PAPER ROOT STUDIOS

GROUP 1

CLOUD ARCHITECTURE PROJECT

Presented By: - Dharmil Patel

COMPANY OPERATION

Paper Root Studios is a paper manufacturing company that is exclusively based in Canada. They currently store their data on-premises but are exploring the possibility of migrating to the cloud for improved accessibility and scalability.

Their data sources are:

E-commerce transactions

Geo-location file

Sales data

Website tracking

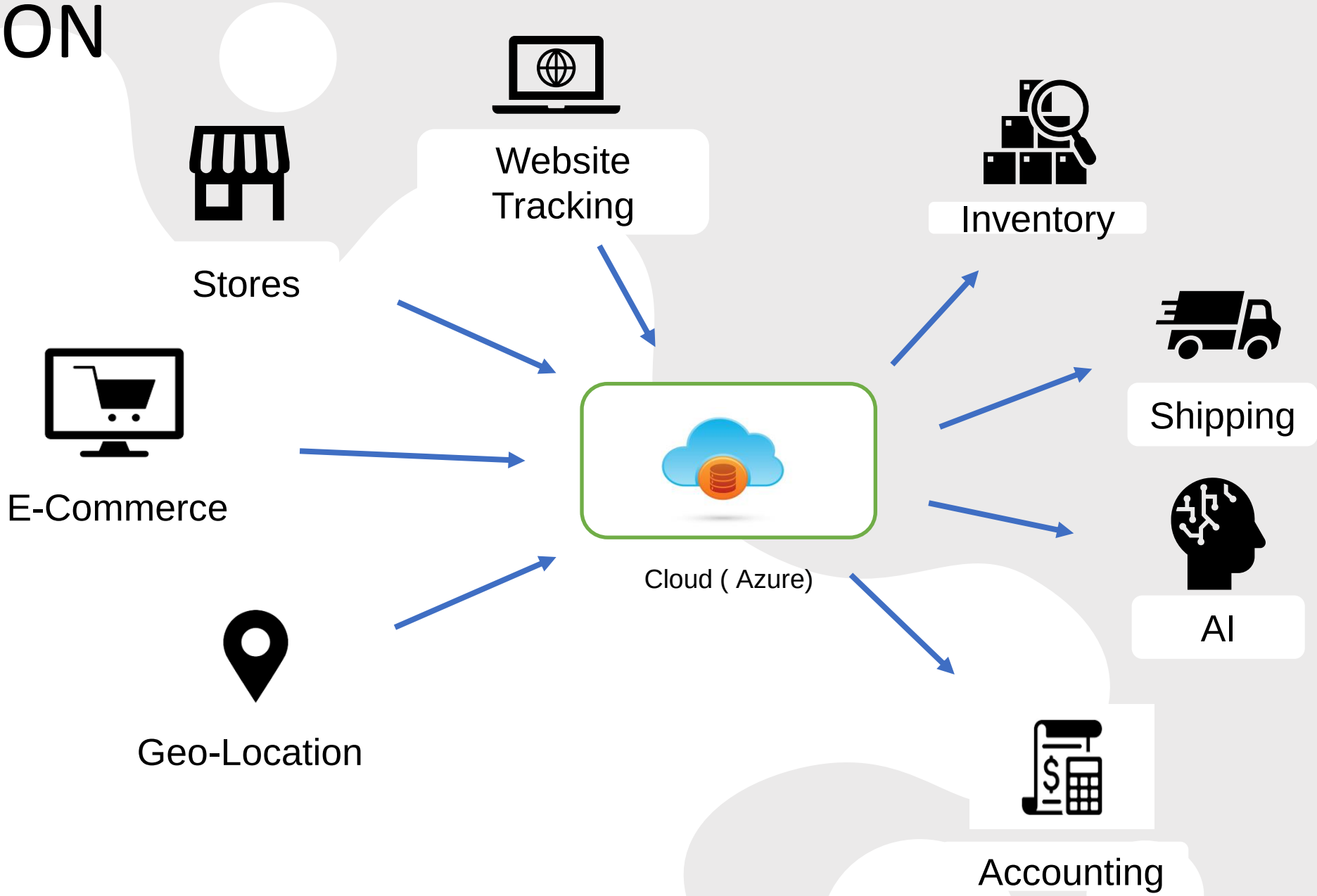
PROBLEM

- Disaster Vulnerability: On-premises data carries the risk of no accessibility during service outages.
- Limited Data Visibility: Managing on-premises data presents challenges in control, potentially resulting in security risks and compliance issues.
- Current Storage Is Not Cost-effective
- No Scalability

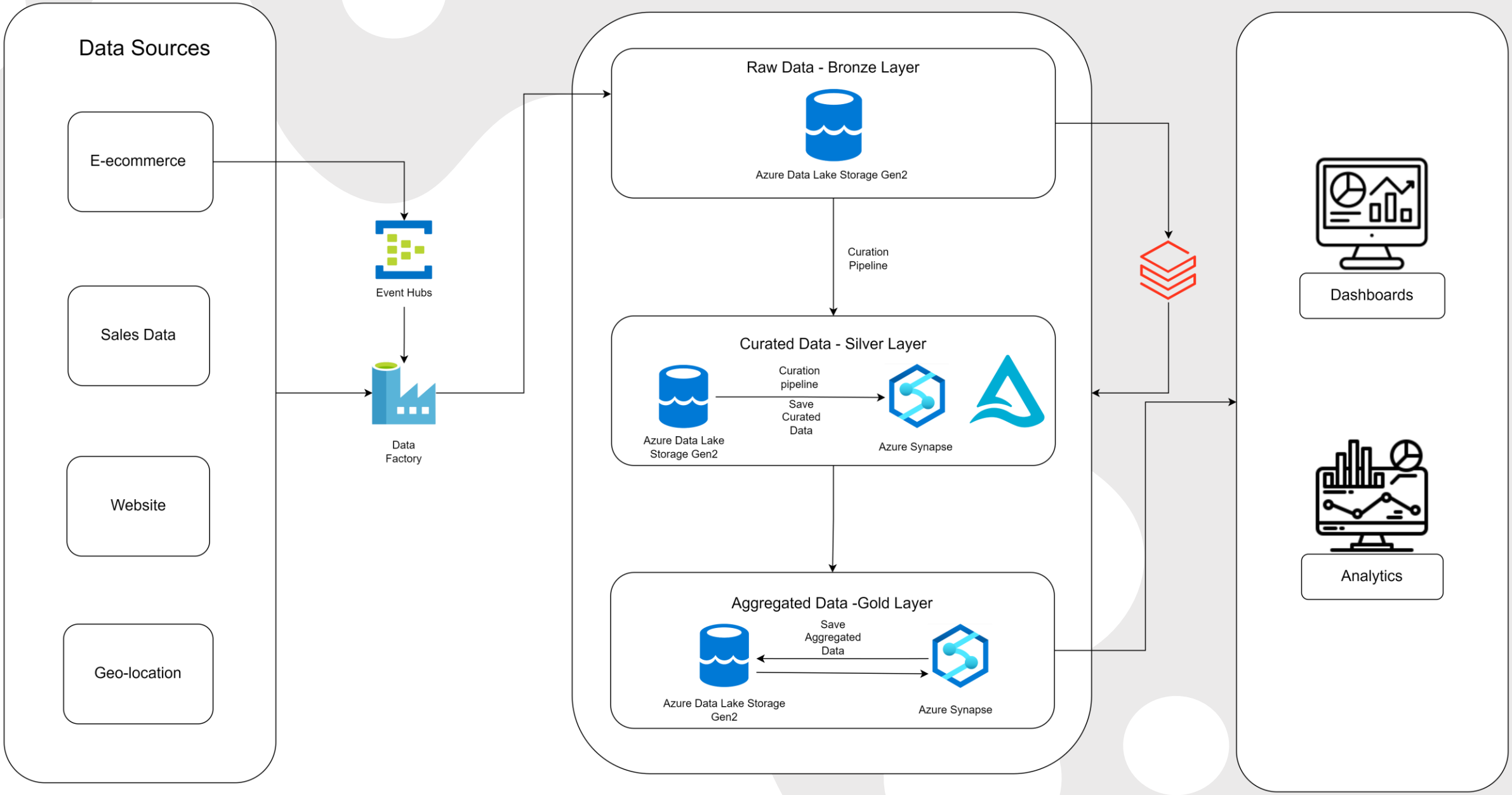
OBJECTIVES

- Migrate Sales data to Cloud (Azure) to make it,
  - Scalable
  - Cost-Efficient
  - Enable real-time insights and decision-making
  - Seamless online transaction
  - Unified storage system
  - Secure

VISION


Website Tracking
Stores
Inventory
Shipping
E-Commerce
Cloud (Azure)
AI
Geo-Location
Accounting

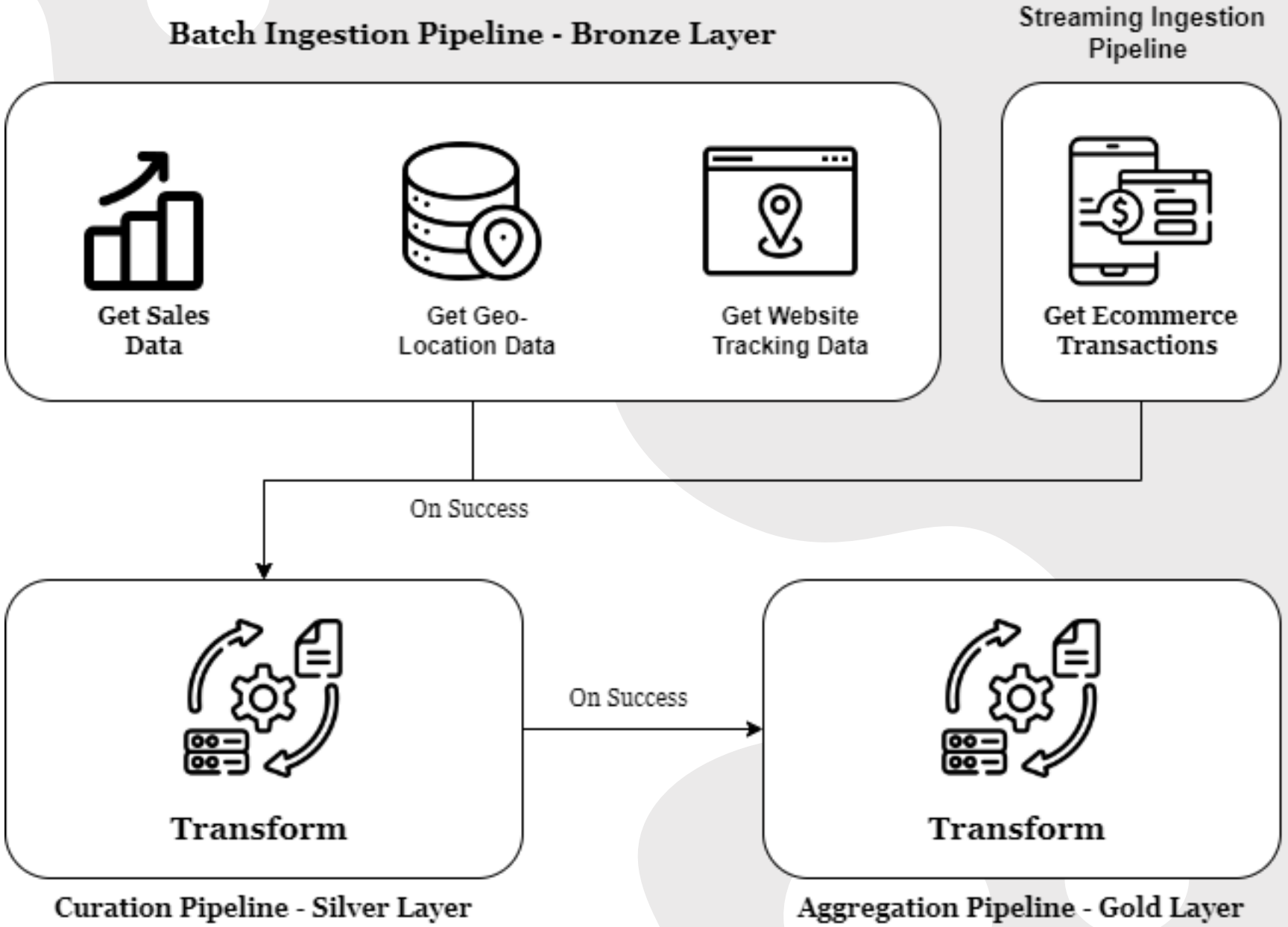
PROPOSED ARCHITECTURE



PIPELINE STRATEGY:

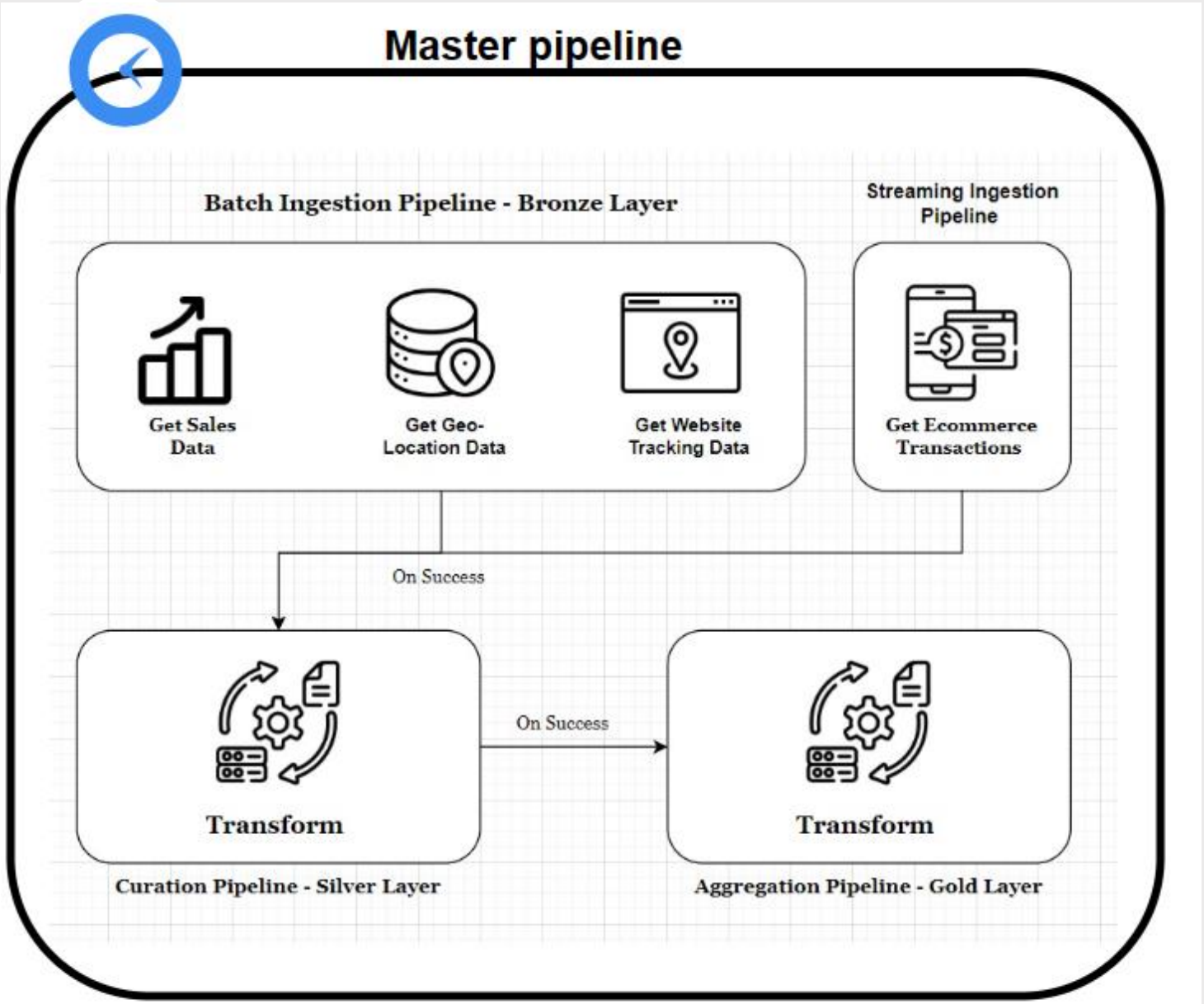
- The ingestion, curation and gold pipeline will run as one complete unit and each pipeline will work by invoking the ingestion pipeline, pulling data from our sources, then we will invoke the curation pipeline which will clean our data and merge it in the Delta lake tables. Finally, the aggregation pipeline summarizes data for our end users.

THE DEVELOPMENT



APPROACH

- For our company we will be deploying our pipeline, using the event-based approach by creating a master pipeline that controls the run of other pipelines based on their success and failure, only after the prior pipeline is completed does the next start, with little or no risk of overlapping. Once we have all three pipelines working in Azure Data Factory, we schedule the master pipeline for automated execution for every one hour start date and time recurrence.


Master pipeline
Batch Ingestion Pipeline - Bronze Layer
Streaming Ingestion Pipeline
Get Sales Data
Get Geo-Location Data
Get Website Tracking Data
Get Ecommerce Transactions
On Success
Transform
On Success
Transform
Curation Pipeline - Silver Layer
Aggregation Pipeline - Gold Layer

STRATEGY FOR DEALING WITH PIPELINE FAILURE

- Setting a safe timeout value
- Retry interval: Setting a higher value for retries covers against transient failures that are, typically, related to network glitches and unavailability.
- If the hourly triggered pipeline failed to run due to database unavailability, we can decide to bypass the failed activity ,terminate the entire pipeline or let the entire pipeline fail and rerun after seeking customer’s approval
- Set alert mechanisms to enable quick reaction to failures

ADDRESSING CHANGE IN THE LAKEHOUSE

SCHEMA EVOLUTION

- It is very likely that our schema for incoming file or table changes. What do we do if this happens? We will use Delta Lake because it safeguards our data by enforcing schema validation and adjusting the schema without the need for table versioning or rewriting prior data with the dummy column. It automatically compares the schema of the incoming data against prior data. If a discrepancy is detected in Delta Lake, an exception is raised by default.

DATA SHARING

- While data sharing enables organizations to realize revenue diversification using data monetization, there are problems associated with data sharing methods like emails and hard drives such as insecurity, lack of control and tracking. So, we have a solution, we will employ our organization to share data using AZURE DATA SHARE.

CONCLUSION

- By implementing the cloud architecture plan we proposed for Paper Root Studios, their online transactions will be smooth and hassle-free. This will not only help them save costs but also ensure that their data is secure. Additionally, the plan will provide them with actionable insights, which will enable them to make informed business decisions and achieve their goals more efficiently.

REFERENCES

- Draw.io
- Microsoft
- Sait slides

THANK YOU!