



SYNOPSIS
ON
Build a log Analytics Solution on AWS

L

Submitted By-

Adesh Kumar (171500015)

Ishan Rathi (171500144)

Ayush Kumar Singh (171500075)

Submitted To-

Dr. Manoj Varshney

ACKNOWLEDGEMENT

We would like to express my sincere gratitude to my mentor- **Dr. Manoj Varshney** for providing their invaluable guidance, comments and suggestions throughout the course of the project. We would also thank our college in the project who helped us in making this project.

Which help us learning about virtual environment of **AWS** and helping us to build our mini project “**Build a Log Based Analytics Solution on AWS**”.

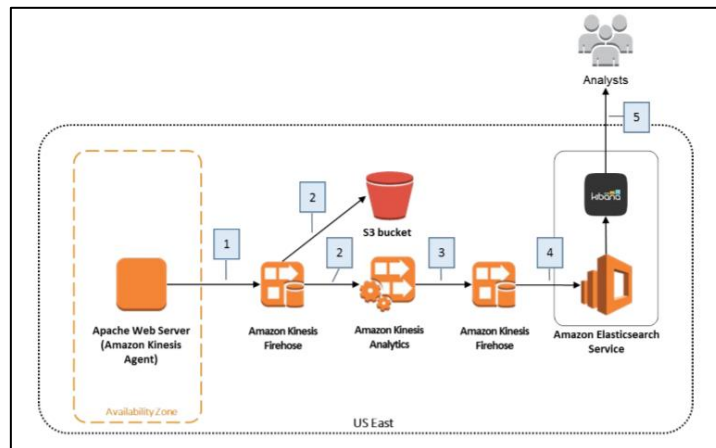
INTRODUCTION

Amazon Kinesis Analytics is the easiest way to process streaming data in real time with standard SQL without having to learn new programming languages or processing frameworks. Amazon Kinesis Analytics enables you to create and run SQL queries on streaming data so that you can gain actionable insights and respond to your business and customer needs promptly.

This tutorial walks you through the process of ingesting streaming log data, aggregating that data, and persisting the aggregated data so that it can be analyzed and visualized. You will create a complete end-to-end system that integrates several AWS services. You will analyze a live stream of Apache access log data and aggregate the total request for each HTTP response type every minute. To visualize this data in near real-time, you will use a user interface (UI) tool that will chart the results.

ARCHITECTURE

One of the major benefits to using Amazon Kinesis Analytics is that an entire analysis infrastructure can be created with a serverless architecture. The system created in this tutorial will implement Amazon Kinesis Firehose, Amazon Kinesis Analytics, and Amazon Elasticsearch Service (Amazon ES). Each of these services is designed for seamless integration with one another. The architecture is depicted below.



The web server in this example will be an Amazon Elastic Compute Cloud (EC2) instance. You will install the Amazon Kinesis Agent on this Linux instance, and the agent will continuously forward log records to an Amazon Kinesis Firehose delivery stream (step 1). Amazon Kinesis Firehose will write each log record to Amazon Simple Storage Service (Amazon S3) for durable storage of the raw log data (step 2), and the Amazon Kinesis Analytics application will continuously run an SQL statement against the streaming input data (step 2). The Amazon Kinesis Analytics application will create an aggregated data set every minute and output that data to a second Firehose delivery stream (step 3). This Firehose delivery stream will write the aggregated data to an Amazon ES domain (step 4). Finally, you will create a view of the streaming data using Kibana to visualize the output of your system (step 5).

USE OF THE PROJECT

- Just like Google trace us, in the same way we can also have the traces of every user of Amazon Web Services.
- We can design our product by analysing user logs.
- We can implement this on the streaming sites like Netflix, Ullu, AltBalaji, Amazon Prime Videos, etc.

FUTURE SCOPE

- In News Channels
- We can notice websites trends
- Log analytics is a common big data use case that allows you to analyze log data from websites, mobile devices, servers, sensors, and more for a wide variety of applications such as digital marketing, application monitoring, fraud detection, ad tech, gaming, and IoT.

FUNCTIONAL SPECIFICATION

Amazon EC2: Amazon EC2 provides the virtual application servers, known as instances, to run your web application on the platform you choose. EC2 allows you to configure and scale your compute capacity easily to meet changing requirements and demand. It is integrated into Amazon's computing environment, allowing you to leverage the AWS suite of services.

Amazon Kinesis Firehose: Amazon Kinesis Firehose is a fully managed service for delivering real-time streaming data to destinations such as Amazon S3, Amazon Redshift, or Amazon ES. With Firehose, you do not need to write any applications or manage any resources. You configure your data producers to send data to Firehose and it automatically delivers the data to the destination that you specified.

Amazon S3: Amazon S3 provides secure, durable, and highly-scalable cloud storage for the objects that make up your application. Examples of objects you can store include source code, logs, images, videos, and other artifacts that are created when you deploy your application. Amazon S3 makes it is easy to use object storage with a simple web interface to store and retrieve your files from anywhere on the web, meaning that your website will be reliably available to your visitors.

Amazon Kinesis Analytics: Amazon Kinesis Analytics is the easiest way to process and analyse streaming data in real-time with ANSI standard SQL. It enables you to read data from Amazon Kinesis Streams and Amazon Kinesis Firehose, and build stream processing queries that filter, transform, and aggregate the data as it arrives.

Amazon Elasticsearch Service: Amazon ES is a popular open-source search and analytics engine for big data use cases such as log and click stream analysis. Amazon ES manages the capacity, scaling, patching, and administration of Elasticsearch clusters for you while giving you direct access to the Elasticsearch API.

SOFTWARE SPECIFICATION

Technology Implemented: AWS, Cloud Computing, Virtualization

Language Used: Sql

Web Browser: Any Browser

HARDWARE REQUIREMENT

Processor: i3 Processor and above

Operating System: Windows

RAM: 4 GB and above

Hard Disk: 500 GB and above

CONCLUSION

Log analytics is a common big data use case that allows you to analyze log data from websites, mobile devices, servers, sensors, and more for a wide variety of applications such as digital marketing, application monitoring, fraud detection, ad tech, gaming, and IoT. In this project, you will use Amazon Web Services to build an end-to-end log analytics solution that collects, ingests, processes, and loads both batch data and streaming data, and makes the processed data available to your users in analytics systems they are already using and in near real-time. The solution is highly reliable, cost-effective, scales automatically to varying data volumes, and requires almost no IT administration.