WHITE PAPER

Six Patterns of Big Data and Analytics Adoption: The Importance of the Information Architecture

Sponsored by: Oracle

Dan Vesset
June 2015

IDC OPINION

The requirements for access to and analysis of Big Data to derive value from it are evolving fast and so are the perspectives on what it takes to make Big Data effective in the enterprise. Only five years ago, the advice was somewhat adversarial and superficial:

- Download Hadoop and try it out — avoid the costly mistakes of data warehouse deployments.
- Focus first and foremost on data visualization to provide end users with self-service.
- Install an open source analytics tool and use it locally on your data.

What these and other similar "recommendations" ignored was the importance of an enterprise infrastructure, data management, and governance.

Clearly, the excitement about Big Data is justified. The body of research — both academic and commercial — showing the benefits and competitive advantage of the data-driven culture has continued to grow in recent years. But the hype surrounding Big Data must also be balanced with pragmatic advice for organizations struggling in the fast-changing environment of more technology choices, budget constraints, and limited supply of highly skilled professionals.

The reality is that Hadoop can offload data warehouse history and detail, but it will not eliminate the data warehouse. New tools should be focused on end-user productivity, but Big Data maturity doesn't mean only deploying a new discovery or reporting tool — it requires a blend of descriptive and predictive analytics, business expertise and a supportive organizational culture, modern technology and the ability to create bridges to existing technology and, importantly, a business need and an architecture.

In short, Big Data technology alone is not an architecture. If not handled with care, Big Data can easily add to the complexity of a corporate IT environment. And that is not the purpose. Big Data is merely the latest aspect of a comprehensive and, hopefully, integrated enterprise-class information management capability. Having a Big Data strategy and architecture does not mean that there should be an attempt to deploy a single enterprisewide project, nor does it mean that the architecture locks the organization into a chosen set of technology for the long term.

Emerging best practices clearly point to the need for a Big Data strategy and architecture that promotes incremental deployments and agility in development and adoption of new technology components. As more companies progress along the Big Data and analytics maturity curve, leading organizations are differentiated by doing just that, balancing stability and agility by providing controlled empowerment to business analysts, data scientists, and all the other decision makers in the organization.

These organizations ensure that their capabilities in people, technology, process, and data are aligned to maximize positive outcomes. They value not only their data scientists but also their Big Data and analytics architects and developers. 60% of large organizations report that the lack of data scientists hinders Big Data and analytics project success, but the same percentage of organizations cite a lack of business intelligence (BI) and analytic applications developers, and 45% cite a lack of staff responsible for designing, deploying, and managing the information management software platform. There are thousands of Big Data or data architect positions open at organizations just in the United States.

The Big Data architecture matters because it enables organizations to:

- Translate the vision and business goals into incremental projects that characterize an agile approach to development
- Be proactive in using Big Data technology and all the relevant internal and external data to drive better decisions or to monetize the data externally
- Provide a balance between long-term stability of Big Data assets and the need for innovation to leverage the latest and greatest technology, people, processes, and data

IN THIS WHITE PAPER

In this white paper, IDC describes lessons learned from interviews and surveys of organizations engaged in Big Data initiatives and the patterns of adoption they have followed to expand existing or initiate new Big Data projects to create value for their organizations. The document highlights the importance of the Big Data architecture to drive improvements and innovation in customer interactions, operational efficiency, and compliance and risk management, among a wide range of business goals and desired outcomes. This white paper utilizes previously published IDC research frameworks such as the IDC Big Data and Analytics Opportunity Matrix and the IDC Big Data and Analytics MaturityScape.

Finally, this white paper highlights Oracle Corp.'s Big Data architecture, technology, and services, as well as Oracle customer examples utilizing these offerings.

Throughout the white paper, we use the term *Big Data.* It is not meant as a reference to a specific volume, variety, or velocity of data. Instead, the idea is that Big Data use cases, whatever they are, are "all inclusive" of supporting technologies and processes across the information architecture, such as data infrastructure, integration, quality, life-cycle management, analytics, business intelligence, security, and analytic applications.

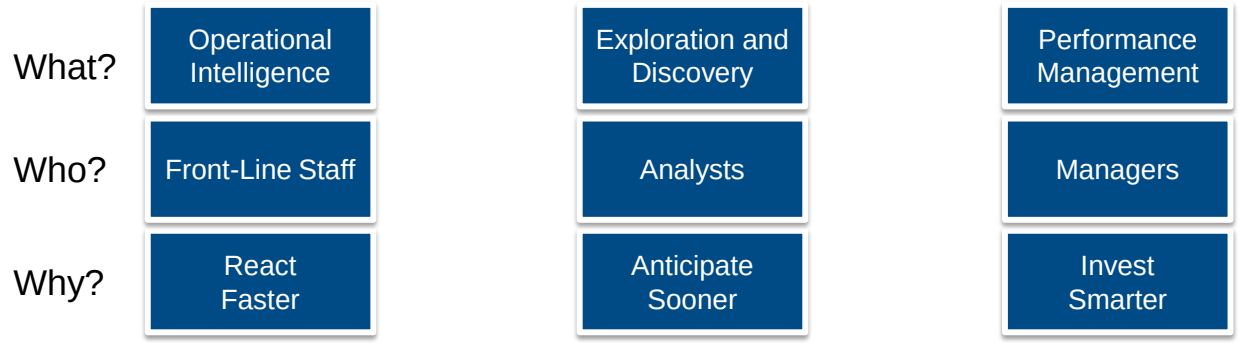
FROM STRATEGY TO EXECUTION

Big Data has become a top agenda item for a growing number of executives. At the same time, hype about technology and inflated promises of outcomes abound – both of which ignore the real challenges faced by organizations of all sizes. Managers in many organizations believe they have what it takes to harness the power of Big Data for improving data-driven decision making, yet they are missing the competency to address the range of technology, staffing, process, and data requirements involved. With the opportunity to unlock the value of Big Data to accelerate innovation, drive optimization, and improve compliance using data-driven decision making comes the need to demonstrate value, navigate expanding technology alternatives, change business processes, and ensure the availability of skilled staff.

One of the building blocks of Big Data solutions for any organization is the Big Data architecture, which needs to account for a mix of technology to address three primary use cases, as shown in Figure 1.

FIGURE 1

Common Big Data Use Cases, User Personas, and Motivations



Source: IDC, 2015

Big Data Use Cases

Data, analytics, and decision making are not the exclusive domain of data scientists. Everyone from executives to line-of-business staff makes decisions. Some are strategic decisions; others are operational or tactical decisions. IDC identifies three primary Big Data use cases with associated user personas and motivations for deriving value from Big Data:

- **Operational intelligence.** With a focus on real-time operations, this use case addresses the high-velocity attribute of Big Data, with streaming data technology and event processing to enable decisions "on the fly." This use case is driven by the goal of creating sense-and-respond operational execution: monitoring for specific events or listening for predictors of an impactful change and then orchestrating a rapid and appropriate response. Architecturally, we can think of data flowing simultaneously along two paths: real-time pipelining to facilitate real-time operations and batch-oriented processes that capture and stage data at various states.

- **Exploration and discovery.** Exploratory work looks to discover signals and nonobvious relationships or patterns among broader data sets to uncover new insights that can positively impact decision making and organizational performance. Once discovered and validated, newly discovered signals are operationalized to improve operational intelligence and monitored for ongoing quality using performance management processes. Discovery can be visual, collaborative, and algorithmic. High-volume data sets require rapid pivoting for exploration — sometimes orthogonally connected. Discovery requires data processing optimization, flexible schemas, statistical modeling, fast processors, and lots of memory.
- **Performance management.** This use case addresses strategic decisions about past performance as well as planning functions. This is the traditional domain of the data warehouse for descriptive analytics and reporting. Adding Big Data can increase the timeliness of business reporting and nuanced precision with larger data sets and new data sources. Examples include the addition of reputation risk assessment, based on social media monitoring, into broader risk metrics or the move toward real-time performance management.

A modern Big Data architecture needs to ensure that all relevant technology is made available based on the unique requirements of the aforementioned use cases as well as the connections among them.

Although the three use cases represent a high-level categorization of dozens of specific opportunities to use Big Data solutions, they highlight the reality of a need to address a growing number of requirements, technologies and, most importantly, *change* that is affecting organizations of all sizes. IDC research shows that over the past two years, 60% of organizations have started using new analytic techniques, 33% have started using new data types, and 32% have introduced new metrics or key performance indicators (KPIs). Decision makers and analysts at these organizations are asking new questions and questioning the status quo. Their expectations for the type of interaction they should have with data have changed, and organizations better be ready to address these expectations with a modern Big Data architecture.

That is still not the case today for the majority of organizations. IDC research also shows that only about 10% of employees from across organizational levels are fully satisfied with the Big Data technology resources available to them to support analysis and decision making. At best, this percentage grows to 30% when those "somewhat" satisfied are added into the mix. This leaves 70% of users who are complaining — either vocally or with their wallets.

Ensuring that an enterprisewide Big Data approach supports the wide variety of business needs requires the following:

- A Big Data strategy and architecture that describes the organization's vision for the role that data, analytics, and technology can play in enabling a data-driven decision-making culture
- Big Data technology that is highly performant, scalable, secure, and governed and that supports the stated design
- Supportive organizational culture and human resources, including executive focus on data-driven decisions; availability of internal or external analytic, data management, development, and maintenance skills; and data, analytics, and decision-making governance policies and procedures
- A collaborative approach between central IT, line-of-business, and analytics groups, where the value each group contributes is based on the group's core competency, coordinated across the enterprise using clearly articulated policies and procedures
- An appropriate — agile — development approach that recognizes the core competencies of the various stakeholders in the broader technology development process

The right Big Data architecture is formed or limited by the business context, where the top-level business objectives remain stable but the means to achieve these objectives are different. For example, retailers have focused on and continue to focus on merchandizing and price optimization; telecommunications firms have focused on and continue to focus on customer churn prevention and network optimization; and healthcare providers, educational institutions, and government agencies have focused on and continue to focus on providing services to their constituents — yet how they do it or can do it today is vastly different from the common practices of just a few years ago.

With a plethora of emerging information management technologies, it is easy to be focused on the technology itself; entranced by a shiny new object — and search for problems to solve. Of course, new technologies often solve problems that were never solvable or affordably solvable before. However, the better path to embrace new technologies is to always stay close to the business context — whether strategic, operational, or tactical. Planning for Big Data is not just about mastering new technologies but, more importantly, about establishing an architecture vision that is aligned with the business' analytical needs.

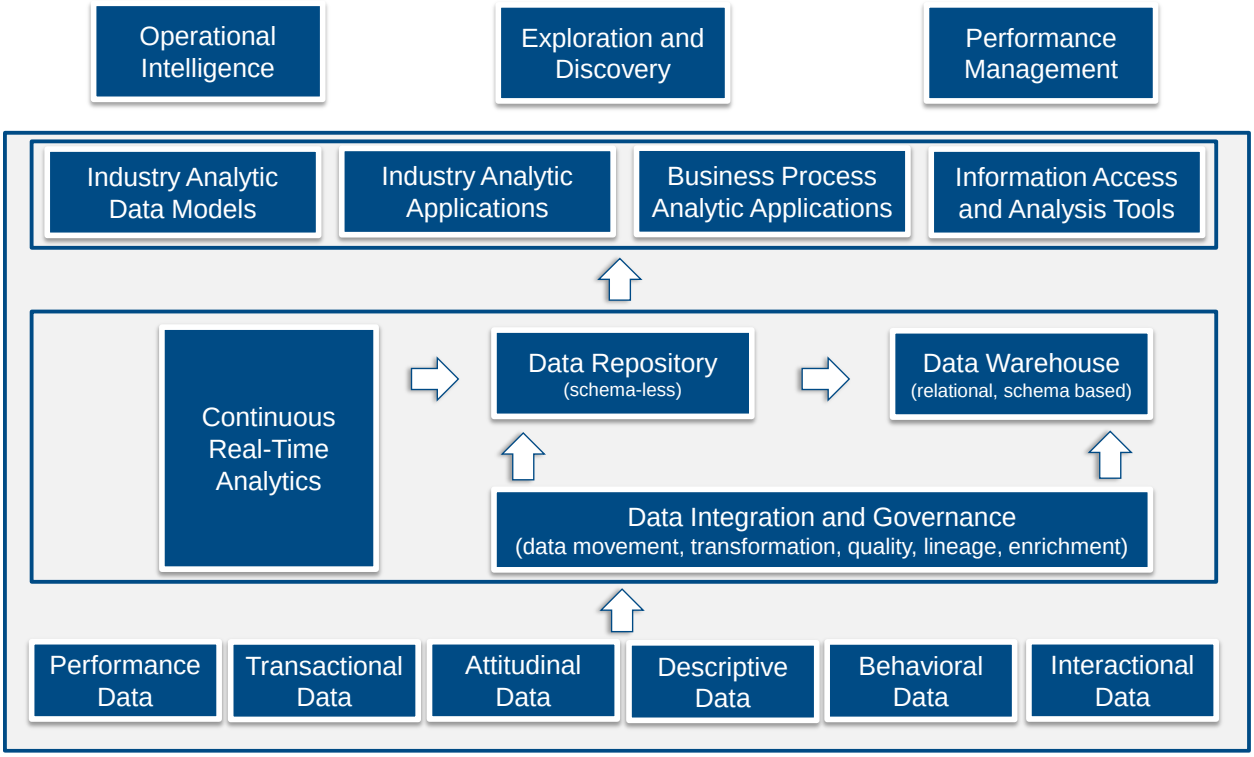
A conceptual view of the overall Big Data architecture is shown in Figure 2.

Most organizations today have multiple data warehouses, data marts, data caches, and operational data stores. One of the reasons why many of these organizations struggle to deliver value from data is because of the number of possible integration points among the number of different data management and analysis technologies. Where to start or how to proceed is not always clear. The hype surrounding Big Data in recent years has spread an increased sense of confusion about the technology and the process for deriving value from data.

But even as some organizations struggle to deliver value from Big Data, others have experienced success in leveraging Big Data to their competitive advantage. What follows are the lessons learned from some leading North American and European companies as they have charted new Big Data architectures, adopted new Big Data and analytics technology and, as a result, reaped business value in the form of higher revenue, lower costs, and increased productivity. Each of the six organizations has experienced a unique pattern of Big Data solution adoption guided by existing information architectures.

FIGURE 2

Conceptual View of the Big Data Architecture



Source: IDC, 2015

Patterns of Adoption

Just as Big Data does not mean a single technology or a single use case, there is no single path to start or expand an existing Big Data architecture. Some organizations may want to supplement their relational data warehouse with a Hadoop-based data store (a data reservoir or data lake — whatever your preferred metaphor), while others may want to deploy streaming event analytics. Some may want to focus on a narrowly scoped data exploration project, while others may want to operationalize their findings based on data-driven analytical models. There is no one right path.

Six Big Data use cases at organizations across industries that illustrate various architectural approaches for modernizing their information management platforms are discussed in the sections that follow (see Table 1).

The use cases differ in terms of goals, approaches, and outcomes, but they are united in that each company highlighted in the case studies has a Big Data strategy based on clear business objectives and an information technology architecture that allows it to stay focused on moving from that strategy to execution.

TABLE 1

Patterns of Adoption

Case	Industry	Motivation	Scope
1	Banking	Transformational modernization	Transform core business processes to improve decision-making agility and transform and modernize supporting information architecture and technology.
2	Retail	Agility and resiliency	Develop a two-layer architecture that includes a business process–neutral canonical data model and a separate layer that allows agile addition of any type of business interpretation or optimization.
3	Investment banking	Complementary expansion	Complement the existing relational data warehouse with a Hadoop-based data store to address a near-real-time financial consolidation and risk assessment.
4	Travel	Targeted enablement	Improve a personalized sales process by deploying a specific, targeted solution based on real-time decision management while ensuring minimal impact on the rest of the information architecture.
5	Consumer packaged goods	Optimized exploration	Enable the ingestion, integration, exploration, and discovery of structured, semistructured, and unstructured data coupled with advanced analytic techniques to better understand the buying patterns and profiles of customers.
6	Higher education	Vision development	Guarantee architectural readiness for new requirements that would ensure a much higher satisfaction level from end users as they seek to leverage new data and new analytics to improve decision making.

Source: IDC, 2015

Motivation #1: Transformational Modernization

More regulation, faster decision cycles, new interactions with customers, and a legacy IT architecture – these are just a few of the daily business pressures and technology issues facing financial services firms. The first organization we studied was a major European bank that recognized the need to transform a number of its core business processes to improve decision-making agility, and in doing that, the organization saw the need to transform and modernize its supporting information architecture and technology.

Approach

To achieve a number of its new business goals, the bank recognized the need to become a much more data-driven organization. The bank referred to this new initiative as its "data first" strategy. It also recognized this strategy was not achievable with its existing architecture – largely because much of the key data required was stored in a mainframe and was impossible to access in an agile way.

The bank engaged Oracle because of Oracle's Big Data architecture development experience. Oracle's Innovation Workshop methodology created an environment that allowed business and IT to align the bank's interests while focusing on the bank's transformational agenda. The first phase of the process involved working with the bank's business and IT stakeholders to define and prioritize key business challenges. In the second phase, the "as is" state was assessed for each business use case. In the third phase, an initial solution architecture was developed along with a deployment plan. By using workshops in this way, the bank was able to drive consensus and momentum into the project, which would have been impossible in any other fashion.

- Customer experience was one of the core themes that flowed from the first workshop, and it quickly became clear that existing solutions were not going to meet much-needed requirements, most notably because of the lack of agility in business intelligence processes and technology. For example, it typically took six months from initial request by the business to the deployment of a new key performance indicator at the bank. There was also a lack of any self-service capability to allow analysts to upload local market data so it could be analyzed.
- The need to support stress testing for regulatory compliance was also seen as critical at the bank — while in the past, the data the bank was required to supply to regulators was very limited. This is no longer the case, because agile access to all the data in the bank became a must-have to supply timely answers to auditors' questions.
- Yet another challenge faced by the bank was the lack of guaranteed business continuity because of subpar high-availability and disaster recovery (HA/DR) capabilities. Like many banks, the data required to meet the real business need was "locked up" in a mainframe, which while highly reliable required offloading of data to make the data usable and actionable. These offloads were done to populate many downstream siloed data marts for fraud, risk, marketing, and other functions. These data marts lacked interoperability and created unnecessary maintenance costs because each data mart had its own development, support, and management teams. Business users and analysts often bypassed IT to gain access to their data with a range of tools that further exacerbated the technology maintenance challenge. To add urgency to the situation, the bank required some new capabilities to be in place within four months — this would have been impossible using its existing solutions, so a new approach was required.

The outcome of the third workshop was a new architecture that called for:

- A Hadoop-based data repository to manage the full spectrum of data required by the bank, including data incrementally fed from the mainframe
- An enterprise information store using the relational database to manage the structured enterprise data as well as the aggregated data from the data repository
- Additional reporting and discovery tools

To simplify and speed up the solution deployment to meet the bank's aggressive timelines, the bank chose to use Oracle's Engineered Systems, including Oracle Big Data Appliance (BDA) for Hadoop, Oracle Exadata for the database, and Oracle Exalytics for the analytical and reporting components.

The bank recognized that it did not have skills on the Hadoop platform to manage the onboarding of data and also that the tools and techniques on Hadoop were rapidly changing. To manage around these issues, the bank adopted Oracle's Data Factory concept. The Data Factory abstracts above individual ETL tools to allow users to define ETL flows onto and between Hadoop and the enterprise information store, which are then managed and controlled through the Data Factory. In this way, the bank was able to rapidly enable a number of completely new data sources from branch office logs, specific complex mainframe data and sensors data that count visitors to bank branches, and social media data.

Benefit

The bank is now able to form a consolidated view of all relevant data. Changes that previously took months to complete are now deployed in days. To support the data discovery needs of the newly appointed chief data officer and an expanded team of data scientists, as well as business analysts, the bank is now able to rapidly provision data discovery "sandboxes." Business analysts have access to data through an expanded fit-for-purpose software portfolio, which includes Oracle Endeca Information Discovery and Oracle Business Intelligence Suite. The bank attributed much of the success of the new information management approach to Oracle's Innovation Workshops and the Data Factory approach to data integration and governance.

The "appliance"-based approach was also an important consideration. By using prebuilt building blocks, the bank was able to develop and deploy the solution quickly — in less than four months, the bank was able to go from the initial concept to a full implementation of the initial project phase. In this way, the bank has created value and additional momentum, transforming what it does.

At a time when other organizations have been talking about Hadoop and wondering which open distribution to use, this bank implemented an operational system including a full disaster recovery site. The inclusion of new HA/DR capabilities in the design of the overall new solution was critical to the bank because it wanted to ensure the new system had the same high reliability that the bank was used to getting from the mainframe.

Motivation #2: Agility and Resiliency

Retail businesses not only are extremely competitive but also are challenged with low profit margins and tightly managed expenses and staffing levels. The need to continuously improve the customer buying experience is paramount. To that end, retail companies depend on comprehensive analytics to make smart decisions and increase customer loyalty and profit margins.

In this environment, IT needs to ensure that architecture decisions persevere over time — in terms of relevancy and to minimize disruption as business conditions inevitably change. An ad hoc approach to new Big Data technology deployments may address isolated business needs in the short term but will result in greater IT complexity and cost as well as a disconnect from broader, enterprisewide business needs.

Approach

One such organization is a United Kingdom–based retailer that set for itself the challenge of designing a flexible data warehouse that could support its ever-changing needs. The original motivation for the project came from an overstocking issue that started to have a material impact on share price. However, while the retailer knew that it urgently needed to get a handle on the core information in the business, such as sales and inventory, it also knew that changes already in flux to core systems, such as its retail management and store planning systems, were going to make design very difficult.

The retailer adopted Oracle's reference architecture for Information Management to manage this change issue. In fact, the retailer not only adopted the reference architecture but substantially enhanced it by contributing its best practices. As there is clearly a difference in the rate of change between the core data contained in the data warehouse and the reporting needs to meet the ever-changing business process requirements, the reference architecture separates these components into two separate layers:

- In the first layer, data is stored in a canonical model that is business process neutral.
- The second layer adds an access and performance layer that allows any type of business interpretation or optimization to be added quickly.

Through this approach, even though almost all the operational applications that created data were changed within a short time, the core business model remained unchanged. As mentioned previously, the business was also rapidly evolving and adding new analytic applications such as a stock forecasting system to support new business processes. These requirements created new analytic and reporting needs, which could also be quickly met through changes to existing structures or through the addition of new structures. These changes are simplified because they are all based on the underlying data in the canonical model.

Over time, the canonical model has also been able to accommodate new data originating from a wide range of additional sources as the retailer expanded its operations with new product ranges, suppliers, and channels. The business also added a store loyalty card system with all that entails. Even these changes have been readily included in the data warehouse, which is also now used to support real-time operational queries from the company's call center operations and customers viewing their account details online.

An important concept maintained by the data warehouse team has been the delivery of a "single version of the truth." With the extension of the data warehouse to include query support for operational applications, the team has moved on to now talk of a "single version of the question." That is, exactly the same query should run over the same data to yield the same result, regardless of the BI or operational tool used.

As the company has expanded into the use of mobile apps for in-store shopping and the use of beacons, the resulting volumes of data have required the company to deploy new types of information management technologies, such as Hadoop and NoSQL. The fit-for-purpose strategy for each data structure is an important operational decision, but what is more important is that new data stores and structures can be easily incorporated into the existing integrated enterprise data model. As it turned out, the retailer's mobile applications are supported by four operational data stores (relational, NoSQL, Hadoop, and the file system). However, the new data structures for discovering marketing opportunities, assessing customer sales history, and evaluating marketing campaign results were aggregated into the canonical model of the company's data warehouse without any issues.

Benefits

At the height of the Big Data hype curve, in 2015, we have heard proclamations of the death of the data warehouse. Behind this shortsighted assertions are two faulty assumptions. One assumption is that a data warehouse is a static form of a database designed at a point in time for a given data set. The other assumption is that the concept of a data warehouse based on a relational database has no place in today's world of Big Data where the schema-on-read approach to data ingestion and management will triumph over all other methods.

With regard to the first assumption, while data warehouses can be designed without flexibility, they don't have to be. Having a persistent canonical data model to rely upon was the key to ongoing flexibility at this retailer. With regard to the second assumption, a modern Big Data architecture presumes that specialized technologies excel at given requirements. In the case of Big Data, high-volume data ingest and storage have different requirements than high concurrency query design. Analytical platforms serve many purposes, and it is more likely that the two environments work better together than one replacing the other.

The benefits of the resilient architecture and value proposition of proper design at this retailer are lower total cost of ownership and flexibility in supporting new data sources, new processes, and new systems. Although much of the new data and most of the new systems were not envisioned originally, the data model at this retailer had been designed in a business-neutral way, resulting in minimal need

for its modification over many years. This has allowed the business and IT to focus on the things that matter — driving insight and commercial value from the retailer's data. This retailer is a perfect example of the change that a growing number of organizations are experiencing in this era of Big Data — the retailer has started using new data sources and types, introduced new analytics, and started to use new metrics and KPIs. All this has been made possible by the combined attributes of long-term resiliency and agility of the information architecture.

Motivation #3: Complementary Expansion

In this case, the goal of the organization was to complement the existing relational data warehouse with a Hadoop-based data store to address a near-real-time financial consolidation and risk assessment.

Approach

One large investment bank set out to address its needs for a new financial consolidation and risk assessment process — the old one being plagued with latency and often an inability to access raw data for compliance reporting. There is a critical need at the bank for low latency in financial consolidation, regulatory reporting, risk profile assessment, and scenario modeling, which required incorporation of large volumes of data from many hundreds of systems across the company. Among the issues facing the bank was the inability to use the data warehouse as the system of record for original, raw data for regulatory compliance because the data in the data warehouse was cleansed or, in some ways, transformed as part of the load process.

The bank needed to modernize its architecture by extending the capabilities provided by the existing relational data warehouse. It adopted Oracle's Information Management reference architecture and deployed a Hadoop-based raw data store as a landing space for a variety of data from a broad range of applications. From there, relevant data is pushed into an Oracle Exadata-based data warehouse, including a range of summaries that are used for downstream reporting applications. To tighten the link between the relational data warehouse and Hadoop, the bank adopted Oracle Big Data SQL, which allows queries to join data across both platforms efficiently, instead of having to move data to a single platform. As part of this architecture, Hadoop (HDFS) is now used as the system of record for original, raw data. The bank originally adopted MapReduce to transform data for consumption by downstream processes. While the need for transformation has been substantially reduced through the use of Oracle Big Data SQL, some requirements still remain. The bank realizes that maintaining millions of lines of MapReduce is not sustainable in the longer term and is currently investigating the use of Spark.

As part of introducing Hadoop into the organization, the team has had to learn a lot about this open source technology ecosystem. One of the challenges has been the frequent change of Apache open source projects compared with similar commercial software. As a result, the bank is revisiting its Big Data open source technology decisions every six months — not a trivial task, and one that has required an internal organizational adjustment to constant change.

The logical and physical architecture decisions at the bank are the responsibility of its architecture board, including the chief architect, domain architects, and architects from Oracle. The board also reviews ongoing change requests and recommendations from internal business users and IT.
One of the primary goals for this board is to correctly identify the optimal purpose for each component of the architecture. "What is this Hadoop cluster used for? When and why should data move into a data warehouse? Which open source technology should be used, and when?" are some of the many questions facing the architecture board. When there have been internal "battles" between the proponents of relational and nonrelational technologies — because of historical preferences, differences in staff skills, or

the lack of experience with specific technologies — the bank's architecture board has relied on a quantifiable assessment methodology that reviews each technology's fit for purpose based on variables such as amount of data, types of queries, availability, latency, and recoverability. This approach has taken "gut feel" and emotion out of internal debates and resulted in a better understanding of the value of both relational and nonrelational technologies within the bank's information architecture.

Benefit

With emerging technologies, you need to be agile in order to reap the benefits offered while also staying as true as possible to a set of base architectural principles to avoid unforeseen consequences that will most likely lead to additional complexity and cost over time.

Let's face it, there is a real difference between a "legacy culture" in an organization and a "legacy architecture." Organizations need to experiment and take some risks with nascent technology in order to move forward and benefit from new developments, but that should not mean forgetting about the architectural basics.

Not everything turns out as anticipated. By taking an architectural approach and first building a logical architecture, rather than starting from a technology view, the bank has been able to maintain a stable solution delivering tangible business value in the face of a changing physical architecture. To protect yourself, design a logical architecture distinct from the physical architecture and determine a means by which architectural and technology choices can be evaluated (dispassionately) — in the end it can save the day.

Motivation #4: Targeted Enablement

In this case, the goal of the organization was to improve its personalized sales process by fully utilizing real-time and historical data to improve operational decisions. The organization was looking to ensure that it could deploy a specific, targeted solution based on real-time decision management while ensuring minimal impact on the rest of the information architecture.

Approach

Big Data is not only about large volumes of social media data or Internet of Things (IoT) data. For some organizations, the key to deriving value from data is enabling real-time data monitoring, analysis, and decision recommendations. Use cases amenable to technology with these functionalities include those that fall into the operational intelligence category and involve a large number of tactical decisions that can benefit from automation. When making a large number of decisions, even a small improvement in outcomes can have a significant impact on the bottom line of your company.

One customer that needed to improve its personalized sales process is a cruise ship operator. The company wanted to increase its shore excursions and onboard goods and services sales. The company carries over 1 million customers each year, and therefore, the decision-making process to address the personalized preferences of customers could be achieved only through automation — not only because of the sheer number of decisions but also because the company was not in a position to hire a large number of data scientists.

The company turned to Oracle Real-Time Decisions (Oracle RTD) as a solution for its operational intelligence needs. This decision management software product combines a range of functionality, including user-defined business rules management, automated segmentation, and real-time predictive analysis.

The company used Oracle RTD to evaluate data from onboard purchasing, reservations, and CRM (by Oracle Siebel) applications. The software assesses a customer's history, transactions, lifetime value, demographics, and customer service information. It uses this historical information in combination with real-time data feeds and analysis of the customer's current activity on the company's Web site or with the call center to optimize recommendations that balance purchase propensity with a range of other important factors. While this may seem simple, the key factor for the company was Oracle RTD's ability to adapt decision in real time to ensure every seat is filled on all excursions, not just the popular excursions, as well as manage the time taken to handle each call in the call center and simplify navigation on the Web site. With Oracle RTD, the company is able to provide its customers with better service while maximizing revenue and minimizing costs such as the number of call center agents required.

Benefit

These and other similar decision management cases focused on operational intelligence highlight the need for the information architecture to address not only volume but also velocity of the wide variety of Big Data. In this Big Data technology adoption pattern, the deployment of real-time analytics and decision management software was highly targeted to a specific business process and complementary to the existing relational data warehousing or non-relational data store technologies.

Similar examples exist across industries. For instance, a consumer high-tech manufacturer has been using Oracle RTD to automate upsell and cross-sell recommendations. It uses these recommendations to drive content on its direct sales Web site as well as within its 20,000-employee call centers. After the initial deployment of the software, this company experienced a 10% increase in the close rate among call center agents supported by the outputs of Oracle RTD in the first year.

Motivation #5: Optimized Exploration

In this use case, the goal of the organization was to enable the ingestion, integration, exploration, and discovery of a wide variety of structured, semistructured, and unstructured data coupled with advanced analytic techniques in an effort to better understand the buying patterns and profiles of its customers.

Approach

It's not always easy to make a case for a new data exploration project and investment in advanced and predictive analytics. On the surface, such investments make a lot of instinctive sense, but unlike making an ROI case based on variables such as lower-cost storage, greater insight is not something that's easily translatable into increased revenue, lower costs, or better productivity or innovation. With this use case, IT had to go to the business and find real business problems — and that's a scary thing for many IT groups.

At a consumer packaged goods (CPG) company, architects facilitated such an interaction between IT and the business to address new questions about what, when, where, how, and why consumers buy the company's products; how to influence purchasing patterns; and what the profile of a consumer is. These are not new questions in the CPG industry, but how they are answered has changed significantly with the availability of new data and new technologies.

The company had a lot of new and traditional data sets available for analysis, but the data was not integrated and often was siloed on dedicated systems. The company had looked at integrating unstructured data with its structured data in the data warehouse for many years, but it was a daunting challenge, with the data structures of the unstructured data often changing. In addition, it typically could take up to six months to add new data types to the data warehouse, and the variability of new

data sources from social media, Web sites, and other unstructured and semistructured data sources made it nearly impossible to integrate with the data warehouse.

The architecture team worked with the business to understand the types of business questions the business typically needed to answer and the data sources the business typically used for analysis. As a result, the architecture team identified new analytic opportunities the business needed to address to provide better insight about customers and to provide a competitive advantage for the business.

To address the business requirements, the company deployed an Oracle Big Data Appliance and loaded and integrated over 25 new and traditional data sets, which had never been done before at this company. Because no complex data modeling or ETL processes were required, the company was able to integrate the data sets in a matter of weeks, with the added benefit of being able to quickly load and integrate new data types as needed. In addition, because of the availability of the large volume of storage with the Oracle BDA, the company was able to effectively load much more granular data sets, which were typically measured in terabytes and were considered to be too expensive to load to the data warehouse because of storage costs.

With all the key data sets loaded and integrated, analysts were able to run a variety of advanced and predictive analytics against the data using a large variety of analytic techniques including R-based clustering, correlations, regressions, time series, and linear forecasting analytics, to name a few. As the data scientists and business analysts explored the data, new patterns and relationships emerged in the data that had never been seen before, and new questions began to surface. In an effort to analyze and explain the emerging patterns and relationships, data scientists, business analysts, architects, and the business were all brought together to explain the results in terms the business could understand, consume, and act on.

Among the organizational and individual behavior challenges this CPG company had to overcome were linguistic differences and communication challenges between data scientists and line-of-business staff, especially around "translating" outputs of analytics for business. The data scientists tended to speak in terms of standard deviations, correlations, and regressions, which the business did not always understand. The business analysts were able to bridge some of the gaps, but in the end, the data scientists had to learn to speak the language of the business and the business needed to learn the language of analytics to translate the analytics into actionable insight.

Benefit

The CPG company not only gained a valuable new technology solution to support its data exploration and discovery needs but also established new processes and collaboration methods to support the same. It was able to incorporate previously unused internal data as well as external social, consumer, and other structured and unstructured data sources to ensure that all relevant data, big or small, was available to analysts and line-of-business decision makers through a technology solution that supports highly variable, ad hoc analytic workloads.

Motivation #6: Vision Development

In this case, the goal of the organization was to use an architectural development process for Big Data in order to expand from an existing data warehousing environment to a next-generation data warehouse in light of opportunities arising from the availability of new data types and sources. The IT group wanted to understand what the impact of this new data and processing would be on its information architecture. IDC research shows that only 8% of business users are completely satisfied with the speed of the IT group's

response to their Big Data and analytics needs. The university's IT group wanted to be ready for new requirements to ensure a much higher satisfaction level from its users. In other words, it was looking to be proactive in delivering value to its internal customers and needed guidance to get there.

Approach

Not every organization — whether commercial or nonprofit — has the resources and the expertise to redefine its architecture to rapidly address the evolving data management and analysis needs. Many organizations turn to external parties, including consulting firms and technology vendors, for help. One of the sources for information architecture design is Oracle's team of enterprise architects who have developed and rely on the Oracle Architecture Development Process (OADP). Oracle provides these resources to many organizations in order to help them get the most out of Oracle technology and to integrate Oracle and other technologies into effective solutions to meet organizations' long-term goals.

One of the organizations that turned to Oracle for just such assistance was a United States–based university. It already had a very mature data warehouse implementation, and its IT leadership was considering additional improvements to the university's information architecture. These improvements included taking advantage of Big Data sources both within the university and from external sources, as well as using new Big Data discovery tools to explore the data. One use case included collecting student sentiment data from social media Web sites.

The university characterized its existing information architecture as being technology driven rather than being business case driven. In this case, the IT group was ahead of the business side of the organization and wanted to start using Big Data as part of expanding its data warehouse capabilities.

The university engaged with Oracle enterprise architects to define the "business" context, articulate its architecture vision, assess the current state, define the future state, and develop a road map and change management process. Oracle's architects spent five weeks learning about the university's current state and then prepared the architecture recommendation. It's important to note the architecture in this case refers not only to the technology architecture but also to the business architecture, application architecture, and information architecture. The latter, which is the focus of this white paper, included the definition of:

- Business capabilities, such as data-driven decision making and unified information access, tied to organizational goals of improved effectiveness, quality at scale and rapid innovation through shared services, data governance, and analytics
- Information capabilities, such as data security, business intelligence, data integration, data stewardship and quality, optimized information platform, and multistructured data
- Transformation goals, such as providing advanced analytics on new data types and sources, more pervasive use of the solution throughout the university, and more integrated and cross-functional information

As part of the process, the team developed the conceptual, capability, and logical views of the information architecture, which incorporated current and proposed technology from multiple vendors and defined batch, real-time, and microbatch data flows.

Benefit

As a result of engaging in a structured enterprise architecture design process, the university was able to ensure a tight linkage between its proposed technology and information architectures and business goals. It came away with an iterative road map with 1- to 18-month goals that build on each other and

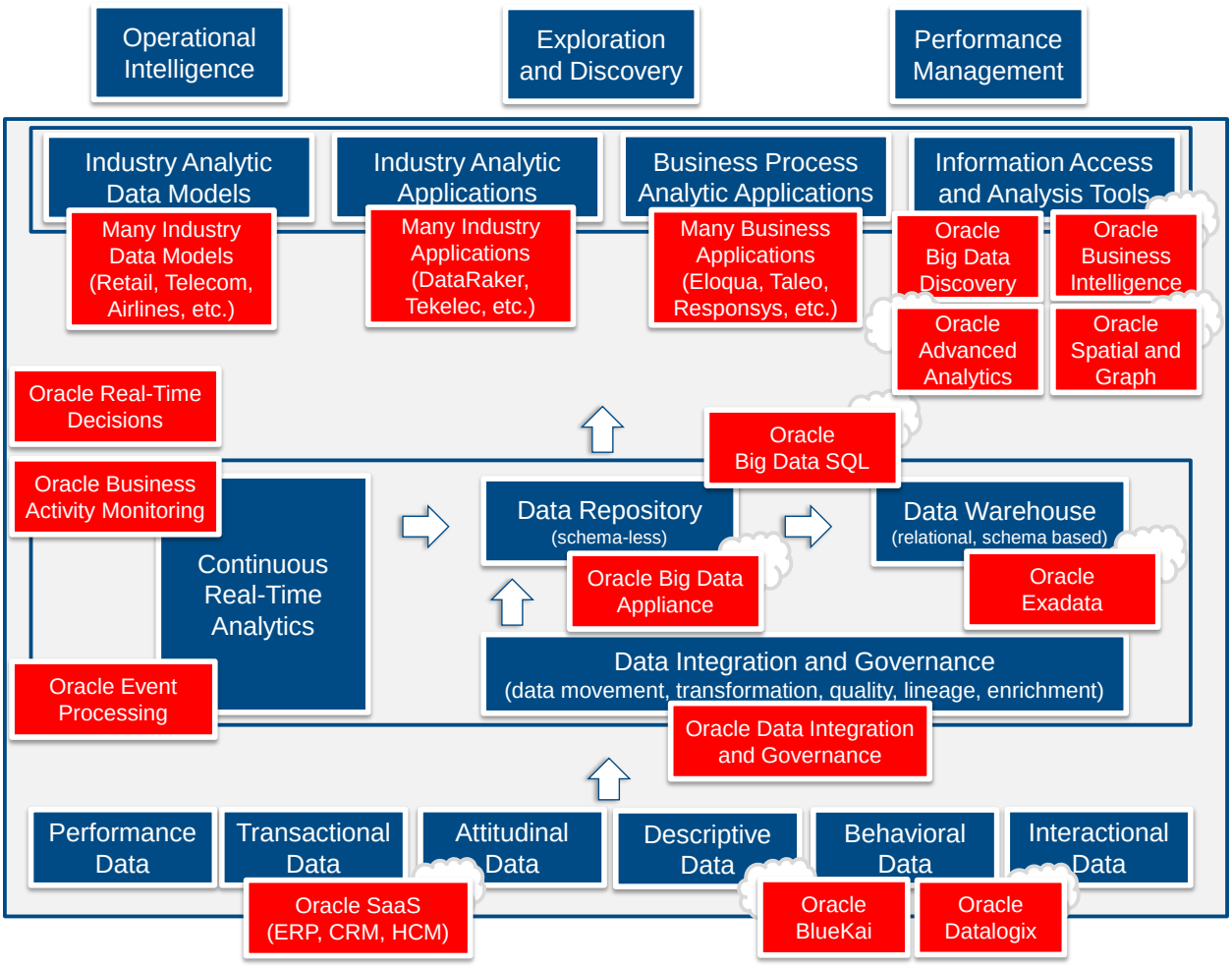
ensure business value at each step. When organizations engage in a process rather than just purchase a collection of products for a specific project, they have a more business-driven and often longer-term view. The architecture matters because it translates the vision and business goals into incremental projects that characterize an agile approach to development.

Oracle in the Big Data Market

Oracle offers a range of Big Data technology components and solutions that its customers are using to address their Big Data needs. In addition, the company offers Big Data architecture design and other professional services that can assist organizations on their path to addressing evolving Big Data needs. Figure 3 shows a sample of Oracle's products mapped to the conceptual architecture in Figure 2.

FIGURE 3

Oracle Big Data and Analytics Software — A Sample of Products



Note: The cloud images identify solutions that are available in the cloud as well as on-premises.

Source: IDC, 2015

The following list includes some of the specific Oracle products in categories representing the three primary Big Data use cases. Note that neither Figure 3 nor the following bullet points provide an exhaustive view of all related Oracle products or services. In addition, the functionality of some of the products is applicable in more than one use case category.

- **Operational intelligence.** To support operational intelligence, Oracle offers products such as Oracle Event Processing, Oracle Real-Time Decisions (a decision management platform that supports next-best action capabilities), and Oracle Business Activity Monitoring. In addition, Oracle architects assist customers in evaluating and deploying open source projects such as Apache Kafka and Spark on the Oracle Big Data Appliance.
- **Exploration and discovery.** To support exploration and discovery workloads, Oracle offers products such as the Oracle Big Data Appliance, in which it bundles Cloudera Enterprise, Oracle R Distribution, and the Oracle NoSQL Database. In addition, Oracle GoldenGate for Big Data provides real-time data ingestion, and Oracle Data Integrator for Big Data orchestrates data movement, transformations, native code generation, and Oracle Metadata Management for data lineage and governance into and away from the discovery platform. The Oracle portfolio also includes Oracle Big Data Discovery, which is a purpose-built search, exploration, discovery, and transformation tool operating directly with Hadoop on Oracle BDA for business users. Oracle Big Data SQL provides Oracle SQL access to Oracle BDA and enables runtime joining of data between Oracle BDA and Oracle Exadata. Oracle Big Data Connectors allow high-speed data movement across both platforms.
- **Performance management.** Oracle offers a range of query, reporting, multidimensional analysis, enterprise search, and financial and strategy management analytic applications. These software tools and applications include specific products such as Oracle Business Intelligence Enterprise Edition, Oracle Endeca, Oracle Essbase, and Oracle Financial Performance Management. In the data management layer, Oracle offers Oracle Exadata, which either incorporates or has options for Oracle Database, in-memory processing, Oracle Advanced Analytics, Oracle Spatial and Graph, Oracle Data Quality, and Oracle Data Redaction, as well as prebuilt data models for specific industries.

Oracle Big Data Architecture Services

In addition to software and hardware components and engineered systems that combine and optimize both software and hardware, Oracle provides comprehensive Big Data architecture development services based on the Oracle Architecture Development Process. This practice, made up of enterprise architects who use workshops and assessment methodologies, guides organizations through the six stages of OADP, including business context evaluation, architecture vision, current state assessment, future state definition, road map development, and change management. Some of the outputs of OADP are conceptual, logical, and physical views of the information architecture and road map to guide the organization for the foreseeable future.

In each of the six cases described in this white paper, customers utilized Oracle professional services either to support architectural development using their OADP approach or to collaborate more closely by using Oracle's services in an architectural board capacity for a longer-term view of the architecture.

Challenges and Opportunities for Oracle

Oracle is not alone in having a service for helping clients and prospects define their Big Data architecture or in providing a range of technology and services offerings to execute Big Data projects. In this competitive and fast-moving market, Oracle must continue to build on the many innovations and key acquisitions that have expanded its portfolio of Big Data solutions. Oracle also needs to assess the pace of cloud adoption for these solutions and ensure that it continues to bring to market in-demand Big Data cloud services.

At the same time, Oracle's position in the market offers the company the opportunity to guide existing and new clients toward a new, modern Big Data architecture that takes a compressive view of all data, multiple information access and analysis use cases, and needs of different user personas. Many of the nonrelational information management and analysis technologies are only now maturing to the point where the large majority of organizations across industries are deploying them in production. Much of the opportunity in the market still remains untapped, and Oracle is in a strong position to address the evolving needs of end users globally.

LESSONS LEARNED

Managers in many companies believe they have what it takes to harness the power of Big Data for improving data-driven decision making, yet they may be missing the competency to address the range of technology, staffing, process, and data requirements involved. With the opportunity to unlock the value of Big Data to accelerate innovation, drive optimization, and improve compliance comes the need to demonstrate value, navigate expanding technology alternatives, recreate business processes, and ensure the availability of appropriately skilled staff.

Henry David Thoreau said, "If you have built castles in the air, your work need not be lost; that's where they should be. Now put the foundations under them." The information foundation and architecture on which it is based is a key building block of these capabilities. In conducting IDC's research through interviews and surveys with customers highlighted in this white paper and others, we have found the following best practices related to the information architecture for successful Big Data initiatives:

- Secure executive sponsorship that emphasizes the strategic importance of the information architecture and ensure that the information architecture is driven by business goals.
- Develop the information architecture in the context of the business architecture, application architecture, and technology architecture — they are all related.
- Create an architecture board with representation from the IT, analytics, and business groups, with authority to govern and monitor progress and to participate in change management efforts.
- Design a logical architecture distinct from the physical architecture to protect the organization from frequent changes in many of the emerging technologies. This enables the organization to maintain a stable logical architecture in the face of a changing physical architecture.
- Consider the range of big use cases and end-user requirements of Big Data. Big Data is not only about exploration of large volumes of log data by data scientists.
- Recognize that the adoption of a comprehensive Big Data solution or any of its components can happen according to a range of different patterns. Each organization should choose the pattern that is appropriate, given existing technology, processes, and human resources.

- Transform the information architecture into incremental, actionable projects rather than trying a "big bang" technology deployment approach for Big Data.
- Look for external help, when needed, for the development of the information architecture. Often an external party can help facilitate internal collaboration between different internal groups as well as assess and recommend new technologies.
- Even at the early stages of a project when evaluating technologies, always consider the full range of functional and nonfunctional requirements that will most likely be required in any eventual deployment. Bolting them on later will drive costs and delays and may require a technology reevaluation. This is yet another reason why an architecture-led approach is important.

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

Global Headquarters

5 Speen Street
Framingham, MA 01701
USA
508.872.8200
Twitter: @IDC
idc-insights-community.com
www.idc.com