Streaming vs. Batch Processing

Overview

Streaming differs from batch processing by providing real-time data handling. It enables immediate responses to events, crucial in scenarios like healthcare, fraud detection, and online retail. Azure Event Hubs serves as the primary service for data ingestion in streaming, allowing multiple applications to connect and process data efficiently.

Key Points

Benefits of Real-Time Data Processing

- **Immediate Responses**: Essential for scenarios requiring quick reactions, such as healthcare and finance.
- **Enhanced Decision-Making**: Allows for decisions based on live data, improving responsiveness.

Use Cases

- **Healthcare**: Real-time monitoring of patients for immediate responses to critical health changes.
- **Fraud Detection**: Monitoring transactions in real-time to prevent unauthorized activities.
- **Manufacturing and Logistics**: Monitoring equipment conditions and traffic in real-time to prevent failures and optimize operations.

Azure Event Hubs

- **Data Ingestion**: Serves as the entry point for data, allowing multiple applications to connect and process data.
- **Scalability**: Handles millions of events per second with low latency.
- **Integration**: Compatible with Apache Kafka, enhancing flexibility and streamlining data management.

Real-Time Data Processing

- **Immediate Analysis**: Allows for the analysis of events as they occur, enhancing responsiveness.
- **Tools**: Utilizes tools like Power BI for real-time data visualization and feedback.

Data Ingestion and Serving

- **Event Sources**: Continuous data generation from applications and IoT devices.
- **Data Serving**: Processed data can be served through real-time dashboards or stored in a data lake for further analysis.

Event Hub Management

- **Namespaces and Pricing Tiers**: Understanding the creation of namespaces and different pricing tiers is crucial for effective utilization.
- **Consumer Groups**: Represent downstream applications accessing data, affecting how events are processed and managed.

Event Retention

- **Retention Period**: Defines how long events are stored, ranging from one hour to three months based on the pricing tier.
- **Capture Feature**: Automatically stores events in a data lake or blob storage for long-term analysis.

Throughput Units

- **Data Ingress and Egress**: Determines how much data can be processed per second, impacting performance and costs.
- **Auto Inflate**: Automatically increases throughput units based on workload needs.

Authorization and Permissions

- **Shared Access Policies**: Define permissions for managing, sending, or listening to events.
- **Consumer Groups and Check pointing**: Ensure data integrity and help applications resume work seamlessly after interruptions.