

DOCUMENTO DE ARQUITETURA DETALHADA: DATA LAKEHOUSE PARA ANÁLISE

Tecnologias: PostgreSQL/Supabase, N8N, FastAPI, Python

Arquitetura: Medallion (Bronze, Silver, Gold)

Novembro de 2025

1 VISÃO GERAL

Este projeto estabelece uma arquitetura moderna de Data Lakehouse utilizando PostgreSQL (Supabase) como repositório analítico, orquestrada pelo N8N. O objetivo primário é desacoplar a análise de dados estratégicos do sistema transacional (ERP Consinco, simulado pela API FastAPI), fornecendo uma Fonte Única da Verdade (Single Source of Truth) para as equipes de Business Intelligence (BI), Planejamento e Finanças.

1.1 Contexto de Negócio Varejista

O foco é transformar dados transacionais de alta cardinalidade (Vendas, Histórico de Estoque) em métricas que suportam decisões críticas de varejo:

- **Gestão de Compras:** Análise de desempenho de fornecedores e necessidade de reposição de produtos.
- **Otimização de Estoque:** Previsão de ruptura (status_estoque CRÍTICO/BAIXO) e identificação de excesso de estoque (status_estoque EXCESSO) para liberar capital.
- **Performance de Vendas:** Acompanhamento de tendências de faturamento e ticket médio por categoria e período.

1.2 Fluxo da Informação – Medallion Architecture

A arquitetura segue rigorosamente o padrão Medallion Architecture (Bronze, Silver, Gold) para garantir qualidade, rastreabilidade e performance.

Camadas da Arquitetura Medallion

Camada	Propósito	Transformação	Detalhes Técnicos
Bronze	Arquivamento Legal e Histórico Bruto. Serve como backup exato do dado de origem para auditoria e reproprocessamento.	Ingestão (EL): Carregamento direto do payload JSON.	Uso de JSONB para flexibilidade de schema e hash_dados para atomicidade e prevenção de duplicidade.

Camadas da Arquitetura Medallion

Camada	Propósito	Transformação	Detalhes Técnicos
Silver	Qualidade e Normalização dos Dados. Dados estruturados, confiáveis e unificáveis. Fonte de dados mestra.	Limpeza e Normalização (T): Conversão de tipos, aplicação de regras de negócio e validação de chaves.	Tabelas em 3NF com Chaves Primárias e Estrangeiras para consistência relacional.
Gold	Entrega de Valor de Negócio (BI). Foco em performance de leitura e métricas prontas para consumo do usuário final.	Agregação e Modelagem (L): Criação de Materialized Views (MVs) e Data Marts para agregação de alta performance.	MVs otimizadas com Índices Únicos para latência de consulta sub-segundo em dashboards.

2 DECISÕES ESTRATÉGICAS E TÉCNICAS

2.1 Estratégia de Orquestração (N8N e Sequenciamento)

Decisões de Orquestração

Decisão	Motivação Técnica	Motivação de Negócio
Orquestrador N8N	A interface Low-Code/No-Code facilita a manutenção por Analistas de Dados, evitando o desenvolvimento (Time-to-Market) de novas integrações comerciais.	Agilidade e Redução de Custos. Diminui a dependência de Desenvolvedores para pequenas tarefas de rotina, aumentando a autonomia da equipe de BI.
Workflow Sequencial	O fluxo é estritamente ordenado (Fornecedores, Produtos, Vendas) porque todas as Chaves Estrangeiras (FKs) devem existir na Camada Silver antes de carregar as tabelas transacionais que dependem delas.	Garantia de Integridade Referencial. Impede que vendas sejam registradas para produtos ou fornecedores inexistentes, assegurando a confiabilidade dos relatórios.
Registrar status_final dos Logs (ETL)	Atendimento ao requisito de auditoria. O log na tabela etl_logs permite registrar o status da execução (duracao_ms, status_final e momento_carga)	Contratos de Serviço (SLA). Permite ao time de BI e Operações monitorar o Service Level de atualização das cargas de dados, garantindo

para registros processados por lote.

informações atualizadas e confiáveis.

2.2 Arquitetura do Data Lakehouse (PostgreSQL)

Decisões Arquiteturais do Data Lakehouse

Decisão	Motivação Técnica	Motivação de Negócio
Bronze: JSONB	O campo JSONB armazena o dado bruto. Oferece flexibilidade de schema (ideal para APIs que mudam layout), índices (queries por MD5) impedem a ingestão do mesmo payload duas vezes, garantindo a Atomicidade.	Arquivamento e Evolução. Mantém um registro auditável e inalterável do dado bruto, permitindo reproduzir análises sem reprocessar dados históricos.
Silver: Aplicação de 3NF e Tipagem	A Silver força a conversão de todos os campos (ex: preço em NUMERIC, datas em TIMESTAMP) aplicando o modelo relacional (3ª Forma Normal).	Confiabilidade da Análise. Remove inconsistências (como número em texto ou erro de parsing) que são o maior gerador de erros analíticos. Silver é a fonte única para relatórios.
Gold: Materializar Views (MVs)	MVs pré-calculam e persistem agregações (SUM, AVG), movendo a carga de trabalho do momento da consulta para o momento do cálculo (do ETL).	Performance Crítica de BI. Garante que relatórios financeiros que exigem estoque e faturamento sejam entregues em latência próxima de zero, suportando relatórios de primeira necessidade.
Gold: Uso de COALESCE	Aplicação de lógica no SQL para substituir NULL e métricas vazias sem agregações.	Exatidão Financeira. Garante que, se um produto não teve vendas em um período, será listado com 0 em vez de gerar erro e confusão para analistas.
Performance: CONCURRENTLY	Este comando permite que a Materialized View seja atualizada sem bloquear a leitura da versão antiga, mantendo a disponibilidade da camada Gold.	Alta Disponibilidade. Garante 100% de uptime dos dashboards, crucial durante o horário comercial para relatórios financeiros.

3 ESTRUTURA DETALHADA DAS CAMADAS E ENTIDADES

3.1 Camada Bronze (Rastreabilidade e Flexibilidade)

Estrutura da Camada Bronze

Tabela	Entidade	Estrutura de Dados	Tipo de Carga	Propósito
bronze_dados_bruto	Particionada	JSONB (Payload completo)	FULL LOAD	Arquivo legal e base para reproprocessamento em caso de erro nos dados.
etl_logs_etl	Logs	Registro com duracao_ms, status_final, registros_processados	Incremental Workflow N8	Monitoramento do SLA de entrega de dados atualizados.

3.2 Camada Silver (Normalização e Qualidade)

Estrutura da Camada Silver

Tabela	Chave Primária	Entidades Relacionadas	Propósito de Negócio
silver_fornecedores	fornecedor_id	silver_produtos	Mestra de Compras e Logística.
silver_produtos	produto_id	silver_vendas, silver_estoque_atual	Catálogo central; define categoria, preço base e relacionamento com o fornecedor.
silver_vendas	venda_id	silver_produtos	Transacional de Vendas. Base para KPIs de performance de Mix de produto e vendedor.
silver_estoque_atual	produto_id	silver_produtos	Posição de Estoque no momento do

			snapshot. Usada para calcular o capital de giro no estoque.
silver_estoque_historico	id	silver_produtos	Transacional de Movimentação. Usado para análise de métricas, previsão e preferências de capital.

3.3 Camada Gold (Métricas de Negócio Otimizadas)

Estrutura da Camada Gold

Objeto Gold	Tipo	Descrição da Métrica de Negócio
gold_vendas_por_produto	MV	Análise de Mix de Vendas. Fornece faturamento_total, total_unidades_vendidas e ticket_medio por produto. Essencial para identificar itens mais lucrativos (SKUs A).
gold_vendas_por_periodo	MV	Análise de Tendência Mensal/Sazonal. Agregado por mês, ideal para gráficos de evolução e comparação de performance anualizada (Year-over-Year).
gold_estoque_resumo	MV	KPI Crítico de Inventário. Combina posição atual (mínimo/máximo) e preço para categorizar o status_estoque (CRÍTICO, BAIXO, OK, EXCESSO). Usado para priorizar a reposição.
gold_alertas_estoque	View	Visão de Exceção para Compras. Filtra o estoque_resumo para mostrar apenas os SKUs com status em (CRÍTICO, BAIXO, EXCESSO), servindo de ação para equipe de reposição.