

PO12: Przetwarzanie Big Data 22/23-Z

Strona główna / Moje kursy / WiIT / Informatyka / Stacjonarne / I stopień / Semestr 7 [WiIT-Inf-st-I] / PO12: Przetwarzanie Big Data 22/23-Z / 3. HDFS, YARN / HDFS, YARN - zadania - zarejestruj swoje rozwiązania tutaj

Rozpoczęto	piątek, 21 października 2022, 21:40
Stan	Ukończone
Ukończono	piątek, 21 października 2022, 21:45
Wykorzystany czas	4 min. 54 sek.
Punkty	29,00/30,00
Ocena	9,67 pkt. na 10,00 pkt. możliwych do uzyskania (96,67%)

Nawigacja w teście

1	2	3	4	5	6
7	8	9			

Pokaż jedną stronę na raz

Zakończ przegląd

09

First, use the terminal to check the following values related to the entire HDFS file system

a.

What is its total size?

39335568

b.

How many blocks does it store?

3

c.

How many files does it have?

3

d.

What is the average block size?

13111856

e.

What is the default replication factor?

2

f.

Could the replication factor be higher with the current Hadoop cluster configuration (consider the number of data nodes in HDFS)?

No

Pytanie 1

Poprawnie

Punkty: 6,00 z 6,00

Oflaguj pytanie

10

Check how many blocks our *input* directory contains.

Odpowiedź: 3

Poprawna odpowiedź to: 3

Pytanie 2

Poprawnie

Punkty: 1,00 z 1,00

Oflaguj pytanie

11

There are additional questions about the station.csv.

a.

How many blocks does it contain?

1

b.

How many racks are used by blocks of this file?

1

c.

How many data nodes are used by the blocks of this file?

2

d.

What is the average replication factor of blocks of this file?

2

Pytanie 3

Poprawnie

Punkty: 4,00 z 4,00

Oflaguj pytanie

13

Use the first command (hdfs dfsadmin -report) to get answers to the following questions:

a.

How many DataNodes are used?

2

b.

How much data space is still available (*DFS Remaining*) at cluster level?

74716868608

c.

How much data space is available (*Configured Capacity*) on each DataNode?

52589998080

Pytanie 4

Poprawnie

Punkty: 3,00 z 3,00

Oflaguj pytanie

14

Use the second command (hdfs getconf -confKey <key>) to get answers to the questions below.

a.

Where (in what directory on the local file system) is data managed by the DataNodes stored (dfs.datanode.data.dir)?

/hadoop/dfs/data

b.

What is the maximum number of threads that support data transfer between clients and the data node (dfs.datanode.max.transfer.threads)?

hdfs getconf -confkey dfs.datanode.max.transfer.threads

4096

c.

What is the maximum number of files managed by NameNode? Assume the maximum size of NameNode Java heap size is 1 GB and the replication factor is 1.

Tips on how to calculate it you can find at: https://www.cloudera.com/documentation/enterprise/5-8-x/topics/admin_nn_memory_config.html

3333333

Pytanie 5

Poprawnie

Punkty: 3,00 z 3,00

Oflaguj pytanie

15

Use the Google Cloud Platform console and page related to your Dataproc cluster. Click on the *Web Interfaces* tab, and run the page with the NameNode interface available at the *HDFS NameNode* link.

Use this interface to answer the following questions

a.

How many DataNodes are active?

2

b.

What is the actual size of *Max Heap Memory*? For the actual replication factor, calculate the maximum number of files that the NameNode can manage.

Maximum number of files: 480000000

Note that you can browse the HDFS file system using this interface.

Pytanie 6

Częściowo poprawnie

Punkty: 1,00 z 2,00

Oflaguj pytanie

19

Answer the following questions

a.

How much memory (**MB**) is allocated to all containers (yarn.nodemanager.resource.memory-mb)?

6144

b.

What is the smallest amount of memory allocated per one container (yarn.scheduler.minimum-allocation-mb)?

1

c.

What percentage of physical processors on the work node can be used by containers (yarn.nodemanager.resource.percentage-physical-cpu-limit)?

100

d.

How many virtual CPU cores have been declared (yarn.nodemanager.resource.cpu-vcores)?

2

Pytanie 7

Poprawnie

Punkty: 4,00 z 4,00

Oflaguj pytanie

29

Answer the following questions:

a.

What was the name of the MapReduce application (Job Name)? Where did it come from? Where can you change it?

b.

Under which queue was this application executed?

Name of the queue: default

c.

What is the current *finalStatus* of this application?

Name of the *finalStatus*: SUCCEEDED

d.

How many Map task *units* have been used?

1

e.

How many Reduce task *units* have been used?

3

Pytanie 8

Poprawnie

Punkty: 4,00 z 4,00

Oflaguj pytanie

30

Go to the Counters page (*Job*->*Counters*) for this job

a.

How many bytes have been read from HDFS?

5466

b.

How many bytes have been saved to HDFS?

121

c.

How many bytes have been saved to local file systems?

986677

Pytanie 9

Poprawnie

Punkty: 3,00 z 3,00

Oflaguj pytanie

Zakończ przegląd