# **INTRODUCTION TO ELASTIC CLOUD**



Opensource software by definition earns no money for its creators, since it's free. The primary sponsor then hopes larger companies will sign up for the enterprise product, for which they charge money. That way the user gets offer support, consulting, and it unlocks different features.

For example, ElasticSearch is free, but with the enterprise edition you get machine learning and other features, for a free. Here we show you the Elastic Cloud, hosted version of ElasticSearch.

Opensource companies recently have begun to host clusters themselves. This make this cheaper for its users that buying the product outright. It lets users pay monthly instead of up front.

The CFO loves that, as the company can write off operating costs immediately instead of having to amortize them over time. That's a fancy way of saying they can reduce their taxes sooner than later.

But for technical people like us, it makes the deployment and maintenance easier.

Here we discuss how ElasticSearch has done this with the Elastic Cloud. In short, the takeaway message is that you work with it the same as if you had installed it yourself. The only difference is you don't have ssh access to the instances. But you don't need that since you can use ssh from any other computer to get to those.

# **Getting Started**

You can get a free 14 day Elastic Cloud trial at <u>https://cloud.elastic.co/.</u> It's easy to get started, you just create a Deployment and ElasticSearch installs a cluster for you at either Amazon AWS or the Google Cloud as the screen below shows.

The screen look familiar to those of you who have used AWS. It asks you for what region you want

your servers and what size hardware, meaning EC2 template.

Name your deployment		
Give your deployment a name		
Salact a cloud platform		
Select a cloud platform		
Select a cloud platform Pick your cloud and let us handle th	ne rest. No additional accounts required.	
Select a cloud platform Pick your cloud and let us handle th aws	re rest. No additional accounts required.	
Select a cloud platform Pick your cloud and let us handle th	te rest. No additional accounts required.	

You can select

more memory, a necessity in any production use, storage and nodes on the customization screen. As you can see that is not possible with the trial version.

Trial user? The trial includes more than enough to get you started with the Electic Steck. Larger deployments	require a credit card to anlock and are not free.
E Data 1 configuration	Summary
awa.deta.highio.l3 (ves meet measur An IIO splimited Easticsearch Instance running on at AWS IS.	Version VIII ES detamemery 1:00
Redifictences	ES data storage 248 05 Total memory 8.8 08 Total storage 243 08
100 100 400 100 1500 1900 5000 × 1 + 408	Architecture
3 GB MM         \$30 GB strengt         > 1 mmle         > 3 mmms         > 8 GB MMH         246 GB strengt           > User setting overrides	
Machine Learning Tconfiguration	Zere 1
arvs.ml.m5 Insertine Learning An Electrosearch machine learning instance running on an AWS m5.	Ther

automatically rolls out a the ElasticSearch servers and the Kibana front end. It gives you an internet domain name so anyone in your organization can access it. It install the SSL certificates and adds a password.

Save your password:

#### Generated user

You can use the credentials below to login to Elasticsearch or Kibana. Make sure to save the password somewhere as this is the only time we can show it to you.

APM Server secret token
cluster on Elastic Cloud. Learn more
The Cloud ID simplifies sending data to your
Get started with Beats and Logstash quickly.
XXXXXXXXXXX
Cloud ID
X0000000000
Password
elastic
Username

XXXXXXXXXXXXX

From the main screen, click **Copy Endpoint** it will give you

the URL for your instance. For example, my URL is https://58571402f5464923883e7be42a037917.eu-central-1.aws.cloud.es.io:9243

Deployments	180714	
Analytics	Analytics	
Easticsearch	Digitigenet many	Supervised car
Enapolismi BPI Consets	Analytics Ball	· riestfu
Kitala Albai	Digitization residen	
Although	475.1	
Derformente		
Custam plugina	Lawet   B Copy Endpoint URL	1410/4/101100/101100/001100
Account	🖉 Kibana 🗄	interest for the former
Help	Launch 2 Copy Bridgoon URL	

Once you get a

deployment (aka cluster) you can check its status with curl, if you are used to using the command line, as shown below.

First save pwd=**"userid:password"**, used for basic authentication, in an environment variable. Then run the curl:

```
export pwd="elastic:"
```

```
curl --user $pwd -H 'Content-Type: application/json' -XGET
https://58571402f5464923883e7be42a037917.eu-central-1.aws.cloud.es.io:9243/_c
luster/health?pretty
```

Here it shows all is good (green) and I that have 2 data nodes.

```
{
  "cluster_name" : "58571402f5464923883e7be42a037917",
  "status" : "green",
  "timed out" : false,
  "number of nodes" : 3,
  "number_of_data_nodes" : 2,
  "active_primary_shards" : 6,
  "active_shards" : 12,
  "relocating_shards" : 0,
  "initializing shards" : 0,
  "unassigned shards" : 0,
  "delayed unassigned shards" : 0,
  "number of pending tasks" : 0,
  "number of in flight fetch" : 0,
  "task max waiting in queue millis" : 0,
  "active_shards_percent_as_number" : 100.0
}
```

And you can see the cluster status from the main screen <a href="https://cloud.elastic.co/deployments">https://cloud.elastic.co/deployments</a>. Click on the name you gave the deployment. Here I called mine **Analytics**.

Nginamata Lariar pingin Mat	Welcome to your 14-day triat help your her also prove to the start fractioners and there are serving, and much next, if call along if you way that to go the proving.	1			
	Deployments				
	Analytics				
	Ball Provident      management      manag				

You can see that it

has created ElasticSearch, Kibana, and APM, which is their monitoring tool.

c		weroweigh	ealions
Depleyments Analytica	Analytics		
Electronershi Long	Beptoyment name	Dephysionett of allow	
Snapshots	Analytics Tensme	e Healthy	
Kibana	Bapicymant version		
APM Activity	1211		
Security Performance	Applications	Omel ID	
Custom plugins	Elasticearch	Ansily156512201727vdHUb6ChiL#235y5j62732555cg5pby01000 3970g49416078072200g42141270y7747855c8y53007387718x5y	
Account	Kitest	2009%245Fjh1F2%32j%b0D811Dg1%4==	û.
Help	Laumon   Copy Endpoint URL		
	APM		
	Launch   Capy APAr Server URL 🗎		

Here you can see

that I have 5 VMs. Or they could be running in containers. I am not sure.

There's not a lot of memory on these nodes. Which, if you have worked with AWS before, makes sense as the larger the machine the higher the subscription fees.

central-ta	eu-certital-15	eu-pentral-to
AVELAPIN.04 Indexce FE VELS	ANVE MAGTER.RM	AVELONIA.MONICIO Indena FO VIII A OR BASA Asta Realer algère Ingeni
AVIS.DATA.HODHO.D Interce II 1222 A SEBAM Ora Report Report	JVM memory pressure 10% Disk weege 0% of 2-08	John memory pressure 75 Drak usager 0% of 130 08
JAM memory pressure 7%	Stop routing	Stop routing

If you click Launch

ElasticSearch, it simply runs the curl command that gives you information about the instance. But basic authentication is turned on. So the nginx web server that serves as the front end will ask you to login:

#### Sign in

https://58571402f5464923883e7be42a037917.eucentral-1.aws.cloud.es.io:9243

Username	elastic		
Password	•••••		
		Cancel	Sign In

Then in the browser it gives you the

```
same data we extracted using curl above.
```

#### {

```
"name" : "instance-000000001",
  "cluster name" : "58571402f5464923883e7be42a037917",
 "cluster uuid" : "x4gHtdPbTiiFlcxWkjDJsQ",
  "version" : {
    "number" : "7.1.1",
   "build flavor" : "default",
   "build type" : "tar",
    "build hash" : "7a013de",
   "build date" : "2019-05-23T14:04:00.380842Z",
   "build_snapshot" : false,
   "lucene version" : "8.0.0",
   "minimum wire compatibility version" : "6.8.0",
   "minimum_index_compatibility_version" : "6.0.0-beta1"
 },
 "tagline" : "You Know, for Search"
}
```

Here is that curl again Notice what we have put the **userid:password** into the environment variable **pwd** to avoid having to type that each time.

export pwd="elastic:"

```
curl --user $pwd
https://58571402f5464923883e7be42a037917.eu-central-1.aws.cloud.es.io:9243/
```

# Look Some (a lot of Data) Into the Environment

Now, let's stress this environment and show you how to use it by loading it with the extremely large FDA (American Food and Drug Administration) data on drug reactions. Will we use this later to build a machine learning model to try to prective the side effects of a particular drug.

### wget

https://download.open.fda.gov/drug/event/all\_other/drug-event-0004.of-0004.js

# unzip drug-event-0004-of-0004.json.zip

There are 4 datasets you can download. As you can see it is quite large > 1 GB when unzipped. And this is just JSON text.

- 1. -rwxrwxrwx 1 ubuntu ubuntu 33M Jun 4 10:51 drug-event-0001-of-0004.json
- 2. -rwxrwxrwx 1 ubuntu ubuntu 79M Jun 5 10:51 drug-event-0002-of-0004.json
- 3. -rwxrwxrwx 1 ubuntu ubuntu 211M Jun 5 10:52 drug-event-0003-of-0004.json
- 4. -rwxrwxrwx 1 ubuntu ubuntu 637M Jun 4 10:52 drug-event-0004-of-0004.json

We can try to use the bulk loader against this data. But ES complains, saying the JSON file is malformed. Plus files 3 and 4 are larger than 100 MB, which is the maximum size the ES bulk loader will handle:

```
curl --user $pwd -H 'Content-Type: application/x-ndjson' -XPOST
'https://58571402f5464923883e7be42a037917.eu-central-1.aws.cloud.es.io:9243/0
/_bulk?pretty' --data-binary @drug-event-0004-of-0004.json
```

It will give this error:

```
{
    "error" : {
        "root_cause" : )\n at "
        }
    ],
```

# Load the Data Using Python

So we can load it the slow way by running this python code below. This program will run slow because it connects to the API one time for each JSON record. It will take days to load it into this instance. But we can load enough of it in a few hours to run some analytics against it, which we will illustrate in subsequent posts.

Change the open statement to pick up the name of the file you downloaded:

```
open('drug-event-0004-of-0004.json')
```

And the URL of your instance. Keep **/fda/\_doc** on the end, as **fda** is the instance we will use and **/fda/\_doc** is the index type.

```
url =
"https://58571402f5464923883e7be42a037917.eu-central-1.aws.cloud.es.io:9243/f
da/_doc/"
```

Here is the code:

import json

```
import requests
import uuid
def fdict(d):
    ky =
    headers = { 'content-type': 'application/json' }
    url =
"https://58571402f5464923883e7be42a037917.eu-central-1.aws.cloud.es.io:9243/f
da/ doc/"
    data={}
    for i in ky:
        data = d
    response = requests.put(url + str(uuid.uuid4()) , headers=headers,
json=data, auth=('elastic', '
You can list some of the data like this:
curl --user $pwd -H 'Content-Type: application/json' -XGET
https://58571402f5464923883e7be42a037917.eu-central-1.aws.cloud.es.io:9243/fd
a/?pretty
And list all indices in this cluster to get the document count:
curl --user $pwd
                  - XGET
https://58571402f5464923883e7be42a037917.eu-central-1.aws.cloud.es.io:9243/ c
at/indices?v
As you can see, the data is large. 47,939 documents is 1.9 GB of storage.
health status index
                            uuid
                                                     pri rep docs.count
docs.deleted store.size pri.store.size
                       DRQnRqnmQsaRcsxHVIT1pg 1 1
green open fda
                                                             47939
                                                                               0
            1003.6mb
1.9gb
```

#### Kibana

Now let's look at the same data with Kibana. Click on the Kibana link from the main screen. In our case it is

https://78f7165792d54709b8ec37f3d80ed854.eu-central-1.aws.cloud.es.io:9243/app/kibana

If you're new to Kibana you have to create an **Index Pattern** before you look at any of the data. which you will do using the Discover screen.

The Index Pattern is the name of the index, in this case **fda**, or a wildcard that matches the index name, like **fda**\*.

You click the **discover** button in Kibana to see this.

Ζ 😐	Home		
8			
Ø Elises	ver Use these solutions to quickly turn your da	ita into pre-built dashboards and monitorin	g systems.
2	_		
a	<u> </u>	î	
8	ADM	Looping	
8	APM automatically collects in-	ingest logs from popular data	Col
	depth performance metrics and	sources and easily visualize in	operat
à	errors from inside your anninations	preconfigured dashboards.	nun
ar i	approximeter.		_
8	Add APM	Add log data	Ŀ
J			
÷	Add sample data	Upload data	tran log file Wher

open Kibana you create the first Index Pattern. ElasticSearch will parse the JSON documents it finds there and show you the schema it found. The schema is the Index Pattern.



Kibana will churn

for a few seconds and then show you the fields it has auto discovered.

transportant the						
Overlagement     Overlagement     Index Uncycle Polices	nter part				* 0 .	•
Rohar, Joba Ratishar BO Upphale Asabijari	Tris page lists every Districterents. To che	Teld in the Marindes and I inge a field type, use the G	he field's associated con lectics earch Mapping AP	n lygen as read På	raked by	
	Press (103)	Ecologiesi Pietalis (C)	Bounce (Warrs (0)			
E Kibana Index Palacite	Q; Filter				All field types	
Saved (Object)) Several	tere .	lan.	format described	in Assessed	. Contraction	
Reporting Advanced Delivery	м	string	•	•		ē
	John	string		•		ē
Legriash .	.007	1.000	r		1	ē.
Pipelino	10.000					i.
- Bosta	2,94	noord	•	•		٢.
Genra Management	sompanynumis	string	•		1	e.
© Security	companyments.tarywork	4				e
L'anno	fulfillespecifies therein	aning.				į.

Below it shows the

first few records. Below we click on one record to expand it.

<b>K</b>	E Discover	0
0	5,858 115	
	New New Open Diare b	ngasi
10.	riters Brann	54. O Per
10	© - + Aat Bar	
÷.	***	0
4	Galaxiani Kalala	<sup>3</sup> suppopulati of effort to boost of consignmentations; of constant: - Safetting effort or the first of the second se
0	Australia factor P	Profile Your Constrainty ( Profile State Strainty ( ) West State ( ) State Stat State State S
	1.8	"METER" L. "MANAGALAN": 1 "METERSTOCK", "METER
8	1 _mas	2 magageate: 2010205 reception/cont; HJ marine; - arrival / 50252age8.tectorial / princpares.pail/Euclid. (
ъ.	1.50	Enterpretation and a second s second second sec
0	1.00949508	BELLERICH, Appellal's ("Advances-Appelle") ("MARGES"), Application-Appelle ("MARGES", "MARGES", "MARGES, "MARGES", "MARGES, "MARGES", "MARGES, "MARGES", "MARGES, "MARGES, "MARGES", "MARGES, "MARGES", "MARGES, "MARGES, "MARGES", "MARGES, "MARGES, "MARGES, "MARGES, "MARGES, "MARGES, "MARGES, "MARGES, "MARGES", "MARGES, " MARGES, "MARGES, "M
. 0	1 Millegetheritete	2 merende interfetet versietendenen af merenen i beiter i beiter enter i enterenen er blinning i
2	4 patienting	PCHEVENING PERFERONMENT ACTO FAIL Development of senter. enderstantiation: FR-Fails the receivers: 201
φ.	<ul> <li>participation missimple.</li> </ul>	print day ( Mapping Street, 1997, Mapping Street, 197, Mapping Street, 1988). Solution, Statement, Mapping Street, 1988
۰	1 patientpatientischipa-	"Angele Solar 1 (1998) 187, "Angel verse store and an angele solar 1 (1998) 197, "Angele Solar Angele Solar and Sol
+	1 judient, judientersatiope	2 manyodi di 100 kato manihistori ili materi - inter i fablimetterinini i riserenti additatio Each JSON record

from the FDA is enormous. A small part looks like this:



From here you could connect logstash to it, to ingest some application or hardware logs. We will explain that and how to use analytics in subsequent posts.