

AI-Powered Data Quality Monitoring

with Amazon Bedrock

Hackathon Project Submission

Built with AWS Bedrock, Lambda, and Serverless Architecture

Executive Summary

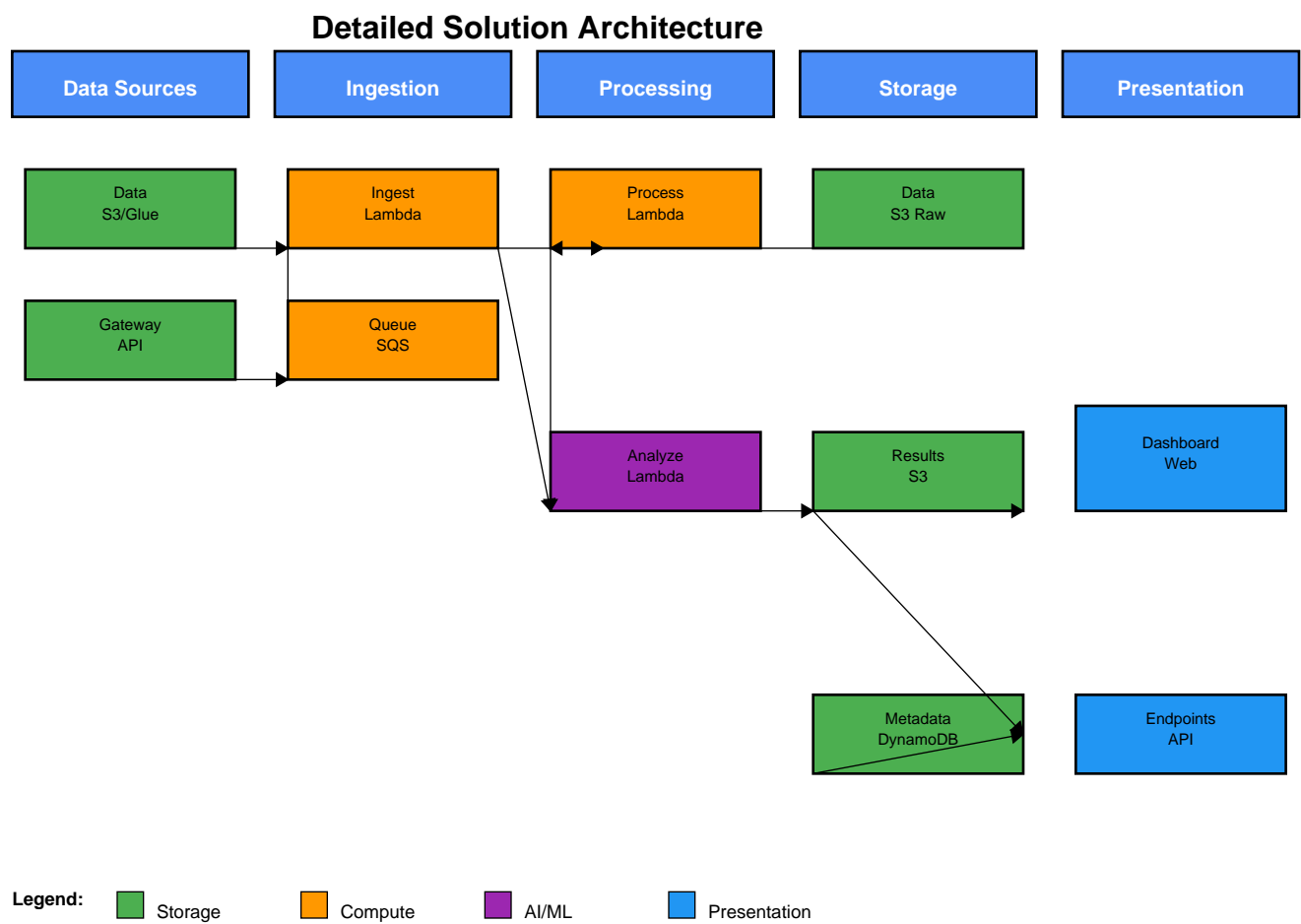
The AI-Powered Data Quality Monitoring solution addresses the critical challenge of maintaining high-quality data in modern data lakes. By leveraging Amazon Bedrock's Claude 2.1 model, this solution provides intelligent, automated data quality assessment and anomaly detection capabilities that go beyond traditional rule-based approaches.

Key Features:

- AI-powered data quality analysis with confidence scoring
- Automated anomaly detection and pattern recognition
- Interactive dashboard for real-time monitoring
- Serverless architecture for cost efficiency
- Seamless integration with AWS data services

Solution Architecture

The following diagram illustrates the high-level architecture of the solution:



Key Components:

- **Data Sources:** S3/Glue tables containing the data to be analyzed
- **Data Quality Pipeline:** AWS Lambda function that processes and analyzes data
- **Amazon Bedrock:** AI/ML service providing natural language understanding
- **Web Dashboard:** Interactive interface for monitoring and analysis

The solution is built on a serverless AWS architecture:

Component	Description
AWS Lambda	Core processing with Python 3.9 runtime
Amazon Bedrock	Claude 2.1 model for AI/ML capabilities
Amazon S3	Data storage and static website hosting
AWS Glue	Data catalog and schema management
Amazon Athena	Interactive query service
Amazon CloudWatch	Monitoring and logging

Technical Implementation

AI/ML Integration:

The solution leverages Amazon Bedrock's Claude 2.1 model for natural language processing and understanding. The system processes data quality metrics and generates human-readable insights with confidence scores.

Key Technical Components:

Data Quality Analyzer: Automated profiling and validation of data against quality rules

Anomaly Detection: Statistical analysis combined with AI to identify unusual patterns

Remediation Advisor: AI-powered suggestions for fixing data quality issues

Interactive Dashboard: Real-time visualization of data quality metrics

Low-Level Architecture

The following diagram shows the detailed technical architecture and data flow:

Component	Description	Technology
Data Ingestion	Collects and validates input data	AWS Lambda, S3 Events, SQS
Data Processing	Executes quality checks and analysis	AWS Lambda, Pandas, PySpark
AI Analysis	Processes data with Claude 2.1	Amazon Bedrock, Claude 2.1
Storage	Stores raw data and results	Amazon S3, Parquet/JSON
Orchestration	Manages workflow execution	AWS Step Functions
Monitoring	Tracks system performance	Amazon CloudWatch, X-Ray
API Layer	Handles external requests	Amazon API Gateway
Frontend	User interface for monitoring	HTML/JS, Chart.js, S3 Website

Data Flow:

1. Data is ingested from S3/Glue and queued for processing
2. Lambda functions process the data in parallel batches
3. Data quality rules are applied and metrics are collected
4. Anomalies are detected using statistical methods
5. AI analysis provides additional insights via Claude 2.1
6. Results are stored in S3 and indexed for quick retrieval
7. The dashboard pulls and visualizes the latest results

Interactive Dashboard



- ### Key Achievements:

- ### Future Enhancements:

- Expand support for additional data sources and formats
- Implement automated remediation workflows
- Add support for custom quality rules via UI
- Enhance anomaly detection with additional ML models