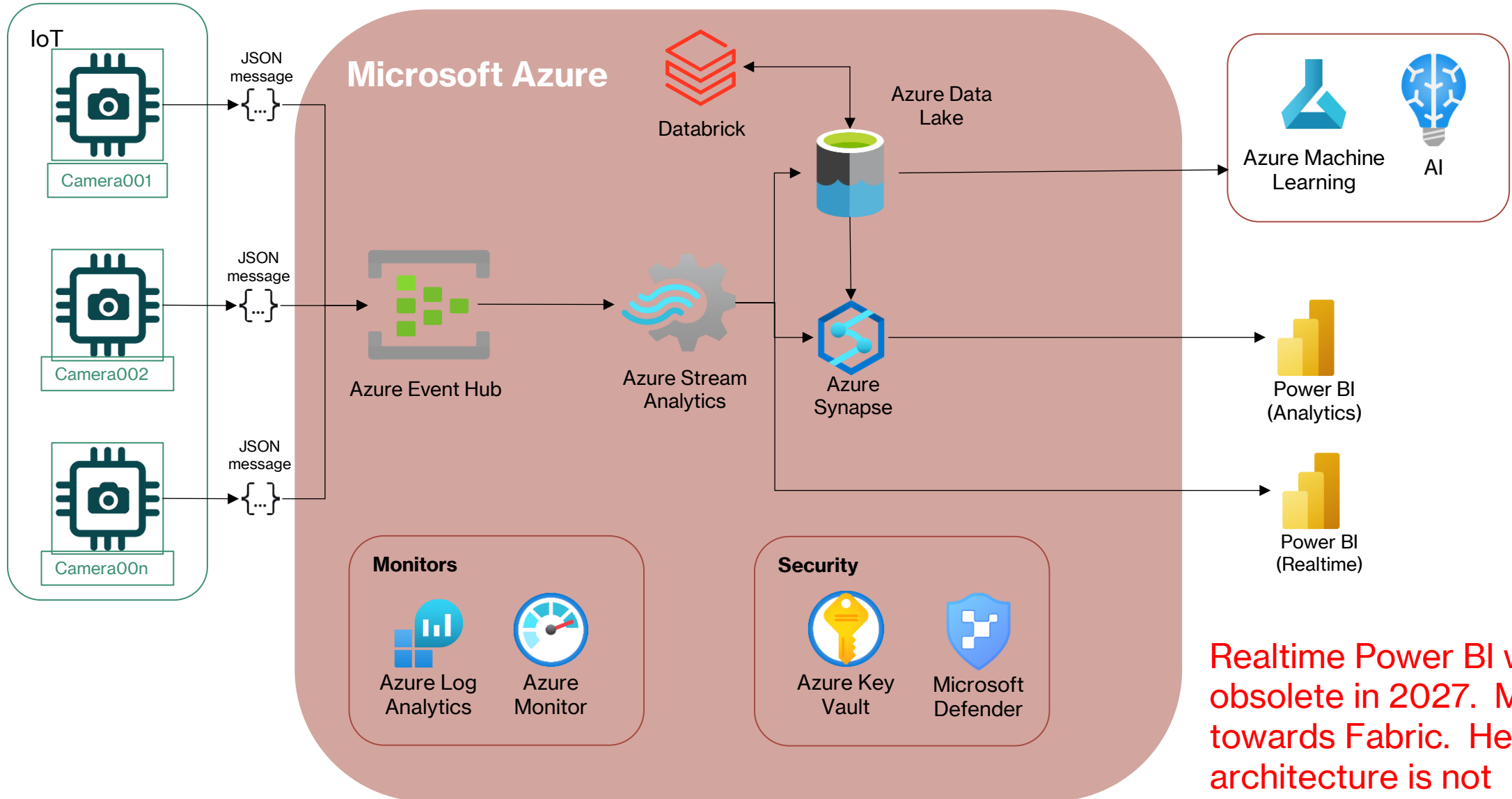




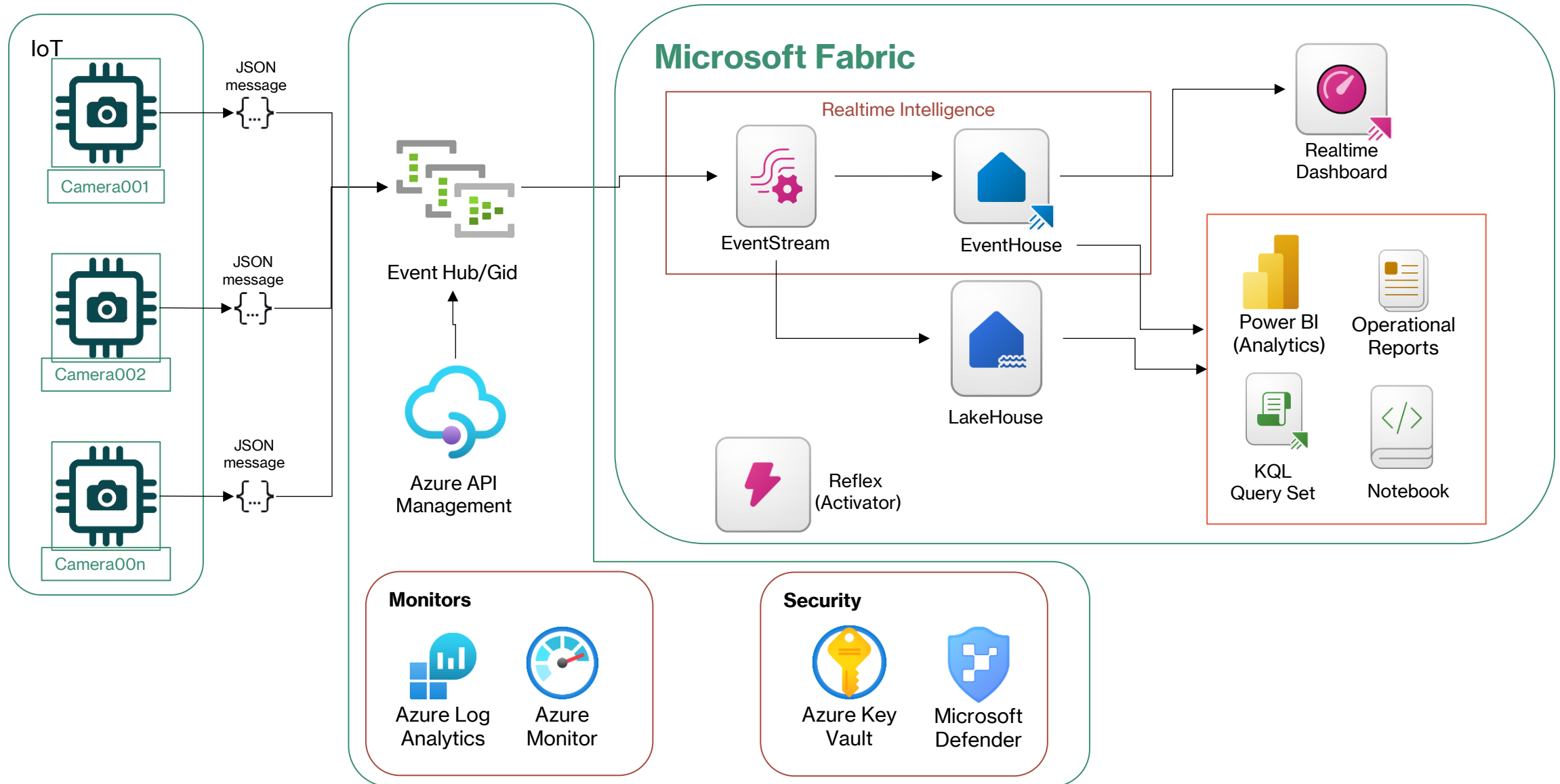
# **Technical Test – Question 3**

# Architecture Diagram

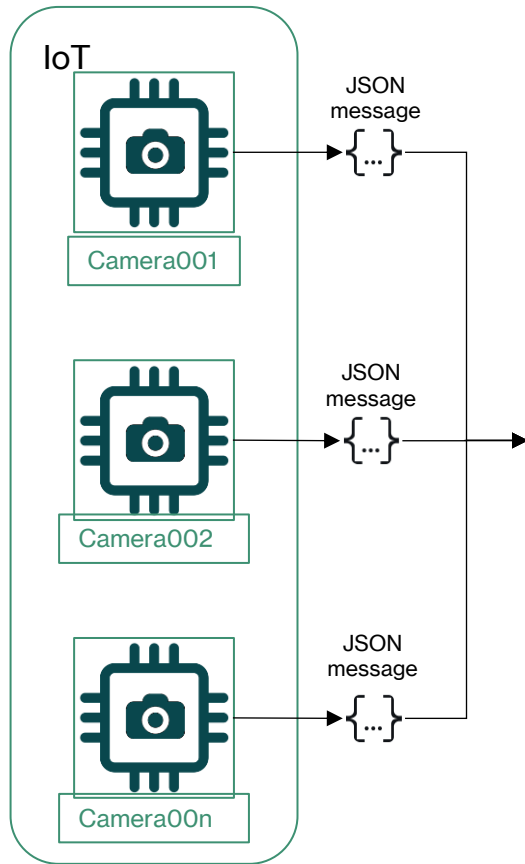


Realtime Power BI will be obsolete in 2027. Moving towards Fabric. Hence this architecture is not recommended.

# Architecture Diagram



# IoT (Camera)



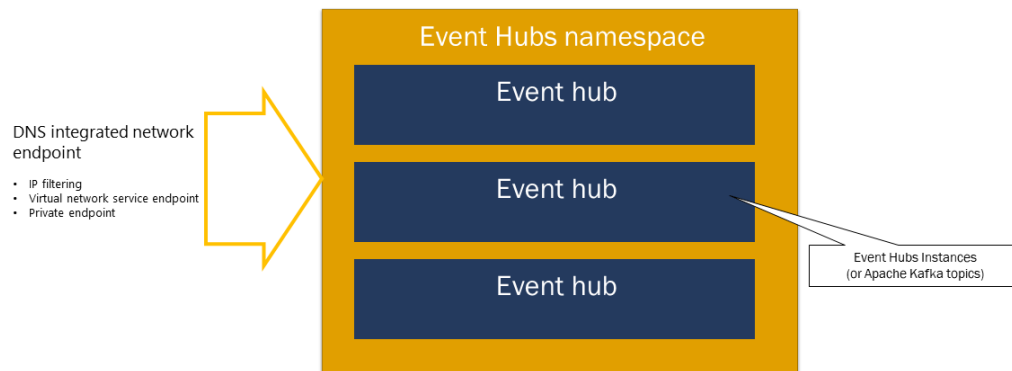
AI cameras are installed at various locations to monitor events and activities. Equipped with built-in AI capabilities, these cameras can process video feeds into JSON messages containing data such as events, timestamps, camera IDs, and locations. They are connected to the internet to push these messages in real-time. In case of internet connectivity issues, the AI cameras can store the messages and send any unsent messages once the connection is restored. Up to 1 millions of messages might be expected across all deployed cameras per second.

# Azure Event Hub



Azure Event Hubs is a scalable event processing service that ingests and processes large volumes of events and data, with low latency and high reliability. It is a native data-streaming service in the cloud that can stream millions of events per second, with low latency, from any source to any destination. Event Hubs is compatible with Apache Kafka. It enables you to run existing Kafka workloads without any code changes. Azure Schema Registry in Event Hubs provides a centralized repository for managing schemas of event streaming applications.

Event Hubs is a multi-protocol event streaming engine that natively supports Advanced Message Queuing Protocol (AMQP), Apache Kafka, and HTTPS protocols. Because it supports Apache Kafka, you can bring Kafka workloads to Event Hubs without making any code changes. You don't need to set up, configure, or manage your own Kafka clusters or use a Kafka-as-a-service offering that's not native to Azure.



An Event Hubs namespace is a management container for event hubs (or topics, in Kafka parlance). It provides DNS-integrated network endpoints and a range of access control and network integration management features such as [IP filtering](#), [virtual network service endpoint](#), and [Private Link](#).

# Azure – API Management



Azure API Management is a hybrid, multicloud management platform for APIs across all environments. As a platform-as-a-service, API Management supports the complete API lifecycle.

- **Centralized Logging:** API Management can log events to Event Hub, capturing detailed information about API requests and responses. This is useful for auditing, analytics, and monitoring
- **Scalability:** Event Hub is designed to handle large volumes of data, acting as a buffer between your API Management service and the infrastructure that processes the logs. This ensures that logging does not impact API performance
- **Integration:** You can use the log-to-eventhub policy in API Management to send messages in a specified format to Event Hub for online or offline analysis
- **Security:** API Management supports managed identities and connection strings for secure access to Event Hub

# Azure – Monitor & Log Analyzer



Azure Monitor is a comprehensive product designed for monitoring and managing the performance and health of your applications, infrastructure, and network. It provides end-to-end observability across Azure, multicloud, and hybrid environments. The Azure monitor is deployed to monitor Event Hub and Microsoft Fabric.

- **Data Collection:** Azure Monitor collects metrics and logs from various sources, including applications, virtual machines, containers, databases, and networking events
- **Analysis and Visualization:** It offers tools for analyzing and visualizing data, helping you understand how your applications are performing and identify issues
- **Alerts and Automation:** You can set up alerts based on specific conditions and automate responses to system event
- **Integration:** Azure Monitor integrates with other Azure services and third-party tools, enhancing its capabilities



Azure Log Analytics is a tool within Azure Monitor that allows you to query and analyze log data from various sources. Here are some key features:

- **Log Querying:** Run and edit log queries against data in the Azure Monitor Logs store.
- **Data Analysis:** Filter, sort, and aggregate log data, making it easier to derive insights and troubleshoot issues
- **Integration:** Integrates with other Azure services like Microsoft Sentinel, Microsoft Defender for Cloud, and Logic Apps, as well as tools like Power BI and Excel
- **Workspace Management:** A Log Analytics workspace is a data store where you can collect log data from Azure and non-Azure resources. You can manage data retention, access, and transformation within this workspace

# Azure – Key Vault & Defender



Azure Key Vault is a cloud service designed to securely store and manage sensitive information such as API keys, passwords, certificates, and cryptographic keys.

- **Secure Storage:** Secure environment for storing secrets and keys, ensuring they are protected from unauthorized access
- **Access Control:** Control access to secrets and keys using Azure Active Directory (AAD) and role-based access control (RBAC)
- **Integration: Key** Vault integrates with various Azure services, allowing you to use stored secrets and keys in your applications without exposing them
- **Automated Management:** Supports automated key rotation, backup, and recovery, simplifying the management of cryptographic keys



Microsoft Defender for Cloud (formerly known as Azure Defender) is a comprehensive cloud-native application protection platform (CNAPP) designed to secure your cloud and hybrid environments.

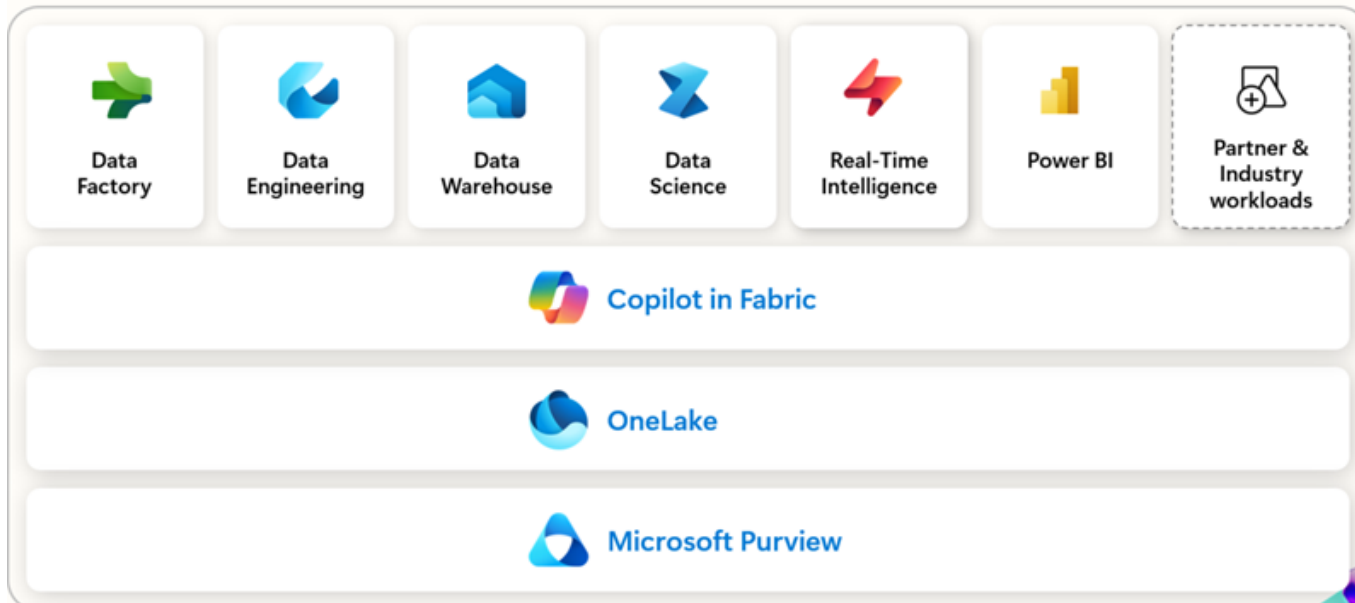
- **Cloud Security Posture Management (CSPM):** Continuously assesses your cloud environment to identify and remediate vulnerabilities, ensuring your resources are configured securely
- **Cloud Workload Protection (CWP):** Provides advanced threat protection for various workloads, including virtual machines, containers, databases, and storage
- **Threat Detection and Response:** Monitors your environment for suspicious activities and provides actionable alerts to help you respond to potential threats
- **Event Hub:** Detect anomalies and potential threats in your event streams, ensuring the security of your real-time data processing
- **Microsoft Fabric:** Integrate Defender for Cloud with Fabric to secure your data science and analytics workloads. It helps protect your data, notebooks, and machine learning models from potential threats



# Microsoft Fabric



Microsoft Fabric is a unified enterprise-ready, end-to-end cloud-based analytics platform that delivers data management and analytics as a software as a service (SaaS). It unifies data movement, data processing, ingestion, transformation, real-time event routing, and report building. The cloud-native architecture handles increasing data volumes and complexities without compromising performance or security.



Fabric integrates workloads like Data Engineering, Data Factory, Data Science, Data Warehouse, Real-Time Intelligence, Industry solutions, Databases, and Power BI into a SaaS platform.

The **OneLake** is the foundation for all Fabric workloads. It's built into the platform and serves as a single store for all organizational data.

# Microsoft Fabric : Real-Time Intelligence



Real-Time Intelligence can be used for data analysis, immediate visual insights, centralization of data in motion for an organization, actions on data, efficient querying, transformation, and storage of large volumes of structured or unstructured data. The Real-Time Hub seamlessly connects time-based data from various sources using no-code connectors, enabling immediate visual insights, geospatial analysis, and trigger-based reactions that are all part of an organization-wide data catalog

# Microsoft Fabric : Real-Time Intelligence



Eventstreams in Microsoft Fabric provide a no-code solution for capturing, transforming, and routing large volumes of real-time events to various destinations. They support multiple data sources and destinations, including connectors to external sources like Apache Kafka, database change data capture feeds, AWS Kinesis, and Google Pub/Sub.

With Eventstreams, you can perform filtering, data cleansing, transformation, windowed aggregations, and duplicate detection to shape the data as needed. Additionally, content-based routing allows you to send data to different destinations based on filters. The derived eventstreams feature enables the creation of new streams from transformations and aggregations, which can be shared with consumers in the Real-Time hub.



Eventhouses provide a solution for handling and analyzing large volumes of data, particularly in scenarios requiring real-time analytics and exploration. They're designed to handle real-time data streams efficiently, which lets organizations ingest, process, and analyze data in near real-time. Eventhouses provide a scalable infrastructure that allows organizations to handle growing volumes of data, ensuring optimal performance and resource use. Eventhouses provide unified monitoring and management across all databases and per database. Eventhouses are tailored to time-based, streaming events with structured, semistructured, and unstructured data. You can get data from multiple sources, in multiple pipelines (For example, Eventstream, SDKs, Kafka, Logstash, data flows, and more) and multiple data formats. This data is automatically indexed and partitioned based on ingestion time.

# Microsoft Fabric : Lakehouse



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 solution for data engineering and analytics. A lakehouse combines the scalability of a data lake with the performance and structure of a data warehouse, providing a unified platform for data storage, management, and analytics.

There are several ways to interact with the Lakehouse and the data within:

- **The Lakehouse explorer:** The explorer is the main Lakehouse interaction page. You can load data in your Lakehouse, explore data in the Lakehouse using the object explorer, set MIP labels & various other things.
- **Notebooks:** Data engineers can use the notebook to write code to read, transform and write directly to Lakehouse as tables and/or folders.
- **Pipelines:** Data engineers can use data integration tools such as pipeline copy tool to pull data from other sources and land into the Lakehouse.
- **Apache Spark job definitions:** Data engineers can develop robust applications and orchestrate the execution of compiled Spark jobs in Java, Scala, and Python.
- **Dataflows Gen 2:** Data engineers can use Dataflows Gen 2 to ingest and prepare their data

# Lakehouse - Medallion Architecture



Medallion Architecture is a data design pattern used to logically organize data in a lakehouse, aiming to incrementally and progressively enhance the structure and quality of data as it moves through each layer. This approach is similar to the different staging areas used in a typical ETL design. The typical design includes

- **Bronze Zone** – This is the zone that stores raw data from external source systems. Data is ingested in its original form without any cleaning or validation. It serves as the initial landing zone for all incoming data. Data retention will be based on requirements and typically should not exceed more than latest 6 months where old data shall be archived into cold data zone.
- **Silver Zone** – This is the zone that stores data that undergoes cleaning and validation processes to ensure its accuracy and consistency. Data retention will be depending on transformation requirements and typically only stores latest copy if assist with having performance efficiency.
- **Gold Zone** – This is the zone that stores Aggregated data, Enriched data, Business Intelligence data and Master data. Data retention will be dependent on business needs and analysis requirements.

# Lakehouse – Ingestion & Data Storage Framework

All data stored in each zones are organized accordingly to ensure Performance, Accuracy & Auditability, Maintainability of ingestion. Data shall be encrypted in transit and at rest at all zone.4

Consideration:



**Performance** – Ensure that data is appended rather than frequently updated. Data retention can impact performance, so it should be planned and designed for an optimal timeframe. Avoid per record processing and use batch record processing for large data processing whenever possible. Data partitioning shall be implemented to efficiently improve performance.

**Accuracy & Auditability** – Data stored should be accurate. Data validation rules shall be implemented during data ingestion at different stages to ensure accuracy. Source and Source record identity is recommended to be stored at each record to trace back to original stage of data. Data ingestion timestamp shall be

**Maintainability** – Data storage should be designed such that it can be maintained easily. Maintainability includes scaling up and out minimizing changes & migration. Data Ingestion logic should be designed to minimize rework and stackable for any enhancement.

# Microsoft Fabric : End-Users Tools



A real-time dashboard in Microsoft Fabric is a dynamic tool that allows you to visualize and monitor data as it flows through your system.

- **Interactive Visuals:** Real-time dashboards consist of tiles, each representing a visual based on a Kusto Query Language (KQL) query. These visuals update automatically to reflect the latest data
- **Data Sources:** Connect to various data sources within your workspace, enabling you to pull in data from different databases and services
- **Customization:** Customize the appearance and behavior of your dashboard tiles to suit your needs. This includes selecting different visual types and formatting options
- **Auto Refresh:** Dashboards can be set to auto-refresh, ensuring that the data displayed is always up-to-date
- **Real-Time Insights:** Designed for event-driven scenarios, real-time dashboards provide immediate visual insights into your data, helping you make quick, informed decisions



KQL Queryset in Microsoft Fabric is a tool used to run, view, and customize query results on data from various sources, such as Eventhouse, KQL databases, and Azure Data Explorer clusters.

- **Data Exploration:** Explore data from different sources like Eventhouse, KQL databases, and Azure Data Explorer clusters using Kusto Query Language (KQL)
- **Custom Queries:** Write and run custom KQL queries to retrieve specific data, perform aggregations, and analyze trends
- **Multi-Source Analysis:** Supports querying across multiple data sources, enabling comprehensive data analysis and correlation
- **Interactive Results:** Provides an interactive results grid to view and customize the output of queries

# Microsoft Fabric : End-Users Tools



A notebook in Microsoft Fabric is a web-based interactive environment used primarily for developing Apache Spark jobs and machine learning experiments

- **Interactive Surface:** A rich, interactive surface where data scientists and engineers can write code, visualize data, and document their work using Markdown
- **Multi-language Support:** You can use multiple languages within a single notebook, including PySpark (Python), Spark (Scala), Spark SQL, and SparkR
- **Enhanced Authoring:** Notebooks come with IDE-style IntelliSense, syntax highlighting, error markers, and code completions to help you write code efficiently
- **Data Exploration and Processing:** Easily explore and process data with intuitive low-code experiences. You can analyze data across various formats like CSV, JSON, Parquet, and Delta Lake
- **Security and Execution:** Notebooks can be executed interactively, as part of a pipeline, or via a scheduler, each running under different security contexts



Operational  
Reports



Power BI is the primary visualization layer in Microsoft Fabric. It allows users to quickly transform data from a lake into stunning visuals embedded



Paginated reports are useful when performing Operational reporting. Designed for creating highly formatted, pixel-perfect reports that are ideal for printing or sharing as PDFs