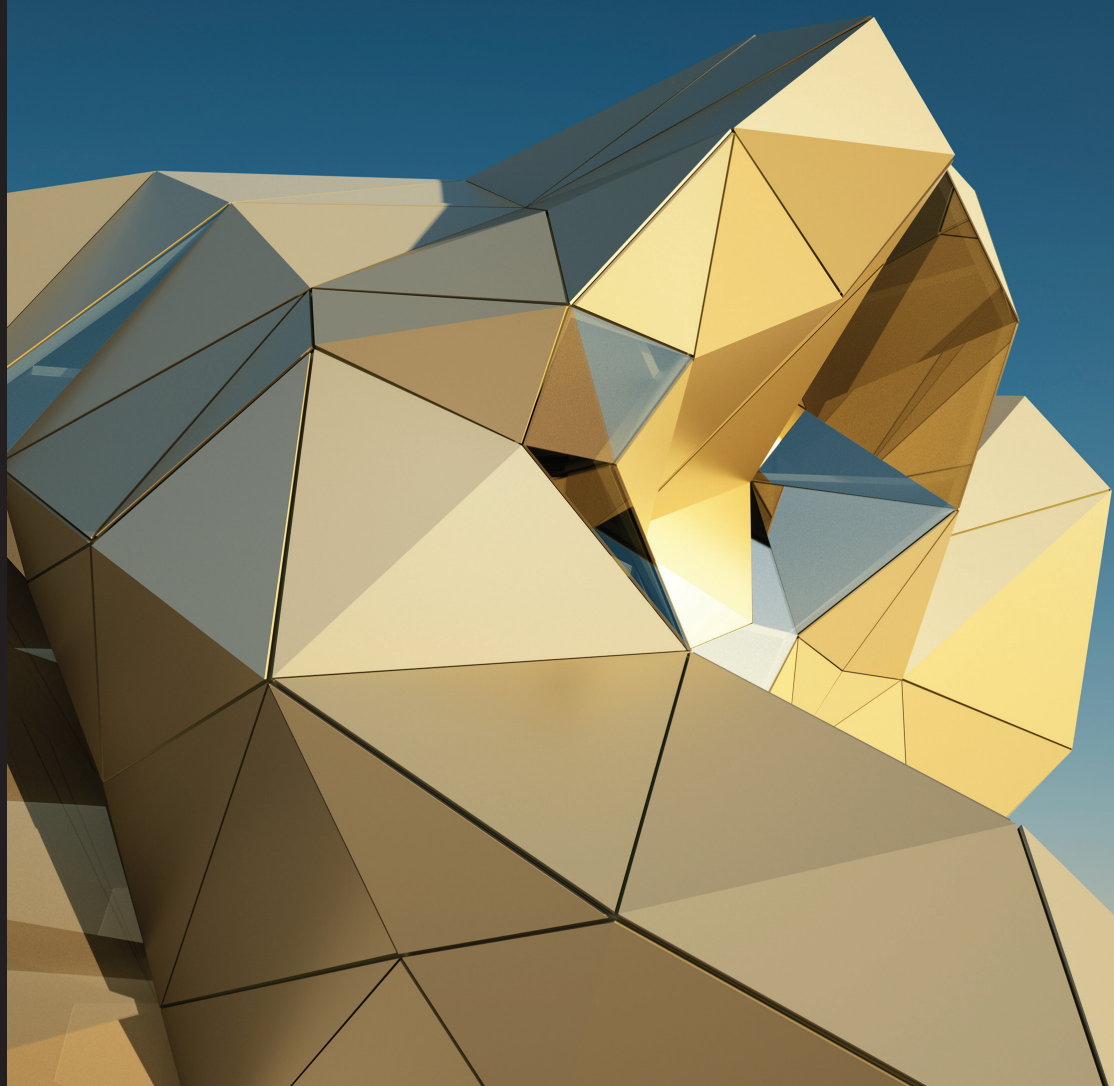
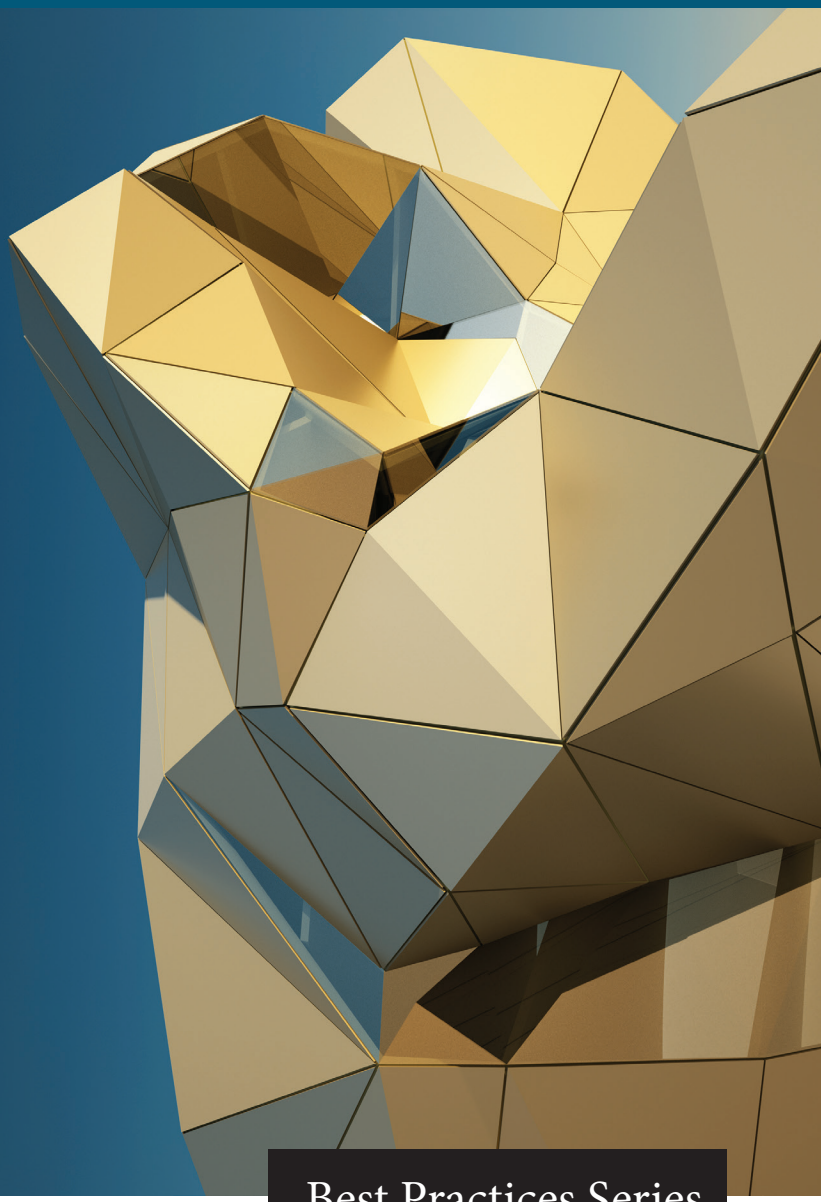
database
TRENDS AND APPLICATIONS

Best Practices Series

# MOVING TO A MODERN DATA ARCHITECTURE





# BUILD TO CHANGE, NOT TO LAST

## Eight Steps to Building a Modern Data Architecture

Best Practices Series

MODERN DATA ARCHITECTURE doesn't just happen by accident, springing up as enterprises progress into new realms of information delivery. Nor is the act of planning modern data architectures a technical exercise, subject to the purchase and installation of the latest and greatest shiny new technologies. Rather, the design and creation of modern data architectures is an uplifting process that brings in the whole enterprise, stimulating new ways of thinking, collaborating, and planning for data and information requirements. It's an opportunity for business decision makers to sit down with IT colleagues and figure out what kind of business they want to be in, what kinds of information they seek to propel that business forward, and what needs to be done to capture and harness that information.

One thing is clear: The old models of data architecture aren't enough for today's data-driven business demands. An architecture designed a decade ago, that rapidly and seamlessly moves data from production systems into data warehouses, for example, may not be capable of meeting the needs of today's real-time, data-driven enterprises.

Architecture is more important than ever because it provides a road map for the enterprise to follow. Without a well-planned, careful, deliberate approach to data architecture, another type of architecture rises to take its place—a "spaghetti architecture" approach that occurs when every business unit or department sets out to buy its own solutions.

Here are the essential components that need to go into building a modern data architecture:

### WORK WITH BUSINESS USERS TO IDENTIFY THE TYPES OF DATA THAT ARE THE MOST VALUABLE

The purpose of good data architecture is to bring together the business and technology sides of enterprises to ensure they are working to a common purpose. To be of value, information needs to have a high business impact. This data may have been within enterprise data environments for some time, but the

means and technologies to surface such data, and draw insights, have been prohibitively expensive. Today's open source and cloud offerings enable enterprises to pull and work with such data in a cost-effective way.

## MAKE DATA GOVERNANCE A FIRST PRIORITY

Working closely with the business side requires guarantees that data not only be of value, but that it is also well-vetted. The process of identifying, ingesting, and building models for data needs to assure quality and relevance for the business. Responsibility for data must be established—whether it's individual data owners, committees, or centers of excellence.

## BUILD SYSTEMS TO CHANGE, NOT TO LAST

A key rule for any data architecture these days it is not wedded in any way to a particular technology or solution. If a new solution comes on the market —the way NoSQL arose a few years back—the architecture should be able to accommodate it. The types of data coming into enterprises can change, as do the tools and platforms that are put into place to handle them. The key is to design a data environment that can accommodate such change.

## DEVELOP A REAL-TIME FOUNDATION

A modern data architecture needs to be built to support the movement and analysis of data to decision makers and at the right time it is needed. Also, it's important to focus on real-time from two perspectives. There is the need to facilitate real-time access to data, which could be historical; and there is the requirement to support data from events as they are occurring. For the first category, existing infrastructure such as data warehouses have a critical role to play. For the second, new approaches such as streaming analytics are critical. Data may be coming from transactional applications, as well as devices and sensors across the Internet of Things and mobile devices. A modern data architecture needs to support data movement at all speed—whether it's sub-second speeds, or with 24-hour latency.

*One thing is clear: The **old models** of data architecture **aren't enough** for today's data-driven demands.*

## BUILD SECURITY INTO THE FOUNDATION

A modern data architecture recognizes that threats are constantly emerging to data security, both externally and internally. These threats are constantly evolving—they may be coming through email one month, and through flash drives the next. Data managers and architects are in the best and most knowledgeable position to understand what is required for data security in today's environments.

## DEVELOP A MASTER DATA MANAGEMENT STRATEGY

With a master data management repository, enterprises have a single "gold copy" that synchronizes data to applications accessing that data. The need for an MDM-based architecture is critical—organizations are consistently going through changes, including growth, realignments, mergers, and acquisitions. Often, enterprises end up with data systems running in parallel, and often, critical records and information may be duplicated and overlap across these silos. MDM also assures that applications and systems across the enterprise have the same view of a customer, versus disparate or conflicting pieces of data.

## POSITION DATA AS A SERVICE

Many enterprises have a range of databases and legacy environments, making it challenging to pull information from various sources. Access is enabled through a virtualized data services layer that standardizes all data sources—regardless of device, applicator, or systems. Data as a service is by definition a form of internal cloud, in that data—along with accompanying data management platforms, tools, and applications—are made available to the enterprise as reusable, standardized services. The potential advantage of data as a service is that processes and assets can be prepackaged based on corporate or compliance standards and made readily available within the enterprise cloud.

## OFFER SELF-SERVICE ENVIRONMENTS

With self-service, business users can configure their own queries and get the information or analyses they want, or conduct their own data discovery, without having to wait for their IT or data management departments to deliver the information. In the process, data application can reach and serve a larger audience than previous generations of more limited data applications. The route to self-service is providing front-end interfaces that are simply laid out and easy to use for business owners. In the process, a logical service layer can be developed that can be re-used across various projects, departments, and business units. IT still has an important role to play in a self-service-enabled architecture—providing for security, monitoring, and data governance. There is a new generation of tools and templates now available from vendors that enable users to explore datasets with highly visual, even 3D views, that can be adjusted, re-adjusted, and manipulated to look for outliers and trends. ■

*—Joe McKendrick*

MarkLogic®

# Powering the Data-Centered Data Center

Modern enterprises face increasing pressure to deliver business value through technological innovations that leverage all available data. Unfortunately, the relational databases that enterprises have relied on for the last 40 years are now too limiting and inflexible. Relational databases drive application-centric architectures that require pre-defined schemas and ETL processes that eventually lead to complex and expensive architectures that are difficult to maintain. Successful organizations are embracing a better approach: *the data-centered data center.*

The data-centered approach focuses on the data, its use, and its governance through its lifecycle. It is architected to allow a single management and governance model, and to bring the applications to the data, rather than copying data to the applications. With the right technology, organizations can build a data-centered data center that minimizes data duplication, provides consistent data governance, enables flexibility in application development, scales up and down to match demand, and manages data securely and cost-effectively.

MarkLogic enables the data-centered data center, having the only NoSQL database that is powerful, agile, and trusted enough to act as a single platform for all of an organization's data. MarkLogic employs a flexible shared-nothing architecture, runs on commodity hardware in any environment, handles both operational and analytical workloads, and secures your data with certified, government-grade security.

## YESTERDAY'S ARCHITECTURE: LOTS OF ETL

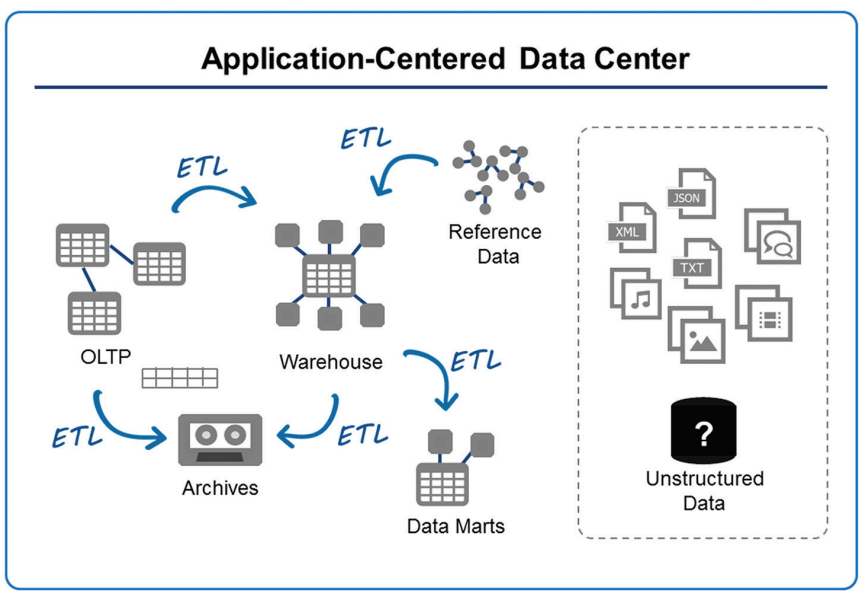
The world used to be simple. Technology limitations meant less data was stored, and there were fewer choices in terms of how to store and manage data. If you needed to build an application, there was a basic recipe that could be followed. You would begin by designing the application, figuring out all the data that would be involved from start to finish. Requirements would be developed up front and you would figure out all the queries you would need to run. Next, you would design the schema and indexes, and then you would build the database. You could then perform the process of extracting, transforming, and loading the data into your new database. After that, you could build the application, write the queries, develop the business logic, and design the user interface.

But, what if the data or requirements changed?

New data or requirements meant beginning the process all over again. It would mean developing a new schema, a new database, and a new application. If there was a required change, resources were consumed very quickly—and each time the process became slower, schemas grew more complex, and ETL more brittle. Perhaps a data warehouse, data mart, or a Master Data Management system would solve the problem—but only briefly. Soon, new requirements to tie data sources together or new data flowing in from outside of the organization would quickly multiply the problem out of proportion.

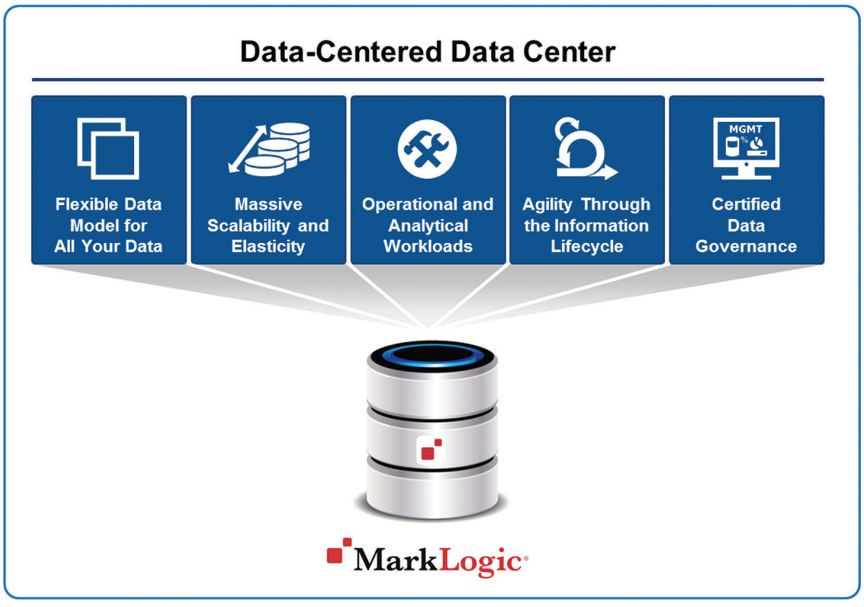
With big data, the problem gets worse. It is often impossible to determine in advance what might be useful to the business. Multiple types of structured and unstructured data are being created and streaming in—from new sources as well—from various new lines of business, mergers and acquisitions, social media, connected devices, and more. Before long, copies upon copies of data are floating around the organization.

What is the security model? How do you manage backup and availability for various systems? How do you apply data



retention and disposal policies? Where is the master? Who is editing the data? Soon, control is lost.

Organizations find themselves in this scenario frequently, and it's easy to understand why. From a business perspective, applications are the visible face of the data and applications appear to hold the value. Application by application, our data centers have become more and more complex and expensive, and the architecture becomes a barrier to success.




## TODAY'S ARCHITECTURE: FLEXIBLE AND ADAPTIVE

Today's modern architecture is refocusing on the most important asset: data. By focusing on the data, you are able to make smarter choices about how to store and manage it. Rather than copying the data all around and maintaining multiple disparate systems, you can manage the data in one place, in its most useful form, and make the application come to the data. After all, the data will surely outlive the application.

To adopt a data-centered data center, you have to start with the right technology. With MarkLogic, you get a single database platform that can store and manage massive volumes of heterogeneous data, perform both transactions and analytics on that data, and still maintain the security and reliability of a trusted legacy database—all the critical elements that you need to have a data-centered data center.

## FLEXIBLE DATA MODEL FOR ALL YOUR DATA

The first problem that needs to be solved is the proliferation of various data types. MarkLogic is designed to handle today's data variety—including documents, relationships, and metadata. MarkLogic is the only database that can store and rapidly query XML, JSON, RDF, and geospatial data—providing a single powerful platform for all of your data. MarkLogic can ingest the two most popular forms of document data—XML and JSON—and can also ingest other sorts of data, from RDF relationships to text, geospatial data, binary video files, and PDFs—without the need for conversion. With MarkLogic, you start with more answers available because you start with better data.

## MASSIVE SCALABILITY AND ELASTICITY

MarkLogic addresses the problem of volume by employing a flexible, shared-nothing, scale-out architecture to handle data across the information lifecycle—without breaking the bank. Unlike relational databases that scale up with larger infrastructure purchases, MarkLogic is designed to cluster across commodity hardware that is much less expensive than the "big iron" required to scale relational databases. If you have more data coming in, you can just add another node to the cluster. And, it's very easy to scale back down and redistribute the workload as necessary. MarkLogic runs production applications with over a hundred nodes, billions of documents, and hundreds of gigabytes of data. But, it is also easy to start small with a three node cluster that can be spun up on Amazon Web Services—using MarkLogic's pre-configured AMIs—in just a few minutes.

## MIX OPERATIONAL AND ANALYTICAL WORKLOADS

In a data-centered data center, the database should be able to handle any type of workload, whether operational or analytical. But traditionally, databases were optimized for one or the other, not both. Gartner even has two different quadrants for each type—one for operational databases and one for data warehouses. With search built-in as a ▶▶

MarkLogic

core component, MarkLogic is designed to handle both operational workloads *and* analytic queries. ACID transactions, memory-buffered writes, and distributed architecture make MarkLogic a fast transactional database. Complex transactions are not a problem, and production applications regularly clock tens of thousands of transactions per second. But it's not just speed—it's also about complexity—and MarkLogic's composable indexes can serve results to complex queries with tens of thousands of operators across huge amounts of data in sub-second time.

## MANAGE THE INFORMATION LIFECYCLE WITH TIERED STORAGE

Data is not static. It must be moved around as its usefulness changes over time. This means the database must be flexible enough to move the data from an operational tier where it's available for low-latency transactions on down to an archive tier where it may only be used later for batch analytics. Traditional databases are not designed to efficiently move data between storage tiers, and data older than 90 days is often stored on tape archives never to be seen again. But, MarkLogic's tiered storage solves this problem by efficiently storing and managing data across the information lifecycle. You can cost-effectively meet SLAs using expensive fast storage such as SSDs for your most recent operational data, and move older data to less expensive storage such as Amazon S3 or the Hadoop Distributed File System for archival purposes.

## MOVE FASTER WITHOUT SACRIFICING DATA GOVERNANCE

MarkLogic was designed from the start to be an enterprise database, and includes all of the enterprise features necessary to run at the heart of the data-centered data center, including government-grade security, high availability and disaster recovery, and ACID transactions. The "move faster and break things" mantra that is often embraced by young startups does not apply to MarkLogic. MarkLogic has been hardened by over a decade of development. And, MarkLogic has been proven in some of the most demanding, mission-critical applications, by some of the most demanding customers in the world. MarkLogic runs at the heart of large investment banks, major healthcare organizations, and classified government systems.

## LEADING ORGANIZATIONS WITH REAL RESULTS

The data-centered data center is not a future aspiration—leading organizations have already implemented it and they are achieving real results. One example is with a Top Five Investment Bank that faced complexity, performance, and cost problems with their operational trade store at the heart of their $87 trillion derivatives business. They had an army of DBAs to maintain 20 Sybase databases that were showing poor performance and instability, and were expensive to maintain. They could not analyze their total market exposure and launch new products in an aggressive marketplace, causing them to be at risk and lag behind competitors.

After implementing MarkLogic, the bank had a simplified, unified platform to power a data-centered data center. MarkLogic's schema-agnostic data model allowed the bank to store data from multiple front-office systems in a single unified system, eliminating data silos. The bank was able to achieve more agility and better risk management because they could develop applications faster. They could easily ingest new data sources and still maintain an integrated, unified view of trade data in order to monitor exposure in real-time. The bank also achieved a massively reduced cost of ownership. Instead of using 20 Sybase databases, they used one single MarkLogic database. And, subsequently, they were able to reduce the number of DBAs required from about 10 to one, helping them save millions of dollars. Furthermore, the bank's new system had incredible performance, handling over 1,600 requests per second while still being available to run ad-hoc analytics as necessary.

## CHANGING ECONOMICS, NEW OPPORTUNITIES

The economics of the data center are changing. Today, we have cheaper hardware, cheaper disk storage, and cloud providers that are lowering prices every day. The biggest cost of information technology is no longer with storage, compute power, or data transfer speeds—it is in the time and development effort of building new applications and the opportunity cost of not moving at the speed of the business. The data-centered data center makes organizations more agile. It makes it easier to find new applications for available data and helps deliver applications more quickly. It lowers the overall IT cost and provides a competitive advantage.

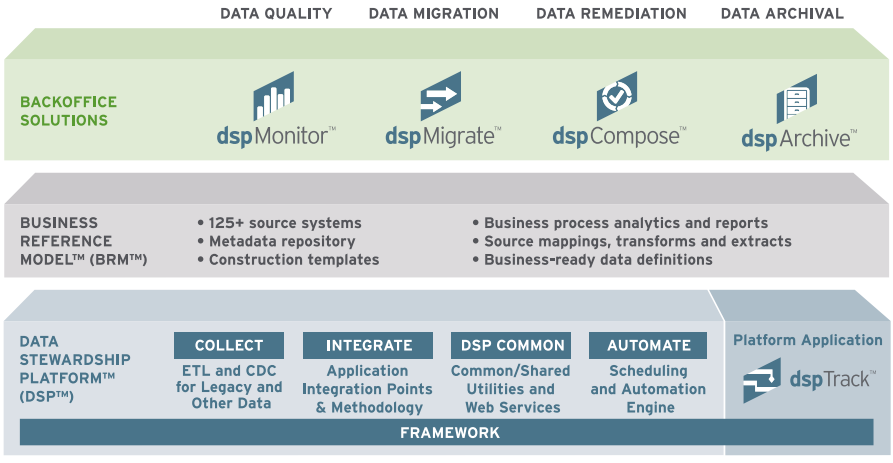
MarkLogic truly stands alone in its capability to power the data-centered data center, as shown by the fact that it is the only technology to appear in three different Gartner Magic Quadrants. MarkLogic is in the Gartner Magic Quadrant for Operational DBMS, Enterprise Search, and Data Warehouse DBMS. In the Magic Quadrant for Operational DBMS, MarkLogic was positioned as a Leader—the only NoSQL database with this distinction. Furthermore, in the Gartner report, *Critical Capabilities for Data Warehouse Database Management Systems* (August 2014), MarkLogic was rated as the number-one operational data warehouse by customers, beating out leaders such as Teradata, Microsoft, and Oracle. With MarkLogic, organizations are able to reimagine their applications because they start by reimagining their data. ■

MARKLOGIC
www.marklogic.com

BackOffice ASSOCIATES

# Information Governance

## *Making the Most of Modern Data Architectures*




**In the world of IT,** there is a never-ending stream of intricate steps to access, combine, update, visualize and overall massage business data as it flows in from remote, legacy or transactional systems and out through analytics and reporting. These steps are necessary to correctly return the right data to the right person at the right time. With the advent of non-structured data, these tasks become more complex and time-consuming, even with the amount of automation available to handle them. Just the size of datasets are becoming too large to handle in ways past—new storage pools, new integration styles and new methods for determining data's value now add significant learning curves to the IT world.

In order to take advantage of new data systems, modern architectures and analytic platforms available, and still return correct data at the right time, it is imperative that businesses take an "Information Governance" approach to their data. When engaging both business users and IT in the activity of determining data value, remediating data errors and establishing solid processes around data quality through the use of automation, a number of key milestones are achieved:

1. ***Knowledge about data and its value is collaboratively shared and managed.***
2. ***Time and effort to identify and remediate data errors is reduced.***
3. ***Programs and projects that depend on clean data achieve faster time-to-value.***

An Information Governance approach, using a data stewardship platform as the underlying coordination and connective tissue for data and activities, enables both business users and IT to unlock the power of their data. A data stewardship platform delivers both the process and automation to help business users quickly and easily see and dissect data errors, and provides click-through rules editing and management. A platform connects all the information critical to data management, and enables IT to understand and manage complex data issues with simplified views of any data activity. It provides on-the-fly reporting configurability, saving hours and days of reporting management.

Data stewardship is a difficult task and requires underlying knowledge of data value across domains in an organization, plus expertise in managing data remediation across many business applications as well as source data systems. A myriad of tools are available for each step in stewardship tasks to remediate data errors, but independent tools require IT to install, integrate and update as necessary. A data stewardship platform reduces that number of tools into one system with automated activities such as mapping and project management, standard integration to modern data platforms and architectures, common access to data design for both business users and IT, ongoing reusability of data designs and system types, and an intuitive web interface that is easily managed by any stakeholder across the organization.

With technology becoming more and more pervasive and available through the Cloud, organizations need to address business issues such as global access to information, consolidation of remote systems and data, divestiture of identified data due to merger/acquisition activity and commonality of systems and process across a distributed workforce, using methodology that accommodates the greatest number of users with the most efficiency, while enabling fast updates that mirror the speed of business. This can only happen through Cloud systems. A data stewardship platform with content available and collaborated on through the Cloud is the answer.

### A DATA STEWARDSHIP PLATFORM

- Makes complex data issues simple for both business users and IT
- Reduces the number of steps and tools to achieve data quality or governance
- Provides an easy, intuitive and collaborative workspace with built-in process
- Reduces reporting workload
- Provides full visibility across global activities and projects with metrics and dashboards

## POWERED BY THE DATA STEWARDSHIP PLATFORM

The Data Stewardship Platform delivers a cohesive way to optimize data quality, migration and governance initiatives. ■

BACKOFFICE ASSOCIATES
www.boaweb.com

Couchbase

# The Tenants of Modern Data Architecture

**What is a modern data architecture?** It's a data architecture created to support future requirements. It's flexible, but it's powerful.

A modern data architecture leverages polyglot persistence. Today, a single database can no longer meet enterprise requirements. The requirements of interactive applications, enterprise, mobile, and web are met by operational databases. The requirements of business intelligence applications are met by analytical databases and/or Apache Hadoop distributions. Finally, off-the-shelf and legacy application requirements are met by relational databases. While a NoSQL database like Couchbase Server is a general-purpose database, it works well with Apache Hadoop distributions for analysis and Apache Lucene distributions for full text search.

A modern data architecture supports agile development. Today, interactive applications are developed by reading from and writing to NoSQL databases like Couchbase Server. These databases support flexible data models, enabling developers to define the data model on the fly and without delay. They are no longer constrained by schemas and change requests.

A modern data architecture is provided as a service. Today, enterprises expect IT to provide database-as-a-service. The consumers are no longer the applications. They are the business units. The database infrastructure is abstracted away from the consumers. The database can be deployed to bare-metal servers, virtual machines, or cloud infrastructure. The database can be deployed to multiple regions via multiple data centers. Regardless of the physical deployment, business units consume databases as a service, not as infrastructure.

A modern data architecture enables a flow of data. There is data at rest: the data stored in an operational or analytical database. There is data in motion: the data moving through messaging systems, the data moving through stream processing systems. A modern data architecture must enable and understand the flow of data. The data may pass through a message queue and a stream processor before it is stored in an operational database. Next, it may be imported, incremental or batch, into an analytical database and/or Apache Hadoop distribution. However, the flow may not be linear. It may be circular. The data may be refined in an analytical database in order to update the data in an operational database. The analytical database may extract data and import it into the operational database to improve operational efficiency. The data will pass through more than one system, and it may pass through a system more than once. It may be processed, filtered, transformed, and more as it flows from the producer to the analyst and back.

A modern data architecture is adaptive. The data no longer defines the applications. The applications define the data. While relational databases support off-the-shelf and legacy applications, NoSQL databases like Couchbase Server enable new applications and new features to define new types of data: social, scientific, energy, telemetry, sensor, geo-location, and more. The data may evolve without having to make changes to the data architecture.

A modern data architecture is elastic. In a global economy within the information age, enterprises may experience immediate hyper-growth or sustained long-term growth. An online retailer may experience increased requirements before, during, and after Black Friday. A video game developer may experience explosive growth when a game goes viral. A payment process may experience sustained growth as the service is made available in more and more markets. As the enterprise moves away from mainframes, it is not moving away from large databases engineered to scale up. It is adopting distributed databases running on commodity hardware.

In conclusion, a modern data architecture leverages polyglot persistence, supports agile development, is provided as a service, enables a flow data, is adaptive, and is elastic. By doing so, it enables enterprises to meet future requirements with less overhead, less cost, and less time.

Couchbase Server is a scalable, enterprise-grade NoSQL database. It supports a flexible data model via documents and JSON. It integrates with Apache Hadoop distributions, and is certified by Cloudera for Cloudera Enterprise. It integrates with LucidWorks and Elasticsearch products. It can be integrated with Apache Storm. It is a key component of real-time big data architectures at eBay, PayPal, LivePerson, LinkedIn, and more. It's the foundation of a modern data architecture. ■

COUCHBASE
www.couchbase.com

snowflake

# Moving to a Modern Data Architecture

The explosion and rapid evolution in data and analytics is forcing businesses to reexamine their data warehousing solutions to determine what is needed to manage and leverage their data to strategic advantage.

## THE CHANGING NATURE OF DATA

It used to be the case that most data you wanted to analyze came from sources in your data center: transactional systems, enterprise resource planning (ERP) applications, customer relationship management (CRM) applications, and the like. The structure, volume, and rate of the data were all fairly predictable and well known.

Today a significant and growing share of data—application logs, web applications, mobile devices, and social media—comes from outside your data center, even outside your control. That data constantly evolves and as a result frequently uses newer, more flexible data formats such as JSON and Avro. That increases demands on both the systems themselves and on the people who manage and use them.

## THE NEED FOR A NEW DATA WAREHOUSE

As companies seek solutions that can accommodate the changes in data and how it is used, they're frustrated to find that most current solutions simply can't keep up. For example, traditional data warehousing solutions were designed for structured, predictable data. Yet companies today also see opportunity to gain insight from machine-generated data, i.e., data generated by web servers, cloud applications, sensors and more. At the same time, newer "big data" platforms like Hadoop are not data warehouses and require specialized expertise that is in very short supply.

As a result, companies are demanding a solution that can work with all business data, handle all their reporting and analytics workloads, and support all of their users in a single solution.

That requires a system that can natively handle both structured and semi-structured data such as weblogs, XML and JSON; scale up and down on the fly to support any number and intensity of workloads; make that data accessible through SQL so that they can leverage their current skills and tools; and do so at a fraction of the cost of traditional systems.

## SNOWFLAKE TAKES A FRESH APPROACH

Snowflake was founded by a team of veteran experts in database technology who saw the opportunity to reinvent the data warehouse to meet these needs. They shared a vision to leverage new and disruptive technology including cloud infrastructure to bring together all users, data and workloads in a single data warehouse without compromising performance, flexibility or ease of use.

The result: the Snowflake Elastic Data Warehouse, a new SQL data warehouse built from the ground up for the cloud to deliver the power of data warehousing, the flexibility of big data platforms and the elasticity of the cloud—at 90% less than on-premises data warehouses.

What makes Snowflake's solution unique?

- **Data warehousing as a service.** Snowflake eliminates the pains associated with managing and tuning a database. That enables self-service access to data so that analysts can focus on getting value from data rather than on managing hardware and software.
- **Multidimensional elasticity.** Snowflake's elastic scaling technology makes it possible to independently scale users, data and workloads, delivering optimal performance at any scale. Run any number of workloads concurrently, including both loading and querying, because every user and workload can have exactly the resources needed, without contention.
- **Single service for all business data.** Snowflake brings native storage of semi-structured data into a relational database that understands and fully optimizes querying of that data. Query structured and semi-structured data in a single system.

Importantly, because Snowflake was built for standard SQL, it delivers these capabilities in a way that takes advantage of the tools and skills companies already have.

## THE CLOUD AS THE FOUNDATION

Cloud infrastructure is the perfect platform for constructing the ideal data warehouse. It delivers near-unlimited resources, on demand, in minutes, and you only pay for what you use.

A data warehouse built for the cloud eliminates the obstacles and pains of deploying and managing infrastructure so that companies can focus on using their data rather than on dealing with infrastructure. In addition, the cloud is the natural integration point for data because a rapidly increasing share of the data comes from applications and systems outside the datacenter—cloud applications like Salesforce, web applications, mobile devices, sensors, and more.

## A NEW STANDARD IN DATA WAREHOUSING

The Snowflake Elastic Data Warehouse brings sorely-needed innovation to data warehousing. Its unique architecture and capabilities offer a highly efficient, cost-effective, and easily manageable data warehouse service designed for today's business needs. ■

SNOWFLAKE
www.snowflake.net @snowflakeDB

IBM Cloudant

# Data Management in the Big Mobile and Internet of Things Era

## DATA MANAGEMENT IN THE BIG MOBILE AND INTERNET OF THINGS ERA

Today's applications are expected to manage a variety of structured and unstructured data, accessed by massive networks of users, devices, and business locations, or even sensors, vehicles and Internet-enabled goods.

As the need for data access expands to the network edge, most databases are still grounded in a central data center. DBAs continue to toil, as they have for decades, moving databases to more powerful hardware, or bigger clusters, and constantly re-designing them in order to accommodate your business growth.

IBM Cloudant is different. It enables applications to operate at this scale, non-stop, by distributing database read/write access out to the edge—like a CDN for your data.

## CLOUDANT—THE INDUSTRY'S FIRST GLOBAL DATA DELIVERY NETWORK

Cloudant is the first data management platform to leverage the availability, elasticity, and reach of the Cloud to create a global data delivery network (DDN) that enables applications to scale larger and remain available to users wherever they are.

- Distributes readable, writable copies of data to multiple locations or devices.
- Users read and write the closest available data source.
- Synchronizes data continuously via filtered, multi-master replication.
- Stores data of any structure as self-describing JSON documents.
- Integrates via a RESTful API.
- Provides full-text search, geo-location services, and flexible, real-time indexing.
- Is monitored and managed for you 24x7 by the big data experts at Cloudant.
- Built on open source. API-Compatible with Apache CouchDB™.

## IDEAL USE CASES FOR CLOUDANT

Companies of all sizes use Cloudant to manage data for many types of large or fast- growing web and mobile apps in ecommerce, online education, gaming, financial services, networking, and other industries. Cloudant is best suited for applications that require an operational data store to handle a massively concurrent mix of low-latency reads and writes. Its data replication and synchronization technology also enables continuous data availability, as well as off-line app usage for mobile or remote users.

As a JSON document store, Cloudant is ideal for managing multi- or un-structured data. Advanced indexing makes it easy to enrich applications with location-based (geospatial) services, full-text search, and real-time analytics.

## SAYING GOODBYE TO MESSY DATA

At Cloudant, our aim is to help you get more from your live data. Data in Cloudant is highly durable and fault-tolerant. Every write is immediately committed to disk, and Cloudant's distributed architecture stores multiple copies of data across different physical infrastructure. All copies do reads *and* writes. It is flexible and scalable and has many different APIs for finding and managing data.

We are excited to announce the integration between Cloudant and the IBM dashDB beta. dashDB is IBM's fully managed data warehouse service in the cloud. Now, you can easily replicate data from Cloudant to dashDB for deeper offline reporting and context. dashDB enables predictive analytics across all data. ■

Cloudant is the first data source to integrate with dashDB. To try both products, sign up at **https://cloudant.com/dashdb-signup/**

CLOUDANT, AN IBM COMPANY
https://cloudant.com