

# Better Understand Web Log Data with Real-Time Dashboards Fueled by AWS and Power BI

Noah Becker, Sharang Jindal, Alok Kumar,  
Anuj Rewale, Patrick Seng, Pushkar Vengurlekar





# Project Definition

- Web log data contains an enormous amount of insights for organization across all corners of the business from IT to Marketing
- These insights can be applied across industries and provide important information related to security, products, and changing consumer preferences
- Our solution uses a combination of tools in a structured pipeline to **collect web log data, filter and analyze data, and produce interpretable, informative dashboards**
- All done in real time in order to provide visibility to both IT and marketing functions within the organization



# Problem Statement

- Web log data from Iranian e-commerce company Zambil
  - Data conforms to Apache standards
- Zambil hopes to gain a better understanding of how its users are interacting with the website
- Additionally, as they continue to grow, the IT staff has concerns with their ability to effectively monitor online activity





# Value

The value of a live web log analysis tool can be seen across an organization. Our design will support two critical components of any business:

## **IT Function:**

- Logging user activity across website
- Monitoring site performance
- Anomaly detection for potential attacks

## **Marketing Team:**

- Gain better understanding of customer profiles
- Track changing consumer preferences



# Competitive Landscape

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



# Methods





# Solution Pipeline

- Overall pipeline contains streaming collection, RDBMS based storage for the collected data, and platform for dashboarding and visualization
- These components are optimized to work seamlessly with one another across AWS and Microsoft platforms



**Web Log  
Data**



**amazon  
Kinesis**



**amazon  
REDSHIFT**



**Power BI**



**Personalized  
Dashboard**



# Amazon Kinesis

- Amazon Kinesis provides real-time ingestion and processing capabilities for streaming data
- Amazon Firehose is a powerful component of Kinesis that provides seamless integration with other AWS products including Redshift
- Web log data uploaded into Kinesis streams in real time







# Amazon Redshift

- Amazon Redshift is a cloud based data warehousing platform that supports standard SQL and BI applications
- Redshift can handle petabyte-scale using columnar storage
- From Firehose, web log data is stored in Redshift where parallel querying capabilities are utilized to enable simultaneous visualizations





# Microsoft Power BI

- Microsoft Power BI is an analytics and visualization tool fully adaptable and customizable to meet specific business needs
- Supports multiple dashboards facing various functional groups
- Power BI is directly integrated with our Redshift databases via DirectQuery, allowing for dashboards updated in real time





# Value



# IT Dashboard Overview

The IT dashboard contains important information allowing the IT team to monitor activity with four distinct graphical components:

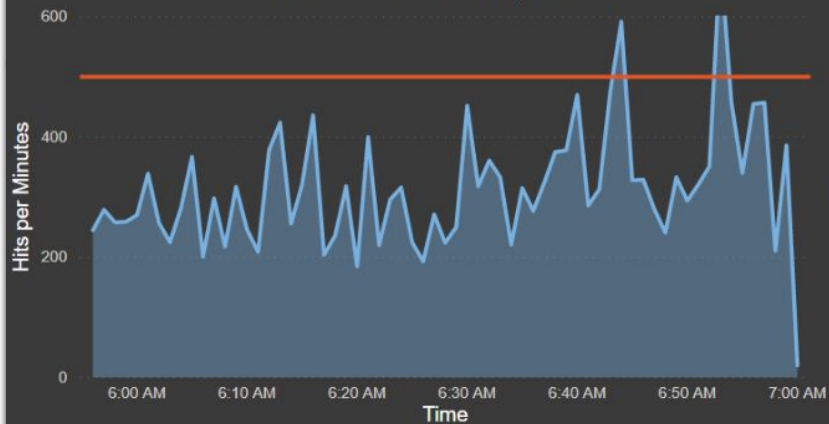
**Website Load Analyzer** - Tracks total traffic on website in real-time allowing IT team to proactively assess overload risks

**Top IP Hits** - Monitors hits for each IP address to track users that may have a suspicious amount of attempts in a certain timeframe

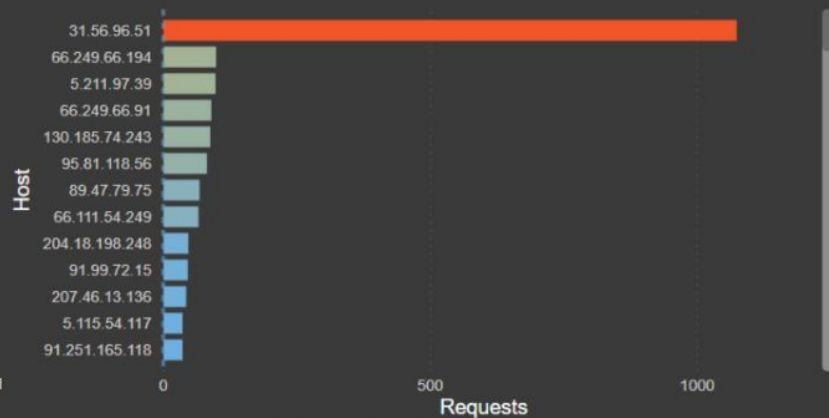
**Erroneous Requests** - Tracks status codes to determine where on the website users may be experiencing performance issues

**Hits by Device Category** - Displays most common device types used for accessing website allowing IT to optimize site for most common access platforms

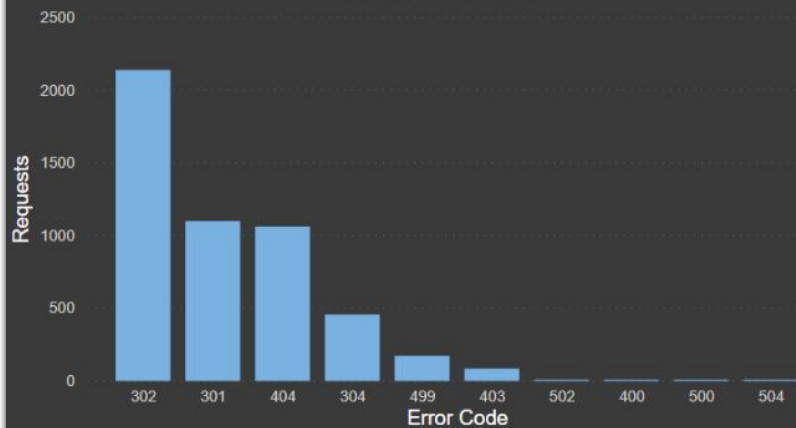
### Website Load Analyzer



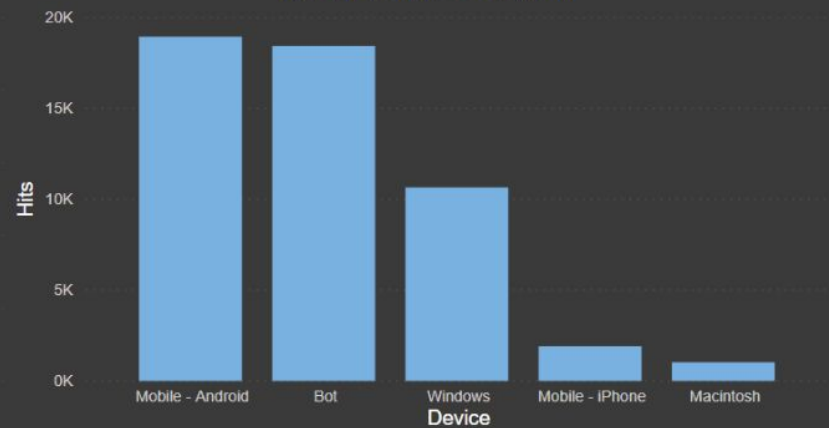
### Top IP hits in last 5 minutes



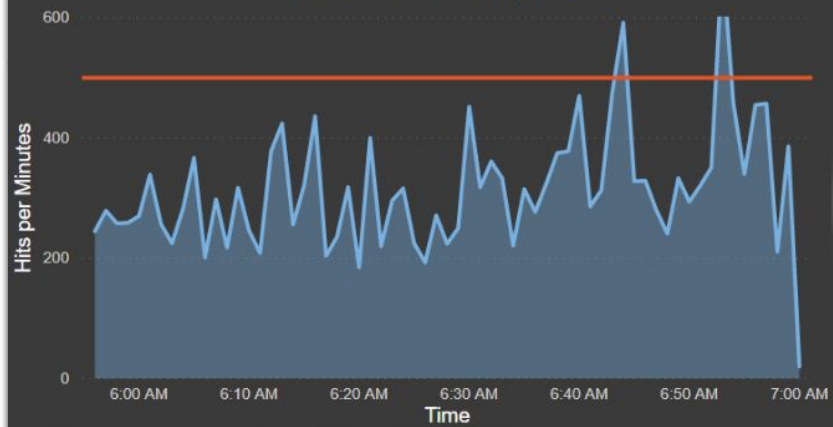
### Erroneous requests



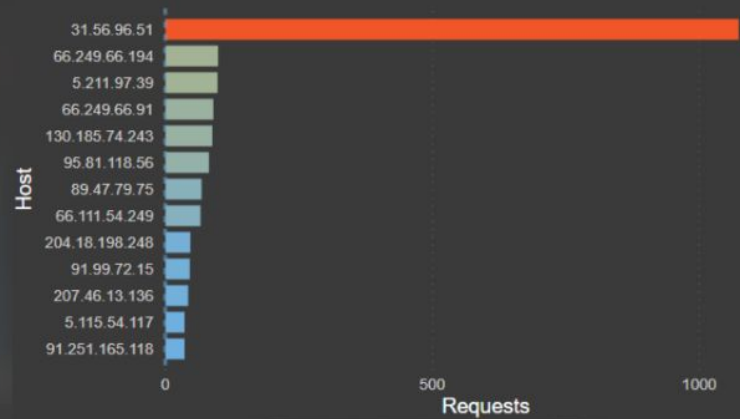
### Hits by Device Category

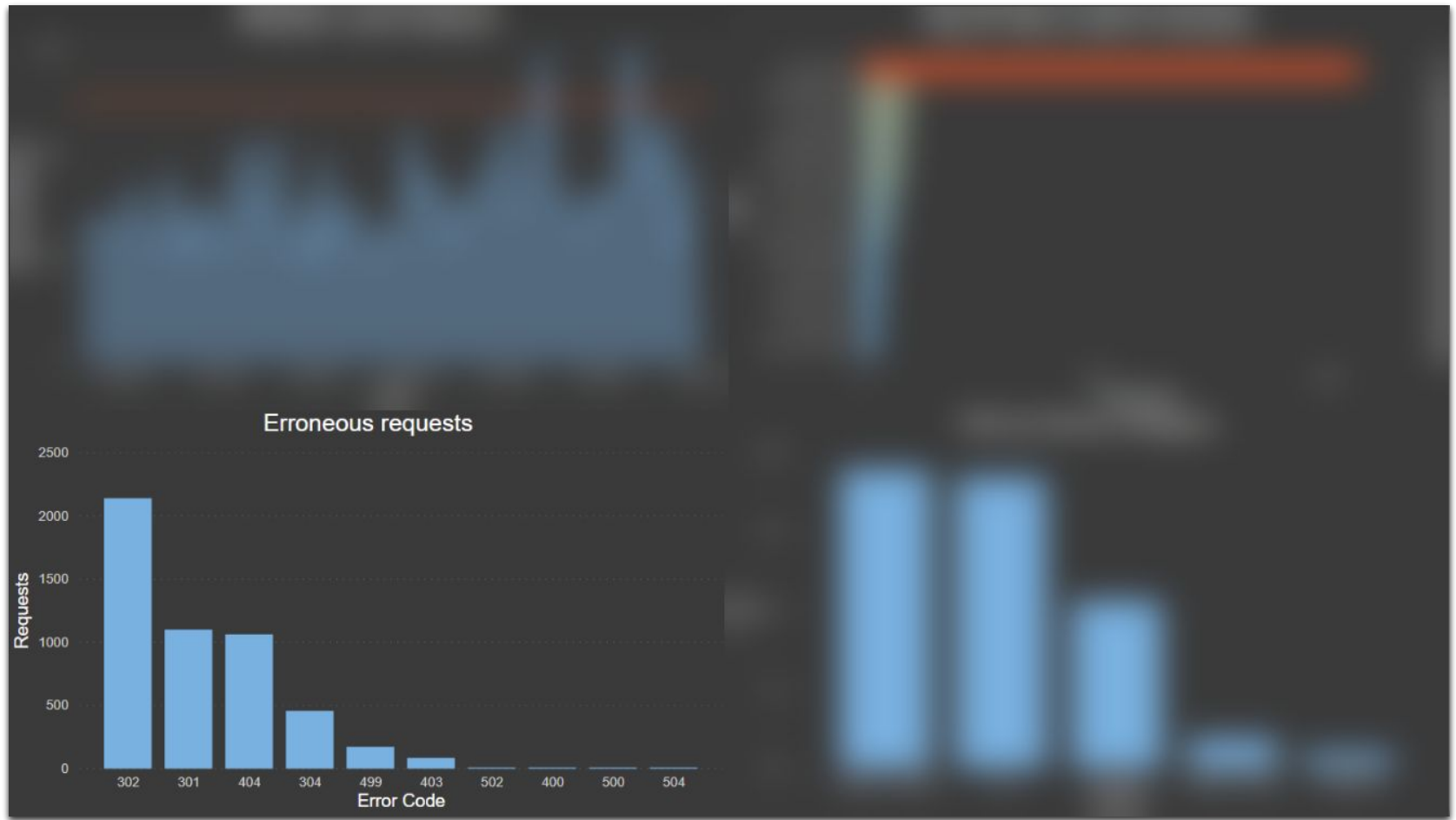


## Website Load Analyzer



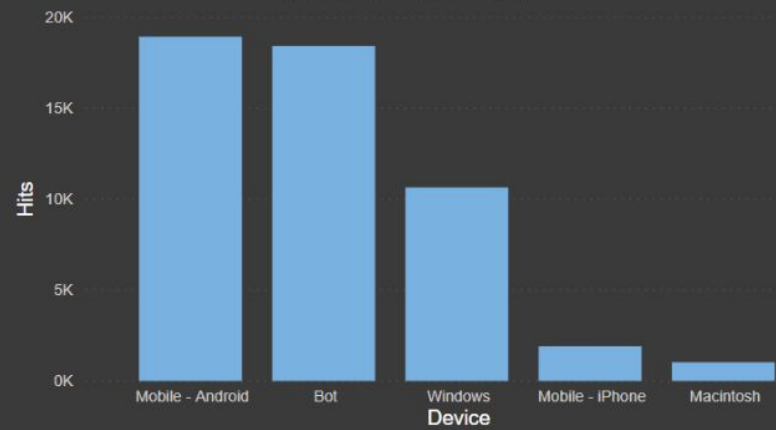
### Top IP hits in last 5 minutes







Hits by Device Category





# Marketing Dashboard Overview

The Marketing dashboard contains important information that enables the marketing team to track consumer attributes and preferences:

**Location of Users** - An interactive map of user access locations provides the business important information that can drive marketing decisions

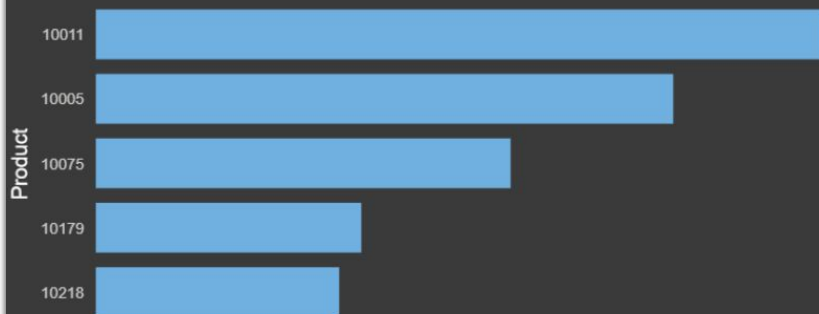
**Top Products** - Most viewed products on the website can help the business understand consumer preferences beyond their purchases alone

**Popularity by Device** - Provides marketing team an overview of mobile vs. desktop access volumes which can help them make resource allocation and promotional decisions

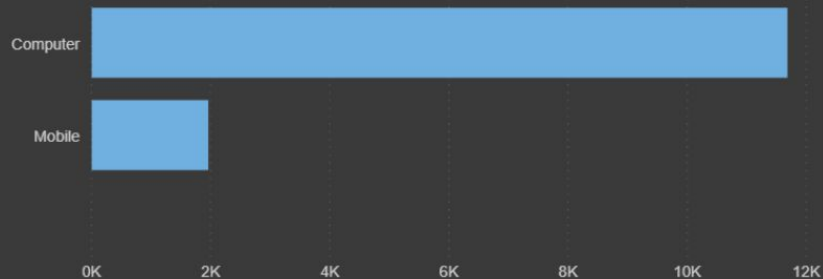
## Location of Users



## Top Products

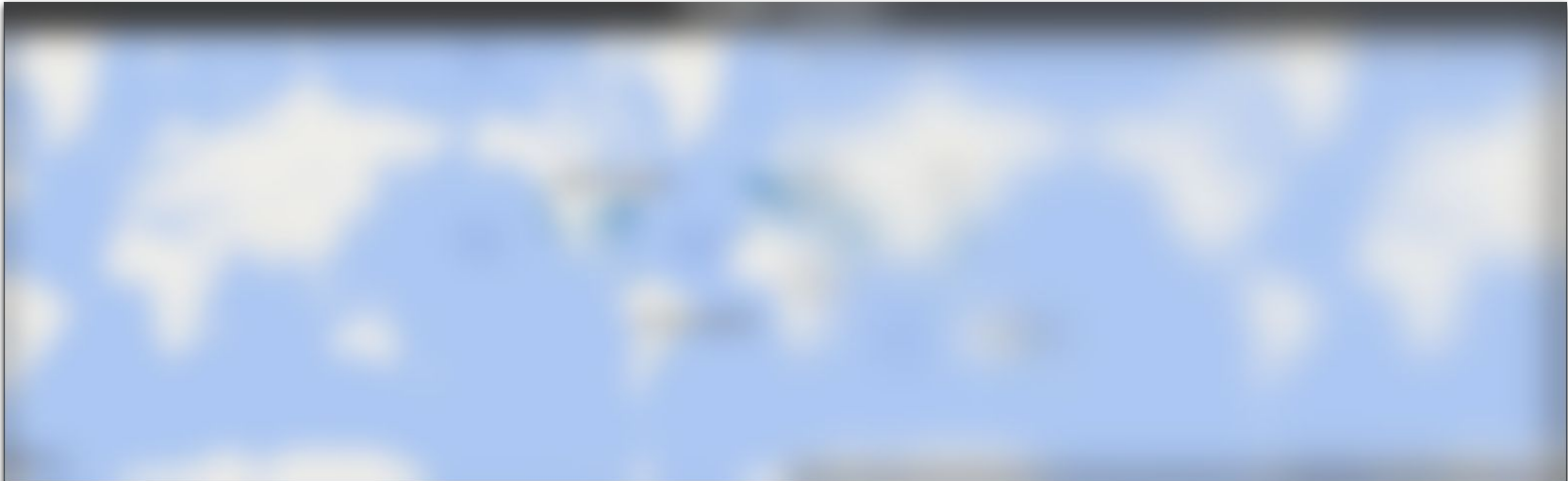


## Popularity by Device

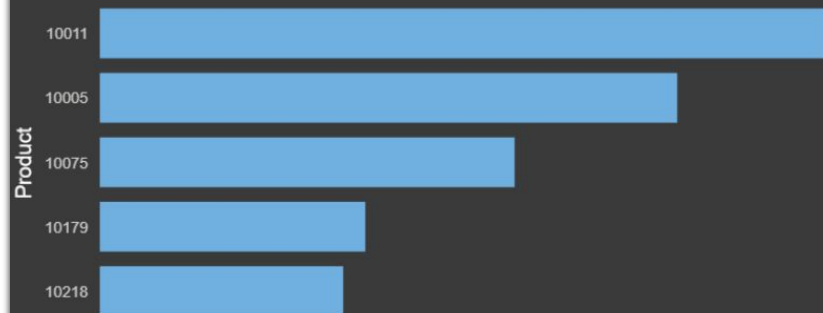


## Location of Users





Top Products







# Conclusion





# Business Impact

Our solution helps the IT team monitor their website proactively, identifying potential performance issues and malicious behavior

Comprehensive dashboards enable the marketing team to better understand their customers' changing behavior and preferences

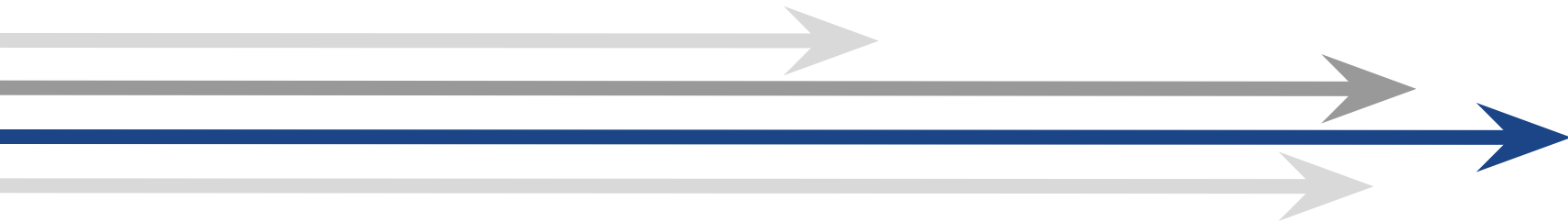






## Next Steps

- Extend core functionality to other business segments such as finance, supply chain, and management.
- Additional features:
  - Customizable trigger alerts based on specified conditions
  - Enhanced anomaly detection for other fraudulent behaviors
  - Clickstream flow visualizations to better identify customer pain points in the online experience





# References

<https://aws.amazon.com/kinesis/>

<https://aws.amazon.com/redshift/>

<https://powerbi.microsoft.com/en-us/>

<https://www.loggly.com/>

<https://goaccess.io/>

<https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/3QBYB5>