Better Understand Web Log Data:

Leveraging Amazon Web Services in conjunction with Power BI to Fuel Real-Time Dashboards

Abstract:

Web log data contains an enormous amount of information and insights for an organization if utilized properly. Iranian e-commerce company Zanbil has concerns regarding its ability to effectively monitor online activity as it continues to grow. Additionally, it hopes to gain a better understanding of how its users are interacting with the website. Using Amazon Kinesis, Amazon Redshift, and Microsoft Power BI, we are able to provide Zanbil real-time dashboards using web log data in standard Apache form. Our solution provides visibility of the user experience and flags potentially fraudulent behavior as it happens.

Technology Pipeline:

Amazon Kinesis:

This service enables real-time ingestion and processing for streaming data. Through the Firehouse component, users have the ability to seamlessly integrate data streams with other AWS products including storage options like Redshift. Kinesis boasts low latency and an easily scalable solution for high throughput data.

In this project, web log data is uploaded into Kinesis streams in real-time and then passed to Redshift for storage.

Amazon Redshift:

Redshift is a cloud-based data warehousing platform. It has support for standard SQL and various BI applications. Using columnar storage format, it is able to handle data in the scale of petabytes. In addition, this service has massively parallel querying capabilities to select data from various views of the database simultaneously.

Our product utilizes Redshift to house web log data and its parallel querying functionality to update figures in real time.

Microsoft Power BI:

This platform fuels intuitive analytics and visualizations in a user-friendly interface. It is fully customizable to meet any business needs. Additionally, Power BI allows for multiple dashboards facing different functional groups. Since 2016, integration with data sources from AWS is supported.

In our project, streaming data read in Redshift is integrated directly with Power BI through DirectQuery. Graphics update automatically with new data.

Team No.
19

Noah Becker (becko910@umn.edu)
Sharang Jindal (jinda038@umn.edu)
Alok Kumar (kumar581@umn.edu)
Anuj Rewale (rewal006@umn.edu)
Patrick Seng (sengx021@umn.edu)
Pushkar Vengurlekar (vengu002@umn.edu)

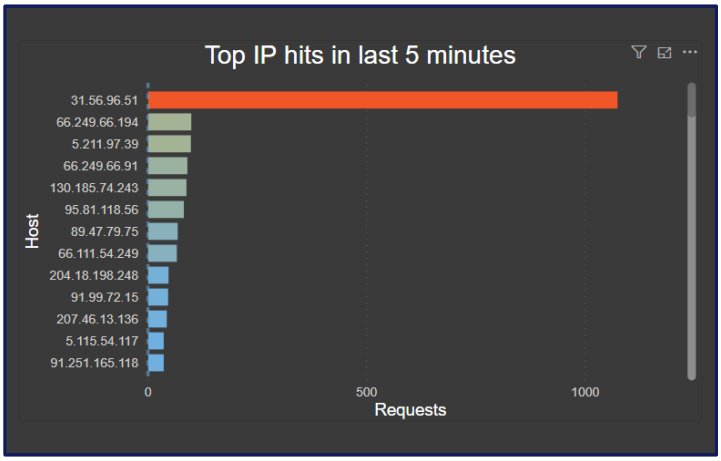
Better Understand Web Log Data:

Leveraging Amazon Web Services in conjunction with Power BI to Fuel Real-Time Dashboards

IT Dashboard:

Our solution offers an IT team the ability to monitor user web activity proactively; identifying and alerting the team with potential performance issues in addition to malicious behavior.

- Website Load Analyzer
- Top IP Hits
- Erroneous Requests
- Hits by Device Category



Marketing Dashboard:

Our solution enables a marketing team the ability to better understand changing behaviors and preferences of customers. Promotional and resource allocation decisions can be made based on various visualizations provided.

- Location of Users
- Top Products
- Popularity by Device

Product Comparison:

	Our Solution	LOGGLY	GoAccess
Streaming Support	✓	✓	✓
Dashboarding Capabilities	✓	✓	✓
Built-in Anomaly Detection	✓	✓	✗
Fully Customizable	✓	✗	✗
Petabyte Level Scalability	✓	✗	✗

Team No. 19

Noah Becker (becko910@umn.edu)
Sharang Jindal (jinda038@umn.edu)
Alok Kumar (kumar581@umn.edu)
Anuj Rewale (rewal006@umn.edu)
Patrick Seng (sengx021@umn.edu)
Pushkar Vengurlekar (vengu002@umn.edu)