



# Data Engineer Portfolio Project Ideas Guide

Eight portfolio-worthy data engineering projects you can build and showcase.

## Beginner Projects

### 1. Batch ETL pipeline

Extract data from a public API, clean and transform it with Python, and load it into a SQLite or Postgres database on a schedule.

### 2. CSV-to-warehouse pipeline

Build a pipeline that loads messy CSV files into a cloud warehouse (Snowflake/BigQuery free tier), handling schema validation and duplicate removal.

## Intermediate Projects

### 1. Airflow-orchestrated pipeline

Recreate your batch ETL project, but orchestrate it with Airflow — including retries, alerts, and scheduled runs.

### 2. Dimensional data model

Take a raw transactional dataset and design a star schema for analytics, then build the transformation logic with dbt.

### 3. Data quality monitoring pipeline

Add automated data quality checks (nulls, duplicates, schema drift) to an existing pipeline, with alerts when checks fail.

## Advanced Projects

### 1. Real-time streaming pipeline

Use Kafka (or a managed alternative) to process a real-time data stream and load aggregated results into a dashboard-ready table.

### 2. End-to-end analytics platform

Combine ingestion, transformation, warehousing, and a BI tool (e.g. Metabase) into one cohesive, documented system.

### 3. Cost-optimized pipeline redesign

Take an existing inefficient pipeline (your own or a public example) and redesign it for lower compute cost



and faster runtime, documenting the before/after.

## How to Present These Projects

For every project, write a short README covering: the problem, your architecture decisions (and alternatives you considered), how you handled failure cases, and what you would do differently at scale. This is what separates a tutorial-follow-along from a real portfolio piece in interviews.

