

**Question 4:** - This is an optional question

Megah Holdings Berhad is a diversified holding company in a few industries. Each industry has different best practices and different ERP implementation methods.

The revenue optimization team requires a dashboard which displays and analyses daily sales at the end of business day from three business units.

Describe a data pipeline with the following input sources:

**a. Retail company ERP - Real time API in XML**

- **Real-time data pipeline** - These data pipelines are optimized to process millions of events at scale, in real-time. Because of this, you can gather, analyse, and store large amounts of information. Streaming/real-time pipelines are useful when processing data from a streaming source, such as telemetry from connected devices or data from financial markets. When it is ingested in real-time, then it is ingested immediately as soon as it arrives

The process of gathering and transmitting data from source systems in real-time solutions such as Change Data Capture (CDC) is known as Real-Time Data Ingestion. This is one of the widely used Data Ingestion Types used especially in streaming services.

CDC continuously monitors transactions as well as redo logs and moves changed data without trying to interfere with database workload. Real-time ingestion is critical for time-sensitive use cases such as stock market trading or power grid tracking, where organizations must react quickly to new data.

Real-time Data Pipelines are also necessary for quickly making operational choices and defining and acting on new insights. In real-time data ingestion, as soon as data is generated, it is extracted, processed, and stored for real-time decision-making. For example, data obtained from a power grid must be continuously monitored to ensure power availability

**b. Chicken Broiler/Farm ERP - Hourly batch file generation in FTP server**

- **Batch data pipelines** – These pipelines are used to process data in batches. Batch processing is usually leveraged when a business wants to move high volumes of data at regular intervals. Batch processing jobs typically run on a fixed schedule (for instance, every 24 hours) or in some instances, once the volume of data reaches a given threshold. This process allows analysts to integrate large amounts of data and provide a reliable and fast decision model.

When it is ingested in batches using the Ingestion pipeline, it is ingested in some chunks at a periodic time interval.

With Batch-based Data Ingestion types, data can be collected by the ingestion layer based on simple schedules, trigger events, and any other logical ordering. When a company needs to collect specific data points daily or simply does not require data for real-time decision-making, batch-based ingestion is beneficial.

### c. Trading company ERP - Manual Excel files download

- **Cloud data pipeline** - These pipelines are designed to work only with cloud-based data destinations, sources, or both. They are hosted directly in the cloud. As such they enable businesses to save money on expert resources and infrastructure.

These pipelines are designed to work only with cloud-based data destinations, sources, or both. They are hosted directly in the cloud. As such they enable businesses to save money on expert resources and infrastructure.