Architecture Diagrams

Traceability with Serverless Analytics

Traceability with Serverless Analytics: Architecture Diagrams

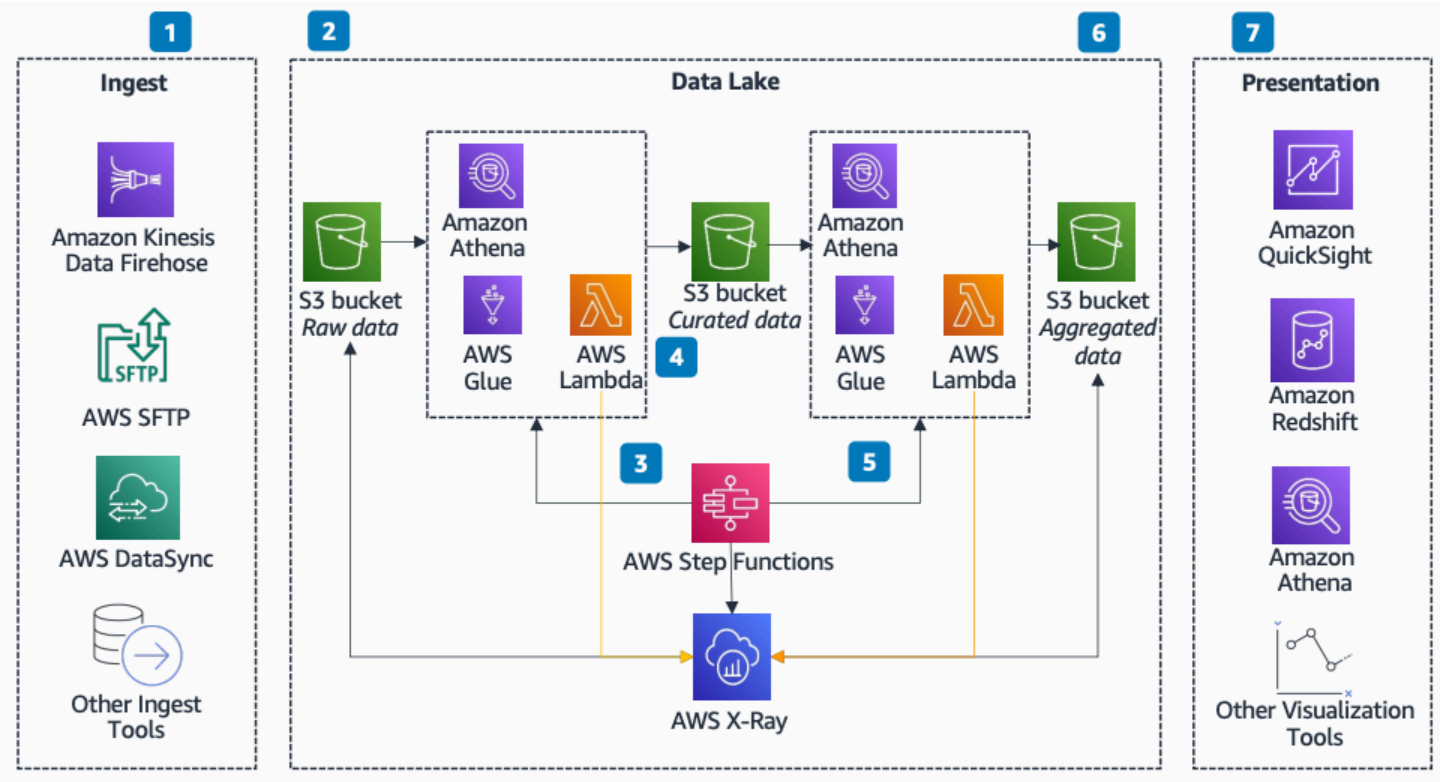
Table of Contents

Tracing Data Lake Ingest and Processing

Publication date: **July 15, 2021 (Diagram history)**

This architecture shows how to trace your data lake ingestion and processing using AWS X-Ray. You can emit traces from key checkpoints in the flow to diagnose data lake processing lifecycle issues and identify performance bottlenecks.

Tracing Data Lake Ingest and Processing Using AWS X-Ray



The following steps describe the architecture:

1. Diverse producers deliver data into the data lake built on Amazon Simple Storage Service.
2. The data first lands in an Amazon S3 bucket for raw data. This is the first trace since AWS X-Ray enables trace messages for Amazon S3 Event Notifications.
3. Data lake curation blocks prepare the data for analytics. AWS Step Functions triggers these processing blocks. AWS X-Ray integrates with Step Functions to help you identify performance bottlenecks and troubleshoot requests that resulted in an error.

4. AWS X-Ray can trace AWS Lambda functions. Lambda runs the X-Ray daemon and records a segment with details about the function invocation and results. For further instrumentation, bundle the X-Ray SDK with the function to record outgoing calls and add annotations and metadata.
5. Data curation continues with the Step Functions state machine triggering the next curation block. Like the first block, X-Ray can capture Step Functions and Lambda functions.
6. The data lands in an aggregated bucket ready for analytics. This is the last trace in the data lake ingestion flow. With traces emitting from all key checkpoints, you can diagnose data lake processing lifecycle issues.
7. Diverse analytics tools such as Amazon Quick Sight, Amazon Redshift, and Amazon Athena enable you to visualize the data from the data lake built on Amazon S3.

Further reading

For additional information, refer to the following resources:

- AWS Architecture Icons
- AWS Architecture Center
- AWS Well-Architected
- Quick editions

Diagram history

To be notified about updates to this reference architecture diagram, subscribe to the RSS feed.

Change	Description	Date
Initial publication	Reference architecture diagram first published.	July 15, 2021
Initial publication	Reference architecture diagram first published.	July 15, 2021

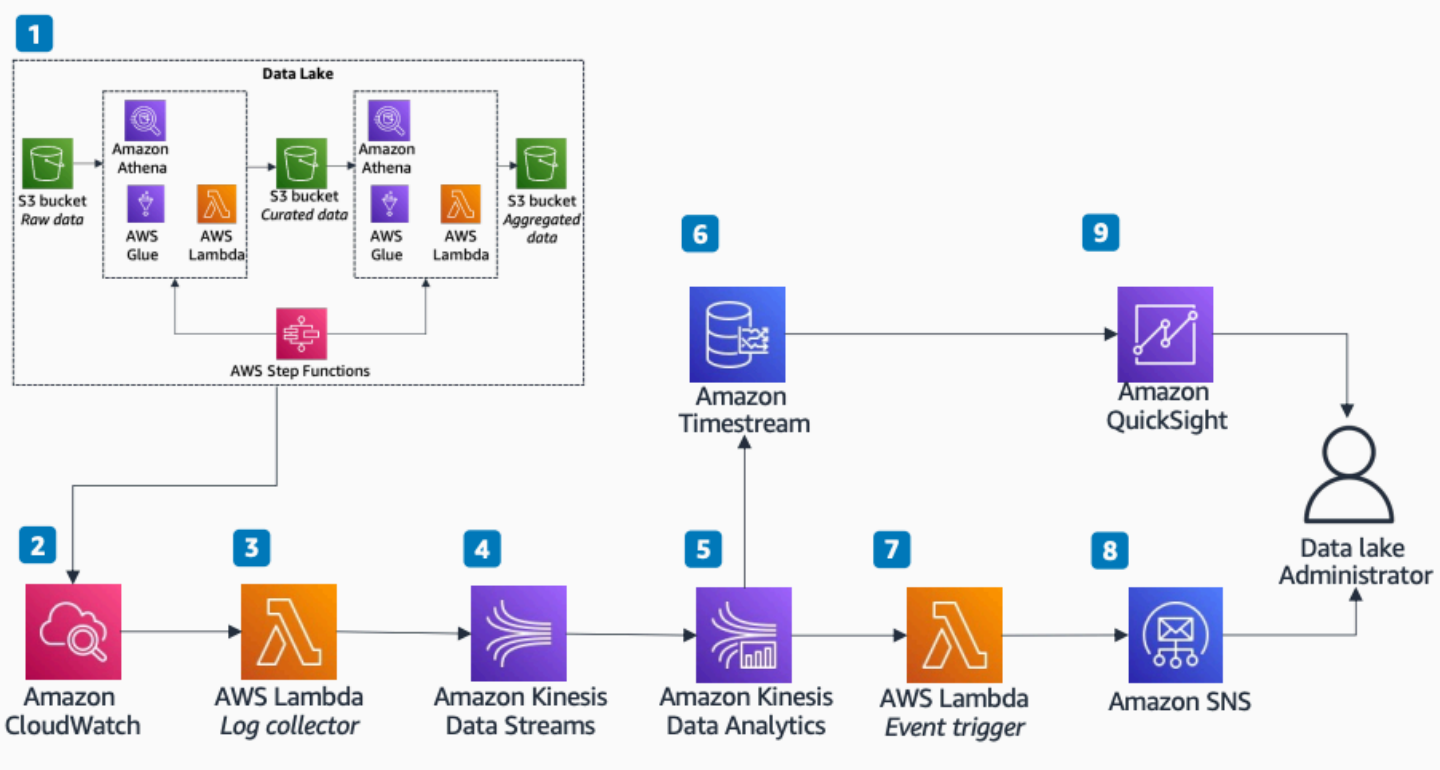
Note

To subscribe to RSS updates, you must have an RSS plugin enabled for the browser you are using.

Data Lake Ingest and Processing with Downstream Analytics

This architecture shows how to monitor data processing pipelines within the data lake using downstream analytics. You can publish relevant logs, metrics, and events to detect abnormal behaviors and visualize trends in near real time.

Data Lake Ingest and Processing with Downstream Analytics



The following steps describe the architecture:

1. Data processing pipelines within the data lake produce data. Any number of source pipelines can produce in this step.
2. The data pipeline components publish relevant logs, metrics, or events to Amazon CloudWatch.
3. An AWS Lambda function fetches data from CloudWatch and pushes pre-processed data to Amazon Kinesis Data Streams.
4. The Kinesis data stream is the central storage for near real-time streaming data.

5. Amazon Kinesis Data Analytics analyzes data in near real time and saves aggregated metrics to an Amazon Timestream database for further analysis. A Kinesis Data Analytics query can also detect abnormal behaviors of data in the pipeline.
6. Amazon Timestream supports time series analytics and defines time series as a native data type. It supports advanced aggregates, window functions, and complex data types such as arrays and rows.
7. The Amazon Kinesis Data Analytics threshold query sends near real-time analytics to Lambda.
8. Based on the processed data from Kinesis Data Analytics, the Lambda function can send real-time alerts using Amazon Simple Notification Service.
9. Amazon Quick Sight dashboards visualize trends such as number of successful pipelines, error rate, error type distribution, and more.

Further reading

For additional information, refer to the following resources:

- AWS Architecture Icons
- AWS Architecture Center
- AWS Well-Architected

Diagram history

To be notified about updates to this reference architecture diagram, subscribe to the RSS feed.

Change	Description	Date
Initial publication	Reference architecture diagram first published.	July 15, 2021
Initial publication	Reference architecture diagram first published.	July 15, 2021

Note

To subscribe to RSS updates, you must have an RSS plugin enabled for the browser you are using.