Tasks

- ☒ Read End-to-End Data Pipeline.pdf
- ☒ Analyze project requirements from PDF
- ☒ Analyze data structure (Parquet files)
- ☒ Create initial Implementation Plan
- ☐ **Phase 1: Infrastructure Setup (Terraform)**
  - ☐ Setup Terraform structure
  - ☐ Create S3 buckets (Raw/Curated)
  - ☐ Create Kinesis Data Stream
  - ☐ Create Kinesis Firehose Delivery Stream
  - ☐ Create IAM Roles
- ☐ **Phase 2: Machine Learning (Local Training)**
  - ☐ Create ML training script (`ml/train_model.py`)
  - ☐ Train Isolation Forest on Mac M4
  - ☐ Save model to .joblib and upload to S3
- ☐ **Phase 3: Stream Processing (Docker + ECS Fargate)**
  - ☐ Create Dockerfile for Consumer App
  - ☐ Create ECR Repository
  - ☐ Create ECS Cluster & Task Definition
  - ☐ Create ECS Service with Auto-Scaling
  - ☐ Create Application Load Balancer (ALB)
- ☐ **Phase 4: Frontend (Streamlit)**
  - ☐ Create Streamlit Dashboard (`dashboard/app.py`)
  - ☐ Implement Summary Metrics (Total Anomalies, Active Wells)
  - ☐ Implement Real-time Anomaly Table
  - ☐ Integrate Altair/Plotly Time-Series Charts
  - ☐ Connect to Athena using boto3/awswrangler
  - ☐ Create Power BI Portal/Link in Streamlit
  - ☐ Create Dockerfile for Dashboard
  - ☐ Deploy Dashboard to ECS Fargate
- ☐ **Phase 5: Business Intelligence (Power BI)**
  - ☐ Install Athena ODBC Driver
  - ☐ Connect Power BI Desktop to Athena
  - ☐ Create Executive Dashboard
  - ☐ Publish Report (Optional)
- ☐ **Phase 6: CI/CD Pipeline (GitHub Actions)**
  - ☐ Create `.github/workflows/deploy.yml`
  - ☐ Configure AWS Credentials in GitHub Secrets
  - ☐ Automate Terraform Plan
  - ☐ Automate ECR Push
- ☐ **Phase 7: Project Teardown (Cost Management)**
  - ☐ Create `scripts/destroy.sh`
  - ☐ Create S3 Emptying Script (Terraform prerequisite)
  - ☐ Conduct Full Teardown Test to ensure $0 cost
- ☐ **Phase 8: Operations (The Switch)**

□ Create `scripts/project_up.sh` (One-click deploy)
□ Create `scripts/project_down.sh` (One-click destroy)