

Title: Building a Data Engineering Team for CRED

Team Structure (First 3 Hires):

1. **Data Engineer:** Focus on pipelines, dbt, BigQuery, and CI/CD.
2. **Data Analyst/ML Engineer:** Build dashboards and feature sets.
3. **Data Ops / QA:** Ensures data quality, monitoring, and SLA compliance.

Key Standards & Processes:

- **Naming Conventions:** stg_ for staging, fact_ for fact tables.
- **Data Quality:** Automated dbt tests for nulls, uniqueness, and row counts.
- **SLAs:** Define data availability windows for dashboards and ML pipelines.
- **Lineage & Documentation:** dbt docs for lineage and model documentation.
- **CI/CD:** GitHub Actions or similar for automated dbt runs and tests.

Balancing Delivery vs Platform Investment:

- Prioritize features that directly impact revenue and user experience.
- Incremental investment in modular pipelines to handle future growth.
- Document everything to enable future team scaling.

Future Roadmap for 10× Data Volume:

- Partition large fact tables by event_date.
- Use clustering for frequent query columns (user_id, event_type).
- Introduce streaming ingestion (Pub/Sub or Kafka).
- Implement feature store (Feast) for ML pipelines.
- Enable cost monitoring and query optimization for BigQuery.