

क्लाउड कंप्यूटिंग और बिग डेटा का परिचय

इस पाठ में निम्नलिखित विषय शामिल हैं:

- ☐ स्पार्क
- ☐ हड्डप
- ☐ कुबेरनेट्स
- ☐ डॉकर
- ☐ प्लिंक
- ☐ मोंगोडीबी

[illegible]

आधिकारिक वेबसाइट के अनुसार, Spark बड़े पैमाने पर डेटा के विश्लेषण के लिए एक इंजन है। Spark एक लाइब्रेरी का सेट है। यह Redis की तरह सर्वर और क्लाइंट में विभाजित नहीं होता है। Spark केवल क्लाइंट साइड पर उपयोग किया जाता है। आधिकारिक वेबसाइट से नवीनतम संस्करण डाउनलोड किया गया है, `spark-3.1.1-bin-hadoop3.2.tar`।

```
$ tree . -L 1
```

LICENSE
NOTICE
R
README.md
RELEASE
bin
conf
data
examples
jars
kubernetes
licenses
python
sbin
yarn

11 डायरेक्टरीज़, 4 फ़ाइलें

```
(libraries)

, , Python (dependency library) `pip install pyspark`

```shell
$ pip install pyspark
Collecting pyspark
 Downloading pyspark-3.1.1.tar.gz (212.3 MB)
 | | 212.3 MB 14 kB/s
Collecting py4j==0.10.9
 Downloading py4j-0.10.9-py2.py3-none-any.whl (198 kB)
 | | 198 kB 145 kB/s
Building wheels for collected packages: pyspark
 Building wheel for pyspark (setup.py) ... done
 Created wheel for pyspark: filename=pyspark-3.1.1-py2.py3-none-any.whl size=212767604 sha256=0b8079e8...
 Stored in directory: /Users/lzw/Library/Caches/pip/wheels/23/bf/e9/9f3500437422e2ab82246f25a51ee480a4...
Successfully built pyspark
Installing collected packages: py4j, pyspark
Successfully installed py4j-0.10.9 pyspark-3.1.1
```

इंस्टॉल हो गया है।

यह वेबसाइट पर कुछ उदाहरण देख रहा है।

```
./bin/run-example SparkPi 10
```

ओह, मैंने अभी डाउनलोड किए गए इंस्टॉलेशन पैकेज में मौजूद प्रोग्राम को चलाने की कोशिश की। लेकिन एक त्रुटि हुई।

```
$./bin/run-example SparkPi 10
21/03/11 00:06:15 WARN NativeCodeLoader: - ... , -
21/03/11 00:06:16 INFO ResourceUtils: spark.driver
21/03/11 00:06:16 WARN Utils: 'sparkDriver'
```

□□□□□ एक तेज़ और सामान्य प्रसंस्करण इंजन है जो □□□□□□ डेटा के साथ संगत है। यह □□□□□ या □□□□□□ के स्टैंडअलोन मोड के माध्यम से □□□□□□ क्लस्टर में चल सकता है, और यह □□□□□, □□□□□□, □□□□□□□□□□□, □□□□□, और किसी भी □□□□□□□ □□□□□□□□□□□□□ में डेटा को प्रोसेस कर सकता है। इसे बैच प्रोसेसिंग (□□□□□□□□□□□ के समान) और नए वर्कलोड जैसे स्ट्रीमिंग, इंटरएक्टिव क्वेरीज़, और मशीन लर्निंग को संचालित करने के लिए डिज़ाइन किया गया है।

hadoop कई बार आया है। spark depends hadoop को ँँँँँँ करने के बाद, मुझे यह वाक्य मिला। ऐसा लगता है कि यह Hadoop फॉर्मेट के डेटा पर निर्भर करता है। आइए पहले Hadoop पर शोध करें।

ँँँँँँ

आधिकारिक वेबसाइट को संक्षेप में देखने के बाद, इसे इंस्टॉल करने का प्रयास करते हैं।

```
brew install hadoop
```

इंस्टॉलेशन के दौरान, आइए इसके बारे में थोड़ा जानें।

ँँँँँँ ँँँँँँँ सॉफ्टवेयर लाइब्रेरी एक ऐसा फ्रेमवर्क है जो सरल प्रोग्रामिंग मॉडल का उपयोग करके कंप्यूटरों के क्लस्टर पर बड़े डेटा सेट के वितरित प्रसंस्करण की अनुमति देता है। यह सिंगल सर्वर से हजारों मशीनों तक स्केल करने के लिए डिज़ाइन किया गया है, जहां प्रत्येक मशीन स्थानीय गणना और भंडारण प्रदान करती है। हार्डवेयर पर निर्भर होने के बजाय उच्च उपलब्धता प्रदान करने के लिए, यह लाइब्रेरी स्वयं एप्लिकेशन लेयर पर विफलताओं का पता लगाने और उन्हें संभालने के लिए डिज़ाइन की गई है, इस प्रकार कंप्यूटरों के एक क्लस्टर के ऊपर एक अत्यधिक उपलब्ध सेवा प्रदान करती है, जहां प्रत्येक मशीन विफलताओं के प्रति संवेदनशील हो सकती है।

ँँँँँँँ एक ऐसा फ्रेमवर्क है जो वितरित डेटासेट को प्रोसेस करने के लिए बनाया गया है। ये डेटासेट कई कंप्यूटरों पर वितरित हो सकते हैं। इसे एक बहुत ही सरल प्रोग्रामिंग मॉडल का उपयोग करके प्रोसेस किया जाता है। इसे एकल सर्वर से हजारों मशीनों तक स्केल करने के लिए डिज़ाइन किया गया है। हार्डवेयर की उच्च उपलब्धता पर निर्भर होने के बजाय, यह लाइब्रेरी एप्लिकेशन लेयर पर ही त्रुटियों की जांच और उन्हें संभालने के लिए डिज़ाइन की गई है। इस तरह, उच्च उपलब्धता वाली सेवाओं को क्लस्टर में तैनात किया जा सकता है, भले ही क्लस्टर में शामिल हर कंप्यूटर विफल हो सकता हो।

```
$ brew install hadoop
```

```
:
```

```
homebrew-core
```

```
homebrew-cask
```

```
`brew update` , :
```

```
git -C /usr/local/Homebrew/Library/Taps/homebrew/homebrew-core fetch --unshallow
```

```
git -C /usr/local/Homebrew/Library/Taps/homebrew/homebrew-cask fetch --unshallow
```

```
GitHub
```

```
Homebrew/homebrew-core Homebrew/homebrew-cask
```

```
==> Downloading https://homebrew.bintray.com/bottles/openjdk-15.0.1.big_sur.bottle.tar.gz
```

```
: /Users/lzw/Library/Caches/Homebrew/downloads/d1e3ece4af1d225bc2607eaa4ce85a873d2c6d43757
```

```
==> Downloading https://www.apache.org/dyn/closer.lua?path=hadoop/common/hadoop-3.3.0/hadoop-3.3.0.tar.gz
```

```
: /Users/lzw/Library/Caches/Homebrew/downloads/764c6a0ea7352bb8bb505989fdee1b36dc628c2dcd6
```

```
Java JDK , :
sudo ln -sf /usr/local/opt/openjdk/libexec/openjdk.jdk /Library/Java/JavaVirtualMachines/openjdk.jdk
```

यदि आपको अपने `~/.bashrc` में `~/.bash_profile` को पहले रखने की आवश्यकता है, तो निम्नलिखित कमांड चलाएँ: `bash echo 'export PATH="/usr/local/opt/openjdk/bin:$PATH"' >> /Users/lzw/.bash_profile`

कंपाइलर्स को ंंकूने के लिए आपको यह सेट करने की आवश्यकता हो सकती है: ंंकूने ंंकूने=""-  
 ////////////////////"

==> सारांश `/opt/anaconda3/anaconda3/anaconda3/15.0.1: 614 फ़ाइलें, 324.9MB ==> anaconda3 इंस्टॉल कर रहा है`  
`/opt/anaconda3/anaconda3/anaconda3/3.3.0: 21,819 फ़ाइलें, 954.7MB, 2 मिनट 15 सेकंड में बनाया गया ==> 1 निर्भर को`  
 अपग्रेड कर रहा है: `anaconda3 3.3.3 -> 3.6.3_1 ==> anaconda3 3.3.3 -> 3.6.3_1 को अपग्रेड कर रहा है ==> डाउनलोड कर रहा है`  
`anaconda3://anaconda.anaconda.org/anaconda/anaconda3/3.6.3/anaconda3/anaconda3-3.6.3-000.000.00 ==> डाउनलोड कर रहा है`  
`anaconda3://anaconda.anaconda.org/anaconda/anaconda3/3.6.3/anaconda3/anaconda3-3.6.3-000.000.00 से #####`  
 100.0% त्रुटि: `brew link` चरण सफलतापूर्वक पूरा नहीं हुआ फ़ॉर्मूला बनाया गया, लेकिन `/opt/anaconda3` में सिमलिक नहीं किया गया `anaconda3` को सिमलिक नहीं किया जा सका लक्ष्य `/opt/anaconda3/opt/anaconda3` का एक सिमलिक है। आप इसे अनलिक कर सकते हैं: `anaconda3 anaconda3 anaconda3`

लिंक को बाध्य करने और सभी संघर्षरत फ़ाइलों को ओवरराइट करने के लिए: 0000 0000 -000000000 00000

सभी फ़ाइलों को सूचीबद्ध करने के लिए जो हटा दी जाएंगी: 0000 0000 -0000000000 -000-000 000000

[illegible]

00000000 0000-0000 है, जिसका अर्थ है कि इसे /0000/000000 में 00000000 नहीं किया गया है, क्योंकि यह 000000 के java 00000000 को छिपाता है।

यदि आपको अपने `PATH` में `openjdk` को सबसे पहले रखने की आवश्यकता है, तो निम्नलिखित कमांड चलाएँ: `bash` `echo`  
`'export PATH="/usr/local/opt/openjdk/bin:$PATH"' >> /Users/lzw/.bash_profile`

कंपाइलर को `brew` इंस्टॉल करने के लिए आपको यह सेट करने की आवश्यकता हो सकती है: `export CXX="g++-7"`  
`export CC="clang-7"`

```
brew install maven
brew link --overwrite maven
```

(यह कमांड `brew` को लिंक करने और पहले से मौजूद फाइलों को ओवरराइट करने के लिए उपयोग की जाती है।)

Hadoop सफलतापूर्वक इंस्टॉल हो गया है।

## मॉड्यूल

इस प्रोजेक्ट में निम्नलिखित मॉड्यूल शामिल हैं:

- `hadoop` मॉड्यूल: अन्य `hadoop` मॉड्यूल का समर्थन करने वाले सामान्य उपयोगिताएँ।
- `hadoop-hdfs` मॉड्यूल: एक वितरित फ़ाइल सिस्टम जो एप्लिकेशन डेटा तक उच्च-श्रृंखला पहुंच प्रदान करता है।
- `hadoop-yarn` मॉड्यूल: जॉब शेड्यूलिंग और क्लस्टर संसाधन प्रबंधन के लिए एक फ्रेमवर्क।
- `hadoop-mapreduce` मॉड्यूल: बड़े डेटा सेट के समानांतर प्रसंस्करण के लिए एक `MapReduce`-आधारित सिस्टम।
- `hadoop-common` मॉड्यूल: `hadoop` के लिए एक ऑब्जेक्ट स्टोर।

यह कहा जा रहा है कि ये मॉड्यूल हैं। इसे टाइप करने पर `hadoop` दिखाई देगा:

```
$ hadoop
: hadoop [OPTIONS] SUBCOMMAND [SUBCOMMAND OPTIONS]
 hadoop [OPTIONS] CLASSNAME [CLASSNAME OPTIONS]
 CLASSNAME - Java

OPTIONS:
 --config dir Hadoop
 --debug
 --help
 buildpaths
```

hostnames list[,of,host,names]

hosts filename

loglevel level log4j

workers

SUBCOMMAND :

:

daemonlog /

:

archive Hadoop

checknative Hadoop

classpath Hadoop jar

conftest XML

credential

distch

distcp

dtutil

envvars Hadoop

fs

gridmix ,

jar <jar> jar : YARN "yarn jar" ,

jnipath java.library.path

kdiag Kerberos

kerbname auth\_to\_local

key KeyProvider

rumenfolder rumen

rumentrace rumen

s3guard S3

trace Hadoop

version

:

```
kms KMS ,
registrydns DNS
```

```
SUBCOMMAND -h
```

आधिकारिक वेबसाइट पर कुछ उदाहरण दिए गए हैं।

```
$ mkdir input
$ cp etc/hadoop/*.xml input
$ bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-3.2.2.jar grep input output 'dfs[a-z]'
$ cat output/*
```

ध्यान दें कि `share/hadoop/mapreduce/hadoop-mapreduce-examples-3.2.2.jar` मौजूद है। इसका मतलब यह हो सकता है कि कुछ उदाहरण फ़ाइलें हमें नहीं मिली हैं। अनुमान है कि Homebrew के माध्यम से इंस्टॉल करने पर ये फ़ाइलें नहीं मिलती हैं। हमने आधिकारिक वेबसाइट से इंस्टॉलेशन पैकेज डाउनलोड किया है।

```
$ tree . -L 1
.
├── LICENSE-binary
├── LICENSE.txt
├── NOTICE-binary
├── NOTICE.txt
├── README.txt
├── bin
├── etc
├── include
├── lib
├── libexec
├── licenses-binary
├── sbin
└── share
```

`share` डायरेक्टरी दिखाई दी है। हालांकि, क्या Homebrew वास्तव में इन अतिरिक्त फ़ाइलों को शामिल नहीं करता है? Homebrew के इंस्टॉलेशन डायरेक्टरी को ढूंढें।

```
$ type hadoop
hadoop /usr/local/bin/hadoop
```

```
$ ls -alrt /usr/local/bin/hadoop
lrwxr-xr-x 1 lzw admin 33 Mar 11 00:48 /usr/local/bin/hadoop -> ../Cellar/hadoop/3.3.0/bin/hadoop
$ cd /usr/local/Cellar/hadoop/3.3.0
```

यह /usr/local/Cellar/hadoop/3.3.0/libexec/share/hadoop के अंतर्गत प्रिंट की गई डायरेक्टरी ट्री है।

```
$ tree . -L 2
```

```
.
├── client
│ ├── hadoop-client-api-3.3.0.jar
│ ├── hadoop-client-miniclustert-3.3.0.jar
│ └── hadoop-client-runtime-3.3.0.jar
├── common
│ ├── hadoop-common-3.3.0-tests.jar
│ ├── hadoop-common-3.3.0.jar
│ ├── hadoop-kms-3.3.0.jar
│ ├── hadoop-nfs-3.3.0.jar
│ ├── hadoop-registry-3.3.0.jar
│ ├── jdiff
│ ├── lib
│ ├── sources
│ └── webapps
├── hdfs
│ ├── hadoop-hdfs-3.3.0-tests.jar
│ ├── hadoop-hdfs-3.3.0.jar
│ ├── hadoop-hdfs-client-3.3.0-tests.jar
│ ├── hadoop-hdfs-client-3.3.0.jar
│ ├── hadoop-hdfs-httpfs-3.3.0.jar
│ ├── hadoop-hdfs-native-client-3.3.0-tests.jar
│ ├── hadoop-hdfs-native-client-3.3.0.jar
│ ├── hadoop-hdfs-nfs-3.3.0.jar
│ ├── hadoop-hdfs-rbf-3.3.0-tests.jar
│ ├── hadoop-hdfs-rbf-3.3.0.jar
│ ├── jdiff
│ ├── lib
│ ├── sources
│ └── webapps
```



## mapreduce

- hadoop-mapreduce-client-app-3.3.0.jar
- hadoop-mapreduce-client-common-3.3.0.jar
- hadoop-mapreduce-client-core-3.3.0.jar
- hadoop-mapreduce-client-hs-3.3.0.jar
- hadoop-mapreduce-client-hs-plugins-3.3.0.jar
- hadoop-mapreduce-client-jobclient-3.3.0-tests.jar
- hadoop-mapreduce-client-jobclient-3.3.0.jar
- hadoop-mapreduce-client-nativetask-3.3.0.jar
- hadoop-mapreduce-client-shuffle-3.3.0.jar
- hadoop-mapreduce-client-uploader-3.3.0.jar
- hadoop-mapreduce-examples-3.3.0.jar
- jdifff
- lib-examples
- sources

## tools

- dynamometer
- lib
- resourceestimator
- sls
- sources

## yarn

- csi
- hadoop-yarn-api-3.3.0.jar
- hadoop-yarn-applications-catalog-webapp-3.3.0.war
- hadoop-yarn-applications-distributedshell-3.3.0.jar
- hadoop-yarn-applications-mawo-core-3.3.0.jar
- hadoop-yarn-applications-unmanaged-am-launcher-3.3.0.jar
- hadoop-yarn-client-3.3.0.jar
- hadoop-yarn-common-3.3.0.jar
- hadoop-yarn-registry-3.3.0.jar
- hadoop-yarn-server-applicationhistoryservice-3.3.0.jar
- hadoop-yarn-server-common-3.3.0.jar
- hadoop-yarn-server-nodemanager-3.3.0.jar
- hadoop-yarn-server-resourcemanager-3.3.0.jar
- hadoop-yarn-server-router-3.3.0.jar

```

hadoop-yarn-server-sharedcachemanager-3.3.0.jar
hadoop-yarn-server-tests-3.3.0.jar
hadoop-yarn-server-timeline-pluginstorage-3.3.0.jar
hadoop-yarn-server-web-proxy-3.3.0.jar
hadoop-yarn-services-api-3.3.0.jar
hadoop-yarn-services-core-3.3.0.jar
lib
sources
test
timelineservice
webapps
yarn-service-examples

```

आप देख सकते हैं कि बहुत सारे jar पैकेज हैं।

```

$ mkdir input
$ ls
bin hadoop-config.sh hdfs-config.sh libexec sbin yarn-config.sh
etc hadoop-functions.sh input mapred-config.sh share
$ cp etc/hadoop/*.xml input
$ cd input/
$ ls
capacity-scheduler.xml hadoop-policy.xml hdfs-site.xml kms-acls.xml mapred-site.xml
core-site.xml hdfs-rbf-site.xml httpfs-site.xml kms-site.xml yarn-site.xml
$ cd ..
$ bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-3.2.2.jar grep input output 'dfs[a-z.]'
JAR : /usr/local/Cellar/hadoop/3.3.0/libexec/share/hadoop/mapreduce/hadoop-map
$
$ bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.0.jar grep input output 'dfs[a-z.]'
2021-03-11 01:54:30,791 WARN util.NativeCodeLoader: - ... , -
2021-03-11 01:54:31,115 INFO impl.MetricsConfig: hadoop-metrics2.properties
2021-03-11 01:54:31,232 INFO impl.MetricsSystemImpl: 10
...

```

वेबसाइट के उदाहरण के अनुसार टाइप करें। ध्यान दें कि bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-3.2.2.jar grep input में, jar पैकेज से पहले वर्ज़न नंबर है। इसलिए हमें इसे अपने 3.3.0 से बदलना होगा।

लॉग का अंत:

2021-03-11 01:54:35,374 INFO mapreduce.Job: map 100% reduce 100%  
2021-03-11 01:54:35,374 INFO mapreduce.Job: Job job\_local2087514596\_0002  
2021-03-11 01:54:35,377 INFO mapreduce.Job: : 30

FILE: =1204316  
FILE: =3565480  
FILE: =0  
FILE: =0  
FILE: =0

#### Map-Reduce

Map =1  
Map =1  
Map =17  
Map =25  
=141  
=0  
=0  
=1  
=25  
=1  
=1  
=2  
=1  
=0  
=1  
GC (ms)=57  
( )=772800512

BAD\_ID=0  
CONNECTION=0  
IO\_ERROR=0  
WRONG\_LENGTH=0  
WRONG\_MAP=0  
WRONG\_REDUCE=0

=123



डिबग कमांड्स को एक्जीक्यूट करने के लिए डिबग एडमिन चलाएं एडमिन क्लाइंट चलाएं राउटर-आधारित फेडरेशन का प्रबंधन करें डरेज़र कोडिंग चलाएं फ़ाइलसिस्टम चेकिंग यूटिलिटी चलाएं एडमिन क्लाइंट चलाएं या एक्सपोर्टेड वैल्यूज़ प्राप्त करें एडिट्स फ़ाइल पर ऑफ़लाइन एडिट्स व्यूअर लागू करें ऑफ़लाइन व्यूअर लागू करें लेगेसी पर ऑफ़लाइन ब्लॉक स्टोरेज पॉलिसीज़ को सूचीबद्ध/प्राप्त/सेट/संतुष्ट करें

:

और आवश्यक लाइब्रेरीज़ प्राप्त करने के लिए आवश्यक क्लास पथ प्रिंट करता है फ़ाइल सिस्टम पर एक फ़ाइल सिस्टम कमांड चलाता है गणित पर्यावरण चर प्रदर्शित करता है से एक प्रतिनिधि टोकन प्राप्त करता है कॉन्फ़िगरेशन से कॉन्फ़िग मान प्राप्त करता है उन समूहों को प्राप्त करता है जिनसे उपयोगकर्ता संबंधित हैं वर्तमान उपयोगकर्ता द्वारा स्वामित्व वाले सभी स्नैपशॉटबल डायरेक्टरीज़ को सूचीबद्ध करता है एक डायरेक्टरी के दो स्नैपशॉट्स के बीच अंतर या वर्तमान डायरेक्टरी सामग्री और एक स्नैपशॉट के बीच अंतर दिखाता है संस्करण प्रिंट करता है

:

क्लस्टर संतुलन उपयोगिता चलाएं डेटानोड चलाएं राउटर चलाएं-  
 दिए गए नोड पर डिस्क के बीच डेटा समान रूप से वितरित करें सर्वर चलाएं, गेटवे  
 जर्नलनोड चलाएं ब्लॉक प्रतिकृति को स्टोरेज प्रकारों के बीच ले जाने के लिए उपयोगिता चलाएं-  
 नेमनोड चलाएं 3 संस्करण 3 गेटवे चलाएं पोर्टमैप सेवा चलाएं  
 सेकेंडरी नेमनोड चलाएं बाहरी स्टोरेजपॉलिसीसंतुष्टि चलाएं फेलओवर कंट्रोलर डेमन चलाएं  
 , जब बिना पैरामीटर्स के या - के साथ इनवोक किया जाता है, तो मदद प्रिंट कर सकता है।

```
``python
from pyspark.sql import SparkSession

spark = SparkSession.builder.master("local[*]")\
 .config('spark.driver.bindAddress', '127.0.0.1')\
 .getOrCreate()
sc = spark.sparkContext
```

यह कोड एक बनाता है और उसे spark नाम से इनिशियलाइज़ करता है। master("local[\*]") का उपयोग करके यह स्थानीय मशीन पर सभी उपलब्ध कोर का उपयोग करता है। spark.driver.bindAddress को 127.0.0.1 पर सेट

किया गया है, जो ड्राइवर को लोकलहोस्ट पर बाइंड करता है। अंत में, `getOrCreate()` मेथड का उपयोग करके `sc` वेरिएबल को बनाया या पहले से मौजूद सत्र को प्राप्त किया जाता है। `sc` वेरिएबल को संदर्भित करता है, जो एप्लिकेशन का मुख्य प्रवेश बिंदु है।

```
text_file = sc.textFile("a.txt")
counts = text_file.flatMap(lambda line: line.split(" ")) \
 .map(lambda word: (word, 1)) \
 .reduceByKey(lambda a, b: a + b)
counts.saveAsTextFile("b.txt")
```

यह कोड `a.txt` फ़ाइल में एक सरल शब्द गणना (प्रोग्राम है। यह `a.txt` फ़ाइल से पाठ पढ़ता है, प्रत्येक शब्द को गिनता है, और परिणाम को `b.txt` फ़ाइल में सहेजता है।

ध्यान दें कि `.config('spark.driver.bindAddress', '127.0.0.1')` सेट करना महत्वपूर्ण है। अन्यथा, आपको Service 'sparkDriver' could not bind on a random free port. You may check whether configuring an appropriate binding address जैसी त्रुटि मिल सकती है।

हालांकि, इस समय फिर से एक त्रुटि उत्पन्न हुई।

Caused by: org.apache.spark.api.python.PythonException: Traceback (most recent call last):

```
File "/usr/local/lib/python3.9/site-packages/pyspark/python/lib/pyspark.zip/pyspark/worker.py", line 4
 raise Exception(("Python in worker has different version %s than that in " +
Exception: Python 3.8 3.9 , PySpark - PYSF
```

यह दर्शाता है कि विभिन्न संस्करणों के Python चलाए गए हैं।

`.bash_profile` को संशोधित करें:

```
PYSPARK_PYTHON=/usr/local/Cellar/python@3.9/3.9.1_6/bin/python3
PYSPARK_DRIVER_PYTHON=/usr/local/Cellar/python@3.9/3.9.1_6/bin/python3
```

फिर भी वही त्रुटि आ रही है। कुछ जानकारी प्राप्त करने के बाद, यह संभव है कि spark चलाते समय यह पर्यावरण चर लोड नहीं हुआ हो, और टर्मिनल के डिफ़ॉल्ट पर्यावरण चर का उपयोग नहीं किया गया हो।

कोड में निम्नलिखित सेटिंग्स की आवश्यकता है:

```
import os
```

## डिफॉल्ट वातावरण सेट करें

```
__HADOOP_HOME__ = '/usr/hadoop/elasticmapreduce/elasticmapreduce3.9/3.9.1_6/usr/hadoop/elasticmapreduce3'
__HADOOP_CONF__ = '/usr/hadoop/elasticmapreduce/elasticmapreduce3.9/3.9.1_6/usr/hadoop/elasticmapreduce3'
```

```
```shell
```

```
$ python sc.py
```

```
21/03/11 02:54:52 WARN NativeCodeLoader: native-hadoop ... , builtin-
```

```
Spark log4j : org/apache/spark/log4j-defaults.properties
"WARN"
```

```
sc.setLogLevel(newLevel) SparkR , setLogLevel(newLevel)
```

```
PythonRDD[6] at RDD at PythonRDD.scala:53
```

इस समय b.txt फ़ाइल जनरेट हो गई है।

```
b.txt
```

```
_SUCCESS
```

```
part-00000
```

```
part-00001
```

खोलो।

```
$ cat b.txt/part-00000
```

```
('college', 1)
```

```
('two', 1)
```

```
('things', 2)
```

```
('worked', 1)
```

```
('on,', 1)
```

```
('of', 8)
```

```
('school,', 2)
```

```
('writing', 2)
```

```
('programming.', 1)
```

```
("didn't", 4)
```

```
('then,', 1)
```

```
('probably', 1)
```

```

('are:', 1)
('short', 1)
('awful.', 1)
('They', 1)
('plot,', 1)
('just', 1)
('characters', 1)
('them', 2)
...

```

सफलता! क्या यह परिचित नहीं लगता? यह ठीक वैसा ही है जैसा Hadoop उदाहरण में था।

```

$ cat output/*
1   dfsadmin

```

(यहाँ कोड ब्लॉक है, इसलिए इसे अनुवादित नहीं किया गया है।)

ये फ़ाइलें HDFS कहलाती हैं। यहाँ Spark का उपयोग शब्दों की गिनती करने के लिए किया गया है। कुछ ही पंक्तियों में, यह बहुत सुविधाजनक लगता है।

□□□□□□□□□□

अगला कदम Kubernetes को समझना है, जिसे k8s भी कहा जाता है, जहां बीच के 8 अक्षरों को 8 से संक्षिप्त किया गया है। यह एक ओपन-सोर्स सिस्टम है जो कंटेनर एप्लिकेशन्स को स्वचालित रूप से डिप्लॉय, स्केल और मैनेज करने के लिए उपयोग किया जाता है।

kubect1 कमांड लाइन टूल का उपयोग □8□ क्लस्टर पर कुछ कमांड चलाने के लिए किया जाता है। इसका उपयोग एप्लिकेशन को डिप्लॉय करने, क्लस्टर संसाधनों को देखने और प्रबंधित करने, और लॉग्स देखने के लिए किया जा सकता है।

□□□□□□□□ का उपयोग करके भी इंस्टॉल किया जा सकता है।

```

brew install kubect1

```

(नोट: यह एक कमांड है जिसे टर्मिनल में चलाया जाता है। इसे हिंदी में ट्रांसलेट करने की आवश्यकता नहीं है।)

लॉग आउटपुट:

```

==>      https://homebrew.bintray.com/bottles/kubernetes-cli-1.20.1.big_sur.bottle.tar.gz
==>      https://d29vzk4ow07wi7.cloudfront.net/0b4f08bd1d47cb913d7cd4571e3394c6747dfbad7ff114c558
##### 100.0%
==>      kubernetes-cli-1.20.1.big_sur.bottle.tar.gz

```


==>

Bash :

/usr/local/etc/bash_completion.d

==>

/usr/local/Cellar/kubernetes-cli/1.20.1: 246 , 46.1MB

इंस्टॉल हो गया है।

```
$ kubectl version --client
```

```
Client Version: version.Info{Major:"1", Minor:"20", GitVersion:"v1.20.1", GitCommit:"c4d752765b3bbac223"
```

```
$ kubectl
```

```
kubectl Kubernetes
```

अधिक जानकारी के लिए यहां जाएं: <https://kubernetes.io/docs/concepts/overview/kubernetes-concepts/>

बेसिक कमांड्स (शुरुआती): `kubectl` फ़ाइल या संसाधन से एक संसाधन बनाएं। `kubectl` एक रिप्लिकेशन कंट्रोलर, सर्विस, डिप्लॉयमेंट या पॉड लें और इसे एक नई `kubernetes` सर्विस के रूप में एक्सपोज़ करें। `kubectl` क्लस्टर पर एक विशिष्ट इमेज चलाएं। `kubectl` ऑब्जेक्ट्स पर विशिष्ट फ़ीचर्स सेट करें।

बेसिक कमांड्स (इंटरमीडिएट): `kubectl` संसाधनों का दस्तावेज़ीकरण `kubectl` एक या अनेक संसाधनों को प्रदर्शित करें `kubectl` सर्वर पर एक संसाधन को संपादित करें `kubectl` फ़ाइलनाम, `kubectl`, संसाधन और नाम, या संसाधन और लेबल चयनकर्ता द्वारा संसाधनों को हटाएं

`kubectl` `kubernetes`: `kubectl` किसी संसाधन के रोलआउट का प्रबंधन करें `kubectl` `kubernetes`, `kubernetes` या `kubernetes` `kubernetes` के लिए एक नया आकार सेट करें `kubernetes` `kubernetes`, `kubernetes`, या `kubernetes`-`kubernetes` को स्वचालित रूप से स्केल करें

क्लस्टर प्रबंधन कमांड्स: `kubernetes` सर्टिफिकेट संसाधनों को संशोधित करें। `kubernetes`-`kubernetes` क्लस्टर की जानकारी प्रदर्शित करें। `kubernetes` संसाधन (`kubernetes`/मेमोरी/स्टोरेज) उपयोग प्रदर्शित करें। `kubernetes` नोड को अनुसूचित नहीं करने योग्य के रूप में चिह्नित करें। `kubernetes` नोड को अनुसूचित करने योग्य के रूप में चिह्नित करें। `kubernetes` रखरखाव की तैयारी में नोड को ड्रेन करें। `kubernetes` एक या अधिक नोड्स पर टेंट्स को अपडेट करें।

समस्या निवारण और डिबगिंग कमांड: `kubernetes` किसी विशिष्ट संसाधन या संसाधनों के समूह का विवरण दिखाएं `kubernetes` पॉड में कंटेनर के लॉग प्रिंट करें `kubernetes` चल रहे कंटेनर से जुड़े `kubernetes` कंटेनर में एक कमांड निष्पादित करें `kubernetes`-`kubernetes` एक या अधिक स्थानीय पोर्ट्स को पॉड में फॉरवर्ड करें `kubernetes` `kubernetes` `kubernetes` सर्वर के लिए एक प्रॉक्सी चलाएं `kubernetes` कंटेनर से और कंटेनर में फ़ाइलें और डायरेक्टरीज़ कॉपी करें `kubernetes` प्राधिकरण का निरीक्षण करें `kubernetes` वर्कलोड और नोड्स के समस्या निवारण के लिए डिबगिंग सत्र बनाएं

उन्नत कमांड्स: `kubernetes` लाइव संस्करण और लागू किए जाने वाले संस्करण के बीच अंतर दिखाएं `kubernetes` फ़ाइलनाम या `kubernetes` के माध्यम से कॉन्फ़िगरेशन को संसाधन पर लागू करें `kubernetes` संसाधन के फ़िल्ड को अपडेट करें `kubernetes` फ़ाइलनाम या `kubernetes` के माध्यम से

संसाधन को बदलें `helm` प्रायोगिक: एक या कई संसाधनों पर विशिष्ट स्थिति की प्रतीक्षा करें। `helm` एक निर्देशिका या दूरस्थ `helm` से एक `helm` लक्ष्य बनाएं।

सेटिंग्स कमांड्स: `helm` किसी संसाधन पर लेबल अपडेट करें `helm` किसी संसाधन पर एनोटेशन अपडेट करें `helm` निर्दिष्ट शेल (`helm` या `helm`) के लिए शेल पूर्णता कोड आउटपुट करें

अन्य कमांड्स: `helm`-`helm` सर्वर पर समर्थित `helm` संसाधनों को प्रिंट करें `helm`-`helm` सर्वर पर समर्थित `helm` संस्करणों को “`helm`/`helm`” के रूप में प्रिंट करें `helm` `helm` फ़ाइलों को संशोधित करें `helm` प्लगइन्स के साथ इंटरैक्ट करने के लिए उपयोगिताएँ प्रदान करें `helm` क्लाइंट और सर्वर संस्करण जानकारी प्रिंट करें

उपयोग: `helm` [फ्लैग्स] [विकल्प]

किसी दिए गए कमांड के बारे में अधिक जानकारी के लिए “`helm` -`helm`” का उपयोग करें। सभी कमांड्स पर लागू होने वाले वैश्विक कमांड-लाइन विकल्पों की सूची के लिए “`helm` `helm`” का उपयोग करें।

```
``yaml
apiVersion: apps/v1
kind: Deployment
metadata:
  name: nginx-deployment
spec:
  selector:
    matchLabels:
      app: nginx
  minReadySeconds: 5
  template:
    metadata:
      labels:
        app: nginx
    spec:
      containers:
        - name: nginx
          image: nginx:1.14.2
          ports:
            - containerPort: 80
```

यह `nginx` फ़ाइल एक `nginx` `nginx` को परिभाषित करती है जो `nginx` कंटेनर को डिप्लॉय करती है। यह `helm`-`helm` `nginx-deployment` नाम से जाना जाएगा और यह `nginx:1.14.2` इमेज का उपयोग करेगा। कंटेनर 80 पोर्ट पर

लिसन करेगा। minReadySeconds सेटिंग यह सुनिश्चित करती है कि कंटेनर कम से कम 5 सेकंड तक तैयार रहेगा।

```
```shell
```

```
$ kubectl apply -f simple_deployment.yaml
```

```
localhost:8080
```

```
-
```

```
?
```

```
$ kubectl cluster-info
```

क्लस्टर समस्याओं को और डीबग और डायग्नोस करने के लिए, 'minikube status' का उपयोग करें। सर्वर localhost:8080 से कनेक्शन अस्वीकार कर दिया गया - क्या आपने सही होस्ट या पोर्ट निर्दिष्ट किया है?

```
```shell
```

```
$ start.sh
```

```
Kubernetes ...minikube : v1.8.1  
: cbda04cf6bbe65e987ae52bb393c10099ab62014
```

```
* minikube v1.8.1 Ubuntu 18.04
```

```
* none
```

```
* localhost (CPUs=2, Memory=2460MB, Disk=145651MB) ...
```

```
* OS Ubuntu 18.04.4 LTS
```

□ minikube v1.17.3 को minikube v1.9.3 पर तैयार कर रहे हैं ...

- minikube.kubernetes.io/minikube=cbda04cf6bbe65e987ae52bb393c10099ab62014

□ minikube क्लस्टर लॉन्च कर रहे हैं ...

□ ऐडऑन्स सक्षम कर रहे हैं: minikube-addons, minikube-addons

□ लोकल होस्ट वातावरण को कॉन्फिगर कर रहे हैं ...

□ हो गया! minikube अब "minikube" का उपयोग करने के लिए कॉन्फिगर है

□ 'minikube' ऐडऑन सक्षम है minikube शुरू हो गया है

```
```shell
```

```
$ kubectl version --client
```

```
Client Version: version.Info{Major:"1", Minor:"20", GitVersion:"v1.20.1", GitCommit:"c4d752765b3bbac223"
```

```
$ kubectl version
```

```
Client Version: version.Info{Major:"1", Minor:"20", GitVersion:"v1.20.1", GitCommit:"c4d752765b3bbac223"
```

```
localhost:8080
```

```
-
```

```
?
```

दस्तावेज़ के अनुसार, Minikube को पहले इंस्टॉल करना आवश्यक है।

```
$ minikube start
minikube v1.16.0 11.2.2
minikube 1.18.1 ! : https://github.com/kubernetes/minikube/releases/tag/v1.18.1
 , : 'minikube config set WantUpdateNotification false'
```

□ स्वचालित रूप से □□□□□□□□□□ ड्राइवर का चयन किया गया □ □□ बूट इमेज डाउनलोड हो रहा है ... > □□□□□□□□-□1.16.0.□□□.□□□256: 65 □ / 65 □ [— — — —] 100.00% ? □/□ 0□ > □□□□□□□□-□1.16.0.□□□: 212.62 □□□ / 212.62 □□□ [ ] 100.00% 5.32 □□□ □/□ 40□ □ कंट्रोल प्लेन नोड □□□□□□□□ को क्लस्टर □□□□□□□□ में शुरू किया जा रहा है □ □□□□□□□□□ □1.20.0 प्रीलोड डाउनलोड हो रहा है ... > □□□□□□□□□-□□□□□□□-□8□-□8-□1....: 491.00 □□□ / 491.00 □□□ 100.00% 7.52 □□□ □ □□□□□□□□□□ □□ बनाया जा रहा है (□□□□□=2, □□□□□□□=4000□□, □□□□□=20000□□) ... □ इस □□ को □□□□□□://□8□.□□□.□□ तक पहुंचने में समस्या हो रही है □ नई बाहरी इमेज को पुल करने के लिए, आपको एक प्रॉक्सी कॉन्फ़िगर करने की आवश्यकता हो सकती है: □□□□□□://□□□□□□□□□□.□□□□.□8□.□□/□□□□□/□□□□□□□□□□□□/□□□□□□□□□□□□□□□□/□□□□□□□□□□ □ □□□□□□□□□□ □1.20.0 को □□□□□□□□ 20.10.0 पर तैयार किया जा रहा है ... □ सर्टिफिकेट और कुंजियाँ जनरेट की जा रही हैं ... □ कंट्रोल प्लेन को बूट किया जा रहा है ... □ □□□□ नियम कॉन्फ़िगर किए जा रहे हैं ... □ □□□□□□□□□□ कंपोनेंट्स की जाँच की जा रही है... □ एडऑन्स सक्षम किए गए: □□□□□□□□□□-□□□□□□□□□□□□□□□□□, □□□□□□□□□□-□□□□□□□□□□□□□□□□□ □ हो गया! □□□□□□□□ अब डिफ़ॉल्ट रूप से “□□□□□□□□□□” क्लस्टर और “□□□□□□□□□□” नेमस्पेस का उपयोग करने के लिए कॉन्फ़िगर हो गया है

20

kube-system	etcd-minikube	0/1	0	74s
kube-system	kube-apiserver-minikube	1/1	0	74s
kube-system	kube-controller-manager-minikube	1/1	0	74s
kube-system	kube-proxy-g2296	1/1	0	60s
kube-system	kube-scheduler-minikube	0/1	0	74s
kube-system	storage-provisioner	1/1	1	74s

minikube का डैशबोर्ड खोलने के लिए।

```
$ minikube dashboard
```

...

...

...

...

<http://127.0.0.1:50030/api/v1/namespaces/kubernetes-dashboard/services/http:kubernetes-d>

The screenshot shows the Kubernetes Dashboard interface. The top navigation bar includes the Kubernetes logo, a dropdown menu set to 'default', a search bar, and icons for adding new resources and notifications. The left sidebar lists various Kubernetes resources: Overview (selected), Cron Jobs, Daemon Sets, Deployments, Jobs, Pods, Replica Sets, Replication Controllers, Stateful Sets, Services, Ingresses, Config Maps, and Persistent Volume Claims. The main content area is divided into two sections. The top section, titled 'Service', contains a table of Services. The table has columns for Name, Namespace, Labels, Cluster IP, Internal Endpoints, External Endpoints, and Creation Time. The 'kubernetes' service is listed with a green checkmark, namespace 'default', labels 'component: apiserver' and 'provider: kubernetes', cluster IP '10.96.0.1', and internal endpoints 'kubernetes:443 TCP' and 'kubernetes:0 TCP'. The bottom section, titled '配置和存储' (Configuration and Storage), contains a table of Config Maps. The table has columns for Name, Namespace, Labels, and Creation Time. The 'kube-root-ca.crt' Config Map is listed with namespace 'default' and creation time '2 minutes ago'.

000000 1: 080

कैसे बंद करें।

```
$ minikube
```

```
minikube Kubernetes
```

बेसिक कमांड्स: `minikube` एक स्थानीय `Kubernetes` क्लस्टर शुरू करता है `minikube` एक स्थानीय `Kubernetes` क्लस्टर की स्थिति प्राप्त करता है `minikube status` चल रहे स्थानीय `Kubernetes` क्लस्टर को रोकता है `minikube delete` एक स्थानीय `Kubernetes` क्लस्टर को हटाता है `minikube dashboard` `Kubernetes` क्लस्टर के अंदर चल रहे `Kubernetes` डैशबोर्ड तक पहुंच प्रदान करता है `minikube ssh` `minikube` को पॉज़ करता है `minikube` `minikube` को अनपॉज़ करता है

`minikube` `minikube`: `minikube`-`minikube` के `minikube` डेमन का उपयोग करने के लिए वातावरण को कॉन्फ़िगर करें `minikube`-`minikube` के `minikube` सेवा का उपयोग करने के लिए वातावरण को कॉन्फ़िगर करें `minikube` `minikube` में एक स्थानीय छवि को जोड़ें, हटाएं, या पुश करें

कॉन्फ़िगरेशन और प्रबंधन कमांड: `minikube` `minikube` ऐडऑन को सक्षम या अक्षम करें `minikube` स्थायी कॉन्फ़िगरेशन मानों को संशोधित करें `minikube` वर्तमान प्रोफाइल (क्लस्टर) प्राप्त करें या सूचीबद्ध करें `minikube`-`minikube` `minikube` या पोर्ट परिवर्तन के मामले में `minikube`-`minikube` को अपडेट करें

नेटवर्किंग और कनेक्टिविटी कमांड्स: `minikube` किसी सेवा से कनेक्ट करने के लिए एक `minikube` रिटर्न करता है `minikube` `minikube`-`minikube` सेवाओं से कनेक्ट करता है

उन्नत कमांड्स: `minikube` निर्दिष्ट डायरेक्टरी को `minikube` में माउंट करता है `minikube` `minikube` वातावरण में लॉग इन करें (डिबगिंग के लिए) `minikube` क्लस्टर संस्करण से मेल खाने वाले `minikube` बाइनरी को चलाएं `minikube` अतिरिक्त नोड्स को जोड़ें, हटाएं, या सूचीबद्ध करें

समस्या निवारण कमांड्स: `minikube`-`minikube` निर्दिष्ट नोड की `minikube` पहचान कुंजी पथ प्राप्त करें `minikube`-`minikube` निर्दिष्ट नोड की `minikube` होस्ट कुंजी प्राप्त करें `minikube` निर्दिष्ट नोड का `minikube` पता प्राप्त करें `minikube` स्थानीय `minikube` क्लस्टर को डीबग करने के लिए लॉग्स लौटाएं `minikube`-`minikube` वर्तमान और नवीनतम संस्करण संख्या प्रिंट करें `minikube` `minikube` का संस्करण प्रिंट करें

अन्य कमांड: `minikube` किसी शेल के लिए कमांड पूर्णता उत्पन्न करें

किसी दिए गए कमांड के बारे में अधिक जानकारी के लिए “`minikube` -`minikube`” का उपयोग करें।

```
`minikube stop`
```

```
`kubernetes`
```

```
```shell
```

```
$ kubectl cluster-info
```

```
Kubernetes https://192.168.99.100:8443
```

```
KubeDNS https://192.168.99.100:8443/api/v1/namespaces/kube-system/services/kube-dns:dns/proxy
```

क्लस्टर समस्याओं को और डीबग और निदान करने के लिए, ‘`minikube` `minikube`-`minikube`’ का उपयोग करें।

```
`https://192.168.99.100:8443` , :
```

```
```json
```

```
{
 "kind": "Status",
 "apiVersion": "v1",
 "metadata": {

 },
 "status": "Failure",
 "message": "Forbidden: \"/system:anonymous\" \"/\" ",
 "reason": "Forbidden",
 "details": {

 },
 "code": 403
}
```

`https://192.168.99.100:8443/api/v1/namespaces/kube-system/services/kube-dns:dns/proxy` पर जाएं:

```
{
 "kind": "Status",
 "apiVersion": "v1",
 "metadata": {

 },
 "status": "Failure",
 "message": "Forbidden: \"/kube-dns:dns\" \"/system:anonymous\" API \"/\" \"/kube-system\" ",
 "reason": "Forbidden",
 "details": {
 "name": "kube-dns:dns",
 "kind": "services"
 },
 "code": 403
}
```

आइए अभी उस कॉन्फिगरेशन को आज़माएं।

```
$ kubectl apply -f simple_deployment.yaml
deployment.apps/nginx-deployment
```

थोड़ी समस्या है। हालांकि, यहां तक हमने kubernetes को चला लिया है। पहले इसे समाप्त करते हैं। बाद में फिर खेलेंगे।

```
$ minikube stop
```

```
"minikube" ...
1
```

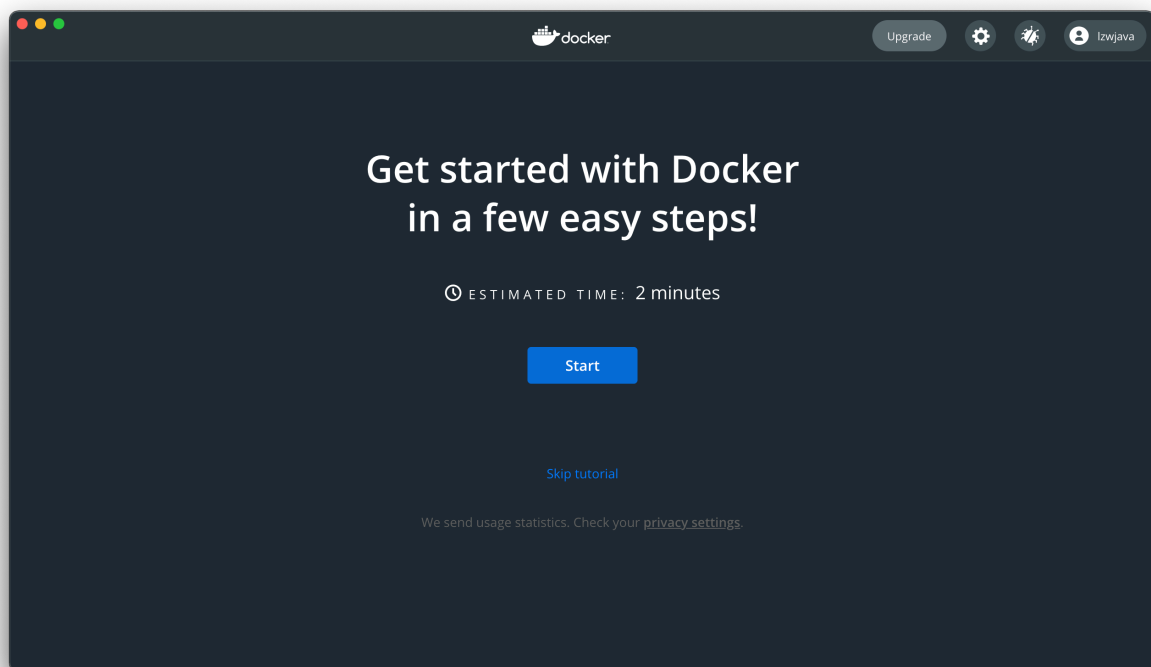
समाप्ति की जाँच करें।

```
w$ minikube dashboard
```

```
, : "minikube start"
```

□□□□□□

Docker एक कंटेनर प्लेटफॉर्म भी है, जो आधुनिक एप्लिकेशन बनाने, साझा करने और चलाने में तेजी लाने में मदद करता है। आधिकारिक वेबसाइट से एप्लिकेशन डाउनलोड करें।



□□□□□□ 2: □□□□□□

क्लाइंट का उपयोग करते समय थोड़ा लैग हो रहा है। चलो कमांड लाइन का उपयोग करते हैं।

```
$ docker
```



उपयोग: `docker [विकल्प] कमांड`

कंटेनरों के लिए एक स्वयं-पर्याप्त रनटाइम

विकल्प: `-v` `वॉल्यूम` क्लाइंट कॉन्फिग फ़ाइलों का स्थान (डिफ़ॉल्ट `"/var/lib/docker/volumes/"`) `-o`, `--opt` `वॉल्यूम ड्राइवर` से कनेक्ट करने के लिए उपयोग करने वाले कॉन्टेक्स्ट का नाम (`docker-engine` और `docker` के साथ सेट डिफ़ॉल्ट कॉन्टेक्स्ट को ओवरराइड करता है) `-l`, `--label` डिबग मोड सक्षम करें `-p`, `--port` कनेक्ट करने के लिए डेमन सॉकेट `(/var/run/docker.sock)` `-H`, `--host` `लॉगिंग स्तर` सेट करें (`"json"` | `"journald"` | `"syslog"` | `"gelf"` | `"fluentd"`) (डिफ़ॉल्ट `"json"`) `-u` `यूजर` का उपयोग करें; `--rm` द्वारा निहित `--no-rm` केवल इस `docker` द्वारा हस्ताक्षरित प्रमाणपत्रों पर भरोसा करें (डिफ़ॉल्ट `"/var/lib/docker/certs/"`) `--prune` `प्रमाणपत्र फ़ाइल का पथ` (डिफ़ॉल्ट `"/var/lib/docker/certs/"`) `--storage-driver` `कुंजी फ़ाइल का पथ` (डिफ़ॉल्ट `"/var/lib/docker/storage/"`) `--storage-opt` का उपयोग करें और रिमोट को सत्यापित करें `-s`, `--seccomp-profile` संस्करण जानकारी प्रिंट करें और बाहर निकलें

प्रबंधन कमांड्स: `docker * docker` (`docker` `0.9.1-0.10.3`) `docker` बिल्ड्स प्रबंधित करें `docker * docker` `के साथ बिल्ड करें` (`docker` `0.5.1-0.10.3`) `docker` कॉन्फिगरेशन प्रबंधित करें `docker` कंटेनर प्रबंधित करें `docker` कॉन्टेक्स्ट प्रबंधित करें `docker` इमेज प्रबंधित करें `docker` इमेज मैनिफेस्ट और मैनिफेस्ट सूचियाँ प्रबंधित करें `docker` नेटवर्क प्रबंधित करें `docker` नोड्स प्रबंधित करें `docker` प्लगइन्स प्रबंधित करें `docker * docker` `के साथ बिल्ड करें` (`docker` `0.5.0`) `docker` सीक्रेट्स प्रबंधित करें `docker` सर्विसेज प्रबंधित करें `docker` स्टैक्स प्रबंधित करें `docker` प्रबंधित करें `docker` प्रबंधित करें `docker` इमेज पर ट्रस्ट प्रबंधित करें `docker` वॉल्यूम प्रबंधित करें

कमांड्स: `docker` एक चल रहे कंटेनर से स्थानीय मानक इनपुट, आउटपुट और त्रुटि स्ट्रीम को जोड़ें `docker` से एक इमेज बनाएं `docker` कंटेनर में हुए परिवर्तनों से एक नई इमेज बनाएं `docker` कंटेनर और स्थानीय फाइल सिस्टम के बीच फाइल/फोल्डर कॉपी करें `docker` एक नया कंटेनर बनाएं `docker` कंटेनर के फाइल सिस्टम पर फाइल या डायरेक्टरी में हुए परिवर्तनों का निरीक्षण करें `docker` सर्वर से रियल टाइम इवेंट्स प्राप्त करें `docker` एक चल रहे कंटेनर में कमांड चलाएं `docker` कंटेनर के फाइल सिस्टम को `docker` आर्काइव के रूप में निर्यात करें `docker` एक इमेज का इतिहास दिखाएं `docker` इमेज की सूची दिखाएं `docker` एक `docker` से सामग्री आयात करके फाइल सिस्टम इमेज बनाएं `docker` सिस्टम-व्यापी जानकारी प्रदर्शित करें `docker` ऑब्जेक्ट्स पर निम्न-स्तरीय जानकारी दिखाएं `docker` एक या अधिक चल रहे कंटेनर को समाप्त करें `docker` `आर्काइव` या `docker` से एक इमेज लोड करें `docker` `रजिस्ट्री` में लॉग इन करें `docker` `रजिस्ट्री` से लॉग आउट करें `docker` कंटेनर के लॉग प्राप्त करें `docker` एक या अधिक कंटेनर के सभी प्रक्रियाओं को रोकें `docker` कंटेनर के पोर्ट मैपिंग या एक विशिष्ट मैपिंग की सूची दिखाएं `docker` कंटेनर की सूची दिखाएं `docker` `रजिस्ट्री` से एक इमेज या रिपॉजिटरी डाउनलोड करें `docker` `रजिस्ट्री` में एक इमेज या रिपॉजिटरी अपलोड करें `docker` कंटेनर का नाम बदलें `docker` एक या अधिक कंटेनर को पुनरांश करें `docker` एक या अधिक कंटेनर हटाएं `docker` एक या अधिक इमेज हटाएं `docker` एक नए कंटेनर में कमांड चलाएं `docker` एक या अधिक इमेज को `docker` आर्काइव के रूप में सहेजें (डिफ़ॉल्ट रूप से `docker` पर स्ट्रीम किया जाता है) `docker` `docker` पर इमेज खोजें `docker` एक या अधिक रुके हुए कंटेनर शुरू करें `docker` कंटेनर(र) के संसाधन उपयोग सांख्यिकी का लाइव स्ट्रीम प्रदर्शित करें `docker` एक या अधिक चल रहे कंटेनर को रोकें `docker` एक टैग `docker` बनाएं जो `docker` को संदर्भित करता है `docker` कंटेनर की चल रही प्रक्रियाओं को प्रदर्शित करें `docker` एक या अधिक कंटेनर के सभी प्रक्रियाओं को अनपॉज़ करें `docker` एक या अधिक कंटेनर की कॉन्फिगरेशन अपडेट करें `docker` `docker` संस्करण जानकारी दिखाएं `docker` एक या अधिक कंटेनर के रुकने तक ब्लॉक करें, फिर उनके एग्जिट कोड प्रिंट करें

करें

किसी कमांड के बारे में अधिक जानकारी के लिए 'docker gettings-started -troubleshooting' चलाएं।

docker gettings-started के बारे में अधिक मदद पाने के लिए, हमारे गाइड्स देखें: <https://docs.docker.com/getting-started/troubleshooting/>

```
```shell
$ docker run -d -p 80:80 docker/getting-started
'docker/getting-started:latest'
latest: docker/getting-started
aad63a933944:
b14da7a62044:
343784d40d66:
6f617e610986:
Digest: sha256:d2c4fb0641519ea208f20ab03dc40ec2a5a53fdfbcca90bef14f870158ed577
Status: docker/getting-started
815f13fc8f99f6185257980f74c349e182842ca572fe60ff62cbb15641199eaf
docker:                                :                                : listen tcp 0.0.0.0:80: bind:
```

पोर्ट बदलें।

```
$ docker run -d -p 8080:80 docker/getting-started
45bb95fa1ae80adc05cc498a1f4f339c45c51f7a8ae1be17f5b704853a5513a5
```

ब्राउज़र खोलें, यह दर्शाता है कि हमने docker को सफलतापूर्वक चला दिया है।

कंटेनर को रोकें। पहले प्राप्त किए गए ID का उपयोग करें।

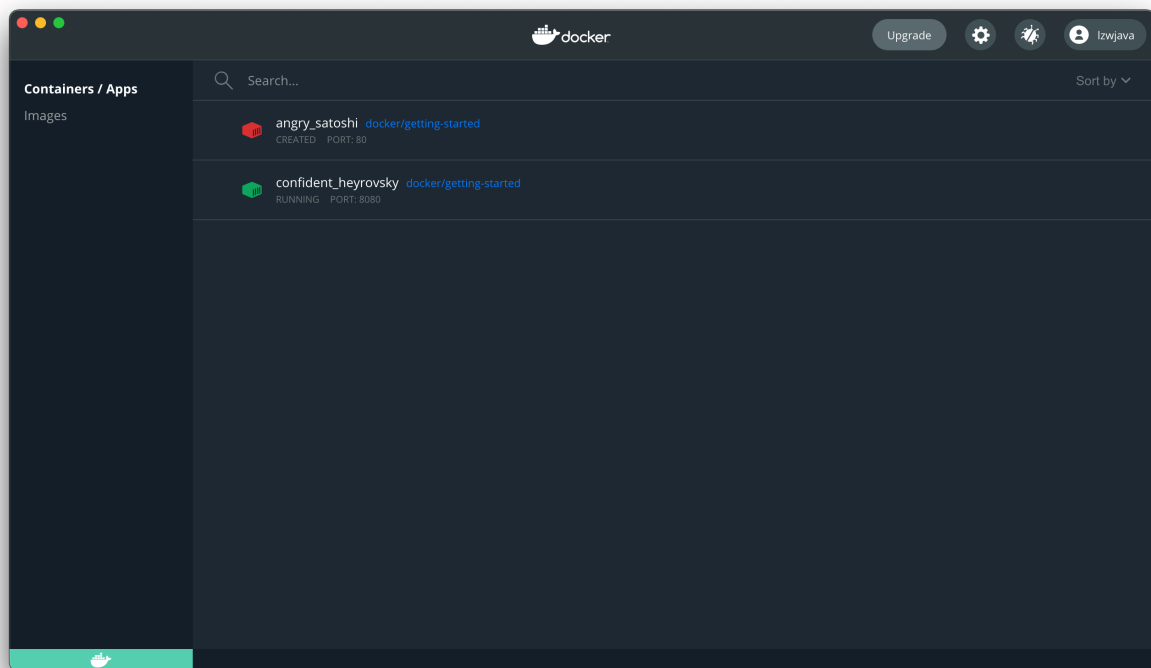
```
$ docker stop 45bb95fa1ae80adc05cc498a1f4f339c45c51f7a8ae1be17f5b704853a5513a5
45bb95fa1ae80adc05cc498a1f4f339c45c51f7a8ae1be17f5b704853a5513a5
```

इस समय वेबसाइट खोलना संभव नहीं था।

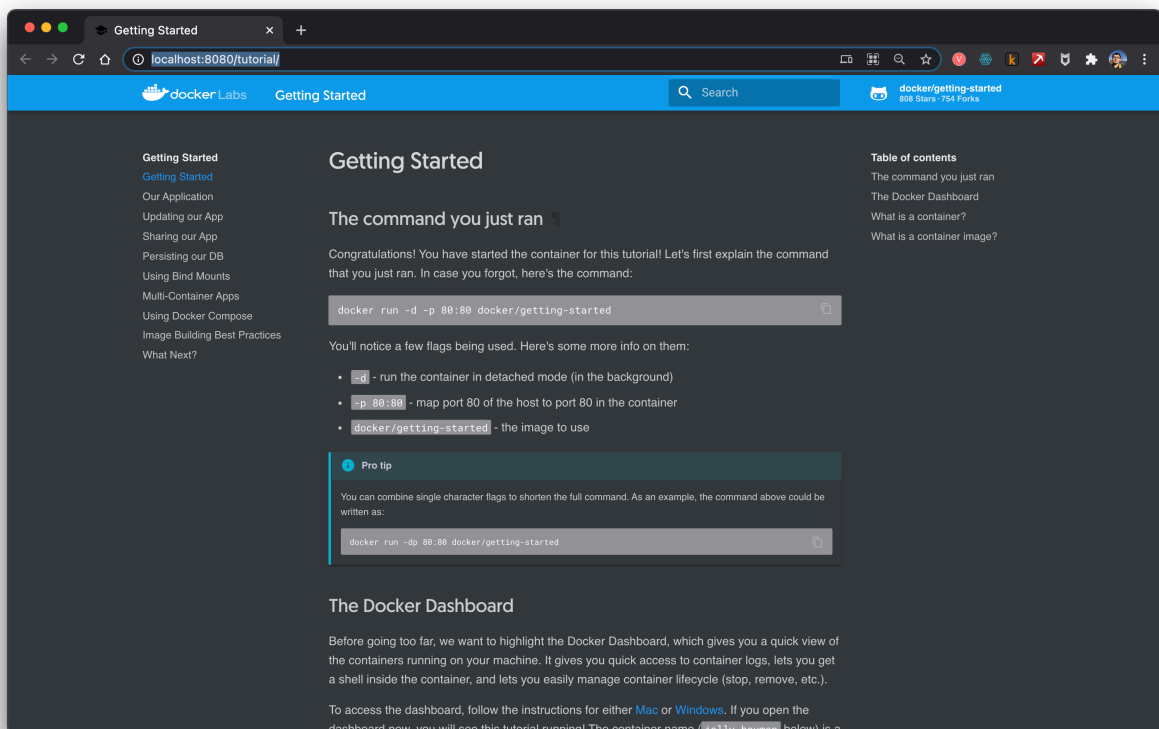
इससे पता चलता है कि docker एक वर्चुअल मशीन की तरह है।

docker

वेबसाइट खोलें।



000000 3: 000000_000



000000 4: ब्राउज़र

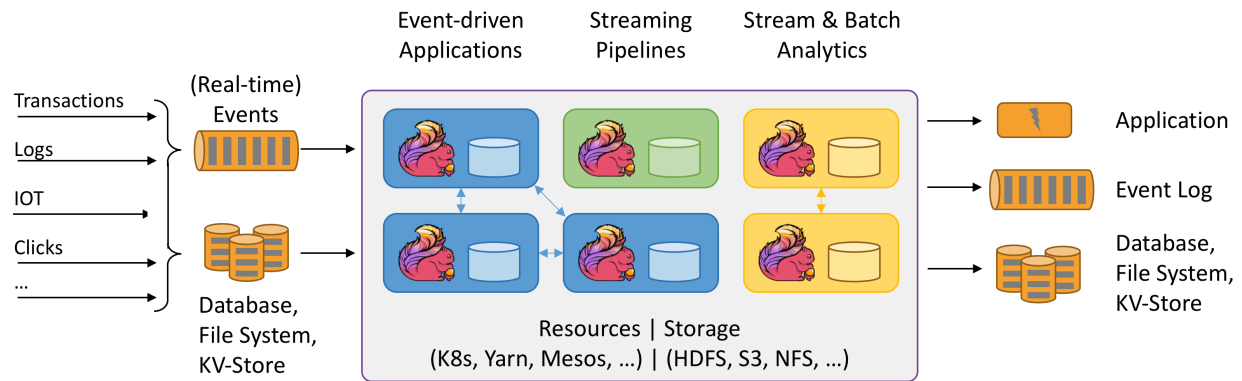


Figure 5: Real-time data processing architecture

Flink को डेटा स्ट्रीम के Stateful कंप्यूटेशन के रूप में जाना जाता है। Stateful का क्या मतलब है? अभी तक समझ में नहीं आया। ऊपर दिया गया चित्र बहुत दिलचस्प है। आइए इसे आज़माएं।

यह कहा जाता है कि वातावरण की आवश्यकता है।

```
$ java -version
java version "1.8.0_151"
Java(TM) SE Runtime Environment (build 1.8.0_151-b12)
Java HotSpot(TM) 64-Bit Server VM (build 25.151-b12, mixed mode)
```

आधिकारिक वेबसाइट से नवीनतम संस्करण flink-1.12.2-bin-scala_2.11.tar डाउनलोड करें।

```
$ ./bin/start-cluster.sh

lzwjava
lzwjava

$ ./bin/flink run examples/streaming/WordCount.jar
WordCount
--input
stdout --output
JobID 60f37647c20c2a6654359bd34edab807

JobID 60f37647c20c2a6654359bd34edab807
: 757

$ tail log/flink-*-taskexecutor-*.out
```

```
(nymph,1)
(in,3)
(thy,1)
(orisons,1)
(be,4)
(all,2)
(my,1)
(sins,1)
(remember,1)
(d,4)
```

```
$ ./bin/stop-cluster.sh
(pid: 41812)      lzwjava
```

हाँ, सफलतापूर्वक शुरुआत की। यह देखा जा सकता है कि यह Spark के समान है।

काइलिन

आधिकारिक वेबसाइट खोलने के लिए।

काइलिन काइलिन एक ओपन सोर्स, वितरित विश्लेषणात्मक डेटा वेयरहाउस है जो बिग डेटा के लिए डिज़ाइन किया गया है; यह बिग डेटा युग में काइलिन (ऑनलाइन विश्लेषणात्मक प्रसंस्करण) क्षमता प्रदान करने के लिए बनाया गया था। काइलिन और काइलिन पर मल्टी-डायमेंशनल क्यूब और प्रीकैलकुलेशन तकनीक को नवीनीकृत करके, काइलिन लगातार बढ़ते डेटा वॉल्यूम के बावजूद लगभग स्थिर क्वेरी गति प्राप्त करने में सक्षम है। क्वेरी विलंबता को मिनटों से सब-सेकंड तक कम करके, काइलिन ऑनलाइन एनालिटिक्स को बिग डेटा में वापस लाता है।

काइलिन काइलिन आपको 3 चरणों में सब-सेकंड लेटेंसी पर अरबों पंक्तियों को क्वेरी करने की सुविधा देता है।

1. काइलिन पर एक काइलिन/काइलिन काइलिन की पहचान करें।
2. पहचाने गए टेबल्स से काइलिन बनाएं।
3. काइलिन-काइलिन का उपयोग करके क्वेरी करें और काइलिन, काइलिन या काइलिन काइलिन के माध्यम से सब-सेकंड में परिणाम प्राप्त करें।

यह मूल रूप से बड़े डेटा के विश्लेषण की एक परत है। इसका उपयोग करके आप बहुत तेज़ी से खोज सकते हैं। यह एक पुल की तरह काम करता है।

दुर्भाग्य से, यह वर्तमान में केवल Linux वातावरण में उपयोग किया जा सकता है। बाद में इसे फिर से संभालूंगा।

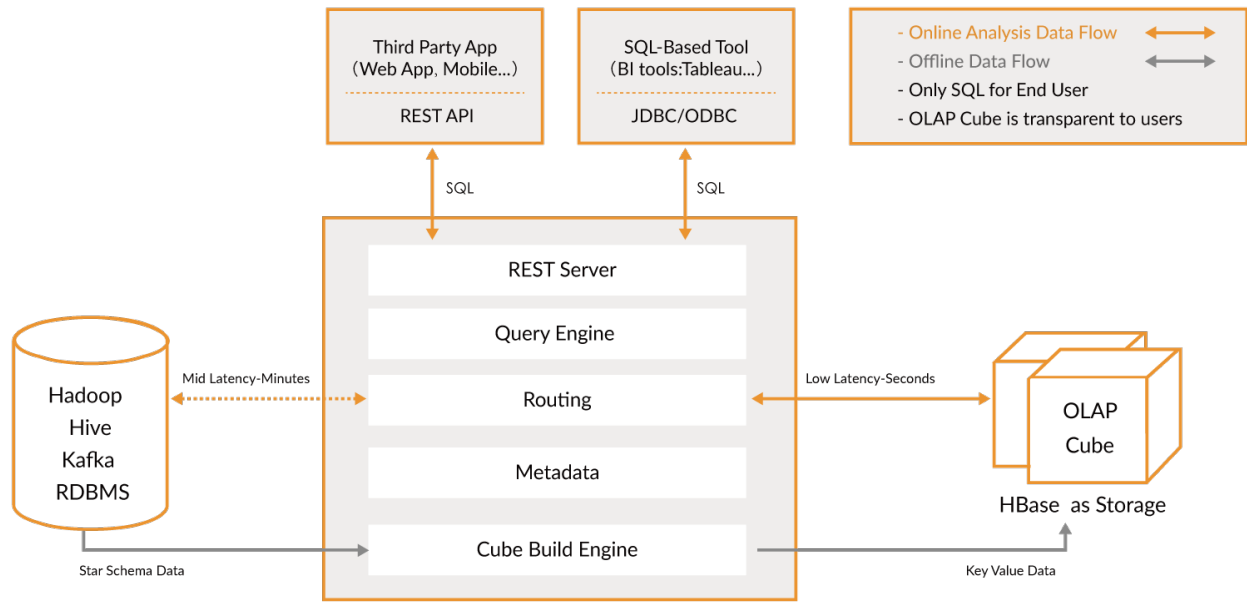


Figure 6: Data Cube Architecture

Figure 6

यह भी एक प्रकार का डेटाबेस है। इसे इंस्टॉल करके देखें।

```
$ brew tap mongodb/brew
==> mongodb/brew
'/usr/local/Homebrew/Library/Taps/mongodb/homebrew-brew'
remote:      : 63,
remote:      : 100% (63/63),
remote:      : 100% (62/62),
remote: 566 ( 21), : 6 ( 1), - : 503
      : 100% (566/566), 121.78 KiB | 335.00 KiB/s,
      : 100% (259/259),
11      (39 , 196.2KB)

$ brew install mongodb-community@4.4
==> mongodb/brew mongodb-community
==> https://fastdl.mongodb.org/tools/db/mongodb-database-tools-macos-x86_64-100.3.0.zip
##### 100.0%
==> https://fastdl.mongodb.org/osx/mongodb-macos-x86_64-4.4.3.tgz
##### 100.0%
==> mongodb/brew/mongodb-community : mongodb-database-tools
```

```

==> mongodb/brew/mongodb-community      : mongodb-database-tools
      : `brew link`
      ,      /usr/local
bin/bsondump
      /usr/local/bin/bsondump
mongodb
      :
      brew unlink mongodb

```

लिंक को फोर्स करने और सभी संघर्ष करने वाली फ़ाइलों को ओवरराइट करने के लिए: `rm -rf /usr/local/bin/bsondump`

सभी फ़ाइलों को सूचीबद्ध करने के लिए जिन्हें हटाया जाएगा: `find /usr/local/bin -type f -exec rm -f {} \;`

संभावित टकराव वाली फ़ाइलें हैं: `/usr/local/bin/bsondump -> /usr/local/bin/mongodb-3.0.7/bin/bsondump`
`/usr/local/bin/mongodb-3.0.7/bin/bsondump -> /usr/local/bin/mongodb-3.0.7/bin/bsondump`
`-> /usr/local/bin/mongodb-3.0.7/bin/bsondump /usr/local/bin/mongodb-3.0.7/bin/bsondump ->`
`/usr/local/bin/mongodb-3.0.7/bin/bsondump /usr/local/bin/mongodb-3.0.7/bin/bsondump -> /usr/local/bin/mongodb-3.0.7/bin/bsondump`
`/usr/local/bin/mongodb-3.0.7/bin/bsondump -> /usr/local/bin/mongodb-3.0.7/bin/bsondump`
`/usr/local/bin/mongodb-3.0.7/bin/bsondump -> /usr/local/bin/mongodb-3.0.7/bin/bsondump`
`-> /usr/local/bin/mongodb-3.0.7/bin/bsondump ==> सारांश 0 /usr/local/bin/mongodb-3.0.7/bin/bsondump-100.3.0: 13 फ़ाइलें, 15400, 11 सेकंड में बनाई गई ==> /usr/local/bin/mongodb-3.0.7/bin/bsondump`
स्थापित कर रहा है त्रुटि: `brew link` चरण सफलतापूर्वक पूरा नहीं हुआ फ़ॉर्मूला बन गया, लेकिन `/usr/local/bin` में सिमलिक नहीं किया गया है `/usr/local/bin` को सिमलिक नहीं किया जा सका लक्ष्य `/usr/local/bin/mongodb-3.0.7/bin/bsondump` का एक सिमलिक है। आप इसे अनलिंक कर सकते हैं: `rm -f /usr/local/bin/bsondump`

लिंक को फोर्स करने और सभी संघर्ष करने वाली फ़ाइलों को ओवरराइट करने के लिए: `rm -rf /usr/local/bin/bsondump`

सभी फ़ाइलों को सूचीबद्ध करने के लिए जो हटा दी जाएंगी: `find /usr/local/bin -type f -exec rm -f {} \;`

संभावित संघर्ष करने वाली फ़ाइलें हैं: `/usr/local/bin/bsondump -> /usr/local/bin/mongodb-3.0.7/bin/bsondump`
`/usr/local/bin/mongodb-3.0.7/bin/bsondump -> /usr/local/bin/mongodb-3.0.7/bin/bsondump`
`-> /usr/local/bin/mongodb-3.0.7/bin/bsondump ==> सावधानियाँ /usr/local/bin/mongodb-3.0.7/bin/bsondump-100.3.0: 13 फ़ाइलें, 15400, 11 सेकंड में बनाया गया ==> /usr/local/bin/mongodb-3.0.7/bin/bsondump`
को अभी शुरू करने और लॉगिन पर पुनः आरंभ करने के लिए: `rm -rf /usr/local/bin/bsondump`
या, यदि आपको बैकग्राउंड सेवा की आवश्यकता/इच्छा नहीं है तो आप सिर्फ चला सकते हैं: `rm -rf /usr/local/bin/bsondump`
`/usr/local/bin/mongodb-3.0.7/bin/bsondump ==> सारांश 0 /usr/local/bin/mongodb-3.0.7/bin/bsondump-100.3.0: 13 फ़ाइलें, 15400, 11 सेकंड में बनाया गया ==> सावधानियाँ /usr/local/bin/mongodb-3.0.7/bin/bsondump-100.3.0: 13 फ़ाइलें, 15400, 11 सेकंड में बनाया गया ==> /usr/local/bin/mongodb-3.0.7/bin/bsondump`
को अभी शुरू करने और लॉगिन पर पुनः आरंभ करने के लिए: `rm -rf /usr/local/bin/bsondump`
या, यदि आपको बैकग्राउंड सेवा की आवश्यकता/इच्छा नहीं है तो आप सिर्फ चला सकते हैं: `rm -rf /usr/local/bin/bsondump`

/usr/local/Cellar/mongodb/3.0.7

```
```shell
```

```
$ brew unlink mongodb
```

```
/usr/local/Cellar/mongodb/3.0.7 ... 11
```

```
$ mongod --version
```

```
db version v4.4.3
```

```
Build Info: {
```

```
 "version": "4.4.3",
```

```
 "gitVersion": "913d6b62acfb344dde1b116f4161360acd8fd13",
```

```
 "modules": [],
```

```
 "allocator": "system",
```

```
 "environment": {
```

```
 "distarch": "x86_64",
```

```
 "target_arch": "x86_64"
```

```
 }
```

```
}
```

फिर mongod चलाकर डेटाबेस सर्वर शुरू करें। हालांकि, पहली बार शुरू करते समय यह कहा गया कि /data/db मौजूद नहीं है। हम एक डायरेक्टरी बनाते हैं, ~/mongodb, जहां डेटाबेस फाइलों को सहेजा जाएगा।

```
$ mongod --dbpath ~/mongodb
```

(यह कोड ब्लॉक है, इसे अनुवादित नहीं किया जाना चाहिए।)

आउटपुट होगा:

```
{"t":{"$date":"2021-03-11T18:17:32.838+08:00"},"s":"I", "c":"CONTROL", "id":23285, "ctx":"main","msg":"\n"}
{"t":{"$date":"2021-03-11T18:17:32.842+08:00"},"s":"W", "c":"ASIO", "id":22601, "ctx":"main","msg":"\n"}
{"t":{"$date":"2021-03-11T18:17:32.842+08:00"},"s":"I", "c":"NETWORK", "id":4648602, "ctx":"main","msg":"\n"}
{"t":{"$date":"2021-03-11T18:17:32.842+08:00"},"s":"I", "c":"STORAGE", "id":4615611, "ctx":"initandlisten","msg":"\n"}
{"t":{"$date":"2021-03-11T18:17:32.842+08:00"},"s":"I", "c":"CONTROL", "id":23403, "ctx":"initandlisten","msg":"\n"}
{"t":{"$date":"2021-03-11T18:17:32.843+08:00"},"s":"I", "c":"CONTROL", "id":51765, "ctx":"initandlisten","msg":"\n"}
...
```



देखा जा सकता है कि सभी JSON प्रारूप में हैं। [] में सभी डेटा फ़ाइलें JSON प्रारूप में सहेजी जाती हैं। इसके बाद, एक और टर्मिनल टैब खोलें।

```
$ mongo
MongoDB shell version v4.4.3
connecting to: mongodb://127.0.0.1:27017/?compressors=disabled&gssapiServiceName=mongodb
Implicit session: session { "id" : UUID("4f55c561-70d3-4289-938d-4b90a284891f") }
MongoDB server version: 4.4.3

:
2021-03-11T18:17:33.743+08:00:
2021-03-11T18:17:33.743+08:00: --bind_ip <address>
2021-03-11T18:17:33.743+08:00: rlimits
2021-03-11T18:17:33.743+08:00: currentValue: 4864
2021-03-11T18:17:33.743+08:00: recommendedMinimum: 64000

MongoDB - , (, CPU, ,)
MongoDB , URL , URL MongoDB
, : `db.enableFreeMonitoring()`
, : `db.disableFreeMonitoring()`
```

इसके बाद, आप डेटा इन्सर्ट करने और डेटा क्वेरी करने का प्रयास कर सकते हैं।

```
> db.inventory.insertOne(
... { item: "canvas", qty: 100, tags: ["cotton"], size: { h: 28, w: 35.5, uom: "cm" } }
...)
{
 "acknowledged" : true,
 "insertedId" : ObjectId("6049ef91b653541cf355facb")
}
>
> db.inventory.find()
{ "_id" : ObjectId("6049ef91b653541cf355facb"), "item" : "canvas", "qty" : 100, "tags" : ["cotton"], "size" : { "h" : 28, "w" : 35.5, "uom" : "cm" } }
```

## अंत में

यहाँ तक। बाद में हम अन्य टूल्स पर हाथ आजमाएंगे। हम ये सब कर रहे हैं, इसका क्या मतलब है। शायद पहले एक संरचना बनाना है। हर काम की शुरुआत मुश्किल होती है, और हमने शुरुआत में ही इन सभी को एक बार में कर लिया। इससे हमें आत्मविश्वास मिला है, और अब आगे, हम इन सॉफ्टवेयर को और अधिक एक्सप्लोर करेंगे।

## अभ्यास

□ छात्र ऊपर दिए गए तरीके से समान रूप से एक्सप्लोर करें।