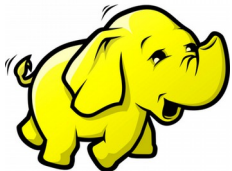


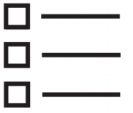
Cloudera Administrator Apache Hadoop Parte 03-1

Cloudera Manager



Marco Reis
<http://marcoreis.net>

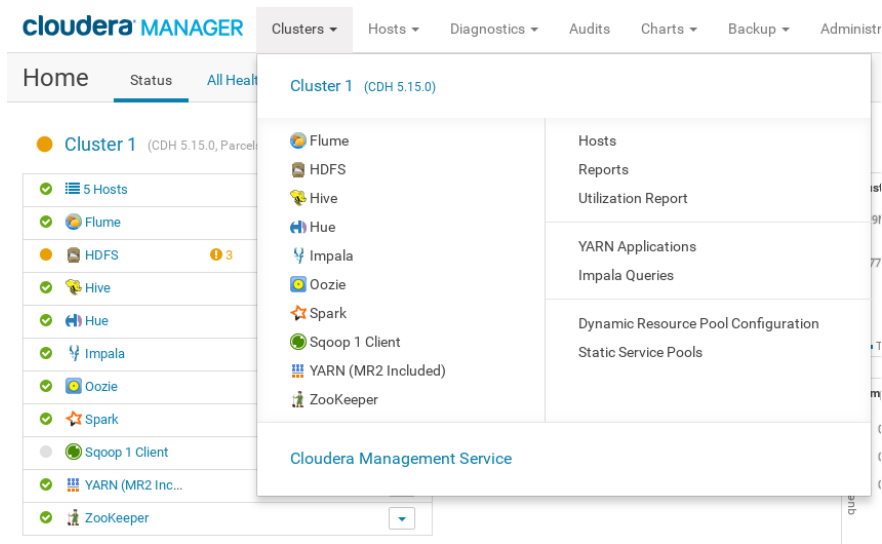
Agenda



- Cloudera Manager
 - Motivação
 - Arquitetura
 - Recursos
 - Navegação
 - Serviços
 - Configuração inicial

Cloudera Manager

- Interface web para administração centralizada do CDH
 - Métricas, alertas e configurações
- Um cluster Hadoop é formado por centenas de processos interligados que precisam de coordenação
 - A administração de um cluster Hadoop é complexa
- Permite verificar e configurar cada componente do cluster, seja ele um serviço ou um host
- Vamos começar com a configuração dos principais serviços



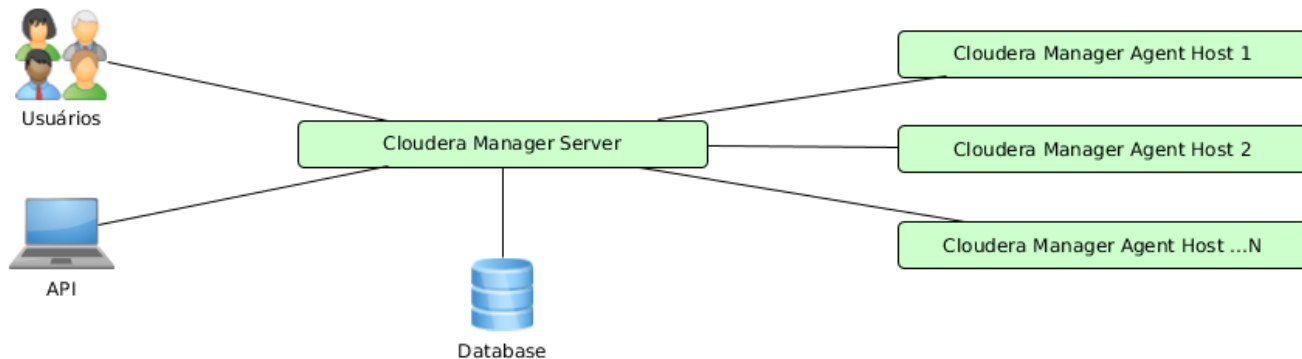


Arquivos de configuração

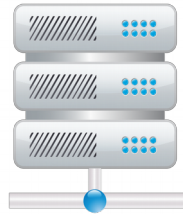
- A configuração do ecossistema Hadoop é feita a partir de arquivos específicos para cada serviço
- O CM facilita o gerenciamento dos arquivos, mas é bom saber o que acontece na realidade
- No Hadoop, os arquivos estão no diretório `/etc/hadoop/conf`, no Hive estão em `/etc/hive/hadoop/conf` e no Impala em `/etc/impala/conf`
- Os arquivos mais importantes são:
 - `core-site.xml`
 - `hdfs-site.xml`
 - `yarn-site.xml`
 - `mapred-site.xml`
 - `hive-site.xml`

Arquitetura do CM

- SCM Server
 - Serviço master do Cloudera Manager
- SCM Agent
 - Roda em cada um dos nós do cluster

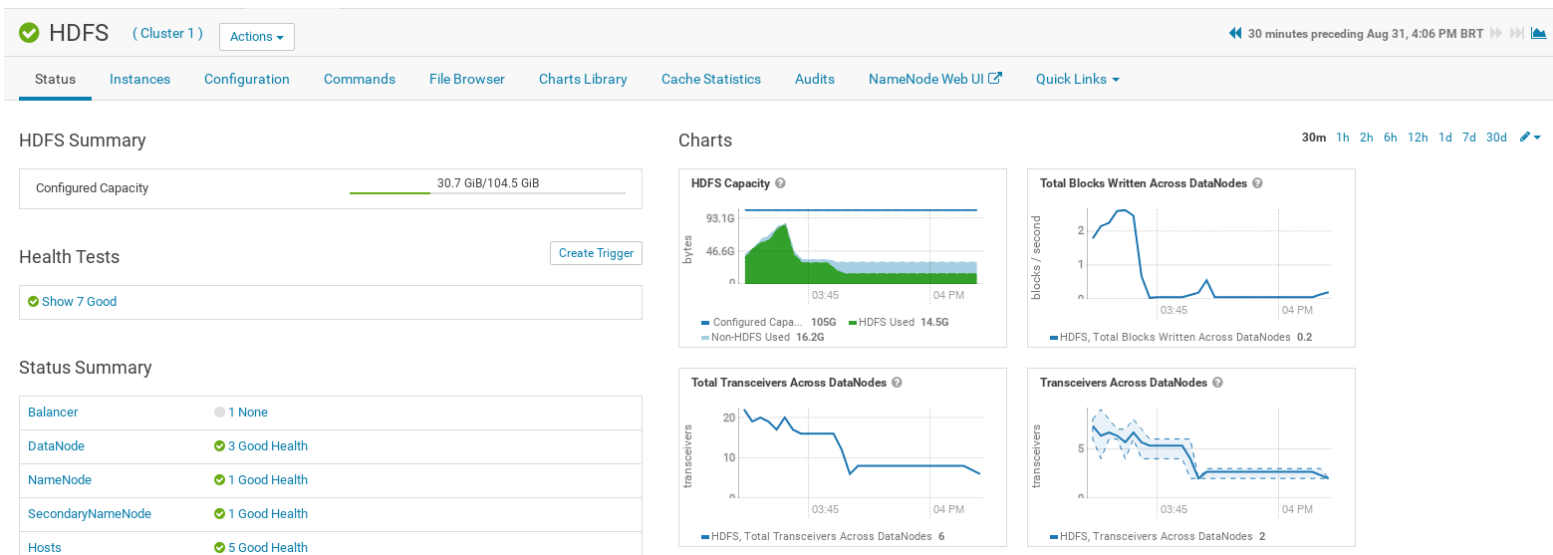


HDFS



Status

- Monitoramento do sistema de arquivos, como a capacidade, a performance, a saúde e alertas



NameNode Web UI

- Clique na aba NameNode Web UI
- A interface do NameNode está disponível em <http://headnode.lab:50070>
- O overview mostra o status e onde está rodando o NameNode

Overview 'headnode.lab:8020' (active)

Started:	Tue Sep 04 18:39:08 -0300 2018
Version:	2.6.0-cdh5.15.1, r2d822203265a2827554b84cbb46c69b86ccca149
Compiled:	Thu Aug 09 13:22:00 -0300 2018 by jenkins from Unknown
Cluster ID:	cluster14
Block Pool ID:	BP-1747176098-192.168.122.10-1536092355685

Summary

- O Summary mostra o status da segurança, safemode, quantidade de arquivos, memória etc
- O safemode (read-only) é ativado enquanto o NN carrega o fsimage na memória e pode demorar em clusters maiores

Summary

Security is off.

Safemode is off.

876 files and directories, 745 blocks = 1,621 total filesystem object(s).

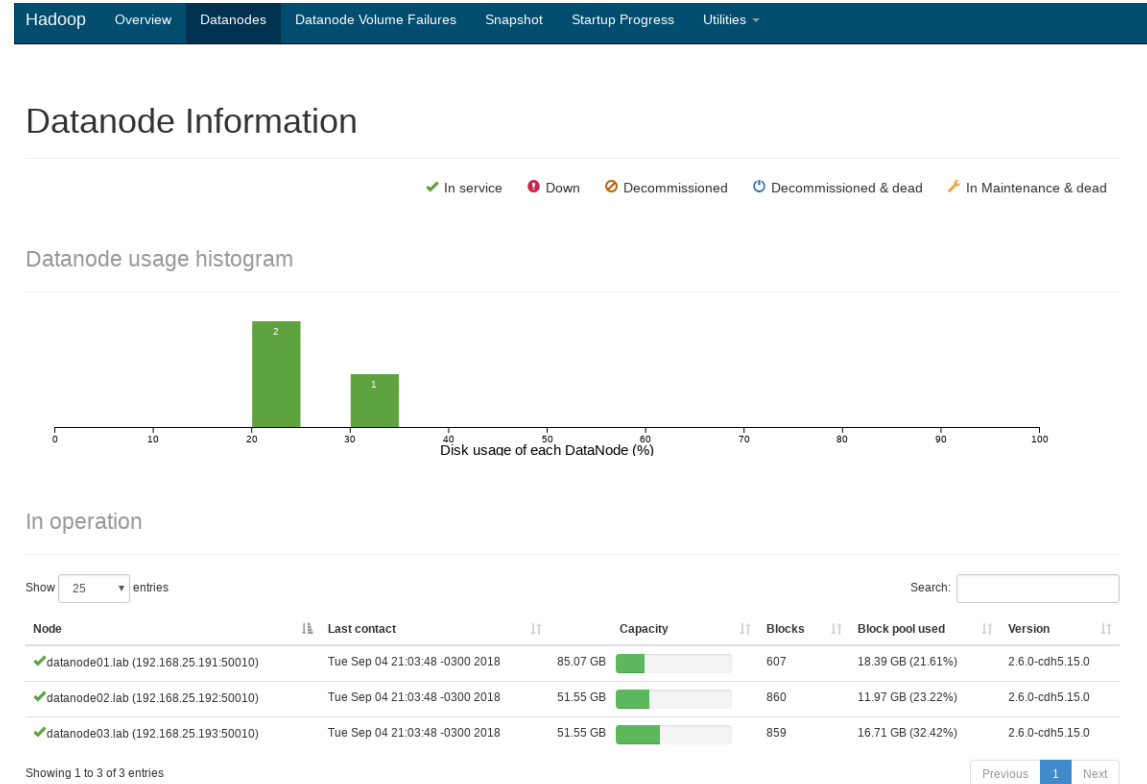
Heap Memory used 37.54 MB of 245.56 MB Heap Memory. Max Heap Memory is 245.56 MB.

Non Heap Memory used 44.19 MB of 45.13 MB Committed Non Heap Memory. Max Non Heap Memory is 130 MB.

Configured Capacity:	101.51 GB
DFS Used:	1.73 GB (1.71%)
Non DFS Used:	14.36 GB
DFS Remaining:	79.63 GB (78.44%)
Block Pool Used:	1.73 GB (1.71%)
DataNodes usages% (Min/Median/Max/stdDev):	1.71% / 1.71% / 1.71% / 0.00%
Live Nodes	3 (Decommissioned: 0, In Maintenance: 0)
Dead Nodes	0 (Decommissioned: 0, In Maintenance: 0)
Decommissioning Nodes	0
Entering Maintenance Nodes	0
Total Datanode Volume Failures	0 (0 B)
Number of Under-Replicated Blocks	0
Number of Blocks Pending Deletion	0
Block Deletion Start Time	Tue Sep 04 18:39:08 -0300 2018
Last Checkpoint Time	Tue Sep 04 18:08:14 -0300 2018

Datanodes

- A aba Datanodes mostra os datanodes, o último contato com o host, a capacidade e a utilização dos discos do cluster



HDFS Instances

- Gerencia os hosts e as funções associadas
 - Permite adicionar novas funções para os hosts (Add Role Instances)
- Selecione um host e verifique as ações disponíveis
 - Na figura podemos ver que o serviço está parado, sendo que uma das opções é iniciá-lo

Actions for Selected (1) ▾		Migrate Roles	Add Role Instances	Role Groups	
<input type="checkbox"/>	Role Type	State	Host	Commission State	Role Group
<input type="checkbox"/>	Balancer	N/A	edgenode.lab	Commissioned	Balancer Default Group
<input type="checkbox"/>	 DataNode	Started	datanode01.lab	Commissioned	DataNode Group 1
<input type="checkbox"/>	 DataNode	Started	datanode03.lab	Commissioned	DataNode Default Group
<input type="checkbox"/>	 DataNode	Started	datanode02.lab	Commissioned	DataNode Default Group
<input type="checkbox"/>	 HttpFS	Started	edgenode.lab	Commissioned	HttpFS Default Group
<input type="checkbox"/>	 NameNode (Active)	Started	headnode.lab	Commissioned	NameNode Default Group
<input checked="" type="checkbox"/>	 SecondaryNameNode	Stopped	edgenode.lab	Commissioned	SecondaryNameNode Default Group

Outros menus

- Commands
- File Browser
- Charts
- Cache Statistics
- Audits

Running Commands

No Results Found.

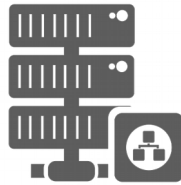
Recent Commands

Command
Deploy Client Configuration
Deploy Client Configuration
Start
Stop
Deploy Client Configuration
Deploy Client Configuration



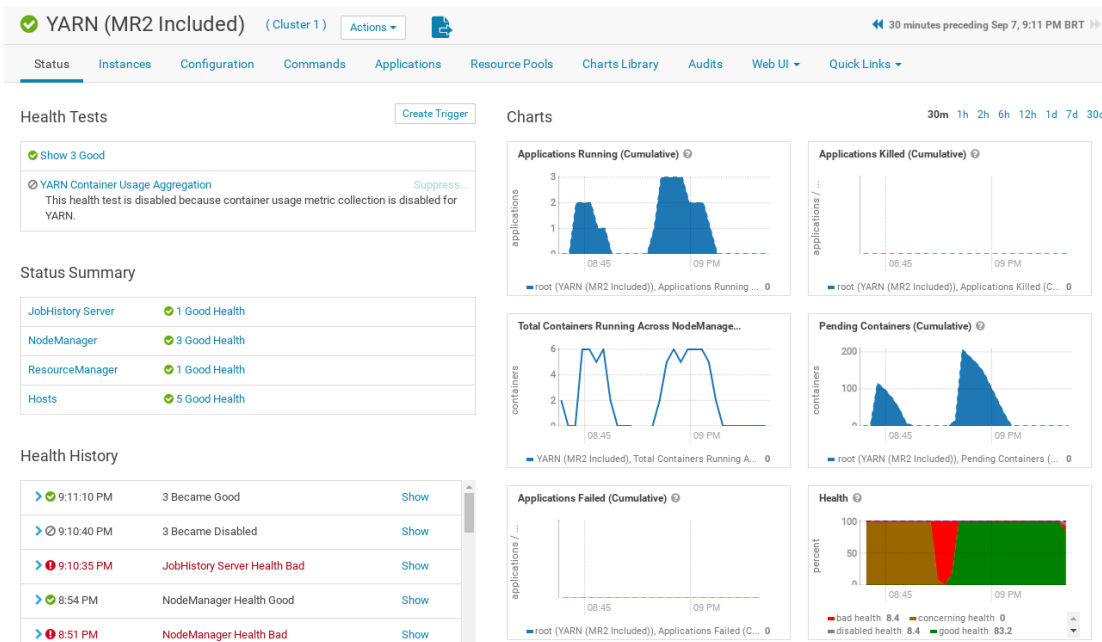
Status	Instances	Configuration	Commands	File Browser	Charts Library	Cache Statistics	Audits	NameNode Web UI	Quick Links
/ Edit									
Showing 1 to 2									
Name	Owner	Group	Last Modified	Size	Mode				
tmp	hdfs	supergroup	08/27/2018 5:15 PM	—	drwxrwxrwx				
user	hdfs	supergroup	08/26/2018 8:38 AM	—	drwxr-xr-x				

YARN



Status

- Mostra informações sobre a saúde do YARN, bem como gráficos sobre as aplicações








Instances

- Mostra as instâncias e as funções associadas
- Permitir operações como iniciar e parar as funções, entrar em modo manutenção e desabilitar o serviço
- O filtro na esquerda permite selecionar as máquinas com base em parâmetros

Filters

- STATUS
 - Concerning Health 2
 - Good Health 3
- COMMISSION STATE
- MAINTENANCE MODE
- RACK
- ROLE GROUP
- ROLE TYPE
- STATE
- HEALTH TESTS

Actions for Selected ▾Add Role InstancesRole Groups

<input type="checkbox"/>	Role Type	State	Host	Commission State	Role Group
<input type="checkbox"/>	 JobHistory Server	Started	edgenode.lab	Commissioned	JobHistory Server Default Group
<input type="checkbox"/>	 NodeManager	Started	datanode03.lab	Commissioned	NodeManager Group 1
<input type="checkbox"/>	 NodeManager	Started	datanode01.lab	Commissioned	NodeManager Default Group
<input type="checkbox"/>	 NodeManager	Started	datanode02.lab	Commissioned	NodeManager Default Group
<input type="checkbox"/>	 ResourceManager (Active)	Started	headnode.lab	Commissioned	ResourceManager Default Group

YARN Applications



- Um cluster Hadoop é usado basicamente para rodar aplicações de análise de dados e o YARN é um dos principais componentes do ecossistema
- O CDH tem várias aplicações de exemplo para demonstração, como as que estão listadas abaixo
- A seguir estão os comandos para executar a aplicação em instalações do CDH com parcels e packages respectivamente
 - `$ yarn jar /opt/cloudera/parcels/CDH/lib/hadoop-mapreduce/hadoop-mapreduce-examples.jar pi 100 1000`
 - `$ yarn jar /usr/lib/hadoop-mapreduce/hadoop-mapreduce-examples.jar pi 100 1000`

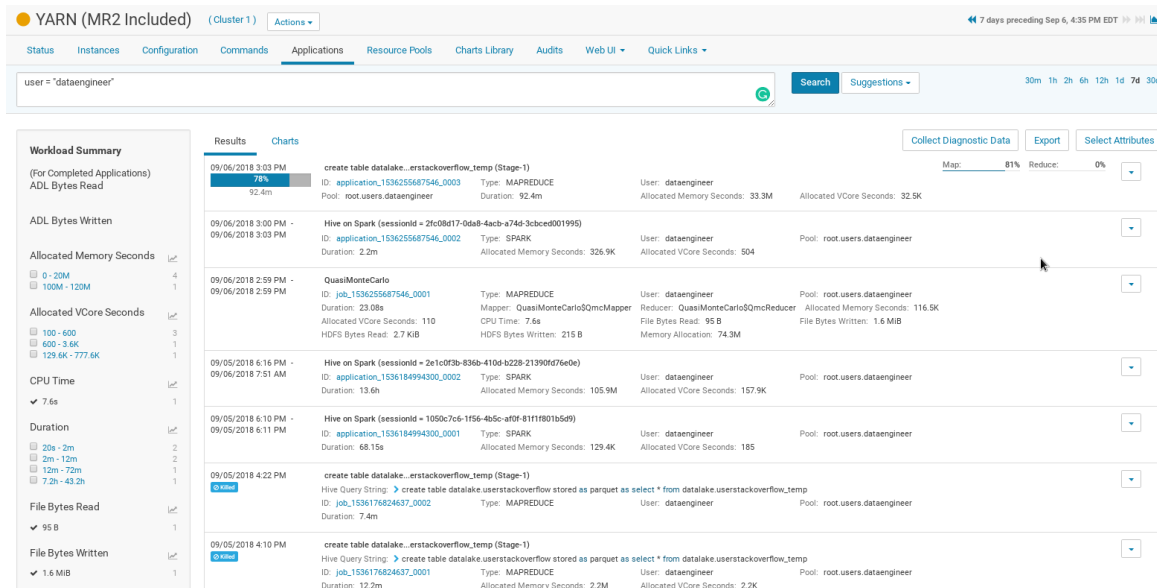
Aplicações Hadoop



- Um dos trabalhos do administrador do cluster é executar e controlar aplicações Hadoop
- O analisador-hadoop.jar processa os dados do bolsa família com o MapReduce
- O código está disponível em <https://github.com/masreis/analizador-hadoop>
- Para executar a aplicação, é necessário autenticar no Kerberos com o usuário dataengineer e usar o YARN
 - \$ kinit dataengineer
 - \$ yarn jar analisador-hadoop.jar net.marcoreis.hadoop.mapreduce.parte2.TotalPorMunicipioDriver /user/dataengineer/dados/bolsafamilia/ /user/dataengineer/saida/bolsafamilia/
- A saída do programa mostra a data, o município, o código do IBGE e o valor total
- Exemplo:
 - 201801 GUAJARA-MIRIM-0001 522890

Applications

- Mostra as aplicações em execução e concluídas
 - Permite forçar a parada das aplicações bem como a análise dos recursos utilizados



Gráficos de utilização

- Mostra graficamente a utilização do cluster e a execução das aplicações
 - Memória, vcores, CPU, duração, bytes lidos e gravados, HDFS etc.



Resource Pools

- Mostra os recursos disponíveis e alocados no cluster, agrupados em cada resource pool
 - Número de processadores, memória, containers etc.

Status

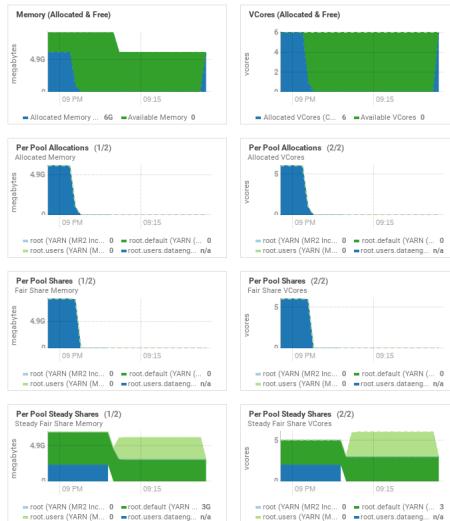
YARN is using 6 vcores and 6 GiB of memory.

Configuration

Name	Allocated Memory	Allocated VCores	Allocated Containers	Pending Containers
root	0 B	0	0	0
root.default	0 B	0	0	0
root.users	0 B	0	0	0
root.users.dataen...	6 GiB	6	6	117

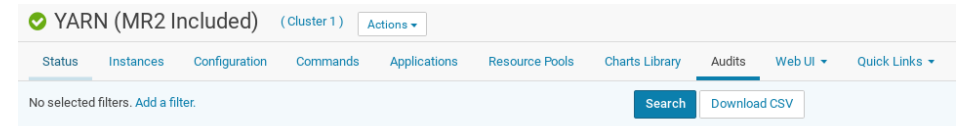
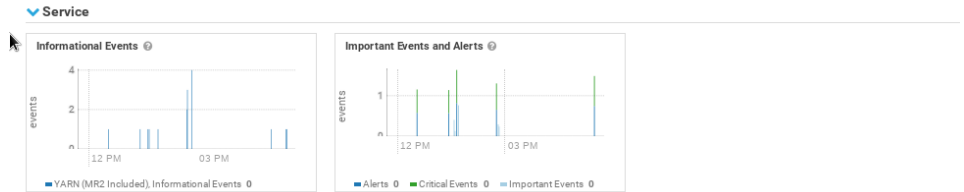
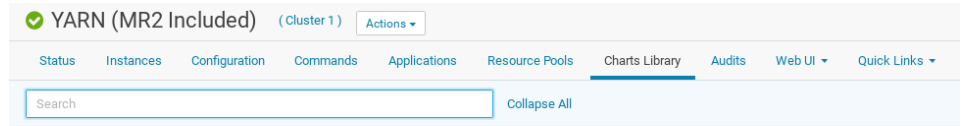
Displaying 1 - 4 of 4

Charts



Outros menus

- Charts
- Audits



September 6, 2018 2:52 PM	START Service: yarn	Succeeded command Start on service yarn
September 6, 2018 2:52 PM	START User: admin IP Address: 172.17.73.91 Service: yarn	Started command Start on service yarn
September 6, 2018 2:52 PM	YARNRECOMMISSION Service: yarn	Succeeded command YarnRecommission on service yarn
September 6, 2018 2:52 PM	Recommissioned Service: yarn Role: yarn-NODEMANAGER-a38adf8d221fd983daabf4dc34e76413	
September 6, 2018 2:52 PM	YARNRECOMMISSION	Started command YarnRecommission on service yarn

Web UI

- O ResourceManager (RM) mostra as métricas do cluster, como por exemplo:
 - Aplicações submetidas, pendentes, em execução e completadas
 - Containers em execução
 - Memória disponível e em uso
 - Processadores disponíveis e em uso
- Depois de submetida, a aplicação entra na fila de execução e pode ter os estados:
 - NEW, NEW_SAVING, SUBMITTED, ACCEPTED, RUNNING, FINISHED, FAILED e KILLED
- Os filtros na esquerda permitem ver em detalhe o comportamento das aplicações, desde os parâmetros de configuração até os bytes gravados
- O RM mantém os dados da execução, incluindo as tarefas de Map e Reduce, memória consumida, tempo de execução etc.
 - Os logs disponíveis na aplicação podem ser usados para analisar e identificar erros ou para otimizar a execução
- Após a finalização, o histórico da aplicação vai para o HistoryServer

- Cluster

- About Nodes
- Applications
- New SubMITTED
- New SAVING
- Accepted RUNNING
- Finished FAILED
- Killed
- Scheduler
- Tools

Cluster Metrics

Apps Submitted		Apps Pending		Apps Running		Apps Completed		Containers Running		Memory Used		Memory Total		Memory Reserved		VCores Used		VCores Total		VCores Reserved	
2	0	1	1	1	2	2 GB	9 GB	0 B	3	6	0	3	6	0	3	6	0	3	6	0	3

Cluster Nodes Metrics

Active Nodes		Decommissioning Nodes		Decommissioned Nodes		Lost Nodes		Unhealthy Nodes		Rebooted Nodes	
3	0	0	0	0	0	0	0	0	0	0	0

User Metrics for dr.who

Apps Submitted		Apps Pending		Apps Running		Apps Completed		Containers Running		Containers Pending		Containers Reserved		Memory Used		Memory Pending		Memory Reserved		VCores Used		VCores Pending		VCores Reserved	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

Show 20 entries

ID	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU Vcores	Allocated Memory MB	Reserved CPU Vcores	Reserved Memory MB	Progress	Tracking UI
application_1536269447911_0002	dataengineer	Hive on Spark (sessionId=c8f0-ac-bb7c-454b-b368-33911f954686)	SPARK	root.users.dataengineer	Thu Sep 6 18:42:22 -0300 2018	N/A	RUNNING	UNDEFINED	2	3	2048	0	0	<div></div>	ApplicationMaster
application_1536269447911_0001	dataengineer	select * from datablake.usersstackover.'Res' (Stage-1)	MAPREDUCE	root.users.dataengineer	Thu Sep 6 18:41:33 -0300 2018	Thu Sep 6 18:42:00 -0300 2018	FINISHED	SUCCEEDED	N/A	N/A	N/A	N/A	N/A	<div></div>	History

Showing 1 to 2 of 2 entries

Web UI

- O HistoryServer (HS) guarda o histórico de execuções das aplicações

▼ Application	Retired Jobs										
About Jobs	Show 20 ▼ entries										
Tools	Search:										
	Submit Time ↕	Start Time ↕	Finish Time ↕	Job ID ↕	Name ↕	User ↕	Queue ↕	State ↕	Maps Total ↕	Maps Completed ↕	Reduces Total ↕
											Reduces Completed ↕
	2018.09.06 17:41:33 EDT	2018.09.06 17:41:40 EDT	2018.09.06 17:42:00 EDT	job_1536269447911_0001	select * from datalake.userstackover...Reis' (Stag	dataengineer	root.users.dataengineer	SUCCEEDED	2	2	0
	2018.09.06 16:51:34 EDT	2018.09.06 16:51:40 EDT	2018.09.06 16:52:00 EDT	job_1536255687546_0004	select * from userstackoverflow wher...Reis' (Stag	dataengineer	root.users.dataengineer	SUCCEEDED	2	2	0
	2018.09.06 15:03:12 EDT	2018.09.06 15:03:18 EDT	2018.09.06 16:49:01 EDT	job_1536255687546_0003	create table datalake...erstackoverflow_temp (Stag	dataengineer	root.users.dataengineer	SUCCEEDED	11	11	0
	2018.09.06 14:59:06 EDT	2018.09.06 14:59:13 EDT	2018.09.06 14:59:37 EDT	job_1536255687546_0001	QuasiMonteCarlo	dataengineer	root.users.dataengineer	SUCCEEDED	10	10	1
	2018.09.05 16:22:23 EDT	2018.09.05 16:22:29 EDT	2018.09.05 16:29:50 EDT	job_1536176824637_0002	create table datalake...erstackoverflow_temp (Stag	dataengineer	root.users.dataengineer	KILLED	0	0	0
	2018.09.05 16:09:53 EDT	2018.09.05 16:10:02 EDT	2018.09.05 16:22:12 EDT	job_1536176824637_0001	create table datalake...erstackoverflow_temp (Stag	dataengineer	root.users.dataengineer	KILLED	0	0	0
	2018.09.05 15:07:49 EDT	2018.09.05 15:07:55 EDT	2018.09.05 15:08:23 EDT	job_1536172432786_0005	QuasiMonteCarlo	dataengineer	root.users.dataengineer	SUCCEEDED	10	10	1
	Submit Time	Start Time	Finish Time	Job ID	Name	User	Queue	State	Maps Tot	Maps Comple	Reduces To
											Reduces Con
	Showing 1 to 7 of 7 entries									First	Previous 1 Next Last

Impala



Impala Queries

- Além das abas básicas, como Status e Instances, o Impala mostra o histórico de consultas com um sumário para análise da performance a partir da memória utilizada, tempo de duração etc.

Impala (Cluster 1) Actions

Status Instances Configuration Commands **Queries** Charts Library Best Practices Audits Web UI Quick Links

Search for Impala queries, e.g. 'rows_produced > 1000', or press space to start typedhead. Search Suggestions

Workload Summary

(For Completed Queries)

Aggregate Peak Memory Usage

- 57.2 MB - 76.3 MB
- 76.3 MB - 95.4 MB
- 114.4 MB - 133.5 MB
- 133.5 MB - 152.6 MB

Duration

- 7ms - 42ms
- 252ms - 1.51s
- 9.07s - 54.69s

HDFS Bytes Read

- 114.4 MB - 171.7 MB
- 342.3 MB - 400.5 MB

Threads CPU Time

- 0ms - 700ms
- 1.4s - 2.1s
- 2.1s - 2.8s
- 2.8s - 3.5s

User

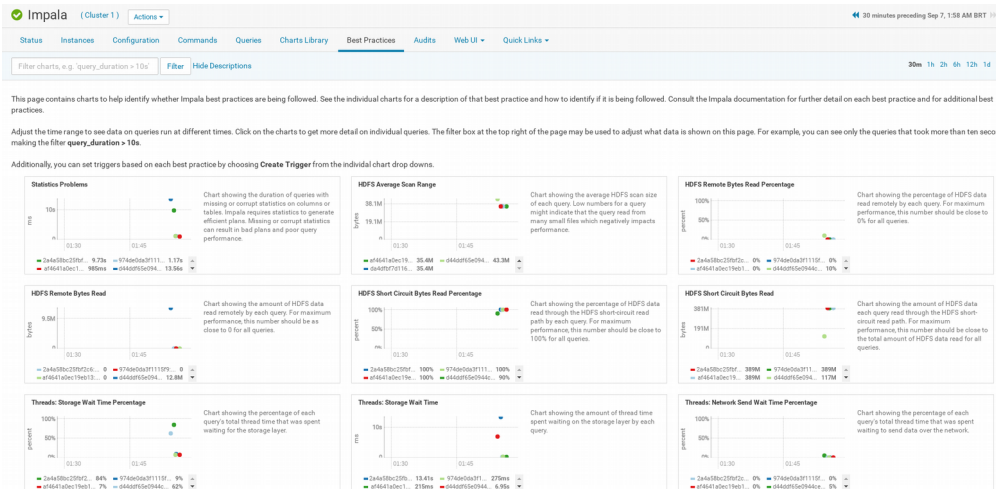
- dataengineer

Results **Charts**

09/07/2018 1:56 AM - 09/07/2018 1:56 AM	select * from userstackoverflow where displayname = 'Marco Reis'	User: dataengineer	Database: datalake	Query Type: QUERY	Aggregate Peak Memory Usage: 90.1 MB	HDFS Bytes Read: 389.3 MB	Coordinator: datanode01.lab	Threads CPU Time: 2.06s	Duration: 985ms
09/07/2018 1:55 AM - 09/07/2018 1:55 AM	select * from userstackoverflow order by downvotes desc limit 100	User: dataengineer	Database: datalake	Query Type: QUERY	Aggregate Peak Memory Usage: 129.6 MB	HDFS Bytes Read: 389.3 MB	Coordinator: datanode01.lab	Threads CPU Time: 2.95s	Duration: 1.17s
09/07/2018 1:55 AM - 09/07/2018 1:55 AM	select * from userstackoverflow order by downvotes desc limit 10	User: dataengineer	Database: datalake	Query Type: QUERY	Aggregate Peak Memory Usage: 150.9 MB	HDFS Bytes Read: 389.3 MB	Coordinator: datanode01.lab	Threads CPU Time: 3.13s	Duration: 1.15s
09/07/2018 1:55 AM - 09/07/2018 1:55 AM	select * from userstackoverflow order by downvotes limit 100	User: dataengineer	Database: datalake	Query Type: QUERY	Aggregate Peak Memory Usage: 91.6 MB	HDFS Bytes Read: 389.3 MB	Coordinator: datanode01.lab	Threads CPU Time: 2.8s	Duration: 9.73s
09/07/2018 1:54 AM - 09/07/2018 1:54 AM	select * from userstackoverflow limit 100	User: dataengineer	Database: datalake	Query Type: QUERY	Aggregate Peak Memory Usage: 70.9 MB	HDFS Bytes Read: 129.9 MB	Coordinator: datanode01.lab	Threads CPU Time: 38ms	Duration: 13.56s
09/07/2018 1:54 AM - 09/07/2018 1:54 AM	show tables	User: dataengineer	Database: datalake	Query Type: DDL			Coordinator: datanode01.lab		Duration: 7ms
09/07/2018 1:53 AM - 09/07/2018 1:53 AM	use datalake	User: dataengineer	Database: default	Query Type: DDL			Coordinator: datanode01.lab		Duration: 35ms
09/07/2018 1:53 AM - 09/07/2018 1:53 AM	show databases	User: dataengineer	Database: default	Query Type: DDL			Coordinator: datanode01.lab		Duration: 293ms

Best Practices

- A aba de Best Practices apresenta estatísticas de utilização dos recursos para a execução das consultas
- A partir desses painéis é possível analisar e melhorar performance das consultas



Spark



History Server Web UI

- A única aba importante é a do History Server, que é similar à do HistoryServer do YARN



Event log directory: hdfs://headnode.lab:8020/user/spark/applicationHistory

Showing 1-13 of 13

1

App ID	App Name	Started	Completed	Duration	Spark User	Last Updated
application_1536286602831_0002	Hive on Spark (sessionId = 076acd25-3850-40fa-83bd-2b47160307cf)	2018/09/06 23:27:54	2018/09/06 23:40:14	12 min	dataengineer	2018/09/06 23:40:15
application_1535716419142_0001	Hive on Spark (sessionId = 69a41efb-98e1-4848-9f81-9dd225b40d20)	2018/08/31 09:14:19	2018/08/31 09:15:21	1.0 min	dataengineer	2018/08/31 09:15:21
application_1535412066373_0001	Hive on Spark (sessionId = d08e3df6-0cfd-437d-a4f5-168aa196ab3d)	2018/08/27 20:24:08	2018/08/28 07:23:32	11.0 h	dataengineer	2018/08/28 07:23:33
application_1535409918354_0001	Hive on Spark (sessionId = 09d53677-a39b-4756-92a2-86a649145950)	2018/08/27 20:03:14	2018/08/27 20:04:46	1.5 min	dataengineer	2018/08/27 20:04:46
application_1535156973059_0012	Hive on Spark (sessionId = 9951937a-d412-435c-aa4a-cb2b2486e16b)	2018/08/24 23:58:13	2018/08/25 00:20:31	22 min	dataengineer	2018/08/25 00:20:31
application_1535156973059_0010	Hive on Spark (sessionId = 9951937a-d412-435c-aa4a-cb2b2486e16b)	2018/08/24 23:40:38	2018/08/24 23:57:40	17 min	dataengineer	2018/08/24 23:57:41
application_1535156973059_0009	Hive on Spark (sessionId = 83779b6e-cb6b-4534-b39a-80c166ec2af4)	2018/08/24 22:53:14	2018/08/24 23:39:13	46 min	dataengineer	2018/08/24 23:39:14
application_1535156973059_0008	Hive on Spark (sessionId = af0586f1-b5fe-4b10-9906-02af635af748)	2018/08/24 22:33:19	2018/08/24 22:34:37	1.3 min	dataengineer	2018/08/24 22:34:37
application_1535156973059_0007	Hive on Spark (sessionId = aa3b6e1f-c326-406a-8296-d2b97d670993)	2018/08/24 22:19:31	2018/08/24 22:22:17	2.8 min	dataengineer	2018/08/24 22:22:17
application_1535070440048_0011	Hive on Spark (sessionId = c1d34665-9cbf-4825-a9b5-a2f12d028552)	2018/08/23 23:35:19	2018/08/24 00:10:46	35 min	marcoreis	2018/08/24 00:10:47
application_1535070440048_0010	Hive on Spark (sessionId = 978d2c44-3623-4266-85bc-7a619249221f)	2018/08/23 23:08:57	2018/08/23 23:34:02	25 min	marcoreis	2018/08/23 23:34:02
application_1535070440048_0005	Hive on Spark (sessionId = dddb7497-b284-4bf9-8114-15dca2168250)	2018/08/23 21:59:07	2018/08/23 23:07:37	1.1 h	hive	2018/08/23 23:07:38
application_1535070440048_0002	Hive on Spark (sessionId = 80b451dd-9af2-4e21-ba76-d267b3acc32b)	2018/08/23 21:50:20	2018/08/23 21:55:57	5.6 min	hive	2018/08/23 21:55:58

[Show incomplete applications](#)

Cloudera Manager



CM Users

- A administração dos usuários está disponível no menu superior Administration → Users
- O CM define funções com níveis de acesso diferentes, como por exemplo:
 - Auditor
 - Read-only
 - Dashboard
 - Full administrator
- Na figura abaixo foram criados alguns usuários de exemplo com os perfis indicados

Actions for Selected ▾ [Add User](#)

<input type="checkbox"/> Username	⬆ Role	⬆ User Type
<input type="checkbox"/> dashboard	Dashboard User	Cloudera Manager
<input type="checkbox"/> configurator	Configurator	Cloudera Manager
<input type="checkbox"/> admin	Full Administrator	Cloudera Manager

Showing 1 to 3 of 3 entries

Reports

- O CDH gera relatórios de utilização para download em formato CSV e XLS
- As opções incluem informações sobre:
 - Utilização de disco
 - Acesso a diretórios
 - Aplicações YARN
 - Consultas Impala
- Está disponível no menu Clusters → Reports

cloudera MANAGER

[Support](#)[admin](#)[Clusters](#)[Hosts](#)[Diagnostics](#)[Audits](#)[Charts](#)[Backup](#)[Administration](#)

Reports (Cluster 1)

Disk Usage (HDFS)

Title	Download
Current Disk Usage By User	CSV XLS
Current Disk Usage By Group	CSV XLS
Current Disk Usage By Directory	CSV XLS
Historical Disk Usage By User	CSV XLS
Historical Disk Usage By Group	CSV XLS
Historical Disk Usage By Directory	CSV XLS
Directory Usage	

User Access (HDFS)

Title	Download
Directory Access By Group	CSV XLS

Applications (MR2 jobs included) ([YARN \(MR2 Included\)](#))

[Configure Aggregated Metrics](#)

Hosts

- O menu superior Hosts → All Hosts permite a visualização do estado do cluster no nível dos servidores
- Mostra detalhes como nome, IP, estado, disco e memória
- Atenção para o menu Filters, que permite selecionar os hosts a partir de diversos parâmetros
- Permite fazer manutenção de host ou removê-lo do cluster

Hosts

Configuration Add New Hosts to Cluster Re-run Upgrade Wizard Inspect All Hosts

Search

Filters

- STATUS
 - Good Health 5
- CLUSTERS
- CORES
- COMMISSION STATE
- LAST HEARTBEAT
- LOAD (1 MINUTE)
- LOAD (5 MINUTES)
- LOAD (15 MINUTES)
- MAINTENANCE MODE
- RACK
- SERVICES
- HEALTH TESTS

Actions for Selected





































Status	Name	IP	Roles	Commission State	Last Heartbeat	Load Average	Disk Usage	Physical Memory	Swap Space
Good Health	datanode01.lab	192.168.122.12	5 Role(s)	Commissioned	12.81s ago	0.00 0.02 0.00	28.7 GiB / 77.5 GiB	942.7 MiB / 3.9 GiB	0 B / 960 MiB
Good Health	datanode02.lab	192.168.122.13	5 Role(s)	Commissioned	12.8s ago	0.00 0.01 0.05	30.5 GiB / 77.5 GiB	1016.1 MiB / 3.9 GiB	2.1 MiB / 960 MiB
Good Health	datanode03.lab	192.168.122.14	6 Role(s)	Commissioned	12.76s ago	0.08 0.05 0.07	15 GiB / 38.3 GiB	1.1 GiB / 3.9 GiB	1.8 MiB / 960 MiB
Good Health	edgenode.lab	192.168.122.11	16 Role(s)	Commissioned	12.78s ago	0.11 0.59 0.69	10.8 GiB / 38.3 GiB	2.3 GiB / 3.9 GiB	3.5 MiB / 960 MiB
Good Health	headnode.lab	192.168.122.10	9 Role(s)	Commissioned	12.79s ago	0.47 0.38 0.38	16.6 GiB / 38.3 GiB	5.7 GiB / 11.7 GiB	0 B / 960 MiB

Columns: 10 Selected

Roles

- O menu Hosts → Roles mostra as funções associadas a cada host

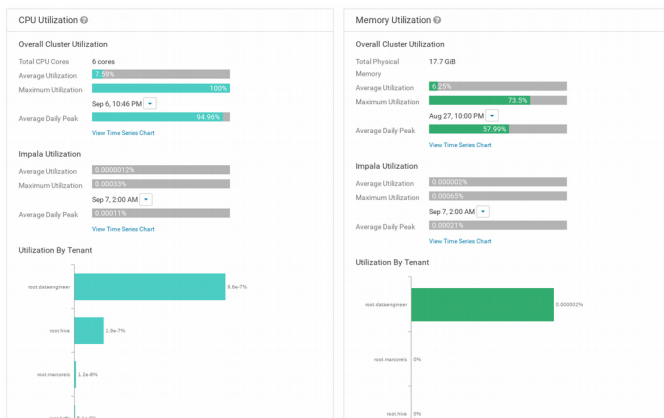
Cluster 1

Hosts	Count	Roles
datanode03.lab	1	 DN  G  ID  G  NM  S
datanode[01-02].lab	2	 DN  G  ID  G  NM
edgenode.lab	1	 A  B  HFS  SNN  HMS  HS2  LB  HS  KTR  ICS  ISS  OS  HS  G  JHS  S
headnode.lab	1	 NN  AM  AP  ES  HM  RM  SM  RM  S

This table is grouped by hosts having the same roles assigned to them.

Utilization Report

- O relatório de utilização lista o percentual de memória e CPU para cada usuário
- A partir destes dados é possível definir uma política de limitação de uso para cada usuário
 - Opção para mostrar relatórios sobre o YARN e Impala



Cloudera Management Service

- O Cloudera Management Service centraliza as funções de gerenciamento do CDH
 - Activity Monitor: atividades do MapReduce
 - Host Monitor: métricas e estado dos hosts
 - Service Monitor: métricas e estado do YARN e Impala
 - Event Server: grava eventos ocorridos no cluster e permite executar consultas
 - Alert Publisher: envia alertas para os eventos
 - Report Manager: relatórios sobre o histórico da utilização do cluster
- Selecione o menu Clusters → Cloudera Management Service

Alert Publisher



- Os alertas são eventos considerados importantes que ocorrem durante a utilização do cluster
- Um evento pode ser informacional, importante, advertência ou crítico
 - Advertências não podem ser ignoradas, principalmente aquelas sobre espaço em disco
- São visualizados na página principal do CM, mas podem ser configurados para enviar mensagens para o administrador
 - É necessário ter acesso a um servidor de e-mail
- Configure o nível de alerta de acordo com sua conveniência
 - Atenção para o Maximum Batch Interval e Logging Threshold

The screenshot shows the Cloudera Manager interface for configuring the Alert Publisher. The top navigation bar includes 'cloudera MANAGER', 'Clusters', 'Hosts', 'Diagnostics', 'Audits', 'Charts', and 'Backup'. Below this is the 'Cloudera Management Service' header with an 'Actions' button. The main navigation tabs are 'Status', 'Instances', 'Configuration' (selected), 'Commands', 'Charts Library', 'Audits', and 'Quick Links'. A search bar is present below the tabs.

The 'Configuration' page is divided into two main sections: 'Filters' and 'Alerts'.

Filters:

- SCOPE:** A list of services with their counts. 'Alert Publisher' is highlighted with a count of 63.
- CATEGORY:** A list of categories with their counts. 'Advanced' is selected with a count of 9.


Alerts:

- Alerts: Enable Email Alerts:** A checkbox that is checked.
- Alert Publisher Default Group:** A dropdown menu showing 'Alert Publisher Default Group'.
- Alerts: Mail Server Protocol:** A dropdown menu showing 'smtp'.
- Alerts: Mail Server Hostname:** A text input field containing 'webmail.marcoreis.net'.
- Alerts: Mail Server Username:** A text input field containing 'ma@marcoreis.net'.
- Alerts: Mail Server Password:** A text input field containing '*****'.
- Alert: Mail From Address:** A text input field containing 'ma@marcoreis.net'.
- Alerts: Mail Message Recipients:** A text input field containing 'reis_marco@hotmail.com'.
- Alert Publisher: Maximum Batch Interval:** A dropdown menu showing '1' and 'day(s)'.
- Alert Publisher Logging Threshold:** A dropdown menu showing 'INFO'.

Mensagem de alerta

- Na imagem ao lado vemos um exemplo de mensagem de alerta enviada por e-mail
- A mensagem ao lado indica que um serviço (ZooKeeper) parou e outro voltou a funcionar

[Cloudera Alert] 2 Alerts since 1:17:11 AM

 ma@marcoreis.net
Via 07/09/2018, 04:24 AM
Unread

Health test changes: 1 Became Bad, 1 Became Good
Time: Sep 7, 2018 1:17:11 AM

[View Details on headnode.lab](#)

Monitor Startup: true
Cluster: cluster
Cluster Display Name: Cluster 1
Service: zookeeper
Service Display Name: ZooKeeper
Service Type: ZooKeeper

Health Test Results:

Health Test Name	Event Code	Severity	Content
ZOOKEEPER_CANARY_HEALTH	Service health test bad	Critical	The health test result for ZOOKEEPER_CANARY_HEALTH has become bad: The ZooKeeper service canary failed for an unknown reason.

Health test changes: 1 Became Bad, 1 Became Good
Time: Sep 7, 2018 1:24:12 AM

[View Details on headnode.lab](#)

Monitor Startup: true
Cluster: cluster
Cluster Display Name: Cluster 1
Service: zookeeper
Service Display Name: ZooKeeper
Service Type: ZooKeeper

Health Test Results:

Health Test Name	Event Code	Severity	Content
ZOOKEEPER_CANARY_HEALTH	Service health test bad	Critical	The health test result for ZOOKEEPER_CANARY_HEALTH has become bad: The ZooKeeper service canary failed for an unknown reason.

Dúvidas?

Marco Reis
<http://marcoreis.net>