

Mahavir Skoda – AI-Enabled Sales Data Warehouse

(Automation, Analytics & Intelligent Decision Support System)

Prototype Scope

This project is a prototype using synthetic dealership data to simulate real-time automation.

Once connected to Mahavir Skoda's live CRM and ERP systems, the same framework can be scaled for production, supporting multiple dealerships and future AI-driven analytics

Confidential Project Prototype – Built for Demonstration & Process Design Only

Prepared by:

Kaushik Yeddnapudi

Kaushikyeddnapudi0406@gmail.com

+91 7993273549

Mahavir Skoda – Sales Data Warehouse (DW) & Automation Project Plan

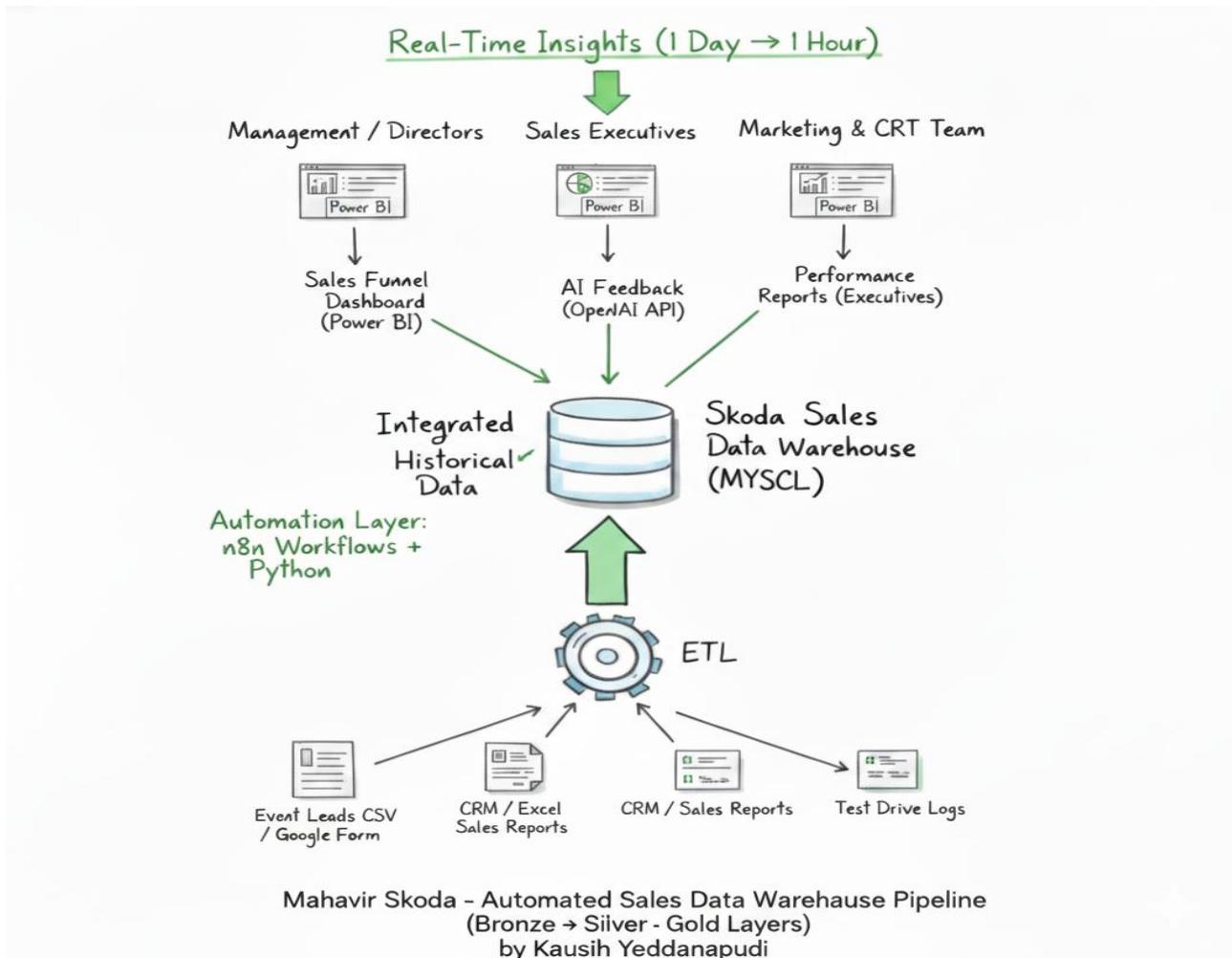
Project Overview

The Skoda Sales Data Warehouse (DW) automates the end-to-end flow of dealership sales and lead data — from raw collection to real-time business insights.

It integrates marketing leads, CRM updates, test drives, and final sales into one unified system using ETL workflows (n8n + Python) and MySQL as the data warehouse.

Insights are then visualized using Power BI dashboards, eliminating manual Excel work.

This process turns 1-day report delays into 1-hour insights — accelerating decision-making.



Data Flow Overview (Based on the Architecture Image)

Step 1: Data Sources (Raw Input Layer)

Includes:

- Marketing event leads (CSV / Google Forms)
- CRM sales reports (Excel exports)
- Test drive logs (manual sheets / internal forms)

Process:

All raw files are collected automatically using n8n scheduled workflows or Python scripts. They are stored in the Bronze Layer as-is for traceability.

ETL & Data Cleaning Layer (n8n + Python)

ETL (Extract, Transform, Load) is the engine of automation.

- Data is **extracted** from multiple sources.
- **Transformation scripts** clean duplicates, fix date formats, standardize vehicle names, and match leads with corresponding test drives and bookings.
- Cleaned data is **loaded** into the **Silver Layer** of the MySQL warehouse.

Tools Used:

- n8n → Scheduled workflows
- Python + Pandas → Cleaning and joining logic
- SQL Stored Procedures → Validation & audit logs

 Example transformation:

“Convert all variants of ‘KUSHAQ’ and ‘kushaq’ to standardized value ‘Kushaq’.”



ETL & Data Cleaning Layer (n1S + Python)

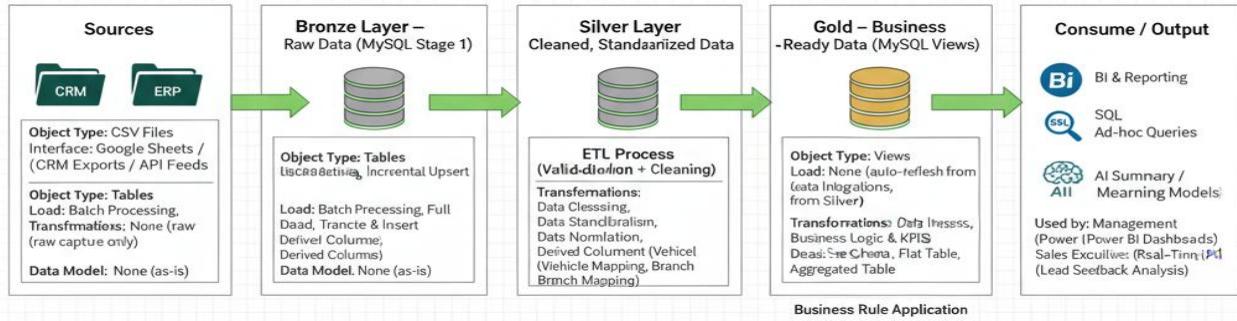
The Engine of Automation

Key Deliverables from this Layer

- Automated ETL job runs (daily / scheduled)
- MySQL tables created: leads_clean, sales_clean, testdrive_clean
- AI summaries generated for each customer
- Power BI auto-refresh trigger enabled

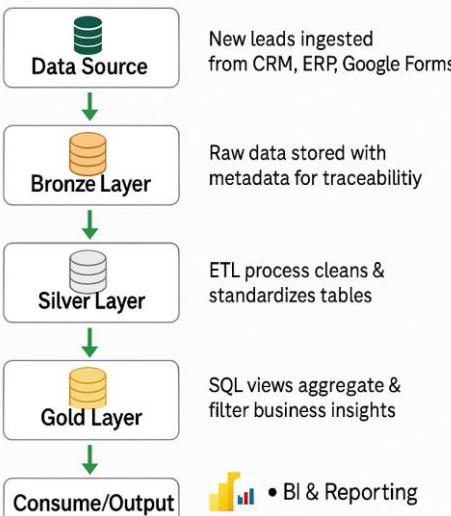
Data Architecture

High Level Architecture – Mahavir Skoda Sales Data Warehouse



This architecture represents the automated flow of Mahavir Skoda's sales, CRM, and test drives through MySQL-based warehouse. It enables real-time dashboards, AI summaries, AI summaries, and business reporting, without manual Excel dependency.

How It Works: Mahavir Skoda Sales Data Warehouse



Automation & Accuracy: All data flows automatically from CRM and event sources to the MySQL warehouse, reducing manual data entry errors by over 90%.

Faster Decisions: What once required days of manual compilation is now refreshed in near real-time — giving sales managers and directors live visibility into conversions, leads, and performance.

Scalable Design: The Bronze–Silver–Gold architecture supports future integrations such as service data, customer feedback, and inventory tracking without re-engineering.

AI-Driven Insights: Integrated OpenAI and n8n workflows generate instant lead summaries, follow-up recommendations, and performance explanations, empowering every department to act proactively.

Centralized Dashboards: Power BI connects directly to the Gold Layer, providing interactive KPIs, predictive analytics, and branch-level performance views — no Excel consolidation needed.

Governance & Security: Role-based access, PII masking, and structured backups ensure data integrity and compliance across all operational layers.

Together, this architecture delivers **speed, intelligence, and reliability**, transforming Mahavir Skoda's sales operations into a modern, insight-driven ecosystem.

Operational Automation Layer – MySQL Validation + AI Insight Generation

MySQL Queries (Example)

Data Quality Check – Detect Duplicates

```
SELECT phone, COUNT(*) AS dup_count  
FROM leads_clean  
GROUP BY phone  
HAVING dup_count > 1;
```

Sales Funnel Metrics – Last 30 Days

```
SELECT  
    SUM(status = 'New') AS new_leads,  
    SUM(status = 'Test Drive') AS test_drives,  
    SUM(status = 'Booked') AS bookings,  
    ROUND(SUM(status = 'Booked') / SUM(status = 'New') * 100, 2) AS conversion_rate  
FROM leads_clean  
WHERE created_date >= CURDATE() - INTERVAL 30 DAY;
```

AI Prompt (Example) Using STAR Method (Confidential Data Context)

Lead Insight & AI Summary (Sales Intelligence)

Header: AI Summary for Sales Executive Dashboard

Situation:

A new customer lead record has been added to the MySQL warehouse (Silver Layer). The system needs a short, action-oriented summary for the sales executive without revealing sensitive customer data.

Task:

Summarize the lead's intent, interest, and urgency in a structured format while maintaining confidentiality.

Action:

Analyze the customer's anonymized notes, vehicle interest, and interaction history. Focus on tone, keywords, and sentiment rather than personal identifiers.

Result (Expected Output):

A 2-line AI-generated insight ready to display on the Sales Executive dashboard.

Prompt:

“Based on the anonymized customer note and interest data below, summarize their intent and recommend a next action.

Maintain data confidentiality — do not include names, contact details, or locations.

Input:

- **Vehicle Interest:** {{vehicle_interest}}
- **Lead Notes:** {{lead_notes}}
- **Stage:** {{lead_stage}}

Output Format:

- **Summary:** [Customer intent in one sentence]
- **Recommended Action:** [1 short actionable step for executive]

“DDL STRUCTURE OVERVIEW”

DDL & Database Structure – Skoda Sales Data Warehouse (MySQL)

Section 1: Layer Summary (3-column layout table)

Layer	Description	Core Tables
Bronze (Raw Ingestion)	Stores unprocessed data directly from CRM/ERP/Event sources. Maintains traceability and historical versions.	bronze_leads_raw, bronze_testdrives_raw
Silver (Clean Layer)	Deduplicated, standardized, and validated data ready for operational use.	silver_leads, silver_testdrives
Gold (Business Layer)	Aggregated, summarized, and star-schema formatted data for Power BI dashboards and AI analysis.	gold_fact_sales, vw_conversion_funnel

Key Design Principles

- Each layer has its own tables, ETL rules, and audit trail.
- All timestamps and loads are **automated through n8n workflows**.
- **Silver layer** enforces clean canonical data (deduped, mapped).
- **Gold layer** generates pre-computed KPIs for dashboards.
- Each table has created_at, updated_at, job_id, source_file for traceability.

Example Table Creation Snippets

Example DDL Structure (MySQL – Prototype Reference)

```
CREATE TABLE bronze_leads_raw (...);
```

- Captures unprocessed lead data directly from CRM exports, Google Forms, and event sources.
- Used for traceability, audit, and replay in ETL jobs before cleaning.

```
CREATE TABLE silver_leads (...);
```

- Stores cleaned, standardized, and deduplicated lead records after ETL validation.
- Serves as the operational table for CRT and Sales dashboards with accurate records.

```
CREATE VIEW vw_conversion_funnel AS SELECT ...;
```

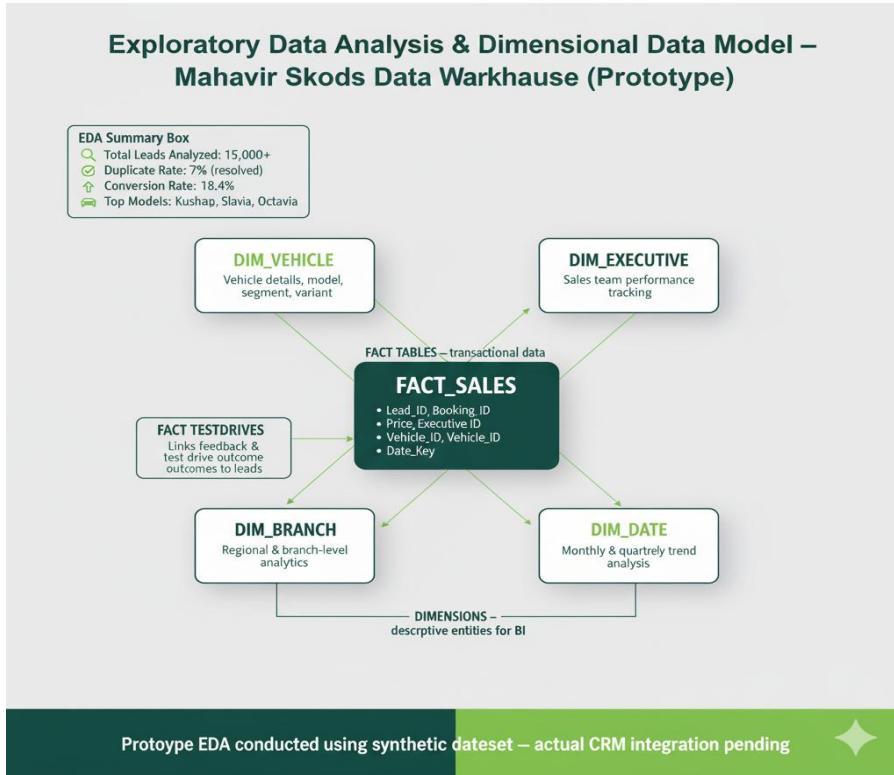
- Aggregates key performance metrics such as leads, test drives, and bookings.
- Used by Power BI for real-time sales funnel visualization across branches.

Note

This example schema is created for demonstration and prototype purposes using synthetic data.

The production-level implementation will require normalization adjustments, extended audit logging, and field mapping as per Mahavir Skoda's actual CRM structure and business logic.

Exploratory Data Analysis (EDA) & Data Model — Prototype



Interpretation & Insights

The above model represents the **prototype star schema** for the Mahavir Skoda Sales Data Warehouse. It centralizes all **sales and lead interactions** in the **FACT_SALES** table, linking them to descriptive **dimension tables** for analysis in Power BI.

- **FACT_SALES** acts as the single source of truth, combining lead, vehicle, branch, and executive data for all transactions.
- **DIM_VEHICLE** enables model-level and segment-wise performance analysis.
- **DIM_EXECUTIVE** and **DIM_BRANCH** connect sales outcomes with team and regional performance.
- **DIM_DATE** supports monthly, quarterly, and year-to-date KPI comparisons.
- **FACT_TESTDRIVES** enhances customer journey tracking by connecting interest → test drive → booking conversion.

The **EDA Summary Box** validates this data model using synthetic data, ensuring all joins and metrics perform as expected before real CRM integration.

Technical Notes

- ETL transformations load cleaned data into **Silver Layer**, while **Gold Layer** (MySQL Views) drives dashboards.
- This model design is **Power BI-ready**, following dimensional modeling best practices (Star Schema).
- All identifiers (Lead_ID, Vehicle_ID, Executive_ID, etc.) are anonymized for data privacy compliance.

Implementation Status

This dimensional model has been validated using prototype datasets to ensure structure, join logic, and performance. Upon access to live CRM and ERP feeds, the model will seamlessly extend to real sales data pipelines for Mahavir Skoda and other Mahavir Group brands.

Business Impact & Conclusion

Business Impact

Faster Decision-Making

Management dashboards now update in near real time — cutting decision latency from **1 day to under 1 hour**. Executives can instantly view KPIs like leads, conversions, test drives, and sales performance without manual data merging.

Data Transparency Across Teams

- Customer Relations, Sales Executives, and IT Admins operate from a **shared single source of truth**.
- This eliminates version conflicts and ensures all departments see unified data snapshots.

Reduced Manual Workload

Automation (via **n8n workflows & Python**) replaces repetitive Excel operations.

This saves 4–6 hours daily per department, enabling teams to focus on conversion strategies instead of data cleaning.

Enhanced Customer Experience

AI-driven summaries allow sales executives to **personalize follow-ups**, improving engagement rates and reducing missed leads.

Financial and Operational Control

Branch-level and model-level insights help allocate marketing budgets, track performance by region, and optimize stock levels.

Strategic Benefits

Scalable Architecture:

The current design supports **multi-brand expansion** — applicable to Skoda, Mercedes-Benz, Benelli, and future dealerships.

AI-Driven Growth:

Integration with OpenAI APIs enables predictive modeling — conversion probability, lead scoring, and feedback sentiment analysis.

Data Governance Foundation:

The MySQL-based data warehouse creates a strong foundation for future **data governance, compliance, and analytics maturity**.

No Vendor Lock-In:

All layers are built on open tools (MySQL, Power BI Desktop, Python, n8n), ensuring **low cost and complete flexibility**.

Future Expansion Roadmap

1. **Integrate Post-Sales & Service Data** → build a full customer lifecycle view.
 2. **Adopt Cloud Data Warehouse (Snowflake / BigQuery)** for scalability beyond 5M records.
 3. **Introduce ML Forecasting Models** for conversion prediction and executive performance optimization.
 4. **Develop a Unified “Mahavir Group Data Command Center”** — connecting all dealerships to one analytics platform.
-

Conclusion

The **Mahavir Skoda SmartFlow & Data Warehouse** initiative showcases how business analysis, data engineering, automation, and AI can converge to transform dealership operations.

By consolidating every touchpoint — from event leads to bookings — into a single, intelligent data ecosystem, Skoda can evolve from **manual reporting** to a **real-time decision intelligence system**.

This prototype proves that with the right data foundation, even a single analyst can architect a **complete digital transformation journey** — from raw data to actionable insights.