Berger Paints India Limited
Business Requirements Document (BRD)

Prepared by: Ankit Sharma
ankitsharma@bergerpaints.in
Date: November 2025

END-TO-END CDC + ELT DATA PIPELINE USING SNOWFLAKE DYNAMIC TABLES

Confidential — For Internal Use Only

COVER PAGE SUMMARY

Company Name: Berger Paints India Limited

Registered Office: Berger House, 129 Park Street, Kolkata, West Bengal, 700017

CIN: L51434WB1923PLC004793

Website: https://www.bergerpaints.com

PROJECT ABSTRACT

This document defines the business and technical requirements for Berger Paints India Limited's modernization of its enterprise data platform. The initiative leverages Snowflake's Dynamic Tables, Streams, and Tasks to implement a secure, scalable, and near-real-time Change Data Capture (CDC) based ELT pipeline for unified analytics and operational efficiency.

1. EXECUTIVE SUMMARY

Berger Paints India Limited (BPIL) is a leading paint manufacturer operating across multiple divisions — Retail, Industrial, Projects, Dealers, Brand Stores, and Local Shops.

The current legacy batch ETL process causes data delays, manual effort, and inconsistencies.

The proposed CDC + Dynamic Table ELT pipeline in Snowflake will transform data management, providing near real-time insights, automation, and improved governance.

The project will deliver:

- **90%** reduction in reporting latency
- Centralized data visibility
- End-to-end automation and error tracking
- Stream-based change capture for all major entities

2. BUSINESS OBJECTIVES

Objective	Description	Outcome / KPI
1. Real-Time Insights	Enable near real-time analytics for all divisions.	Data freshness < 15 minutes.
2. Unified Data Platform	Establish a single source of truth across business units.	Consistent and trusted reporting.
3. Automation	Remove manual dependencies using Snowpipe & Tasks.	Fully automated ingestion and refresh.
4. Governance & Monitoring	End-to-end lineage, audit, and error visibility.	100% transparency and reliability.

Objective	Description	Outcome / KPI
5. Scalability & Performance	Handle 10M+ records/day with minimal maintenance.	Scalable cloud-native system.

3. CURRENT CHALLENGES

Challenge	Description	Impact
Batch Latency	Manual loads run nightly.	Reports delayed by 24 hours.
No CDC Logic	Entire tables reloaded each cycle.	Wasted compute, high cost.
No File Tracking	Duplicate or partial files undetected.	Data integrity issues.
No Alerts or Monitoring	Failures unnoticed until reporting.	Trust erosion in analytics.
Siloed Tools	Multiple scripts across systems.	Poor maintainability.

4. PROPOSED SOLUTION

Berger Paints will adopt a Snowflake-native ELT pipeline architecture using:

- Internal Stages (STG): Secure file loading area.
- Streams: Track row-level changes (CDC).
- MERGE Procedures: Upsert into curated CORE tables.
- Dynamic Tables: Auto-refresh downstream data.
- Tasks: Automate ingestion orchestration.
- Monitoring Layer: Log activities, alerts, and file hashes.

Key Benefits:

- ✅ Real-time insights
- ✅ Full automation
- ✅ Audit & Governance

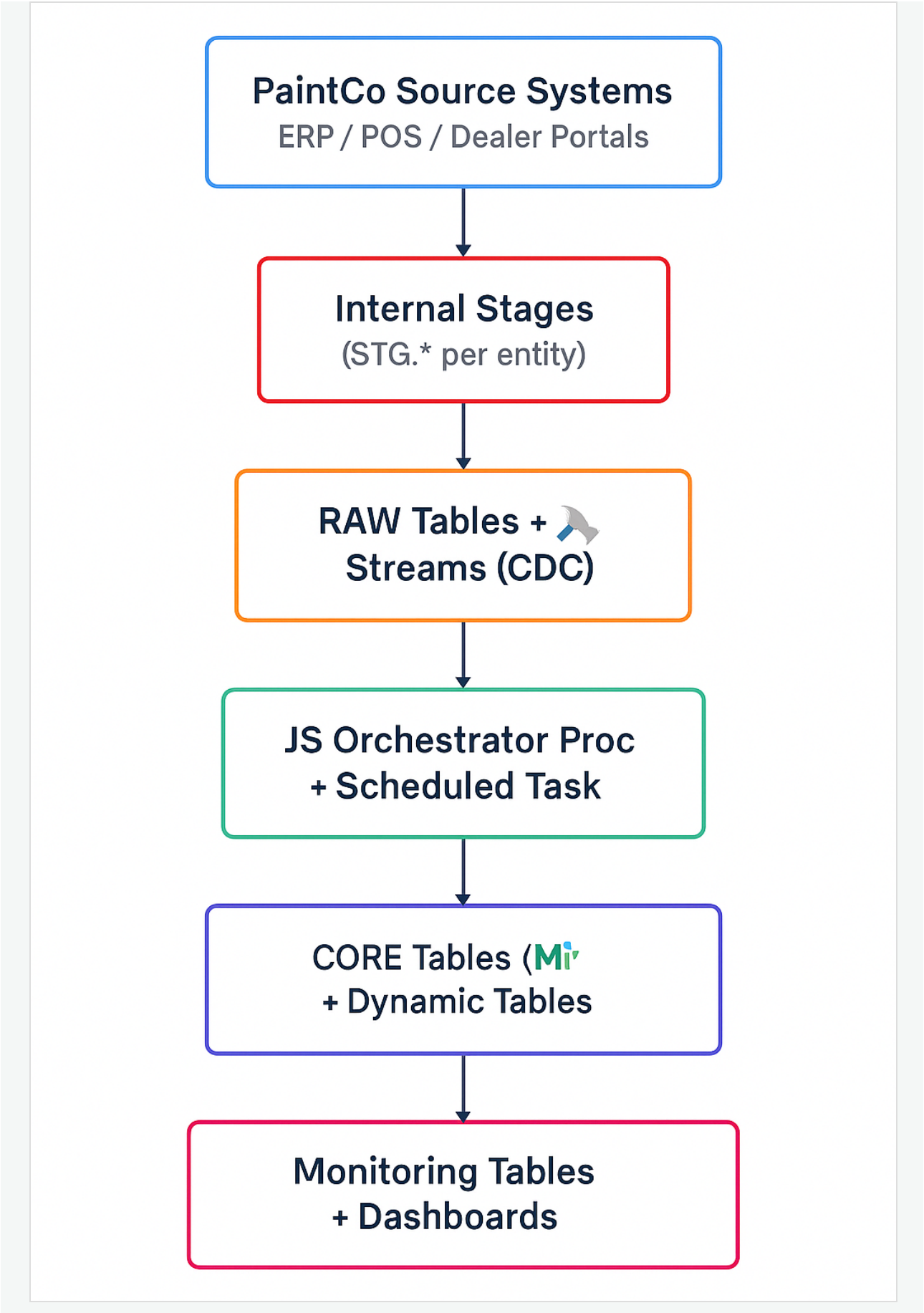
- ✅ Cost efficiency

5. SOLUTION ARCHITECTURE

5.1 LAYERED DESIGN

Layer	Purpose	Schema
Staging Layer	Store incoming CSVs (per entity).	STG
Raw Layer	Initial data load.	RAW
CDC Layer	Streams on raw tables for incremental capture.	RAW
Transformation Layer	MERGE into cleansed, deduplicated tables.	CORE
Analytics Layer	Dynamic Tables auto-refresh for BI use.	CORE
Monitoring Layer	Track files, hashes, logs, and alerts.	MONITORING

5.2 ARCHITECTURE DIAGRAM


PaintCo Source Systems
ERP / POS / Dealer Portals
Internal Stages
(STG.* per entity)
RAW Tables +
Streams (CDC)
JS Orchestrator Proc
+ Scheduled Task
CORE Tables (
+ Dynamic Tables
Monitoring Tables
+ Dashboards

6. DATA DOMAINS & ENTITIES

Domain	Entities / Tables	Purpose
Customer Management	CUSTOMERS, DEALERS, DISTRIBUTORS, LOCAL_SHOPS	Manage customer and partner data
Product & Inventory	PRODUCTS, INVENTORY, SUPPLIERS	SKU, stock, and vendor info
Sales Operations	SALES, PROMOTIONS	Track transactions and offers
Projects & Industrial	PROJECTS, INDUSTRIAL_CLIENTS	Capture project & industrial B2B data
Procurement & Logistics	PURCHASE_ORDERS, SHIPMENTS	Manage supply chain operations
Monitoring & Control	PROCESSED_FILES, PIPE_LOG, ALERT_QUEUE	File tracking and alerts

7. FUNCTIONAL REQUIREMENTS

ID	Requirement	Description
FR-01	Automated File Ingestion	Files ingested from internal stages automatically.
FR-02	Real-Time Load Trigger	Snowpipe REST or Tasks initiate ingestion.
FR-03	Stream-Based CDC	Streams capture incremental data changes.
FR-04	File Tracking	MD5 hash validation avoids duplicate loads.
FR-05	Error Handling	Move bad files to @STG.ERRORS/.
FR-06	Logging & Audit	Store events in PIPE_LOG for traceability.
FR-07	Alert Notifications	Alerts on failure via webhook or email.
FR-08	Dynamic Table Refresh	Automated refresh for BI datasets.
FR-09	Archive Success Files	Move completed files to @STG.PROCESSED/.

8. NON-FUNCTIONAL REQUIREMENTS

Category	Requirement	Target
Performance	Data freshness SLA	< 15 min
Scalability	Handle 10M+ rows/day	✅
Security	Internal-only data movement	✅
Availability	Task success uptime	99.9%
Maintainability	Centralized SQL/JS control	✅
Auditability	File trace coverage	100%

9. IMPLEMENTATION PHASES

Phase	Description	Deliverables
Phase 1: Foundation	Create DB, schemas, file formats.	Setup scripts
Phase 2: Ingestion & CDC	Build RAW tables, Streams, MERGEs.	Working CDC pipeline
Phase 3: Dynamic Tables	Deploy BI layer.	Auto-refresh DWH tables
Phase 4: Monitoring	Add logs, alerts, dashboards.	Monitoring schema
Phase 5: UAT & Go-Live	End-to-end testing & deployment.	Signed-off release

10. SUCCESS METRICS

KPI	Definition	Target
Data Latency	File arrival → analytics visibility	<15 minutes
Data Accuracy	Source vs. CORE match	>99.9%
Pipeline Reliability	Task success rate	>99%
Duplicate Prevention	File hash validation	100%
Alert Responsiveness	Error alert time	<5 minutes

11. TOOLS & TECHNOLOGIES

Category	Technology Used
Cloud Platform	Snowflake Enterprise Edition
CDC Engine	Streams + MERGE Procedures
Orchestration	Snowflake Tasks
Transformation	SQL + JavaScript Stored Procedures
Monitoring	PIPE_LOG, PROCESSED_FILES, ALERT_QUEUE
Visualization	Tableau / Power BI / Streamlit

12. RISK ASSESSMENT & MITIGATION

Risk	Impact	Mitigation
Stream Overflow	Missed CDC data	Schedule merges often
Duplicate File	Redundant load	MD5 hash tracking
Task Failure	Missed load	Retry + Alerts
High Volume	Cost spike	Auto-suspend policy
Data Quality	Wrong source data	Validation checks

13. DELIVERABLES

- Full SQL Build Script: All schemas, tables, and tasks
- JS Orchestrator: Orchestration procedure
- Monitoring Framework: Logging and alerts
- Snowpipe REST Example: JWT integration
- CSV Datasets: 15 entity datasets
- Runbook: Operational guide
- BRD Document: This report

14. GLOSSARY OF TERMS

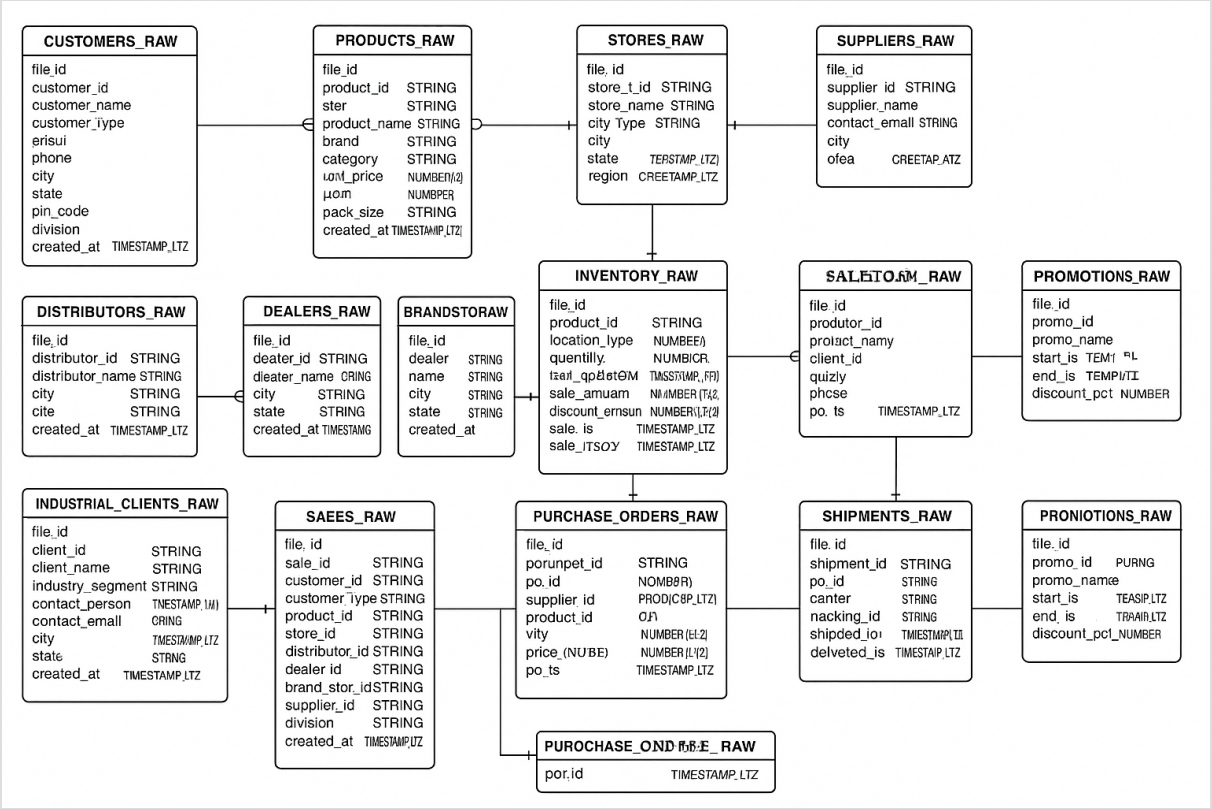
Term	Definition
CDC (Change Data Capture)	Captures incremental data changes.
Dynamic Table	Automatically refreshes dependent data.

Term	Definition
Stream	Tracks changes (insert/update/delete).
Task	Schedules SQL or JS jobs.
Stage	Snowflake internal storage for data.
Snowpipe	Continuous ingestion mechanism.
MERGE	Upsert operation for CDC.
MD5 Hash	File integrity and uniqueness check.
Monitoring Layer	Logs, alerts, and file metadata schema.
Pipeline Orchestrator	Procedure managing the ELT flow.

15. APPROVAL SIGN-OFFS

Role	Name	Signature	Date
Business Sponsor	—		
IT Data Head	—		
Data Engineering Lead	—		
BI / Analytics Lead	—		

APPENDIX: DATA MODEL OVERVIEW



Confidential – Prepared by Ankit Sharma • ankitsharma@bergerpaints.in