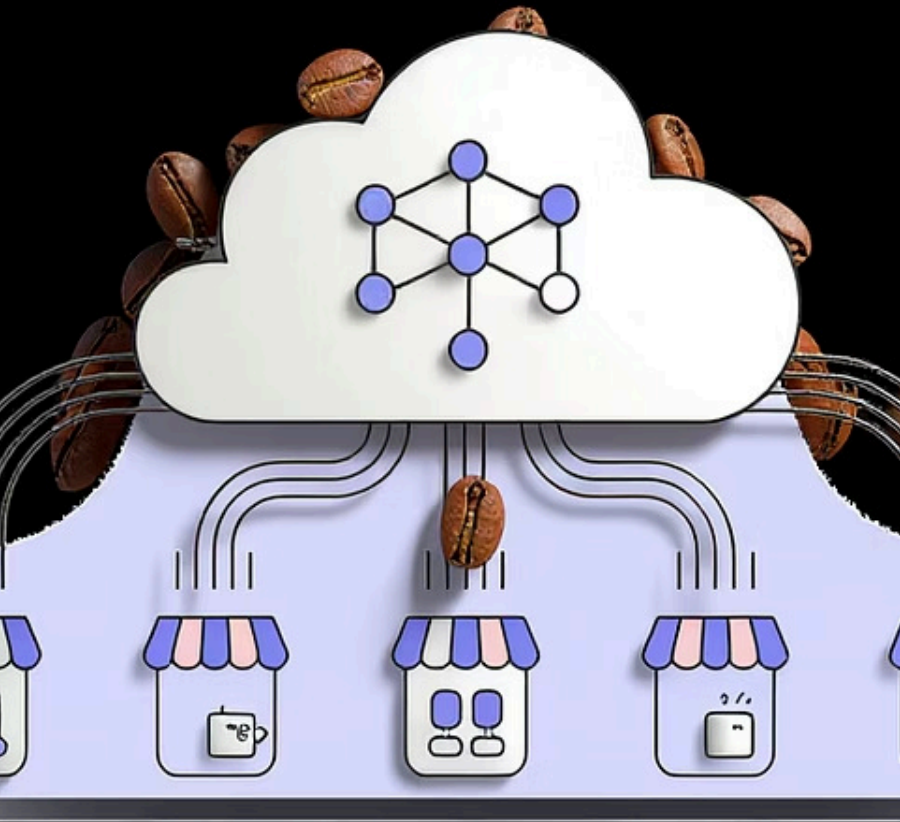Ana-LatteX

Ana-LatteX – Cloud ETL & Monitoring Update

Building a scalable, cloud-based ETL pipeline for centralised analytics and operational monitoring across all café branches
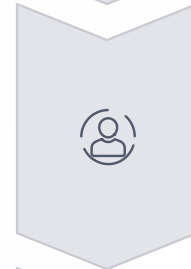
Project Vision

Creating a comprehensive data pipeline to transform branch operations into actionable insights:

Branch Data Collection

CSVs from each branch uploaded nightly to **S3** buckets

ETL Processing

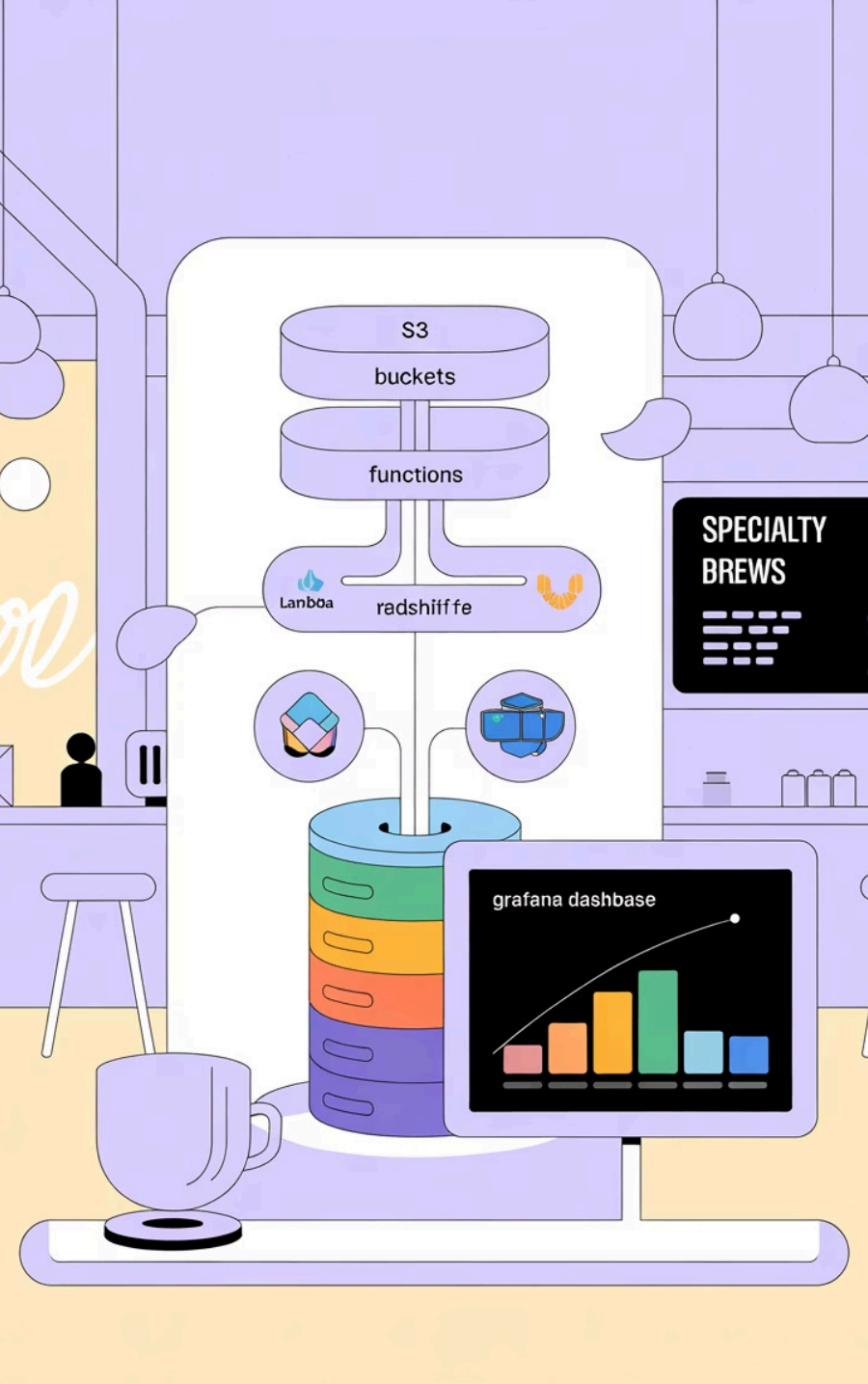**Lambda** functions process data through automated ETL pipeline

Centralised Storage

Transformed data stored in **Redshift** for analytics and querying

Visualisation

Grafana dashboards display business and operational metrics

Sprint 2 Goals: ETL in Cloud

1

Infrastructure as Code

Deploy cloud infrastructure using **CloudFormation** for consistent, repeatable deployments

2

Serverless ETL

Migrate local ETL code to **AWS Lambda** for scalable, event-driven processing

3

Automated Triggers

Configure Lambda to trigger automatically on **S3 CSV uploads** from branches

4

Data Warehouse Integration

Connect Lambda to **Redshift** for centralised data storage and analytics

5

Maintain PoC Pipeline

Keep **local PoC pipeline** independent for debugging and development

Sprint 2 Deliverables: Ticket Highlights

Ticket	Deliverable	Status / Impact
1	Deployment bucket stack *(ana-lattex-deployment-bucket)*	Completed - Automates Lambda & ETL deployments
2	Lambda triggered by S3	Completed - Processes CSVs automatically; logs row count
3	ETL CloudFormation stack *(ana-lattex-etl-stack)*	Completed - Fully IaC deployment; prints "hello team"
4	Redshift Integration	Completed - Lambda connects securely in private subnet; local POC intact

Sprint 2 Demo: Key Proof Points

CSV Upload Trigger

Demonstrate automated Lambda activation when sample CSV is uploaded to S3 bucket
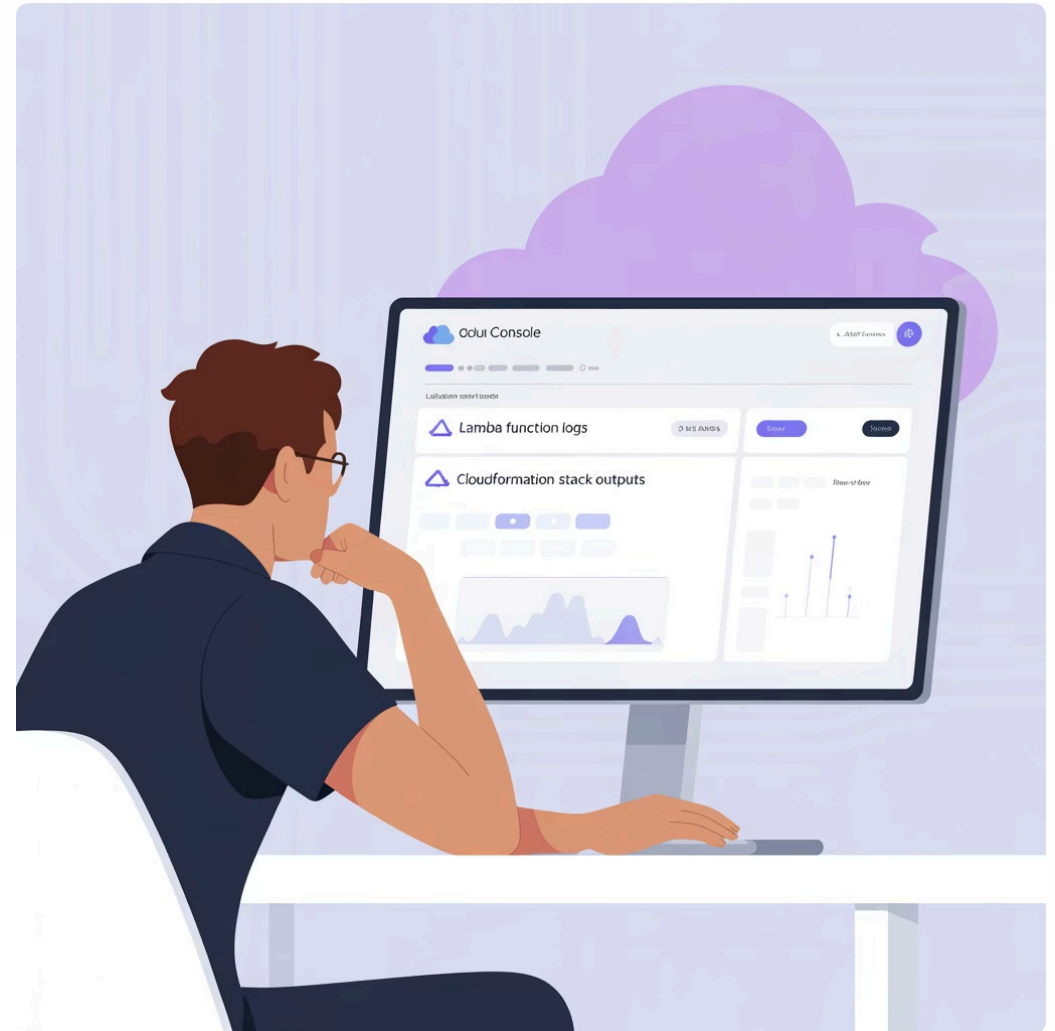
Row Count Logging

Show CloudWatch logs confirming accurate processing of uploaded data
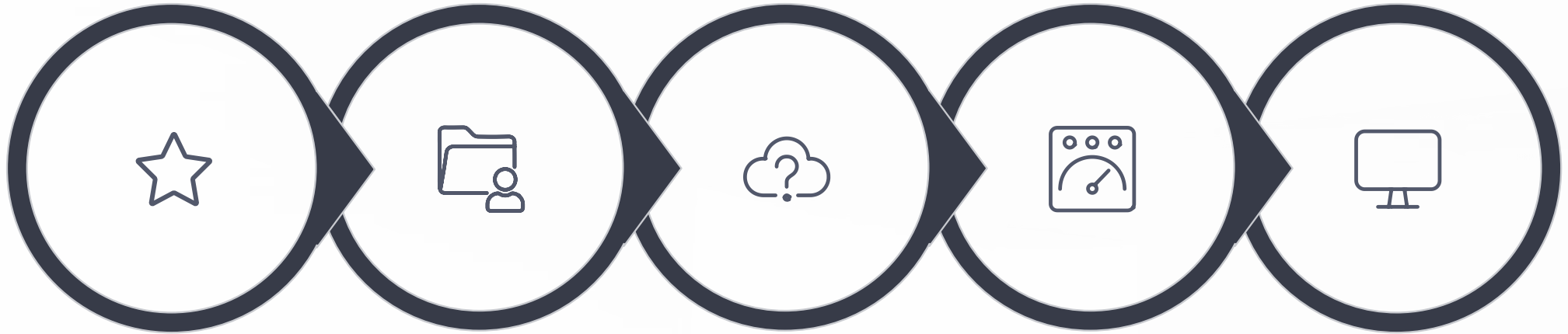
CloudFormation Stacks

Review ETL infrastructure deployed through IaC templates

Local PoC Validation

Confirm local proof-of-concept pipeline remains functional

Sprint 3 Goals: Monitoring & Visualisation

Deploy Grafana

Connect to Redshift

CloudWatch Integration

Business Dashboards

Operational Dashboards

- Deploy **Grafana** on EC2 using CloudFormation for consistent, repeatable setup

- Establish secure connections to both **Redshift** and **CloudWatch** data sources

- Implement version-controlled dashboards for comprehensive monitoring

Sprint 3 Ticket Highlights

Ticket 1: EC2 + Grafana Setup

User Story: Developers can visualise application and data performance

Impact: Secure login/password management; consistent environment

Ticket 2: Data Source Connections

User Story: Data analysts can explore metrics from multiple sources

Impact: Dashboards backed up in GitHub; version-controlled

Ticket 3: Business Dashboards

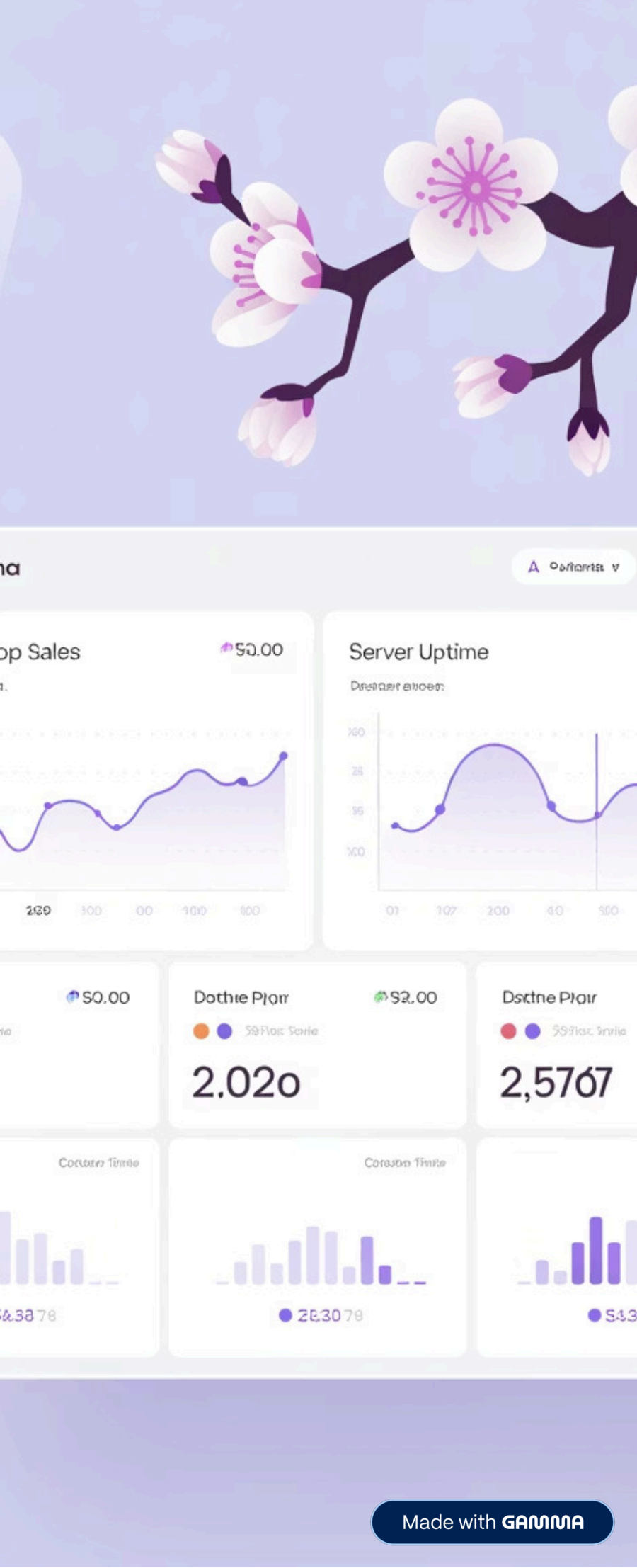**User Story:** Track products sold per day, revenue per branch

Impact: Data-driven decisions; branch performance comparison

Ticket 4: Operational Dashboards

User Story: Monitor Lambda & EC2 performance metrics

Impact: Proactive maintenance; reduced downtime

Technical Highlights

Cloud-Ready ETL Architecture

Scalable pipeline capable of handling multiple branches with consistent processing

Infrastructure as Code

CloudFormation deployment for Lambda, S3, EC2, Grafana ensures reproducibility

Event-Driven Automation

S3 → Lambda → ETL → Redshift pipeline triggered automatically by data uploads

Secure Access Controls

Redshift & Grafana access managed via SSM & Password Manager

Development Continuity

Local PoC pipeline & unit tests remain unaffected by cloud migration

Version-Controlled Visualisations

Grafana dashboards exportable & stored in GitHub repository

Benefits to Business

Centralised Insights

All branch data consolidated in Redshift for comprehensive analysis and reporting

Operational Visibility

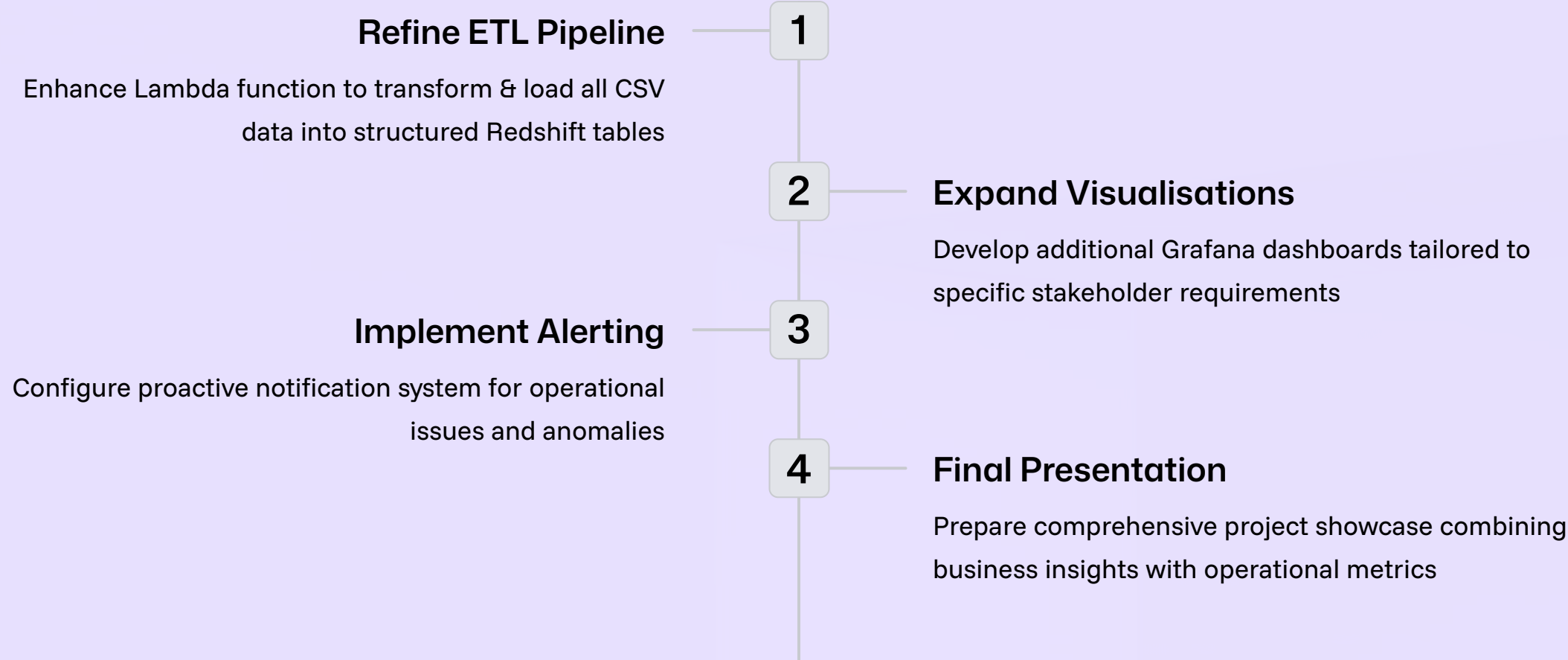Real-time monitoring of Lambda & EC2 performance ensures system reliability

Scalable Solution

Architecture designed to easily accommodate new branches and additional data sources

Accelerated Decision-Making

Visual dashboards provide immediate insights into sales trends and revenue patterns

Next Steps

Refine ETL Pipeline — 1

Enhance Lambda function to transform & load all CSV data into structured Redshift tables

2 — **Expand Visualisations**

Develop additional Grafana dashboards tailored to specific stakeholder requirements

Implement Alerting — 3

Configure proactive notification system for operational issues and anomalies

4 — **Final Presentation**

Prepare comprehensive project showcase combining business insights with operational metrics

Our cloud ETL and monitoring solution provides the foundation for data-driven decision making across Ana-LatteX's café operations, with key components now in production.