Architecture Patterns for the Next-generation Data Ecosystem

Abstract

Transforming IT systems, specifically regulatory and compliance reporting applications has become imperative in a rapidly evolving global scenario. Business transparency and confidentiality, information security, improved data quality for monitoring and compliance, and advanced predictive capabilities form the focus of regulatory activity in the financial services industry. Financial institutions need to consistently meet these objectives in the normal course as also during stress or crisis scenarios.

Traditionally, enterprise data management solutions with traditional data warehouses and Business Intelligence (BI) applications served as the foundation for banks' information needs. However, new technologies in the Big Data space demand a reengineer of these platforms to meet the ever-growing demand for better performance, scalability, and availability. This paper discusses the challenges inherent to the traditional data ecosystems, the architectural approaches available to address these challenges, and the roadmap for building next-generation data ecosystems.

Why Traditional Data Ecosystems Won't Work Now

Financial institutions are extensively data-driven and need to preserve several years of transactional data for analysis and regulatory compliance. To meet this requirement, firms utilize relational technologies and BI tools. However, a broadened regulatory compliance scope needs financial institutions to link their risk and compliance programs with customer experience management strategies, and integrate the data sourced from diverse systems into a central repository.

Regulators conduct stress tests on banks by simulating various risk scenarios, to check if the bank can withstand adverse economic conditions. These simulations typically generate terabytes of data. Analysis of humongous amounts of both structured and unstructured data is key to effective risk assessment and monitoring. Financial institutions must therefore address the data challenges pertaining to volume, velocity, and variety.

The Big Data Promise

Big Data technologies can help address a majority of these data challenges. Traditional data ecosystems that comprise a staging layer, an operational data store, an enterprise data warehouse, and a data mart layer have coexisted with Big Data technologies. However, most financial institutions are now building and developing advanced Big Data platforms that utilize emerging analytics technologies. Firms have started to create landing and processing zones for enterprise-wide data, external data feeds, and unstructured datasets. This requires massively scalable but inexpensive data storage solutions. Here are some of the most common practices that financial institutions have adopted to leverage Big Data technologies:

Augment the traditional data layer

Using data warehouses with relational or appliance technologies for traditional structured data poses challenges to real-time or near real-time decision making. In addition, it is not ideal for integrating unstructured data coming from social media sites, web logs, and smart ATMs.

To address these challenges, Big Data technologies can be used along with the traditional data warehouse layer consisting of structured data. Additionally, the NoSQL data store can be used for unstructured and semi-structured data where data can be processed either in real-time or in batch mode. This approach ensures seamless integration with the existing technology landscape and efficient use of expensive data warehouse appliances. If you choose this approach, make sure to have a comprehensive roadmap to assimilate the traditional data layer into the next-gen data ecosystem. This will help you to support existing functionalities and create new ones, while keeping the cost of services low.

Complement the existing ETL layer

Existing Extract, Transform, and Load (ETL) routines are not suitable to crunch huge amounts of data, since they work in a batch mode. This results in delayed time-to-market and high storage as well as infrastructure costs. Big Data addresses these challenges by changing the current ETL format to Extract, Load, and Transform (ELT) using the Hadoop Distributed File System (HDFS). With this approach, data can be offloaded to the staging area in the Big Data (HDFS) platform. Big Data scripts (MapReduce or Apache Hive or Apache Pig) can also be used to perform data transformation and aggregation processes. The aggregated data then moves to the Database Management System (DBMS) and data warehouse appliances within the traditional data warehouse.

Conduct exploratory analysis

Advanced Big Data technologies help analyze unstructured data from social media platforms and other web-based channels. A combination of Big Data technologies and Natural Language Processing (NLP) based text analytics is best suited for customer sentiment analysis using these datasets. Big Data technologies offer a problem-specific or exploratory analytics platform to effectively analyze granular structured, unstructured, and semi-structured data. This provides a new dimension to analytics, as it incorporates both structured and unstructured data, and seamlessly integrates with the existing analytics tools.

Enhance business performance

Financial institutions have started using real-time analytics on the data from ATM machines, credit card processing systems, and social media sites. These real-time data points offer an understanding of the trading patterns as mandated by regulations in the capital markets, and help identify credit card frauds through timely alerts.

What the Next-generation Data Ecosystem Will Be Like

In the age of digital banking, live chat with banking staff, 24x7 call centers, video chat with advisors and specialists, smart ATMs, and location-based services are gaining prominence. Each of these initiatives creates large volumes of complex, unstructured text, audio, and video data. To handle the increased range of data and operations, financial institutions require a scalable data ecosystem.

The Big Data backed data ecosystem enables process and storage of large volumes of structured, semi-structured, and unstructured data generated by internal and external systems. What follows is a better understanding of customers' needs and behaviors, and improvement in the overall customer experience.

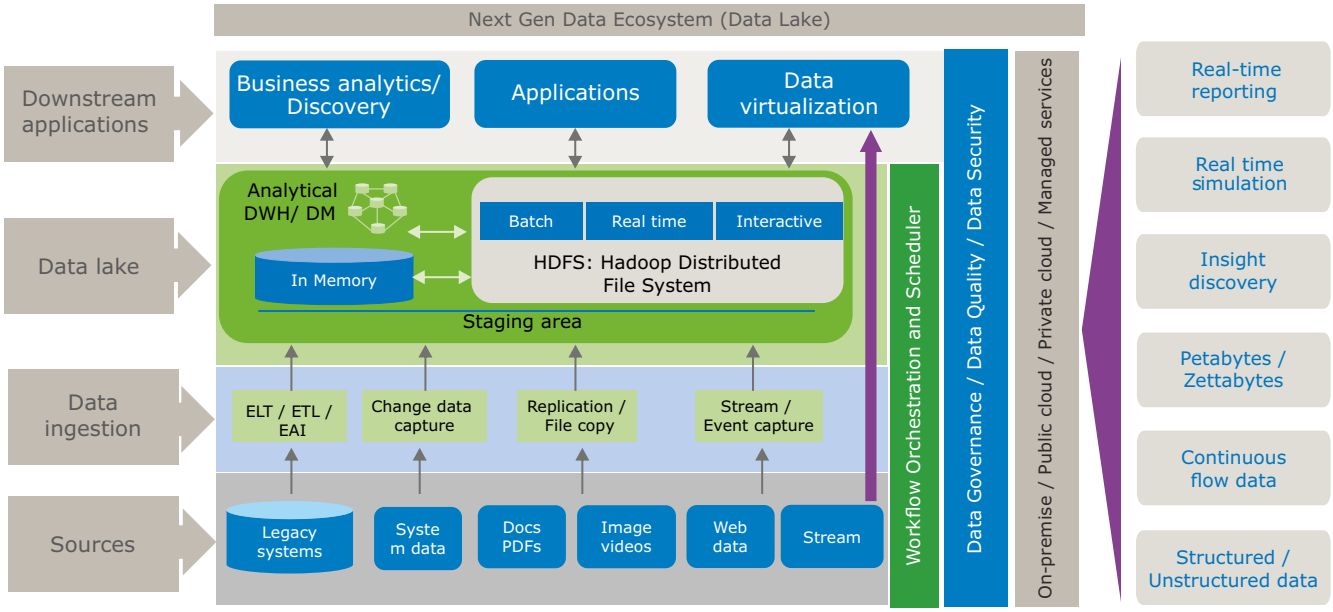

Figure 1: Proposed Architecture of the Next-generation Data Ecosystem (Source: TCS Internal)

In the next-generation data ecosystem (see Figure 1), a Big Data platform serves as the core data layer that forms the data lake. This data lake is populated with different types of data from diverse sources, which is processed in a scale-out storage layer. In the data ingestion layer, data is moved or ingested into the core data layer using a combination of batch or real-time techniques.

Downstream applications such as business analytics and enterprise systems can consume data or information from the data lake by using various services such as Data as a Service, analytical query services, content services, spatial services, and virtualization. The workflow orchestration and scheduler functions specify the sequence for execution of jobs based on their dependencies on the completion of other jobs; the scheduler specifies the starting time for each job. Figure 2 depicts the way the next-generation data ecosystem encapsulates the characteristics of both the traditional data layer (DWH) and the Big Data ecosystem.

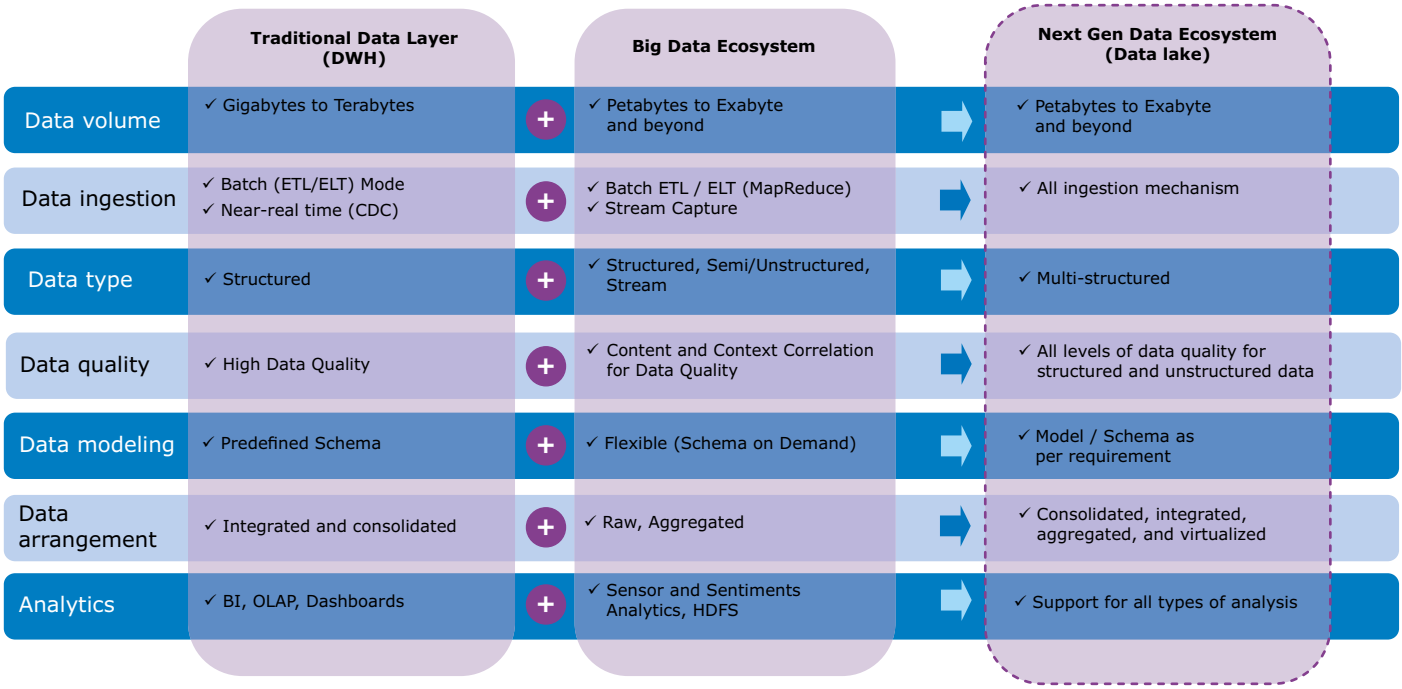

Figure 2: Characteristics of the Next-generation Data Ecosystem
(Source: TCS Internal)

The Way Forward

The new data ecosystem will require firms to institute data governance and stewardship with data interoperability, data trustworthiness, and data security as key capabilities.

Financial institutions will need a comprehensive enterprise-wide data governance framework, the salient features of which are:

Big Data governance council: Identify new roles for implementing the Big Data initiatives and include them in the existing governance council.

Data quality and trustworthiness: Set up processes to enhance the quality of unstructured data coming from unconventional sources.

Metadata management: Ensure data interoperability and define the nature and purpose or context of the structured and unstructured data in the data lake environment.

Data security and privacy: Incorporate effective data authentication, authorization, and encryption capabilities across the next-generation data ecosystem layers.

Information lifecycle management: Define clear policies for the retention, migration, deletion, and archival of data over its entire lifecycle, based on usage pattern across the ecosystem layers.

Financial institutions must evolve their data and information architecture to successfully establish a next-gen data ecosystem. This will enable them to discover business insights and drive quick and responsive business decisions.

About The Authors

Kathirvelan Subramanian

Kathirvelan Subramanian heads the BI CoE's Solution group of TCS' Banking and Financial Services (BFS) business unit. He has over 20 years of experience across business strategy and development, conceptualization and solution development in the Information Management (IM) space, and business consulting. Subramanian's areas of expertise include IM transformation, competency building in emerging technologies, partner collaboration, as well as presales and delivery support. He has a Master's in Computer Applications from the Madurai Kamarajar University, Tamil Nadu, India, and is a PMP® certified professional.

Deepak Srivastava

Deepak Srivastava is a Domain Consultant with the BI CoE of TCS' BFS unit. He has over 23 years of experience with rich experience in the BI and IM area. Prior to his current role, Srivastava has been a data warehouse architect, solution consultant, and delivery manager for a variety of TCS' client projects. His areas of expertise include data and information strategy, architecture development, and large program management.

Contact

Visit the Consulting & System Integration page on www.tcs.com

Email: global.consulting@tcs.com

Blog: #Enterprise Insights

Subscribe to TCS White Papers

TCS.com RSS: http://www.tcs.com/rss_feeds/Pages/feed.aspx?f=w
Feedburner: http://feeds2.feedburner.com/tcswhitepapers

About Tata Consultancy Services Ltd (TCS)

Tata Consultancy Services is an IT services, consulting and business solutions organization that delivers real results to global business, ensuring a level of certainty no other firm can match. TCS offers a consulting-led, integrated portfolio of IT and IT-enabled, infrastructure, engineering and assurance services. This is delivered through its unique Global Network Delivery Model™, recognized as the benchmark of excellence in software development. A part of the Tata Group, India's largest industrial conglomerate, TCS has a global footprint and is listed on the National Stock Exchange and Bombay Stock Exchange in India.

For more information, visit us at www.tcs.com

TCS Design Services I M I 07 I 17