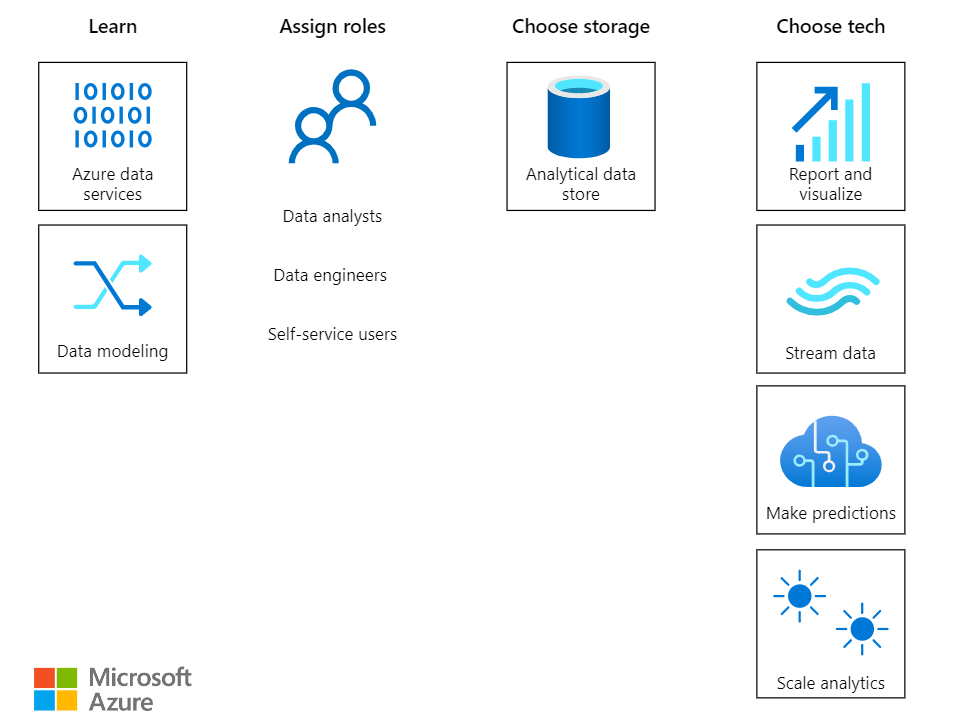
Cloud Computing Essentials with Azure and AWS

# **2.1 Getting Started with Microsoft and Azure**

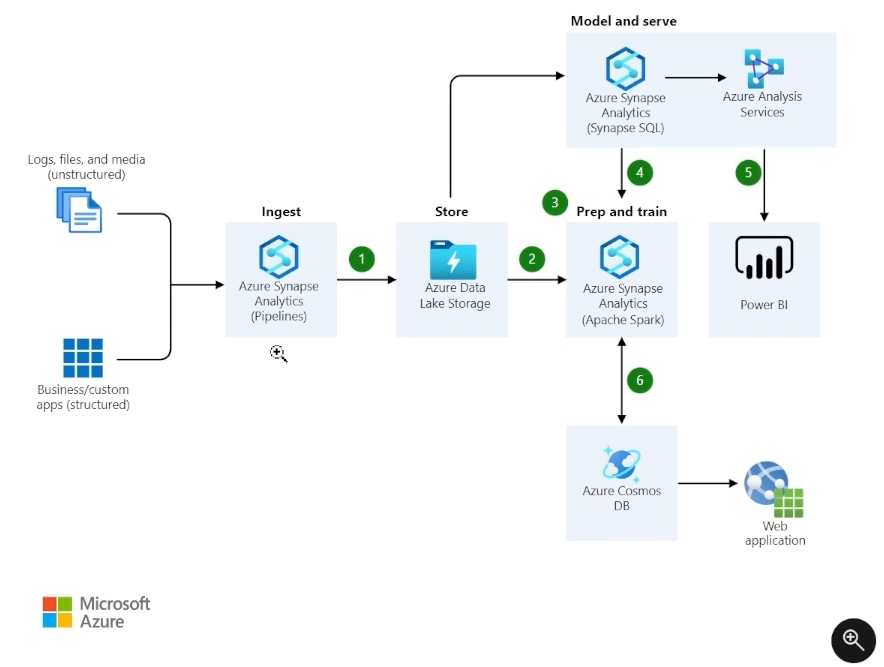
## Azure UI Walkthrough

* **Azure Portal**:
  + Web-based interface that allows you to build, manage, and monitor Azure resources.
* **Azure Portal Walkthrough**:
  + **Home Dashboard**: Upon signing in, the Home Dashboard provides quick access to your resources and services. You can customize this dashboard by adding or removing tiles to suit your workflow.
  + **Navigation Menu**: Located on the left, this menu offers access to all Azure services. You can pin frequently used services for quicker access.
  + **Resource Management**: Use the "All Resources" and "Resource Groups" options to view and manage your assets. Resources can be organized into groups for better management.
  + **Search Bar**: At the top, the search bar helps you quickly locate resources, services, and documentation.
  + **Notifications and Settings**: The top-right corner displays notifications for ongoing operations and provides access to settings where you can configure preferences.
  + **Cloud Shell**: Access Azure CLI or PowerShell directly from the portal using the Cloud Shell icon, enabling command-line management of your resources.
* **Azure Service Categories**:
  + **AI + Machine Learning**: Develop intelligent applications with services like Azure Machine Learning for building models and Cognitive Services for adding capabilities such as vision, speech, and language understanding.
  + **Analytics**: Gather, process, and analyze data using services like Azure Synapse Analytics for data warehousing and Azure Stream Analytics for real-time data processing.
  + **Compute**: Access scalable computing resources with Virtual Machines for flexible workloads and Azure Functions for serverless computing.
  + **Containers**: Deploy and manage containerized applications using Azure Kubernetes Service (AKS) and Azure Container Instances for easy container deployment.
  + **Databases**: Utilize managed database services such as Azure SQL Database for relational data and Azure Cosmos DB for NoSQL options.
  + **Developer Tools**: Enhance development workflows with Azure DevOps for CI/CD pipelines and Visual Studio App Center for mobile app development.
  + **DevOps**: Implement continuous integration and delivery using Azure Pipelines and monitor applications with Azure Monitor.
  + **Hybrid + Multicloud**: Extend Azure services to on-premises environments with Azure Arc for resource management across hybrid and multicloud setups.
  + **Identity**: Manage user identities and access with Microsoft Entra ID (formerly Azure Active Directory) for secure authentication and authorization
  + **Integration**: Connect applications and data with services like Azure Logic Apps for workflow automation and Azure API Management for publishing APIs.
  + **Internet of Things (IoT)**: Connect and manage IoT devices using Azure IoT Hub and analyze IoT data with Azure IoT Central.
  + **Management and Governance**: Maintain control over your Azure environment with Azure Policy for compliance and Azure Cost Management for expense tracking.
  + **Media**: Deliver high-quality media content using Azure Media Services for video encoding and streaming solutions.
  + **Migration**: Simplify the transition to Azure with Azure Migrate, offering tools and guidance for moving workloads to the cloud.
  + **Mixed Reality**: Create immersive experiences with Azure Mixed Reality services, including Azure Spatial Anchors for spatial context in applications.
  + **Mobile**: Build and deploy mobile apps using Azure App Service and engage users with push notifications through Azure Notification Hubs.
  + **Networking**: Connect and manage network resources with Azure Virtual Network for isolated networks and Azure Load Balancer for distributing traffic.
  + **Security**: Protect your environment with Azure Security Center for threat management and Azure DDoS Protection against distributed denial-of-service attacks.
  + **Storage**: Store and access data at scale with Azure Blob Storage for unstructured data and Azure File Storage for managed file shares.
  + **Virtual Desktop Infrastructure**: Enable secure remote work with Azure Virtual Desktop, providing virtualized desktops and applications.
  + **Web & Mobile**: Develop and host web applications using Azure App Service and reach users on any device with responsive designs.



## Azure Data Workflows and Architecture

* **Azure Tech Stack (Data Analytics - Basic)**:



**1. Ingest Data**

* **Purpose**: Gather raw data from various sources like logs, files, media (unstructured), and business/custom applications (structured).
* **Tool**: Azure Synapse Analytics (Pipelines)
  + Orchestrates and automates data movement.
  + Supports ETL (Extract, Transform, Load) and ELT (Extract, Load, Transform) processes.
  + Handles diverse data types, making it suitable for both structured and unstructured data.
* **Outcome**: Data is moved into Azure for processing and storage.

**2. Store Data**

* **Purpose**: Persist raw and processed data for later analysis and modeling.
* **Tool**: Azure Data Lake Storage
  + Scalable and secure storage optimized for big data analytics.
  + Supports hierarchical namespaces, making data organization efficient.
  + Works seamlessly with analytics tools like Azure Synapse and Spark.
* **Outcome**: Data is securely stored and accessible for further processing.

**3. Prep and Train**

* **Purpose**: Process, clean, and analyze data for machine learning or advanced analytics.
* **Tool**: Azure Synapse Analytics (Apache Spark)
  + Apache Spark is used for large-scale data processing and analytics.
  + Provides built-in libraries for machine learning and data preparation.
  + Allows distributed computing for efficient handling of large datasets.
* **Outcome**: Data is prepared for insights generation and model training.

**4. Model and Serve**

* **Purpose**: Build analytical models and serve data insights to end-users or downstream applications.
* **Tools**:
  + **Azure Synapse Analytics (Synapse SQL)**:
    - SQL-based analytics engine for querying and analyzing data.
    - Offers on-demand and provisioned resources for flexibility.
  + **Azure Analysis Services**:
    - Provides enterprise-grade models for data analysis.
    - Enables semantic layer creation for simplified data visualization and reporting.
* **Outcome**: Analytical models are created and made ready for querying or direct visualization.

**5. Visualize and Analyze**

* **Purpose**: Display insights and enable data-driven decisions.
* **Tool**: Power BI
  + Connects to Azure Synapse and other sources for live analytics.
  + Allows users to create interactive reports and dashboards.
  + Facilitates real-time monitoring of business metrics.
* **Outcome**: Insights are presented visually, aiding decision-making.

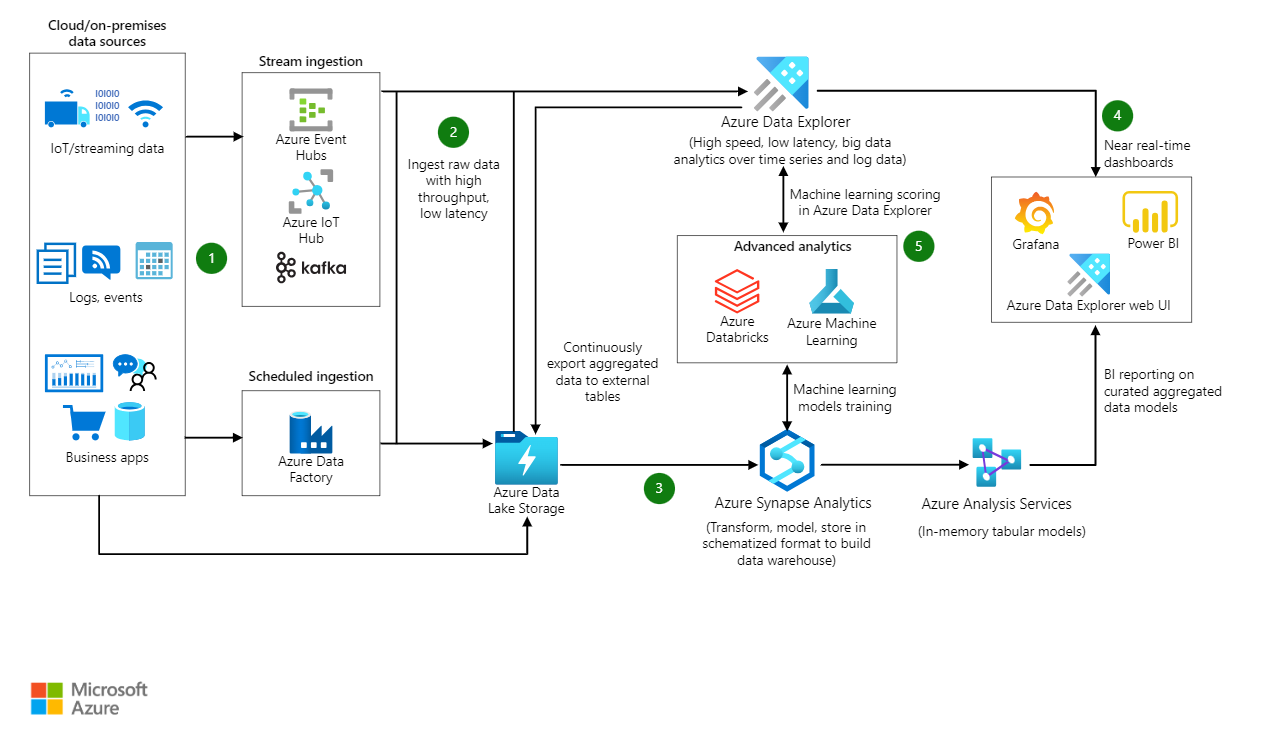
**6. Operationalize**

* **Purpose**: Leverage processed data for operational use and application integration.
* **Tool**: Azure Cosmos DB
  + Globally distributed, NoSQL database designed for scalability and high availability.
  + Supports multiple APIs (e.g., SQL, MongoDB, Cassandra) for diverse application needs.
* **Integration**: Data is used in applications like web apps for real-time functionalities.
* **Outcome**: Data-driven applications deliver value to end-users.

**End-to-End Workflow Explanation**

1. Raw data (structured or unstructured) is ingested into **Azure Synapse Pipelines**.
2. Data is stored in **Azure Data Lake Storage**, ensuring it's secure and scalable.
3. Data is prepared and trained for analysis using **Spark** for large-scale processing.
4. Analytical models are built and served using **Synapse SQL** and **Azure Analysis Services**.
5. Insights are visualized and analyzed through **Power BI** dashboards.
6. Processed data is operationalized in **Azure Cosmos DB** for use in applications like web apps.

* **Azure Tech Stack (Data Analytics - Advanced)**:



**1. Data Sources**

* **Cloud/On-Premises Data Sources**:
  + Includes IoT devices, streaming data, logs, events, and business applications.
  + Data is generated in real-time or batch formats from these sources.

**2. Ingestion**

* **Stream Ingestion**:
  + **Tools**: Azure Event Hubs, Azure IoT Hub, Kafka.
  + Handles high-throughput, low-latency ingestion of real-time data streams.
  + Best for IoT data, event logs, or clickstream data.
* **Scheduled Ingestion**:
  + **Tool**: Azure Data Factory.
  + Used for batch ingestion of data at scheduled intervals.
  + Ideal for structured data from business apps or on-prem systems.

**3. Storage**

* **Azure Data Lake Storage**:
  + Scalable storage for raw, unstructured, and structured data.
  + Acts as the central repository for both real-time and batch-ingested data.
  + Data is stored securely and made accessible for processing.

**4. Real-Time Analytics**

* **Azure Data Explorer**:
  + High-speed and low-latency analytics for time-series and log data.
  + Enables near real-time insights and machine learning scoring directly within the platform.
  + Data from Azure Data Explorer feeds into tools like:
    - **Grafana**: Visualizes time-series data.
    - **Power BI**: Creates interactive dashboards.
    - **Azure Data Explorer Web UI**: For querying and monitoring directly.

**5. Advanced Analytics and Machine Learning**

* **Azure Databricks**:
  + Collaborative data engineering and advanced analytics platform.
  + Supports Spark-based processing for large-scale data transformation.
* **Azure Machine Learning**:
  + Builds, trains, and deploys machine learning models.
  + Integrates with Azure Data Explorer for real-time scoring.
  + Uses data stored in Azure Data Lake Storage for training.

**6. Data Transformation and Warehousing**

* **Azure Synapse Analytics**:
  + Transforms raw data into a structured format for analytics.
  + Builds a data warehouse for storing curated, schematized data.
  + Supports advanced analytics and integration with Power BI.

**7. BI Reporting and Modeling**

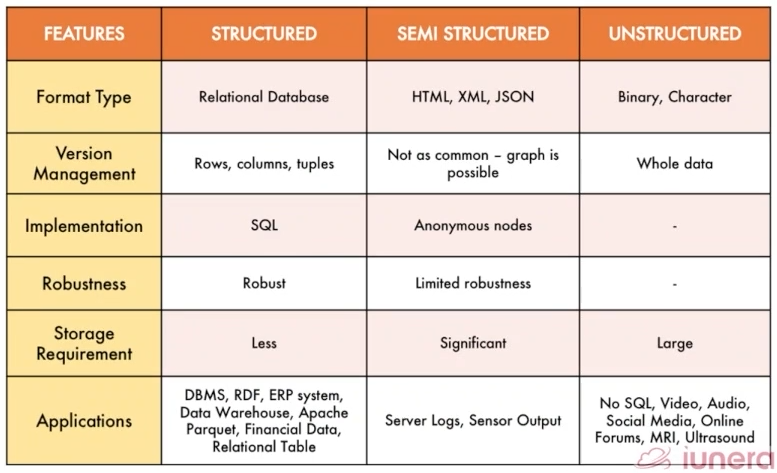
* **Azure Analysis Services**:
  + Creates in-memory tabular models for quick querying and analysis.
  + Enhances performance for large-scale business intelligence (BI) reporting.
* **Power BI and Azure Data Explorer**:
  + Visualizes aggregated data models, making insights actionable.
  + Power BI integrates seamlessly for user-friendly dashboards.

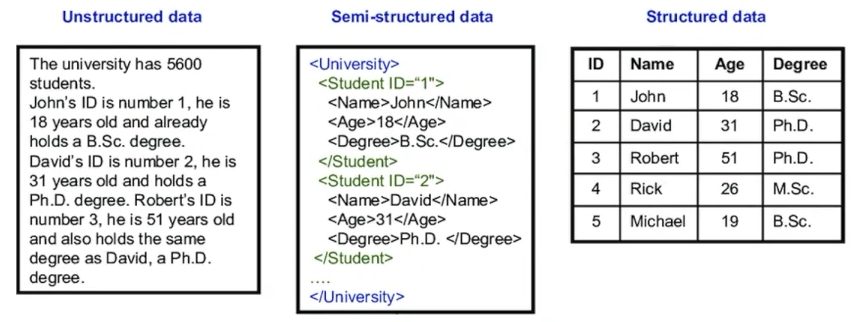
**End-to-End Workflow Explanation**

1. Data Sources generate raw data (real-time or batch).
2. Ingestion pipelines bring data into Azure.
3. Data is stored in Azure Data Lake Storage for secure, scalable access.
4. Azure Data Explorer enables real-time insights and feeds tools like Power BI or Grafana for visualization.
5. Advanced analytics occurs in Azure Databricks and Azure Machine Learning, with models integrated into the system.
6. Transformed data moves into Azure Synapse Analytics, creating a structured warehouse.
7. BI tools like Azure Analysis Services and Power BI provide insights for decision-making.

## Databases, Data Warehouses, and Data Lakes

* **Structured, Semi-Structured, and Unstructured Data**:
  + **Structured Data**: Organized in rows and columns, stored in databases or spreadsheets.
  + **Semi-Structured Data**: Partially organized with consistent patterns, not as strict as relational databases (e.g., JSON, XML).
  + **Unstructured Data**: No fixed format or data model, includes files like images, videos, or emails.





* **Databases vs Data Warehouses**:
  + **Databases**: Store ONLY structured data with strong consistency, reliability, and security.
  + **Data Warehouses**: Handle BOTH structured and semi-structured data for fast querying and analysis.
  + **Performance**: Warehouses offer optimized indexing and queries for large datasets.
  + **Schema**: Both are built around a defined schema.
* **Data Lakes**:
  + Stores large-scale data, especially unstructured data.
  + **Scalability**: Built for flexibility and growth.
  + **Raw Storage**: Provide the ability to keep data in its raw format without predefined schema.
  + **Advanced Analytics**: Perform complex processing on varied, large datasets.
    - **Tools**: Data Bricks
* **Advantages of Databases, Data Warehouses, and Data Lakes**:
  + **Databases**:
    - Ensure strong data consistency, reliability, and security.
    - Offer efficient indexing and fast query optimization.
  + **Data Warehouses**:
    - Optimized for analytical workloads and large dataset analysis.
    - Highly scalable with flexibility for storage and processing.
  + **Data Lakes**:
    - Handle large-scale, diverse data types and workloads.
    - Ideal for unstructured data and large-scale storage.
* **Disadvantages**:
  + **Databases**
    - **Limited Scalability**: Struggles with very large datasets compared to warehouses or lakes.
    - **Structured Data Only**: Requires strict schema, limiting flexibility.
    - **Performance Overhead**: Complex queries on large datasets can slow down performance.
    - **Costly Maintenance**: Requires constant updates, backups, and hardware upgrades.
  + **Data Warehouses**
    - **High Cost**: Expensive to set up, maintain, and scale.
    - **Complex ETL**: Requires extensive ETL processes to load and transform data.
    - **Structured Data Focus**: Not ideal for unstructured or semi-structured data.
    - **Latency**: Batch processing limits real-time insights.
  + **Data Lakes**
    - **Lack of Governance**: Without proper management, can become a "data swamp."
    - **Performance Issues**: Querying raw, unprocessed data can be slow.
    - **Data Quality**: No enforced schema leads to inconsistent or messy data.
    - **Skill Requirement**: Requires advanced tools and expertise for effective use.