Administracja

Agenda

1. Topologia klastrów obliczeniowych w Big Data
2. Metody instalacji narzędzi
3. Zarządzanie i Monitoring

Topologia klastrów

- “Active-Standby”
- “Gold-silver”
- Active - Active ?

Gdzie można uruchamiać?

- Własna serwerownia
- Google Cloud
- Amazon AWS
- Microsoft Azure

Zarządzanie infrastrukturą

- Terraform
- AWS CloudFormation

Terraform

```
provider "aws" {
        region = "eu-west-1"
}

module "ec2" {
        instance_count = 2
        name                        = "example-normal"
        ami                         = "<id_obrazu_do_uruchomienia>"
        instance_type               = "m4.large"
        subnet_id                   = "<subnet_id>"
        vpc_security_group_ids      = ["<security_group_id>"]
        associate_public_ip_address = true
}
```

Zarządzanie konfiguracją

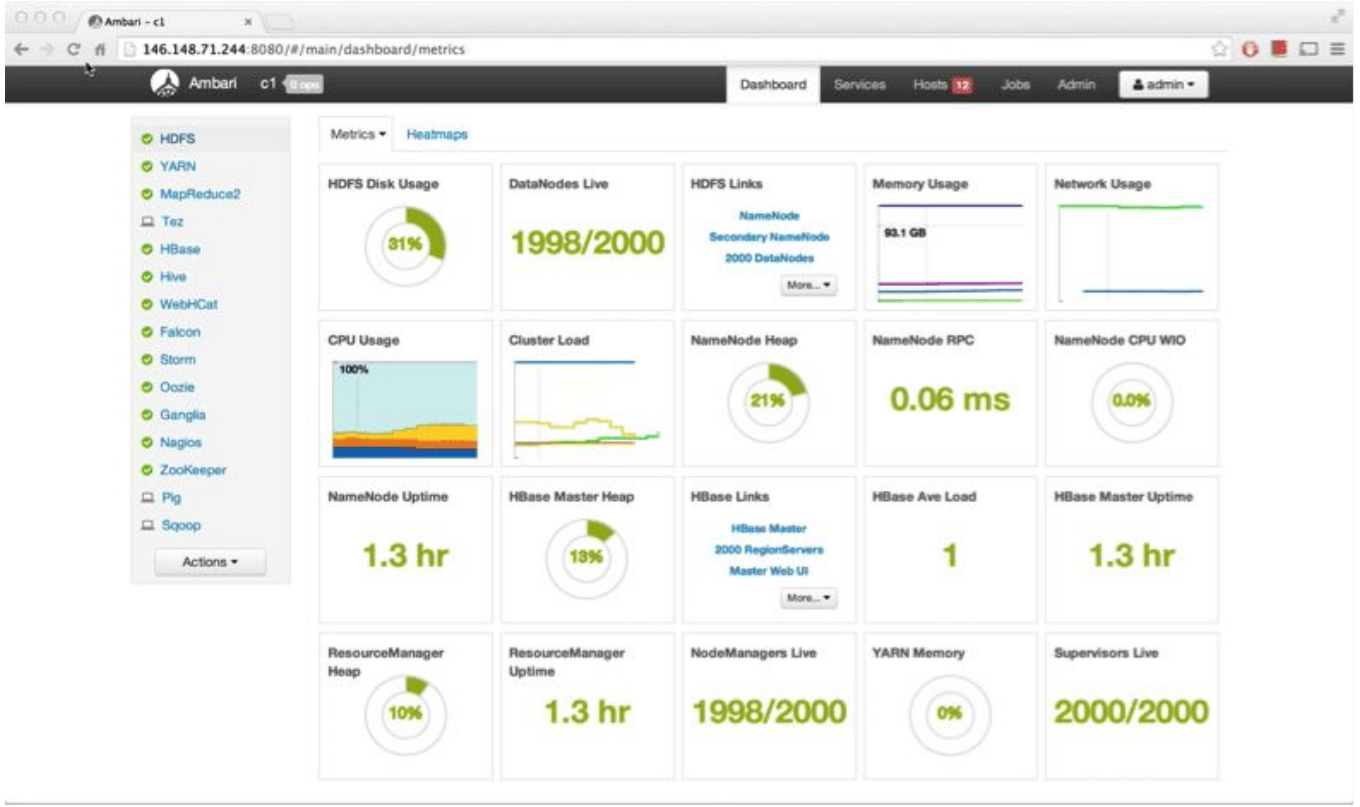
- Puppet
- Chef
- Ansible

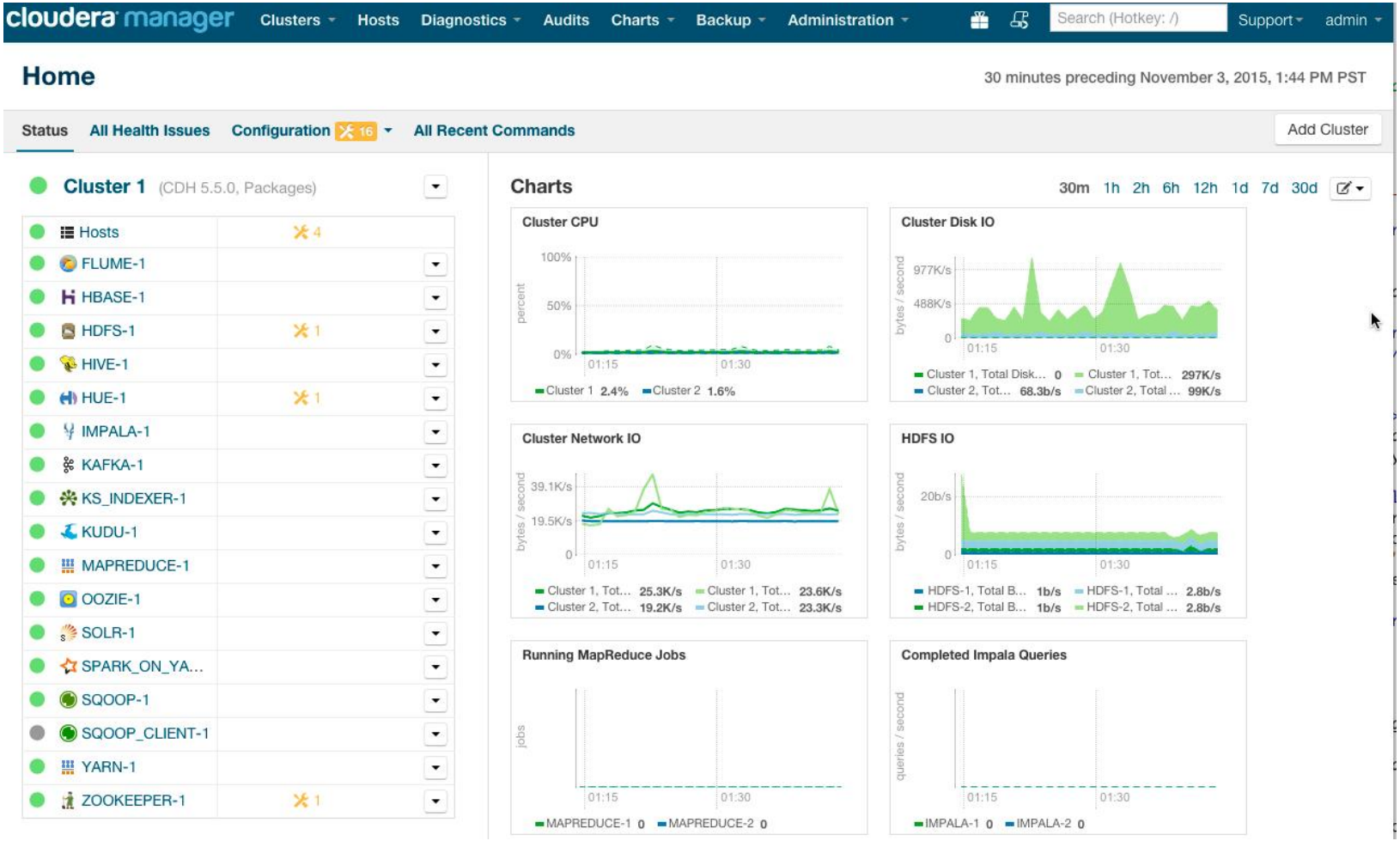
Puppet

```
package { 'mysql-server':
  ensure => installed,
}

service { ‘hadoop-yarn-resourcemanager’:
  ensure => running,
}

user { 'janusz':
  ensure     => present,
  uid        => '1521',
  gid        => '1345',
  shell      => '/bin/bash',
  home       => '/home/janusz'
}
```

Zarządzanie klastrami Hadoop

- Apache Ambari
- Cloudera Manager


Ambari - c1
146.148.71.244:8080/#/main/dashboard/metrics
Ambari
c1
Dashboard
Services
Hosts 12
Jobs
Admin
admin
HDFS
YARN
MapReduce2
Tez
HBase
Hive
WebHCat
Falcon
Storm
Oozie
Ganglia
Nagios
ZooKeeper
Pig
Sqoop
Actions
Metrics
Heatmaps
HDFS Disk Usage
31%
DataNodes Live
1998/2000
HDFS Links
NameNode
Secondary NameNode
2000 DataNodes
More...
Memory Usage
93.1 GB
Network Usage
CPU Usage
100%
Cluster Load
NameNode Heap
21%
NameNode RPC
0.06 ms
NameNode CPU WIO
0.0%
NameNode Uptime
1.3 hr
HBase Master Heap
13%
HBase Links
HBase Master
2000 RegionServers
Master Web UI
More...
HBase Ave Load
1
HBase Master Uptime
1.3 hr
ResourceManager Heap
10%
ResourceManager Uptime
1.3 hr
NodeManagers Live
1998/2000
YARN Memory
0%
Supervisors Live
2000/2000


cloudera manager
Clusters
Hosts
Diagnostics
Audits
Charts
Backup
Administration
Search (Hotkey: /)
Support
admin
Home
30 minutes preceding November 3, 2015, 1:44 PM PST
Status
All Health Issues
Configuration 16
All Recent Commands
Add Cluster
Cluster 1 (CDH 5.5.0, Packages)
Hosts 4
FLUME-1
HBASE-1
HDFS-1 1
HIVE-1
HUE-1 1
IMPALA-1
KAFKA-1
KS_INDEXER-1
KUDU-1
MAPREDUCE-1
OOZIE-1
SOLR-1
SPARK_ON_YA...
SQOOP-1
SQOOP_CLIENT-1
YARN-1
ZOOKEEPER-1 1
Charts
30m 1h 2h 6h 12h 1d 7d 30d
Cluster CPU
100%
50%
0%
percent
01:15
01:30
Cluster 1 2.4% Cluster 2 1.6%
Cluster Disk IO
977K/s
488K/s
0
bytes / second
01:15
01:30
Cluster 1, Total Disk... 0
Cluster 1, Tot... 297K/s
Cluster 2, Tot... 68.3b/s
Cluster 2, Total ... 99K/s
Cluster Network IO
39.1K/s
19.5K/s
0
bytes / second
01:15
01:30
Cluster 1, Tot... 25.3K/s
Cluster 1, Tot... 23.6K/s
Cluster 2, Tot... 19.2K/s
Cluster 2, Tot... 23.3K/s
HDFS IO
20b/s
0
bytes / second
01:15
01:30
HDFS-1, Total B... 1b/s
HDFS-1, Total ... 2.8b/s
HDFS-2, Total B... 1b/s
HDFS-2, Total ... 2.8b/s
Running MapReduce Jobs
jobs
01:15
01:30
MAPREDUCE-1 0
MAPREDUCE-2 0
Completed Impala Queries
queries / second
01:15
01:30
IMPALA-1 0
IMPALA-2 0

Zabezpieczenie danych

- Non secure mode
- Kerberos
- HiveServer

Monitorowanie

- Działanie procesów (czy np. odpowiada na porcie?)
- Statystyki systemu operacyjnego
- Utylizacja zasobów OS
- Monitoring JMX (procesy Javy)
- Testy funkcjonalne

Jakie parametry monitorować?

- Ilość bloków
- Ilość plików
- Utylizacja pamięci / YARN
- Ilość aplikacji w różnych stanach
- Ilość błędnych aplikacji oraz kontenerów
- Użycie pamięci + GC dla procesów HDFS+YARN

Narzędzia do metryk

- “Time Series DB”
- Graphite
- InfluxDB
- OpenTSDB
- Grafana


Big Dashboard
Zoom Out
Oct 4, 2015 11:29:09 to Oct 4, 2015 14:13:23 UTC
Logins
172
Sign ups
263
Sign outs
268
Support calls
80
Memory / CPU
memory
cpu
logins
logins
logins (-1 hour)
Memory / CPU
memory
cpu
logins
logins
logins (-1 hour)
server requests
web_server_01
web_server_02
web_server_03
web_server_04
client side full page load
avg
upper_25 6 ms
upper_50 75 ms
upper_75 411 ms
upper_90 974 ms
upper_95 1.288 s
server requests
web_server_01
web_server_02
web_server_03
web_server_04