

**“SOFA-SO GOOD”**

**LAKEHOUSE**

**ARCHITECTURE**

**(Design Phase Document)**



# **INDEX**

- 1.) Introduction
  - a. Why it works?
  - b. Why Lakehouse Architecture?
- 2.) Objective
- 3.) Purpose
- 4.) Mission and Goals
  - a. Mission
  - b. Goals
- 5.) Design Phase
- 6.) Overview of the Lakehouse Architecture
  - a. Data Inputs
  - b. Data Outputs
  - c. List of data users

# 1. Introduction

Sofa-so good cloud-based architecture is a modern, scalable solution that combines data lake flexibility with data warehouse speed, enabling seamless management of massive unstructured data and rapid analysis within a unified, user-friendly platform.

## 1.1 Why It Works:

- **Scalable & Adaptive:** Seamlessly expands to handle growing data volumes and workloads.
- **Secure & Controlled:** Implements role-based access for data security and governance.
- **Optimized & Efficient:** Enhances performance while minimizing resource usage.

## 1.2 Why Lakehouse? It's Simply Smarter.

### ➤ **Best of Both Worlds:**

- Scale effortlessly with data lake to store vast, unstructured data.
- Get speedy, precise insights like a data warehouse for structured analysis.

### ➤ **One Platform, Endless Possibilities:**

- Process data in real-time or batches—whatever your business needs.

### ➤ **Streamlined and Cost-Effective:**

- Grow with confidence, knowing the architecture scales affordably as your data grows.

## 2. Objective:

To build a robust and scalable data infrastructure that enables Sofa-so-good to gain easy access to business data.

## 3. Purpose:

- Centralize and integrate disparate data sources.
- Automate data processing and transformation for efficiency.
- Enable easy data retrieval for advanced analytics and reporting.
- Provide a single source of truth for all business data.

## 4. Mission & Goals

### a. Mission:

To design and implement a high-performance, reliable data engineering pipeline that ensures accurate, readily accessible, and consistently available data, empowering Sofa-so-good with a foundational infrastructure for future analytics and operational improvements.

### b. Goals:

- **Centralize Data:** Create a unified repository for easy access and informed decision-making.
- **Optimize Data Flow and Pipeline Efficiency:** Improve data workflows by integrating real-time insights and automation for seamless processing and enhanced performance.
- **Future-Proof Data Pipelines:** Build adaptable data engineering systems that prioritize cost-efficiency, optimization, and reliability.

## 5. Design Phase



# Overview of the Lakehouse Architecture

The Lakehouse architecture is a multi-layered system encompassing data collection, processing, and analytics. The key components are as follows:

## **a) Data Inputs**

### **i. Sales Transactional Data:**

- Nature: Detailed records of in-store and point-of-sale transactions, including product details, prices, quantities, dates, customer information, and payment methods.
- Format: CSV, relational database tables (e.g., PostgreSQL, MySQL).
- Ingestion type: Batch ingestion.
- Usage: Sales analysis, forecasting, customer behavior analysis, inventory management, and financial reporting.

### **ii. E-Commerce Data:**

- Nature: Online sales data, including product views, cart additions, checkout processes, order details, customer accounts, and online payment information.
- Format: JSON, relational database tables.
- Ingestion type: Stream ingestion for real-time order processing, and batch ingestion for daily/weekly reports.
- Usage: Online sales analysis, customer journey tracking, website conversion optimization, and personalized online marketing.

### iii. **Website Logs:**

- Nature: Detailed records of user interactions on the Sofa-so-good website, including page views, clicks, search queries, session durations, and device information.
- Format: log files, unstructured data.
- Ingestion type: Stream ingestion for real-time web traffic analysis.
- Usage: Website analytics, user behavior analysis, A/B testing, and marketing campaign performance evaluation.

### iv. **Warehouse Operational Data:**

- Nature: Data related to warehouse inventory, stock movements, shipping, receiving, and order fulfillment processes.
- Format: CSV, database tables.
- Ingestion type: Batch ingestion for daily/weekly inventory updates.
- Usage: Inventory management, stock optimization, order fulfillment tracking, and supply chain analysis.

## b) **Data Outputs**

### i. **Sales Performance Dashboards:**

Real-time sales tracking, product performance, and regional sales analysis.

User needs: Sales and management teams for monitoring sales trends and performance.

### ii. **Inventory Management Reports:**

Stock levels, reorder points, and inventory turnover rates.

User needs: Operations and inventory management teams for optimizing stock levels.

iii. **Customer Segmentation Reports:**

Customer demographics, purchase behavior, and personalized recommendations.

User needs: Marketing and sales teams for targeted campaigns and customer relationship management.

iv. **Sales Forecasting Reports:**

Predictive analysis of future sales based on historical data and market trends.

User needs: Management and sales teams for resource allocation and planning.

v. **Web Analytics Dashboards:**

Website traffic, user engagement, and conversion rates.

User needs: Marketing and web development teams for optimizing website performance.

c) **List of Data Users**

- **Marketing Team:** Uses customer segmentation, web analytics, and sales data for targeted campaigns and personalized marketing.
- **Sales Team:** Uses sales performance dashboards, customer purchase history, and forecast reports to improve sales strategies and customer interactions.



- **Operations Team:** Uses inventory management reports and supply chain analysis to optimize stock levels and warehouse operations.
- **Management Team:** Uses sales performance dashboards, forecast reports, and overall business analytics for strategic decision-making.
- **Customer Service Team:** Uses customer data and purchase history to provide personalized and efficient customer support.
- **Web Development Team:** Uses web analytics and API data to improve website functionality and customer experience.