Audit Logs

Storing audit logs in AWS managed services

Research


Amazon Athena


Amazon Kinesis Firehose


AWS SQS


Amazon RDS


Amazon
DynamoDB


AWS Lambda


Amazon S3


Amazon DocumentDB


Amazon QuickSight

Where to store?

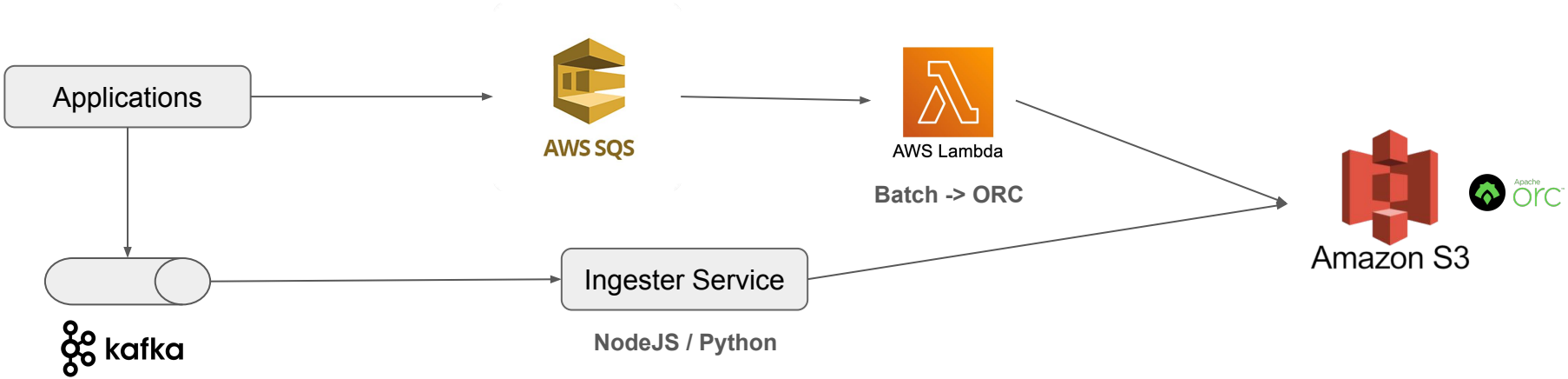
1. Cost efficient
2. Immutability using object lock
3. Secure using IAM policies
4. Performant if proper partitioning and file formats used

What to store?

1. ORC columnar file format
2. High write throughput
3. Better compression
4. Better query performance

How to store?

1. Kafka can be replaced with SQS as a streaming service
2. Applications push audit messages to SQS directly and only proceed execution on successful publishing
3. Can use lambda function to batch & convert logs to S3
4. Commit the messages in streaming service only on confirmation that the data is stored in S3 to ensure data durability
5. Use partitioning strategy Year/Month/Day/Hour



How to Query?

1. There are many query engines and integrations to query from Amazon S3
2. AWS based solutions
   a. Athena + Quick Sight
3. Self hosted solutions
   a. Trino + Apache Superset
   b. ClickHouse + Apache Superset

Next Steps

1. PoC
    a. Implement mechanism for applications to push the audits to SQS service instead of Kafka
    b. Implement Lambda function
    c. Consume messages from kafka, produce ORC files in S3
    d. Integrate query engine and UI
    e. Estimate the AWS cost per audit message
2. Standardize AWS audit stack
3. Think about maintaining two separate audit stacks for AWS and Onprem or aligning them…