DARA AI PLATFORM

Architecture & Delivery Roadmap

AI-Powered Analytics, NLP, and Compliance Intelligence

Document Type	Architecture & Roadmap
Version	1.0
Date	February 14, 2026
Classification	Internal – Confidential

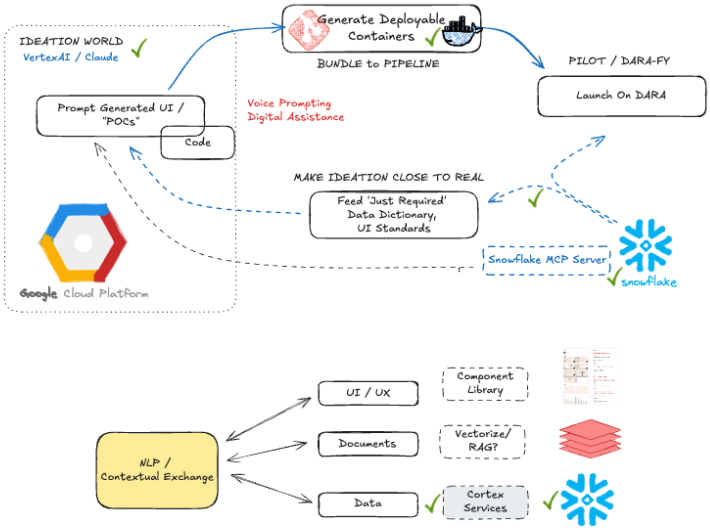
1. Executive Summary

This document outlines the architecture, current state, and delivery roadmap for the DARA AI Platform initiative. The platform leverages Claude, Google Vertex AI, and Model Context Protocol (MCP) integrations to deliver AI-powered analytics, natural language processing, regulatory compliance intelligence, and automated workflow capabilities for the DARA ecosystem.

The initiative targets two critical milestones: a Rocket Demo on April 1, 2026, and the Ignite presentation on May 9, 2026. This document serves as the authoritative reference for technical architecture decisions, capability status, and prioritized deliverables.

2. Current State Assessment

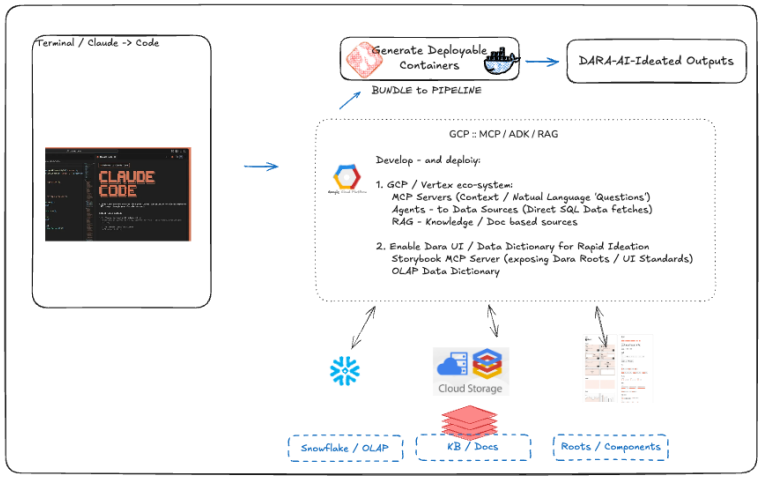
The following table summarizes the capabilities established to date, including their integration points, current maturity, and status.



Capability	Description	Integration	Status
Prompt-to-Ideation	Generative AI tooling enabling prompt-driven ideation workflows using Claude and Vertex AI	Claude / Vertex AI	**Completed**
Ideation based on OLAP Data	Providing limited OLAP Dictionary to Vertex and Claude, produced REAL DARA analytics panels	OLAP / Analytics	**Completed**
Snowflake MCP Server	Validated connectivity pipeline from Snowflake OLAP data sources to Vertex AI via MCP server and Agent framework	Snowflake MCP	**Completed**
Storybook MCP Server	MCP server integration with Storybook to generate code	Storybook MCP	**In Progress**

	aligned with DARA UI standards and component library		
RAG / Document Agent	GCP Document Agent for vectorization and retrieval-augmented generation across DARA User Guides, compliance documents, and regulatory materials	GCP Doc AI	**In Progress**
NLP Layer	Voice / Digital Assistant :: natural language processing layer on top of vectorized document stores for semantic query and response generation	Vertex AI	**In Progress**

3. Technical Architecture



3.1 Platform Components

The DARA AI Platform is built on a modular architecture that integrates multiple AI services and data sources through the Model Context Protocol (MCP) layer. The core components are organized into four tiers: the Data Layer (Snowflake OLAP/Aggregate Marts), the AI Services Layer (Vertex AI, Claude, GCP Document AI), the Integration Layer (MCP Servers, Agents), and the Presentation Layer (DARA UI, Voice/Digital Assistant).

3.2 Data Layer

The data layer provides the foundational data infrastructure for all AI-driven capabilities. Snowflake serves as the primary OLAP data warehouse, hosting Aggregate Marts that power analytics and compliance reporting. Connectivity to the AI Services Layer is established through a validated Snowflake MCP server, enabling direct data retrieval in response to natural language queries.

3.3 AI Services Layer

The AI Services Layer encompasses three primary functions. First, generative analytics powered by Claude and Vertex AI provide prompt-driven insights derived from live data. Second, the RAG pipeline leverages GCP Document AI to vectorize internal documentation (DARA User Guides) and external regulatory/compliance documents, enabling retrieval-augmented generation for contextual, evidence-based responses. Third, the NLP engine provides natural language understanding across data queries, document retrieval, and voice/digital assistant interactions.

3.4 Integration Layer

The MCP (Model Context Protocol) framework serves as the integration backbone, connecting AI services to data sources and UI components. Active MCP servers include the Snowflake MCP for data retrieval, the Storybook MCP for UI code generation aligned with DARA design standards, and the GCP Document Agent for document vectorization and retrieval.

4. Delivery Roadmap & Priority Deliverables

The following deliverables are organized against two key milestones: the Rocket Demo (April 1, 2026) and the Ignite presentation (May 9, 2026).

Workstream	Deliverable	Target
DARA Insights – Analytics	Prompt-generated analytics fetching data directly from OLAP / Aggregate Marts, enabling self-service analytical insights	May 9
NLP – Data Queries	Natural language querying over Snowflake data via MCP, returning structured and conversational responses	May 9
NLP – Document Intelligence	Semantic search and response generation across vectorized DARA and regulatory documents	May 9
NLP – Voice & Digital Assistant	Voice and conversational interface enabling hands-free interaction with platform capabilities	May 9
Compliance Dashboard – Rule Capture	Centralized capture and cataloging of all regulatory rules applicable to the portfolio	May 9
Compliance Dashboard – Test Generation	AI-generated actionable compliance tests derived from regulatory rules	May 9
Compliance Dashboard – Portfolio Analysis	Automated analysis of current portfolio for compliance impact assessment	May 9
Workflow Integration	Intercept call detail anomalies, trigger AI-generated workflow/communication events, and track resolution to closure	**April 1**

5. Milestone Detail

5.1 Rocket Demo – April 1, 2026

The Rocket Demo is the near-term priority milestone, focusing on demonstrating workflow integration capabilities. The demonstration will showcase the platform's ability to detect a call detail anomaly in real time, automatically trigger an AI-generated workflow or communication event in response, and provide end-to-end tracking and reporting on the anomaly resolution lifecycle. This demo serves as proof-of-concept for the platform's operational automation capabilities.

5.2 Ignite – May 9, 2026

The Ignite presentation is the strategic milestone targeting the full breadth of platform capabilities. All workstreams listed in Section 4 are targeted for this date. The presentation will demonstrate DARA Insights analytics, NLP across data and documents, the Regulatory and Compliance Dashboard, and the voice/digital assistant interface.

6. RAG & Document Intelligence Architecture

The Retrieval-Augmented Generation (RAG) pipeline is a critical component underpinning both the NLP and Compliance workstreams. The architecture follows a three-stage process.

6.1 Document Ingestion & Vectorization

Source documents are ingested from two primary corpuses: the DARA User Guide and associated operational documentation, and the Compliance/Regulatory document library. The GCP Document AI Agent processes these documents, extracting text and structural metadata, then generating vector embeddings stored in a vector database for efficient semantic retrieval.

6.2 Retrieval & Response Generation

When a user submits a natural language query, the system performs semantic similarity search against the vector store, retrieves the most relevant document chunks, and passes them as context to the language model for grounded response generation. This ensures that all responses are evidence-based and traceable to source documentation.

6.3 Compliance Application

The Regulatory and Compliance Dashboard leverages this RAG infrastructure to capture regulatory rules, generate compliance tests, and assess portfolio impact. By vectorizing the complete regulatory corpus, the platform enables natural language querying of compliance requirements and automated identification of portfolio exposure to regulatory changes.

7. Risks & Dependencies

Risk / Dependency	Description	Mitigation
Storybook MCP Maturity	MCP server for DARA UI code generation is still under active development	Parallel manual component development as fallback
RAG Vectorization Quality	Quality of vectorized embeddings directly impacts NLP accuracy and compliance coverage	Iterative tuning of chunking strategy and embedding models
April 1 Timeline	Workflow integration demo requires end-to-end anomaly detection, event triggering, and tracking	Early integration testing; scoped demo scenario
Data Access & Governance	Snowflake OLAP access patterns must comply with data governance and security policies	Collaborate with data governance team for access review

8. Next Steps

1. Finalize Workflow Integration demo scenario and begin end-to-end integration testing for the April 1 Rocket Demo.

2. Complete Storybook MCP server configuration to achieve DARA UI Standards compliance in generated code.
3. Expand RAG document corpus to include full regulatory library and validate vectorization quality.
4. Build and validate the Regulatory and Compliance Dashboard with rule capture, test generation, and portfolio analysis workflows.
5. Prepare integrated demonstration materials for Ignite (May 9) covering all four workstreams.