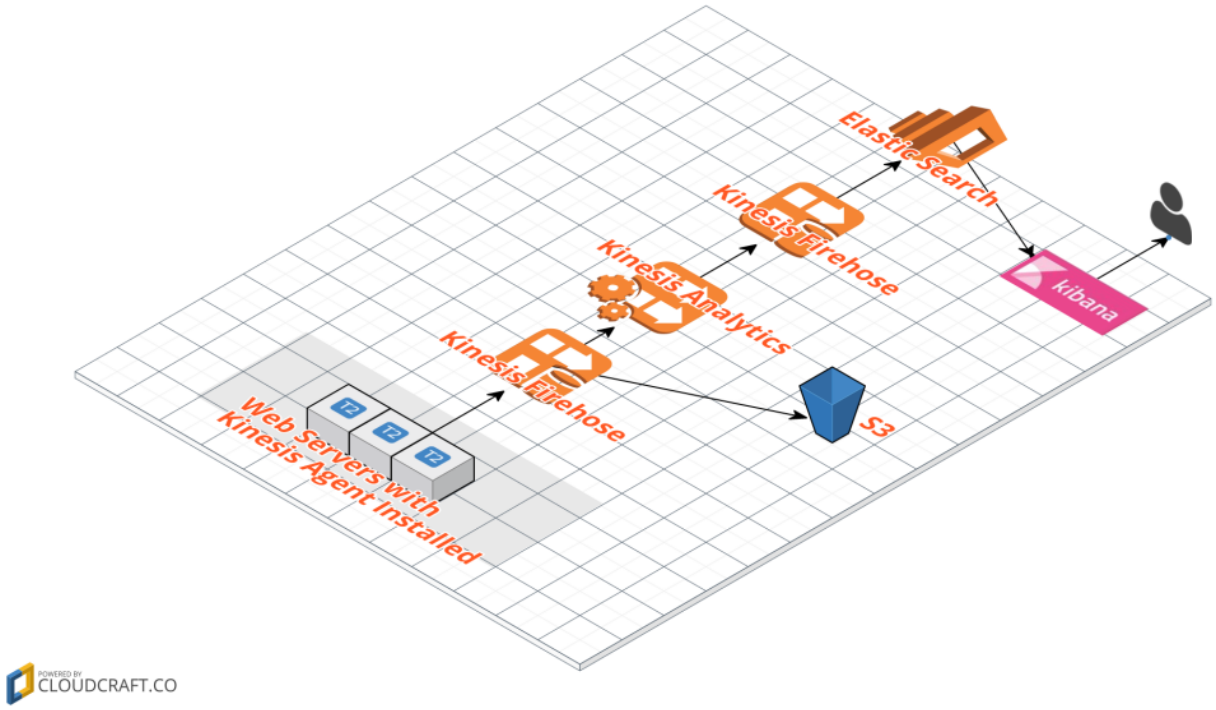
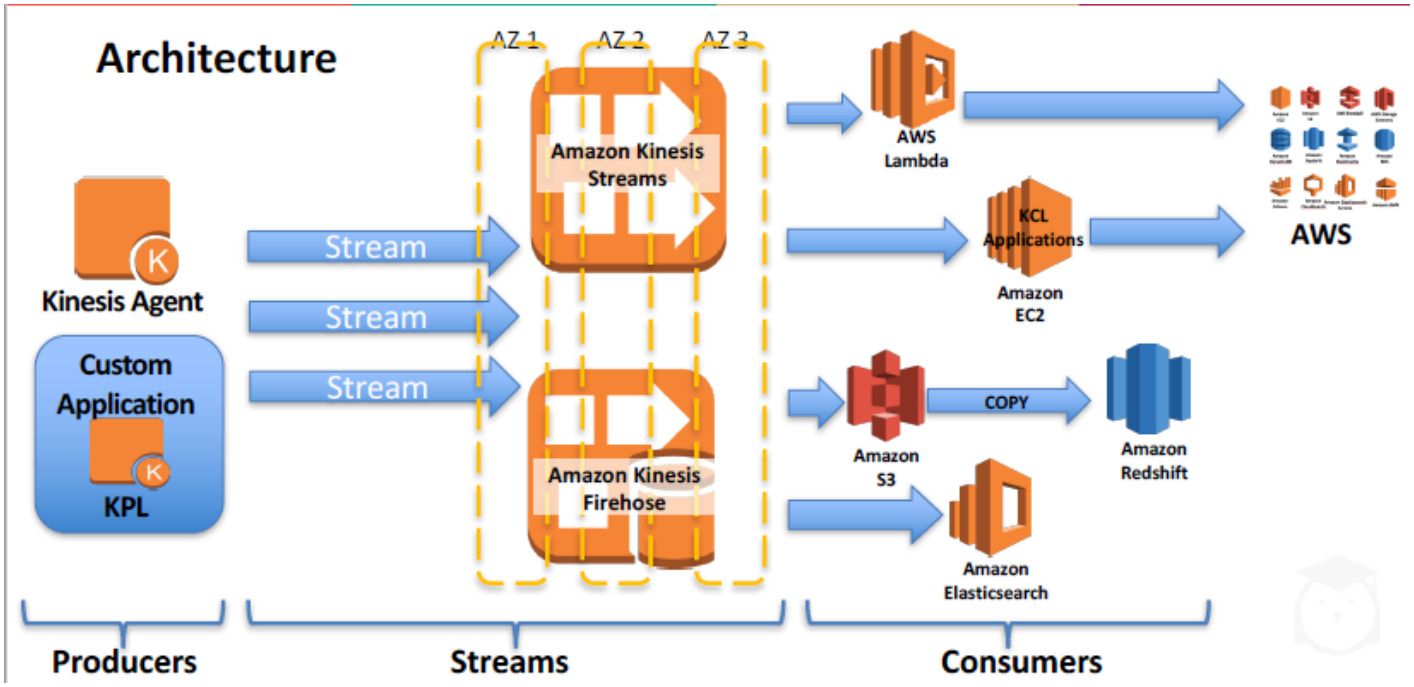
Apache Log Analytics - Kinesis Firehose, Analytics and Elastic Search

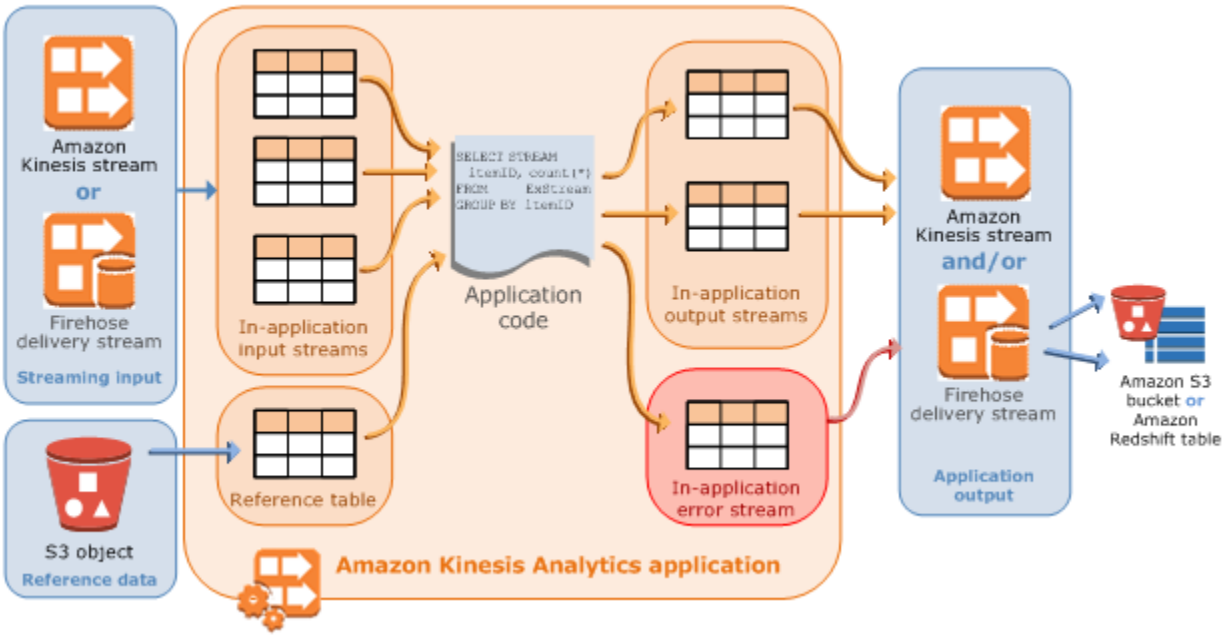


Kinesis Fireshose -



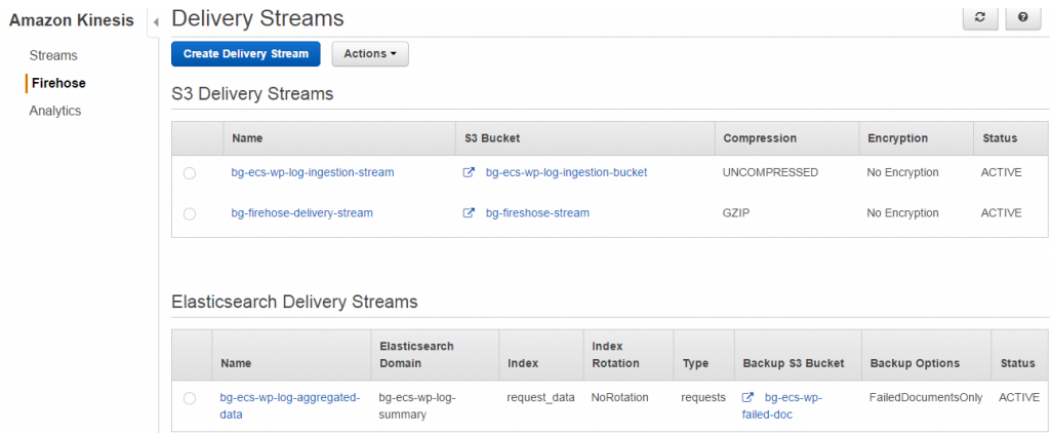
- Firehose can scale to gigabytes of streaming data per second, and allows for batching, encrypting and compressing of data.
- Easiest way to load streaming data into AWS.
- Firehose will automatically scale to meet demand.

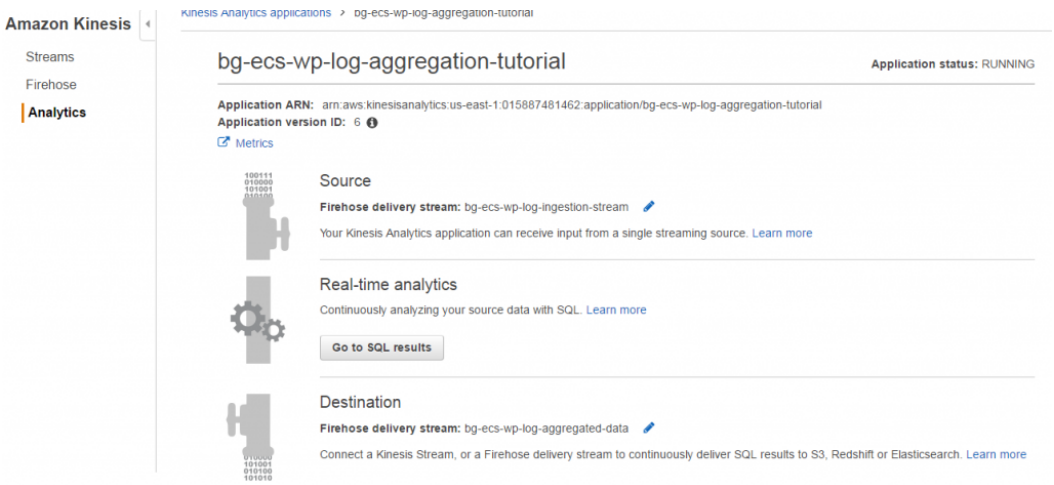
Kinesis Analytics -

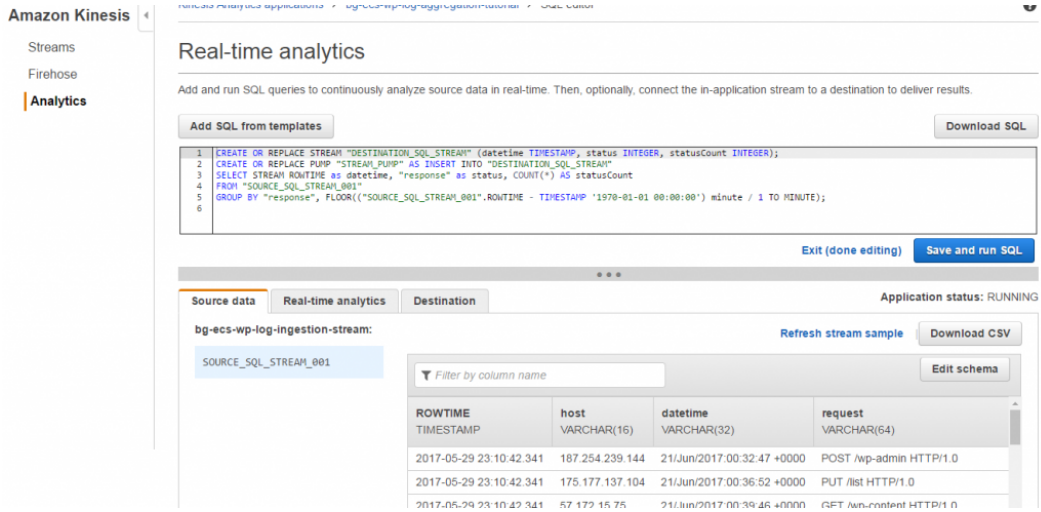


- Amazon Kinesis Analytics enables users to run standard SQL queries over live streaming data
- Readily query Kinesis Stream/Firehose data and export the output to destination like S3
- In-flight analytics, Standard SQL Interface, Elastic scaling up or down
- data dump to S3, ES, and Redshift
- Auto-schema detection
- Pre-build stream processing templates.

Follow Build a Log Analytics Solution on AWS Tutorial to perform this exercise.




Amazon Kinesis
Streams
Firehose
Analytics
Kinesis Analytics applications › bg-ecs-wp-log-aggregation-tutorial
bg-ecs-wp-log-aggregation-tutorial
Application status: RUNNING
Application ARN: arn:aws:kinesisanalytics:us-east-1:015887481462:application/bg-ecs-wp-log-aggregation-tutorial
Application version ID: 6
Metrics
Source
Firehose delivery stream: bg-ecs-wp-log-ingestion-stream
Your Kinesis Analytics application can receive input from a single streaming source. Learn more
Real-time analytics
Continuously analyzing your source data with SQL. Learn more
Go to SQL results
Destination
Firehose delivery stream: bg-ecs-wp-log-aggregated-data
Connect a Kinesis Stream, or a Firehose delivery stream to continuously deliver SQL results to S3, Redshift or Elasticsearch. Learn more


Amazon Kinesis
Streams
Firehose
Analytics
Real-time analytics
Add and run SQL queries to continuously analyze source data in real-time. Then, optionally, connect the in-application stream to a destination to deliver results.
Add SQL from templates
Download SQL
CREATE OR REPLACE STREAM "DESTINATION_SQL_STREAM" (datetime TIMESTAMP, status INTEGER, statusCount INTEGER);
CREATE OR REPLACE PUMP "STREAM_PUMP" AS INSERT INTO "DESTINATION_SQL_STREAM"
SELECT STREAM ROWTIME as datetime, "response" as status, COUNT(*) AS statusCount
FROM "SOURCE_SQL_STREAM_001"
GROUP BY "response", FLOOR(("SOURCE_SQL_STREAM_001".ROWTIME - TIMESTAMP '1970-01-01 00:00:00') minute / 1 TO MINUTE);
Exit (done editing)
Save and run SQL
Source data
Real-time analytics
Destination
Application status: RUNNING
bg-ecs-wp-log-ingestion-stream:
SOURCE_SQL_STREAM_001
Refresh stream sample
Download CSV
Edit schema
Filter by column name
ROWTIME TIMESTAMP
host VARCHAR(16)
datetime VARCHAR(32)
request VARCHAR(64)
2017-05-29 23:10:42.341 187.254.239.144 21/Jun/2017:00:32:47 +0000 POST /wp-admin HTTP/1.0
2017-05-29 23:10:42.341 175.177.137.104 21/Jun/2017:00:36:52 +0000 PUT /list HTTP/1.0
2017-05-29 23:10:42.341 57.172.15.75 21/Jun/2017:00:39:46 +0000 GET /wp-content HTTP/1.0

Real-time analytics

Add and run SQL queries to continuously analyze source data in real-time. Then, optionally, connect the in-application stream to a destination to deliver results.

Add SQL from templates | Download SQL

```
1  CREATE OR REPLACE STREAM "DESTINATION_SQL_STREAM" (datetime TIMESTAMP, status INTEGER, statusCount INTEGER);
2  CREATE OR REPLACE PUMP "STREAM_PUMP" AS INSERT INTO "DESTINATION_SQL_STREAM"
3  SELECT STREAM ROWTIME as datetime, "response" as status, COUNT(*) AS statusCount
4  FROM "SOURCE_SQL_STREAM_001"
5  GROUP BY "response", FLOOR(("SOURCE_SQL_STREAM_001".ROWTIME - TIMESTAMP '1970-01-01 00:00:00') minute / 1 TO MINUTE);
6
```

Exit (done editing) | Save and run SQL

Source data | Real-time analytics | Destination — Application status: RUNNING

In-application streams:

- DESTINATION_SQL_STREAM
- error_stream

Pause results — New results are added every 2-10 seconds. **The results below are sampled.**

Scroll to bottom when new results arrive.

Filter by column name

ROWTIME	DATETIME	STATUS	STATUSCOUNT
2017-05-29 23:11:00.0	2017-05-29 23:11:00.0	200	18831

Source data | Real-time analytics | Destination — Application status: RUNNING

In-application streams:

- DESTINATION_SQL_STREAM
- error_stream

Pause results — New results are added every 2-10 seconds. **The results below are sampled.**

Scroll to bottom when new results arrive.

Filter by column name

ROWTIME	DATETIME	STATUS	STATUSCOUNT
2017-05-29 23:11:00.0	2017-05-29 23:11:00.0	200	18831
2017-05-29 23:12:00.0	2017-05-29 23:12:00.0	200	27919
2017-05-29 23:12:00.0	2017-05-29 23:12:00.0	404	1256
2017-05-29 23:12:00.0	2017-05-29 23:12:00.0	301	1222
2017-05-29 23:12:00.0	2017-05-29 23:12:00.0	500	603
2017-05-29 23:13:00.0	2017-05-29 23:13:00.0	200	27925
2017-05-29 23:13:00.0	2017-05-29 23:13:00.0	301	1223
2017-05-29 23:13:00.0	2017-05-29 23:13:00.0	404	1228
2017-05-29 23:13:00.0	2017-05-29 23:13:00.0	500	624

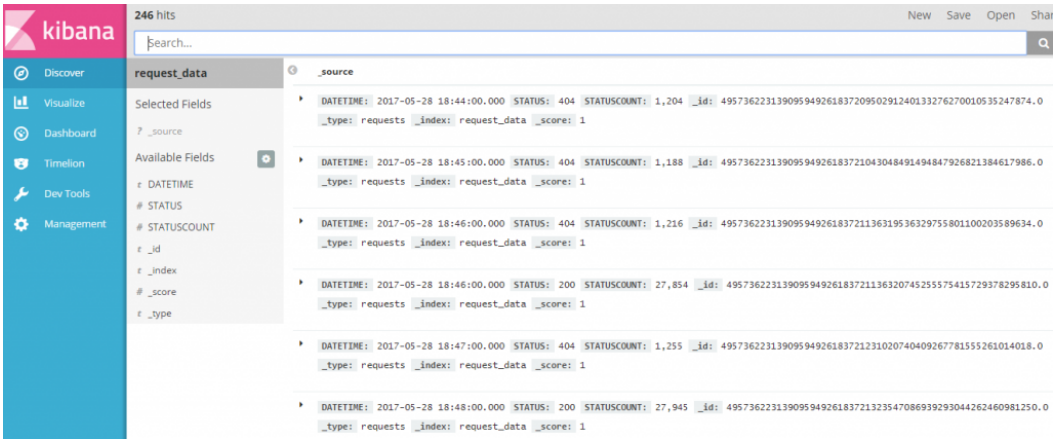
bg-ecs-wp-log-summary

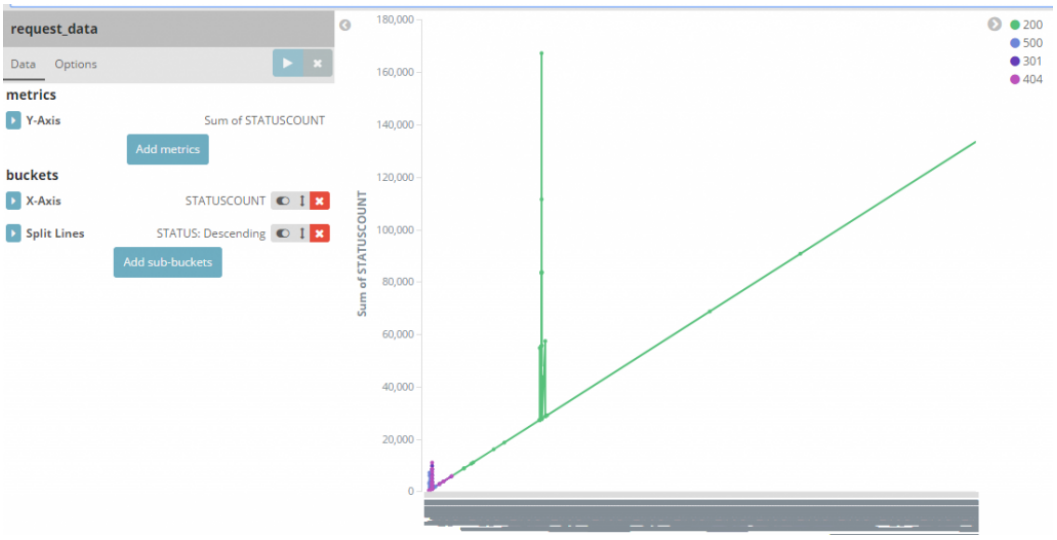
Configure cluster | Modify access policy | Manage tags

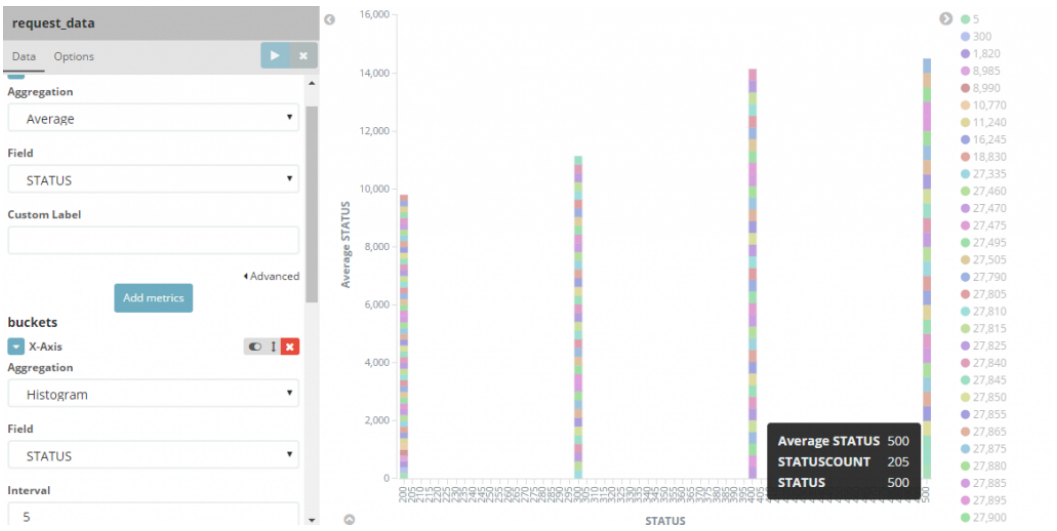
Domain status	Active
Elasticsearch version	5.1
Endpoint	search-bg-ecs-wp-log-summary-jrklgq2gkf3itt3qpc5atygsse.us-east-1.es.amazonaws.com
Domain ARN	arn:aws:es:us-east-1:015887481462:domain/bg-ecs-wp-log-summary
Kibana	search-bg-ecs-wp-log-summary-jrklgq2gkf3itt3qpc5atygsse.us-east-1.es.amazonaws.com/_plugin/kibana/

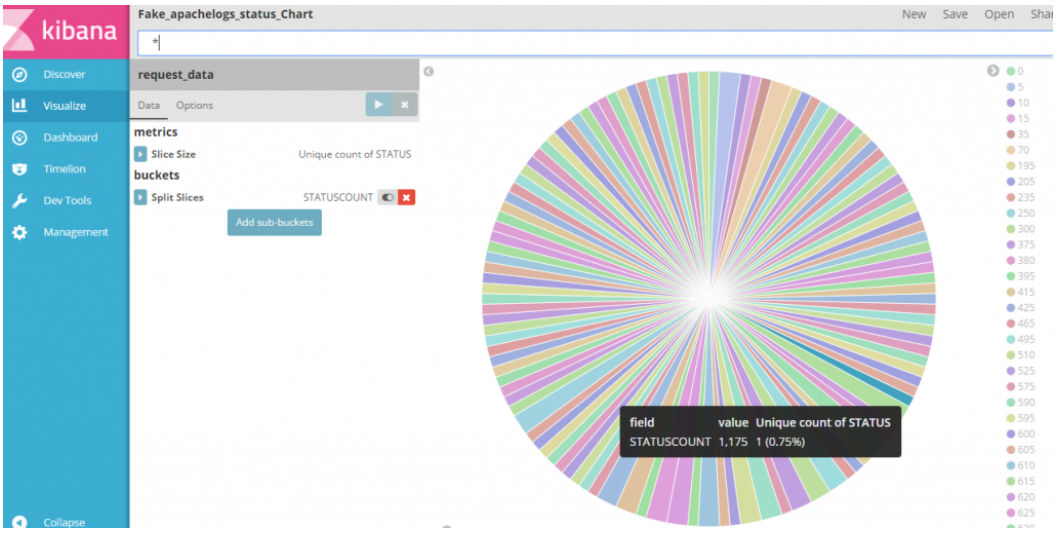
Cluster health | **Indices** | **Monitoring**

Status	Yellow
Number of nodes	1
Number of data nodes	1
Active primary shards	6
Active shards	6
Relocating shards	0




request_data
Data
Options
metrics
Y-Axis
Sum of STATUSCOUNT
Add metrics
buckets
X-Axis
STATUSCOUNT
Split Lines
STATUS: Descending
Add sub-buckets
Sum of STATUSCOUNT
180,000
160,000
140,000
120,000
100,000
80,000
60,000
40,000
20,000
0
200
500
301
404


request_data
Data
Options
Aggregation
Average
Field
STATUS
Custom Label
Advanced
Add metrics
buckets
X-Axis
Aggregation
Histogram
Field
STATUS
Interval
5
Average STATUS
16,000
14,000
12,000
10,000
8,000
6,000
4,000
2,000
0
STATUS
Average STATUS 500
STATUSCOUNT 205
STATUS 500


kibana
Fake_apachelogs_status_Chart
New
Save
Open
Discover
Visualize
Dashboard
Timelion
Dev Tools
Management
Collapse
request_data
Data
Options
metrics
Slice Size
Unique count of STATUS
buckets
Split Slices
STATUSCOUNT
Add sub-buckets
field value Unique count of STATUS
STATUSCOUNT 1,175 1 (0.75%)