

## Comparative Document on Built Solutions

Comparative solutions built across real-time, near real-time, and batch processing scenarios, focusing on key Azure technologies like Event Hub, Stream Analytics, Cosmos DB, Synapse Pools, and Serverless SQL Pool. The goal is to compare how data ingestion, processing, and storage are handled.

### Technologies used for Processing:

- **Real-Time Processing:** Event Hub, Stream Analytics, Cosmos DB/Dedicated SQL Pool.
- **Near Real-Time Processing:** Event Hub, Stream Analytics, Synapse Spark Pool, Cosmos DB.
- **Batch Processing:** Synapse Dedicated SQL Pool, Synapse Spark Pool, Data Lake, Serverless SQL Pool

**1. Real-Time Processing Solution:** Real-time data ingestion and analytics for Event Hubs, where data is streamed continuously and requires immediate insights.

Steps	Technologies Used
<b>Data Ingestion</b>	<b>Azure Event Hub</b> streams real-time data from a device.
<b>Data Processing</b>	<b>Azure Stream Analytics</b> processes the streaming data in real time using SQL-like queries.
<b>Data Storage</b>	The processed data is stored in <b>Dedicated SQL Pool</b> for as the output sink in Stream Analytics to store processed data for real-time querying.

**2. Near Real-Time Processing Solution:** Data provided from a place where slight delays in processing (seconds to minutes) are acceptable for real-time user insights.

Steps	Technologies Used
<b>Data Ingestion</b>	Sample streams real-time data from <b>Event Hubs</b> .
<b>Data Processing</b>	<b>Cosmos DB</b> connected to Azure Synapse Pool using Synapse Link and used <b>Spark Pool</b> or <b>Synapse Dedicated SQL/Serverless Pool</b> for more complex data transformations.
<b>Data Storage</b>	<b>Synapse Dedicated SQL/Serverless SQL</b> Pool for reporting and analysis.

**3. Batch Processing Solution:** Enterprise dataset from a case where large datasets are ingested, processed, and analyzed periodically (e.g., daily reports).

Steps	Technologies Used
<b>Data Ingestion</b>	Data is ingested in batches from <b>Azure Data Lake</b> or external databases
<b>Data Processing</b>	<b>Azure Synapse Dedicated SQL Pool or Serverless Pool and Spark Pool</b> for large-scale processing and transformations.

<b>Data Storage</b>	Processed data is stored in <b>Synapse SQL Pool</b> for reporting, or in Data Lake for further analysis.
---------------------	--

<b>Aspect</b>	<b>Real-Time Processing</b>	<b>Near Realtime Processing</b>	<b>Batch Processing</b>
<b>Definition</b>	Processing data immediately as it arrives (milliseconds/seconds).	Processing data with minimal delay (seconds/minutes).	Processing large datasets after accumulation over time (hours/days).
<b>Azure Services</b>	Event Hub, Stream Analytics, Cosmos DB	Event Hub, Stream Analytics, Synapse Spark Pool, Cosmos DB	Synapse Dedicated SQL Pool, Synapse Spark Pool, Serverless SQL Pool
<b>Use Cases</b>	IoT, financial transactions, live data streaming	Social media analytics, log processing, monitoring systems	Data warehousing, ETL pipelines, periodic reports
<b>Data Ingestion</b>	Continuous streaming from Event Hub or IoT Hub	Streaming with slightly delayed processing	Batch data ingestion from data lakes, databases
<b>Data Processing</b>	Stream Analytics, real-time transformations	Stream Analytics, Spark Pool for slightly delayed processing	Spark Pool, SQL Pool for periodic data transformations
<b>Storage</b>	Cosmos DB, Synapse SQL Pools	Cosmos DB, Synapse SQL Pools	Synapse Dedicated SQL Pool, Data Lake

### Conclusion:

This comparative approach enables participants to understand the key differences between real-time, near real-time, and batch processing solutions using Azure services. By revisiting previous tasks and implementing Azure technologies, participants can develop robust data pipelines that meet various business needs. The final comparative document will serve as a reference for how Azure services can be used in different processing scenarios.