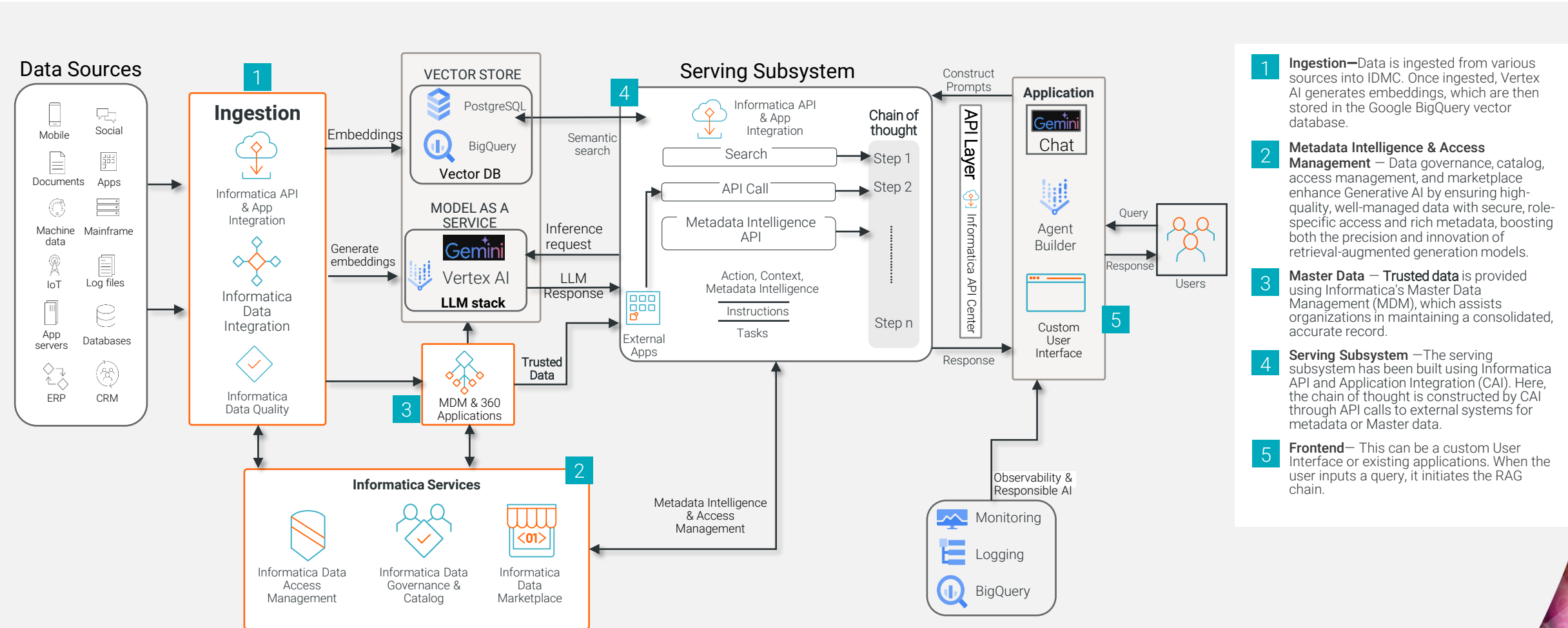


Reference Architecture

GenAI Blueprint for Google Cloud

Retrieval Augmented Generation Framework for GenAI with Informatica on Google Cloud



- 1 Ingestion**—Data is ingested from various sources into IDMC. Once ingested, Vertex AI generates embeddings, which are then stored in the Google BigQuery vector database.
- 2 Metadata Intelligence & Access Management** — Data governance, catalog, access management, and marketplace enhance Generative AI by ensuring high-quality, well-managed data with secure, role-specific access and rich metadata, boosting both the precision and innovation of retrieval-augmented generation models.
- 3 Master Data - Trusted data** is provided using Informatica's Master Data Management (MDM), which assists organizations in maintaining a consolidated, accurate record.
- 4 Serving Subsystem** —The serving subsystem has been built using Informatica API and Application Integration (CAI). Here, the chain of thought is constructed by CAI through API calls to external systems for metadata or Master data.
- 5 Frontend**— This can be a custom User Interface or existing applications. When the user inputs a query, it initiates the RAG chain.