**Lab 2: Building Data Pipelines with Dataflow and Advanced SQL Analytics**

**Part 1: Pre-Pipeline Preparation (30 minutes)**

**Task 1.1: Verify Lab 1 Infrastructure**

📸 **Screenshot Required #1:** BigQuery showing your existing tables

A screenshot of a computer

AI-generated content may be incorrect.

**Task 1.3: Prepare New Data for Pipeline**

A screenshot of a computer

AI-generated content may be incorrect.

**Part 2: Building Your Dataflow Pipeline (60 minutes)**

📸 **Screenshot Required #3:** Running Dataflow pipeline graph view

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.  
📸 **Screenshot Required #4:** Successful pipeline completion status

A screenshot of a computer

AI-generated content may be incorrect.

**Task 2.2: Understand Your Pipeline**



**Task 2.3: Verify Pipeline Output**

Check that your pipeline successfully loaded data:

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

SELECT 'Original' as source, COUNT(\*) as records

FROM `[your-project].[original-dataset].superstore\_sales`

UNION ALL

SELECT 'Pipeline' as source, COUNT(\*) as records

FROM `[your-project].pipeline\_processed\_data.superstore\_transformed`;

A screenshot of a computer

AI-generated content may be incorrect.

**Part 3: AI-Powered SQL Analytics (60 minutes)**

**Task 3.1: Generate Advanced SQL with Gemini**

Now that you have pipeline-processed data, use Gemini to create sophisticated analyses:

**Analysis 1: Time-Series Decomposition**

A screenshot of a computer

AI-generated content may be incorrect.

**Analysis 2: Customer Cohort Analysis**

A screenshot of a computer

AI-generated content may be incorrect.

**Analysis 3: Market Basket Analysis**

A screenshot of a computer

AI-generated content may be incorrect.

**Task 3.2: Optimize Generated Queries**



**Task 3.3: Create Analysis Views**

A screenshot of a computer

AI-generated content may be incorrect.

**Part 4: DIVE Method with Pipeline Data (45 minutes)**

**Task 4.1: Choose a Pipeline-Specific Question**

Select ONE question that leverages your pipeline's capabilities:

1. **Data Freshness:** "What's the impact of processing delays on decisions?"

**Task 4.2: Apply DIVE Framework**

**D - Discover**

**I - Investigate**

**V - Validate**

**E - Extend**

**Task 4.3: Document Pipeline Insights**

****

**Part 5: Advanced Pipeline Features (30 minutes)**

**Task 5.1: Schedule Your Pipeline**

1. **Create Cloud Scheduler Job**

📸 **Screenshot Required #5:** Cloud Scheduler job configuration

A screenshot of a computer

AI-generated content may be incorrect.

**Task 5.2: Monitor Pipeline Performance**

A screenshot of a computer

AI-generated content may be incorrect.

-- Set up alerting (document the configuration)

**Task 5.3: Cost Analysis**

Understand pipeline costs using university credits:

A screenshot of a computer

AI-generated content may be incorrect.

**Task 6.2: Create Lab Summary**



**Task 6.3: Submit to D2L**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |