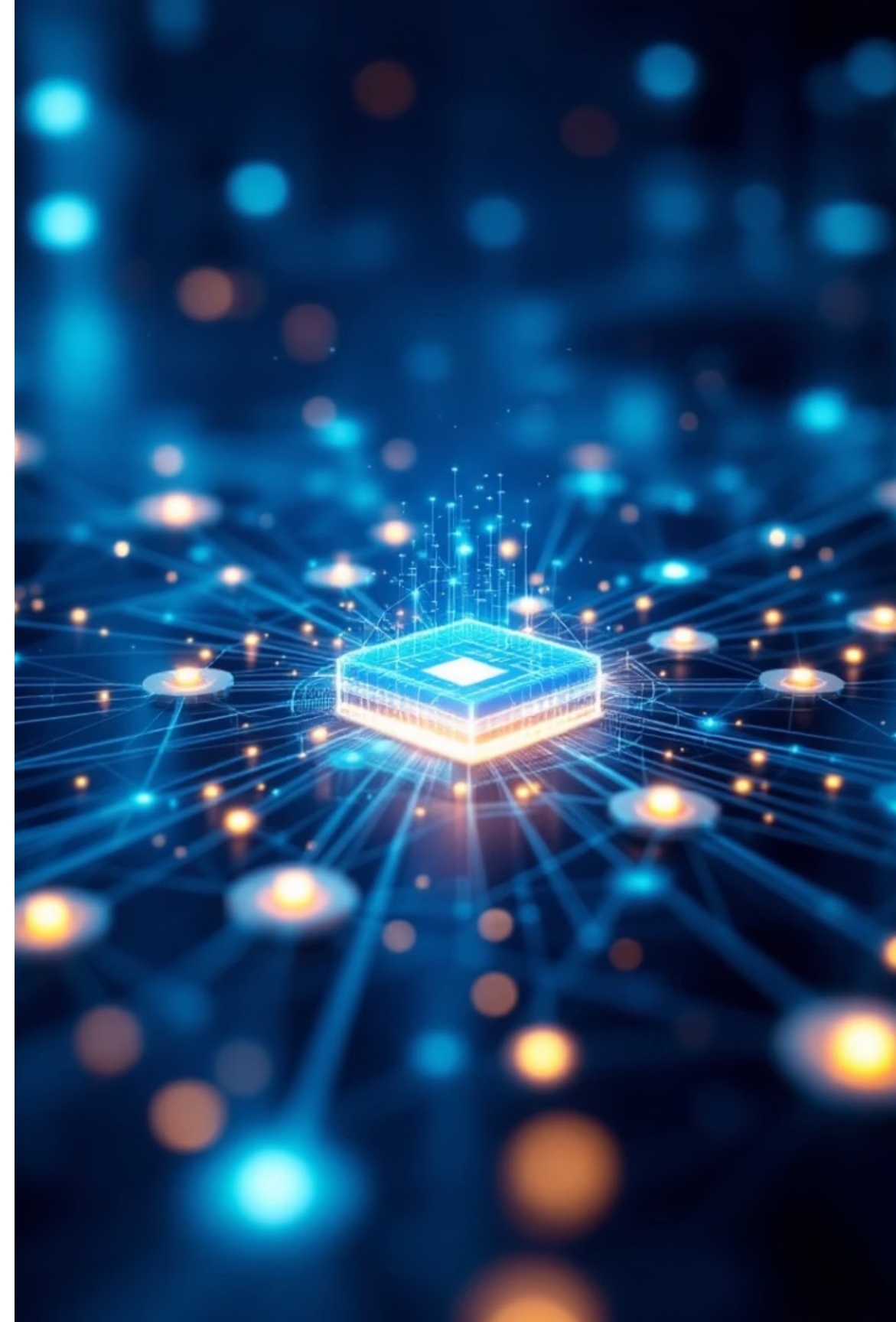


Data Warehouse Design Proposal

Data Engineering Task - Bosch Thermotechnik GmbH

Amparo Iglesias

21.04.2025



Data Model Concept Proposal

Based on a dimensional star schema, supporting fast analytical queries and actionable business insights.

Fact Tables

- Fact_Orders: Tracks sales, pricing, and quantities
- Fact_Delivery: Captures delivery schedules, actuals, and delays



Fact Tables are closely linked to dimension tables, and store the key business metrics

Dimension Tables

- Dim_Customer: Customer details and types
- Dim_Product: Product names and base prices
- Dim_Delivery: Delivery dates and times
- Dim_Order: Order dates and status



Dimension Tables provide descriptive context for facts

Fact Tables

Orders & Deliveries

Fact_Orders

- Order ID, Customer ID, Product ID
- Quantity Ordered, Gross/Net Price
- Total = Quantity × Net Price

Fact_Delivery

- Delivery ID, Scheduled/Actual Dates
- Delay in Minutes
- On-Time Delivery (Yes/No)

Business Insights

- Most sold products
- Average prices
- Revenue per customer
- Logistics performance



Dimension Tables



Dim_Customer

Customer ID, Name,
Type



Dim_Product

Material ID, Product
Name, Base Price



Dim_Delivery

Delivery ID, Date,
Time, Day of Week

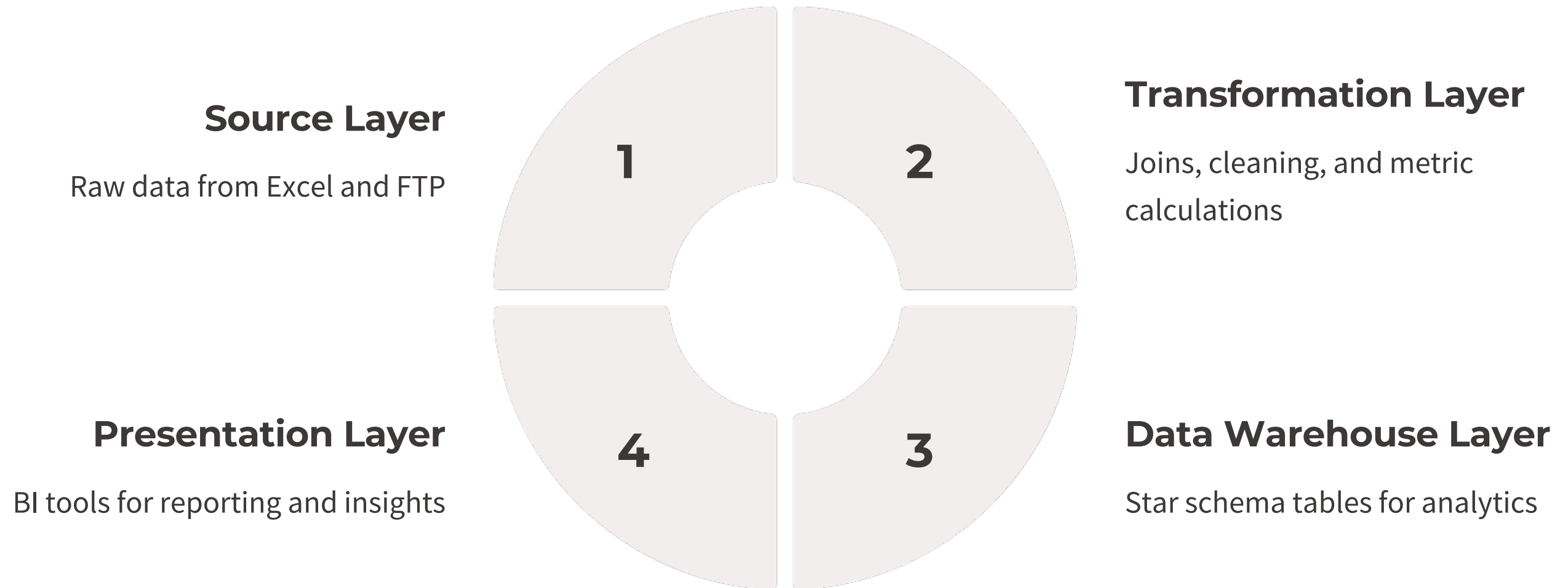


Dim_Order

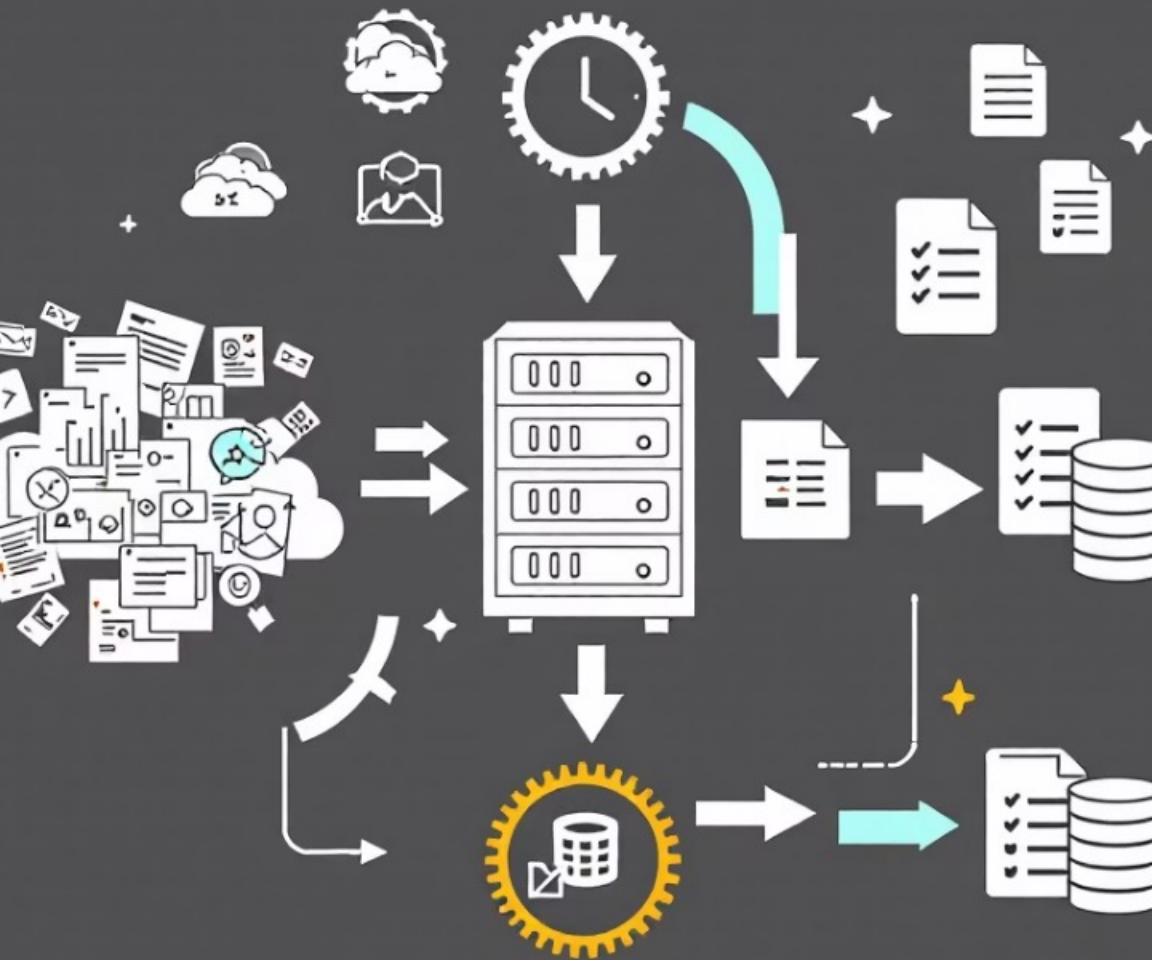
Order ID, Order Date,
Status



Layered Data Warehouse Architecture



Data Integration & Transformation



1

Raw Ingestion

Collect Excel and FTP data

2

Data Cleaning

Normalize formats, resolve inconsistencies

3

Joins & Calculations

Link orders, deliveries, and confirmations

4

Star Schema Build

Populate fact and dimension tables



Key Challenges in Implementation

Data Integration

Combining Excel and FTP data with different formats

Date/Time Consistency

Resolving mixed date formats across systems

Order-Delivery Linking

Handling deliveries with multiple order lines and partial fulfillment

Pricing & Missing Data

Ensuring accurate net price and addressing incomplete records

Open Questions & Next Steps

Data Refresh Frequency?

Determine ETL schedule: daily, hourly, or event-based

BI Tool Selection?

Choose platform for dashboards and reporting

Business Logic Decisions?

Define handling of partial/undelivered orders and regional analysis

Route Optimization?

Create methods to improve delivery efficiency

