indium
Make Technology Work

Data Lakes for Digital Banks

White Paper

www.indiumsoftware.com

Neo banks (Digital banks) are agile and lean, and becoming popular due to their ability to understand customers better and provide a great banking experience. However, as they grapple with huge volumes and varieties of data that can potentially improve their service levels, these banks need well-designed data lakes and visualisation capabilities to leverage.

The Neo Wave

Banking is serious business. Even less than a decade ago, a bank needed to have offices across the country to be able to service clients. Clients needed to fill up forms and provide identity proof to be able to open accounts. Then ATMs (Automated Teller Machines) came up and at least a part of the transaction could be done without going to a branch. Online banking further eased banking, but there was still a brick and mortar bank in the background that processed and kept track.

Then came a wave of fintech companies, changing the way money was handled. From online, cashless transactions, they are now venturing into online only banks. According to a CNBC report, the areas big banks avoided following the financial crisis of 2008 gave lean and young startups to venture into and expand. Their gaining momentum is summarised in the fact that venture capital funding of fintech firms touched $7.5 billion, and cryptocurrency raised $12 billion in 2018 alone.

A Markets and Markets Research report shows that the digital banking platforms market will be worth $5.7 billion by 2023. The digital banking platforms market is expected to grow at a Compound Annual Growth Rate (CAGR) of 11.2 percent from $3.3 billion in 2018 to $5.7 billion by 2023. The customers, especially millennials, are gravitating towards these banks due to the ease of opening an account and the comfort of operating it from anywhere, all their transaction details are easily available with a swipe on their smartphones.

Customer at the Centre

Traditionally, customers had no choice in the matter of banking services, taking the best from what was being offered. But today, technology has put the customer right in the centre, analysing user behaviour and customising services to suit their requirements. To enable building a scalable and innovative bank around data that offers customers with exceptional services, offers and experiences, a data, analytics and AI approach can help. This can help provide:

- Dynamic real-time and/or near real-time insights on dashboards including operational reporting like batch, structured and fixed reports
- Deep intelligence/ insights available, high-level summary reports, if needed, for legal/ regulatory purposes
- Continual data lineage that ensures visibility from data origin to $ impact.
- Scalable & flexible analytics platform optimized for innovation
- Support for data by design, embedded analytics, data discovery and test & iterate principles.

Operating without branches and not needing traditional banking infrastructure, the neo, digital or challenger banks – as they are called - operate without any branches and are leaner as a result.

This initially requires building the operational data store (ODS) and ensuring that core banking operational data is stored in real-time or near-real-time. Operational reporting and dashboards on the ODS layer will provide every business user with access to insights generated from the transactional data.

Data Management for Insights

Data is the key to be able to generate insights that will enable designing better products, services and strategies that will win and help retain customers.

However, data itself can pose a challenge as it is available in multiple formats – structured, unstructured, semi-structured, as text, images, videos and in other formats. Traditional databases are not equipped to handle such a large variety of data, as well as the volume that flows into an organization due to access to external sources as well. So in addition to an Enterprise Data Warehouse to store the structured and transformed data, the neo banks also need a Data Lake to store and manage the unstructured data and the raw data in all formats.

Data Lake vs Data warehouse

A data lake is a central location to store all data, whatever its source or format, whether structured or unstructured, typically built using Hadoop. However data lakes could be built using hybrid platforms to leverage benefits of both traditional and next gen platforms. Some of these hybrid combinations include:

- Any data lake OEM platforms on Cloud + On premises
- Hadoop + Traditional Database/Data warehouse platforms
- Hadoop + Cloud Storage (AWS S3, Azure Blob Storage)

A variety of storage and processing tools can be used to extract value quickly and speed up key organizational decisions. This data architecture is becoming popular due to the growing variety and volume of data, and as enterprises are leveraging mobile, cloud-based applications, and the Internet of Things (IoT) needing big data to enhance their service levels.

The benefits of a data lake include:

- Little or no processing of data to conform to any schema, flexibility, and enabling analysis of stream of data in a timely manner.
- Deriving value from unlimited types of data.
- Storage of any type of structured and unstructured data in their raw form, be it CRM data or social media posts
- Unlimited methods of querying the data Doing away with data silos
- Democratizing access to data through a
- single, unified view of data across the organization

Key Attributes of a Data Lake

Data lakes are becoming essential for enterprise data strategies to handle the 4 V's of data namely volume, velocity, veracity and variety in order to meet the customer expectations, and the rapid globalization of economies. The following three characteristics make data lake desirable:

- Data is stored as a single shared repository within a Distributed File System (DFS). It enables tracking changes to data throughout the data lifecycle, a function useful for compliance and internal auditing purposes.
- Orchestration and job scheduling capabilities enable resource management while providing a central platform for delivering consistent operations, security and data governance tools across the data lake clusters. As a result, analytic workflows can access the data and the computing power as required.
- The in-built applications or workflows consume, process or act upon the data that is preserved as is while tracking any changes. This enables sharing of data across stakeholders, and each to work on it as per their need.

Leveraging for Better Insights

Online banks have been able to score one over traditional banks in being able to provide features that improve customer engagement online.

But to continue to remain relevant and one step ahead of competition, the role of Big Data analytics is becoming clear. The sheer volume of data, its varied formats and the velocity at which it is flowing in has become one of its greatest challenges. Just to get a perspective, the digital data is expected to touch 44 zettabytes (44 trillion gigabytes) by 2020.

Ignoring this volume is unwise. But to sort through it and find out what's useful and what isn't is no mean task. With increase in potentially useful data as lines blur between industry segments and as customer expectations increase, the data that needs to be analyzed will also grow exponentially.

Another area where data storage and management is becoming critical is in compliance and keeping pace with evolving regulations.

They definitely need a unified data platform like the data lake where all data can converge into a data fabric where they can store the variety of data, scale up as required, manage, apply and analyze data quickly and reliably.

Building a data lake – An Indium Software perspective

There are multiple ways to architect, design and build a data lake. However a meticulous and agile approach is required to build a data lake in quick time, with optimum data quality and lower cost.

The following steps largely constitute the process to implement a data lake for a digital bank:

Build a business architecture that describes the follow of information from end customer until the bank's digital platform across varies layers like functional, security, application, data and infrastructure

Build a data model for a data lake that will support operational reporting, dashboards as well as advanced analytics

Develop a technical architecture of the data lake (or hybrid data platform) to establish a standard framework for data movement

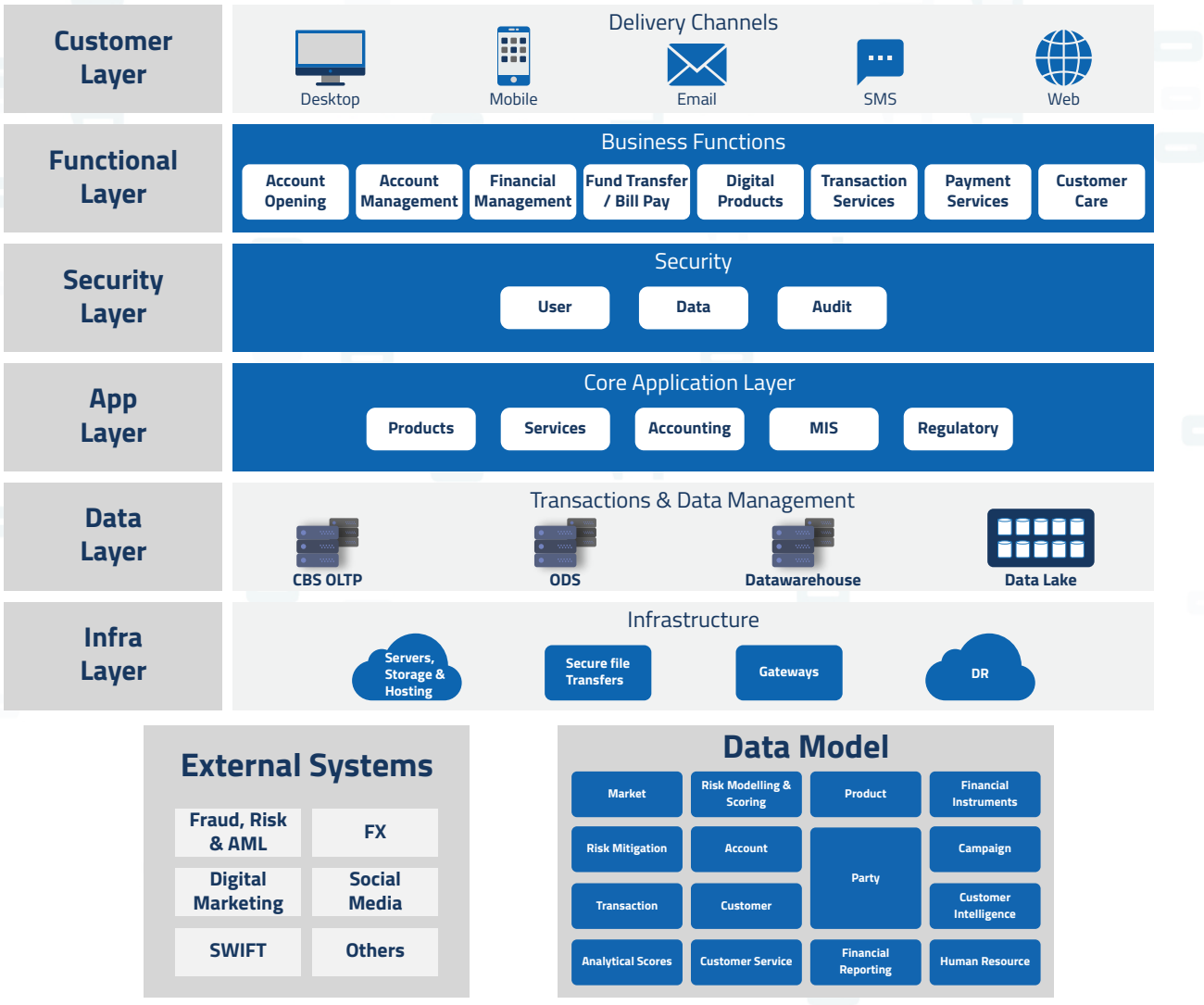
Develop frameworks for data governance, security, integration, master & metadata management

Build and Implement the components like data pipelines, databases, data lake components, reports, analytical models and dashboards

Operationalize the analytical models into the business functions like marketing, customer acquisition, loyalty, etc.,

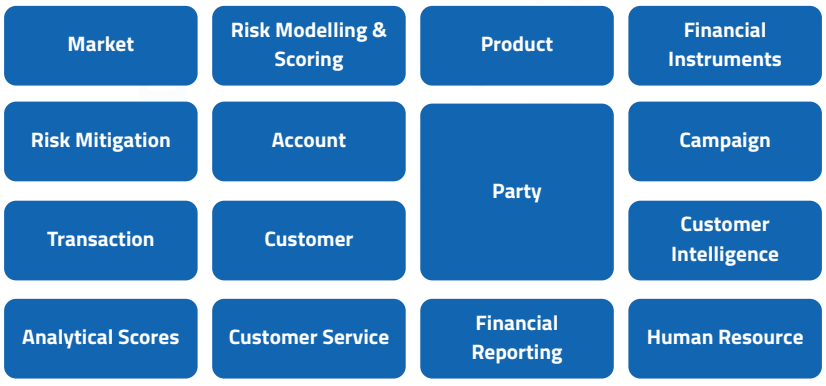
Standard Business Architecture of a digital bank

The following diagram illustrates a standard business architecture of a digital bank. The elements in the diagram are representative and could be customized depending upon each bank:



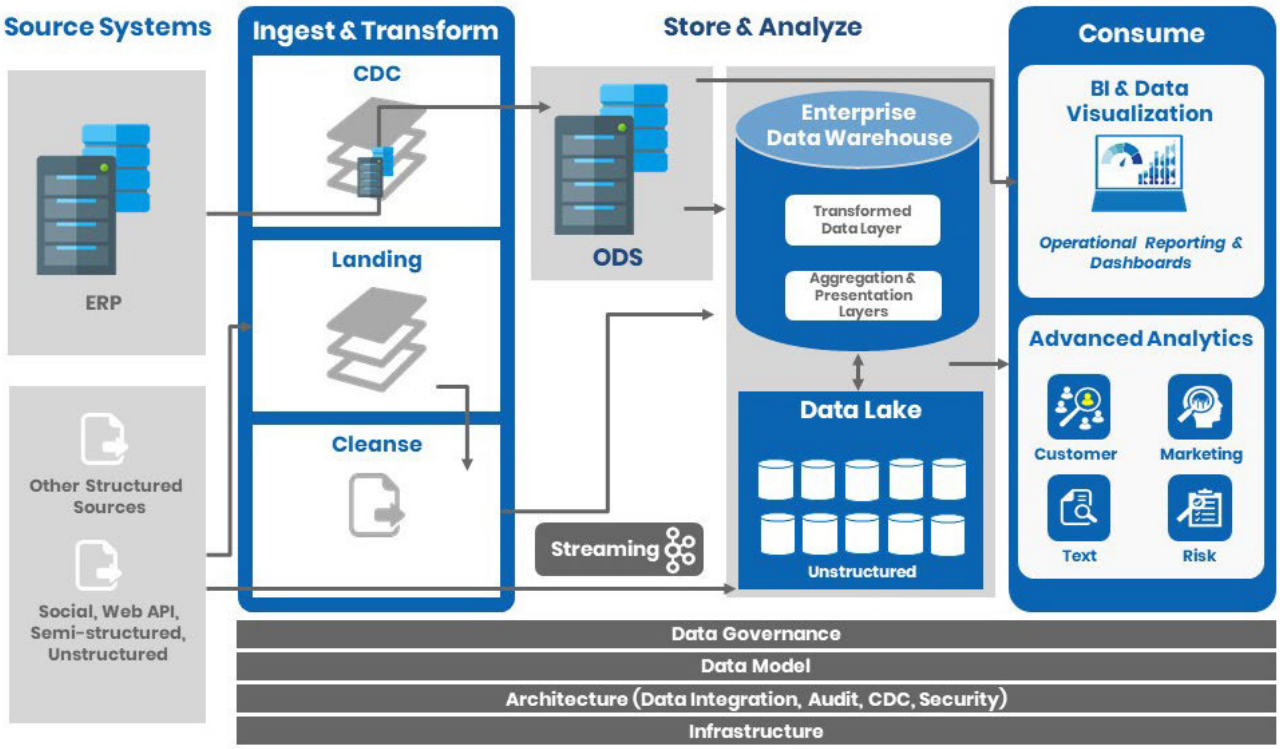
Conceptual data model of a data lake for a digital bank

The following diagram illustrates a conceptual data model of a data lake for a digital bank which can then be further customized based on the needs of the customer. Once the logical data model is finalized, Indium Software will then physicalize it and implement on the data lake platform.



Technical architecture of a data lake for a digital bank

The following diagram illustrates the reference technical architecture of a data lake for a digital bank which can then be further customized based on the needs of the customer:



Indium Software – For Designing Right

Data lakes are the answer to the neo banks' need for managing large volumes of data, for creating apt queries that improve decision making and for enhancing customer experiences. But data lakes are a mere platform. And they come with their own set of challenges.

For one, they can become data swamps if not designed properly, with an understanding of the business, its strategies or its goals. It needs a business-driven approach, and therefore, needs someone who understands the banking business as well as data lakes and Hadoop to be able to design it well. Appropriate engineering skills supported by background and experience are key advantages Indium Software, which has two decades of experience in quality assurance, banking and relevant experience in big data, offers.

This experience enables Indium Software to take into consideration the complexity of data lake solutions from a technical and engineering perspective, expose it to several technical capabilities ranging from self-service data ingest, to data preparation, data profiling, data classification, data governance, data lineage, metadata management, global search and security. This helps in making the data lake effective and efficient.

Fourth is the need to create visualization charts according to business requirements from ODS, Data Warehouse and Data Lake and integrate data visualization with the operational and management process. Lastly, it also requires facilitating changes across People, Process, Data and Technology to be effective and meet its goals. Indium Software offers these capabilities such that online banks can leverage the Big Data through data lakes to create efficient business strategies that assure growth and customer delight

INDIA

Chennai | Bengaluru | Mumbai
Toll-free: 1800-123-1191

USA

Cupertino | Princeton
Toll-free: +1 888 207 5969

UK

London

SINGAPORE

+65 9630 7959

Sales Inquiries
sales@indiumsoftware.com

General Inquiries
info@indiumsoftware.com