouses. Generating vast amounts of data, and storing it in a way that provides later business value without sacrificing the time it takes to

access that data, is often an elusive combination that can eat up a lot of employees' time. That's valuable time that could be spent on innovating and gaining a strategic advantage from your competitors. In this section, you discover three drivers that motivate many technology users to explore what cloud analytics and cloud-based data warehouses can do for them.

Reducing hardware costs with a cloud-based data warehouse

As your business generates and wades through more data, that data has to live somewhere. Data has to live on hardware (just like those of you who save all your digital photos onto a hard drive on your home computer.) When one disc gets full, you need to have another one at-the-ready to start saving data without missing a beat.

Even though costs of purchasing extra hardware have gone down as performance has increased over time, that's still a cash outlay that has to happen. And in the business world, buying your own hardware that's robust enough to store all the data you're generating isn't as cheap as you going to your local big-box discount store and buying a new hard drive. Depending on your hardware needs, it can cost anywhere from a hefty four- to five- (sometimes, even six-!) figure sum to add new machines and storage.

Don't forget to factor in additional costs of time (remember that time is money!) that your specialized employees will have to invest. For instance, they'll need to go through vendor negotiations, receive the physical shipment of hardware, set it up (it's not plug-and-play) and configure it, get the most stable version of system software installed that works with your specific make of hardware, ensure everything is working smoothly, and ensure that your existing configuration recognizes the new hardware. What if those specialized employees could instead spend all those hours enhancing your systems versus just setting up new hardware to achieve parity with what's already in-house?

With all these direct and hidden costs, many technical leaders are adopting cloud data warehousing and cloud analytics to reduce costs. Instead of adding hardware, they're considering extending their existing on-site data warehouse with a cloud-based option.

Improving performance with a cloud-based solution

IT leaders are also examining the performance improvements they can experience in processing data without investing further in hardware. Instead of having to shop for the most optimal hardware configuration that can give them performance improvements without breaking the bank, technology users can rely on cloud vendors to do this research for them (as part of the cloud subscription).

Cloud data warehouse vendors take on the research and vendor negotiations as they regularly upgrade their equipment and configurations to continue providing top-of-the-line systems and service to their customers. Performance metrics and reducing system downtime are crucial to a cloud vendor's existence. Competition in this field also benefits the consumer. A cloud vendor relying on an older stack of hardware and software risks competing with rivals offering faster performance at a competitive price point. And just like with a magazine subscription, if you don't like the service you're getting from a cloud-based application vendor, you can cancel the subscription and switch over to someone else (though of course I recommend a proper cost-benefit analysis before doing anything rash).

Responding faster to requests from business users

A third benefit that intrigues technology users when they research cloud-based data warehouses is the dramatic improvement in getting business requesters the information they need in a lot less time. Additional cost also means that business users have to wait to get their requests fulfilled. Why? Maybe the same co-workers who could run reports for them are the same people setting up the hardware and making sure it's working right. That's a potentially expensive opportunity cost if it's preventing business users from making timely strategic decisions. Alternatively, the person who could run the report for the business requester is herself bottlenecked because of poor system performance issues related to a data warehouse that's bogged down.

Cloud-based data warehouses offload a lot of the legwork that your in-house employees traditionally had to take on (see the earlier

section on costs to adding more hardware). Without those issues, they can focus their time on adding value to the data architecture that most efficiently and quickly provides the data that business requesters want. This time improvement can reduce the behind-the-scenes setup time from months to weeks. And then it can mean the capability for the business to get its information within a day or a few days, versus weeks (if at all).

Opening Doors for Business Users

Business users have been very open to the benefits provided by cloud-based business applications. Cloud-based analytics and underlying cloud-based data warehouse architectures provide vast improvements over traditional methods of data warehousing and BI. In this section, I discuss a few advantages that business teams get when they "go cloud."

Accessing business data from standard reporting tools

Today's cloud-based data warehousing options have often wisely partnered with experienced technology integrators to provide pre-defined connectors to a variety of business systems. Integration experts (like Informatica) provide a portfolio of these connectors to help reduce the time it takes for their customers to get to the data. This means reducing the time it takes to derive insights from the data.

With large cloud-based systems like Amazon Web Services, an integration company like Informatica realizes that many customers will need to connect to the same system. Why have businesses duplicate the same effort to build connectors to the same system, from scratch? That would seem no different than the individual effort each company invested in the traditional data warehouse days.

With pre-built connectors, data can now flow more quickly and consistently from a data warehouse into a reporting application (like Salesforce Analytics Cloud, Birst, or Tableau).

In Figure 2-1, you see an example of how data is presented using Salesforce's Analytics Cloud product.

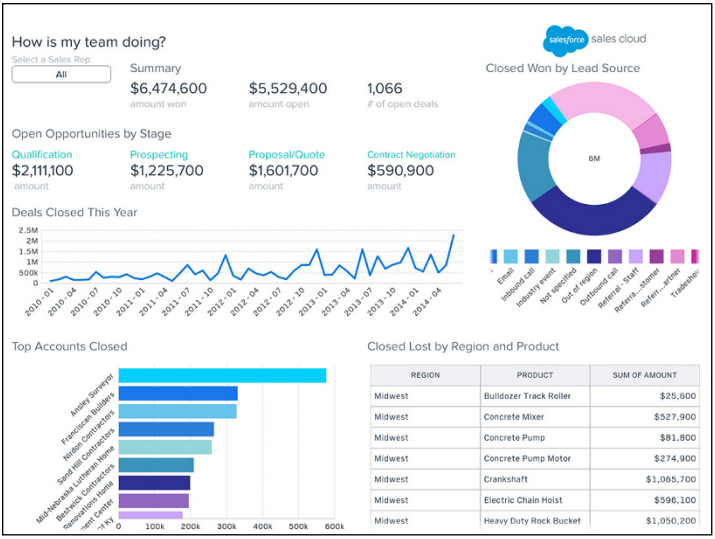

FIGURE 2-1: A Salesforce Analytics Cloud dashboard.

Incorporating analytic functionality into applications

For businesses that sell software or that build internal analytic tools for their employees, access to cloud-based data warehouses gives them the opportunity to embed better analytical capabilities into their own products.

Allowing internal or external customers to build their own reports and run their own analysis on their data would take a monumental effort inside a company. Not only would they have to build the analytics infrastructure internally, but also they'd have to build the hardware and networking infrastructure to allow the entire system to scale, while providing security so that customer X doesn't mistakenly run a report and see customer Y's data.

By using a cloud-based data warehousing infrastructure, companies can now provide analytical features to their own customers without overwhelming investments in hardware, software, networking, and security needed to provide a reliable customer experience.

Providing global customer visibility

By harnessing all existing sources of data in a timely manner, which provides faster access to customer insights via the cloud, technology users can now more feasibly provide business users the once-elusive goal of unified customer information. Unified information allows departments to pull data from one source of truth. Sales can now run reports off the same sales and lead data as Marketing, and Finance can follow the same bookings numbers as Sales. Again, the time freed up for the technical users can now be leveraged to harness all existing sources of data, bringing them together in the cloud and pushing it to various reporting systems.

A head of sales and a CFO can both see bookings information (which is captured in a customer relationship management tool like Salesforce.com), while also seeing revenue information served up from financial systems (since people may pay for what they've bought over several periods of time).

Chief Customer Officers who are in charge of expanding business from existing customers and preventing attrition can now use analytics to track various customer adoption information that collectively acts as an early warning system for customer happiness. Customer satisfaction scores, login information (for customers of web-based subscription products), and usage data can combine to create alerts if certain thresholds are surpassed or not met.

These are just a handful of examples of where business users are using better access to data to improve the service to their customers. What insights would improve the way your company does business today?

Learning from Other Companies

If you've been reading this chapter from beginning to end, you're likely building a case for adopting cloud analytics. So in this section, I give you a few examples of how some businesses today are using cloud analytics to improve their specific needs.

Keeping local efficiencies while gaining global insights

Picture a global financial services firm. It has three independently operated regions: the Americas, EMEA (covering Europe, the Middle East, and Africa), and APAC (Asia-Pacific countries). Each region uses its own technologies and systems to track customer information and financial data. This de-centralized approach allows a region to remain flexible, especially when it comes to addressing regulatory concerns in the area. Business practices are also fine-tuned to the cultural nuances of the region.

Though each region is quite nimble on its own, its separate systems and processes make any global understanding of the business nearly impossible. An enterprise data warehouse approach would require a lot of time to collect all the various reports from each region, and then roll them up to a central authority. It would also require a lot of specialized resources to administer the data warehouse. It would also be quite pricey with building and maintaining an on-site data warehouse. All this meant that the company was very slow in reacting to macroeconomic global forces.

This company turned to cloud analytics to provide a holistic view of global performance without the cost of setting up a traditional data warehouse. With cloud analytics, the company could choose which data sources it wanted to work with first and could take advantage of several pre-built connectors so its team didn't have to build them from scratch. With a cloud-based system, it was also able to use just the amount of analytics storage and computing resources that it wanted. Similarly, the company could add more data connections as the initiative gained credibility. It knew it could always add more storage or resources when it hit certain usage boundaries, and that this could be done with less hassle or cost than in the traditional DW world.

Analyzing client behaviors across new channels

A second example concerns a media company that supported major technology communities. It wanted to better target its users by gaining more insights into their online content consumption behaviors. The company had data from a number of end-user interactions. Some of these were in-person, like at a trade show,

or virtual, with numerous online webinars it held, and volumes of website interactions where people were reading online articles. The various channels through which users accessed this information online came from a variety of mediums, too. People accessed this information via their computers' browsers, from iPhones, iPads, and Android phones.

All this data was extremely fragmented. Event attendance information was kept on Excel spreadsheets, while website traffic information was in an online system, and webinar attendance was in a third-party system.

Knowing that it had to incorporate several data sources at once and faced with an aggressive timeline, this company decided to use cloud analytics to get a pilot up and running sooner rather than later. The purchasing cycle was less time-consuming than what usually went through its procurement department, and the lower cost also meant the company spent less time seeking purchasing approvals.

With cloud analytics, this company was able to consolidate customer demographic and company information with engagement preferences, content download history, event registration preferences, and the customer's influence. The client was able to gain visibility into how its content was consumed during the customer life cycle, which allowed an increase in lead volumes and open rates of targeted marketing emails.

IN THIS CHAPTER

- Identifying the types of data you want to analyze
- Categorizing your data needs
- Evaluating your cloud analytics options

Chapter 3
Following a Cloud Analytics Blueprint

With all the buzz around cloud analytics, you'd think that all you need to do is buy these products to hit the bull's-eye, exposing all the answers your company needs to excel and beat the competition. Not so fast, friends. Though many cloud analytics solutions come with pretty visualizations, a lot of thought goes into getting the data to present itself so nicely. Data visualization is the icing on the cake of well-collected and scrubbed data. You've probably heard the expression, "garbage in, garbage out." That's still the case when it comes to your data quality and how you want to consolidate it to get to some meaningful insights. In this section, I discuss the underlying foundations that need to be established correctly in order to establish meaningful visual intelligence.

Integrating All Your Data

The first step is identifying the sources of data that you wish to consolidate and analyze. What do you need to know? Where is that data captured today, if at all? How do people get access to it? How reliable are the sources of data that house the information you're

trying to access? Before you can derive great analysis from data, you have to collect it all in one place. Consider this the foundation, or first layer, that needs to be built before anything useful can come out of a cloud analytics solution.

Figure 3-1 shows all the participants in a cloud analytics ecosystem.

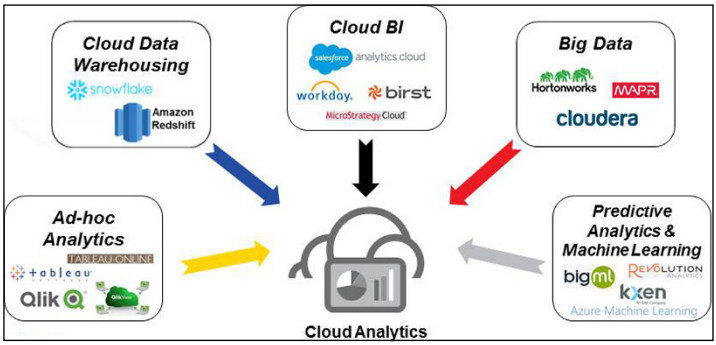

FIGURE 3-1: Different forces are converging on cloud analytics.

TIP

As you get answers to these questions, you'll discover that some repositories of information in your company are extremely accurate, while others, not so much. Be prepared to talk to different levels of people to get an idea how trustworthy that information is. Your head of Sales Operations may say everything's a bed of roses in your CRM system, while the Account Managers may tell you they can't trust anything in it other than the amount of the deals they've won. Plan on assessing how good or bad the data is and get an idea of how it could be improved. Chances are there may be additional projects that happen during or after your cloud analytics implementation that try to improve any issues stemming at the mouth of the input valve (which those empowered to make change may only see after they start getting insights via the visualization of the analytics).

Make sure to ask how often this data needs to be collected, massaged, and analyzed because that'll determine how frequently the data gets put into a data warehouse. In some cases, getting your hands on this data once a month is sufficient for the types of reports you're running. Sometimes having the latest information daily is fine, too. But in other scenarios, maybe a difference in

hours could change your information so greatly that your business users need hourly updates.

Data coming from several systems most likely needs to be massaged a bit so it all comes together and looks like one unified view. You may hear the acronym "ETL" to describe a tool that does this — literally *E*xtracting data, *T*ransforming it, then *L*oading it into a system to be further analyzed.

In Figure 3-2, you see a visualization of information from various data sources converging into an ETL tool that nicely integrates information, cleans it up, and then loads it into a unified system for analysis.

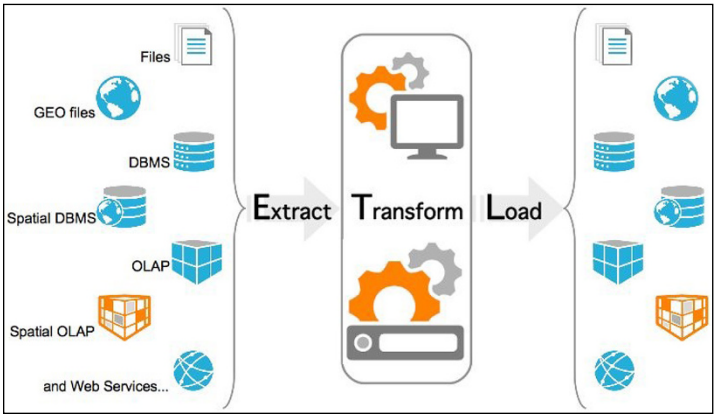

FIGURE 3-2: The flow of data into and out of an ETL tool.

Given the volumes of data being generated based on usage statistics, you may also hear the term *streaming data.* This phrase describes data streaming into a data source and immediately making itself available to stream out real-time answers. In this situation, a continuous stream of data is pouring in and analysis of that data is also happening all the time. Data volumes here are usually much larger than what's passed through a typical ETL tool. This data can come from always-on devices that generate log files, from sensors, and API feeds that are always monitoring data for any unusual activity. Think of devices like the Nest thermostat or your Fitbit device — these are examples of "always-on" devices that are collecting and delivering data.

After you've identified your data sources, know what needs to be cleaned up about that data and the frequency with which you need to analyze this information, you're able to address what services can help you do this. This is the second layer of any cloud analytics solution.

Integration experts like Informatica play a key role at this level by providing tools to help folks more quickly get access to data. In the old days, if you needed access to data in an Excel spreadsheet, a back-end financial system, and a CRM solution, one of the data warehousing experts in your company had to build custom connectors to those data stores. Then he or she would have to spend time writing scripts that took that data and massaged it (or integrated it) so it all spoke the same language before it was ready for analysis.

Now integration companies have done most of the heavy lifting by pre-building those connectors for you to the most commonly used business systems: web-based apps, internal corporate networks and corporate applications, and so on. These companies do the legwork (finger work?) of building the connections by using open APIs or other third-party connectors.

Figure 3-3 displays a graphical user interface that can help users more easily build integrations to various data sources.

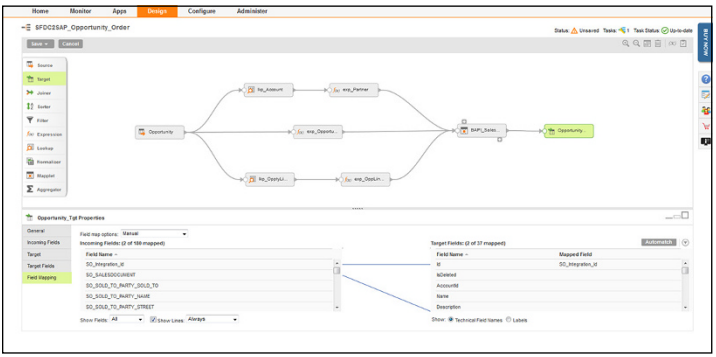

FIGURE 3-3: Informatica's cloud-based design tool.

Additionally, they've made it easier for users to still clean, validate, and enrich the data from all those various sources in order to consolidate everything to provide accurate insights. The majority of actions can be configured using intuitive mouse-clickable

choices for a friendly user interface, again reducing the time each company needs to get up and running.

These companies can also accommodate various frequencies with which you'll need this information captured, processed, and loaded for use.

Finding the Right Cloud Analytics Solution for You

Though cloud analytics has made the overall world of business intelligence and data warehouses less daunting than before, deciding on what cloud solutions to choose should still be taken with care. In this section, I discuss some questions to consider when shopping for a cloud analytics solution for your business.

Categorizing your data needs

In inventorying your data sources, you also want to identify how much of that information is structured data — that's the data that comes from organized data models and is fairly easily captured, stored, queried, and analyzed. Usually that information is stored in a database. This data also most reflects what is being queried from traditional data warehouses — it's that "what if" analysis that looks at patterns in consumption of a business's widgets.

Some cloud analytics options operate in "batch" mode, which is similar to how traditional data warehouses work. One example is Amazon's Redshift, which stores information in structured data stores and is a viable alternative to traditional data warehouses.

However, if you need to analyze real-time data streams, you need to make sure a cloud analytics solution can handle that. In these situations, a solution needs to continuously receive large volumes of data and constantly aggregate and analyze it. Processing all this data is done in real time, which means that the processing is usually lightweight to keep from bogging down any performance.

Log data, social media content, market data feeds, and web behavior can all be captured and processed in real time to monitor any deviations from the norm.

Identifying the primary user audience

Though all cloud analytics solutions market themselves as having user experiences that are easier than traditional BI tools, they still can lean toward more or less technical bents. Some tools like Amazon Redshift are geared more toward technical end-users and assume a basic developer type of skill set and thinking. Perhaps instead of keying in your own SQL query from scratch, you may now have some point-and-click or click-and-drag tools to make your fingers less tired, but having fundamental technical proficiency (to form a query, for example) may still needed.

Other applications like Salesforce.com's Analytics Cloud are geared more toward business users. The apps' configuration wizards and user experience focus more on empowering non-technical users to create their own reports and dashboards.

Familiarizing yourself with pricing models

The new world of cloud analytics also means that different vendors will have different sales and pricing approaches. Yes, cloud analytics vendors offer more cost-effective, flexible, and attractive subscription pricing plans compared with those of traditional data warehousing companies. Also remember that, with the cloud, you don't have to buy the hardware and software to maintain the systems yourself.

Some vendors charge based on various usage metrics — the amount of storage you need and network and processing power that you'll use, for example. This allows you to only pay for what you need, but you should still model out costs over the next few years. Other vendors charge based on the number of users who would be accessing the system. Think about all the people who need to access the solution every day versus employees who just log in occasionally. Does the vendor have a one-size-fits-all user licensing plan, or are there different types of user licenses (and thus, costs)?

Fortunately, cloud application vendors often provide trial periods where you can run your tests and kick the tires. As they say, the devil's in the details. Make use of the free trials and have a game plan as to what you're going to test in each app and why.

Visualizing the Data

After you've spent the time investing in getting the right data into a cloud analytics solution, make sure your visualization layer is easily usable and intuitive for its targeted end-users. Usually these will be both business end-users who want the capability to quickly create quick snapshots of data, as well as the more specialized data warehouse folks who are connecting this layer on top of the data aggregation layer that they've built.

In Figure 3-4, you see how data can be displayed to business users in beautiful, impactful ways. A picture is indeed worth a thousand words, as the old saying goes.

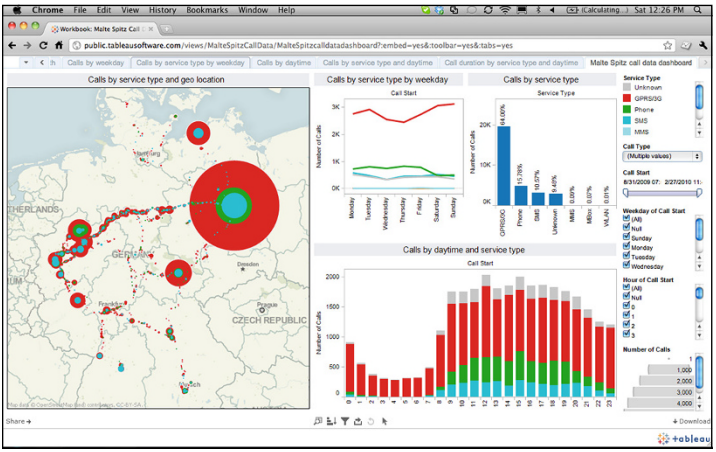

FIGURE 3-4: Visualizing data.

Make sure you have some simple queries and reports or dashboards that you want end-users to try to build. Try setting up some yourself ahead of time and see if the end-users naturally gravitate toward them and want to drill down to get more information, or if they're just more confused.

REMEMBER

Companies like Birst, Domo, GoodData, Salesforce, Tableau, and others provide this visualization layer. Again, sign up for trials whenever possible, and make sure some key business end-users are enlisted to give you feedback on how intuitive they find the app.

IN THIS CHAPTER

- Establishing common ground between big data and cloud analytics
- Understanding differences
- Identifying complementary areas

Chapter 4
Differentiating Big Data from Cloud Analytics

As you continue to learn more about the world of cloud analytics, you'll inevitably run into the term *big data.* Big data usually refers to the aggregate size of all the data that you're trying to capture, not just the size of a single data source. Similar to cloud analytics, this term is often something you hear frequently from industry thought leaders. But what does it really mean? Is it just a different way to describe what cloud analytics is about? No. In this section, I talk about where similarities exist between the two concepts, how they differ, how they complement each other, and why it matters.

Finding Common Ground

Both big data and cloud analytics solutions are anchored on key principles for how a data solution is architected. Data architecture comprises three layers:

- **Foundational data layer:** Resides across multiple systems
- **Processing/computing layer:** Collects all that data from all those systems, works to consolidate and clean everything up, and determines how frequently all this work happens

- **Visualization layer:** Takes all that data and represents it in pretty graphs and charts that make business people happy

Both big data and cloud analytics solutions assume that the data layer accommodates a variety of data sources of different volumes and sizes. Don't forget about unstructured data — it may not fit well into a traditional relational database model but nevertheless is being produced in volumes never seen before, due to innovations in digital information. Capturing unstructured data into traditional enterprise data warehouses may cause a bit of head-scratching for your team. The solution may be so time-consuming and costly that you'd be wise to assess some cloud-based options.

With cloud analytics and the capability to expand your data warehousing needs as your aggregate data needs grow, there are no hard and fast delineations between when some amount of data suddenly becomes big data and is no longer suitable for a cloud analytics solution. Big data can be regarded as just another data source to be processed and analyzed by a modern data warehousing solution.

Both big data and cloud analytics solutions are highly distributed and scalable. Distributed computing is data geek-speak for saying that a whole bunch of computer systems can be used to work on a single problem. Even better, they're all networked together so they can share information about the problem that they're trying to solve. So both big data and cloud analytics take a distributed computing approach to processing and analyzing data. Big data leverages Hadoop, while cloud analytics uses its cloud-based architecture to distribute computing costs across multiple subscribers of its service.

Hadoop is an open-source software framework that you hear often associated with big data. It allows big data to be stored and processed across low-cost distributed hardware systems. It's evolved over time now to include a software ecosystem built to handle massively parallel computing.

The nature of distributed computing allows the architecture (and thus the customers) to scale. Customers can pay as they go and add more processing power as their needs increase. Vendors can

also handle increasingly large amounts of data processing without compromising on performance.

Figure 4-1 is a visualization of how distributed computing allows

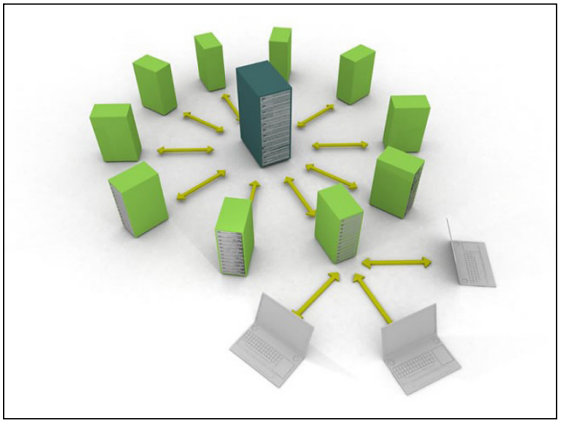

FIGURE 4-1: Distributed computing allows multiple systems to share the workload of one task.

companies to scale and add more machines as their demands grow.

Understanding Key Differences

So how are big data and cloud analytics different? There are three main differences:

- Big data has come to mean not just vast aggregate amounts of unstructured and structured data. It refers to analytic processing of extremely large data sets based on a set of algorithms called MapReduce. MapReduce is part of the Hadoop open-source ecosystem, and is a model that improves the processing time needed to support the analysis of enormous amounts of data. MapReduce was invented by Google, a company that definitely collects and analyzes a vast amount of information all the time. You often hear references to Hadoop or MapReduce when people talk about big data.

In Figure 4-2, you see a typical systems architecture for how MapReduce and Hadoop interact in an environment. Think of the Hadoop foundation as an operating system (including a Hadoop File System), on top of which sits MapReduce, which was the first way one could write programs to use this operating system.

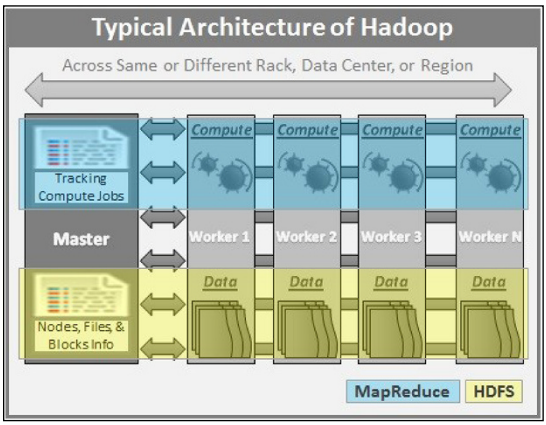

FIGURE 4-2: Understanding how MapReduce and Hadoop interact.

- With big data, the distributed storage of all that data and the computing infrastructure generally resides within a company's own data center. Cloud analytics is, well, in the cloud, and users subscribe to this service in order to outsource the infrastructure costs that go into maintaining a data center. There are cloud-based Hadoop distributions (like Amazon Elastic Map Reduce, Hadoop on Google Compute Engine, and Microsoft Azure's HDInsight), in addition to the on-premise options (provided by vendors like Cloudera, Hortonworks, Pivotal, and Teradata). As with other cloud investments, these choices are unique to your business.
- Big data users and the use cases for big data are usually highly technical in nature. You won't find a lot of business users getting into the weeds here. Think of huge companies or government organizations that need to process a ton of information — Google, Amazon, the CIA, and IRS come to

mind. If you're a business user and you're talking with a big data expert, don't try to fake it. Just do your best to listen. Bottom line, big data experts will likely be focused on the infrastructure and data architecture as a means to an end (which is cloud analytics), whereas those interested in cloud analytics put more of a priority on the potential insights they can gain from data.

Complementing Each Other

No, I don't mean patting each other on the back (that's a compliment with an "i"). In this section, I identify synergies where big data and cloud analytics can assist each other.

Because big data can process and analyze vast amounts of data, some of the processing work can be assigned to the Hadoop machines (also known as a Hadoop cluster), before moving the data into a data warehouse where further processing might occur with other less massive data sources. A cloud analytics solution can accept inputs from multiple data sources and should view big data as just another data source. The visualization layer that business users rely on won't have to change, but they could also benefit from additional input from big data.

For example, structured data from a company's CRM and financial systems can flow into the cloud analytics solution. In addition, if there's a huge amount of website visitor data that also needs to be added, the IT team may decide to first capture that big data in a Hadoop cluster, massage it, and then load that into the cloud analytics solution. Combined with the CRM and financial systems data, that information can then be presented to business users to track behavioral patterns across website actions, sales person interactions, and deal sizes.

Figure 4-3 helps illustrate the terms and relationships between analytics and big data and data warehousing.

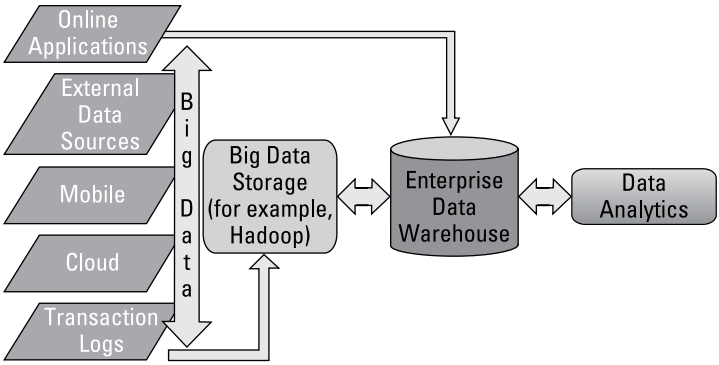

FIGURE 4-3: The relationship between analytics, big data, and data warehousing.

IN THIS CHAPTER

- Describing the Internet of Things
- Reviewing predictive analytics and machine learning

Chapter 5
Looking Forward: The Future of Cloud Analytics

As established companies face updating their enterprise data warehouses to accommodate new forms and volumes of information, and as younger companies begin exploring analytics options that used to be out of their reach, it helps to have a sense of where the world of cloud analytics is headed. In this chapter, I discuss two trends that may affect your long-term technology strategy.

Living in a Connected World

The cartoon show, *The Jetsons*, depicted a family living in the future complete with robot maids and pets, and highly automated home appliances. *Star Trek* showed Captain Kirk communicating with his crew members using a small rectangular device that folded shut when not in use. Dick Tracy, the detective, had a watch that he could talk into. What used to be just wishful thinking in comic strips and science fiction is now reality, thanks to a combination of factors that make smart devices relatively affordable and small enough to be convenient.

Toys, refrigerators, home security systems, home thermostats, cars, phones, heart monitor implants, and watches are just a few items that are designed with unique technology identifiers (like an IP address) that allow them to capture and transfer data without any human intervention. The Internet of Things (IoT) describes this world where a thing can be any piece of hardware, or even a living being, that is connected to other things via the Internet or other wireless technologies.

Figure 5-1 shows how devices and even living things can be interconnected to collect and share data.

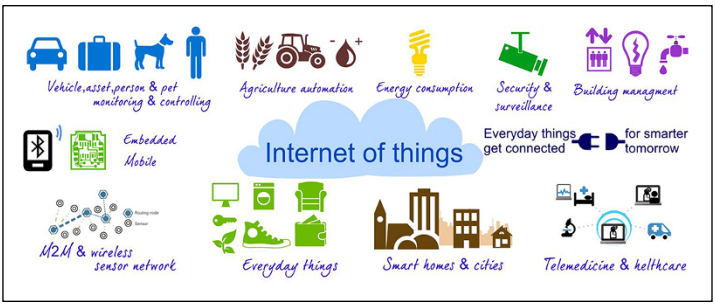

FIGURE 5-1: The Internet of Things connects everything.

Whether you philosophically agree or disagree with empowering any device to track and share data, the IoT isn't going away. All the data captured by these devices needs to reside somewhere, and in aggregate can shed additional light on correlations that can better direct product development strategies. If your company manages hardware today, what plans are there to connect it to the world? Or at least to your help center in case the hardware malfunctions?

Even if your company doesn't manufacture goods, you're probably part of an ecosystem that could benefit from new channels of data capture to drive better customer targeting. For example, if a major grocery store chain had sensors in its shopping carts and baskets to capture information about in which aisles customers spent the most time, that aggregated data could help prioritize where new product promotions should be displayed.

With the advent of the IoT, your company needs to think about how capturing and analyzing information from sensors or devices could add to insights gained from cloud analytics. Businesses today

are exploring these avenues or at least understanding their data warehouse needs in order to process all this new information. After all, the IoT is yet another data source that cloud analytics can take advantage of.

Predicting Behavior through Predictive Analytics

While cloud analytics can empower many end-users and newer companies to benefit from the power of business intelligence, it gives traditional data experts and statisticians a kids-in-a-candy-store experience.

The amount of data that can now be collected means that analysis can deliver statistically significant correlations, and as businesses capture more data over time, they can apply various statistical techniques to analyze current and historical facts and make smarter predictions about future events.

In Figure 5-2, you see the evolution of cloud analytics from its earliest modern days to the present time and beyond.

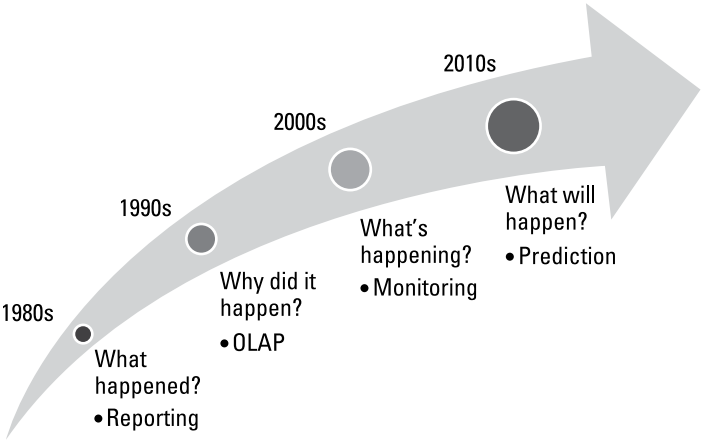

FIGURE 5-2: The evolution of predictive analytics.

Creating models, regression analysis, data mining, and all sorts of other things you should have paid more attention to in

college statistics courses will increase in demand. More than just universities, think tanks and Fortune 100 companies will need help analyzing all this information and making sure the insights derived are sound.

As machines also become more connected, they can also be programmed to learn from vast amounts of data and adapt themselves to act differently without a programmer explicitly having to get involved. "Machine learning" then allows systems to learn from data and then use that data to make predictions or decisions that cause the system to act differently.

Both predictive analytics and machine learning may sound like something from science fiction novels — but this is really happening today. Industries like media, insurance, finance, and consumer retail can all make use of predictive analytics to help their businesses better assess and predict risk and trends. With machine learning, companies that manufacture automobiles, home appliances, and medical devices could all use predictive analytics to anticipate when maintenance is due.

The field of predictive analytics and machine learning are both moving to cloud-based options where traditionally very precise calculations had to be set up in very un-user-friendly applications with great infrastructure costs. Options like Microsoft Azure's machine learning service helps data scientists of any level get started with their models sooner rather than later.

IN THIS CHAPTER

- Knowing what underlying data should be analyzed
- Preparing yourself as you research cloud analytics solutions

Chapter 6
Ten Do's & Don'ts When Evaluating a Cloud Analytics Solution

If you've gotten the green light to seriously start investigating cloud analytics solutions, make sure that the whole exercise is worth your while. The devil's in the details, and it helps to be able to engage in deeper conversations sooner rather than later. This chapter gives you some tips as you begin your journey.

Do Understand Who Wants to Measure What

What business stakeholders are having trouble finding the right insights they need to be successful with their work? Do they know what the cause could be? Try to get specific information about their challenges. If your head of sales is saying she doesn't have any insights into year-over-year win/loss ratios by competitor, that's good detail. If your marketing operations executive isn't able to identify how many average campaign touches a prospect

needs before he ends up purchasing, and thus is shooting in the dark when it comes time to build next year's marketing budget, that's also helpful. Do your Finance managers need insights into how sales revenue is allocated over a fiscal year? This information helps you suss out whether end-users need improved visibility or if there are bigger issues affecting the capture and quality of data. Wanting a cloud analytics solution to understand why reps are losing deals . . . is a little vague and could set unrealistic expectations that a technology will be the cure-all medicine to possible underlying process deficiencies.


TIP

Make sure you understand each business stakeholder's challenge, how it impacts his business, and whether the issue can be addressed with technology or an improved business process.

Do Evaluate the Quality of Your Existing Data

If you're like most companies, your sales organization probably gets gold stars for entering thorough information into your CRM system. Okay, now that we've all had a good laugh, let's admit that this is actually rarely the case. On the other hand, financial systems at your company should hopefully be accurate, and have end users who are motivated to get this data into their system correctly the first time.

As you gather a list of the various data sources that would contribute to the cloud analytics solution, you'll see varying levels of data accuracy. Having this knowledge should give you an understanding of how much processing a cloud analytics solution should be able to do to consolidate data from different systems into one happy data warehouse. For some of the systems where the data may not be as accurate, you'll want to assess whether a technology or changed business process can improve this.

Don't Ignore Your Existing Architecture

You may have nothing in place other than some Excel spreadsheets, or you may have a very complex on-premise enterprise data warehouse. Or your company could have both of those

scenarios if some departments aren't integrated with current data warehouse systems. Don't assume that a new cloud analytics solution is going to automatically replace what you already have (well, unless it's nothing). There are tons of scenarios where a hybrid approach is practical. Having an idea of what approach you'd take to gradually phase in a new solution also prepares you to ask questions about any cloud analytics solution's options for playing nicely with enterprise data warehouses or other internal corporate systems. Even if you don't have anything in place, you'll be armed with an idea of what data and departments will be affected and can prepare a phased rollout approach.

Do Compare Costs between Your Current Architecture and the Proposed Solution

Jot down some rough cost numbers for addressing your business challenges today using your current system, as well as the same costs for a cloud analytics solution. Include potential hardware costs, software costs, data center costs, licensing costs, and so on. Estimate these costs over a few years. Also take into consideration when the solution for the business users can be deployed using your current architecture versus something new.

Do Have a Plan on How to Process All Your Data

How much of the data that's needed is structured versus unstructured? Could it be structured for now but you want to be able to add in unstructured data sources in the future? Make sure all that data can be handled by the solutions you're researching. Do you have an existing ETL solution that can process all this information before sending it to another repository? Or do you also need to investigate an integration solution, too? Or both? Investigate current options and see if the modern user interfaces could help save your resources' time to manually connect to and transform data.

Do Have Some Test Scenarios Ready to Use

Have some basic scenarios ready to use as you evaluate different systems. It could be as easy as wanting to extract some opportunity records from your CRM system, and financial information from your accounting system, merging some data, and then loading it into another environment from which a select number of end-users can attempt to view or build reports and dashboards. Identify the various technical and business users that should be part of the evaluation. Make sure that what's being evaluated across solutions is consistent.

Do Take Advantage of Free Trials

Always take advantage of free trials. This gives you a way to get into the weeds and perform some basic tests without any commitments. Trials often last for 14 or 30 days, so having some test scenarios at the ready will help your team make the most of the time. Test out online help and training options while you're at it. Log a support case and see what the response is like. And, depending on the complexity of your problem and your company's brand recognition, you may even get your trial extended if you ask your sales person nicely.

Don't Skimp on Training Your Team

Even if a cloud analytics solution has helpful online tutorials and training options, you should always ask about training opportunities conducted in real-time. Often these courses are valuable both for establishing the basics and also for having a direct conduit to getting more advanced questions or scenarios answered. Usually the people doing the training have run into a lot more scenarios or gotchas than you have — they're a wealth of extra information, as are insights from fellow class attendees. Training can apply to very technical members of your team, as well as for newly empowered business users.

Do Think about the Possibilities

With the ongoing developments of unstructured data, the Internet of Things, and how increasing amounts of and access to data has reinvigorated predictive analytics and machine learning opportunities, it's not just companies that manufacture hardware that are affected by this. Maybe your company doesn't have plans to track this data in the next year, but can you say the same thing about your industry? Or your competition? What do thought leaders in your industry say? These may be early days for some companies, while others have made significant investments in capturing and mining data like never before. Consider the baby steps that your company should be taking and if your cloud analytics solution is ready to address data needs derived from these emerging fields. Understand what potential there is to unlocking access to more information.

Do Stay Current on Latest Developments

As innovation generates more data, the world of cloud analytics continues to evolve. Both old and new players in the space are competing to establish market share, carving out a space for themselves, and delivering on the vision of visualizing data from any source at any time. Things can change quickly and you should find time to network with peers and talk to vendors in the cloud analytics ecosystem to hear what's got traction. Attending conferences like Informatica World (`www.informaticaworld.com`) is critical, as you can research multiple cloud analytics vendors in one setting, while also getting educated on what tools are available to process and transform all that data.

Notes

Notes

Informatica

Unleash the Disruptive Power of Data for Critical Decisions.

You want intelligent data. The kinds of data that will give you new insights so that you can transform customer experience and engagement, or increase operational efficiencies.

When you empower your employees with the right data they can make pivotal decisions. But that data is in a multitude of applications on-premises and in the cloud.

Informatica helps you supercharge your self-service analytics with integrated data. Informatica Intelligent Cloud Services for cloud analytics provides fast, flexible, and scalable capabilities to integrate any data source. Finally, you can deliver intelligent data to your users' dashboard of choice.

Discover the disruptive power of data.
Go to informatica.com/cloudanalytics to learn more.

Unleash the power of data

Businesses use cloud analytics to discover new insights so that they can make transformative decisions. However, the data that fuels cloud analytics must be accessible, secure, and relevant. Learn how to optimize your business intelligence so that you can respond more quickly to users' demands. Discover how to capture real-time data streams and plan your architecture for cloud analytics.

Inside...

- Know what to look for in Amazon Redshift
- Get the most out of Tableau
- Do more with Salesforce Analytics Cloud
- Create a cloud analytics blueprint
- See the future of cloud analytics
- Find tips for choosing analytics tools

Liz Kao is a veteran of the enterprise software industry and the author of many Dummies books.

Go to Dummies.com®
for videos, step-by-step photos, how-to articles, or to shop!

for dummies®
A Wiley Brand

Also available as an e-book

ISBN: 978-1-119-45952-1
Not for resale

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley’s ebook EULA.