

can you create a simple design document for me to test this feature

Here's a **simple design document** for testing your AI agent's ability to generate ETL pipelines based on a design document. This example focuses on a basic e-commerce data workflow.

Design Document: E-commerce Data ETL Pipeline

Overview

This document outlines the requirements for an ETL pipeline to process e-commerce data. The pipeline should ingest raw data from a Google Cloud Storage (GCS) bucket, perform basic transformations, and load the transformed data into BigQuery tables.

Data Sources

1. **Raw Data:** CSV files stored in a GCS bucket (`gs://ecommerce-data-bucket/orders.csv`).
2. **Data Format:** CSV with columns for `order_id`, `customer_id`, `order_date`, `total_amount`, and `order_status`.

ETL Requirements

1. **Data Ingestion:**
 - Ingest raw CSV data from GCS into a BigQuery table named `ecommerce_data.raw_orders`.
 - Use BigQuery's `LOAD DATA` statement.
2. **Data Transformation:**
 - Filter orders with a status of "completed".
 - Calculate a new column `order_year` based on the `order_date`.
 - Store transformed data in a BigQuery table named `ecommerce_data.transformed_orders`.
3. **Data Loading:**
 - Load transformed data into a final BigQuery table named `ecommerce_data.final_orders`.

Pipeline Orchestration

- Use Cloud Composer to orchestrate the pipeline.
- Schedule the pipeline to run daily at 02:00 AM UTC.

Monitoring and Logging

- Use Cloud Logging to monitor pipeline execution and errors.
- Send notifications via Cloud Pub/Sub when pipeline execution fails.

Example SQL Queries for Transformation

```
-- Transformation Query
CREATE OR REPLACE TABLE ecommerce_data.transformed_orders AS
SELECT
    order_id,
    customer_id,
    order_date,
    EXTRACT(YEAR FROM order_date) AS order_year,
    total_amount,
    order_status
FROM ecommerce_data.raw_orders
WHERE order_status = 'completed';

-- Loading Query
INSERT INTO ecommerce_data.final_orders
SELECT * FROM ecommerce_data.transformed_orders;
```

Testing Criteria

- Ensure all data is correctly ingested from GCS into BigQuery.
- Verify that only completed orders are included in the transformed data.
- Confirm that the final table contains the expected columns and data.

This design document provides a clear set of instructions for your AI agent to generate an ETL pipeline. You can use this as input for your agent to test its ability to parse requirements and create pipeline steps dynamically.