AWS re:Invent

ANT212

Delivering data to your Amazon Elasticsearch Service domain

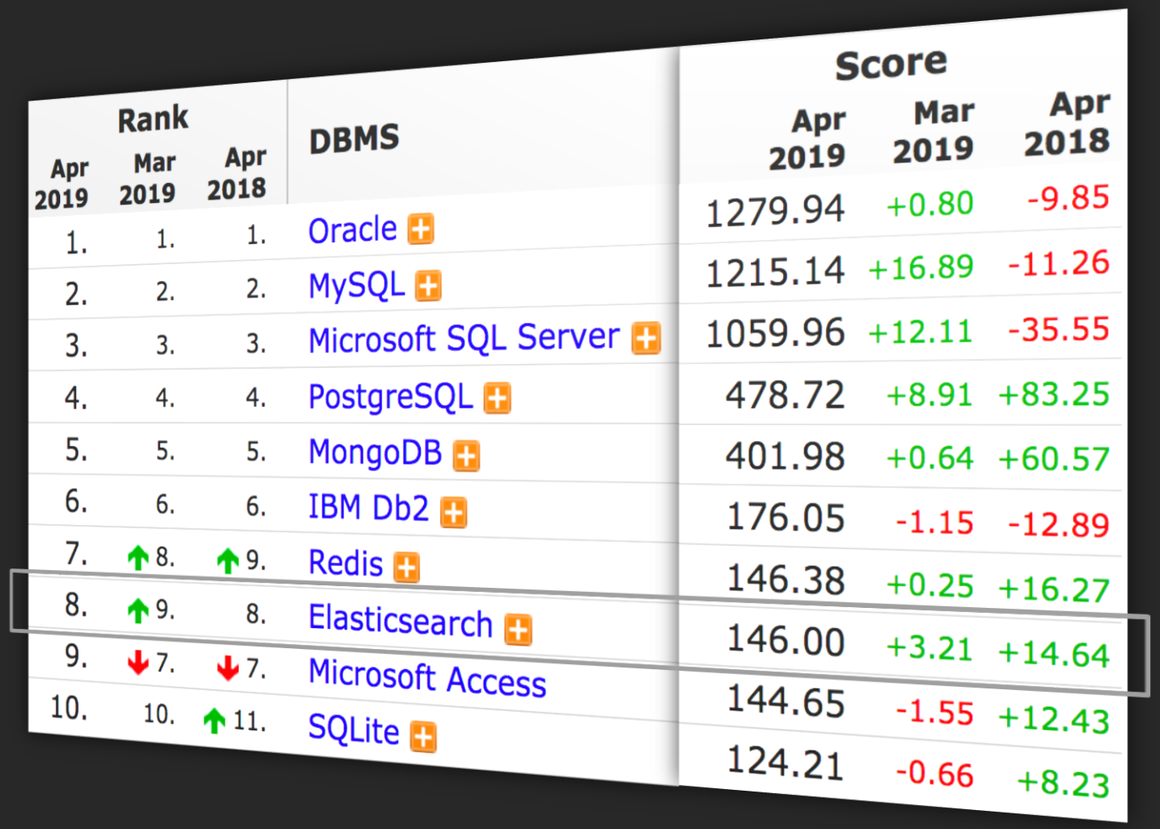
Bhaven Avalani
Senior Software Development Manager
Amazon Elasticsearch Service

Balaji Kannan
Software Development Manager
Amazon Elasticsearch Service

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What is Elasticsearch

- Sometimes referred to as the “ELK Stack” – Elasticsearch, Logstash & Kibana
- Distributed search and analytics engine built on Apache Lucene
- Easy ingestion and visualization
- Developed in Java



Rank			DBMS	Score		
Apr 2019	Mar 2019	Apr 2018		Apr 2019	Mar 2019	Apr 2018
1.	1.	1.	Oracle	1279.94	+0.80	-9.85
2.	2.	2.	MySQL	1215.14	+16.89	-11.26
3.	3.	3.	Microsoft SQL Server	1059.96	+12.11	-35.55
4.	4.	4.	PostgreSQL	478.72	+8.91	+83.25
5.	5.	5.	MongoDB	401.98	+0.64	+60.57
6.	6.	6.	IBM Db2	176.05	-1.15	-12.89
7.	↑ 8.	↑ 9.	Redis	146.38	+0.25	+16.27
8.	↑ 9.	8.	Elasticsearch	146.00	+3.21	+14.64
9.	↓ 7.	↓ 7.	Microsoft Access	144.65	-1.55	+12.43
10.	10.	↑ 11.	SQLite	124.21	-0.66	+8.23

Source: TechCrunch survey of popular open source software from April’17

Agenda – Ingestion into Amazon Elasticsearch Service

- Log Ingestion into Elasticsearch – The Why?
- The Collect-Transform-Buffer-Deliver(CTBD) Design Pattern – The How?
- Demo
- AWS/Open Source Options for CTBD
- Q&A

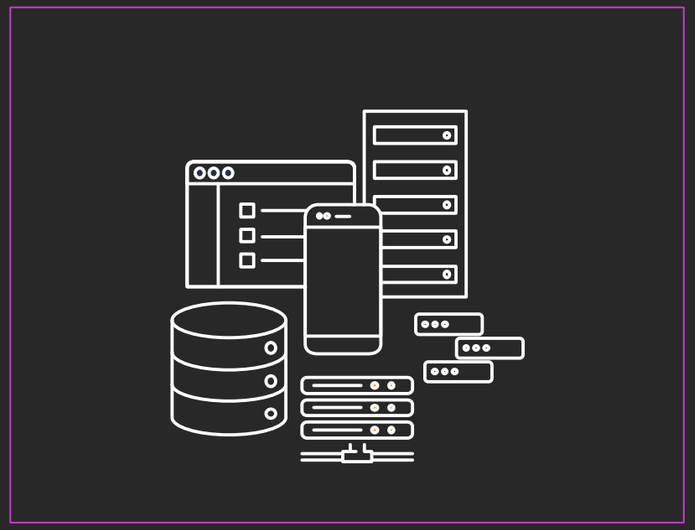
Log Ingestion - Why

Why

- 100s to 1000s of servers
- Emitting 10s of logs files each
- Containing operational(errors, latency, etc.) and security data (logins)
- Need for scalable grep and aggregate
  - And historical data access
  - And corelated data

Machine data driving Elasticsearch growth

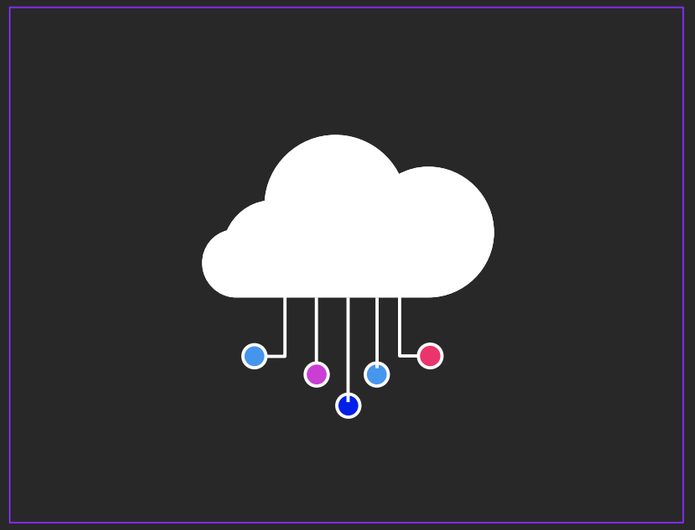
Machine-generated data is growing **10x faster** than business data… Logs, logs, and more logs



IT & DevOps: Databases, Servers, Storage, Networking



Increase in IoT and Mobile Devices: Gaming, Sensors, Web Content



Cloud-based architectures

Source: insideBigData—The Exponential Growth of Data, February 16, 2017

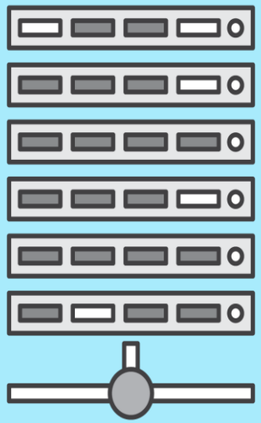
Log Ingestion – How?

Ingestion: Moving logs from your servers to Amazon Elasticsearch Service

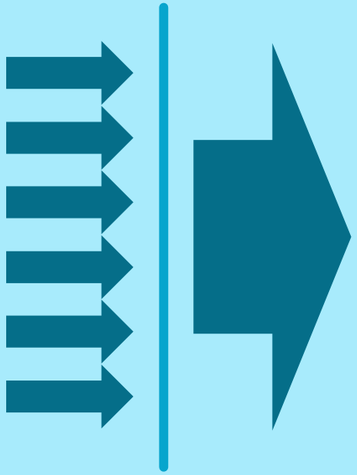
- Collection via agents, logging, etc
  - Syslog, Access Logs, etc.
- Transformation: Convert to a structured format
  - Parse the data: eg <%{POSINT:priority}>
  - Transform the data: kv filter to extract key-value pairs
- Buffer: Batch before inserting
  - Kinesis, Kafka, etc
- Deliver: Batch Load to Elasticsearch

Lab: Syslog and HTTP log ingestion

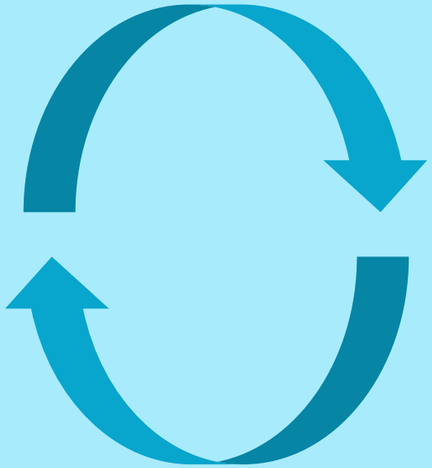
Key Idea



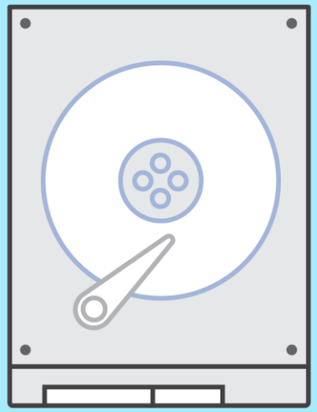
Logs:
http



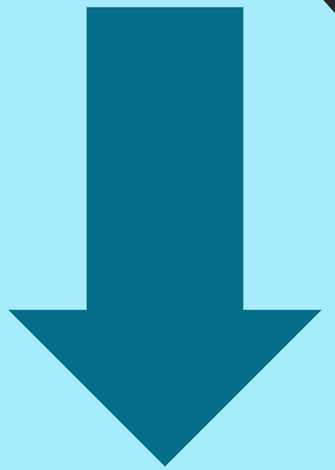
Collect:
Kinesis Agent



Transform:
Kinesis Data Firehose



Buffer:
Kinesis



Deliver

Log Ingestion – LAB

LAB : Ingest http logs onto elastic search

Log Ingestion – Key Thoughts

Ingestion : Moving logs from your servers to Amazon Elasticsearch Service

- Collect, Transform, Buffer, Deliver pattern
- Many AWS/Open Source Options Available for each step
  - Collect: Logstash, Beats, Fluentd, Kinesis Agent..
  - Transform: Logstash, AWS Lambda, AWS Kinesis..
  - Buffer: Kafka, Logstash, Amazon Kinesis, Amazon S3..
  - Deliver: Logstash, AWS Lambda, Amazon Kinesis,..
- Visualize using Kibana

Collect



Kinesis Agent



CloudWatch Logs Agent



Application



Logstash



Beats



Fluentd

Buffer



Kinesis Data Firehose



CloudWatch Logs



Amazon S3



Amazon Elasticache/Redis



Logstash



Kafka



Rabbit MQ

Transform



Kinesis Data Firehose



CloudWatch Logs



Logstash



AWS Lambda

Deliver



Kinesis Data Firehose



CloudWatch Logs



AWS Lambda



Logstash



Worker nodes

Open discussion

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Bhaven Avalani and
Balaji Kannan

avalani@amazon.com
kbalaji@amazon.com

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Please complete the session survey in the mobile app.

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.