

Microsoft Fabric for Streamlined Data Engineering

Role of Shortcuts, Data Mirroring, Copilot

Data engineering is pivotal in steering businesses that harness data for actionable insights, strategic decisions, better customer experience, and innovation. The meteoric growth of cloud solutions and data has rendered traditional on-prem infrastructure with hardware constraints obsolete in addressing scaling challenges. Cloud Data platforms have emerged as a transformative solution, reshaping the execution of data engineering tasks through scalable, cost-effective data pipelines and processing systems. This shift has resulted in sophisticated, scalable, and cost-effective data pipelines and processing systems.

As new-age cloud platforms and tools evolve, data engineering is now evolving through integration with artificial intelligence (AI) and, machine learning (ML). The new Software-as-a-Service (SaaS) offerings help in reducing the provision and maintenance of compute and storage. Organizations are now also trying to leverage Gen AI to help end users jumpstart these data transformation activities.

— The key goals for Data Engineers

Data engineering is essential for managing digital flow and fuels business intelligence (BI). Within this wide-ranging practice, **data engineers handle the critical tasks of:**



Data Ingestion

Diverse and varying volumes of structured, semi-structured, and unstructured data from multiple sources require varying frequency of transformation (real-time or batch loads) with different velocities for effective analysis.



Data Integration

After data is ingested, the data integration process merges the landed data, transforms it, and prepares it by combining data from different data clusters to create a unified view, ensuring consistency and accuracy.



Data Processing

To handle the large volume, diversity, and frequency of data, data engineers have to collect, organize, cleanse, and verify the data across the data landscape. Once data is sanitized, their efforts are spent on analyzing the data and converting it into valuable insights for businesses. To ensure accurate analytical results, data engineers establish frameworks for parallel, real-time processing and uphold data integrity.



Data Quality and Governance

Data governance is crucial for any data ecosystem to reduce data duplication and build a trusted data lake. Data engineers have to connect disparate data silos while ensuring data accessibility and quality for cross-functional insights. Data engineers have to implement real-time monitoring and alerts for data changes. Through this governance, users are empowered to understand and use data effectively thus unlocking the full potential of Data.



Data Pipeline Orchestration

Data engineers have to ensure efficient data movement, data consistency and quality, scalability, performance, monitoring, and alerting for pipeline health anomalies. Building these checks and balances is a cumbersome task as it requires orchestrating various tools and services to ensure it. They also have to ensure that pipelines can scale seamlessly and perform efficiently.

Streamlining data-driven journeys with Microsoft Fabric

A comprehensive data-engineering solution for the new age, Microsoft Fabric connects data across sources, delivers batch or real-time analytics, and empowers organizations with in-built, role-tailored tools to reduce complexities and generate actionable insights. It unifies components from Power BI, Synapse Data Engineering, and Synapse Data Warehousing into a single environment.

The three unique features that make Fabric an indispensable tool for data engineering are **Shortcuts**, **Data Mirroring**, and **Copilot**. Together, they create a cohesive ecosystem to simplify and optimize data ingestion, integration, processing, quality, governance, and data pipeline orchestration. Exploring each feature's customization reveals its transformative impact on data management, batch or real-time processing, advanced analytics, and better business decisions.



One Lake

SaaS Foundation

01

Shortcuts

Shortcuts are objects in Microsoft OneLake that point to other storage locations such as Azure Storage Account, AWS S3, or even OneLake itself. They help simplify workflows, accelerate data engineering tasks, and streamline repetitive processes by providing a simple mechanism to virtualize data enabling reusable sequences of actions, such as data transformations, validations, and integration. Data engineers can use them to build, deploy, and manage data pipelines without extensive coding.

Shortcuts facilitate seamless data orchestration across diverse sources and destinations, empowering organizations to handle complex data workflows efficiently and quickly achieve their data integration objectives.

The data engineering scenarios that are easier to handle with Shortcuts include:

Data Ingestion

Previously data engineers would have to ingest data from various sources (Azure, AWS, on-premises) into a unified data lake. With Microsoft Fabric, data engineers can create shortcuts pointing to existing data sources thus eliminating edge copies of data. This removes latency associated with data copies and staging environments.

Data Integration

Previously, data engineers would have to manually unify data across domains, clouds, and accounts. Through Microsoft Fabric, data engineers can now use shortcuts to create a single virtual data lake by connecting to existing data sources (Azure, AWS, OneLake). They can also manage permissions and credentials centrally. This new approach simplifies data integration and reduces complexity.

Data Warehousing

When dealing with large-scale data processing for your enterprise data warehouse, data engineers can create shortcuts to relevant data tables or files, thus avoiding unnecessary data duplication and hence optimizing batch processing by directly accessing data through shortcuts. This approach through Microsoft Fabric reduces latency and leads to efficient data movement, and simplified orchestration.

Data Transformation

Earlier, data engineers would have to spend time and effort in transforming raw data into usable formats (e.g., Parquet, Delta). However, through Microsoft Fabric, data engineers can now create shortcuts within lakehouses (tables or file folders) and these shortcuts automatically synchronize metadata for Delta/Parquet formats. This enables seamless access to Spark, SQL, and other analytical engines. Through this feature, Microsoft Fabric enables efficient data processing without manual setup/intervention.

Real-time Analytics

In Microsoft Fabric if the requirement is to analyze streaming data in real-time, data engineers can set up shortcuts to streaming data sources and eliminate the need for intermediate copies, thus enabling seamless integration with real-time analytics engines. This helps in achieving faster insights, reduces complexity, and streamlines data access.

Implementation of ML Models

In case there is a need to build predictive models or run machine learning experiments, data engineers can leverage shortcuts to relevant feature datasets, avoiding data duplication during model training and seamlessly integrating with Azure ML or other AI/ML tools. This accelerates model development, leading to better resource utilization, and simplifying data access.

02 Data Mirroring

Microsoft Fabric is a cutting-edge data engineering technology that offers a range of advanced capabilities, with data mirroring being a standout feature. In today's data-driven world, analytics platforms are only as good as the data they operate on. However, with the exponential growth of data across different applications, databases, and warehouses, managing and ingesting this data for analytics and AI becomes a daunting task. Microsoft Fabric's mirroring resolves this issue by allowing seamless access to any database or data warehouse within Fabric. Mirroring eliminates the need to switch clients or deal with proprietary storage formats, and familiar database editors simplify management.

Mirroring technology ensures real-time data replication by capturing every transaction and transforming it into Delta tables in Fabric's OneLake. This results in immediate access to changes in the source data without the need for complex extraction, transformation, and loading (ETL) processes. Every mirrored database has a SQL Analytics Endpoint, housing metadata and pointing to OneLake data. This makes querying effortless for SQL developers and citizen developers. Fabric's built-in advanced analytics and AI capabilities thrive on this unified, real-time data.

Microsoft Fabric's mirroring bridges data silos, accelerates insights, and simplifies the data journey. It revolutionizes data engineering by offering unified data access, real-time data replication, and simplified data warehousing. The technology is a game-changer for data engineers, analysts, and organizations hungry for actionable intelligence.

	Scenario	Solution with Fabric data mirroring	Benefit
Data Ingestion	If the requirement is to continuously ingest data from external databases or data warehouses into your data lake.	<ul style="list-style-type: none"> ▶ Set up mirroring to replicate data snapshots from the source to Fabric's data warehousing. ▶ No need to move the source data; it remains in place Create shortcuts to access mirrored data seamlessly.	Real-time data replication without ETL complexity.
Data Processing	Transforming and processing data for analytics or reporting.	<ul style="list-style-type: none"> ▶ Mirrored data is available in near real-time. ▶ Cross-join and query data directly within Fabric. Optimize data processing without additional data movement.	Faster insights, reduced latency, and simplified processing.
Data Integration	Unify data from various sources (on-premises, cloud, external databases).	<ul style="list-style-type: none"> ▶ Mirroring bridges the gap between external databases and Fabric. ▶ Centralize data access without switching clients. Use shortcuts to integrate mirrored data seamlessly.	Simplified data integration, consistent access, and reduced complexity.

03 Copilot

An AI-powered assistant in Microsoft Fabric, Copilot provides intelligent suggestions and automates code generation for data engineering tasks. It analyzes code patterns, context, and user input to provide relevant recommendations on data transformation functions or pipeline optimizations.

Copilot reduces manual coding efforts, improves code quality, and fosters collaboration between team members. Its ML capabilities enhance the productivity and efficiency of building and managing data pipelines within the Fabric ecosystem.

The data engineering scenarios redefined by Copilot include:



Complex Data Transformations

Copilot assists in creating data pipelines (e.g., Azure Data Factory) and suggests effective algorithms with pre-built functions for complex data transformations, improving development and ensuring accuracy in tasks such as feature engineering or data cleaning.



Optimizing Data Pipelines

This feature analyzes pipeline configurations and offers optimization suggestions - such as parallel processing techniques or resource allocation adjustments - to augment performance and scalability.



Preventing Errors

Copilot suggests potential coding errors or inconsistencies in data engineering. It provides real-time feedback and suggests corrective actions to minimize debugging time and ensure optimized data processing pipelines in production environments.

Copilot's functionalities make it a valuable tool for:



Automated Code Generation and Documentation

Copilot accelerates data engineering tasks by generating code snippets for operations such as data parsing or SQL queries. It also automates the generation of documentation for data flow descriptions and, transformation processes, saving time and ensuring project transparency and maintainability.



Collaborative Development

By suggesting best practices, code optimizations, and alternative solutions, Copilot boosts collaboration among data engineers. It helps engineers grow their data skills and knowledge through a platform that provides a personalized and interactive learning experience, feedback, and guidance.




Onboarding New Team Members

Copilot also helps new team members understand codebases quickly and simplifies data workflow management by providing contextual suggestions and explanations, facilitating faster ramp-up and integration into data engineering projects.

Journey to Microsoft Fabric with LTIMindtree's Sunshine

At LTIMindtree, we help our customers move their data estate seamlessly to Microsoft Fabric with an expedited journey powered by Sunshine. In this customized process, our solution gives them the advantages of automation, cost savings, and risk reduction.

 **Connect with us to learn more about Sunshine**

Migration	Greenfield Development	Finops
<p>On prem Datawarehouse to Azure Automated code conversion and data reconciliation</p> <p>      </p> <p>    </p>	<p>Custom Metadata driven Ingestion</p> <p>Built in Data Quality</p> <p>Automated tool kits for Purview</p> <p>    </p>	<p>Utilization Monitoring</p> <p>Cost and Performance</p> <ul style="list-style-type: none"> • optimization • Recommendations <p>Chargeback reports at sub instance level</p> <p>Supported Services</p> <p>     </p>

60%

Automation

25%

Cost Savings

30%

Time Reduction

30%

Risk Reduction

Sunshine is available in Azure marketplace

ETL Migrate	Informatica to Data bricks or Synapse Spark pool	Semantic Migrate	Azure Analysis services to Migrate Power BI Large Data sets	Report Migrate	Legacy Analytics Services to Power BI Migrate
--------------------	--	-------------------------	---	-----------------------	---

References [Microsoft | What is Microsoft Fabric](https://learn.microsoft.com/en-us/fabric/get-started/microsoft-fabric-overview)
<https://learn.microsoft.com/en-us/fabric/get-started/microsoft-fabric-overview>

Get to Fabric Faster with LTIMindtree



Benefits	Strong Experience & history in delivering Industry ready solutions	Focus and Investment in Microsoft Fabric	Automation first approach to get faster to Fabric with 60%+ faster Time to Value for migration
	Strategic Partnership with Microsoft specifically on Fabric	Joint development of Customer pilots by LTIM Azure CoE backed by Fabric Product Engineering team	

Unleashing Fabric's potential: elevating data engineering efficiency

In conclusion, Shortcuts, Data Mirroring, and Copilot synergize to simplify data engineering. Shortcuts help in automate tasks by connecting to data sources without copying, Data Mirroring ensures real-time synchronization, and Copilot leverages Gen AI capabilities to improve productivity for end users. For enterprises, leveraging these features accelerates data activities, eliminating bottlenecks and fostering innovation, leading to enhanced business agility and a competitive edge in dynamic data environments.



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 82,000+ talented and entrepreneurial professionals across more than 30 countries, LTIMindtree — a Larsen & Toubro Group company — combines the industry-acclaimed strengths of erstwhile Larsen and Toubro Infotech and Mindtree in solving the most complex business challenges and delivering transformation at scale. For more information, please visit www.ltimindtree.com