**YOULearn.com(Your Online University)**

Cloud Data Architecture





**Introduction**

Modern e-learning platforms generate and rely on massive volumes of data to operate effectively. At YOULearn.com, information such as user activity, course materials, financial transactions, and student engagement is constantly being collected and accessed. To handle this efficiently, a robust cloud-based data infrastructure is essential for streamlined storage, fast access, and deep analysis.

This case study presents a scalable cloud solution that brings together diverse data sources, processes them through advanced pipelines, and delivers actionable insights. The primary aim is to boost system efficiency, strengthen data security, and deliver meaningful analytics for students, educators, and administrators.

**Why Choose YOULearn?**

YOULearn is an easy-to-use and powerful online learning platform. It helps students learn better by giving them course suggestions, tracking their progress, and showing clear reports. Teachers and admins also get helpful tools to manage classes and understand how students are doing. YOU Learn works fast, keeps data safe, and can support many users at the same time. It’s built using smart cloud technology, so it’s reliable and ready to grow as more people use it.

**OBJECTIVE**

The YOULearn system was built to manage data in a smart and secure way. It brings all learning data into one place, cleans it, and turns it into useful insights. The goal is to help students, teachers, and admins get real-time information, improve platform speed, and support better decisions through clear reports and dashboards.

**Cloud Architecture for YOULearn:**

* YOULearns cloud architecture uses a Bronze-Silver-Gold model:
* **Bronze**: Stores raw data from sources like user activity and payments.
* **Silver**: Cleans and organizes the data for analysis.
* **Gold**: Contains ready-to-use data for dashboards and reports.

**DESIGN PHASE OF CLOUD ARCHITECTURE**

A screenshot of a computer

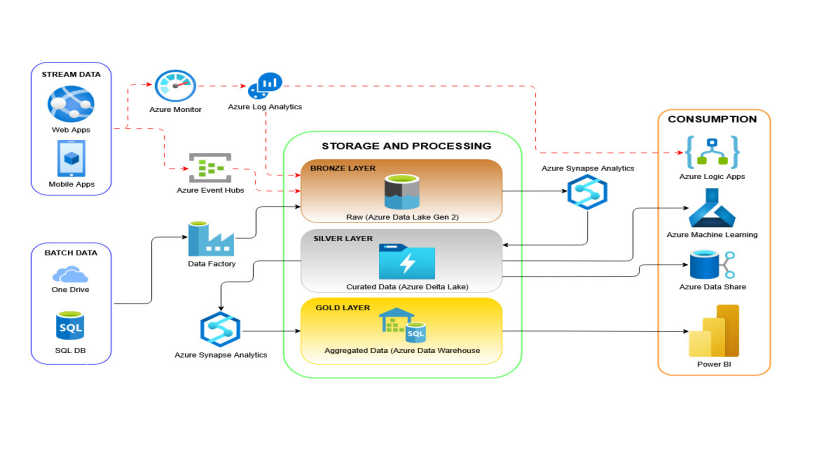
Description automatically generated

**Database Inputs**

1. **User Database**
   * **Description**: Profiles, enrolments, progress
   * **Location**: AWS RDS (PostgreSQL/MySQL)
   * **Type**: Structured (SQL)
2. **Course Content Storage**
   * **Description**: Videos, documents, quizzes
   * **Location**: AWS S3 / Azure Blob
   * **Type**: Unstructured (Multimedia)
3. **Payment Transactions**
   * **Description**: Subscriptions, refunds, invoices
   * **Location**: AWS RDS (with Stripe/PayPal)
   * **Type**: Structured (SQL)
4. **User Engagement Data**
   * **Description**: Completion rates, quiz scores, time spent
   * **Location**: Google Big Query / AWS Redshift
   * **Type**: Structured (SQL)
5. **System Logs**
   * **Description**: Events, errors, performance
   * **Location**: AWS CloudWatch / ELK Stack
   * **Type**: Semi-structured (Logs)

**Database Outputs**

1. **Student Reports & Certifications**
   * **Delivery**: Auto-generated dashboards and certificates
   * **Users**: Students, Instructors
2. **Course Recommendations**
   * **Delivery**: AI-based suggestions
   * **Users**: Students
3. **Revenue Reports**
   * **Delivery**: Financial dashboards
   * **Users**: Business & Finance Teams
4. **Content Performance Metrics**
   * **Delivery**: Analytics dashboards
   * **Users**: Instructors, Content Teams
5. **Instructor Dashboards**
   * **Delivery**: Web-based insights
   * **Users**: Instructors
6. **Marketing Insights**
   * **Delivery**: AI-driven trend reports
   * **Users**: Marketing Teams
7. **System Health Reports**
   * **Delivery**: Real-time monitoring
   * **Users**: IT & DevOps



1. **Data Sources:**

* **Stream Data**: From Web & Mobile Apps via Azure Event Hubs.
* **Batch Data**: From OneDrive, SQL Databases via Data Factory.

1. **Storage & Processing Layers:**

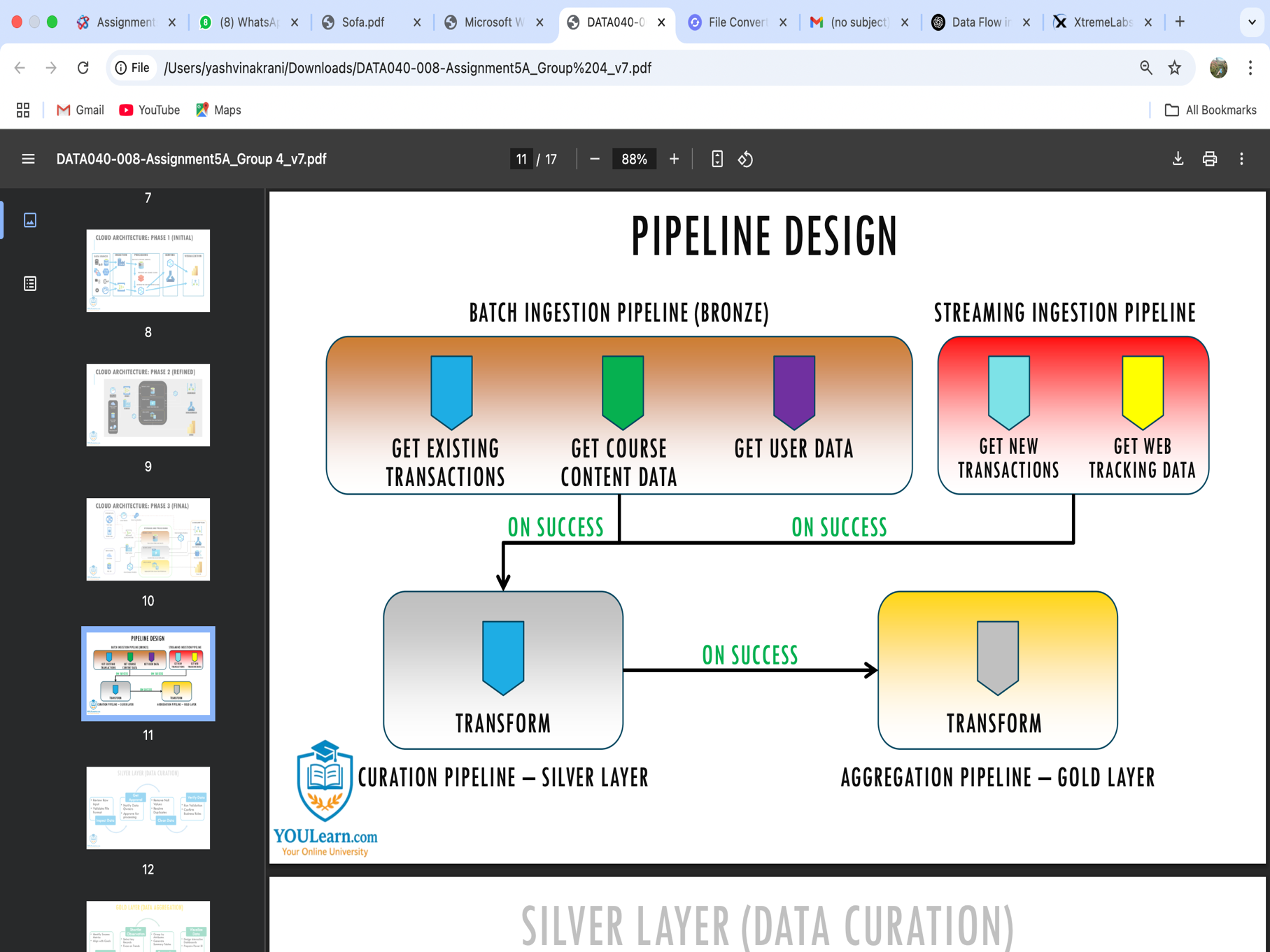
* **Bronze Layer**: Stores raw data in Azure Data Lake Gen2.
* **Silver Layer**: Contains cleaned, curated data in Azure Delta Lake.
* **Gold Layer**: Holds aggregated, high-value data in Azure Data Warehouse.

1. **Analytics & Monitoring:**

* Azure Synapse Analytics and Azure Log Analytics process data across all layers.
* Azure Monitor tracks data flow and system health.

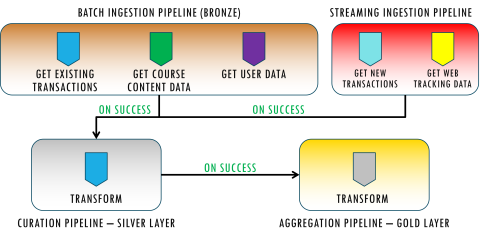
1. **Data Consumption:**

* Tools like Power BI, Azure Machine Learning, Logic Apps, and Azure Data Share use the processed data for reporting, AI modelling, automation, and sharing.



1. **Bronze Layer (Batch Ingestion)**
   * Collects raw data from sources like:
     + Existing transactions
     + Course content
     + User data
2. **Streaming Ingestion Pipeline**
   * Ingests real-time data such as:
     + New transactions
     + Web tracking data
3. **Silver Layer (Curation Pipeline)**
   * Transforms and cleans data from the Bronze and streaming pipelines.
4. **Gold Layer (Aggregation Pipeline)**
   * Further transforms curated data into high-value, aggregated datasets for analytics and reporting.

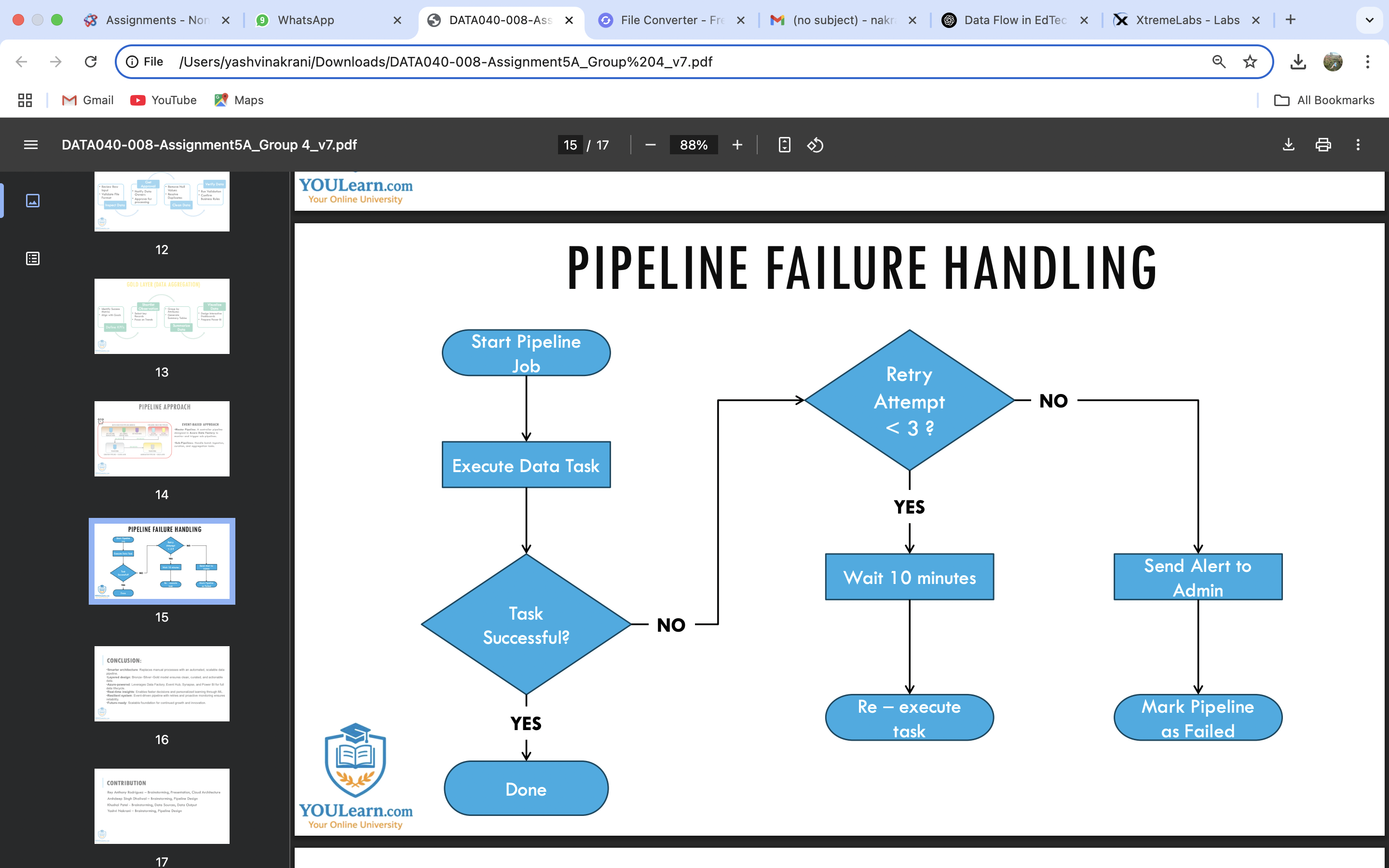
**PIPELINE APPROACH**



**Data Processing Pipeline Overview**

This architecture outlines a layered data pipeline that processes both batch and streaming data using the Bronze, Silver, and Gold layer model:

* **Bronze Layer (Batch & Streaming Ingestion)**:
  + Collects raw data from multiple sources:
    - **Batch**: Transactions, course content, user data
    - **Streaming**: Real-time transactions and web tracking
* **Silver Layer (Curation)**:
  + Transforms and cleans the raw data into a structured format for analysis.
* **Gold Layer (Aggregation)**:
  + Aggregates curated data into refined datasets used for dashboards, reports, and business intelligence.



**PIPELINE FAILURE HANDLING**

When a data pipeline runs, it tries to complete a task.

* If the task works, it finishes successfully.
* If the task fails, the system checks how many times it has tried:
  + If it's tried less than 3 times, it waits 10 minutes and then tries again.
  + If it's already tried 3 times, it sends a message to the admin and stops the pipeline.

This helps the system fix small issues on its own and notifies someone if it needs help.

**CONCLUSION:**

The new cloud data architecture at YOULearn.com replaces manual work with an automated system that is fast, reliable, and scalable. By using the Bronze-Silver-Gold model, the platform can clean, organize, and analyze data more effectively.

With tools like Azure Data Factory, Event Hub, and Power BI, the system now supports real-time insights, smarter decisions, and a better learning experience. It’s built to grow, adapt, and support future innovations in online education.