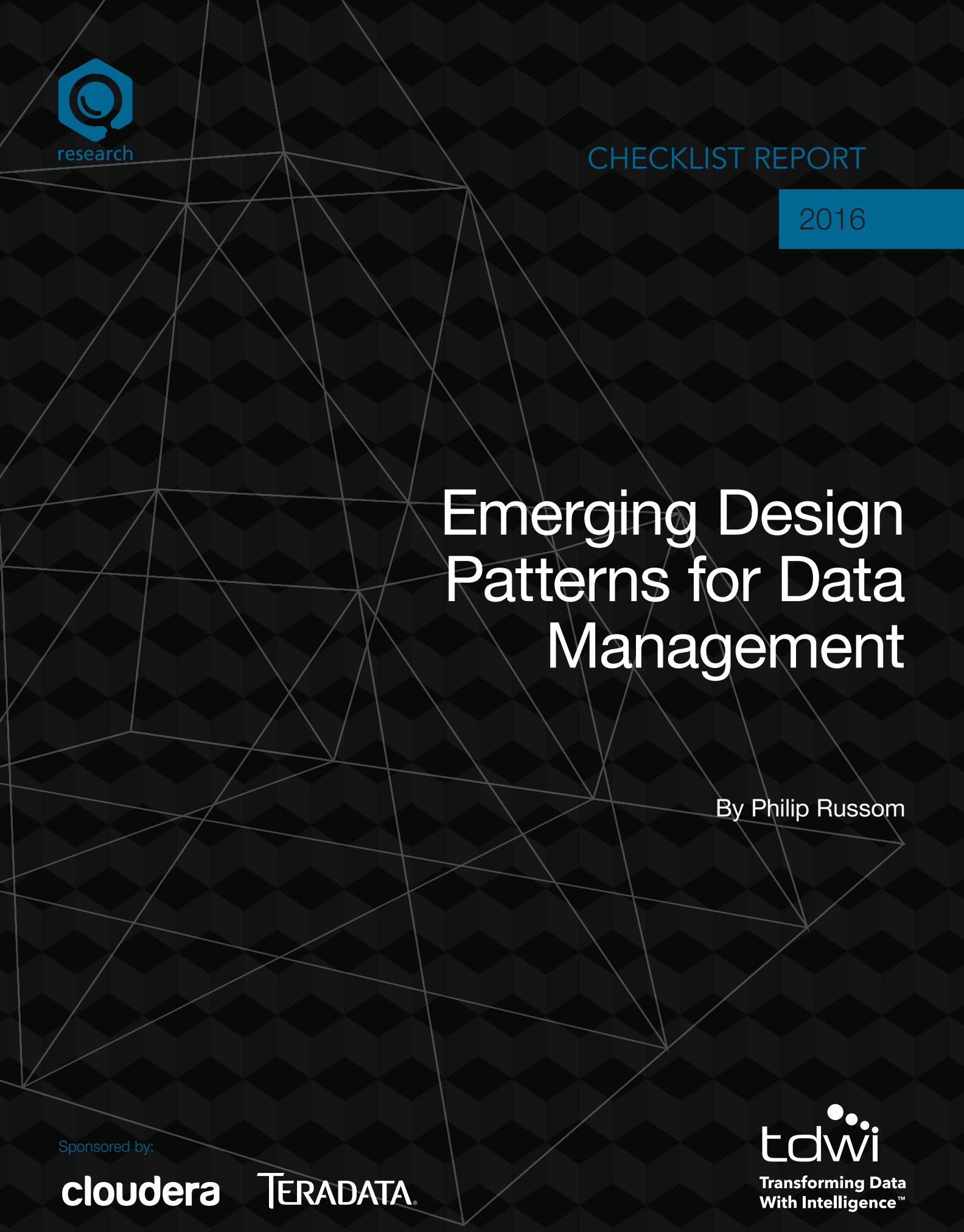

research

CHECKLIST REPORT

2016

Emerging Design Patterns for Data Management

By Philip Russom

Sponsored by:

cloudera TERADATA

tdwi

Transforming Data With Intelligence™

NOVEMBER 2016

TDWI CHECKLIST REPORT

Emerging Design Patterns for Data Management

By Philip Russom

Transforming Data With Intelligence™

555 S. Renton Village Place, Ste. 700
Renton, WA 98057-3295

T 425.277.9126
F 425.687.2842
E info@tdwi.org

tdwi.org

TABLE OF CONTENTS

FOREWORD

Organizations that seek to be data driven are experiencing considerable change of late:

Data itself is evolving. Traditional enterprise data is being joined by new categories of big data, including data from sensors, hand-held devices, machinery, Web applications, and social media. These bring with them new structures, interfaces, and latencies.

Data management is evolving to address the new data. Users are turning to new design patterns on both old and new platforms to accommodate the capture, storage, processing, analytics, and delivery of big data and other new data assets.

Businesses are evolving to take advantage of the new data assets. New data often means new analytical insights and new ways to innovate operations, product design, and marketing. Hence, savvy business managers are eager to accumulate and explore new data in search of opportunities for customer base development, operational excellence, and competitive advantage.

These changes sound like problems, but they are actually opportunities for organizations that can embrace big data, implement new design patterns and platforms for new data, scale to greater volumes and processing loads, and react accordingly via analytics. Here are some representative use cases, each showing how new technologies and new data-driven design patterns can enable new business practices:

Discovery of new insights and opportunities. Because big data usually comes from new sources, TDWI often refers to it as *new data* or *new big data*. The great promise and relevance of new big data is that it can be leveraged to develop new insights, which in turn can help organizations adapt to change in evolving business environments.

Competing on analytics. New design patterns and platforms integrate a broad range of data sources to create unique views into your customer base and marketplace.

Multichannel marketing. A mix of old and new data from websites, call center applications, smartphone apps, and social media can reveal how your customers behave in diverse situations, thereby enabling modern multichannel marketing.

Operational excellence. In an increasingly data-driven business world, big data takes operational analytics and a 360-degree view of customers to a new level.

Analytics with all the data. Given the right design patterns and data platforms, new big data can provide larger and broader data samples, thereby expanding existing analytics for risk, fraud, customer base segmentation, and the complete view of the customer.

Real-time operations. The transportation and logistics industries have transformed in recent years by leveraging sensor data for greater understanding of time-sensitive entities in motion across geographic space; this in turn provides unprecedented customer service, operational efficiency, and competitive advantage. There are even TV commercials about how innovative firms now capture and operationalize real-time data to reduce from weeks to hours the approval of insurance policies and residential mortgages.

Decision-making value from unstructured text. Sentiment analysis (based on human language and other unstructured text) has become almost commonplace as a new insight into customer bases and marketplaces. It is implemented by deploying new design patterns on new data platforms that are conducive to unstructured and multistructured data.

Clearly, there are benefits to be seized, but a number of things have to be in place before the benefits can be realized. Because the most compelling benefits involve managing and analyzing new and big data, most user organizations need to evolve their design patterns and their portfolios of data management platforms, then implement multiple forms of analytics as determined by the priorities of the business.

This Checklist Report will drill into some of the emerging design patterns and platforms for data that modern data-driven organizations are embracing, with a focus on data warehouses and Hadoop, plus a few other data ecosystems (e.g., marketing, supply chain, and Web). The goal of the report is to accelerate users' understanding of new design patterns and data platforms so they can choose and use the ones that best support the new data-driven goals of their organizations.

NUMBER ONE

USE BOTH TRADITIONAL AND EMERGING DATA-DRIVEN DESIGN PATTERNS

A *design pattern* is a rule of thumb—a generalized, repeatable approach to a commonly occurring information technology situation that developers must flesh out (based on current requirements) to create a finished design. In this report, design patterns tend to be localized constructs, such as data schema, models, tables, and record structures. These differ from data architectures, which tend to be large-scale combinations of multiple design patterns and other components, as seen in data warehouse architectures, multichannel marketing, and some applications clusters. Furthermore, the concept of the design pattern originated in applications development, whereas this report is focused on *data-driven design patterns*, especially new and emerging ones used for advanced analytics, exploration and discovery, big data, and multistructured data.

In the world of data management, data developers and modelers have long used data-driven design patterns. For example, data warehouse architectures typically include data marts, time series, dimensional models, and other design patterns. As another example, modern multichannel marketing can employ design patterns for customer profiles and segments, customer masters, and the complete view of the customer. Various data-driven design patterns found across the enterprise are for transactions, bills of material, supply chain documents exchanged with partners, and other operational data.

Note that design patterns are quite diverse in how they capture, organize, and manage data. The data itself is also diverse in terms of structure, container, latency, and volume. As a data management strategy that accommodates diversity, many organizations deploy multiple data platform types so that users can store, manage, and process data on just the right platform for a given design pattern and its use cases. Amid this diversification of data platforms, relational platforms are still relevant and vital, and they are increasingly complemented by Hadoop.[1]

These examples prove that traditional design patterns are all around us, representing common business entities, events, and processes. Yet new design patterns are emerging as big data arrives and new applications come online.

For example, data management professionals continue to use older design patterns but also introduce new ones, such as the logical data warehouse, enterprise data hub, and data lake. In addition, they are evolving older patterns to address new requirements, as when data landing is redesigned to embrace early ingestion or when enterprise data is redesigned to accommodate advanced analytics.

Some emerging design patterns are optimized for new data from the Web, machines, and social media. Others leverage new data characteristics, such as geospatial coordinates and real-time events. As we'll see later, these and other design patterns can be fleshed out in many ways (to accommodate new data types and use cases) and deployed on many types of platforms (from relational databases to Hadoop), as well as in the cloud, on premises, and on hybrid combinations of all the previous.

[1] For a discussion of multiplatform data ecosystems, see *TDWI Best Practices Report: Evolving Data Warehouse Architectures* (2014), online at www.tdwi.org/bpreports.

NUMBER TWO

EMBRACE THE DESIGN PATTERN TREND TOWARD VERY LARGE DATA REPOSITORIES

One of the strongest trends in data management today is toward design patterns that involve very large data repositories, which are used to capture and integrate big data, other new data, and traditional data. In these cases, a large repository stores data containers and other information elements with fidelity to their original schema, content, and condition so that the raw details of the source are retained for a wide range of use cases in data exploration, data integration, and advanced analytics. Given the wide range of data that technical users are capturing, this type of data repository regularly amounts to tens or hundreds of terabytes of data—and sometimes petabytes.

There are a number of terms in the data management community for the current generation of very large data repositories and similar approaches, including *data lake*, *data vault*, and *enterprise data hub*. Furthermore, other data-driven design patterns have traditionally handled large collections of detailed source data, including data warehouses and the many variations of the operational data store (ODS).

Though one raw-data repository can do many things, some users deploy several so each can be optimized for specific data types, workloads, latencies, and so on. Very large data repositories may also be deployed per business unit, geography, platform type, or application technology stack. Finally, note that a very large data repository can be deployed atop various data platforms, including MPP relational DBMSs; however, the large configuration of a DBMS required to make a large raw-data repository functional would be extremely expensive and would limit the diversity of data assumed of a very large raw-data repository. Thus the trend is toward Hadoop as the preferred platform for such repositories.

Many user organizations are already employing very large data repositories for real-world use cases,[2] including:

- **Advanced analytics applications** that are enabled by mining, clustering, statistics, and graph techniques often rely on the rich details of raw source data instead of the cleansed, aggregated, and remodeled data typically found in a data warehouse.
- **Data exploration and discovery** are in vogue with both technical and business users because they must study new big data to understand its technical condition and its potential business value, and most users want to explore in a self-service manner. Incorporating diverse data into a large repository empowers broad exploration as well as rich analytical correlations across data of differing sources and vintages.
- **Extending and modernizing a data ecosystem** can be accomplished by integrating a large repository into its data architecture. TDWI has seen numerous user organizations breathe new life into their relational data warehouse by integrating it with Hadoop.
- **Multistructured data** has confounded many data management teams. The confusion is now ending because Hadoop has arrived as a platform conducive to capturing and processing a very wide range of data types, formats, and structures with diverse latency and ingestion requirements.
- **Data hubs** are one way to organize a very large data repository, and many users are going in this direction. A successful data hub can bring together all these use cases on a single platform so that data can be shared across workloads and teams—from processing to analytics to operationalizing results. A data hub includes the security and governance needed to handle sensitive data, and it makes finding, querying, and extracting data easier and better governed. Yet the hub still respects the repository's mandate of maintaining source fidelity. Accordingly, the primary role of a data hub is to be a manageable and governable distribution point for sharing data among many users, organizations, and applications, which greatly enhances the business value of the very large data repository.

[2] Some of these use cases and others can be enriched with self-service functionality to make them practical for nontechnical and mildly technical users, as explained in "Number Five" of this report.

NUMBER THREE

KNOW HADOOP'S ROLES WITH EMERGING DESIGN PATTERNS

A recent TDWI survey indicates that the number of deployed Hadoop clusters is up 60% over two years.[3] Users are aggressively adopting Hadoop and its ecosystem of data management and analytics components, whether open source, vendor supplied, or user built. These are being applied to a variety of use cases in data warehousing, analytics, and data management.

Although data-driven design patterns and technology platforms are mutually exclusive, Hadoop is a natural fit for big data, new data types and sources, and very large data repositories. This is why most very large raw-data repositories are built atop Hadoop, from which they get several desirable capabilities:

- Hadoop handles the massive volumes of data we assume of big data and enterprise data, and it does so with linear scalability.
- In general, a Hadoop cluster costs less than an equivalent configuration of an enterprise database management system (DBMS), as measured by hardware and software acquisition costs.
- Hadoop can handle any data structure you can put in a file, and many types of new data arrive in files, as in Web server logs and JSON.
- High availability is built into every multinode Hadoop cluster.
- Hadoop supports the SQL functionality and third-party analytics tools that businesses are already accustomed to using for more structured access (e.g., Hive for batch processing and Impala for SQL-based analytics). Although a large repository focuses on raw data, users inevitably need to structure some of that data for better business value, specific analytics applications, data standardization, and evolving requirements.
- Hadoop improves regularly as seen in the recent advances made in vendor distributions and in the broader ecosystem of tools that support Hadoop for metadata management, security, SQL support, and real-time data.

A single Hadoop cluster can support multiple use cases:

Enterprise data hub (EDH). In a recent TDWI survey, among all the practices users could implement atop Hadoop, the one with the greatest anticipated growth over the next three years is the EDH.[4] As Hadoop users mature in their use of large repositories and other data-driven design patterns in a Hadoop environment, they need the governance, orchestration, and light structuring that a data hub can provide if they are to bring multiple use cases together in a multitenant environment.

Data warehouse augmentation. The primary mandate of a data warehouse (DW) is to provision relational and dimensional data for standard reports, dashboards, performance management, and OLAP. These and related purposes are mission-critical for enterprises, and so relational DWs are more relevant than ever. However, when integrated with a relational DW, Hadoop can augment, extend, and modernize the warehouse architecture.

In turn, the presence of Hadoop enables a DW team to also manage and leverage nonrelational data, thereby enabling many new analytics applications and business use cases. Note that, in this hybrid scenario, relational DBMSs and Hadoop complement each other, and one does not replace the other. This complementary alliance is why emerging DW architectures and data-driven design patterns increasingly incorporate both Hadoop and relational platforms.

Data landing and data integration. Hadoop provides much-needed scalability and extra storage capacity for raw data (both old and new) entering the data warehouses and other data ecosystems. In a related area, Hadoop can offload and optimize ETL/ELT and other data integration functions, all within the same system where further analysis can take place. Hadoop users now "push down" certain types of processing for transformation, aggregation, summation, calculation, and so on, similar to how they push relational processing down into the relational warehouse.

Note that ELT processing varies tremendously; in general, relational forms are best done with a relational target, whereas algorithmic forms usually work well on Hadoop. As in other scenarios, Hadoop and relational platforms have differing strengths and therefore complement each other.

Live, online data archive. Most of the raw source data (that might be archived) is in straightforward schema that Hadoop can manage and process easily, though at a much lower price point than a relational platform or dedicated archiving platform. Unlike traditional offline archives, a Hadoop-based archive is online and ready for real-time search, query, and look-ups, which transforms archiving from a cost center to a business opportunity.[5]

Integrated data for exploration. Given the massive volumes and diverse structures that Hadoop can manage, a large repository atop it provides ample, detailed data for broad data exploration and the discovery of new business opportunities, typically in a self-service manner.

(Continues)

[3] See the discussion around Figure 4 in *TDWI Best Practices Report: Hadoop for the Enterprise* (2015), online at www.tdwi.org/bpreports.

[4] See the discussion around Figure 17 in *TDWI Best Practices Report: Hadoop for the Enterprise* (2015), online at www.tdwi.org/bpreports.

[5] For more information about modern archiving, see *TDWI Checklist Report: Active Data Archiving* (2014), online at www.tdwi.org/checklists.

(Continued)

Advanced analytics. This is what the business wants and needs from raw-data repositories and Hadoop, as today's conventional wisdom says that the primary path to business value from big data and other new data sources is through analytics. Hadoop delivers in spades because it supports both set-based analytics (SQL and some other relational functions) and algorithmic analytics (for mining, machine learning, graph, and statistics).

NUMBER FOUR

ALSO CONSIDER VIRTUAL AND FEDERATED DESIGN PATTERNS

By definition, a data-driven design pattern is first and foremost a *logical* construct that describes the general shape and content of a data schema, model, data set, or result set. The data that populates the design pattern may be pre-collected in one physical location or distributed across several. The latter scenario is often called a *virtual* data set, where data that is physically distributed across platforms is integrated on the fly to instantiate a design pattern's intended result set.

The logical data warehouse relies heavily on logical design patterns expressed via technology for data virtualization, federation, views, and indexing. These virtual techniques are also good for imposing just enough structure on subsets within a large raw-data repository (or across multiple platforms), without altering the structure and condition of the original raw data.

For example, the latest generation of data visualization tools (when pointed at raw-data repositories and other data sources) regularly execute federated queries and manage the result sets as materialized views. Visualization tools aside, virtual functionality that is appropriate to many design patterns (regardless of the physical data platforms involved) is also supported in tools for data integration, reporting, analytics, and database management.

Distributed queries (sometimes called federated queries) are naturally assumed in today's multiplatform data warehouse architectures, which regularly mix Hadoop and relational platforms. A large-scale, hybrid data architecture of this sort will manage data on platforms that are best suited to the storage and processing requirements of specific data sets. When data analysts and similar users explore distributed data, they depend heavily on tools that support distributed queries and data federation. Otherwise the scope of discovery is severely limited, which means that business value from the technology is also limited.

Hence, if you want data exploration to encompass all data (thereby avoiding samples), reaching multiple data platforms via a single distributed query becomes mandatory. An added benefit of distributed queries is that they allow the user to leave data in situ—that is, on the platform where it's managed. The query pulls back a small subset of the remote data without needing to copy very large data sets from one platform to another.

NUMBER FIVE

ADOPT NEW PRACTICES FOR BUSINESS ANALYTICS, AS ENABLED BY NEW DESIGN PATTERNS AND THEIR TOOLS

Making a business case for emerging data-driven design patterns, large data repositories, and Hadoop is relatively straightforward. These and associated tools enable new practices that support the goals of the business or lead to organizational advantages.

Self-service data access. A growing number of users want to work fast, at the speed of thought. They want to work autonomously without waiting for IT or a data management team to create unique data sets for them. To make this new practice happen, users need tools and platforms that provide *self-service data access*, especially for large raw-data repositories. While accessing data in a self-service fashion, a user (whether business, technical, or hybrid) tends to build a data set that includes personally discovered information in a model that depicts the problem or opportunity he or she is studying (such as a new form of customer churn or a profile of customers prone to buy certain products).

Self-service data access is a foundation for other self-service tasks, such as data exploration, data prep, visualization, and some kinds of analytics. All these require substantial business metadata (not just technical metadata) if business users are to make sense of emerging design patterns for data.

Data exploration and discovery. Self-service empowers business users to autonomously explore data and discover new facts and relationships that can propel the business. Further, it enables both business and technical users to study and on-board new data as it arises. Therefore, when you design a large raw-data repository, involve business people in the requirements-definition phase to ensure that the data sets they need for self-service data exploration are in the repository.

Data prep. Instead of attempting to use a dauntingly feature-rich data integration platform, some users just need *data prep*, which is a subset of that functionality, presented via an easy-to-use user interface. Data exploration, prep, and analysis often go together in a multistep self-service process. As users access and explore data, they want to prepare a data set quickly and easily based on what they discover, then share the prepped data set with colleagues or seamlessly take it into analytics.

Note that the new practice of data prep does not replace traditional data management best practices that include substantial data integration, quality, complex aggregation, and advanced data structures (as required by most standard reports and data warehouses). The old and new practices complement each other, and so both are required for comprehensive data management that meets the needs of today's diverse user constituencies.

Multigenre analytics. As user organizations dive deeper into analytics, they soon learn that many business problems are best solved with multiple forms of analytics. After all, they learn something unique from each analytical approach; combining approaches expands their insights and opportunities. A Hadoop-based repository allows new big data to be repurposed in many different ways as needed by analytics for mining, graph, machine learning, and statistics. Note that data visualization is also a form of analytics, and it is one of the most appealing for business users at the moment.

Business monitoring. Some design patterns assume the ingestion of streaming data, followed immediately by real-time access. In sophisticated implementations, the Hadoop-based raw-data repository can be a hub for real-time data in support of fast-paced business operations. In these cases, users need to monitor, configure, and access streams of data at close to real time in a self-service fashion. For example, a merchandizing manager might monitor shopping carts in e-commerce and launch micro sales and temporary discounts accordingly. An engineer might monitor a utility grid and reroute service. As a value add, these tasks may be performed via manual self-service tool functions, via tool automation, or both.

NUMBER SIX

ADAPT YOUR DATA MANAGEMENT BEST PRACTICES TO NEW DESIGN PATTERNS

Embrace early ingestion and late processing. As more users want to conduct analytics with relatively fresh data, it's important to have a data-driven design pattern optimized for fast ingestion. That way data is available ASAP for analytics, reporting, exploration, and business monitoring. To speed ingestion, more and more data is being ingested in its original raw state or with minimal improvement up-front; instead, data is processed and improved later by repurposing the captured raw data. Users don't waste as much time as in the past improving data on the off chance it will be used. For new best practices in early ingestion and late processing to work properly, emerging design patterns must handle a wide range of latencies for data ingestion, from traditional overnight batch to true real time.

Early ingestion has its benefits. However, you should not give in to the temptation to "dump" large amounts of arbitrary data into a raw-data repository without proper data governance and metadata management. Instead, you should select data carefully, then document and improve it as it comes in for data lineage and audit, metadata development, volume or node assignment, and compliance. Even within these strictures, a large raw-data repository can still be true to its primary mission, which is to provide detailed source material for exploration, analytical correlation, and future repurposing of data.

Perform more ad hoc data management on the fly, instead of a priori. Instead of taking a week to design a data model, many users now create the model and populate it with data while they explore data in a very large data repository or other new design pattern. This is possible because today's advanced hardware and software have the speed and scale required to process and repurpose data on the fly, at runtime.

Develop metadata as you explore and use the data of a design pattern. Many big data and other new data sources have little or no metadata available. Sometimes the sources are outside your reach, so you cannot extract metadata from them (as with remote sensors). At other times, data arrives in files that may—or may not—have a header that's equivalent to metadata (as with XML, JSON, and most logs). In these cases, you need to develop and manage metadata yourself.

For Hadoop-based data repositories, there are tools available today with which you can deduce new metadata and further develop preexisting metadata on the fly during exploration or at runtime. Note that technical metadata is good, but you'll also need business metadata for the growing number of users who need self-service data access, data exploration, data prep, analytics, and visualization. Finally, be sure that metadata developed for new data-driven design patterns is captured in a repository for future use and sharing, as well as for business consistency and the integration of multiple data platforms.

Govern new big data as you would any data asset. As new data-driven design patterns come online and as new big data arrives, data governance bodies should be involved so that new data and platforms are controlled and standardized. Likewise, as users are trained on the new data and platforms, the training should make them aware of relevant governance policies. Otherwise, new practices with emerging design patterns (such as self-service exploration and analytics) run the risk of violating governance polices for compliance, security, and privacy. In addition, some new data (even when managed in raw form) should comply with enterprise standards for data quality and modeling.

Although data governance is mostly about people and process, modern metadata management can make new data design patterns (whether on Hadoop or relational platforms) more governable (and discoverable for self-service use) via software automation for data lineage and data cataloging. Furthermore, operational metadata can enable the tracking and governance of data accesses for security purposes.

Balance capacity and cost with emerging design patterns. Big data just keeps getting bigger, especially as you widen the range of new sources and store even more raw data. Furthermore, many design patterns incorporate both new data and traditional enterprise data to enable rich analytical correlations across many sources and vintages. Therefore you must select your data platforms and tools carefully so you can scale easily as data volumes and processing loads increase. Yet scaling to greater capacity requirements should be done in a cost-effective manner. A leading reason why Hadoop has come on strong in data management is its ability to scale linearly but with minimal cost. Similarly, many users are turning to cloud-based data platforms (whether relational or Hadoop) because clouds offer automatic elasticity at a reasonable cost.

NUMBER SEVEN

MODERNIZE YOUR DATA MANAGEMENT ARCHITECTURE BY INCLUDING AND INTEGRATING MULTIPLE DATA-DRIVEN DESIGN PATTERNS

Traditional data warehouse (on a relational platform) and a big data design pattern (usually on Hadoop). Seventeen percent of data warehouse professionals surveyed recently by TDWI have Hadoop in production and integrated with their data warehouse. Approximately 36% say they will have Hadoop integrated with the warehouse within three years. The users surveyed also indicated that Hadoop complements their traditional warehouse platform without replacing it.[6]

Hadoop adoption is increasing in data warehousing because some (but not all) design patterns associated with the relational warehouse are a good fit for Hadoop. For example, Hadoop's linear scalability and support for multiple ingestion methods makes it a bigger and better data landing and staging platform. The large amounts of detailed source data that many data warehouse teams retain for advanced analytics are easily stored and processed on Hadoop. Furthermore, Hadoop can modernize a data warehouse by enabling it to capture and process a wide range of data structures from new sources, such as sensors, social media, the Web, and hand-held devices.

Even so, the relational warehouse is still the preferred platform for the design patterns that feed data into standard reports, dashboards, performance management, dimensional data, OLAP, and strategic decision making. Thus the strengths of the relational warehouse and Hadoop are complementary, making the two a perfect match, especially for organizations with stringent requirements for traditional warehousing, as well as for big data, advanced analytics, and multistructured data.

Data-driven design patterns for data integration. Complex flows for data integration (DI) often need a data platform where data is landed as it is ingested into an environment. DI processes may then make passes over the landed data to process and stage it for eventual loading and synchronizing with target systems. The landed and staged data needs design patterns optimized for the requirements of today's fast ingestion, real-time streams, multistructured data, and other mature requirements for staging data. This is true of data warehousing, but CRM, Web, supply chain, and other functional areas within a business have similar data landing/staging technical functions that need to be organized by emerging design patterns.

Some design patterns need to migrate within the larger architecture. For example, the average data warehouse architecture includes several operational data stores and data marts. These may be physically located within the primary instance of the warehouse or on standalone DBMS instances elsewhere. Users with experience with Hadoop in data warehousing tend to migrate these design patterns to Hadoop to take advantage of its low cost, linear scalability, and ability to support a wide range of processing, analytics, and operational use cases via a single cluster. Outside warehousing, organizations practicing multichannel marketing tend to have design patterns per customer channel or CRM/SFA application and customer view; these are inherently redundant and could benefit from migration and consolidation into a large repository on Hadoop.

Designs and platforms for real-time functions are needed, too. With more and more new data coming in streams (from sensors, Web apps, and smartphone apps), users are reacting by optimizing large repositories for early ingestion or real-time interfaces. Such optimization via emerging design patterns is worthwhile when technical users are under pressure to provide fresh data for time-sensitive business processes, such as performance management, e-commerce, facility surveillance, business monitoring, and real-time analytics.

Streams and other new real-time data sources tend to arise suddenly as new applications, partners, and customers come online. To address today's design patterns—and to future-proof against unforeseen ones—it helps to have tools and platforms that support machine learning, metadata deduction (schema-on-read), on-the-fly modeling, and other automation for data developers.

Creative uses. Although one very large data repository can serve multiple purposes for multiple tenants, TDWI is starting to see multiple repositories per department or business function. For example, the marketing repository is especially prominent because the 360-view of customers and analytics for customer base segmentation benefits from highly detailed data.

Furthermore, modern marketers are eager to explore customer data in search of new segments, churn, products of affinity, and opportunities for growing accounts. In these cases, multiple design patterns come together to support differing views of customers. TDWI has found similar emerging design patterns for other departments, such as procurement departments, sales pipeline functions, and security analytics.

[6] See the discussion around Figure 16 in *TDWI Best Practices Report: Data Warehouse Modernization* (2016), online at www.tdwi.org/bpreports.

ABOUT OUR SPONSORS

cloudera

www.cloudera.com

Cloudera delivers the modern data management and analytics platform built on Apache Hadoop and the latest open source technologies. The world's leading organizations trust Cloudera to help solve their most challenging business problems with Cloudera Enterprise, the fastest, easiest, and most secure data platform available for the modern world. Our customers efficiently capture, store, process, and analyze vast amounts of data, empowering them to use advanced analytics to drive business decisions quickly, flexibly, and at lower cost than has been possible before. To ensure our customers are successful, we offer comprehensive support, training, and professional services. Learn more at www.cloudera.com.

TERADATA.

teradata.com

Teradata empowers companies to achieve high-impact business outcomes. Our focus on business solutions for analytics, coupled with our industry leading technology and architecture expertise, can unleash the potential of great companies. Visit teradata.com.

ABOUT THE AUTHOR

Philip Russom, Ph.D., is senior director of TDWI Research for data management and is a well-known figure in data warehousing, integration, and quality, having published over 550 research reports, magazine articles, opinion columns, and speeches over a 20-year period. Before joining TDWI in 2005, Russom was an industry analyst covering data management at Forrester Research and Giga Information Group. He also ran his own business as an independent industry analyst and consultant, was a contributing editor with leading IT magazines, and worked as a product manager at database vendors. His Ph.D. is from Yale. You can reach him at prussom@tdwi.org, @prussom on Twitter, and on LinkedIn at linkedin.com/in/philiprussom.

ABOUT TDWI RESEARCH

TDWI Research provides research and advice for BI professionals worldwide. TDWI Research focuses exclusively on analytics and data management issues and teams up with industry practitioners to deliver both broad and deep understanding of the business and technical issues surrounding the deployment of business intelligence and data warehousing solutions. TDWI Research offers reports, commentary, and inquiry services via a worldwide membership program and provides custom research, benchmarking, and strategic planning services to user and vendor organizations.

ABOUT TDWI CHECKLIST REPORTS

TDWI Checklist Reports provide an overview of success factors for a specific project in business intelligence, data warehousing, or a related data management discipline. Companies may use this overview to get organized before beginning a project or to identify goals and areas of improvement for current projects.