Goodbye Splunk, Hello Amazon Elasticsearch Service

Abhishek Tiwari

Citation: *A. Tiwari*, ”Goodbye Splunk, Hello Amazon Elasticsearch Service”, Abhishek Tiwari, 2015. doi:10.59350/rn6np-tt078

Published on: October 01, 2015

Amazon today announced Amazon Elasticsearch Service (Amazon ES) - a fully managed Elasticsearch service which can support your real-time distributed search requirements. I am not sure about others, but for me a this is a big deal and can be a game changer. I have worked with both Splunk and Elasticsearch.

On the one hand, Splunk is definitely a superior product and well-packaged solution for operational intelligence. On the other hand, Elasticsearch is community driven, it has a strong open-source ecosystem. More than anything Elasticsearch has no licensing cost compared to Splunk. In recent years, Elasticsearch, Logstash, and Kibana (aka ELK) stack has emerged as a powerful and free alternative to Splunk. Just to be clear, Amazon ES is not offering just Elasticsearch but the full ELK stack using pre-installed plugins.

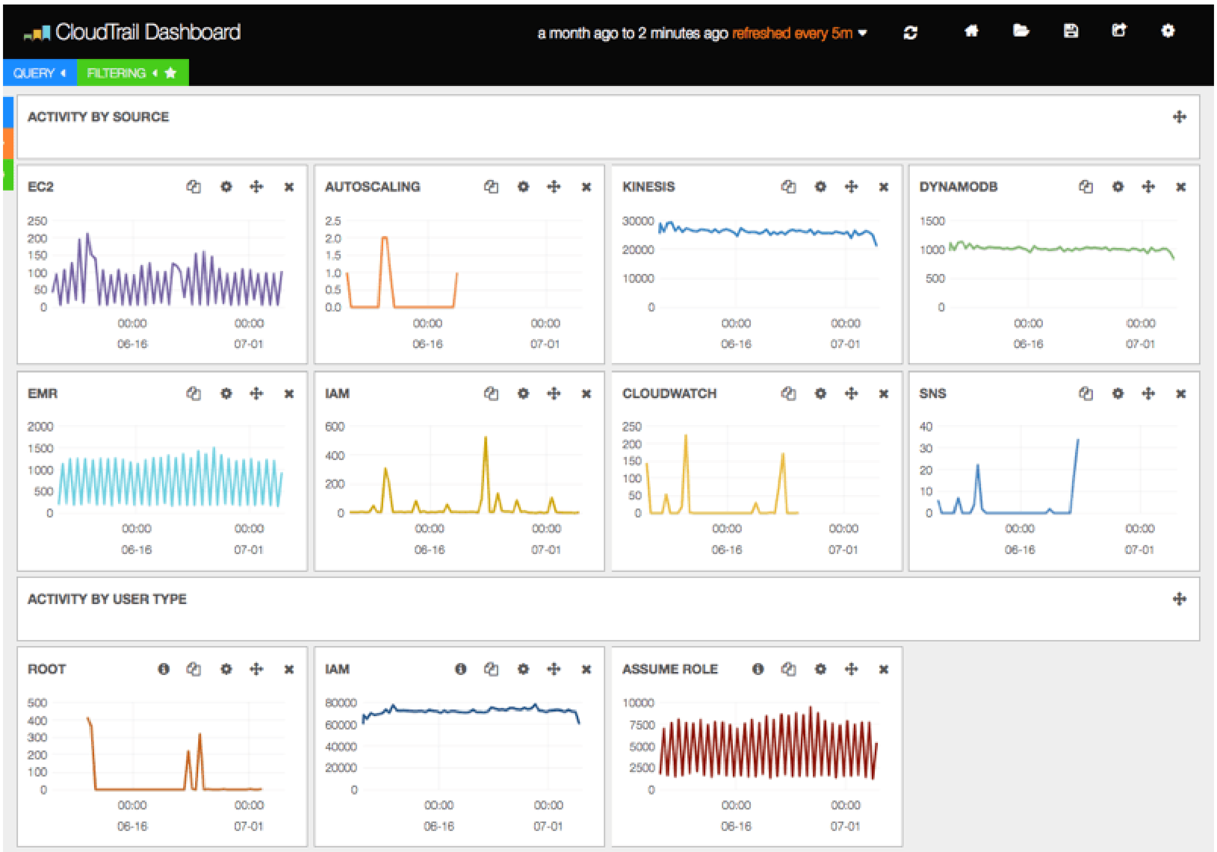


Figure 1: AWS CloudTrail integration with Elasticsearch and Kibana

Killer Features

Amazon ES is managed Elasticsearch with support for high availability, zone awareness, patch management, failure detection and node replacement, backups, and monitoring. Ability to scale up or down based on demand makes Amazon ES more cost effective. You just pay for the resources utilised by Amazon ES such as EC2, data transfer, EBS, S3 but nothing for Elasticsearch itself.

Lambda Powered

Basically, all data ingestion is driven by Amazon Lambda. This comes as no surprise to me, and I think eventually Amazon all Amazon services will be streaming data into Amazon ES via Lambda functions. Lambda functions can be used to pre-transform the data so Elasticsearch indexing is no brainer and less CPU intensive.

Managed

Like Amazon RDS, Amazon ES is simple to deploy and administer. Amazon ES basically takes away time-consuming sys-admin tasks. You can either launch a new cluster or resize your cluster up or down via within minutes using CLI, a single API call, or a few clicks on the AWS Management Console.

Default plugins

Amazon ES comes with the several pre-installed plugins which include Logstash, Kibana4, KIbana3, ICU Analysis and the Kuromoji. In addition, Amazon ES can utilise Logstash plugins such as s3 plugin, DynamoDB input plugin and DynamoDB output plugin. For instance, Logstash s3 plugin allows bulk data load capability from Amazon S3.

Security

Rather than commercial Elasticsearch plugins such as Marvel and Shield, Amazon ES relies on AWS Identity and Access Management (IAM) and CloudWatch to support the data security and monitoring. IAM role based access control is par with the Shield. CloudWatch can monitor the Elasticsearch deployment like Marvel by reporting metrics around cluster, master node and EBS volume. Moreover, configuration related activities can be audited using CloudTrail.

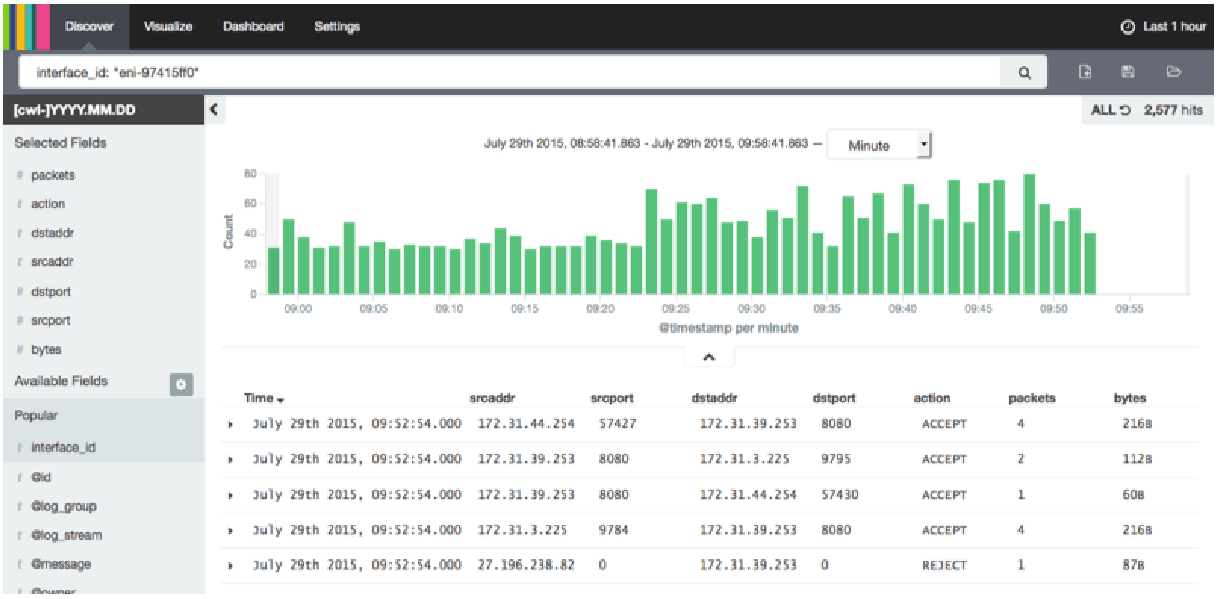
Amazon Integration

It seems like Amazon has been working on integrating various AWS services with Elasticsearch for quite some time now. In the core of these integrations, Amazon is using two of it's key services AWS

Lambda and Amazon Kinesis. Basically, Lambda automatically streams data to Elasticsearch whenever new data is added to AWS service - be it Amazon S3, Amazon Kinesis, Amazon DynamoDB or Amazon CloudWatch. Lambda detects change either via polling (Amazon Kinesis), trigger notification (for instance Amazon S3) or the stream functionality (DynamoDB).

In August this year, Amazon announced DynamoDB integration with Elasticsearch using DynamoDB Streams connector plugin. This integration enabled DynamoDB users to perform efficient real-time queries (structured, full-text, fuzzy and multifield) against DynamoDB data using Elasticsearch.

As part of today's launch of Amazon ES, Amazon also announced Amazon ES integration with CloudWatch Logs. Again this is something existed before via a Kinesis based consumer but now there seems to be a more seamless integration based on Lambda. You can specify subscription filter pattern to identify specific terms or pattern in your CloudWatch log events. There are some out-of-box Kibana dashboards available for VPC Flow, CloudTrail, and Lambda.



interface_id: "eni-97415ff0"

July 29th 2015, 08:58:41.863 - July 29th 2015, 09:58:41.863 —

Time	srcaddr	srcport	dstaddr	dstport	action	packets	bytes
July 29th 2015, 09:52:54.000	172.31.44.254	57427	172.31.39.253	8080	ACCEPT	4	216B
July 29th 2015, 09:52:54.000	172.31.39.253	8080	172.31.3.225	9795	ACCEPT	2	112B
July 29th 2015, 09:52:54.000	172.31.39.253	8080	172.31.44.254	57430	ACCEPT	1	60B
July 29th 2015, 09:52:54.000	172.31.3.225	9784	172.31.39.253	8080	ACCEPT	4	216B
July 29th 2015, 09:52:54.000	27.196.238.82	0	172.31.39.253	0	REJECT	1	87B

Figure 2: Amazon VPC Flow Logs using Elasticsearch and Kibana

Closing thoughts

Starting from today, Amazon ES is available to all customer in all AWS regions. Amazon ES is already integrated with key AWS services which makes it production ready from day one. Current use cases for Amazon ES includes but not limited to real-time application monitoring, streaming analytics, operational intelligence (monitoring and alert), and rich search and navigation experience.