📅 Roadmap: Sales Analytics Data Warehouse (GCP)

Timeline: 18 Sept – 10 Oct (23 days)

🔹 Week 1 (Sep 18 – Sep 24) → *Foundation + Core Pipeline*

Goal: Get raw data flowing into BigQuery fact table

- **Day 1–2 (Sep 18–19): GCS Setup**

    - Create bucket (`gm-sales-raw-data-affan`).

    - Upload `sales_small.csv` to `inbound/sales/`.

    - Test upload/download.

- **Day 3 (Sep 20): BigQuery Dataset**

    - Create dataset `gm_sales` (US region).

    - Create **external table** on CSV.

    - Run `SELECT *` sanity check.

- **Day 4–5 (Sep 21–22): Fact Table**

    - Transform into `fact_sales` with partition by `order_date`.

    - Add casting + cleaning.

    - Run row count + daily totals query.

- **Day 6–7 (Sep 23–24): Repo & Docs**

- ○ Organize SQL into `sql/01_external_fact.sql`.

- ○ Add README with Day 1 summary.

- ○ Write project architecture overview.

◆ **Week 2 (Sep 25 – Oct 1) → *Expansion to Data Warehouse***

Goal: Build a proper star schema + add quality checks

- ● **Day 8–9 (Sep 25–26): Dimension Tables**

  - ○ Create `dim_products`, `dim_customers`, `dim_date`.

  - ○ Link them with `fact_sales`.

  - ○ Draw ER diagram for docs.

- ● **Day 10–11 (Sep 27–28): Data Quality Checks**

  - ○ Add SQL validations: missing order_id, negative sales/profit.

  - ○ Store results in `quality_logs`.

  - ○ Document checks in README.

- ● **Day 12–13 (Sep 29–30): Queries & Aggregations**

  - ○ Write example queries:

    - ■ Profit by category

    - ■ Sales by region

    - ■ Top 10 products

  - ○ Save them into `sql/02_analytics_queries.sql`.

- **Day 14 (Oct 1): Documentation Update**

  - Add star schema diagram.

  - Update README: Day 2 progress.

◆ **Week 3 (Oct 2 – Oct 8)** → *Automation + Visualization*

Goal: Make project impressive for evaluation

- **Day 15–16 (Oct 2–3): Orchestration**

  - Set up **Cloud Composer (Airflow)**.

  - Create a DAG for CSV → external table → fact/dim load.

  - Test with small CSV refresh.

  - Delete environment after testing (to save credits).

- **Day 17–18 (Oct 4–5): Visualization**

  - Connect BigQuery to **Looker Studio**.

  - Create dashboard:

    - Sales trends over time

    - Profit by category

    - Region map

    - Top products/customers

- **Day 19–20 (Oct 6–7): Enhancements**

  - Polish dashboards.

  - Add filtering (e.g., by region/segment).

○ Save screenshots for docs.

- **Day 21 (Oct 8): Repo Finalization**

    ○ Push SQL, docs, screenshots.

    ○ Update README: Day 3 progress.

◆ Wrap-Up (Oct 9–10) → *Polish + Presentation*

Goal: Finish report + prepare demo

- **Day 22 (Oct 9): Report Writing**

    ○ Write Final Report: Problem → Methodology → Results → Future Scope.

    ○ Insert screenshots, diagrams.

- **Day 23 (Oct 10): Presentation**

    ○ Prepare PPT (architecture, pipeline, dashboard).

    ○ Dry-run demo on GCP (GCS + BigQuery + Looker Studio).

    ○ Final Git commit: `"Final project submission"`.

✅ Deliverables by 10 Oct

- **GCS bucket + BigQuery fact & dimensions.**

- **Data quality checks + analytics queries.**

- **Looker Studio dashboard (screenshots + live demo).**

- **GitHub repo (SQL, docs, screenshots, README).**

- **Final Report + PPT for submission.**