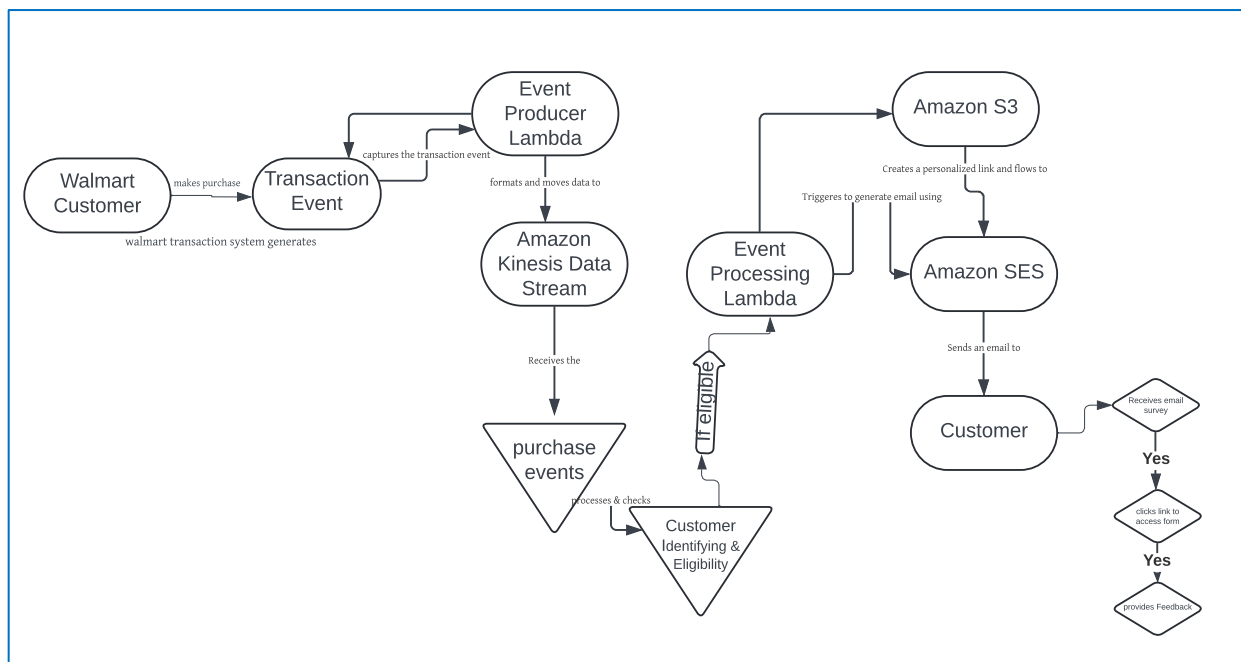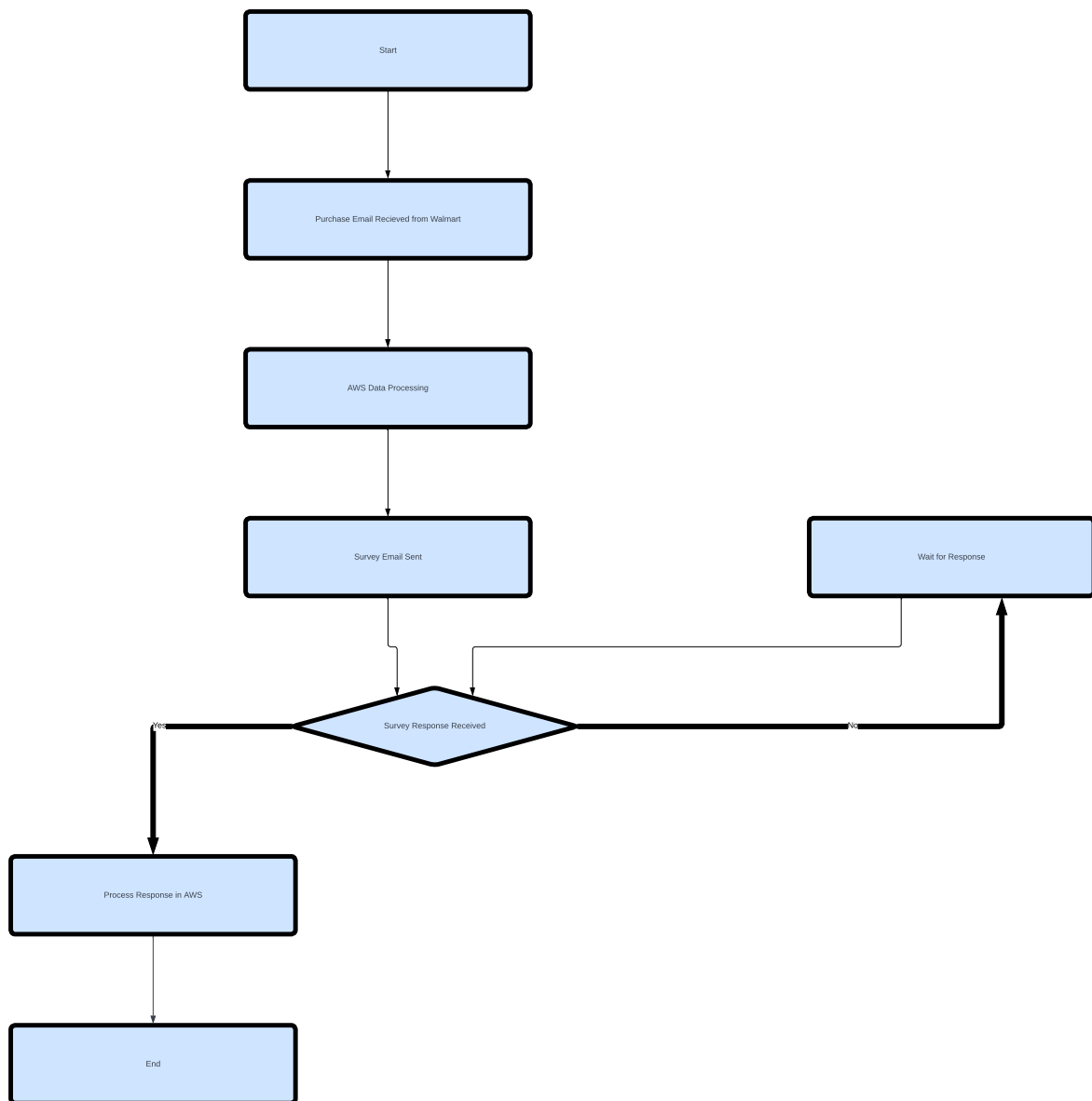# Assignment: AWS

1.)Flow:

1. A customer makes a purchase at a Walmart store, and the Walmart Transaction System generates a transaction event.

2. The Event Producer Lambda captures the transaction event and formats the relevant data.

3. Formatted data is then published to the Amazon Kinesis Data Stream.

4. The Event Processing Lambda subscribes to the Kinesis Data Stream and receives the purchase events.

5. It processes the purchase events, including identifying the customer and determining if they are eligible for an email survey.

6. If eligible, the Event Processing Lambda triggers the generation of an email survey using Amazon SES. The survey includes a personalized link to the survey form hosted on Amazon S3.

7. Amazon SES sends the email survey to the customer's email address.

8. The customer receives the email survey, clicks on the link to access the survey form, and provides feedback.
This architecture provides a scalable and efficient solution for sending email surveys to Walmart customers based on their purchase transactions, utilizing various AWS services for data processing, storage, and email delivery.

```
┌─────────────────────┐
│        Start        │
└─────────────────────┘
           │
           ▼
┌─────────────────────┐
│ Purchase Email      │
│ Recieved from Walmart│
└─────────────────────┘
           │
           ▼
┌─────────────────────┐
│  AWS Data Processing │
└─────────────────────┘
           │
           ▼
┌─────────────────────┐          ┌─────────────────────┐
│  Survey Email Sent  │          │  Wait for Response  │
└─────────────────────┘          └─────────────────────┘
           │                              │
           ▼                              ▼
        ◇ Survey Response Received ◇
   Yes ◀─────────────────────────── No ──▶
           │
           ▼
┌─────────────────────┐
│ Process Response in AWS│
└─────────────────────┘
           │
           ▼
┌─────────────────────┐
│         End         │
└─────────────────────┘
```
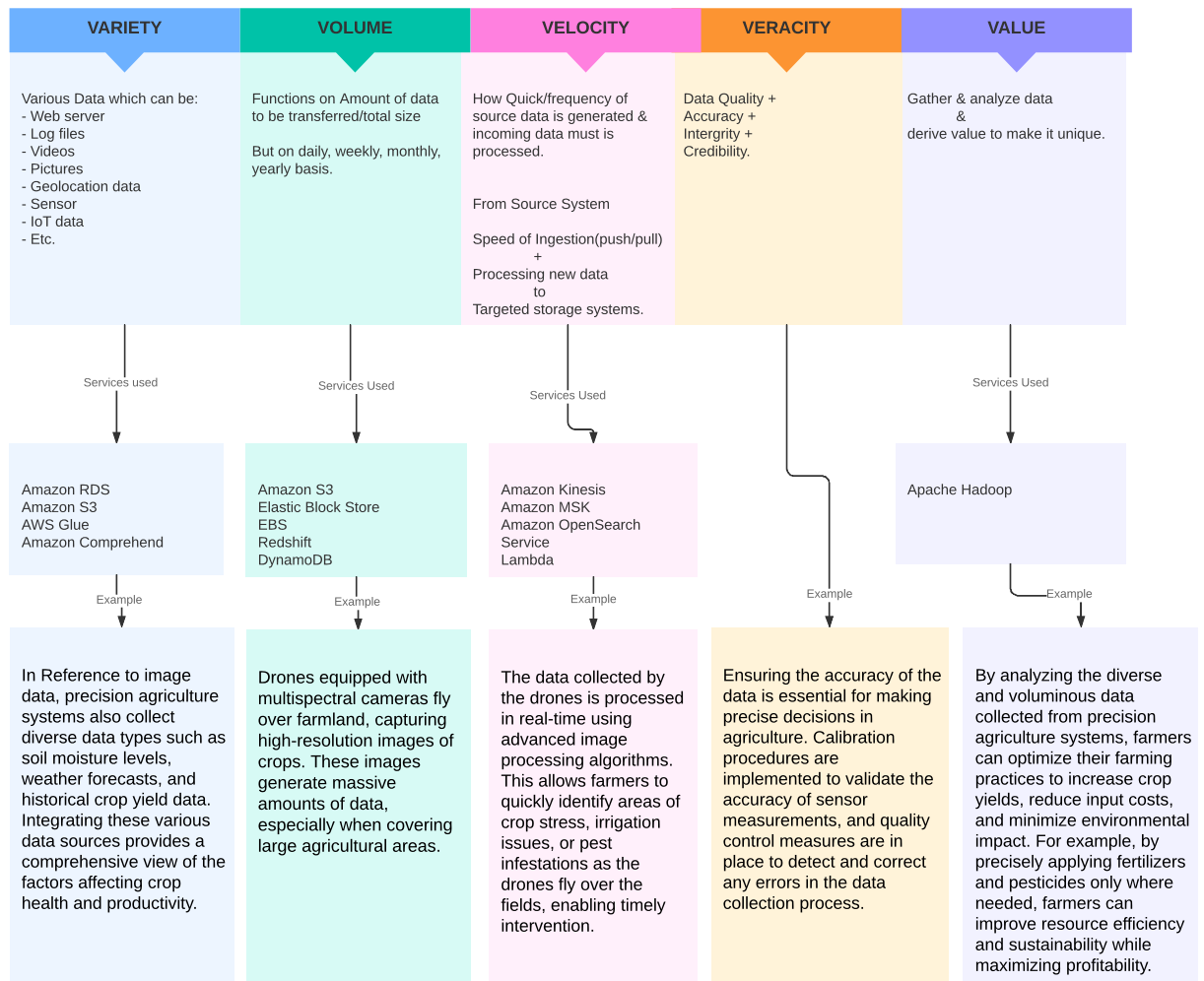
2.)

- Data Realm
- Data Ops
- Data Mesh
- Data Lineage
- Monolith Service Architecture
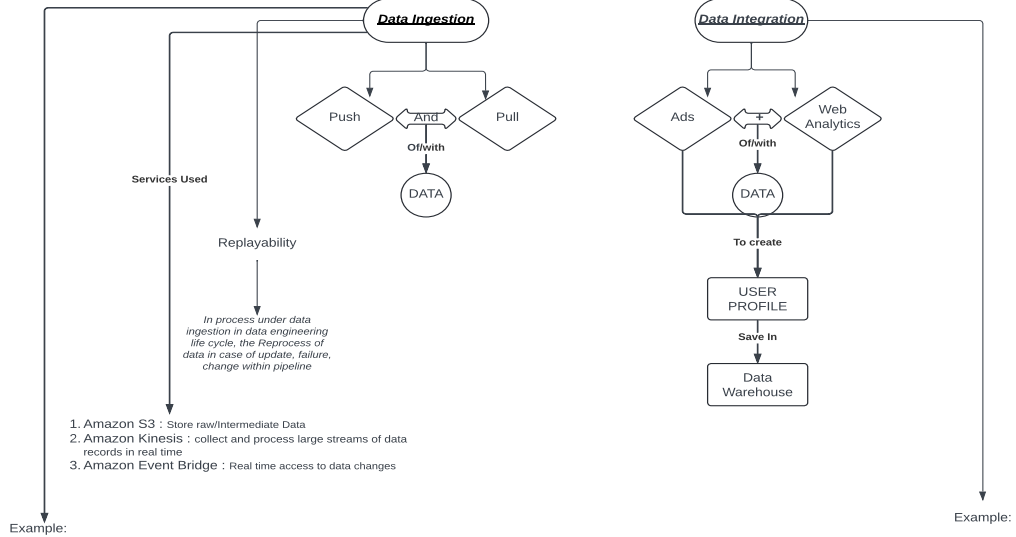- Data Ingestion
- Usage of Aws Services tools

3.)
5 V's of Data

5V'S Of DATA

VARIETY	VOLUME	VELOCITY	VERACITY	VALUE
Various Data which can be: - Web server - Log files - Videos - Pictures - Geolocation data - Sensor - IoT data - Etc.	Functions on Amount of data to be transferred/total size But on daily, weekly, monthly, yearly basis.	How Quick/frequency of source data is generated & incoming data must is processed. From Source System Speed of Ingestion(push/pull) + Processing new data to Targeted storage systems.	Data Quality + Accuracy + Intergrity + Credibility.	Gather & analyze data & derive value to make it unique.
Services used	*Services Used*	*Services Used*		*Services Used*
Amazon RDS Amazon S3 AWS Glue Amazon Comprehend	Amazon S3 Elastic Block Store EBS Redshift DynamoDB	Amazon Kinesis Amazon MSK Amazon OpenSearch Service Lambda		Apache Hadoop
Example	*Example*	*Example*	*Example*	*Example*
In Reference to image data, precision agriculture systems also collect diverse data types such as soil moisture levels, weather forecasts, and historical crop yield data. Integrating these various data sources provides a comprehensive view of the factors affecting crop health and productivity.	Drones equipped with multispectral cameras fly over farmland, capturing high-resolution images of crops. These images generate massive amounts of data, especially when covering large agricultural areas.	The data collected by the drones is processed in real-time using advanced image processing algorithms. This allows farmers to quickly identify areas of crop stress, irrigation issues, or pest infestations as the drones fly over the fields, enabling timely intervention.	Ensuring the accuracy of the data is essential for making precise decisions in agriculture. Calibration procedures are implemented to validate the accuracy of sensor measurements, and quality control measures are in place to detect and correct any errors in the data collection process.	By analyzing the diverse and voluminous data collected from precision agriculture systems, farmers can optimize their farming practices to increase crop yields, reduce input costs, and minimize environmental impact. For example, by precisely applying fertilizers and pesticides only where needed, farmers can improve resource efficiency and sustainability while maximizing profitability.

4.)

Moving data from one place ti another. Point
1 to Point 2

Combining Data from different sources
into new dataset

Data Ingestion

Data Integration

Push

And

Pull

Of/with

DATA

Ads

+

Web
Analytics

Of/with

DATA

To create

USER
PROFILE

Save In

Data
Warehouse

Services Used

Replayability

*In process under data
ingestion in data engineering
life cycle, the Reprocess of
data in case of update, failure,
change within pipeline*

1. Amazon S3 : Store raw/Intermediate Data
2. Amazon Kinesis : collect and process large streams of data
records in real time
3. Amazon Event Bridge : Real time access to data changes

Example:

Example:

- Data ingestion involves collecting real-time data on your music listening behavior.
For example, every time you play a song, skip a track, or add a song to your
playlist, this data is ingested into the system.
- The app may gather data on the time of day you listen to music, the genre of
music you prefer, and the artists you frequently listen to.

- Data integration combines this real-time listening data with other relevant
information to generate personalized recommendations and playlists tailored to
your tastes.
-System integrates your music listening history with data on trending songs,
popular playlists, and new releases. It also takes into account factors like your
location and the weather to offer contextually relevant music suggestions.

data ingestion involves capturing real-time data on your music listening behavior, while data
integration combines this data with contextual information to offer personalized
recommendations and enhance your listening experience.