Reference Architecture Functional Development Challenge: Space-Based Data Acquisition, Storage and Dissemination

In the rapidly evolving domain of space-based data acquisition, the design and implementation of a robust data architecture present multifaceted challenges that must be addressed to ensure seamless collection, storage, processing, and analysis of data. As organizations venture into the utilization of satellite and space sensor data, it becomes imperative to overcome these challenges to harness the full potential of space-based information. CAL FIRE, the California Department of Forestry and Fire Protection, stands at the forefront of needing such advanced data architectures to effectively monitor, manage, and mitigate wildfires across the state.

As CAL FIRE integrates satellite and space sensor data into their operations, they face three critical challenges in data architecture. First, the diversity and volume of data sources and ingestion mechanisms require efficient and scalable solutions to handle real-time data streams. Second, the data storage and processing layers must be designed to manage vast amounts of data while ensuring data integrity, security, and accessibility. Lastly, the data consumption and analytics layers need to facilitate advanced analytics and decision-making processes, enabling CAL FIRE to respond swiftly and effectively to wildfire threats. These challenges highlight the necessity for a comprehensive and resilient data architecture that can support CAL FIRE's mission in safeguarding California's landscapes and communities.

Challenge 1: Data Sources and Ingestion Mechanisms

Objective: Architect, design, develop and prototype a versatile data ingestion mechanism that can handle batch, real-time, and streaming data from various sources, ensuring minimal latency and maximum fidelity.

In the realm of space-based data acquisition, one of the foremost challenges is to design and implement robust data ingestion mechanisms that can effectively handle a multitude of data sources and types. This challenge necessitates the development of versatile systems capable of processing batch data, real-time data streams, and continuous data inputs with minimal latency and maximum fidelity.

The ingestion mechanisms need to accommodate various data formats and sources, including structured data from databases, semi-structured data from logs and APIs, and unstructured data. It is essential to implement error handling and data validation protocols to maintain data quality throughout the ingestion process. This includes designing scalable data pipelines that can adapt to varying data loads.

Deliverables: CAL FIRE will accept submissions starting Friday, August 22, 2025, and ending at midnight on Sunday, October 26, 2025.

Core Technical Deliverables

- **Architectural Blueprint**

  - High-level system architecture diagram

  - Data flow and component interaction overview

  - Justification of chosen technologies for latency/fidelity balance

- **Data Ingestion Prototype**

  - Source adapters/connectors for batch, real-time, and streaming inputs

  - Support for multiple data formats: structured, semi-structured, unstructured

  - Implementation of scalable pipelines

- **Latency & Fidelity Metrics Dashboard**

  - Visualization of data processing latency across ingestion modes

  - Fidelity checks and validation results for ingested data

Reliability & Scalability Assets

- **Error Handling & Validation Framework**

  - Data quality assurance modules

  - Protocols for schema validation, retries, deduplication, and fault tolerance

Documentation & Knowledge Share

- **Technical Documentation**

  - Setup instructions, API references, configuration files

  - Details on supported data formats and sources

- **User Guide**

  - Step-by-step guide for deploying and testing the mechanism

  - Screenshots, sample inputs/outputs

Prize for Challenges

CAL FIRE has secured a $50,000 cash prize from the Gordon and Betty Moore Foundation that will be paid by the Earth Fire Alliance for the winner of the challenge.

Judging Criteria and Methodology for Challenge

Introduction

The following document outlines the judging criteria and methodology for evaluating the deliverables of three distinct challenges. Each challenge focuses on a different aspect of data management, storage, and consumption. The judging criteria are designed to ensure that teams meet all the requirements of the deliverables in each challenge. A numeric scale will be used to judge each deliverable. The challenge will be judged by [Scott Gregory, Phil SeLegue, Ben Rogers, Ann Kapusta, Brian Collins, Sean McFadden, and Chris Anthony. CAL FIRE has not sponsored challenges like this before but based on anecdotal evidence and discussions with industry CAL FIRE expects approximately 100 participants.

Challenge 1: Data Sources and Ingestion Mechanisms

Objective

Architect, design, develop, and prototype a versatile data ingestion mechanism that can handle batch, real-time, and streaming data from various sources, ensuring minimal latency and maximum fidelity.

Deliverables and Judging Criteria - Architectural Blueprint

- High-level system architecture diagram (0-50 points)
- Data flow and component interaction overview (0-10 points)
- Justification of chosen technologies for latency/fidelity balance (0-10 points)

Data Ingestion Prototype

- Source adapters/connectors for batch, real-time, and streaming inputs (0-10 points)
- Support for multiple data formats: structured, semi-structured, unstructured (0-10 points)
- Implementation of scalable pipelines (0-10 points)

Latency & Fidelity Metrics Dashboard

- Visualization of data processing latency across ingestion modes (0-50 points)
- Fidelity checks and validation results for ingested data (0-10 points)

Reliability & Scalability Assets

- Error Handling & Validation Framework (0-10 points)
- Data quality assurance modules (0-10 points)
- Protocols for schema validation, retries, deduplication, and fault tolerance (0-10 points)

Documentation & Knowledge Share

- Technical Documentation (0-10 points)
- Setup instructions, API references, configuration files (0-10 points)
- Details on supported data formats and sources (0-10 points)
- User Guide (0-10 points)
- Step-by-step guide for deploying and testing the mechanism (0-10 points)
- Screenshots, sample inputs/outputs (0-10 points)

Methodology

Each deliverable will be judged on a scale of points based on the completeness, functionality, and quality of the submission. The total score for Challenge 1 will be the sum of the scores for each deliverable, with a maximum possible score of 250 points.