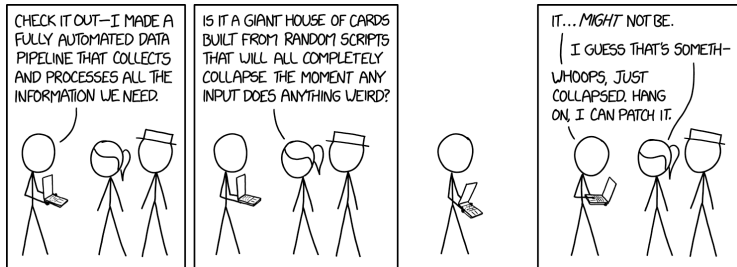


# Data & Analytics Case Study

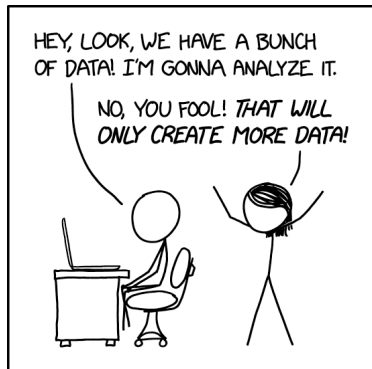
Max van Rooijen (max@pl9.co)

March 11, 2025



# Why Analytics in Comic Design?

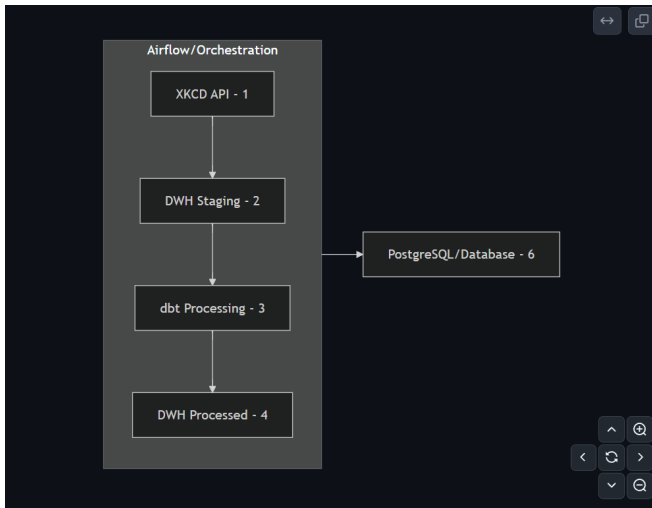
- ▶ Understanding audience preferences and engagement
- ▶ Optimizing storytelling and visual elements
- ▶ Enhancing content personalization and recommendation



# Technical Solution: Batch Processing Pipeline

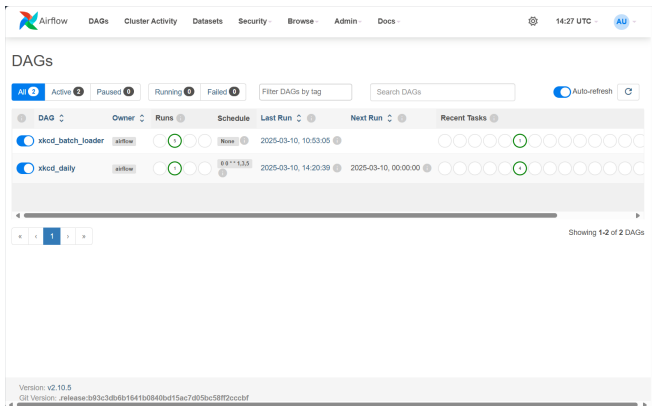
- ▶ Fetching data from the XKCD API
- ▶ Using Apache Airflow to orchestrate workflows
- ▶ Staging data in a PostgreSQL Data Warehouse
- ▶ Transforming raw data using dbt (Data Build Tool)
- ▶ Running data quality checks to ensure integrity

# Solution Overview



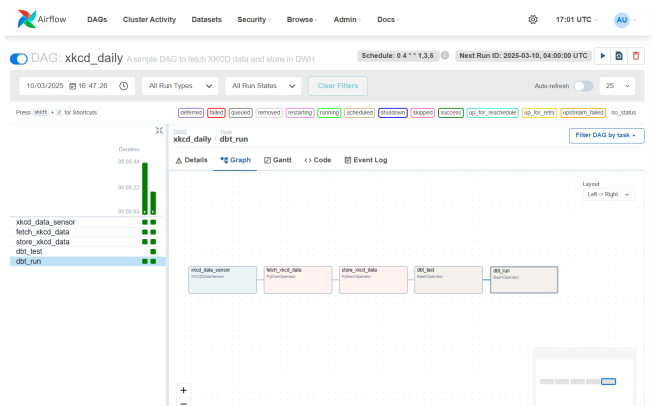
High-level overview of the architecture.

# Pipeline Architecture



Overview of the DAGs in Airflow

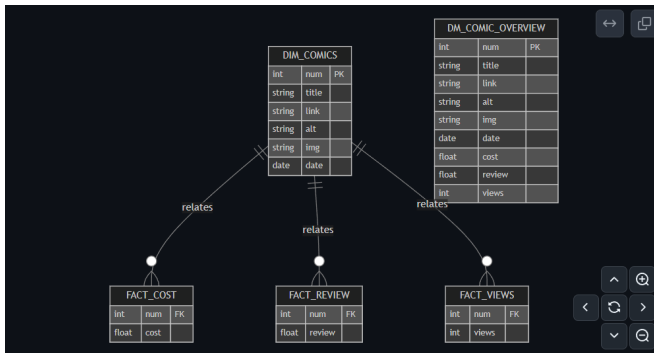
# Pipeline Architecture



Daily run DAG

# Data Model and Insights

- ▶ Dimensional model for structured data storage
- ▶ Fact tables for costs, reviews, and views
- ▶ Aggregated insights for trend analysis



## Future Improvements

- ▶ **Machine Learning Integration:** Use ML models for predictive insights, such as forecasting trends in comic engagement.
- ▶ **Enhanced Data Quality Checks:** Implement more comprehensive tests for duplicates, referential integrity, and outlier detection.
- ▶ **Unit Tests for DAGs:** Add unit tests to Airflow DAGs to ensure robust workflow execution.
- ▶ **More Elaborate Data Model:** Expand the data warehouse schema to include more detailed metadata and user behavior analytics.
- ▶ **Real-time Processing:** Extend the pipeline to support real-time analytics with streaming data solutions like Apache Kafka.
- ▶ **Scalability:** Optimize the data pipeline to handle larger volumes of data and improve performance.
- ▶ **Monitoring and Alerts:** Set up monitoring and alerting to notify the team of any issues or anomalies in the data pipeline.



# Future Business Cases

- ▶ **Optimizing Ad Placement:** Use insights from user engagement to strategically place advertisements in comics.
- ▶ **Personalized Content Recommendations:** Suggest comics based on reader preferences and past interactions.
- ▶ **Subscription & Monetization Strategies:** Identify premium content opportunities based on high-engagement trends.
- ▶ **Predicting Viral Content:** Utilize data to forecast which comics are likely to go viral and maximize their exposure.
- ▶ **Expanding into New Media:** Apply the same analytical insights to animations, graphic novels, and interactive storytelling.
- ▶ **Improving Reader Retention:** Track engagement metrics to refine content strategy and keep readers coming back.