Take-Home Assignment: Policy Lifecycle Pipeline Challenge

**Context**  
You’re joining the team responsible for the Single View of Policy (SVoP). This data product supports enterprise-wide use of policy lifecycle data across analytics, underwriting, and regulatory functions. Your task is to design and document a scalable data pipeline to model raw policy events into a clean, auditable structure.

**Your Task**  
Build a dbt (or SQL + markdown) pipeline to create a final model:  
**policy\_lifecycle\_enriched**

Use the provided **policy\_events.csv** file to:

**1. Model Policy Events**

* Group and order events by policy\_id
* Derive:
  + First quote
  + First bind
  + First cancel
  + Current status
* Calculate durations:
  + Quote to bind
  + Bind to cancel
* Handle:
  + Duplicates
  + Out-of-order events
  + Missing/invalid timestamps
* Partition data by event\_month

**2. SQL & Data Modelling**

* Write modular, high-quality SQL
* Optimise for readability, performance, and maintainability
* Use at least two models (e.g., stg\_policy\_events, policy\_lifecycle\_enriched)
* Include 3–5 data tests (e.g., non-null policy\_id, valid status values)

**3. Documentation**

* Explain logic, assumptions, and edge case handling
* Provide 2–3 sample BI queries
* Use dbt docs or markdown

**4, Scaling Strategy (Markdown)**

If data volume increased 10x monthly, how would you scale and productionise this pipeline?

**Evaluation Criteria**

* **Advanced SQL**: structure, clarity, and logic
* **Data transformation and modelling quality**
* **Documentation & reasoning**
* **Production-readiness**: scalability, auditability
* **Communication and architectural thinking**

**Bonus (Optional)**

* Sketch DAG/refresh logic (e.g., Cloud Composer)
* Propose anomaly detection/logging