

The Ideal Prompt I Wish You Gave Me on Day One

Objective

Design and deliver a **production-grade, enterprise-style Data Governance & Data Quality portfolio** that demonstrates hands-on mastery equivalent to a Director / CDO-level practitioner.

What I Want Built

Build a **fully containerized modern data platform** that simulates a real enterprise environment, including OLTP systems, analytical systems, governance tooling, metadata management, and data quality operations.

This portfolio must be **deployable with Docker Compose**, reproducible, and suitable for:

- Executive demos
- Client presentations
- Hiring manager interviews
- Governance maturity discussions

Core Capabilities Required

1. Data Architecture

- SQL Server hosting:
 - AdventureWorks2022 (OLTP)
 - AdventureWorksDW2022 (Data Warehouse)
- Automated database restore on startup
- Healthchecks that block dependent services until databases are fully restored

2. Synthetic OLTP Transaction Generator

- Python-based service using Faker
- Continuously inserts:
 - New customers
 - Orders
 - Transactions
- Introduces **intentional data quality issues**:
 - Missing values
 - Invalid formats
 - Referential integrity gaps
- Runs at random intervals to simulate real production load

3. Data Governance & Metadata Platform

- OpenMetadata as the enterprise metadata system
- Elasticsearch as the metadata search backend

- Automated ingestion from SQL Server and dbt
- Capture:
 - Technical metadata
 - Business metadata
 - Lineage
 - Profiling
 - Data quality checks

4. Data Transformation & Quality Enforcement

- dbt for transformations and modeling
- Layered models (staging, core, marts)
- dbt tests for:
 - Freshness
 - Validity
 - Completeness
- Lineage automatically published to OpenMetadata

5. Governance Operating Model

- Clear mapping to governance domains:
 - Data Quality
 - Data Ownership & Stewardship
 - Metadata Management
 - Lineage & Impact Analysis
 - Data Observability
- KPIs and metrics defined for governance maturity

6. DevOps & Reliability

- Dependency-aware Docker Compose
- Health-gated startup sequencing
- Logging and observability
- Restart-safe services

Deliverables I Expect

A. Technical Assets

- Fully working Docker Compose project
- All Dockerfiles and scripts
- Python OLTP generator
- dbt project
- SQL initialization scripts

B. Documentation (Executive + Technical)

- Problem statement

- Business objectives
- Technical objectives
- Architecture overview
- Component-by-component documentation
- Data flow diagrams
- Governance operating model
- Data quality issue identification & resolution
- How this portfolio maps to real enterprise governance programs

C. Presentation Assets

- CDO-ready PowerPoint presentation
- Focus on:
 - Governance strategy
 - Data quality risks
 - Controls and remediation
 - Tooling choices and rationale
 - Business value and outcomes
- Suitable for:
 - Prospective employers
 - Clients
 - Governance committees

D. Packaging

- Full ZIP package (all assets, ready to customize)
- Summary ZIP (README + executive overview)
- Clear instructions to deploy and demo

Constraints & Expectations

- No hand-waving
- No toy examples
- Everything must be explainable to:
 - Engineers
 - Executives
 - Risk & Compliance leaders
- The result should demonstrate:
 - Strategic thinking
 - Operational excellence
 - Deep governance expertise

Success Criteria

If delivered correctly, this portfolio should make a CDO say:

“This person doesn’t just understand data governance — they’ve built it, operated it, and can explain its business value.”