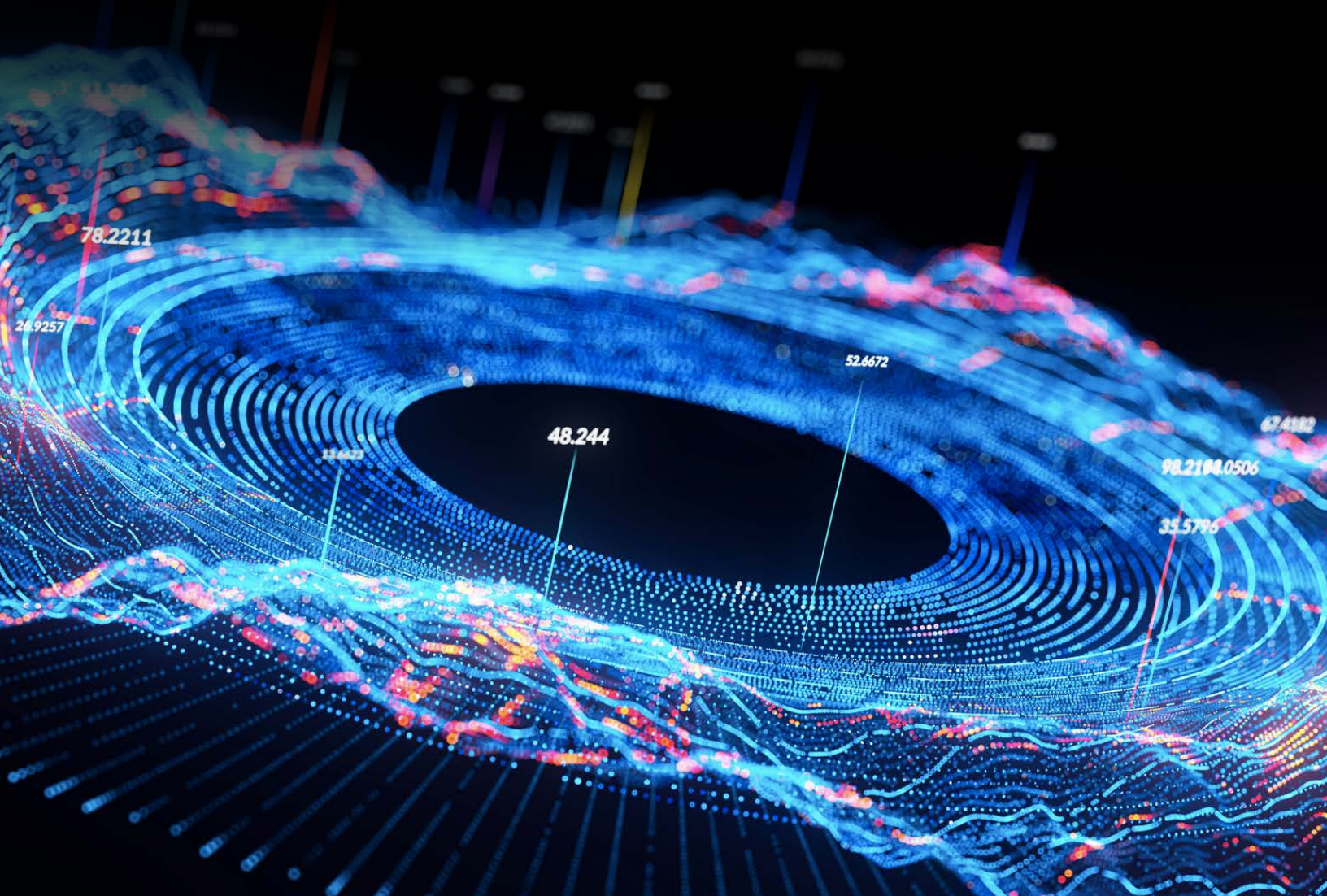


Whitepaper

# Power BI at Scale

## How Microsoft Fabric Handles Big Data Workloads

by Sanjay Biswas, Suman Gupta and Kallol Dutta



# Table of Contents

Glossary	2
Introduction	3
Key data analytics requirements and challenges	3
MS Fabric Power BI - Organization BI analytics solution	6
Copilot (GenAI) Assistance	18
Performance and benefits analysis	15
References	21
About Authors	22

## Glossary

Abbreviation	Full Form
BI	Business Intelligence
MS	Microsoft
Gen2	2 <sup>nd</sup> Generation
GenAI	Generative Artificial Intelligence

# Introduction

The explosion of enterprise data in recent years has pushed traditional business intelligence (BI) tools to their performance limits. As organizations grapple with hundreds of millions of records from diverse sources, delivering accurate, timely insights becomes increasingly complex.

This whitepaper explores how Microsoft Fabric and GenAI-powered Power BI enable scalable, user-friendly analytics on massive datasets. It outlines challenges faced with large-scale analytics and demonstrates how Fabric's unified architecture can overcome them.

This paper is intended for data architects and IT professionals who build enterprise data infrastructure. It is also aimed at data analysts and senior leaders responsible for deriving insights from complex datasets. Lastly, it targets professionals exploring Microsoft Fabric's potential for large-scale data analytics.

## Key data analytics requirements and challenges

Organizations face growing pressure to deliver fast, trusted insights as enterprise data volumes grow exponentially, from ERP systems, IoT sensors, SaaS tools, and third-party sources. However, managing and analyzing such data at scale presents significant challenges: data pipelines become fragile, refresh cycles slow, and legacy BI tools struggle to query multi-billion-row datasets.

## How can data analytics help business growth?

When powered by a unified architecture, data analytics can unlock critical business value. From real-time insight generation to cost optimization, the benefits of enterprise-scale analytics include:

- **AI-powered BI tools:** Provide strategic insights and operational improvements with clear visualizations and dashboards.
- **Customer insights:** Helps understand customer behavior and preferences for better satisfaction and loyalty.
- **Operational efficiency:** Identifies inefficiencies and areas for cost savings and process optimization.
- **Risk management:** Predicts and mitigates risks, ensuring better compliance and security.
- **Competitive advantage:** Offers a competitive edge by enabling quick adaptation to market changes & innovation.
- **Predictive analytics:** Forecasts future trends and outcomes to anticipate market demands and optimize inventory.
- **Personalization and targeting:** Allows precise targeting and personalization in marketing campaigns.

## Key data analysis capability requirements

### Aggregated data reports

Aggregated data reports use summarized tables with significantly lower volume than fact tables. This improves report performance in Power BI, as most calculations are pre-processed. It also reduces strain on the BI infrastructure.

### Non-aggregated data reports

These reports operate on detailed transaction or fact-level data, with some pre-applied calculations. While they improve accuracy, they typically require importing massive datasets into Power BI, posing performance and scalability challenges for IT teams.

### Self-service data models

These models include granular fact and dimension tables that allow users to navigate from summarized to detailed data. Power BI supports this through aggregated table switching, enabling historical data analysis without engineering intervention.

### Static data reports

Static reports use archived or snapshot-based datasets. These are imported into Power BI as one-time loads and can be reused for periodic analysis. This reduces load time and refresh costs but lacks real-time data access.

### Implications for BI Scalability

Among these reporting types, non-aggregated, self-service, and static reports tend to involve large data volumes and frequent access. Organizations often struggle to maintain performance while expanding BI capabilities without a scalable backend.

## Challenges in Scaling Business Intelligence for Large-Scale Data

As organizations grow increasingly data-rich, traditional BI systems struggle to manage the scale, speed, and cost of insights. Below are four key challenges that often impede efficient analytics delivery in high-volume data environments.

## Large Data Volume

Managing analytics on datasets larger than 100 million records often leads to storage, processing, and query performance bottlenecks. Ensuring data quality at scale adds further complexity and time burden.

## Data Access Cost

Running direct queries on large datasets increases load on the database, leading to higher computing costs and slower performance due to real-time resource consumption.

## Redundant Data Models

Import-mode models often duplicate data across multiple models, leading to inefficiencies for storage, processing, and model maintenance. This redundancy impacts load times and overall system responsiveness.

## BI Performance

BI performance depends on the database configuration, network/firewall throughput, and optimization of the source system. Poor setup can result in latency and delayed dashboard responses.

These limitations highlight the urgent need for a unified, high-performance analytics platform that reduces redundancy, manages cost, and scales effortlessly with growing data volumes.



# MS Fabric Power BI - Organization BI analytics solution

## Modern BI analytics logical architecture

The reference architecture below illustrates how Microsoft Fabric and Power BI work together to enable seamless data onboarding, high-speed analytics, and GenAI-powered insights for business and IT users.

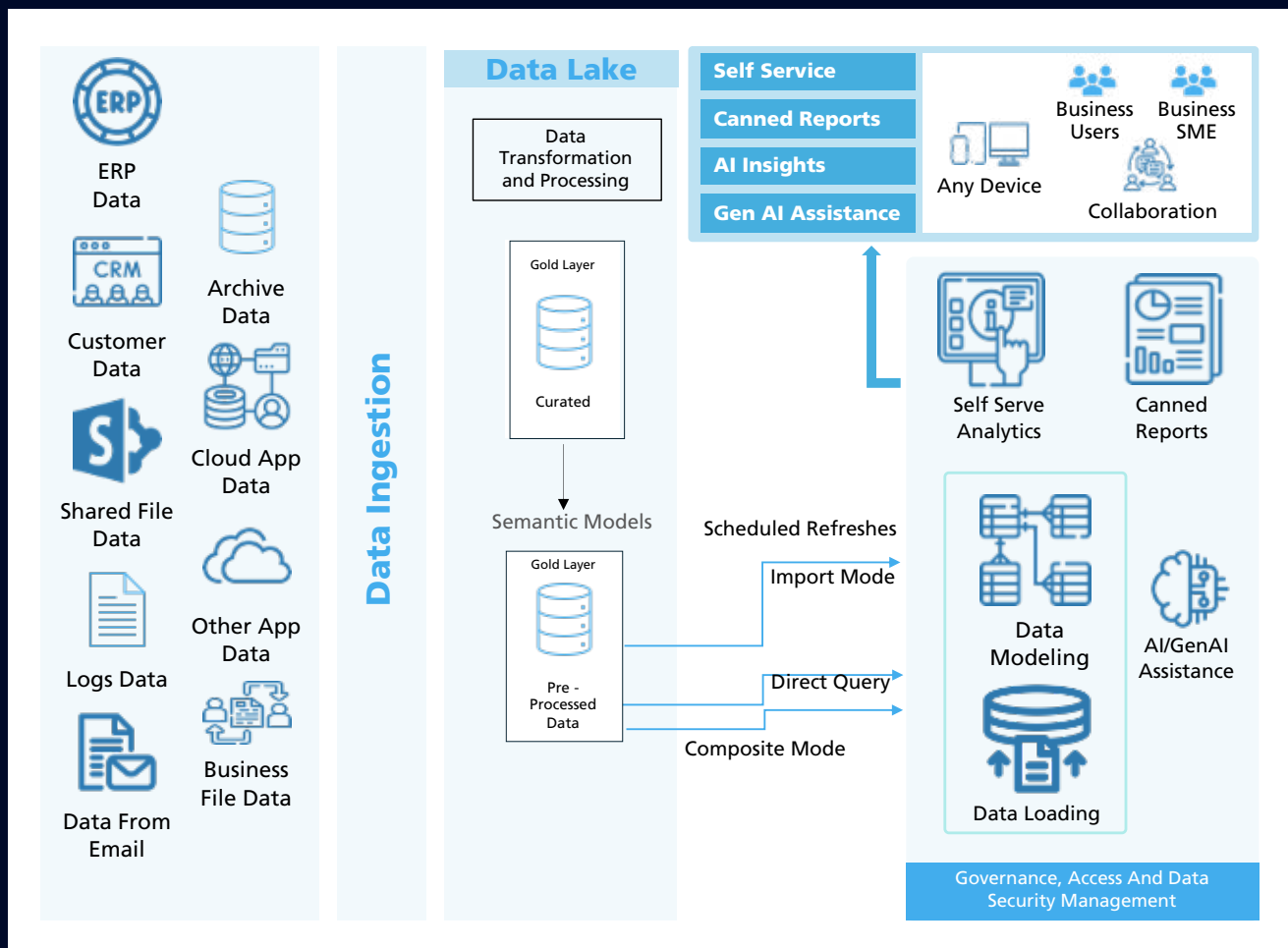


Figure 1: Logical BI application architecture

## Key expectations from the BI application architecture

- **User-centric design:** The interface should be intuitive, with dashboards and visualizations that enhance decision-making without steep learning curves.
- **Real-time access:** Systems must deliver insights from live data with minimal latency.
- **Scalability & high-volume data handling:** It should handle massive data loads and increasing user demands without performance degradation.
- **Unified data integration:** Fabric should seamlessly connect diverse sources, clouds, apps, and shared drives for end-to-end visibility.
- **AI & predictive intelligence:** Support for AI-generated insights, predictive models, and automated summarizations.
- **Extensive data volume analysis:** It must efficiently process and analyze massive datasets, enabling robust data handling and analysis.
- **Customizable and collaborative:** Enable tailored report creation, teamwork, and secure sharing across business functions.



## How to achieve the above architecture using MS Fabric

The diagram below illustrates how Microsoft Fabric's platform services integrate to implement the architecture discussed previously.

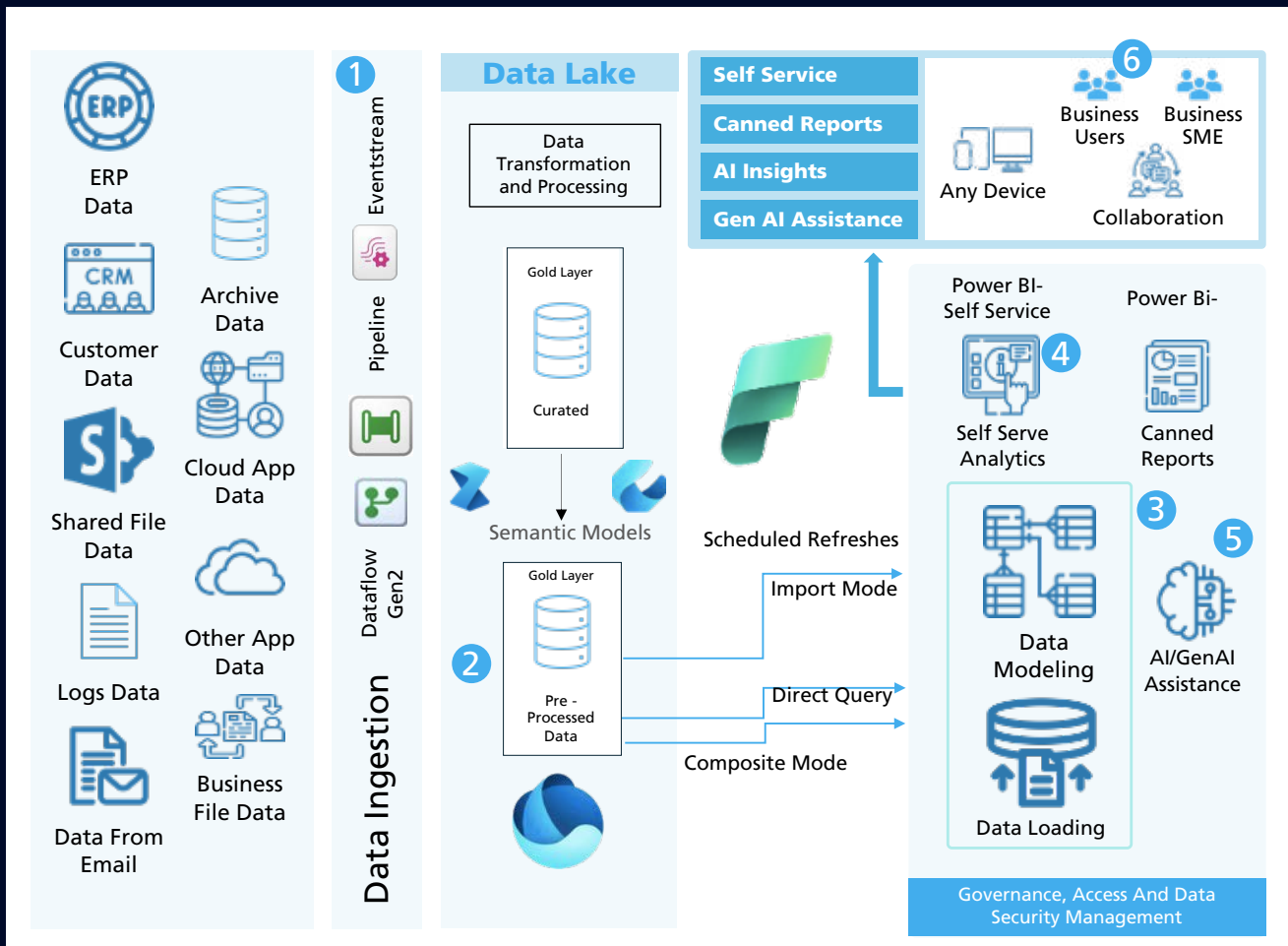


Figure 2: BI application architecture using MS Fabric

## Key highlights of the Microsoft Fabric-enabled BI architecture

The diagram below illustrates how Microsoft Fabric's platform services integrate to implement the architecture discussed previously.

- **Unified data ingestion:** Microsoft Fabric centralizes data from ERP, CRM, file-based, and streaming sources into a single platform, simplifying management and access.
- **Real-time insights:** Power BI supports live dashboards and Direct Query mode for instant data access, ensuring timely decisions.
- **Scalability & flexibility:** The architecture handles massive datasets efficiently using Import, Direct Query, and Composite models.
- **AI-driven analytics:** Integration of Copilot and GenAI accelerates insight generation, predictive modeling, and code-free exploration.
- **Custom dashboards & reports:** Power BI lets users design personalized views and reports aligned with business goals.
- **Governance & security:** Centralized data governance, RLS/OLS, and audit control ensure compliance, secure analytics at scale.
- **Seamless collaboration:** Features like shared workspaces, versioned dataflows, and co-authoring enhance teamwork.

Microsoft Fabric supports various ingestion mechanisms for batch, pipeline-based, and real-time streaming data. The following sections explain how these capabilities power modern BI architecture.

## What is Dataflow Gen2

Dataflow Gen2 is a low-code, Power Query-based transformation engine in Microsoft Fabric. It supports over 300 data and AI transformations. It enables users to build ingestion pipelines using a familiar interface seen in Excel and Power BI. Dataflow Gen2 integrates with services like Dynamics 365, Power Platform, and more.

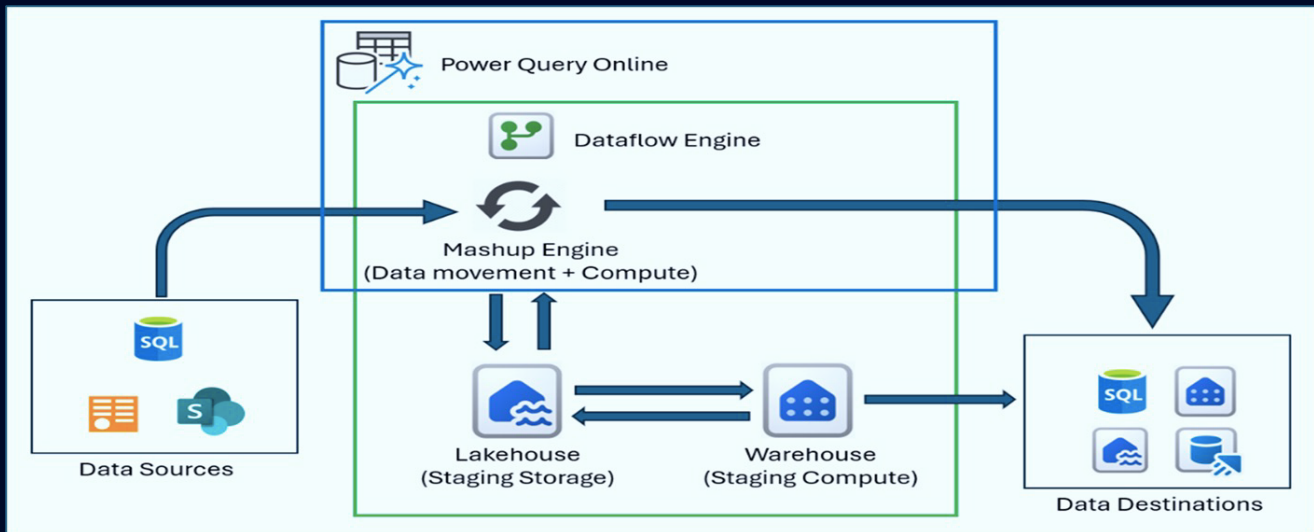


Figure 3: Dataflow Gen2 data flow diagram (<https://learn.microsoft.com/en-us/fabric/data-factory/pricing-dataflows-gen2>)

- **Interface:** Offers a Power Query-based low-code interface for data transformation and ingestion.
- **Large-scale data:** Supports ingestion of terabyte-scale datasets using tools like Copy Activity.
- **Data pipelines:** Easily integrates into pipeline orchestration for multi-step workflows.
- **Incremental refresh:** Only processes changed data, reducing latency and compute cost.
- **AI integration:** Enables Copilot and GenAI features for smart transformations and insight generation.

## What is a pipeline

Pipelines in Microsoft Fabric enable orchestrated, automated data workflows across diverse data sources. They offer flexibility for data transformation and support conditional logic, retries, and failure handling.

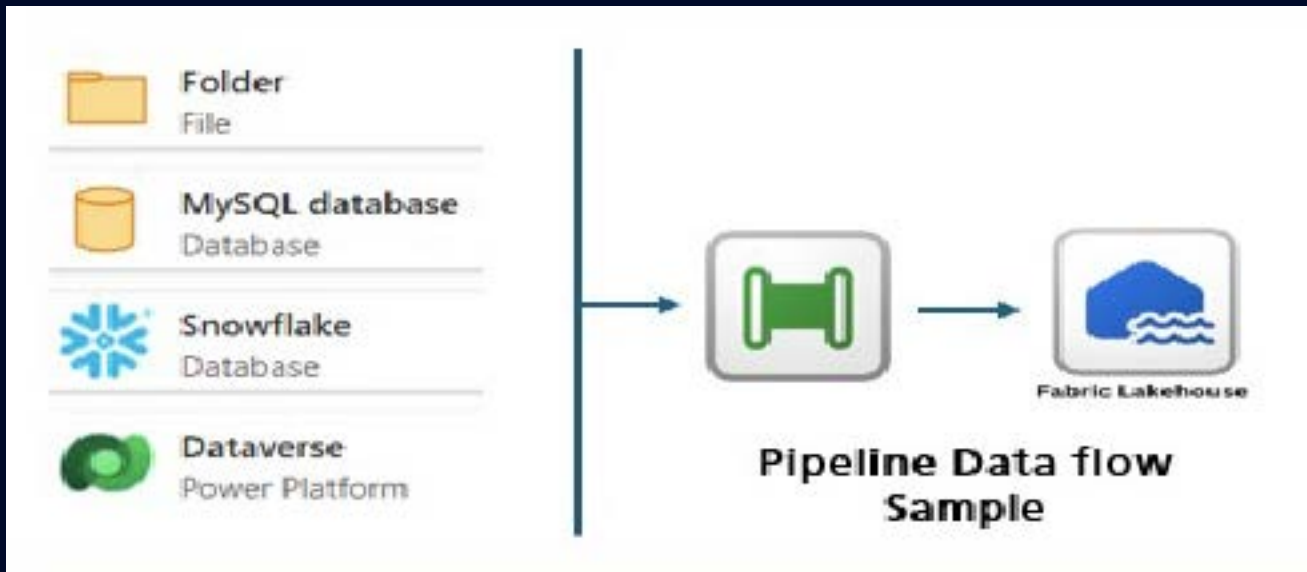


Figure 4: Data flow diagram using pipeline service

#### Data orchestration

Supports complex, multi-source ETL flows using drag-and-drop logic.

#### Automation and monitoring

Provides robust APIs and Fabric-native scheduling for reliable integration and observability.

## What is Fabric Lakehouse

Microsoft Fabric Documentation states, “Microsoft Fabric Lakehouse is a data architecture platform for storing, managing, and analyzing structured and unstructured data in a single location.

It's a flexible and scalable solution that allows organizations to handle large volumes of data using various tools and frameworks to process and analyze that data. It integrates with other data management and analytics tools to provide a comprehensive data engineering and analytics solution.

A Lakehouse combines a data lake's scalability with a data warehouse's performance and structure, providing a unified data storage, management, and analytics platform.

## Features

### Unified platform

- Combines lake-scale storage with warehouse-style querying and structure.
- Enables efficient management of both structured and unstructured data.

### SQL analytics endpoint

- Automatically generates SQL endpoints for direct querying via traditional SQL.
- Simplifies access and boosts analytical flexibility.

### Automatic table discovery

- Auto-registers files ingested into the Lakehouse as analyzable tables.
- Reduces manual intervention and speeds up data readiness.

### Seamless integration

- Connects natively with Power BI and other analytics tools.
- Streamlines workflows across ingestion, modeling, and reporting.

## Power BI semantic data model and data consumption

The diagram below outlines how Power BI models data from the Fabric Lakehouse and serves it through various consumption channels like reports, dashboards, and apps.

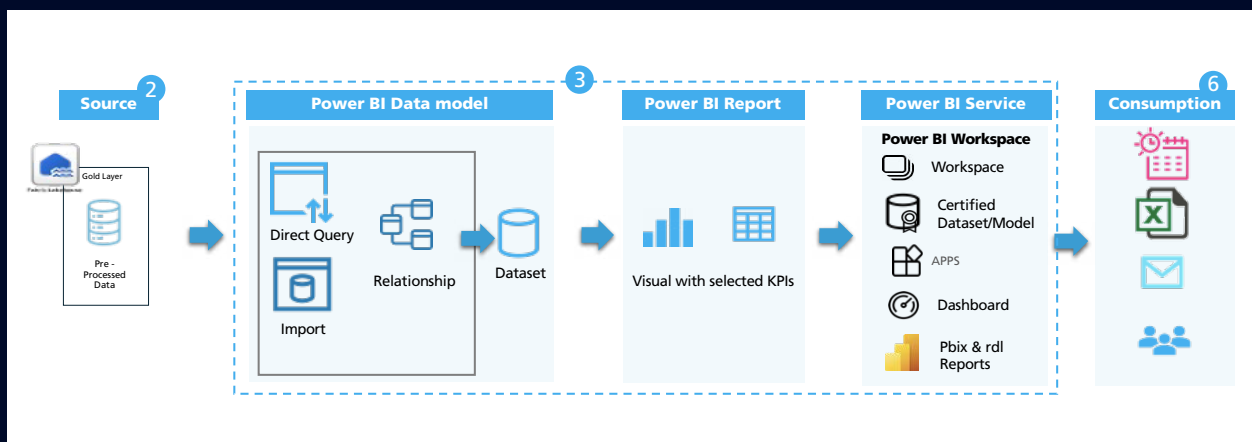


Figure 6: How Power BI data modelled and accessed

## Dedicated tables/views reporting tables and views

- A unified data platform with a dedicated reporting data source includes processed data tables/datasets and pre-join views. It supports data transformation, aggregation, and computations in views necessary for reporting.

## Build Power BI dataset

- Connect data tables via Import, DirectQuery, or Composite modes to build a Power BI data model. Establish relationships between tables, generate KPI metrics/measures using DAX, and use Power BI Datasets as the data layer for Power BI reports, including paginated reports.

## Build interactive dashboards

- User experience-led designs involve creating wireframe-based UX and UI prototypes, applying layout and color theory, and developing visuals on top of the Power BI dataset model. Interactive dashboards are enhanced with slicers, drill-down, and drill-through features to uncover actionable insights.

## Enterprise reporting

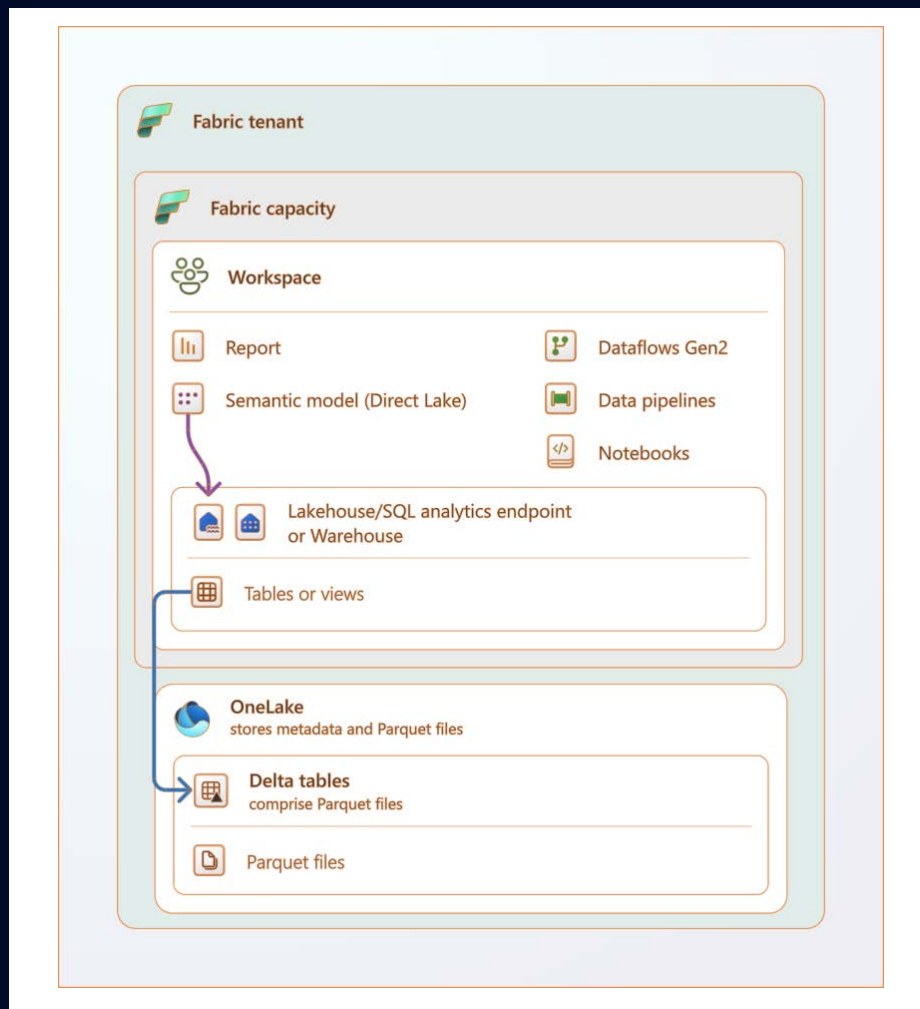
- Power BI Service is a cloud-based platform that facilitates collaboration, sharing, and enterprise distribution. It hosts Power BI reports, supports scheduled data refreshes and report publishing, and allows exports in Excel and PDF formats.

## Business users

- The business team accesses dashboards and reports by logging into the Power BI App and service, enabling self-service report creation for detailed data analysis.

## Power BI Semantic Data Model with Direct Lake

Direct Lake mode is a groundbreaking new engine capability for analyzing massive datasets in Power BI. The technology consumes parquet-formatted files directly from a data lake without having to query a Warehouse or SQL analytics endpoint or import or duplicate data into a Power BI semantic model.



**Figure 7:** Delta tables in Onelake lakehouse (<https://learn.microsoft.com/en-us/fabric/fundamentals/direct-lake-overview>)

## Why choose Direct Lake?

- **No data copy/ETL** - Connects live to your Delta tables in OneLake, avoiding costly data movement and staging.
- **Massive scale** - Handles hundreds of gigabytes to terabytes directly in the lake, without importing the whole dataset into memory.
- **Import-like speed** - Once columns are transcoded into VertiPaq, queries run at in-memory speeds.
- **Agile refresh** - Incremental framing and automatic updates deliver fresh data with minimal overhead.

## How does Direct Lake work with the Semantic Model

### Transcoding (On-Demand Column Loading)

- DAX/MDX query arrives → determine needed columns (including those for relationships/measures).
- If not resident, load complete column data from Parquet-based Delta tables into memory.
- Columns stay cached until they are evicted by framing, inactivity, or capacity pressure, Microsoft Learn.

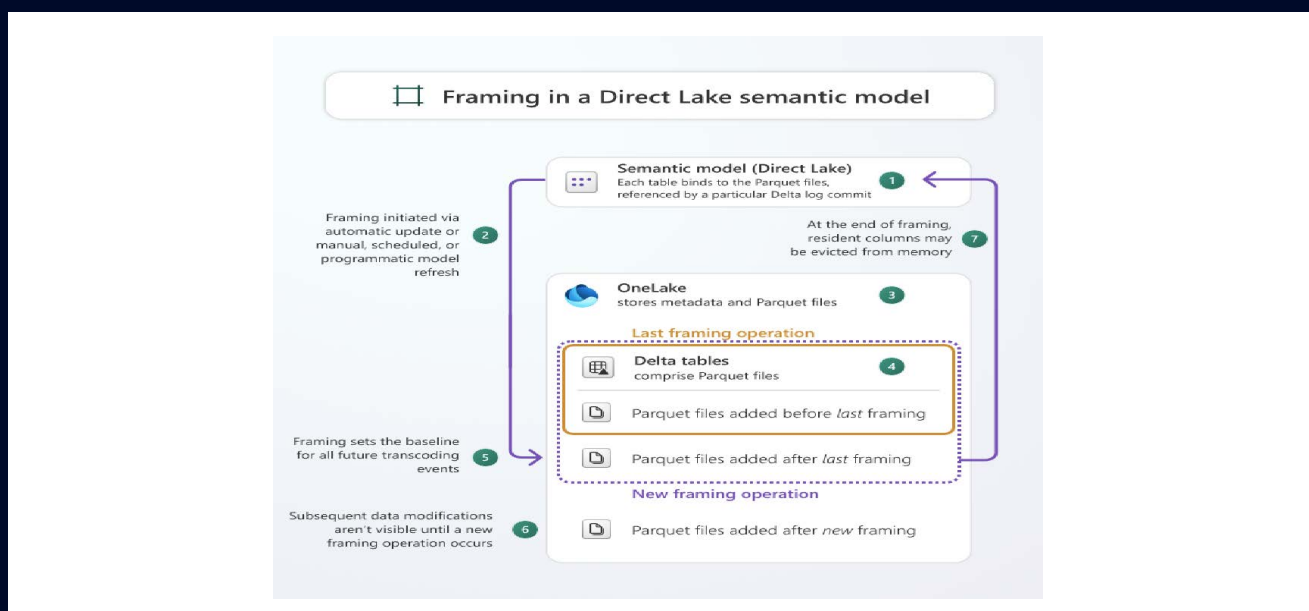
### Framing (Lightweight Refresh)

- Sets a point-in-time baseline by updating the model to reference the latest Parquet files.
- Evicts only changed column segments; preserves dictionaries and minimizes reload cost.
- New data isn't visible until the next framing cycle (manual, scheduled, automatic, or programmatic), Microsoft Learn.

### Direct Query Fallback

- Automatic fallback occurs when the model exceeds capacity guardrails or when querying SQL-endpoint views or tables enforcing RLS.
- Falls back to Direct Query against the SQL analytics endpoint (latest data but slower) – control behavior via the DirectLake behavior property.

## Direct Lake – Direct Benefits



**Figure 8:** How semantic model work with Onelake (<https://learn.microsoft.com/en-us/fabric/fundamentals/direct-lake-overview>)

## Power BI benefits from Direct Lake integration

- **Automatic Semantic Model provisioning** - Every Fabric Lakehouse auto-provisions an SQL analytics endpoint with a default Direct Lake semantic model ready for immediate reporting.
- **On-demand data access with Smart Caching** - Power BI visuals query the semantic model, which fetches only the needed columns from the Lakehouse and caches them in VertiPaq.
- **V-Order optimization for Delta Tables** - Lakehouse tables are physically ordered via V-Order, dramatically speeding up predicate pushdown and range scans.
- **Near real-time updates** - Combine Direct Lake's automatic updates with Power BI's Automatic Page Refresh for dashboards that reflect changes in the Lakehouse in seconds to minutes, not hours.
- **Simplified security & RLS without performance penalty** - Grant report viewers workspace, viewer access, or leverage a fixed identity for non-default models. Define row-level security in the semantic model itself; it avoids SQL-based RLS that would trigger a Direct Query fallback.
- **Reduced operational overhead** - No need for dedicated ETL pipelines to sculpt data for Power BI, model definition, and data access happen in one place.

## How self-service capabilities will add value

The gap between what “business wants” and what “business can do themselves” will be reduced with self-service implementations. Our framework-driven approach for self-service empowers businesses with “insights - made by self” @speed a possibility, without having to understand complex data models, etc.

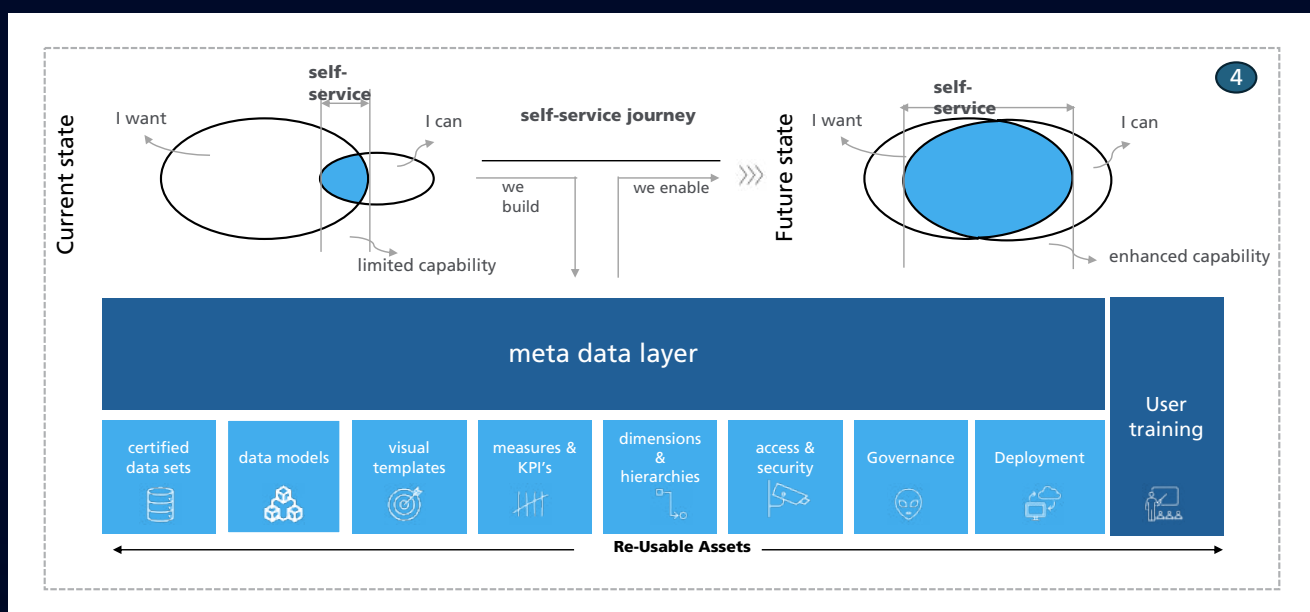


Figure 9: How Self-service analysis help users

## Key benefits and considerations of the Power BI reporting tool

### Key Advantages:

#### Canned Reports:

- Canned reports offer speed and efficiency by being prebuilt and quickly generated, saving time and effort, while also providing consistent formatting and data presentation, which is useful for regular reporting needs. Ease of Use: Users do not need advanced technical skills to generate these reports, making them accessible to a broader audience.
- Reduced Errors: Automated generation reduces the risk of human errors in data compilation and formatting.
- Compliance: They can be designed to meet specific regulatory or compliance requirements, ensuring that all necessary information is included.
- Can be shared with all the respective user base using security and privacy.
- Can save time and infra cost based on the optimized model and report design.

#### Self-Service BI Reports:

- Self-serve reporting tools empower users to design and customize their own reports, reducing IT dependency and enhancing operational efficiency by enabling quick decision-making. Data Democratization: It provides access to data for a broader range of users, enhancing data literacy and enabling more informed decision making across the organization.
- Faster Decision Making: With direct access to data, users can make quicker, data driven decisions.
- Resource Optimization: IT departments can focus on more complex tasks rather than routine report generation.

#### NLP and Generative AI:

- Efficiency: AI can process large volumes of data quickly, providing timely insights and reports.
- Accuracy: AI reduces human error, ensuring more accurate data analysis and reporting.
- Predictive Capabilities: AI can forecast trends and outcomes based on historical data, aiding in proactive decision making.
- Automation: Routine tasks can be automated, freeing up human resources for more strategic activities.
- 24/7 Availability: AI powered assistants can provide support and generate reports at any time, increasing accessibility.

### Points to Remember:

- Nontechnical users might misinterpret data while creating self-service reports, leading to incorrect conclusions. Proper training and accurate data models are key for success.
- Increased access to data can lead to potential data breaches if not properly governed.
- Copilot requires high computation power so concurrent usages might cause higher infrastructure usages and can make the application to run slower.
- AI systems can perpetuate existing biases in data, leading to unfair or skewed results, and they rely on the quality of input data. Poor data can lead to inaccurate outputs.

## Copilot (GenAI) Assistance

Copilot introduces new extensions to Power BI, leveraging Codex to support data analysts and developers. These capabilities include writing DAX formulas, modifying data models, adding visuals, and generating complete reports from natural language prompts. This AI-powered integration boosts productivity and streamlines Power BI data modeling and report generation.

### Conversational data modeling

Copilot understands natural language prompts, enabling easy data model building and modification. Relationships can be created, measures defined, and calculated columns generated effortlessly, reducing manual effort by approximately 10%.

### Intelligent suggestions & auto-complete

Copilot's AI engine is trained on vast amounts of data and insights. Based on the data context, Copilot provides intelligent suggestions and autocomplete options, making the modeling process faster and more efficient. This results in a 30% increase in data insights.

### Automated data profiling

Copilot leverages its deep understanding of data patterns to provide automated data profiling. It helps identify data quality issues, detect outliers, and suggest potential transformations or cleansing steps.

### Data exploration & visualization

Copilot will generate appropriate visuals based on user queries, providing a quicker and more interactive way to explore the data. This will be a foundation for building the required complex report designs, saving approximately 10% of manual effort.

## Key benefits of using AI and Copilot in Power BI

### AI & automation

- Delivers advanced analytics and generate actionable insights.
- Identifies data patterns and presents them in an easy-to-understand format.
- Enables faster, data-driven decisions with deeper visual insights.

### Self-service & exploration

- Adds dynamic values to smart narrative using natural language processing.
- Generates and refines DAX formulas in seconds, no developer knowledge required.
- Applies AI to enhance data analysis and improve accessibility.

### Productivity & speed

- Speeds up visual creation and data exploration using the Q&A feature.
- Accelerate report design by building a ready-to-use foundation.
- Help developers explore and understand data faster to save effort.

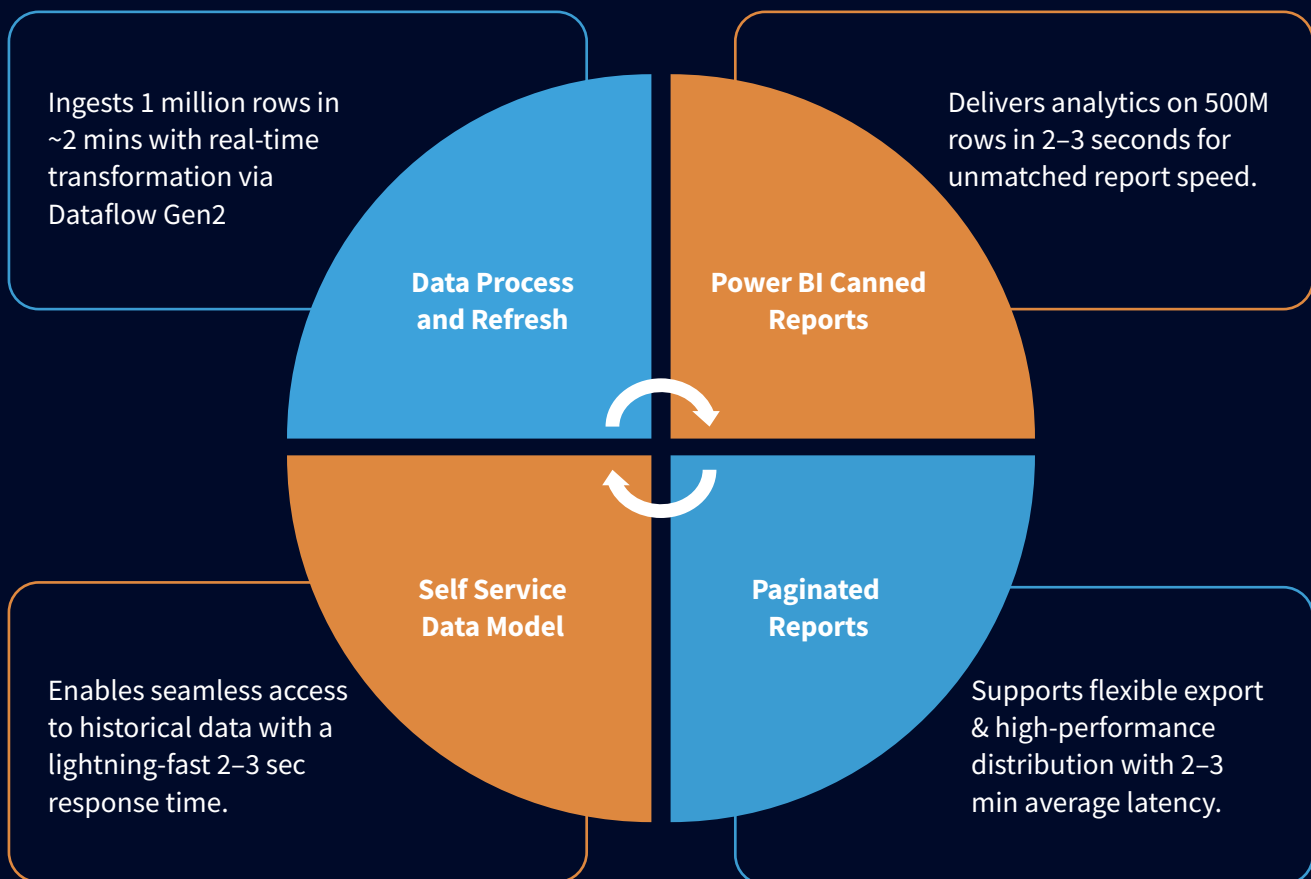
### How Copilot assists the developers and users

- Empower developers and report authors to create data analysis expression (DAX) formulas to reduce development efforts and increase efficiency.
- Copilot can analyze data and recommend/create report pages automatically or based on user requirements (limitations apply).
- It uses natural language commands to modify your data model. A simple command, "add a relationship between the Sales and Date tables" or "create a calculated column for Year-to-Date Sales," and it will execute these tasks.
- It provides interactive recommendations for improving your data model, such as suggesting new measures, optimizing existing ones, and identifying potential issues with relationships or data integrity.
- Copilot can summarize the data model, helps to understand the structure and key elements of the data. to the generated insight can help to suggest and optimize the data model for better performance and accuracy.

## Performance and benefits analysis

The following performance insights are based on a proof of concept (POC) using a massive dataset containing 500 million transaction rows in a fact table. Performance was evaluated across the following dimensions:

1. Canned report access and navigation experience.
2. Self-serve report creation experience.
3. Paginated report performance.
4. Report data extraction, scheduling, and sharing.



## Benefits and considerations

### Benefits

- Combines the best of data lake and data warehouse paradigms.
- Enables high-speed querying with Direct Lake, no added cost.
- Supports real-time data updates in Power BI reports.
- Optimized for handling extensive datasets.
- Automates semantic model updates with minimal effort.

### Considerations

- Microsoft Fabric is evolving and may have limitations in specific use cases.
- Initial setup may require investment, despite long-term cost efficiency.

## References

<sup>1</sup>Gartner, How Data Fabric Can Optimize Data Delivery:

<https://www.gartner.com/en/data-analytics/topics/data-fabric>

<sup>2</sup>Microsoft documentation, Limitations of Microsoft Fabric Data Warehouse, May 19, 2025:

<https://learn.microsoft.com/en-us/fabric/data-warehouse/limitations>

<sup>3</sup>Microsoft documentation, Dataflow Gen2 Architecture, February 27, 2025:

<https://learn.microsoft.com/en-us/fabric/data-factory/pricing-dataflows-gen2>

<sup>4</sup>Microsoft documentation, Getting from Dataflow Generation 1 to Dataflow Generation 2, April 24, 2025:

<https://learn.microsoft.com/en-us/fabric/data-factory/dataflows-gen2-overview>

<sup>5</sup>Microsoft documentation, What is a lakehouse in Microsoft Fabric?, February 4, 2025:

<https://learn.microsoft.com/en-us/fabric/data-engineering/lakehouse-overview>

<sup>6</sup>Microsoft documentation, Fabric Eventstream – overview, March 25, 2025:

<https://learn.microsoft.com/en-us/fabric/real-time-intelligence/event-streams/overview?tabs=enhancedcapabilities>

<sup>7</sup>Microsoft documentation, Ingest data into your Warehouse using data pipelines, April 6, 2025:

<https://learn.microsoft.com/en-us/fabric/data-warehouse/ingest-data-pipelines>

<sup>8</sup>Microsoft documentation, Ingest data into the Warehouse, April 6, 2025:

<https://learn.microsoft.com/en-us/fabric/data-warehouse/ingest-data>

# About the authors

---



## Sanjay Biswas

**BI Architect, Data Architecture and Strategy, LTIMindtree**

Sanjay has 14+ years of expertise in data analytics, BI solutions, migration, and end-to-end implementation. He excels in data visualization strategies that drive data-informed decisions.



## Suman Gupta

**Data Architect, Data Architecture and Strategy, LTIMindtree**

Suman has 15+ years of experience in data analytics, data management, and end-to-end implementation. He specializes in data migration strategies that drive informed decision-making.



## Kallol Dutta

**Data Strategy and Data Integration CoE Head in LTIMindtree**

Kallol has over 25 years of industry experience and has led multiple strategy consulting projects and implementation engagements. He has authored and published point of view and whitepapers across different data concepts and technologies.

**LTIMindtree** is a global technology consulting and digital solutions company that enables enterprises across industries to reimagine business models, accelerate innovation, and maximize growth by harnessing digital technologies. As a digital transformation partner to more than 700 clients, LTIMindtree brings extensive domain and technology expertise to help drive superior competitive differentiation, customer experiences, and business outcomes in a converging world. Powered by 86,000+ talented and entrepreneurial professionals across more than 40 countries, LTIMindtree — a Larsen & Toubro Group company — solves the most complex business challenges and delivers transformation at scale. For more information, please visit <https://www.ltimindtree.com/>.