Financial Services Data & Analytics Newsletter

June 2024

pwc

Table of contents

Introduction

With every click and action performed in this digital world, an immense amount of data is being generated every day. This data, both structured and unstructured, can result in useful insights for businesses. However, in order to gather and analyse insights from it, businesses need a wide array of tools ranging from ingestion, processing, visualisation, analytics and governance. This need, combined with an ever-changing landscape of technologies, leads to the rise of several challenges and complexities – e.g. seamless integration of multiple tech stacks offering best-in-class services, managing and maintaining custom-built frameworks, combining those in one ecosystem, ensuring governance and enabling efficient use and collaboration, and sustained cost-efficient architecture.

To combat these challenges, Microsoft introduced Fabric -a cloud-based software as a service (SaaS) tool that helps businesses to manage all data and analytics tools to turn data into actionable workloads. Microsoft Fabric is a single unified platform for all stakeholders that assimilates all capabilities and provides seamless integration of services with a user-friendly interface and collaboration features to provide various data and analytical solutions.

In this newsletter, we will be discussing Microsoft Fabric in detail, highlighting its features and business use cases. We'll also touch upon industry news from partnerships and alliances between industry stalwarts and new-age companies to offer an improved customer experience and optimise efficiencies of the financial service ecosystem.

Happy reading!

Demystifying Microsoft Fabric

01 Microsoft Fabric

- Microsoft Fabric is an end-to-end data analytics platform designed for unified solutions.

- It offers a seamlessly integrated and user-friendly platform and empowers data engineers, scientists, analysts and business personnel by helping them collaborate on a single platform.

- It operates on a SaaS model that brings simplicity and enables integration to business solutions.
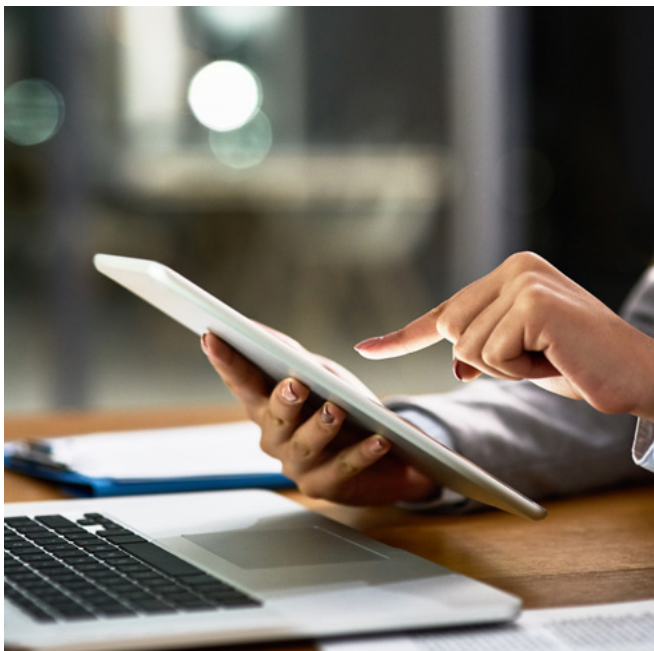
1.1. Key features of Microsoft Fabric and how it solves organisational challenges

End-to-end data and analytics capabilities: Bringing together a broad spectrum of data and analytics capabilities into a unified platform, Microsoft Fabric offers a comprehensive suite of analytical functionalities, covering everything from data movement and storage to data science and business intelligence. Although it is not an entirely new product from Microsoft, it amalgamates the finest aspects of data fabric, data mesh and data hub into one cohesive solution.

OneLake: Like OneDrive, OneLake is included with every Microsoft Fabric tenant, serving as the repository for all your analytics data. Functioning as a unified data lake, OneLake accommodates data of any format or origin, facilitating users' access and analysis from a centralised location. This feature enables organisations to maintain a single data lake encompassing all organisational data, streamlining data management with a singular copy for further utilisation. OneLake establishes a robust foundation centred around data lakes, addressing the challenges of today's fragmented data landscape with a unified storage system.

Real-time capabilities: With the growing complexities of data systems and the diverse source of data, organisations face challenges in integrating and analysing data in real time. To address this, Microsoft Fabric offers synapse real-time analytics, enabling organisations to scale their analytics solutions to meet evolving needs and generate real-time insights. This feature equips organisations with robust capabilities, including automatic data streaming, indexing, and the generation of queries and visualisations.

AI-powered solution: Azure's OpenAI service is embedded into every layer of Microsoft Fabric, empowering users to fully harness their data's potential. This integration allows developers to leverage generative AI on customer data, helping business users in uncovering valuable insights. A standout feature of Fabric is its ability to harness AI for creating data pipelines and flows, generating code and building models – all facilitated by the integration of GPT-powered copilot.

Centralised data management capabilities: With OneLake and a lakehouse-centric architecture, Microsoft Fabric addresses the fragmented landscape of data tools and platforms with its SaaS multi-cloud lakehouse solution. By storing data in the Delta Lake format, it ensures compatibility with any tool capable of reading this format.

1.1. Key features of Microsoft Fabric and how it solves organisational challenges

On-premises data gateway: On-premises data gateway' is a software designed to securely bridge the gap between the client's on-premises environment and the cloud. It helps organisations keep databases on their on-premises networks while integrating them with Microsoft Fabric (cloud) securely. An on-premises data gateway can be installed within a local network environment.

Scalability: As data volumes grow, traditional solutions may struggle to scale efficiently to meet increasing demands without incurring high costs or performance degradation. Fabric, with its **cloud native SaaS architecture** and **elastic scalability**, can handle massive volumes of data and accommodate fluctuating workloads efficiently. This scalability ensures high performance without compromising cost-effectiveness.
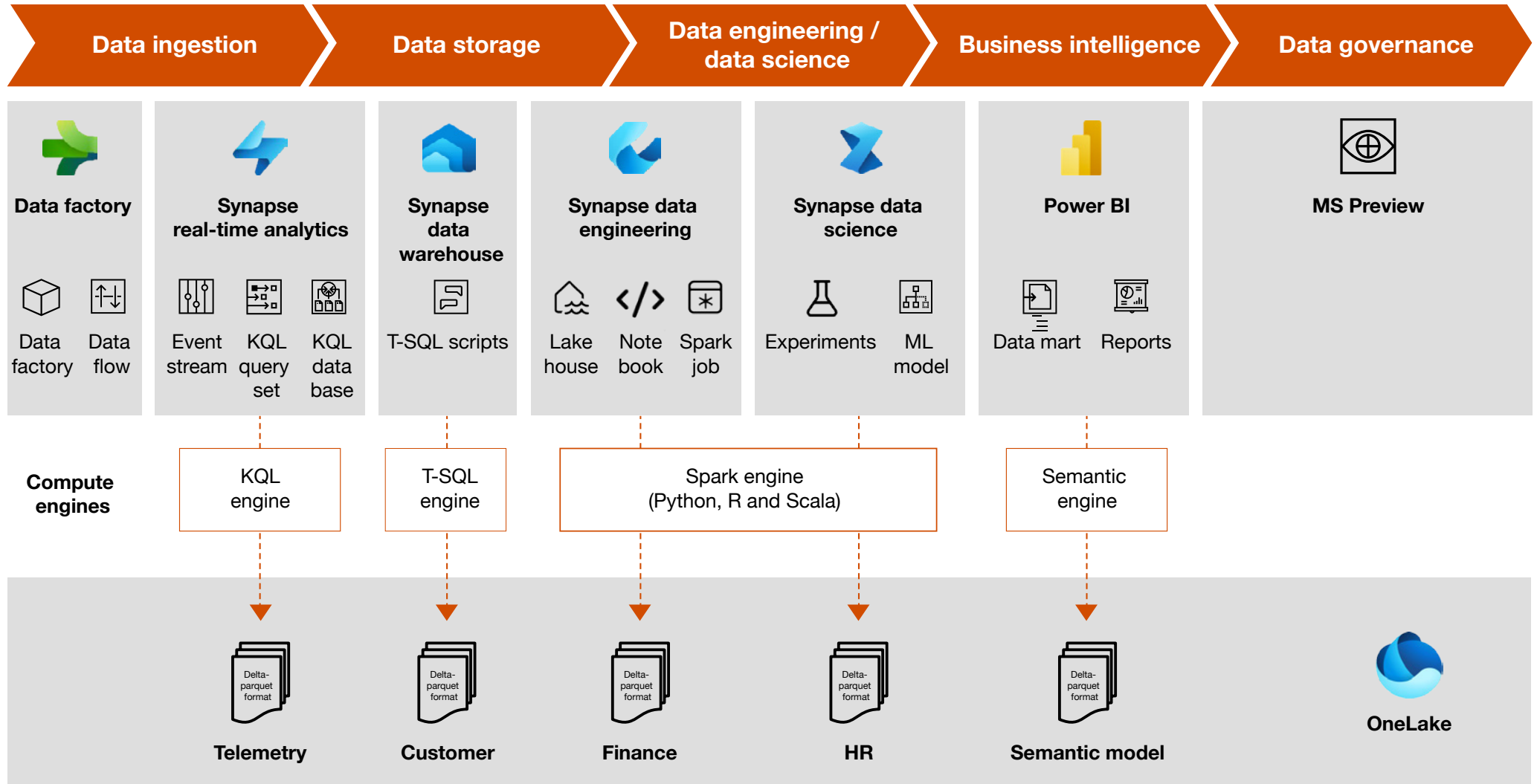
Data security and compliance: Protecting sensitive data and ensuring compliance with regulations (such as Digital Personal Data Protection Act 2023) is a significant challenge for organisations. Fabric's one access control on lakehouse architecture enhances collaboration and streamlines data management while providing uniform security across the organisation by integrating with Azure's comprehensive security services, including Azure Entra ID, Azure Key Vault and Azure Security.

Cost management: Managing the costs of multiple components with data storage, data processing etc., can be challenging – especially with unpredictable workloads. Microsoft Fabric's pricing is based on the total computing and storage utilised and not dependent on the individual component costs and utilisation. This universal computation of Microsoft Fabric allows organisations to save costs and eliminates complexity for managing separate charges for different services.

1.2. Microsoft Fabric architecture

Data ingestion	Data storage	Data engineering / data science	Business intelligence	Data governance

Data factory

Data factory | Data flow

Synapse real-time analytics

Event stream | KQL query set | KQL data base

Synapse data warehouse

T-SQL scripts

Synapse data engineering

Lake house | Note book | Spark job

Synapse data science

Experiments | ML model

Power BI

Data mart | Reports

MS Preview

Compute engines

KQL engine | T-SQL engine | Spark engine (Python, R and Scala) | Semantic engine

Delta-parquet format — **Telemetry**

Delta-parquet format — **Customer**

Delta-parquet format — **Finance**

Delta-parquet format — **HR**

Delta-parquet format — **Semantic model**

OneLake

1.3. Architecture components

Microsoft Fabric streamlines onboarding experiences, workspace interactions, security management, compliance and collaboration, while still offering tailored experiences for business intelligence analysts, data scientists and data warehousing practitioners.
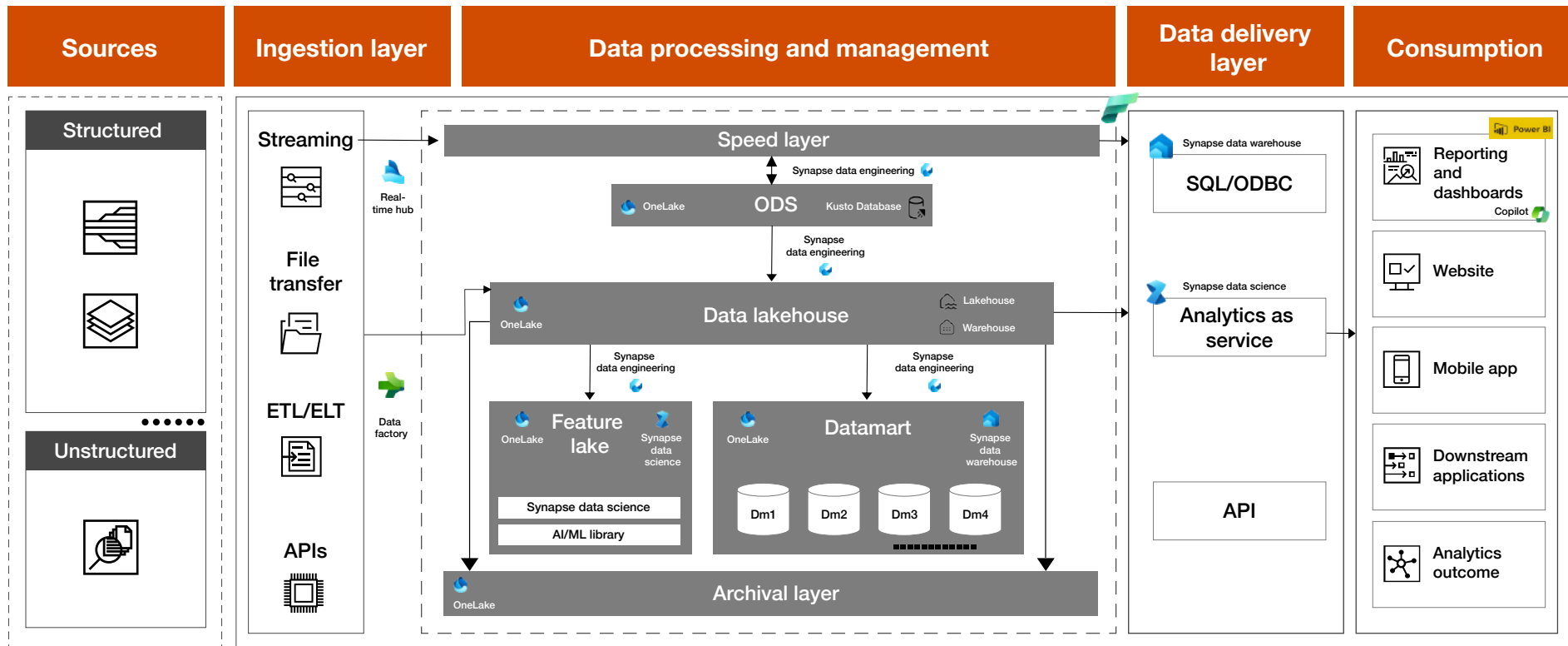
Below are primary Microsoft Fabric components:

Sr. no.	Component name	Capability	Key features

1	Data factory	Data integration: Ingest, prepare and transform data	• Capability to copy fast to move data quickly between locations
			• Two primary features:
			  • Dataflows
			  • Pipelines
			• More than 300 transformations in data flow designer
			• Facilitates building complex and flexible workflows with data pipelines

2	Real-time hub	Real-time data processing to extract data insights and actions	• Versatile solution for data from IoT systems, system logs, free text, semi-structured data • Facilitates easy access, addition, exploration and data sharing to enable broader insights and visual clarity • Event streams: Capture, transform, and route high volumes of real-time events to various destinations with a no-code experience • Processing capabilities like filtering, data cleansing, transformation, windowed aggregations and dupe detection available on event streams
3	Lakehouse	Unified data storage location	• Structured and unstructured data at a single location: OneLake architecture makes data storage, data sharing and exploration a simpler process by centralising the storage in the open data standard of Delta-Parquet format. • Data engineers can use the notebook to write code to read, transform and write directly to lakehouse as tables and/or folders. • Lakehouse explorer: Main lakehouse interaction page where data can be explored using the object explorer with a low-code experience • Existing data storage like Azure Data Lake Storage (ADLS), S3 can be accessed in OneLake using shortcuts without copying the data to OneLake.
4	Kusto database	Storage and data management	• KQL queries for structured, semi-structured and unstructured data to streamline data analysis. • Empowers organisations to perform time-series analysis on large volumes of data

Sr. no.	Component name	Capability	Key features
5	Synapse Data Warehouse	Data storage	• Transactional data warehouse that supports open data format • Data can be cross-queried without making copies of the same. • Automatic scaling of resources • Compute nodes and storage independent of each other
6	Synapse Data Science	Machine learning capabilities	• Easy data ingestion: Data pipelines can access and transform the data into a machine learning consumable format. • Powerful open-source visualisation libraries to enhance the data exploration experience • Data wrangler tool: Generates transformations and automation through generated codex • MLFlow experience: tracks the ML model training, log experiments and models
7	Synapse Data Engineering	Infrastructure capabilities	• Spark platform that enables creation, management and optimisation of infrastructures for collecting, storing, processing and analysing vast volumes of data • Provision to copy–paste data in lakehouse by designing flexible pipelines • Notebooks allow sharing of documents containing live codes, equations and visualisation • With a Fabric notebook, the user can get started with zero set-up effort – i.e. there's no need to spend time configuring environments or installing dependencies, just open the notebook and start coding. • Data Activator is a no-code experience in Microsoft Fabric for automatically taking action when patterns or conditions are detected in changing data.
8	Power BI	Business intelligence	• Data visualisation tool that enables connection with various sources and identification of important metrics in data • Paginated report workflow – design paginated reports to be printed or shared • Power BI semantic model – includes hierarchies, descriptions and relationships to provide deeper domain understanding • Power BI in Fabric has a feature called Direct Lake Mode, which allows it to access underlying data directly from the data lake, providing live data access.
9	Copilot	Data engineering and data science	• Copilot provides capabilities of intelligent code completion, task automation, building data pipelines and developing complex analytical models.Data can be cross-queried without making copies of the same. • It is available for data factory, Power BI, data engineering and data science experiences to provide a range of facilities on report building insight generation and code automation.

02 Microsoft Fabric-based reference architecture

Sources	Ingestion layer	Data processing and management	Data delivery layer	Consumption

Sources

Structured

Unstructured

Ingestion layer

Streaming

Real-time hub

File transfer

Data factory

ETL/ELT

APIs

Data processing and management

Speed layer

Synapse data engineering

OneLake ODS Kusto Database

Synapse data engineering

Data lakehouse

OneLake

Lakehouse

Warehouse

Synapse data engineering

Synapse data engineering

Feature lake

OneLake Synapse data science

Synapse data science

AI/ML library

Datamart

OneLake Synapse data warehouse

Dm1 Dm2 Dm3 Dm4

Archival layer

OneLake

Data delivery layer

Synapse data warehouse

SQL/ODBC

Synapse data science

Analytics as service

API

Consumption

Power BI

Reporting and dashboards

Copilot

Website

Mobile app

Downstream applications

Analytics outcome

Release management and devOps	Copilot and other generative AI features	Scheduling and orchestration
Data governance and data lineage	Logging and monitoring	High availability and recoverability

- The data factory component efficiently handles the ingestion of batch data, orchestrating its entry into OneLake, which serves as a uniform data repository as data lakehouse.

- Data lakehouse in OneLake can be accessed easily through lakehouse endpoint and also by using SQL endpoint, which gives end users easy access choices.

- Real-time hub serves as the conduit for ingesting streaming data, facilitating its seamless integration into downstream processes. Real-time data, dynamically generated and continuously flowing, can be harnessed directly for reporting purposes, and seamlessly stored within the KQL database. This can be referred for real-time dashboard and analytics through Power BI.

- Data consumption materialises through the utilisation of data science notebooks in unified Synapse Data Science module for analytical use cases. Realm of unified data science unfurls its potential through experiments, where algorithms and models are meticulously crafted to extract insights.

- Power BI runs on data lakehouse platform by adopting open Delta Lake and open-source data format parquet which helps to avoid vendor lock-in and reduces data duplication and management. Direct lake mode unlocks incredible performance directly against OneLake, with no data movement. Combining this with the ability for the other analytical engines to read and write data directly in the lake, Fabric will reshape how business users consume high volume data. Power BI datasets in direct lake mode improve query performance that is at par with the import mode, with the real-time nature of DirectQuery. As the data never leaves the data lake, there is no need to manage dataset refresh in Power BI in direct lake mode.

- Power BI provides AI capabilities through copilot, which enables the creation of visuals directly by providing prompts.

- Microsoft Fabric environment works under unified governance.

03 Migration use case: Azure Synapse Analytics and Power BI to Fabric

3.1 Typical migration lifecycle constitutes the following steps:

Step 1	Step 2	Step 3	Step 4	Step 5
Evaluate	Design	Migrate	Govern	Optimise
• Identify migration objective and motivation. • Identify key stakeholders and establish clear business outcomes. • List migration scenarios. • Estimate cost savings. • Assess technical fit for each workload.	• Map current state and future state architecture. • Identify workloads and stock keeping units (SKUs). • Evaluate security, governance, privacy and networking considerations	• Conduct a small proof of concept (POC). • Get help and support during migration and post-migration activities. • Automate wherever possible, leveraging various migration tools and accelerators.	• Monitor and manage cost, security and performance. • Perform governance benchmark assessment.	• Identify opportunities to optimise cost, performance, security and scalability, and operational excellence.

3.2 Key factors to consider while performing migration

3.2.1 Evaluate – data strategy review

- Ensure long-term scalability and cost-effectiveness of data warehouse/lake.

- Take advantage of the new and enhanced functionality of Power BI, Azure and Fabric.

- Revise data-modelling strategy to simplify model development, improve reliability and support complex security requirements.

- Retire components that can be consolidated into other services.

- Improve collaborative development experience and build robust DevOps practice.

- Provide greater administrative control and monitoring.

- De-risk change management process, particularly with respect to shared datasets.

- Identify comparable or equivalent Fabric components and align them with your current infrastructure.

- Evaluate how Fabric can help in cost reduction, optimised resource utilisation, and improved cost management capabilities, when compared to existing data platform architecture.

3.2.2 Design approach

- Ingestion methods dataflows, pipelines, or a hybrid approach

- Location for data to be OneLake – how data will be saved for each layer to be identified

- Naming conventions for all object types

- Transformation methods to be agreed with the client – PowerQuery, SQL, Python etc.

- Platform components and tiers required

3.2.3 Migration approach

- **Lift and shift**
  Lift and shift is a good approach while migrating existing environments with a small number of objects/workloads which do not carry traditional legacy debt artefacts. This approach will work well for workloads that are optimised with current ADLS in Delta-Parquet format, Azure Synapse and Power BI, so the same will not require major changes in Fabric. Built-in API support in Fabric makes it easier to integrate real-time ingestion and AI/ML outcomes with external applications.

- **Phase-wise approach with architectural changes**
  For cases where the legacy warehouse has evolved over a long period of time, while ingestion scripts can be reused or improved with new connectors, OneLake architecture must be completely renewed for new data engineering routines and emerging data models. The new data model can reuse the transformation logics, but it should cater to the features of data democratisation, governance and security aspects by design. This requires a well-thought-out phased approach of slowly decommissioning parts of the legacy warehouse instead of a hard cutover. An important aspect is to prioritise critical use cases/business outcomes that can be implemented on Fabric to derive quick wins and gain stability and trust on the new Fabric architecture. It may so happen that few features on the old set of tools/technology stack may not be directly available in Fabric, but the latter's robust capabilities are designed to accommodate any modern data platform features.

04 **Microsoft Fabric pricing**

Microsoft Fabric's pricing depends on licencing, OneLake storage and mirroring.

a. Licensing

The primary type of Microsoft Fabric is based on two models:

1. **capacity-based licences and**

2. **per-user licences.**

1. Capacity (pay-as-you-go or reservation)

A capacity-based licence provides the infrastructure for Microsoft Fabric. This licence allows us to:

- Use all the Microsoft Fabric features licensed by capacity.

- Create Microsoft Fabric items and connect to other Microsoft Fabric items.

- Save items to a workspace and share them externally for re-use with an appropriate licence.

Capacity licences are split into SKUs. Each SKU provides a set of Fabric resources allowing an organisation to procure as many capacity licences as needed.

Capacity licences offer dependable and consistent performance for organisational workloads. Each capacity licence offers a selection of varied SKUs that provides different resource tiers for storage, memory and computing power that can be procured depending on the type of solution to be deployed.

Capacity licences are further classified into either pay-as-you-go or reservation. A reservation licensing model does not cover storage or networking charges associated with Microsoft Fabric usage; it only covers Fabric compute capacity usage.

2. Per user licences

There are three types of user licences:

- **Free:** A free licence allows users to create and share Fabric content other than Power BI items in Microsoft Fabric, if users have access to Fabric capacity (either trial or paid).

- **Pro:** A Pro licence lets users share Power BI content with other users. Every organisation needs at least one user with a Pro or premium per user (PPU) licence if they intend to use Power BI within Fabric. SKUs smaller than F64 require a Power BI Pro or PPU licence for each user consuming Power BI content. SKUs greater than F64 (larger Fabric capacities) are available for unlimited users with a free licence if they have viewer role on the workspace.

- **Premium per-user (PPU):** PPU licences allow organisations to access Power BI premium features by licencing every user with a PPU licence instead of purchasing Power BI Premium. PPU can be more cost effective when Power BI Premium features are needed for less than 250 users. PPU uses shared capacity across the organisation, which provides the computing power for Power BI operations. PPU licences provide partial access to Microsoft Fabric. If you're using a PPU licence, the only items that you can access in Fabric are the Power BI items.

b. OneLake storage

- OneLake storage: Uniform storage of data which is billed in per GB.

- OneLake cache: The storage is billed for KQL cache storage and data activator data is retained.

- OneLake BCDR: The storage is billed as Business Continuity and Disaster Recovery (BCDR) storage.

c. Mirroring

Mirroring provides a modern way of the erstwhile change-data-capture (CDC) by accessing and ingesting data continuously and seamlessly from databases or data warehouses in Fabric by replicating a snapshot of the database to OneLake and continually keeping the replica in sync in near-real time. User can have free mirroring storage for replicas up to a certain limit based on the purchased compute capacity SKU provisioned by SKU – for example, if a user purchased F64, the user would get 64 free TB worth of storage. OneLake storage is billed only when the free mirroring storage limit is exceeded, or the provisioned compute capacity is paused.

05 When should an organisation move to Microsoft Fabric?

Microsoft Fabric can transform the data analytics landscape by offering a holistic, integrated approach to data management and analysis. Its importance stems from its capacity to offer a unified, collaborative platform that caters to all aspects of the data analytics process. Adopting Microsoft Fabric will help organisations in embracing a platform that offers cutting-edge data analytics technology. This comprehensive tool, along with seamless integration within the Microsoft ecosystem, makes it a potent solution for any organisation's data analytics challenges.

01 Consumer attitudes towards AI in banking: Insights from the Banking Disruption Index

According to the Banking Disruption Index by GFT, 44% of US consumers are receptive to using AI in banking, especially for fraud detection and savings advice, provided they have transparency in its application. The report emphasises that younger generations are more accepting of AI, while older individuals are more skeptical. Traditional banks are advised to focus on high-value AI applications like real-time fraud monitoring and automated financial advice to maintain consumer trust and compete with digital-first institutions.

02 Accenture study highlights AI's transformative potential in banking

Accenture's study suggests that AI could increase bank productivity by up to 30% and operating income by 20%. It evaluated the impact of generative AI on 2.7 million banking employees across 170 roles, finding that AI can streamline tasks like credit analysis and customer service, potentially reducing costs by 1 to 2%. Manoj Singodia of Accenture India emphasised the necessity of adopting a holistic approach to integrate AI into banking value chains. has garnered accolades such as being named 'Bank of the Year' in the BT-KPMG Best Banks and FinTechs survey for three consecutive years.

03 RBI's digital mission advances with new data centre to embrace AI and ML technologies

The Reserve Bank of India (RBI) is advancing its digital mission, with a new Greenfield Data Centre set to be completed by year's end, which will primarily serve as the financial centre for research and capacity building. PayNearby emphasised the importance of sovereign control over this critical infrastructure to mitigate accessibility risks. The RBI is also promoting the use of AI and ML in supervision, risk management and other key areas while cautioning against the potential risks associated with their overuse and the broader economic impacts of climate change.

04 Major financial institutions advocate for blockchain interoperability in new industry report

Deutsche Bank, Citi, Mastercard, Northern Trust, and Centrifuge, in collaboration with Axelar Foundation and Metrika, have released a report emphasising the importance of interoperable blockchain networks in the financial sector. The report addresses blockchain applications' liquidity fragmentation, security, scalability and transparency. These findings will be further discussed at the upcoming Point Zero Forum in Zürich, Switzerland, highlighting the need for multi-chain asset interoperability to cater to diverse blockchain adoptions by clients.

05 RBI to re-evaluate liquidity coverage ratio (LCR) amid rising digital transaction risks

RBI plans to review its liquidity coverage ratio (LCR) framework to enhance the management of liquidity risks in banks, in light of recent banking crises abroad. Prompted by rapid fund withdrawals experienced by banks like Silicon Valley and Signature Bank in the US, the RBI aims to adapt to the challenges posed by increasing digital transactions. The review might include limiting certain online transactions to business hours to prevent potential cash crises.

06 Punjab National Bank (PNB) enhances loan recovery through data analytics and employee inclusivity

PNB has used data analytics to improve loan recovery rates. Atul Kumar Goel, the bank's managing director said that the bank has taken a data-driven approach to customise recovery approaches. As per him, the bank has employed the visually impaired staff in call centres to handle overdue payments. These strategies have helped them to decrease non-performing assets in the farm loan sector from 14–15% to just 0.4%.

07 **Proper data governance is crucial before implementing AI in CRM systems finds Forrester Study**

Forrester's study warns banks against implementing AI in customer relationship management (CRM) systems without proper data strategy and strong data governance as this would lead to inaccurate outcomes and unintended filtration. The report suggests having proper data management and governance strategy to avoid future complications.

08 **RBI implements supervisory actions on banks to enhance digital banking safety**

The RBI has taken regulatory actions against certain financial institutions, including Kotak Mahindra Bank, to address the shortcomings in banks' IT systems. The RBI restricted these banks from onboarding new customers and issuing new credit cards due to frequent system outages. The regulatory steps aim to ensure banks can handle large transaction volumes and maintain quality of customer service.

Contact us

Mukesh Deshpande
Partner, Technology Consulting
mukesh.deshpande@pwc.com

Hetal Shah
Partner, Technology Consulting
hetal.d.shah@pwc.com

Acknowledgements
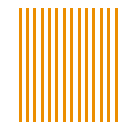
About PwC

At PwC, our purpose is to build trust in society and solve important problems. We're a network of firms in 151 countries with over 360,000 people who are committed to delivering quality in assurance, advisory and tax services. Find out more and tell us what matters to you by visiting us at www.pwc.com.

PwC refers to the PwC network and/or one or more of its member firms, each of which is a separate legal entity. Please see www.pwc.com/structure for further details.

pwc.in