Machine Learning and Data Mining

DFM – Example

Claudio Sartori

DISI
Department of Computer Science and Engineering – University of Bologna, Italy
claudio.sartori@unibo.it

Outline

The Story – Why RetailCo Needs a Data Warehouse

- RetailCo: a national retail chain with hundreds of stores and thousands of products
- Each store runs its own operational system:
  - Sales registers
  - Inventory management
  - Customer loyalty programs
- Data is fragmented, inconsistent, and operational in nature (built for transactions, not analysis)
- Stores may have different software

Challenges Faced by Management

Executives and managers encounter daily challenges:

Fragmented information – Sales reports differ between stores and channels.

Slow decision-making – IT staff must manually gather and reconcile spreadsheets.

Missed opportunities – Managers cannot react quickly to demand, customer preferences, or supply chain issues.

Limited forecasting – Overstock and stockouts occur due to lack of predictive insights.

The Solution – Build a Data Warehouse

- A centralized, integrated, historical repository of corporate data.
- Supports strategic, tactical, and operational decision-making.
- Enables OLAP analysis and data mining / machine learning to uncover:
  - Trends
  - Customer behaviors
  - Performance indicators
- Transforms RetailCo from *data chaos* to *data-driven decision making*.

Outline

High-Level Management Requirements

Executives and Directors want the DWH to deliver:

- Unified view of the business (one version of the truth).
- Strategic KPIs: revenue growth, profitability by product, market basket analysis.
- Customer insights: loyalty program performance, segmentation, churn analysis.
- Forecasting capabilities: sales and inventory demand trends.
- Support for corporate strategy: store expansion and investment decisions.

Medium-Level Management Requirements

Department Heads and Regional Managers require:

Sales monitoring – daily, weekly, monthly views across stores and regions.

Inventory control – stock levels, turnover, slow-moving items.

Promotions analysis – measure effectiveness of campaigns and seasonal boosts.

Supply chain optimization – supplier performance, delivery times, costs.

Customer service monitoring – satisfaction indicators, loyalty card usage.

Outline

Dimensional Fact Model Approach

The DWH will be modeled using the *Dimensional Fact Model (DFM)*:

- Ensures clarity for both business stakeholders and IT specialists.
- Provides intuitive hierarchies and measures for analysis.

Facts – Core Measurable Events

- *Sales Fact*
  - Measures: Quantity Sold, Revenue, Discount, Profit
  - Grain: Each sales transaction (line-item level)
- *Inventory Fact*
  - Measures: Stock level, Replenishment quantity, Turnover rate
  - Grain: Daily snapshot per product per store
- *Promotion Fact*
  - Measures: Campaign cost, Incremental revenue, ROI
  - Grain: Promotion-event per product per store

Dimensions – Perspectives for Analysis

Time – Day → Month → Quarter → Year

Product – Product → Subcategory → Category → Department → Brand

Store – Store → City → Region → Country

Customer – Customer → Segment → Loyalty Tier → Demographics

Promotion – Campaign → Channel → Period

Supplier – Supplier → Region → Type

Sales Fact Schema – Example

Fact – Sales

Measures – Quantity Sold, Revenue, Profit, Discount

Dimensions –

- Time (Day, Month, Year)
- Product (Product, Category, Brand)
- Store (Store, Region)
- Customer (Customer ID, Segment, Loyalty Tier)
- Promotion (Campaign, Channel)

Benefits of the Star Schema

Additivity – Measures like sales and profit can be aggregated across dimensions.

Historical analysis – Supported by time dimension hierarchies.

Flexibility – Slice-and-dice across region, product, or customer groups.

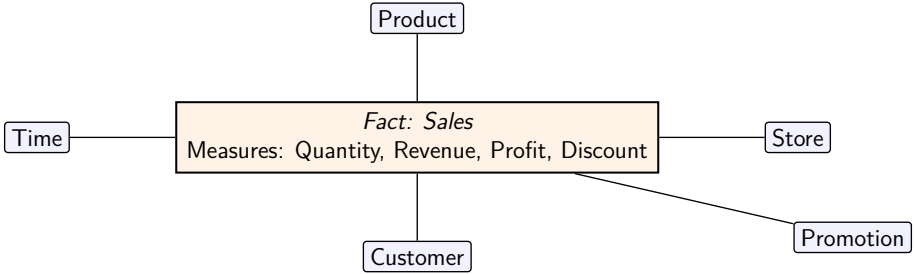
Outline

Final Deliverable

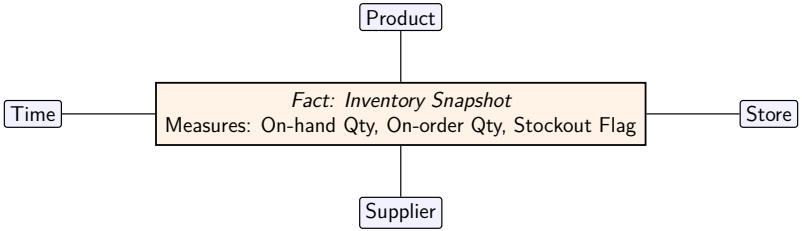
- A retail data warehouse designed around:
  - Sales
  - Inventory
  - Promotions
- Dimensions reflect the company's core perspectives:
  - Time, Product, Store, Customer, Promotion, Supplier
- Provides RetailCo with a scalable, integrated, business-driven BI platform.
- Supports decision-making from the boardroom to the shop floor.

Outline

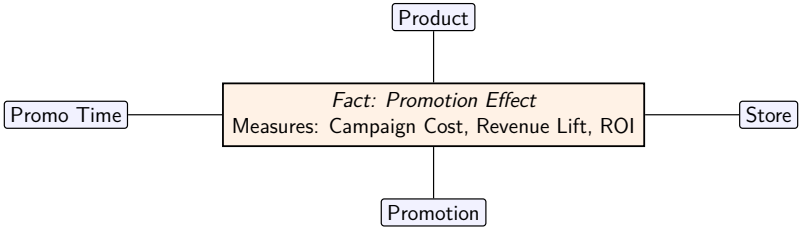
Sales Data Mart – DFM



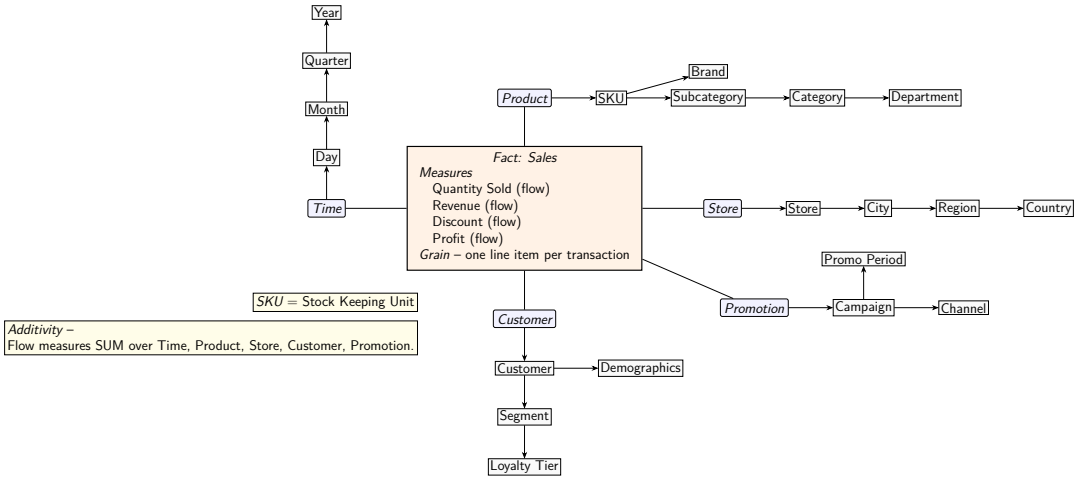
Inventory Snapshot Data Mart – DFM



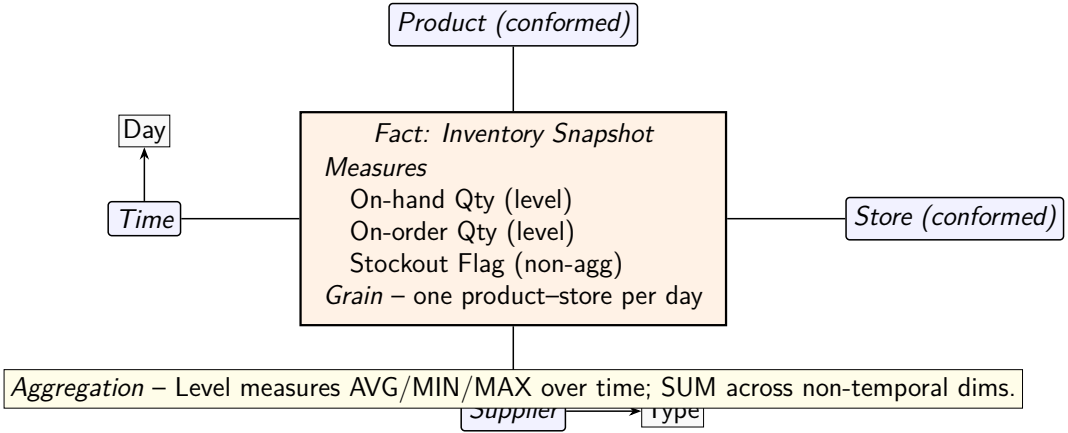
Promotion Effect Data Mart – DFM



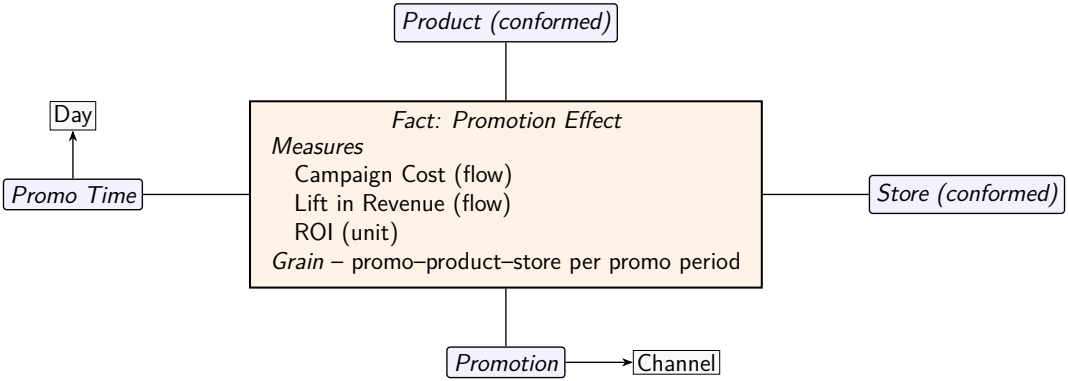
Sales Data Mart – Conformed Star (DFM)



Inventory Snapshot Data Mart – Periodic Snapshot (DFM)



Promotion Effect Data Mart – Event Fact (DFM)



Algebraic measure – ROI requires support measures (Revenue Lift, Cost).

Legend & Modeling Notes

Facts (orange) hold the quantitative measures; grain is explicitly stated.

Dimensions (blue) provide analysis axes with *hierarchies* (arrows).

Conformed dimensions (Product, Store, Time) are reused across data marts.

Additivity

- Flow measures (e.g., Revenue) SUM across all non-temporal dims and over time.
- Level measures (e.g., On-hand Qty) use AVG/MIN/MAX over time; SUM across non-temporal dims.
- Unit measures (e.g., ROI) aggregate with AVG; avoid summation.

Recommended physical design – Star schemas; consider snowflaking only for shared, slowly changing hierarchies.

Outline

Next Steps for Implementation

To ensure successful deployment of the Data Warehouse

ETL Process – Define extraction, transformation, and loading workflows for consistent, cleansed data.

Data Governance – Establish roles, responsibilities, and data quality standards.

Infrastructure – Select scalable architecture (on-premises, cloud, or hybrid).

BI Tools – Integrate with visualization and reporting platforms (e.g., Tableau, Power BI).

Phased Rollout – Start with sales and inventory marts; extend to promotions and customer analytics.

Training – Provide training for managers, analysts, and IT staff to maximize adoption.

From Data Chaos to Data-Driven Decisions