Infosys cobalt

WHITE PAPER

PETABYTE-SCALE DATA MANAGEMENT FOR CLOUD BANKING

Migrating banking data from legacy systems to the cloud is easier said than done, especially if data volumes are in petabyte scale. A change in people, processes and perception is required before banks can begin their journey to the cloud. Navigating these is crucial to a bank's transformation and survival in this customer-centric age.

Infosys®
Navigate your next

Infosys® | Knowledge Institute

Banks view cloud cautiously

Cloud computing has fueled rapid growth and adoption of innovative digital business models. From Uber to Spotify and Netflix to Slack, thousands of digital leaders have built their businesses on cloud software. This is because it enables flexible, scalable, cost-efficient and continuously upgradable capabilities. But for banks, adopting cloud technologies can be a real challenge — not least because mainframes continue to be at the core of their business, given their reliability, almost impregnable nature and ability to manage large volumes of complex tasks. In fact, over 90% of the world's top banks still rely on mainframes.[1] This is despite the fact that mainframes are archaic in nature and present a host of issues.

Challenge of cloud banking

1. Legacy technology

According to an Infosys Knowledge Institute survey, legacy systems currently rank as the third most commonly cited barrier to digital transformation (named by 42% of respondents), but financial services executives expect that they could become the most serious barrier in 2019.[2]

Millions are spent every year to maintain these archaic systems,[3] which are often incompatible with modern technology. These systems act as a barrier to better catering to the needs of the digital customer. The reliance on legacy technology also stops banks from taking advantage of Agile and DevOps ways of working, increased automation, and analytics. According to Gartner estimates, banks need to triple their digital business innovation budgets to modernize legacy applications through 2020.[4]

Experts say that banks spend 80% of their IT budgets on legacy technology maintenance, and a Tier 1 bank could spend up to $300 million a year to update existing software to meet regulatory requirements.[5] But cloud can act as a workaround in reducing costs.[6] For example, the Federal Home Loan Bank of Chicago reduced infrastructure costs by 30% by moving all of its internal production workloads to the cloud.[7]

2. Data management

With data growing by 2,500 petabytes a day,[8] data management is increasingly becoming a concern for banks,[9] not least because of regulatory and security requirements. Regulators require banks to provide detailed reports, additional information on stress testing and information at a granular level.[10] Banks continue to capture data in multiple forms (customer personal information, transaction history, journey maps, market data), and this huge volume of data is now stored in data warehouses and lakes.

Banks store data in various locations, and it's often accessed by different users, creating a siloed approach and multiple layers of data duplication.[11] Singular data warehouses and lakes are formed, giving rise to further issues of storage, meeting regulatory requirements, authenticity, incompatible formats and higher time to compute. These have an indirect bearing on run-the-bank costs.

The data collected by banks must also be analyzed and deployed to facilitate better decisions.[12] However, due to legacy technology, banks are unable to leverage access to this data for analytics and insights and improve customer experiences.

3. Manual data migration

Currently, when a bank decides to move data to a new system — whether it's in the cloud or on-premises — a new project team is set up and works in a silo. Data is migrated from the data lake into the new system, or a new connection is created between them. Because of the silos in the bank, other teams using the data may be unaware of this migration or new connection.

In many instances, multiple teams want the same data and follow a similar process without getting connected to each other. This leads to duplication of data, with multiple copies created in the data lake and transported to different systems either in the cloud or internally.

Banks' manual and siloed approach of ingesting and migrating data is time-consuming. It is also counterproductive, as in this digital age, customers' demand for instant, real-time, and personalized products and services has become the new normal.[13]

4. Lower number of legacy coders

Legacy systems are built on outdated languages such as COBOL. For instance, the U.S. financial sector has over 200 billion lines of COBOL code currently active, and COBOL powers over 90% of ATMs.[14] Despite COBOL's widespread use, today's coders prefer to use new languages that are compatible with artificial intelligence, machine learning or cloud computing. Few people want to learn a language that can talk only to legacy technologies, and coders familiar with COBOL can be well into their 50s or 60s,[15] presenting significant skills shortage challenges in maintaining older technologies.[16] Although upgrades are available, they are still not fit enough to compete or converge with systems of this digital era.

5. Security concerns

Customer data in any form is sensitive for banks. The major cloud vendors, such as Google and Microsoft, have outstanding security expertise, and all are certified compliant with federal data governance standards. However, for years banks have delayed and tried to avoid the issue of modernizing their infrastructure. While they agree that cloud infrastructure offers them the ability to better cater to the needs of the digital customer, they are reluctant to adopt it until they are convinced of the safety and security of their data. They do not have a clear strategy that will help them quickly adopt cloud.

6. Agile and DevOps ways of working

Legacy systems are expensive to maintain and delay a product's time to market.[17] They are also not ideally suited to Agile programming methods, instead relying on waterfall methods that can slow down the production of software and result in less timely feature releases.

Banks that have made the shift to Agile and DevOps ways of working have benefited. In 2012, J.P. Morgan followed a quarterly software-release cycle, and coordination between development and operations was minimal. These quarterly releases heightened risks, were cumbersome and time-consuming, and increased delivery costs. The financial institution decided to embrace Agile and DevOps practices. The software-releases cycle changed from quarterly to 100 releases in 2015, 200 in 2016 and over 400 in 2017.[18]

Capital One's move from waterfall to Agile software development helped cut the time to build new application infrastructure by 99%. DevOps' automation and continuous integration of new code helped speed the bank's development cycles, and releases occurred with increased frequency and higher reliability.[19]

7. Cultural shift

Banks need to undergo a perception change. They must be ready to build a culture that is cost-conscious, customer-conscious and efficiency-conscious.[20] This is easier said than done, as many systems, processes and people have grown with the bank. A Boston Consulting Group assessment revealed that among companies that underwent a digital transformation, the number of profitable enterprises was five times higher among those that focused on a cultural shift compared with those that did not.[21]

How can banks benefit from cloud?

Moving data and applications to the cloud can save banks money. Some say it could cut IT costs by as much as 75%.[22] A large global bank that Infosys partnered with to solve this problem estimated that it could save 50% of costs overall by adopting cloud. Some areas in the bank expect to save 90% of their costs in the transition. To achieve this, the bank is working with Infosys to build a multi-cloud data management system that interfaces with Google, Amazon and Microsoft clouds to migrate its data.[23]

Cloud is pivotal in architecting the bank of the future. Banks can benefit from the move to the cloud not only by reducing costs, but also by capitalizing on cloud's computing capabilities, its ability to scale IT solutions and its reliability. In fact, moving to the cloud can make banks more like the fintech upstarts with which they increasingly compete. Founded in 2015, fintech company Monzo has an infrastructure that is capable of serving 1.7 million customers supported by only 10 people on its infrastructure and reliability team. The bank has 400 core banking microservices on the cloud, which help it deliver value to customers in the form of offerings such as instant balance inquiry and real-time statements.[24]

Issues faced by an Infosys banking client

The bank that Infosys partnered with to build the open source data management platform started off with the classic challenges faced by many of its peers. Project teams worked in silos and used various tools and software. For moving data, Ab Initio was used; for tagging, Collibra; and for scheduling, Control-M. There were nearly 20 moving parts, with licenses attached to each.

The bank generates and moves terabytes of data each day. Each process of data ingestion took eight to 12 weeks to deliver and was cumbersome, time-consuming and counterproductive.

Ushering in the Infosys open source data management platform

To solve this problem, Infosys built an open source, petabyte-scale multi-cloud data ingestion and management platform, part of Infosys Cobalt. It is a metadata-driven data management ecosystem that has been designed to meet an organization's current and future data delivery requirements. The platform allows businesses or functions within banks to move data from a source to a destination in a defined format at an agreed frequency.

The platform was built to solve a host of banking issues, beginning with data management. It provides a central way to monitor all banking data and enable it to be ingested in the cloud without duplication. The platform has also enabled the implementation of a Trusted Source Framework which helps with data lineage. This allows users to track data usage, understand who has made the last change, how the data has been tagged and better manage the single view of the available data.

The architecture behind Infosys' open source data management platform

The data management platform first targets the ingestion problem — taking data and moving it to cloud or on-premises, ingesting it on cloud systems such as Hadoop, Google Cloud Platform (GCP), Amazon Web Services (AWS) and Microsoft Azure.

Second, it focuses on the data management issue. The platform enables automated ingestion of data versus the manual and siloed approach previously followed by banks. It also provides a platform and interface to ingest data centrally.

The platform can replicate data from on-premises to a multi-cloud environment, while supporting batch and near-real-time movement. It was built with the purpose of enabling guaranteed data delivery at scale. The data management platform carries out various functions, removing the need to use multiple types of software for each function and saving licensing costs.

The platform acquires, ingests and transforms data in the following stages:

1. Acquire

Banking transactional data is stored on multiple databases and interchange formats built on legacy mainframe technology. Structured and unstructured data is stored in big data Hadoop platforms, Oracle databases, DB2, etc.

The platform interfaces with the databases to acquire the stored data.

2. Ingest

Once the raw data is pulled in from on-premises, it needs to be stored in a format that is durable and can be easily accessed. The platform's architecture ingests this data in various phases:

- Data Extraction
  - Structured, unstructured or semi-structured data is extracted or replicated from various databases and source systems. The data management platform's easy-to-use interface supports interaction with over 25 source

systems including relational database management systems (e.g., Oracle, Teradata, MS SQL Server, Hadoop, HDFS) and multiple clouds (GCP, Azure, AWS). Data is extracted or replicated in near-real time using the Kafka processing engine, while NiFi is used to process batches.

- Specific datasets can be extracted with the user-friendly interface that enables developers and business users to create ingestion pipelines and track their movement in real time.
- The interface allows seamless movement of data at petabyte scale at an accelerated pace from different on-premises systems to on-premises Hadoop data lakes or to multiple cloud platforms (GCP, AWS, Azure).
- Unlike monolithic applications built as a single unit, the Infosys platform's microservices-driven architecture helps reduce development time. Its suite of services, each independently run and deployable, reduces the dependency on skilled developers to deliver data movement.

• Data Masking

- Data masking helps financial institutions protect restricted and sensitive customer data — including Personally Identifiable Information (PII) — prevent unauthorized data access and avoid unwarranted data exposure. As a result, banking fraud is reduced.
- Data is masked before it is ingested into on-premises locations or the cloud. The platform's Key Management Service (KMS) masks data in real time, and data can be unmasked only for authorized users.
- This helps banks comply with cybersecurity and regulatory requirements and helps them build trust with users that the data access path is secure.

• Data Lineage

- Data lineage shows the life cycle of data, i.e., its origin, where it has moved over time (within on-premises locations or cloud), what actions have been performed on it and its final destination. It helps trace data back to its original source (whether on-premises or in the cloud), reconciles the data, reduces duplication and traces errors back to their sources quickly.
- Another important facet is "explainability." As there are multiple dependencies on specific data, data lineage helps banks explain why certain decisions were made. It also helps banks comply with regulatory requirements of maintaining and managing customer data, e.g., the General Data Protection Regulation (GDPR).
- Infosys' open source data management platform's data lineage is built using Java, Python and D3.js library.

• Job Scheduling

- This helps automatically trigger data transfers on a recurring or an ad hoc basis — daily, weekly, monthly or based on the occurrence of any event.
- Banks need to process large volumes of transactions quickly, without any errors or any downtime. Automating the job scheduling ensures that necessary data is transferred efficiently, to the right place at the right time.
- If a server fails, the disaster recovery procedure is triggered and switches the job loads to the disaster recovery servers.
- The platform's job scheduler uses an event-driven architecture to schedule jobs.

• Data Encryption

- Restricted and sensitive customer data, including PII, is ciphered at multiple levels while at rest and in transit to ensure no data breach can occur. In fact, the Gramm-Leach-Bliley Act (GLBA) in the U.S. also requires institutions to protect customers' Nonpublic Personal Information (NPI).
- The KMS complies with all banking standards. It uses a 4096-bit RSA Key Vault, ensuring tamper-proof protection and encryption of data. This is higher than the industry standard of 256-bit encryption. The platform's KMS has separate modules for data masking and data encryption.
- Data in transit to the cloud is encrypted using Transport Layer Security (TLS) and authenticated using certificates to ensure communication between the client and the server is trusted. Cloud certificates are stored using a double encryption key.

• Real-Time Streaming

- Infosys' open source data management platform is built to move data continuously and in near-real time from source to destination. To accelerate data movement from on-premises to cloud, the platform uses Infosys-developed open source components, NiFi, Kafka and cloud.
- Changes can often represent only a small portion of the total data volume. The data management platform uses Infosys-developed open source components that read logs, and then replicate and mirror changes to the cloud.

- Cloud provides a staging location for data on its journey toward processing, storage and analysis.

3. Publish, Transform and Manage

Stored data is transformed into actionable information, and the results are converted into a format that is easy to draw insights from.

- Data Publishing
  - Data ingested and stored is now cleansed and organized in cloud databases or data warehouses. Also, targeted tables are created in targeted databases based on metadata information provided. The cloud data manager uses Google BigQuery to interact with and analyze huge volumes of stored data.
  - This function is built using Java, Python, Cloud SDK, Cloud native tools and Google BigQuery.
- Data Masking
  - Data is again masked after it is ingested into on-premises locations or the cloud. The KMS masks data in real time, and data can be unmasked only for authorized users.
- Data Profiling
  - Data profiling helps assess the quality and relationship of the available data. Data profiling jobs are undertaken in the platform's Scheduler.
- User Management
  - The user management function authenticates and onboards users based on an organization's Active Directory. It also assigns roles or provides secure access to various features of the platform that users are authorized to work on.
  - The data management platform uses Java Spring Boot, Active Directory, Cloud Identity & Access Management (IAM) and KMS for this functionality.
- Metadata
  - This function stores business and technical metadata for data ingested, processed and scheduled by the platform.
  - It helps trace the data lineage and log and track data, and provides dashboards to track the feed status.
  - Infosys' open source data management platform uses Postgres or MySQL databases to store metadata.
- Data Lineage
  - Data is unmasked after it is ingested into on-premises locations or the cloud. The KMS allows only authorized users to unmask the data.
  - This helps banks comply with cybersecurity and regulatory requirements and helps build trust with users that the data access path is secure.
- Business Glossary
  - The platform captures the business glossary and moves along with the data to the destinations on-premises and in cloud. It tags the business glossary for the data and attributes ingested, and is also helpful for audit purposes.
  - Infosys' open source data management platform uses Java and an Open Source Business Glossary Model to save the business glossary.

Benefits of Infosys' open source data management platform

- With the data management platform, a large global bank reduced its data migration cycle by over 75% and decreased its daily liquidity reporting time by nearly 80%.
- The platform improved data management at the bank, becoming a one-stop shop for data migration. This resulted in reduced cost of storage and operations, and improved data lineage.
- It also helped Infosys' client save $8.5 million on license renewal costs for the Ab Initio software that addressed real-time data processing and application integration.
- With its petabyte-scale transfer ability, Infosys' open source data management platform has delivered significant benefits to the bank and helped move hundreds of applications and petabytes of data to the cloud.

Future of banking with Infosys' open source data management platform

With Infosys' data management platform, banks can now migrate volumes of data back and forth quickly at scale and deploy artificial intelligence and machine learning to analyze data, provide better insights and make better decisions. As a result, banks can begin to act truly as digital-first companies, much like the fintech competitors they increasingly face in the market. As an open source solution, Infosys' data management platform will benefit from the contributions of the open source

community and other banks that choose to test and use it. We hope that in the future, it will become a standard platform that enables traditional large banks to engage with the cloud at petabyte scale.

References

1. "Two-platform IT: Why the mainframe still has its place in the modern enterprise," Information Age, April 12, 2018, https://www.information-age.com/mainframe-modern-enterprise-123471427/
2. "Infosys Digital Radar 2019: Barriers and Accelerators for Digital Transformation in the Financial Services Industry," June 2019, https://www.infosys.com/about/knowledge-institute/insights/Pages/financial-services-industry.aspx
3. "Banks face spiraling costs from 50-year-old IT," Financial News, October 2, 2017, https://www.fnlondon.com/articles/banks-face-spiraling-costs-from-archaic-it-20170912
4. "Infosys Digital Radar 2019: Barriers and Accelerators for Digital Transformation in the Financial Services Industry," June 2019, https://www.infosys.com/about/knowledge-institute/insights/Pages/financial-services-industry.aspx
5. "Banks face spiraling costs from 50-year-old IT," Financial News, October 2, 2017, https://www.fnlondon.com/articles/banks-face-spiraling-costs-from-archaic-it-20170912
6. "Consumers Want An Experience That Legacy Banking Systems Can't Deliver," The Financial Brand, April 2, 2018, https://thefinancialbrand.com/71829/legacy-systems-banking-customer-experience-impact/
7. "AWS Case Study: Federal Home Loan Bank of Chicago," Amazon Web Services, https://aws.amazon.com/solutions/case-studies/FHLBC/
8. "Endless possibilities with data: Navigate from now to your next," Infosys Knowledge Institute, November 2018, https://www.infosys.com/data-analytics/insights/Documents/endless-possibilities-with-data.pdf
9. "Banks Kept On Their Toes By Dizzying Data Management Regulations," PYMNTS.com, December 5, 2017, https://www.pymnts.com/news/b2b-payments/2017/wolters-kluwer-bank-data-management-regulation/
10. "Regulatory Data Management: Data Quality and Integrity Concerns for Asian Banks," Moody's Analytics, April 2019, https://www.moodysanalytics.com/articles/2019/regulatory-data-management
11. "Data Management Challenges For Financial Services," Digitalist Magazine, January 15, 2019, https://www.digitalistmag.com/customer-experience/2019/01/15/data-management-challenges-for-financial-services-06195118
12. "Why aren't banks making the most of data?," Raconteur, June 18, 2019, https://www.raconteur.net/finance/data-management-banking
13. "Digital Banking Transformation: Redefine the Banking Core," Infosys Knowledge Institute, April 2019, https://www.infosys.com/about/knowledge-institute/insights/pages/digital-banking-transformation.aspx
14. "COBOL blues," Reuters, http://fingfx.thomsonreuters.com/gfx/rngs/USA-BANKS-COBOL/010040KH18J/index.html
15. "Do You Know Cobol? If So, There Might Be a Job for You." The Wall Street Journal, September 21, 2018, https://www.wsj.com/articles/do-you-know-cobol-if-so-there-might-be-a-job-for-you-1537550913
16. "Legacy systems are a pain in the bank," Finextra, October 26, 2018, https://www.finextra.com/blogposting/16205/legacy-systems-are-a-pain-in-the-bank
17. "Infosys Digital Radar 2019: Barriers and Accelerators for Digital Transformation in the Financial Services Industry," June 2019, https://www.infosys.com/about/knowledge-institute/insights/Pages/financial-services-industry.aspx
18. "How J.P. Morgan Asset Management went from quarterly to daily releases," TechBeacon, 2018, https://techbeacon.com/devops/how-jp-morgan-asset-management-went-quarterly-daily-releases
19. "On-Demand Infrastructure on AWS Helps Capital One DevOps Teams Move Faster Than Ever," Amazon Web Services, https://aws.amazon.com/solutions/case-studies/capital-one-devops/
20. "Digital Banking Transformation: Redefine the Banking Core," Infosys Knowledge Institute, April 2019, https://www.infosys.com/about/knowledge-institute/insights/pages/digital-banking-transformation.aspx
21. "It's Not a Digital Transformation Without a Digital Culture," Boston Consulting Group, April 13, 2018, https://www.bcg.com/publications/2018/not-digital-transformation-without-digital-culture.aspx
22. "Banks face spiraling costs from 50-year-old IT," Financial News, October 2, 2017, https://www.fnlondon.com/articles/banks-face-spiraling-costs-from-archaic-it-20170912
23. "Banks face spiraling costs from 50-year-old IT," Financial News, October 2, 2017, https://www.fnlondon.com/articles/banks-face-spiraling-costs-from-archaic-it-20170912
24. "How Monzo built a digital bank on AWS for over 500,000 customers," Amazon Web Services, https://aws.amazon.com/solutions/case-studies/monzo/

Authors

Ajay Vij

SVP & Industry Head – Financial Services
Ajay_Vij@infosys.com

Mohammad Faizan

Senior Manager, Client Services – Financial Services
Mohammad_Rahim@infosys.com

Jitendra Raisinghani

Principal Technology Architect – Financial Services
Jitendra_Raisinghani@infosys.com

Sharan Bathija

Senior Consultant – Infosys Knowledge Institute
Sharan_BP@infosys.com

About Infosys Knowledge Institute

The Infosys Knowledge Institute helps industry leaders develop a deeper understanding of business and technology trends through compelling thought leadership. Our researchers and subject matter experts provide a fact base that aids decision making on critical business and technology issues.
To view our research, visit Infosys Knowledge Institute at infosys.com/IKI

Infosys Cobalt is a set of services, solutions and platforms for enterprises to accelerate their cloud journey. It offers over 14,000 cloud assets, over 200 industry cloud solution blueprints and a thriving community of cloud business and technology practitioners to drive increased business value. With Infosys Cobalt, regulatory and security compliance, along with technical and financial governance comes baked into every solution delivered.

For more information, contact askus@infosys.com

Infosys®
Navigate your next