This structured approach for implementing a **Cloud-Based Enterprise Data Architecture** is critical for a Big Four firm, as it links complex engineering work (Data Lake/Mesh) directly to business outcomes (self-service BI, AI enablement).

Here is the comprehensive action plan.

Comprehensive Action Plan: Cloud-Based Enterprise Data Architecture

Section	Content
Preamble/Role	Senior Partner, Big Four Consulting Firm. The company is a global retail corporation with highly fragmented data residing in dozens of legacy systems, severely limiting the ability of marketing and supply chain teams to run advanced analytics or use AI for inventory forecasting.
Core Mandate	Design a comprehensive action plan for implementing a Cloud-Based Enterprise Data Architecture (Data Lake/Mesh). The plan must focus on robust data ingestion, quality management, security, and establishing self-service analytics capabilities for 500 business users.
Objective	Achieve 80% data coverage of Tier 1 sources (Sales, Inventory, Customer Master) and enable self-service analytics for 500 business users by Month 15.
Compelling Why	The strategic imperative is Accelerated Decision Velocity and AI Readiness. The new architecture is projected to reduce the time-to-insight for strategic reports by 60% (from weeks to days), enable 10 new predictive AI/ML use cases (e.g., personalized promotions), and allow for the decommissioning of 5 legacy data marts, yielding an estimated 25% reduction in data infrastructure TCO over five years.
Approach	Phase 1: Data Strategy Alignment & Use Case Prioritization (Months 1-3): Finalize the Target Data Architecture (Lake or Mesh), conduct an inventory of existing data assets, and prioritize the Top 10 Tier 1 data sources and business AI use cases for initial migration. Phase 2: Architecture Design & Platform Build (Months 4-7): Design the security, ingestion, storage, and processing layers on the selected cloud provider. Build the foundational Cloud Data Platform and initial Dev/Test environments. Phase 3: Data Ingestion & Quality Setup (Months 8-12): Build and automate production data pipelines for the 10 prioritized sources. Implement data quality rules and establish the Enterprise Data Catalog (EDC) / Metadata Layer. Phase 4: User Onboarding & Self-Service Enablement (Months 13-15): Finalize security and access controls. Onboard and train the first 500 business users, transition development to the internal Data Engineering team, and establish the maintenance runbook.

Section	Content
Organization	Data Governance Council (DGC): Chaired by the CDO and CFO. Focuses on policy, compliance, and budget approval for data quality remediation. Chief Data Officer (CDO) Office: Centralized function responsible for enterprise data strategy, governance, and the platform roadmap. Data Engineering Team: Responsible for building and maintaining the core data ingestion and transformation pipelines. Analytics CoE (Center of Excellence): Responsible for training, tool standardization, and promoting self-service analytics among business users.
Processes & Governance	Data Quality Remediation Workflow: Implement a 3-stage workflow: 1) Automated detection (via platform rules), 2) Data Steward assignment (for root cause), and 3) Data Owner sign-off on the fix (system change). Metadata Management Standards: Mandate that all ingested data assets be tagged with five required attributes (Owner, Sensitivity, Lineage, Quality Score, Refresh Rate) and logged in the EDC before being released to business users. Secure Data Access Request Process: Implement a formalized, automated process (via the EDC) where users request access to specific data sets, requiring mandatory Data Owner approval and automated compliance checks, e.g., role-based access control (RBAC).
Key Deliverables	Phase 1: Target Data Architecture Blueprint, Prioritized 10 Data Sources/Pipelines List, Initial Security Policy Document. Phase 2: Cloud Data Platform Production Environment Ready, Initial Data Catalog (EDC) Deployed, Core Networking Security Controls Implemented. Phase 3: 10 Production Data Pipelines Automated, Finalized Data Quality Rule Sets, Enterprise Data Dictionary. Phase 4: 500 Users Onboarded and Trained, Self-Service BI Tool Stack Standardized, Maintenance Runbook.
Critical Risks & Mitigation	1. Budget Focus Shifting Away from Foundational Work Risk: Executives lose patience and demand AI products before the data foundation is stable. Mitigation: Tie 50% of the planned AI use case funding to the completion of Phase 3 (Data Quality Setup) to enforce funding of the foundation first. CDO must provide monthly "Health of the Data Platform" reports. 2. Security Misconfiguration in the Cloud Environment Risk: Improper setup of cloud storage or access controls leads to data breaches. Mitigation: Mandate third-party Cyber Audit of the cloud architecture security model before the Phase 3 launch and implement a "Policy as Code" approach to ensure consistent security across all environments. 3. Lack of Data Literacy Among Business Users Risk: Users misinterpret data or fail to use the new self-service tools correctly, leading to incorrect decisions. Mitigation: Implement a tiered, mandatory Data Literacy Certification Program for all 500 target users. Grant access to sensitive data only after certification is complete.

Section	Content
Change Management Plan	Strategy: Promote a "Data as an Asset" culture, shifting ownership from IT to the business. Incentivize Data Sharing: Publicly recognize and reward the Top 3 Data Owners (business leaders) who achieve the highest data quality scores and share the most valuable, compliant data. Communication: Emphasize the democratization of data and the opportunity for users to build their own insights, reducing reliance on IT/Analytics teams.
Crucial Additional Element	Key Data Quality Dimensions to be Measured: 1. Completeness: % of required fields populated for CDE records. 2. Accuracy: % of records matching the source of truth (e.g., validated address data). 3. Consistency: % of transactions using standardized codes/formats across systems. 4. Timeliness: Latency (age) of data ingested vs. the required refresh rate (SLA). 5. Uniqueness: % of records with unique keys (e.g., Customer ID).