Mini Capstone 2 alternative

For Azure:

In this project, we will use an eCommerce dataset to simulate the logs of user purchases, product views, cart history, and the user's journey on the online platform to create two analytical pipelines, Batch and Real-time.

The Batch processing will involve data ingestion, Data Lake architecture, processing, and visualization using Azure Stream Analytics, Azure Data Factory, Azure Data Lake Storage, and Power BI to draw insights regarding the following:
* Unique visitors per day
* During a certain time, the users add products to their carts but don't buy them
* Top categories per hour or weekday (i.e. to promote discounts based on trends)
* To know which brands need more marketing

The Real-time channel involves detecting Distributed Denial of Service (DDoS) and Bot attacks using Azure Functions, Azure Cosmos DB, Azure Monitor, and Azure Service Bus.

What we will do and learn as part of this project:
* How to stream data, by simulating real data producers, such as users browsing and buying products at the online store
    * Determine which products are being sold in near real-time
    * Main technology: Azure Stream Analytics and Python
* How to process incoming streams and create triggering mechanisms;
    * For example: If the same user_id views > 10 products/sec, send an immediate alert -> Possible BOT / DDoS attack
    * Main technology: Python, Apache Flink, Azure Data Lake Storage, Azure Service Bus, Azure Functions.
* How to process hourly data in a batch layer
    * For example:
        * New products every hour from our vendors, to be added to our inventory
        * Main technology: Azure Data Factory tasks, detecting new product files at our vendors' servers, syncing/replicating these to our data store

- Business Intelligence - Dashboards and visualizations
  Main Technologies:
    - Possibly Azure Monitor and Grafana for Real Time
    - Power BI for BI reporting

For GCP:

In this project, we will use an eCommerce dataset to simulate the logs of user purchases, product views, cart history, and the user's journey on the online

platform to create two analytical pipelines, Batch and Real-time.

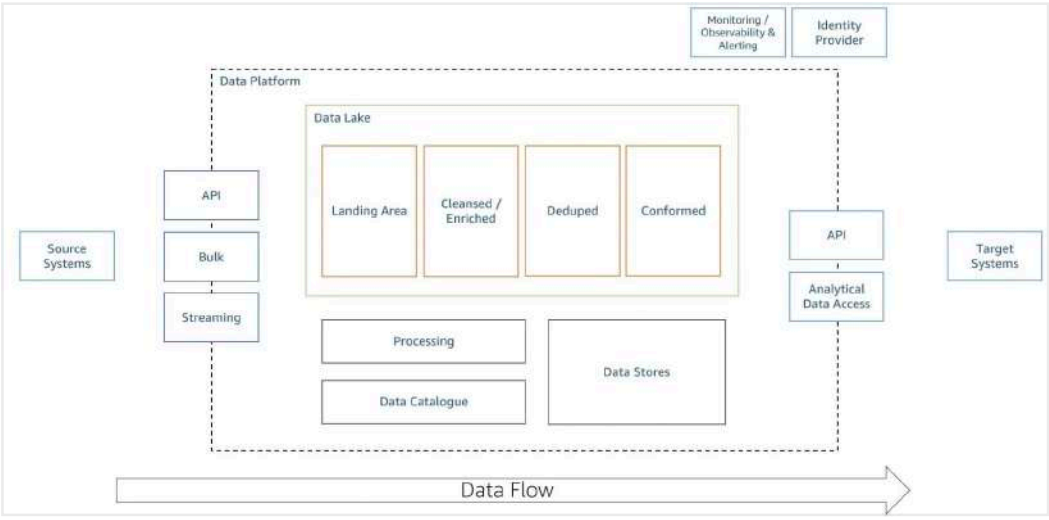
The Batch processing will involve data ingestion, Data Lake architecture, processing, and visualization using Google Cloud Dataflow, Google Cloud Storage, and Google Data Studio to draw insights regarding the following:

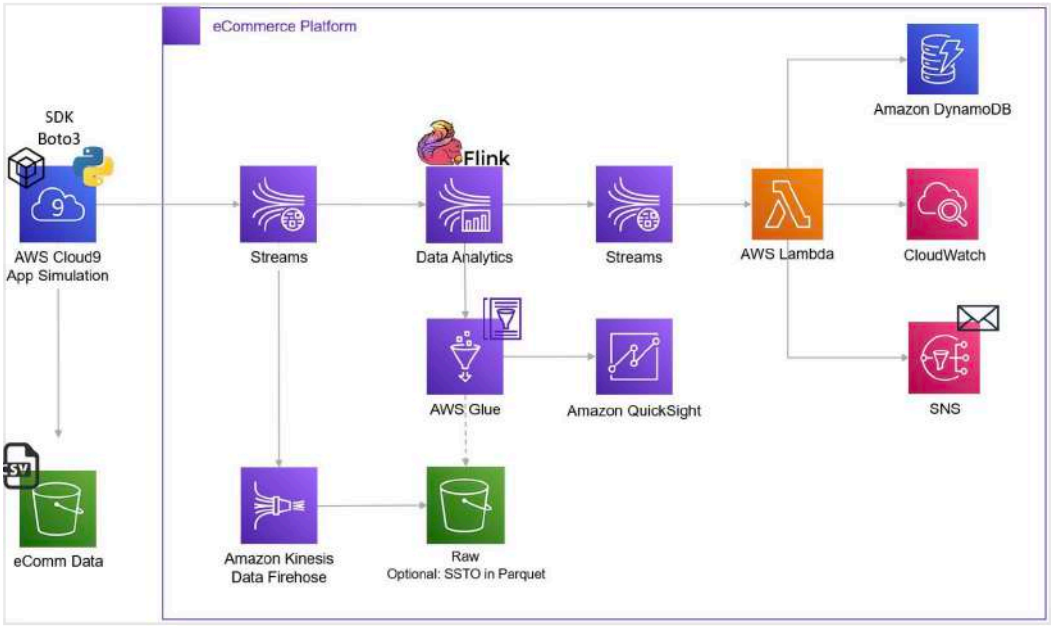
- Unique visitors per day
- During a certain time, the users add products to their carts but don't buy them
- Top categories per hour or weekday (i.e., to promote discounts based on trends)
- To know which brands need more marketing

The Real-time channel involves detecting Distributed Denial of Service (DDoS) and Bot attacks using Google Cloud Functions, Google Cloud Bigtable, Google Cloud Monitoring, and Google Cloud Pub/Sub.

What we will do and learn as part of this project:

- How to stream data, by simulating real data producers, such as users browsing and buying products at the online store
  - Determine which products are being sold in near real-time
  - Main technology: Google Cloud Dataflow and Python
- How to process incoming streams and create triggering mechanisms;
  - For example: If the same user_id views > 10 products/sec, send an immediate alert -> Possible BOT / DDoS attack
  - Main technology: Python, Apache Beam, Google Cloud Storage, Google Cloud Pub/Sub, Google Cloud Functions.
- How to process hourly data in a batch layer
  - For example: New products every hour from our vendors, to be added to our inventory
  - Main technology: Google Cloud Dataflow jobs, detecting new product files at our vendors' servers, syncing/replicating these to our data store
- Business Intelligence - Dashboards and visualizations
  - Main Technologies:
  - Possibly Google Cloud Monitoring and Grafana for Real Time
  - Google Data Studio for BI reporting





For AWS Kinesis Streams:

Azure: Azure Stream Analytics
GCP: Google Cloud Pub/Sub

For AWS Kinesis Data Firehose:

Azure: Azure Data Factory with Data Flow
GCP: Google Cloud Dataflow

For AWS Kinesis Data Analytics:

Azure: Azure Stream Analytics with Real-time Analytics
GCP: Google Cloud Dataflow or Google Cloud Data Analytics

For AWS Glue:

Azure: Azure Data Factory
GCP: Google Cloud Dataflow or Google Cloud Dataprep

For AWS S3 (Simple Storage Service):

Azure: Azure Blob Storage
GCP: Google Cloud Storage

For AWS Quicksight:

Azure: Power BI
GCP: Google Data Studio

For AWS Lambda:

Azure: Azure Functions
GCP: Google Cloud Functions

For AWS DynamoDB:

Azure: Azure Cosmos DB
GCP: Google Cloud Bigtable

For AWS CloudWatch:

Azure: Azure Monitor
GCP: Google Cloud Monitoring

For AWS SNS (Simple Notification Service):

Azure: Azure Service Bus or Azure Notification Hubs
GCP: Google Cloud Pub/Sub