GROUP 2: WEATHER X EBAY ANALYTICS (EAST COAST)

Ayesha Ateeq Burney, Linus Krüer, Benedikt Arndt, Muhammad Afaq Ali Saqib

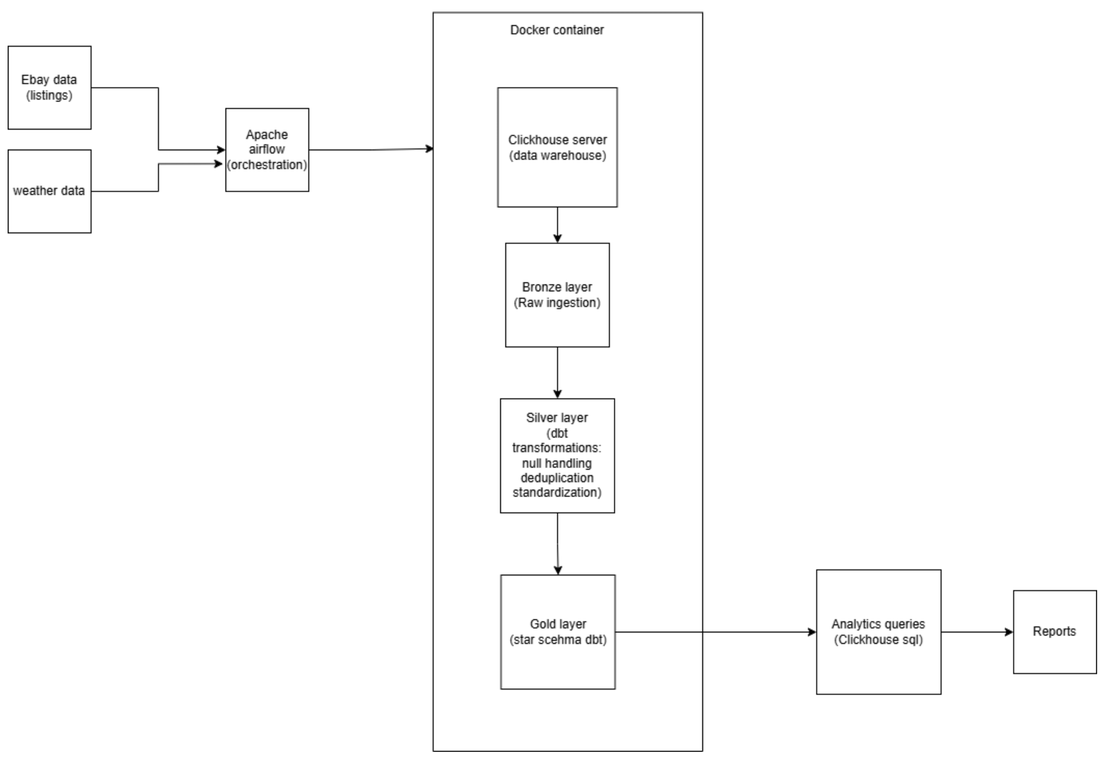
Institute of Computer Science, University of Tartu

UNIVERSITY OF TARTU

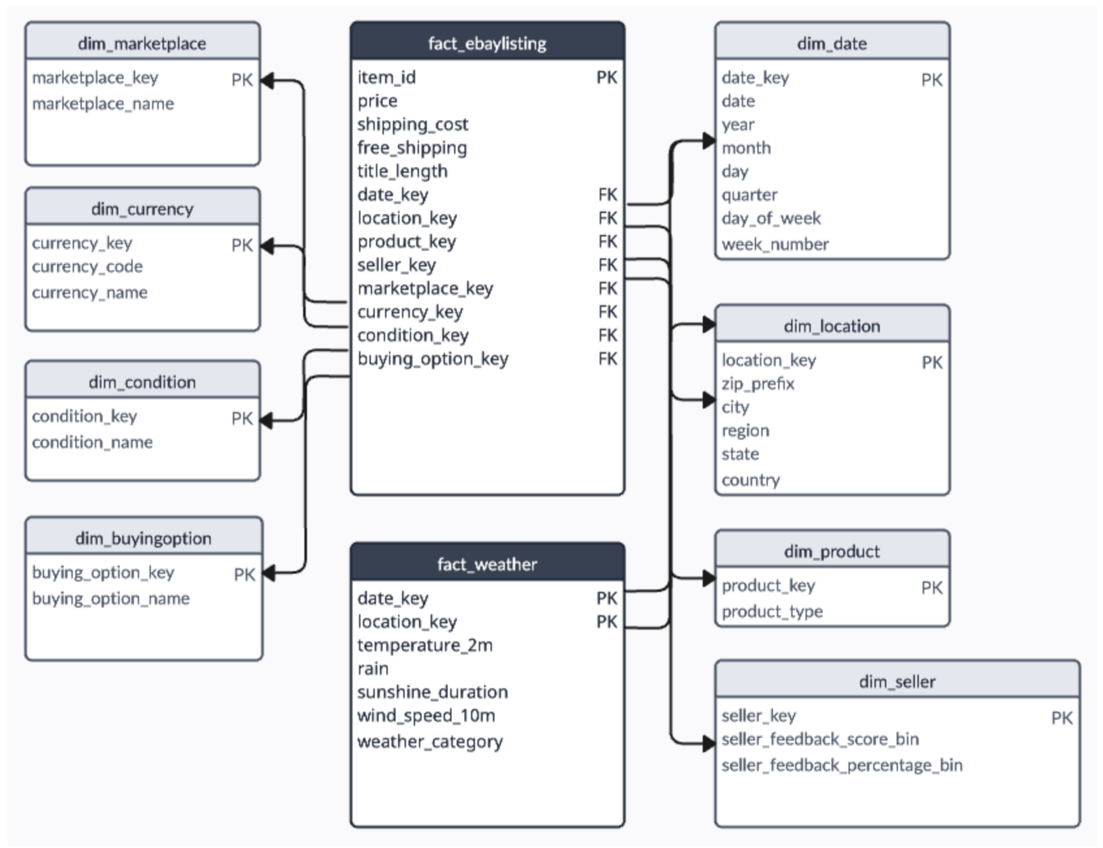
Business context

- We analyze how weather patterns on the US East Coast influence eBay marketplace activity, combining hourly eBay listings with daily weather observations.
- Key KPIs include: average listings per hour, average price changes and the free-shipping rate under different weather conditions.
- The project uses two datasets: Open-Meteo weather data[1](temperature, rain, wind, etc.) and eBay listing data[2](prices, seller metrics, shipping options, product types).

Architecture



Star Schema



Pipelines & Tools

- We implement a full **medallion architecture** using Docker-orchestrated services:
  - **Bronze (Raw Ingestion)**: Airflow PythonOperators fetch Open-Meteo weather + eBay Browse API data into ClickHouse: `bronze_weather`, `bronze.ebay_raw_data`.
  - **Silver (Cleaning & Standardization)**: dbt removes duplicates, casts types and harmonizes fields into `silver_weather`, `silver_ebay_listings`.
  - **Gold (Star Schema)**: dbt models build analytical tables: `fact_weather`, `fact_listings` + 11 `dim_*` tables enabling weather-driven eBay insights.
- **Pipeline orchestration**: Apache Airflow triggers ingestion → dbt Silver → dbt Gold → dbt tests, visible in Graph View DAGs (`unified_data_pipeline`).
- **Data Transformation** (Bronze → Silver):
  - In `bronze.ebay_raw_data`, title strings vary in length and format.
  - Silver transformation creates a `title_length` metric using: `length(title) AS title_length`, enabling listing-quality analytics in Gold layer.
- **Tools in the stack**: Airflow (orchestration), ClickHouse (warehouse), dbt (SQL transformations + tests), Docker Compose (deployment).
- Weather and listings are linked along **date + location**, enabling category demand, pricing and shipping KPIs per weather condition.

Governance & Quality

- Implemented ClickHouse RBAC with two roles: `analyst_full` (unmasked access) and `analyst_limited` (masked serving views).
- Applied multiple masking techniques in `serving_views_masked`: hashing (`item_id`, `seller_location`), bucketing (`price`), banding (`seller_feedback_percentage`) and prefixing (`zip_prefix`) to protect PII while preserving analytical value.
- Added data-quality tests in OpenMetadata: NOT NULL checks (foreign keys), uniqueness checks (surrogate keys) and value-range validation (feedback percentage 0–100).
- Documented tables, columns and relationships in OpenMetadata, including automated ingestion of ClickHouse metadata and registration of Superset dashboards.
- Ensured governance across the pipeline through Iceberg bronze-layer lineage, masked serving layers and metadata-tracked transformations.

Outcome & Lessons

Insights

- **Extreme Cold** conditions drive the highest listing volume (700 listings) but the lowest prices, suggesting urgent demand for warming products.
- **Extreme Heat** shows a 100% free shipping rate, indicating sellers maximize convenience during heat-related product demand.
- Clear **weather-driven product demand patterns** exist: each weather condition drives specific category spikes (e.g., hand warmers for Extreme Cold, sunglasses for Extreme Heat).
- **Premium products** (e.g., air conditioners) show high price volatility during extreme weather, while simple items (e.g., hand warmers) maintain stable pricing.
- **Premium sellers** (Very High feedback, 5000+) dominate during Extreme Cold and Extreme Heat, commanding higher market shares.
- **Extreme Cold listings** have the longest titles and highest quality scores, indicating more detailed product descriptions.
- Major metropolitan areas show different pricing patterns: NYC dominates volume during Extreme Cold, while Miami shows premium pricing during Extreme Heat.

Dashboards

- Pricing trends by weather condition
- How product category demand shifts with weather changes
- Average price comparison: product type vs. weather
- Listing volume by weather code (high vs. low activity)
- Weather's influence on daily listing activity

Lessons Learned

- **Weather data nuances**: Weather conditions had a more complex impact on listings than initially assumed, requiring finer segmentation and additional analysis to uncover meaningful patterns.
- **Data quality is critical**: Unexpected missing values or inconsistencies in the raw data required additional cleaning steps. Implementing automated data quality checks earlier would have saved time and reduced errors.

References

[1] *Open-Meteo Weather API*. 2025. URL: `https://open-meteo.com/` (visited on 09/25/2025).

[2] *eBay REST API Documentation*. 2025. URL: `https://developer.ebay.com/api-docs/static/gs_ebay-rest-getting-started-landing.html` (visited on 09/23/2025).

Github-Repo