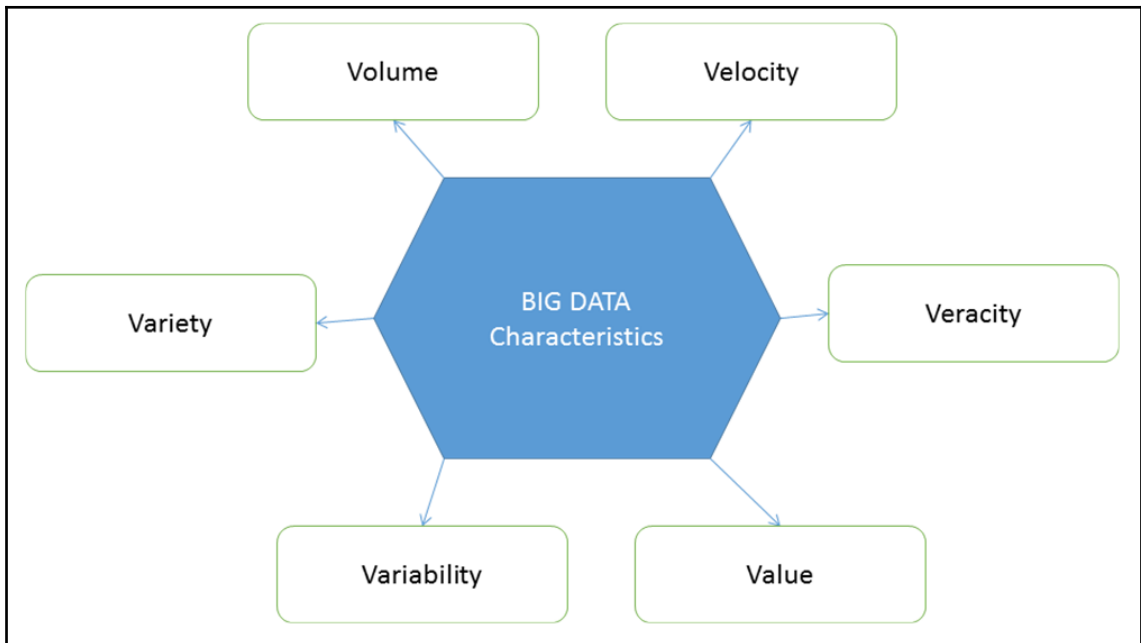
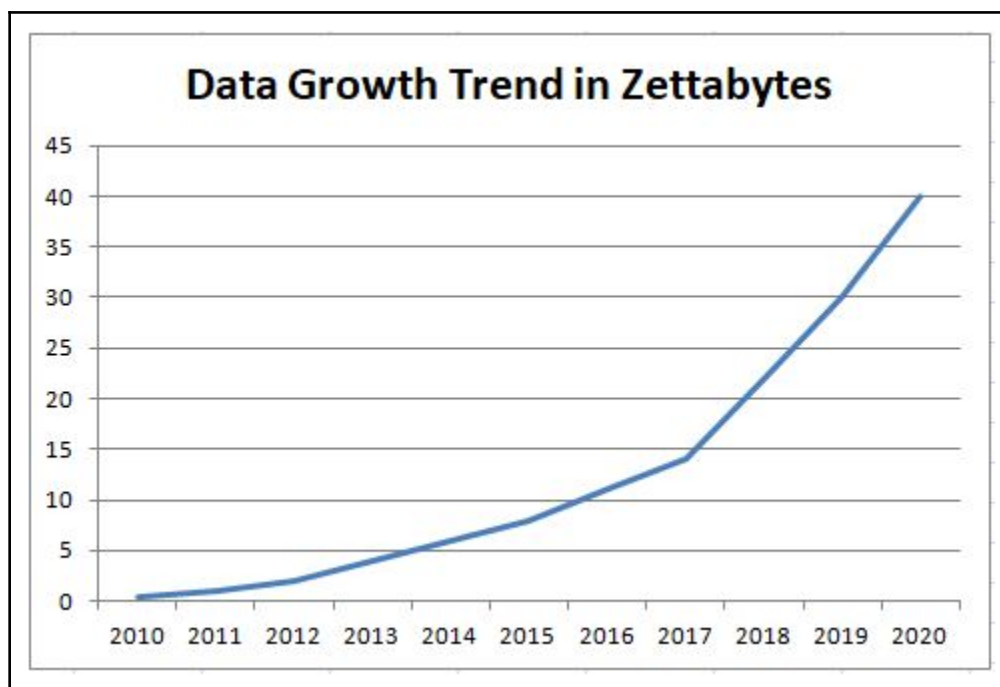
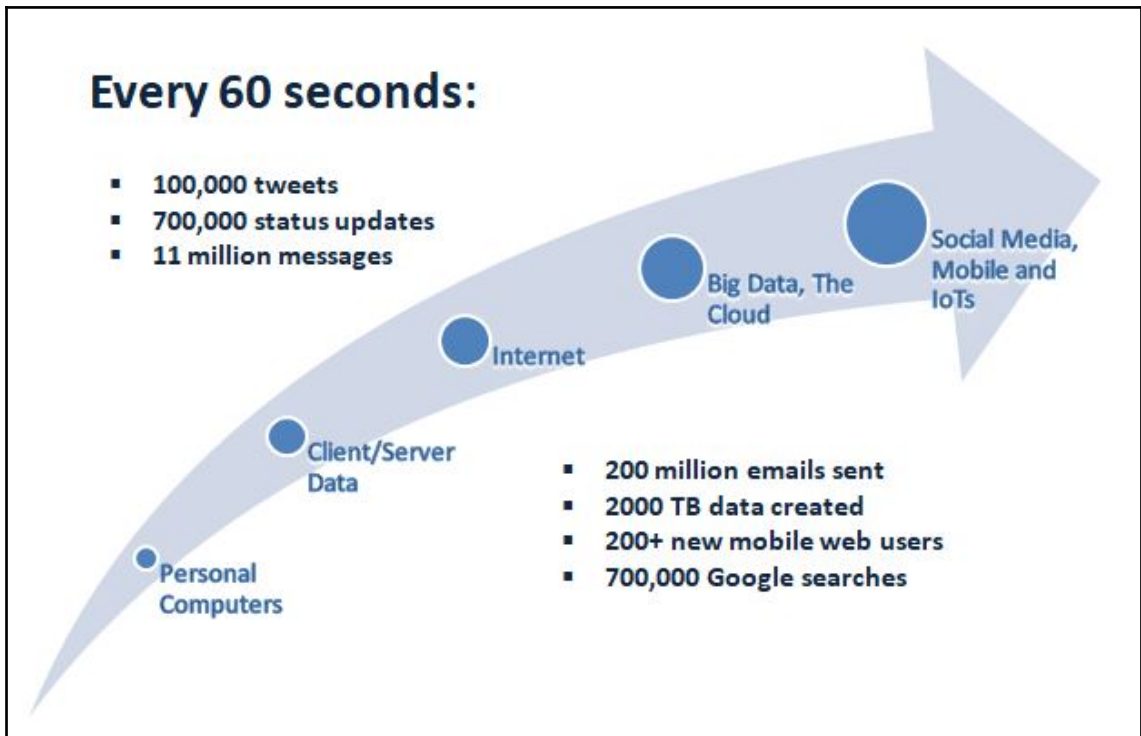


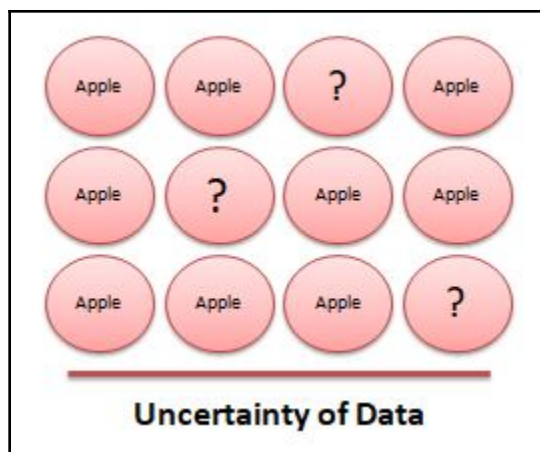
1 Graphics Bundle

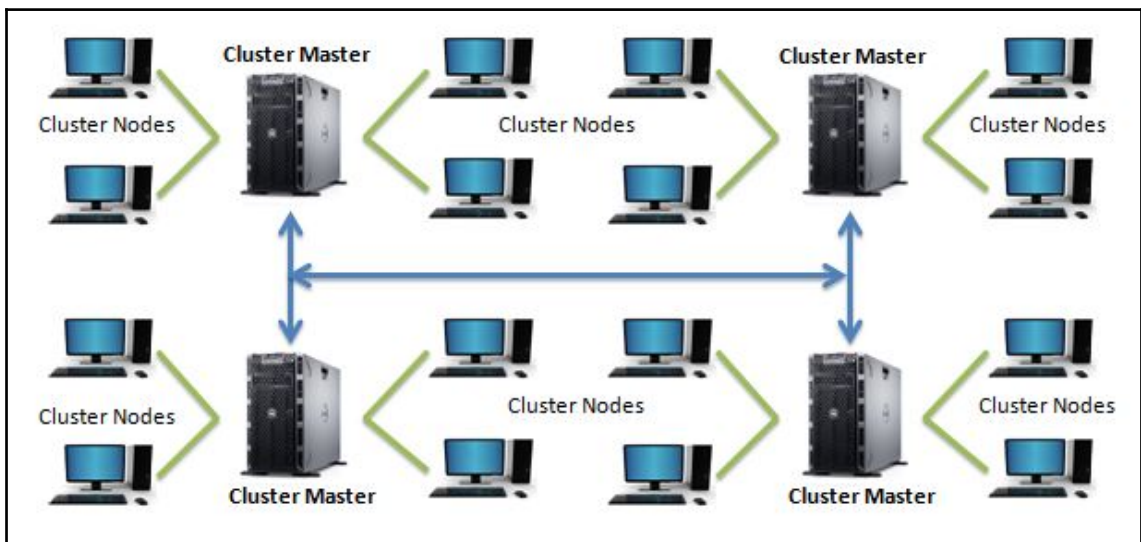
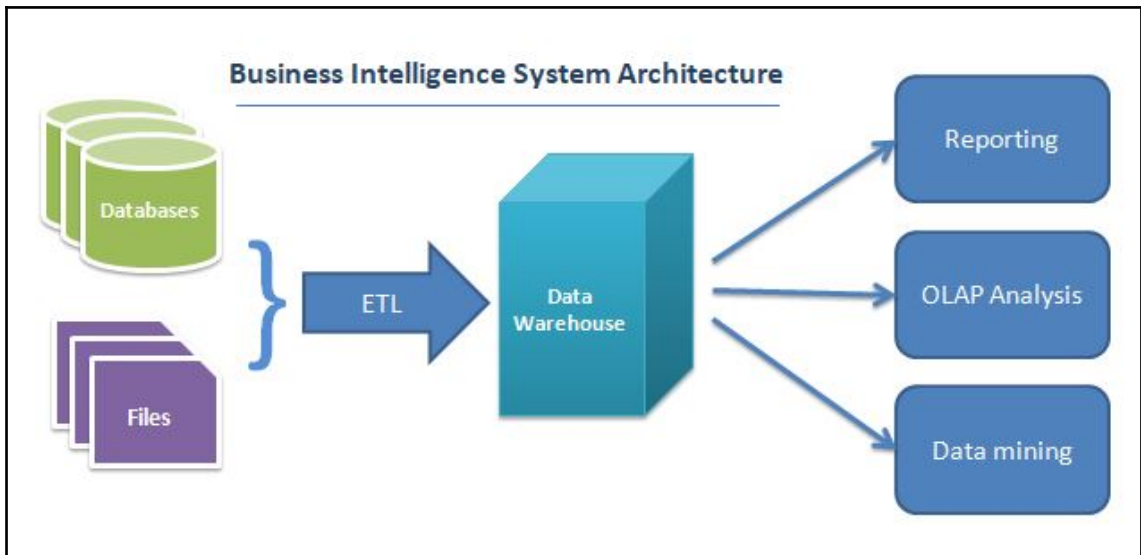
Chapter 1: Why Big Data

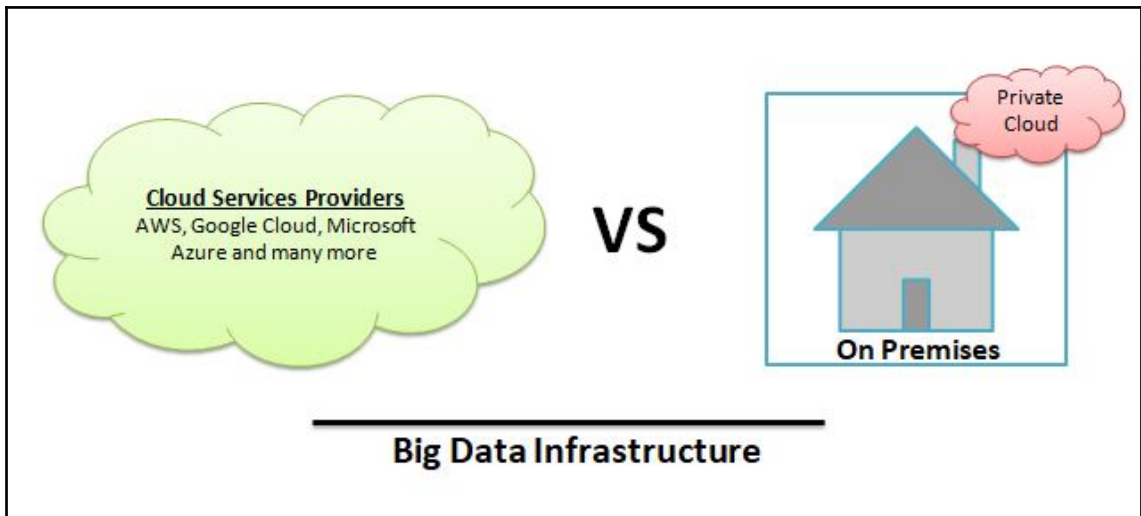






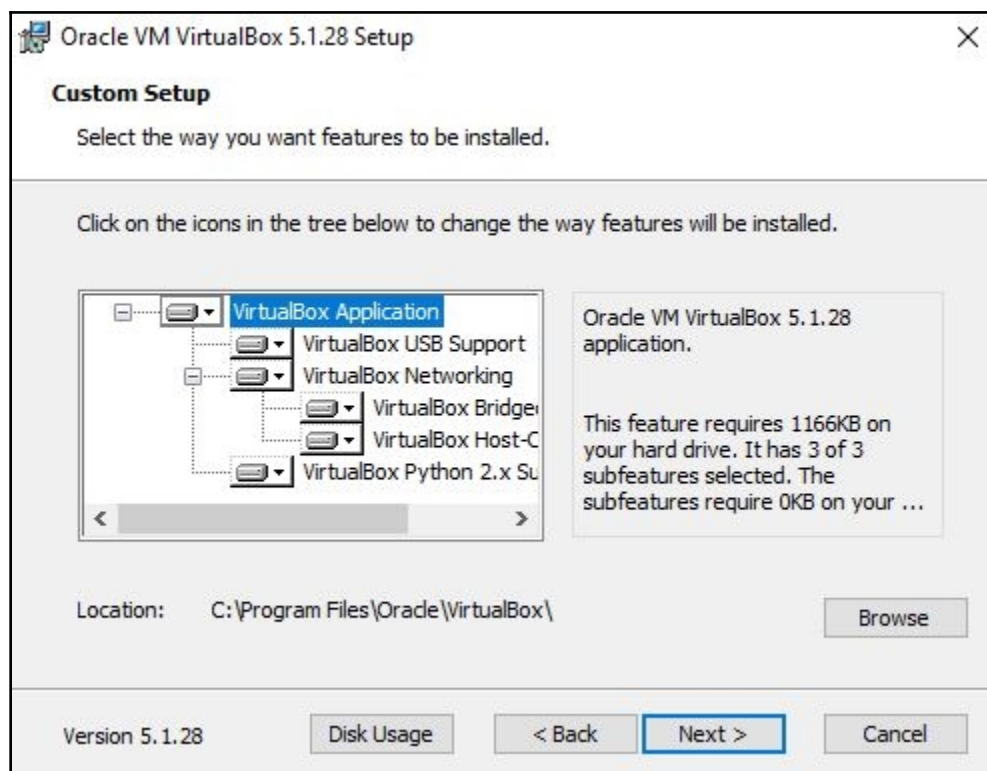


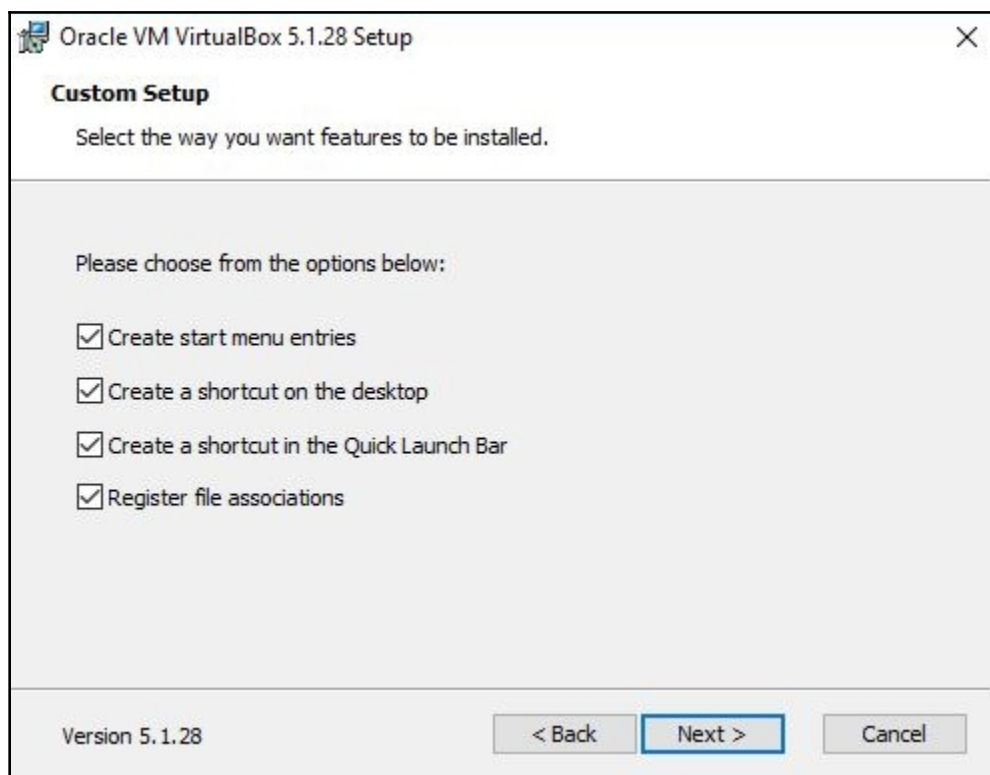




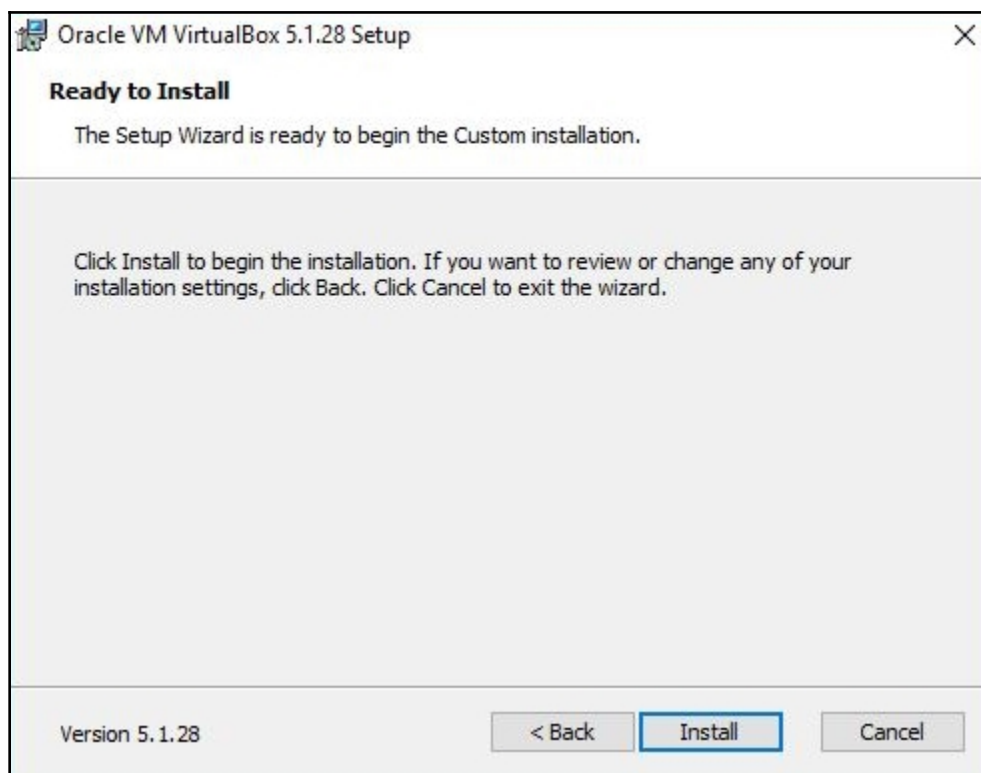
Chapter 2: Big Data Environment Setup



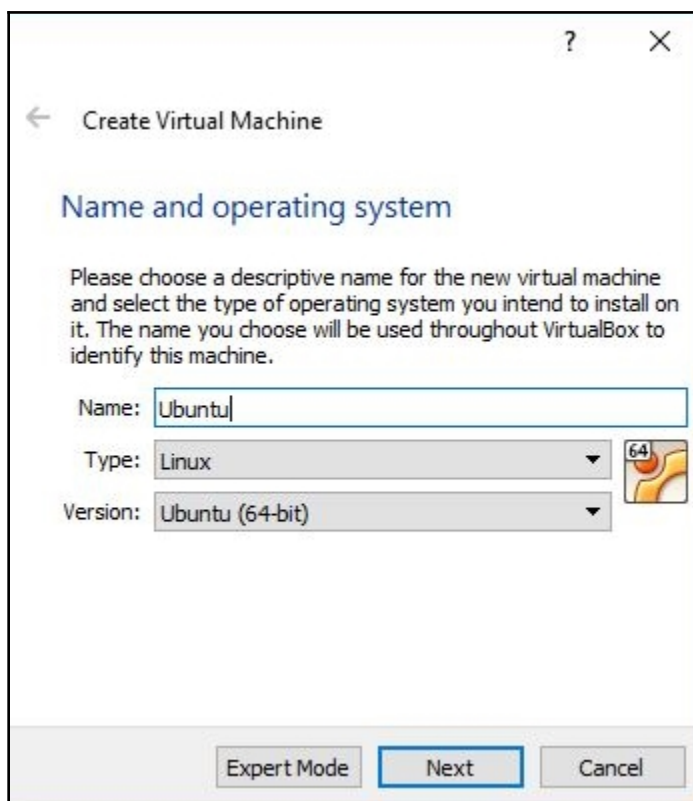


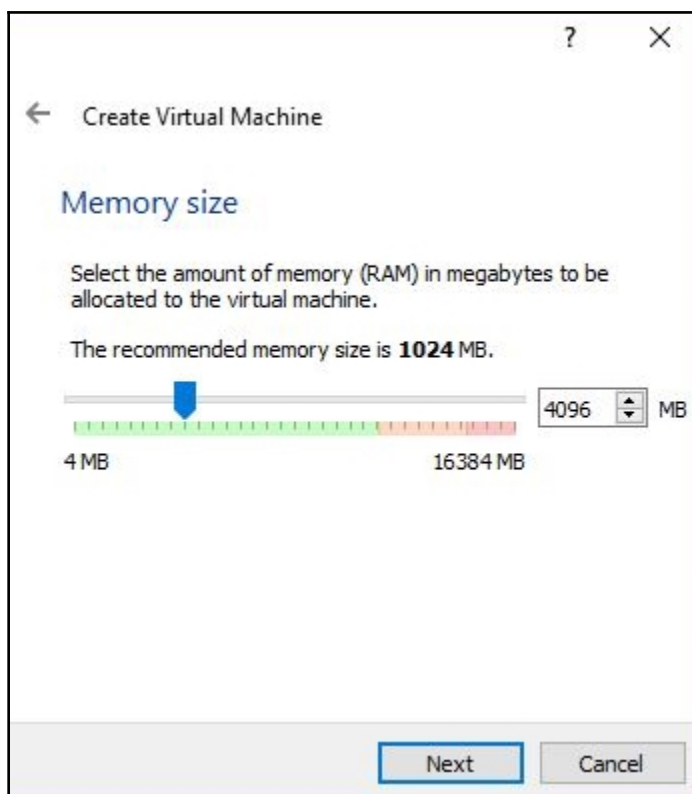


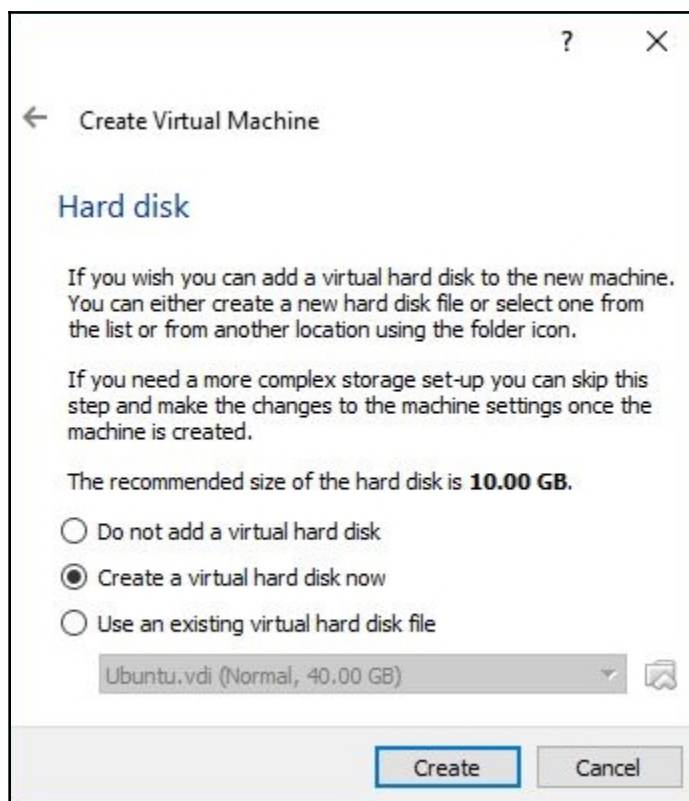


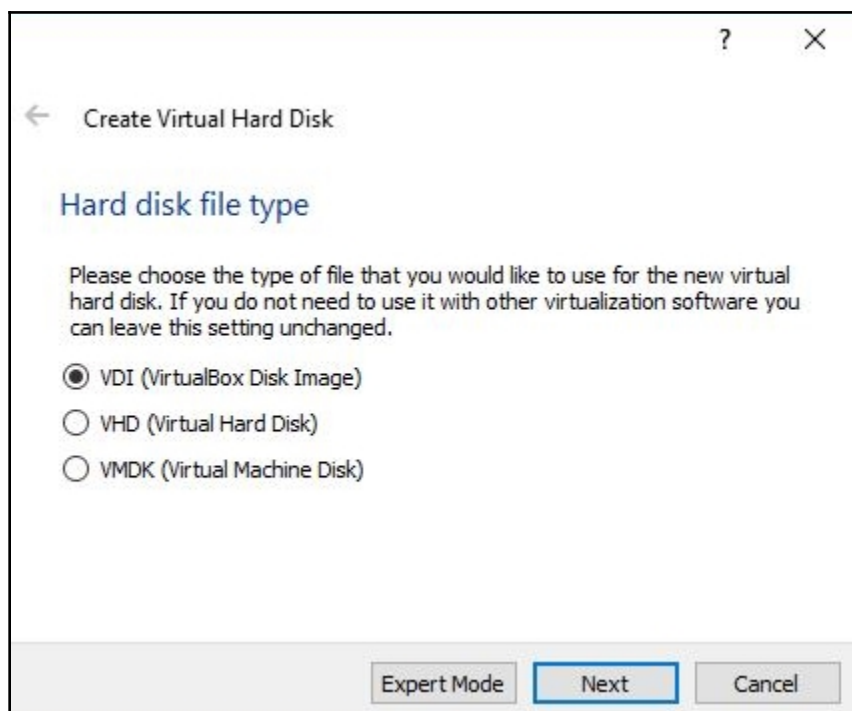


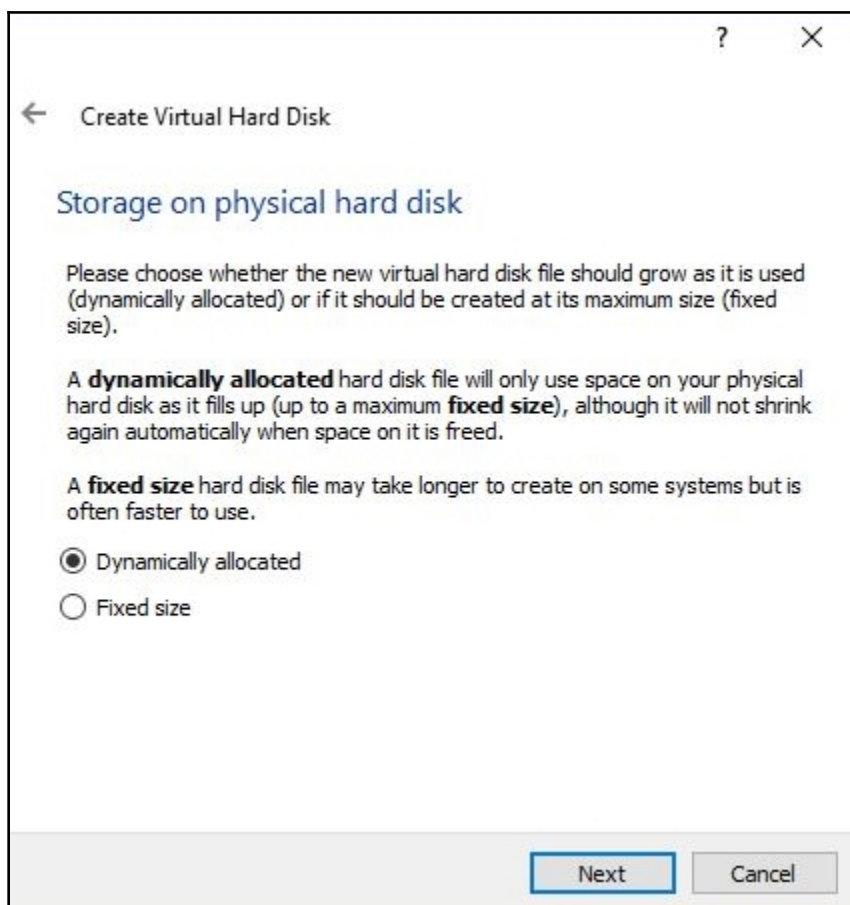


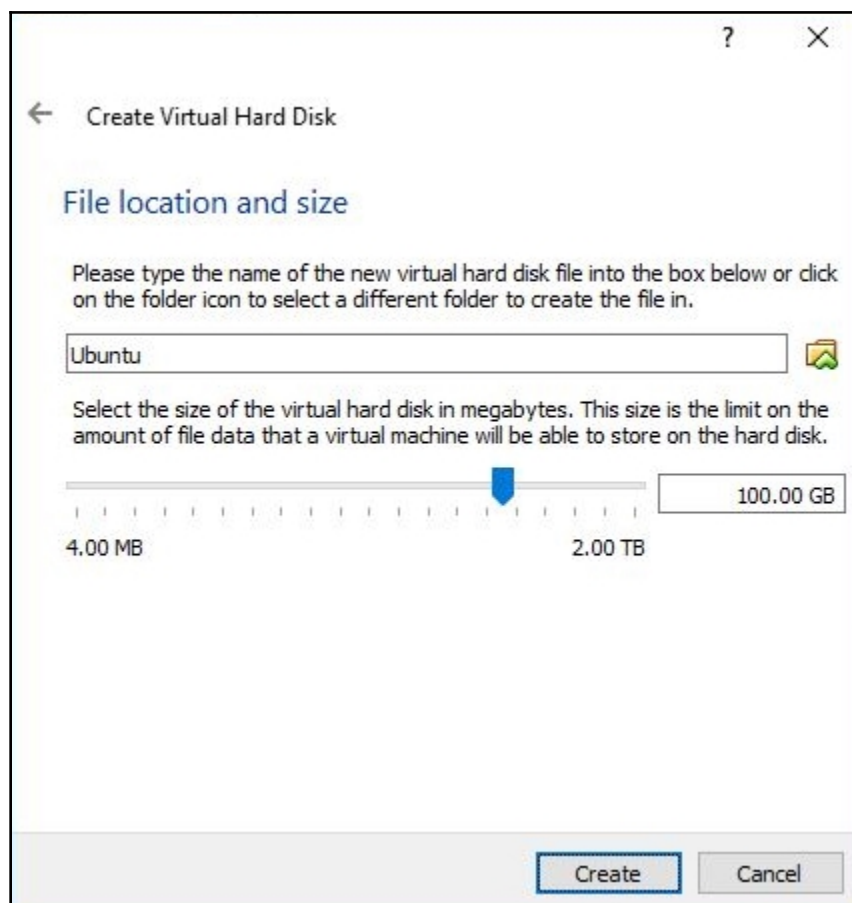


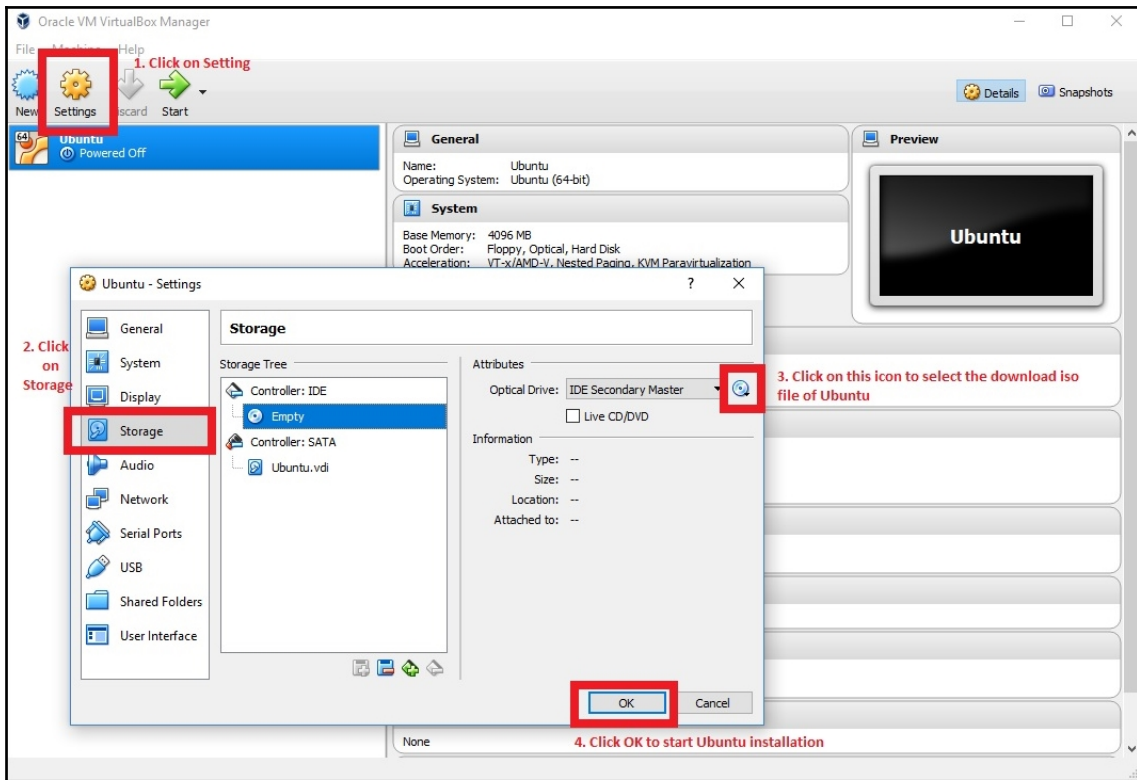


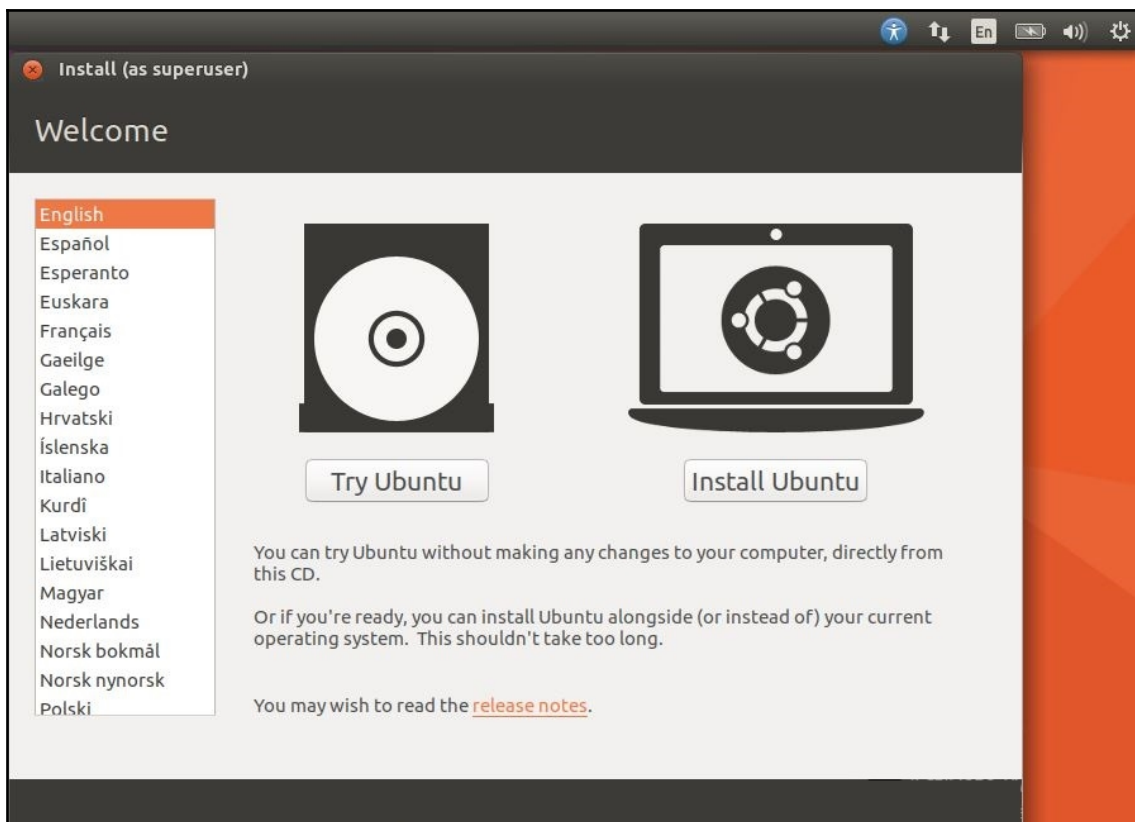


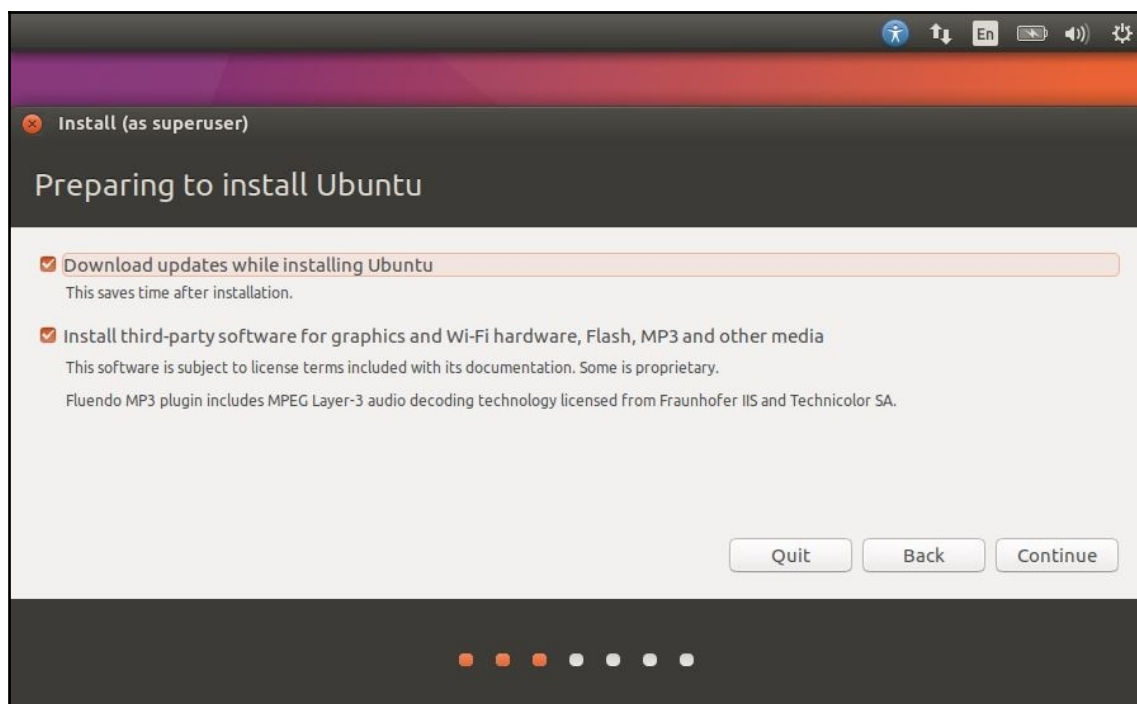


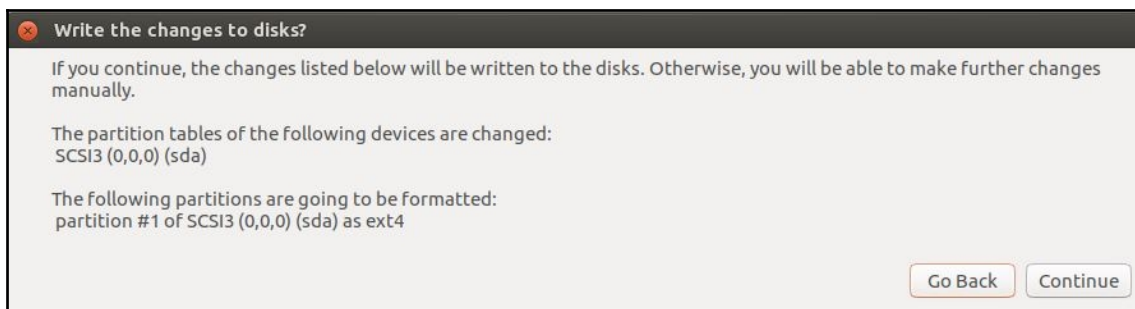
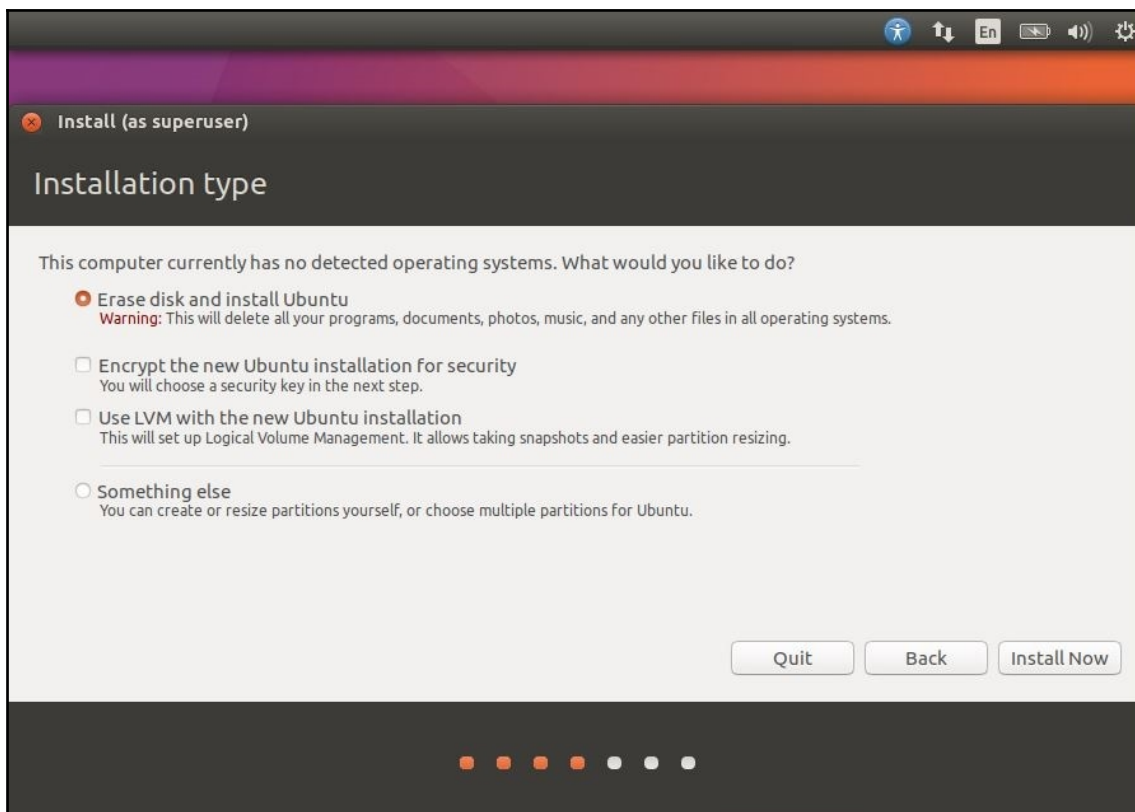


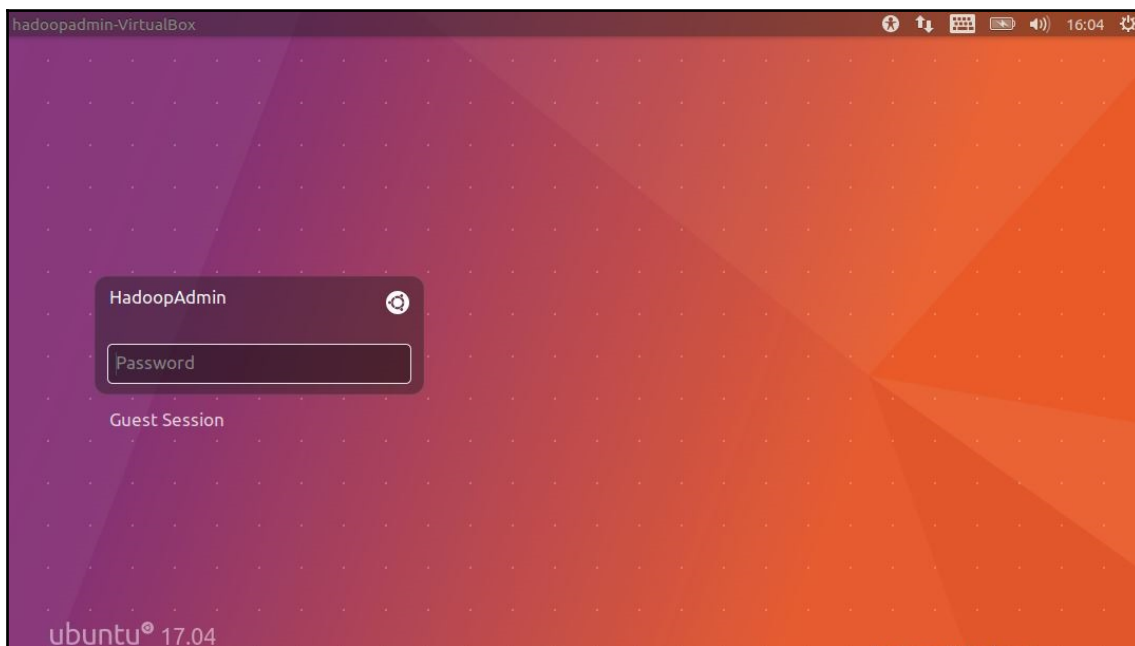
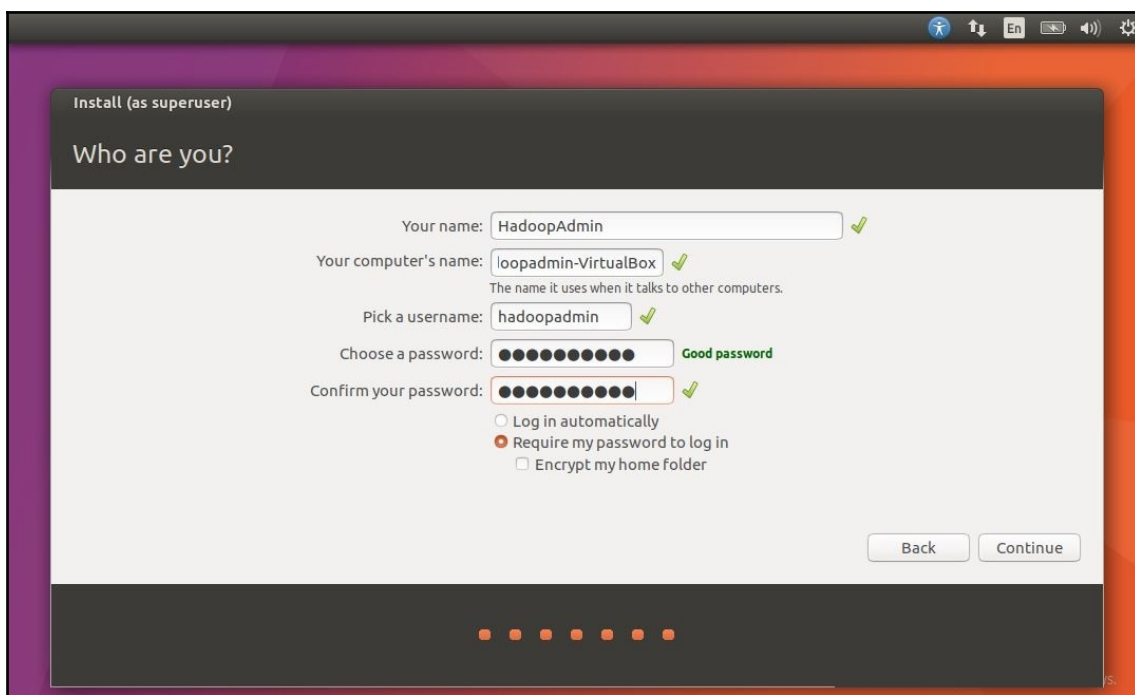


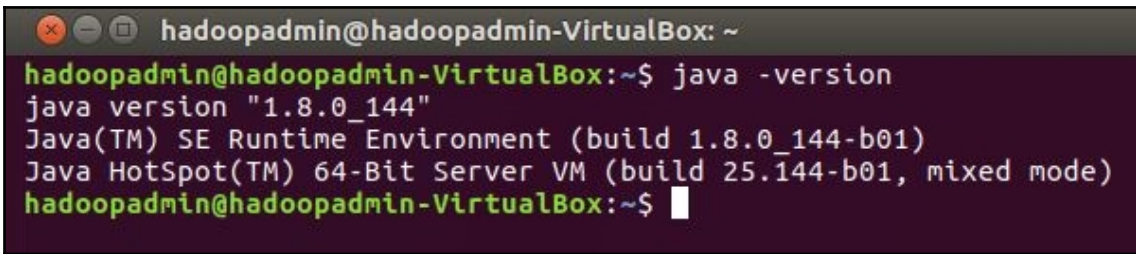






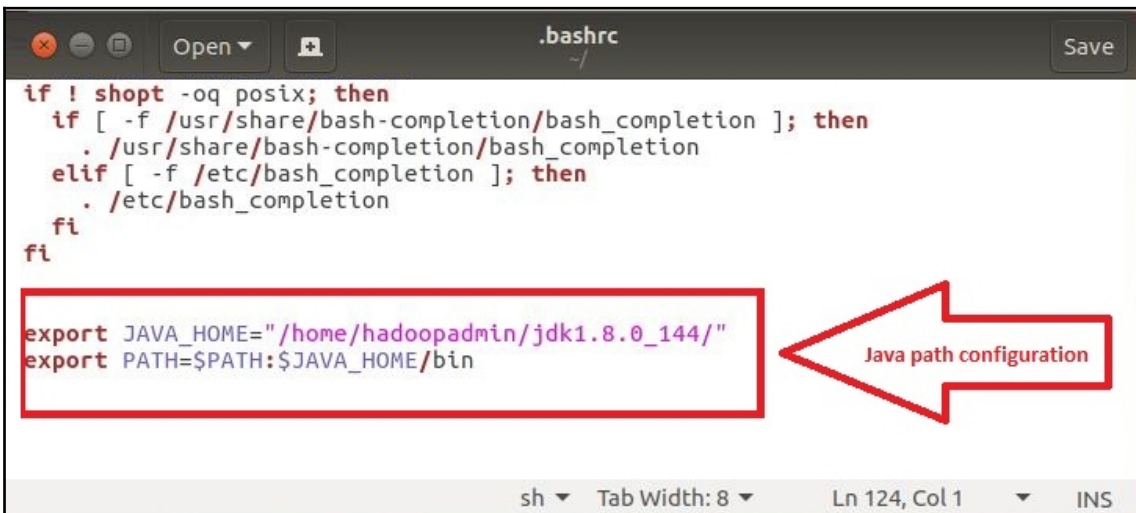






A terminal window titled 'hadoopadmin@hadoopadmin-VirtualBox: ~'. The command 'java -version' has been executed, resulting in the following output: 'java version "1.8.0_144"', 'Java(TM) SE Runtime Environment (build 1.8.0_144-b01)', and 'Java HotSpot(TM) 64-Bit Server VM (build 25.144-b01, mixed mode)'. The prompt is now 'hadoopadmin@hadoopadmin-VirtualBox:~\$'.

```
hadoopadmin@hadoopadmin-VirtualBox: ~  
hadoopadmin@hadoopadmin-VirtualBox:~$ java -version  
java version "1.8.0_144"  
Java(TM) SE Runtime Environment (build 1.8.0_144-b01)  
Java HotSpot(TM) 64-Bit Server VM (build 25.144-b01, mixed mode)  
hadoopadmin@hadoopadmin-VirtualBox:~$
```



An editor window titled '.bashrc' with a 'Save' button in the top right. The file contains shell script code for bash completion and environment variable exports. The last two lines are highlighted with a red rectangle: 'export JAVA_HOME="/home/hadoopadmin/jdk1.8.0_144/"' and 'export PATH=\$PATH:\$JAVA_HOME/bin'. A large red arrow points from this rectangle to the text 'Java path configuration'.

```
.bashrc  
~/  
Save  
  
if ! shopt -oq posix; then  
    if [ -f /usr/share/bash-completion/bash_completion ]; then  
        . /usr/share/bash-completion/bash_completion  
    elif [ -f /etc/bash_completion ]; then  
        . /etc/bash_completion  
    fi  
fi  
  
export JAVA_HOME="/home/hadoopadmin/jdk1.8.0_144/"  
export PATH=$PATH:$JAVA_HOME/bin  
  
sh Tab Width: 8 Ln 124, Col 1 INS
```

Java path configuration

```

hadoopadmin@hadoopadmin-VirtualBox: ~
hadoopadmin@hadoopadmin-VirtualBox:~$ ssh-keygen -t rsa -P ""
Generating public/private rsa key pair.
Enter file in which to save the key (/home/hadoopadmin/.ssh/id_rsa):
Created directory '/home/hadoopadmin/.ssh'.
Your identification has been saved in /home/hadoopadmin/.ssh/id_rsa.
Your public key has been saved in /home/hadoopadmin/.ssh/id_rsa.pub.
The key fingerprint is:
SHA256:hh6KxsybXZC5Y1QDClljr0ITc7981+eYKLIuOVMAbg hadoopadmin@hadoopadmin-VirtualBox
The key's randomart image is:
+---[RSA 2048]-----+
|o.o .                |
|.+. .o .hadoopadmin-VirtualBox:~$ cd ..
|... . . +hadoopadmin-VirtualBox:/home$ ls
|o. . . = . hadoopadmin
|E.. . * S+hadoopadmin-VirtualBox:/home$ cd hadoopadmin/
|oB = + = +hadoopadmin-VirtualBox:~$ ls
|+ + X = o o . . . . . music      public      videos
|  O B.+ o + . . . . . desktop Pictures templates
|  o.B +o... o . . . . . hadoopadmin-VirtualBox:~$ ls
+-----[SHA256]-----+
hadoopadmin@hadoopadmin-VirtualBox:~$

```

```
hadoopadmin@hadoopadmin-VirtualBox: ~  
hadoopadmin@hadoopadmin-VirtualBox:~$ ssh localhost  
The authenticity of host 'localhost (127.0.0.1)' can't be established.  
ECDSA key fingerprint is SHA256:HCW4EqtM/ly0H0vrWv8GDcu3ek1SPxREwD0r55qaaqk.  
Are you sure you want to continue connecting (yes/no)? yes  
Warning: Permanently added 'localhost' (ECDSA) to the list of known hosts.  
Welcome to Ubuntu 17.04 (GNU/Linux 4.10.0-37-generic x86_64)  
  
* Documentation:  https://help.ubuntu.com  
* Management:    https://landscape.canonical.com  
* Support:        https://ubuntu.com/advantage  
  
* CVE-2017-14106 (divide-by-zero vulnerability in the Linux TCP stack) is  
  now fixed in Ubuntu. Livepatch now, reboot later.  
  - https://ubu.one/CVE-2017-14106  
  - https://ubu.one/rebootLater  
  
111 packages can be updated.  
0 updates are security updates.  
  
The programs included with the Ubuntu system are free software;  
the exact distribution terms for each program are described in the  
individual files in /usr/share/doc/*/copyright.  
  
Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by  
applicable law.  
  
hadoopadmin@hadoopadmin-VirtualBox:~$
```



```

hadoopadmin@hadoopadmin-VirtualBox: ~
hadoopadmin@hadoopadmin-VirtualBox:~$ sudo addgroup hadoop
[sudo] password for hadoopadmin:
Adding group `hadoop' (GID 1001) ...
Done.
hadoopadmin@hadoopadmin-VirtualBox:~$ sudo adduser --ingroup hadoop hduser
Adding user `hduser' ...
Adding new user `hduser' (1001) with group `hadoop' ...
Creating home directory `/home/hduser' ...
Copying files from `/etc/skel' ...
Enter new UNIX password:
Retype new UNIX password:
passwd: password updated successfully
Changing the user information for hduser
Enter the new value, or press ENTER for the default
    Full Name []: Hadoop User
    Room Number []:
    Work Phone []:
    Home Phone []:
    Other []:
Is the information correct? [Y/n] Y
hadoopadmin@hadoopadmin-VirtualBox:~$ █

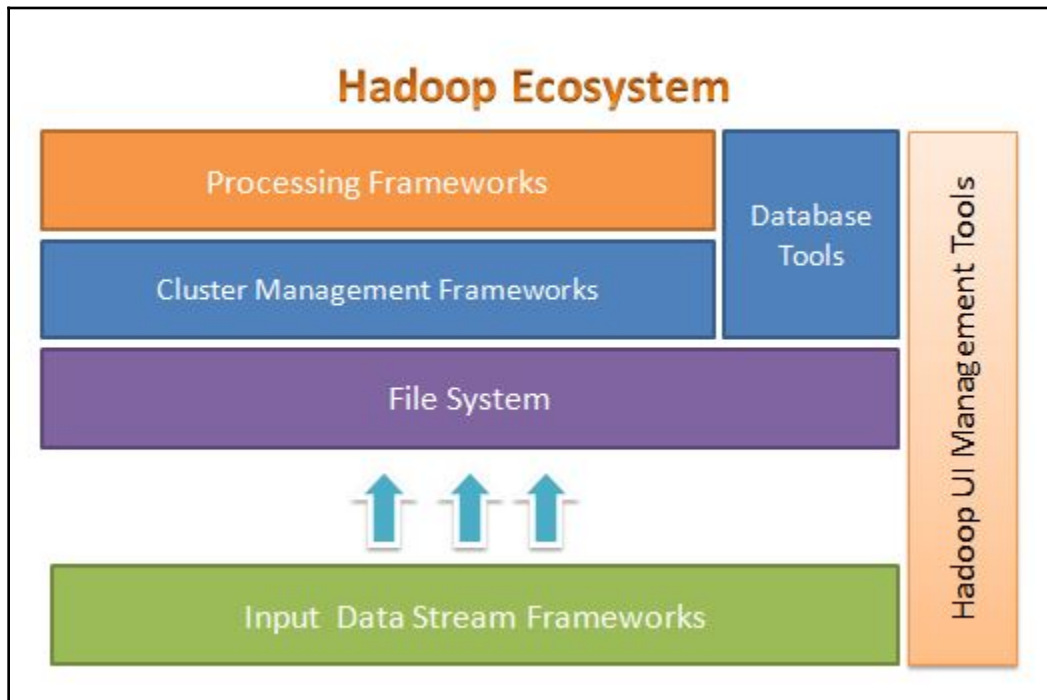
```

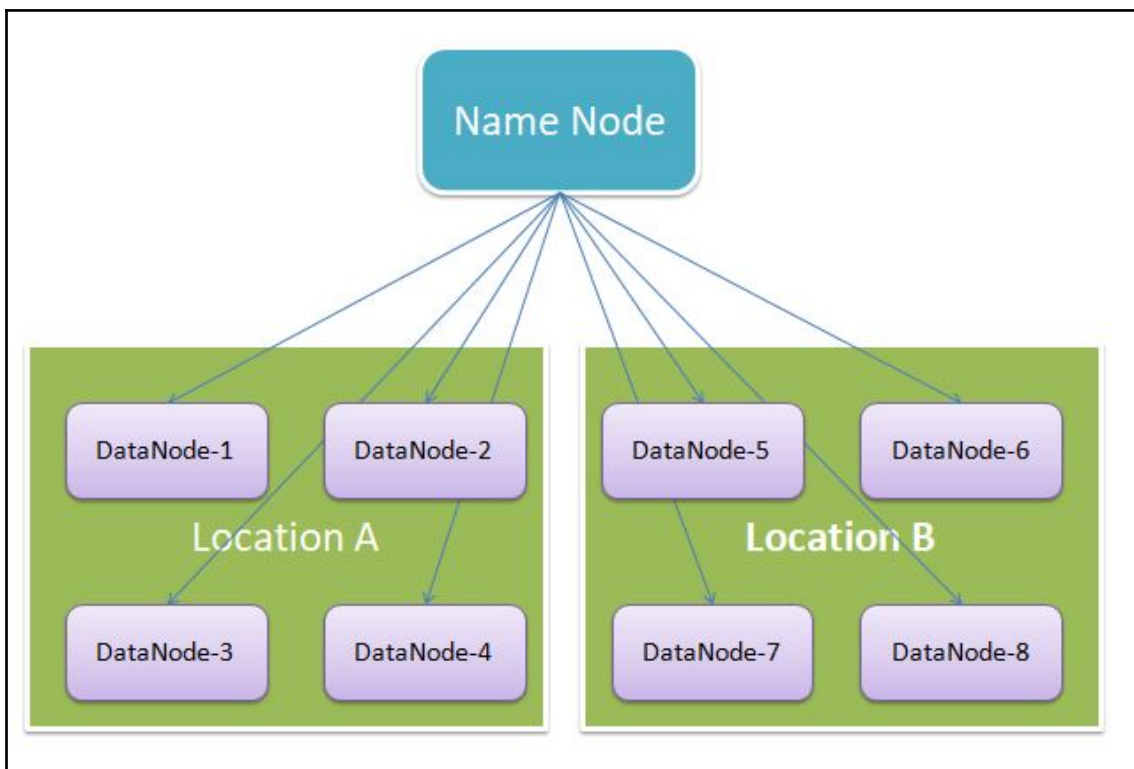
Namenode information x +	
localhost:50070/dfshealth.html#tab-overview	
Search	
Summary	
Security is off.	
Safemode is off.	
1 files and directories, 0 blocks = 1 total filesystem object(s).	
Heap Memory used 45.49 MB of 71.34 MB Heap Memory. Max Heap Memory is 966.69 MB.	
Non Heap Memory used 48.89 MB of 49.78 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.	
Configured Capacity:	97.93 GB
DFS Used:	24 KB (0%)
Non DFS Used:	9.67 GB
DFS Remaining:	83.25 GB (85.01%)
Block Pool Used:	24 KB (0%)
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 0.00%
Live Nodes	1 (Decommissioned: 0)
Dead Nodes	0 (Decommissioned: 0)

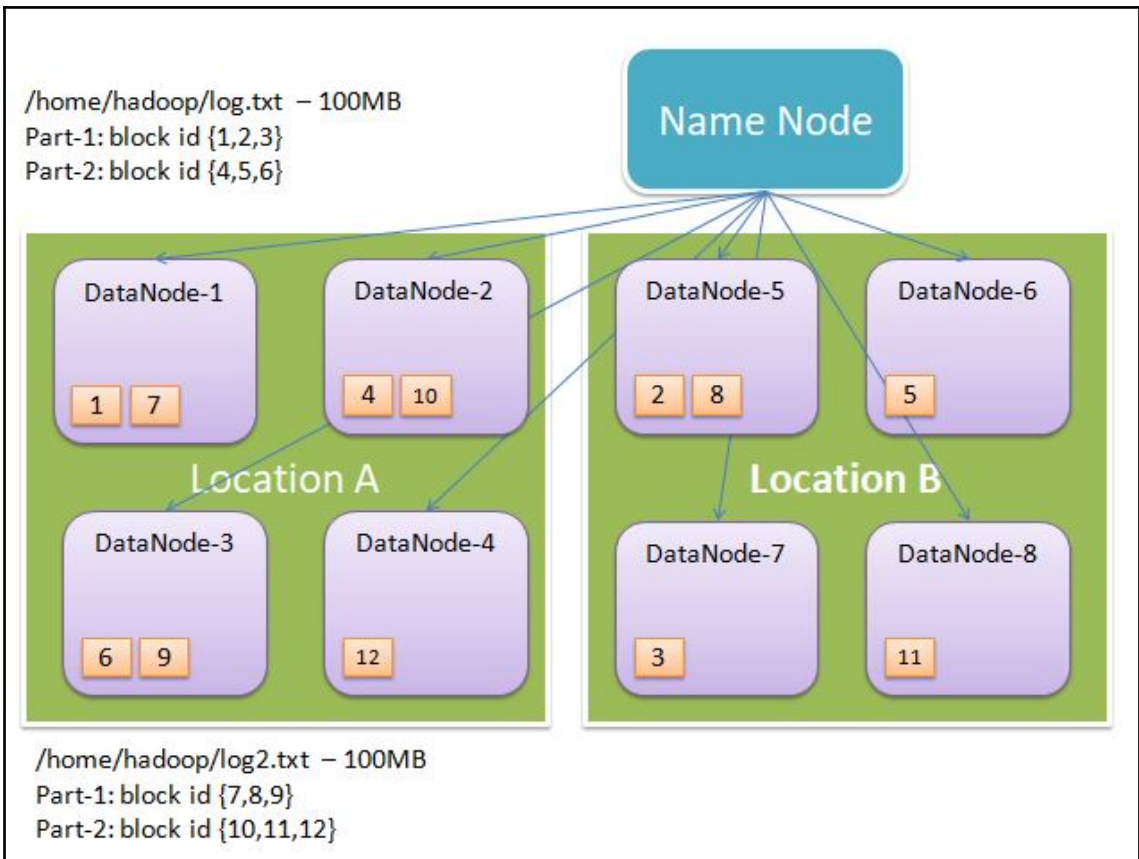
```
hadoopadmin@hadoopadmin-VirtualBox: ~/hadoop-2.8.1
hadoopadmin@hadoopadmin-VirtualBox:~/hadoop-2.8.1$ jps
3140 ResourceManager
3287 NodeManager
2791 DataNode
3609 Jps
2988 SecondaryNameNode
hadoopadmin@hadoopadmin-VirtualBox:~/hadoop-2.8.1$
```

```
hadoopadmin@hadoopadmin-VirtualBox: ~/hadoop-2.8.1
hadoopadmin@hadoopadmin-VirtualBox:~/hadoop-2.8.1$ jps
4272 Jps
hadoopadmin@hadoopadmin-VirtualBox:~/hadoop-2.8.1$
```


Chapter 3: Hadoop Ecosystem







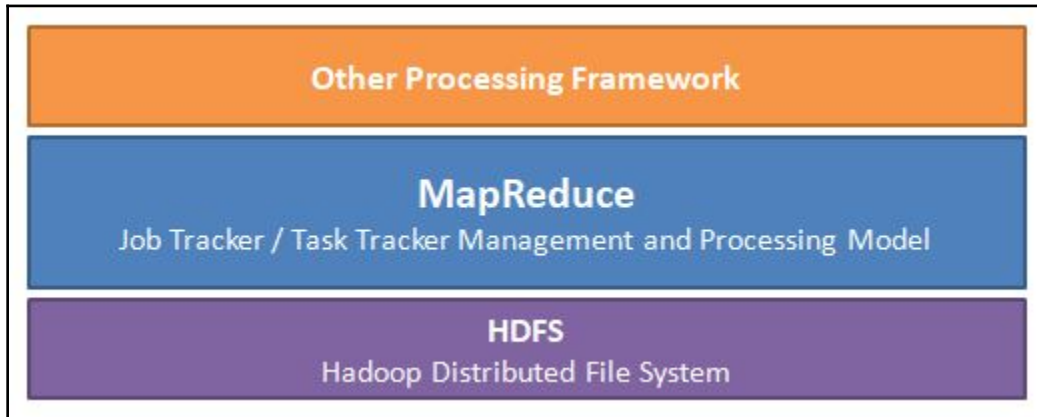
```
hadoopadmin@hadoopadmin-VirtualBox: ~  
hadoopadmin@hadoopadmin-VirtualBox:~$ hdfs dfs -mkdir /hadooptest  
hadoopadmin@hadoopadmin-VirtualBox:~$ hdfs dfs -ls /  
Found 1 items  
drwxr-xr-x  - hadoopadmin supergroup          0 2017-10-26 12:37 /hadooptest  
hadoopadmin@hadoopadmin-VirtualBox:~$
```

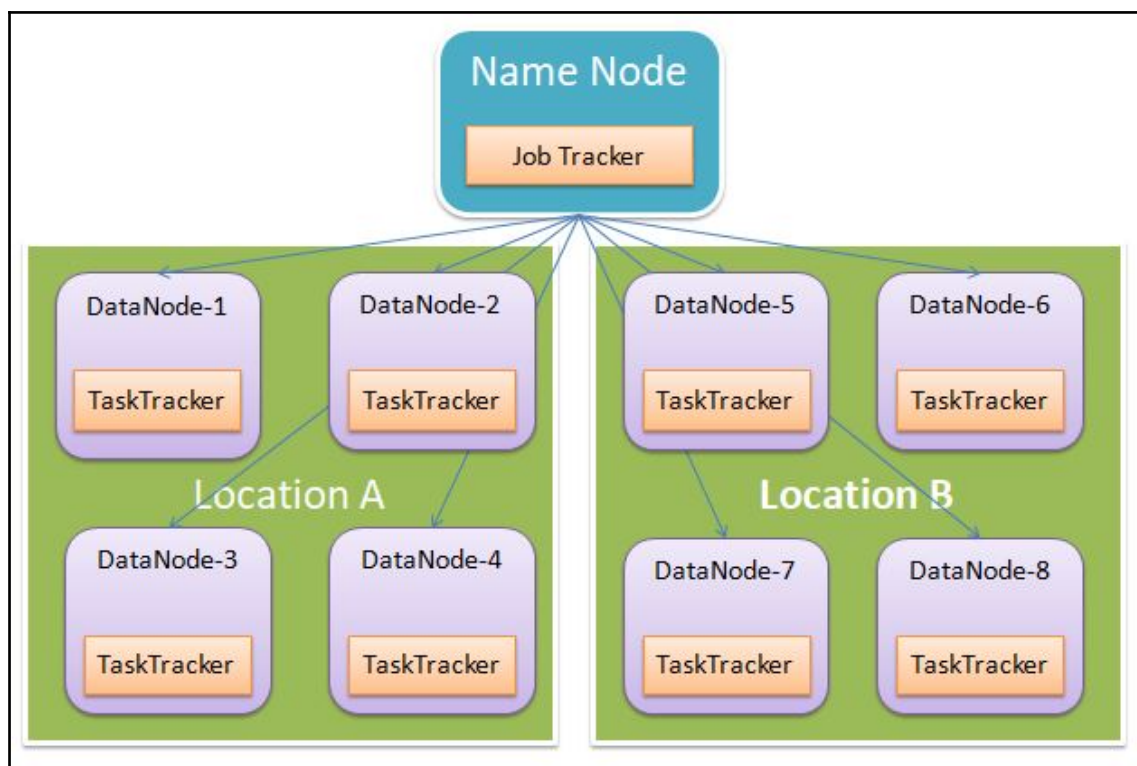
Overview "localhost:9000" (active)

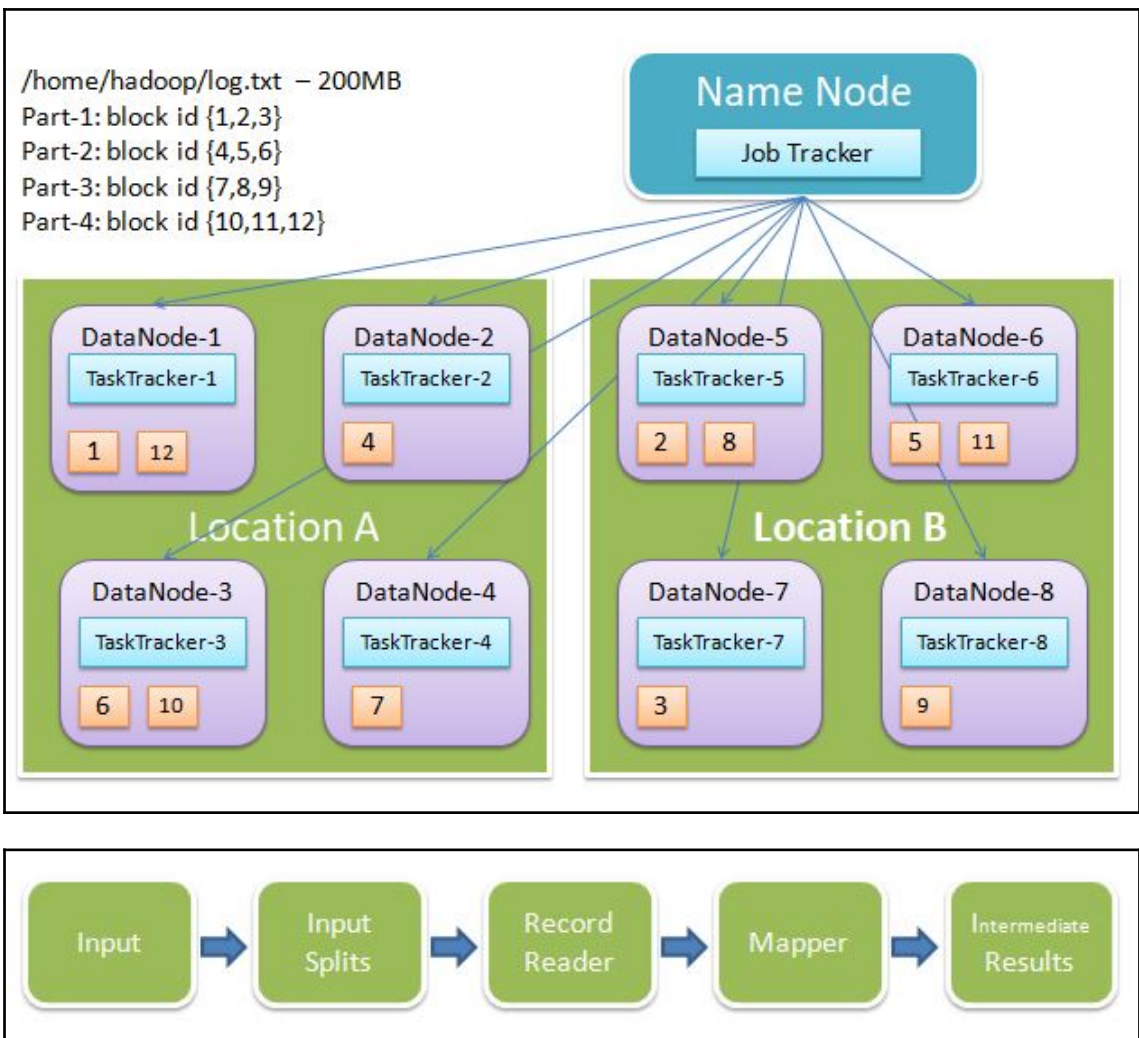
```
hadoopadmin@hadoopadmin-VirtualBox: ~  
hadoopadmin@hadoopadmin-VirtualBox:~$ hdfs dfs -put test.txt /hadooptest/  
hadoopadmin@hadoopadmin-VirtualBox:~$ hdfs dfs -ls /hadooptest  
Found 1 items  
-rw-r--r--    1 hadoopadmin supergroup          45 2017-10-26 12:55 /hadooptest/test.txt  
hadoopadmin@hadoopadmin-VirtualBox:~$
```

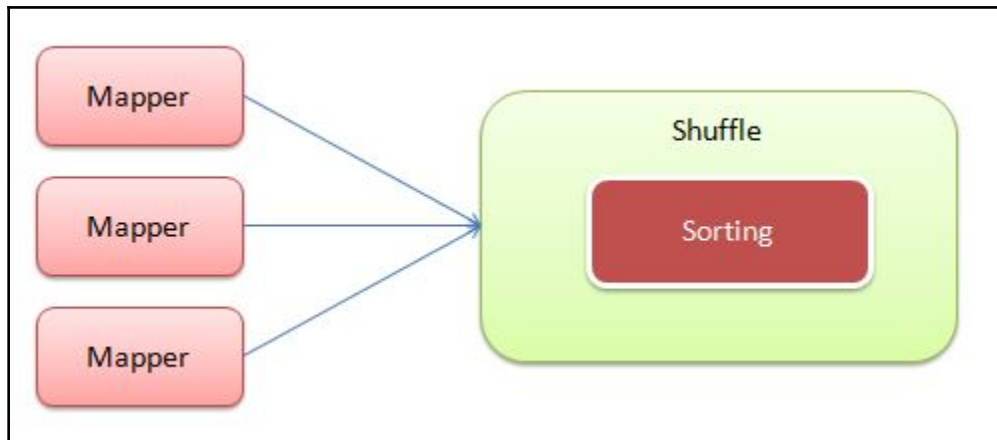
```
hadoopadmin@hadoopadmin-VirtualBox: ~  
hadoopadmin@hadoopadmin-VirtualBox:~$ mkdir hdfsfiles  
hadoopadmin@hadoopadmin-VirtualBox:~$ hdfs dfs -get /hadooptest/test.txt hdfsfiles/  
hadoopadmin@hadoopadmin-VirtualBox:~$ ls hdfsfiles/  
test.txt  
hadoopadmin@hadoopadmin-VirtualBox:~$
```

```
hadoopadmin@hadoopadmin-VirtualBox: ~  
hadoopadmin@hadoopadmin-VirtualBox:~$ hdfs dfs -rm /hadooptest/test.txt  
Deleted /hadooptest/test.txt  
hadoopadmin@hadoopadmin-VirtualBox:~$ hdfs dfs -rmdir /hadooptest  
hadoopadmin@hadoopadmin-VirtualBox:~$
```





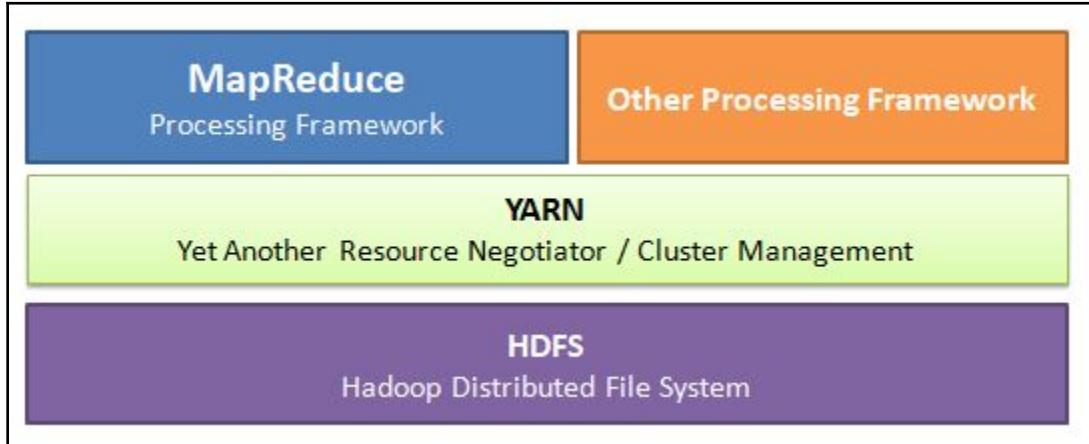


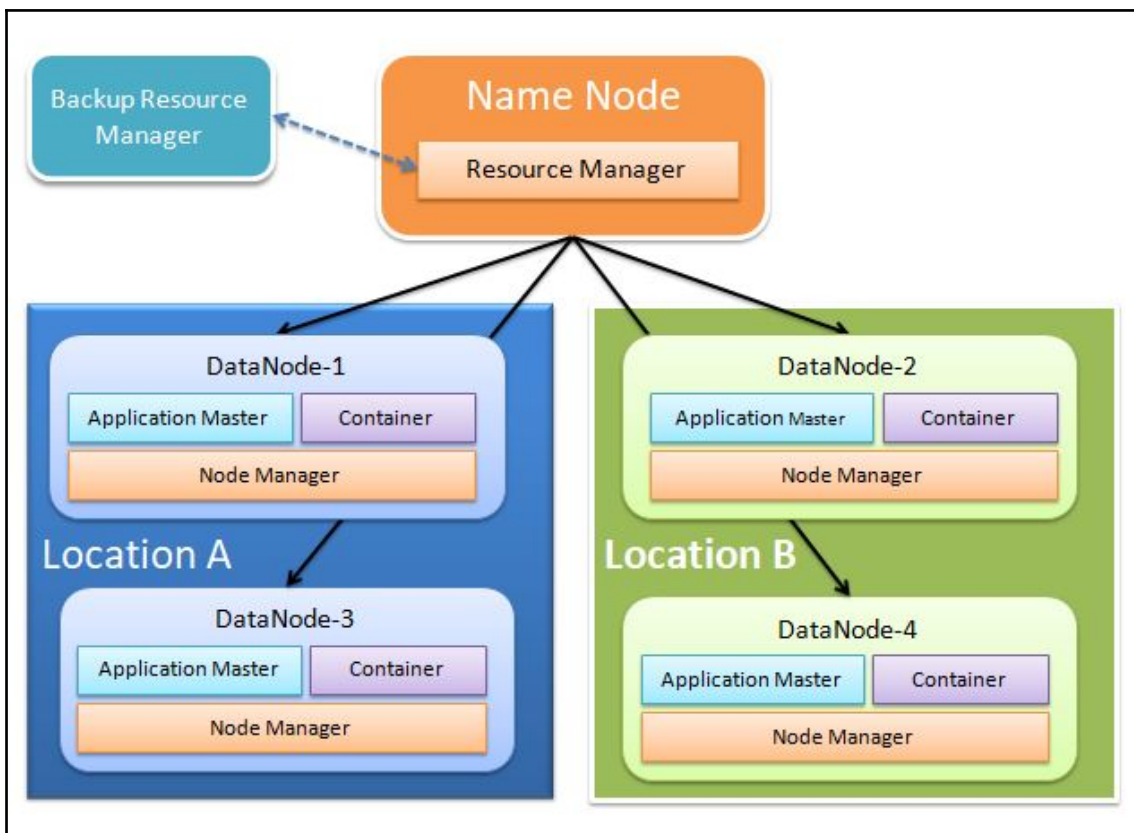


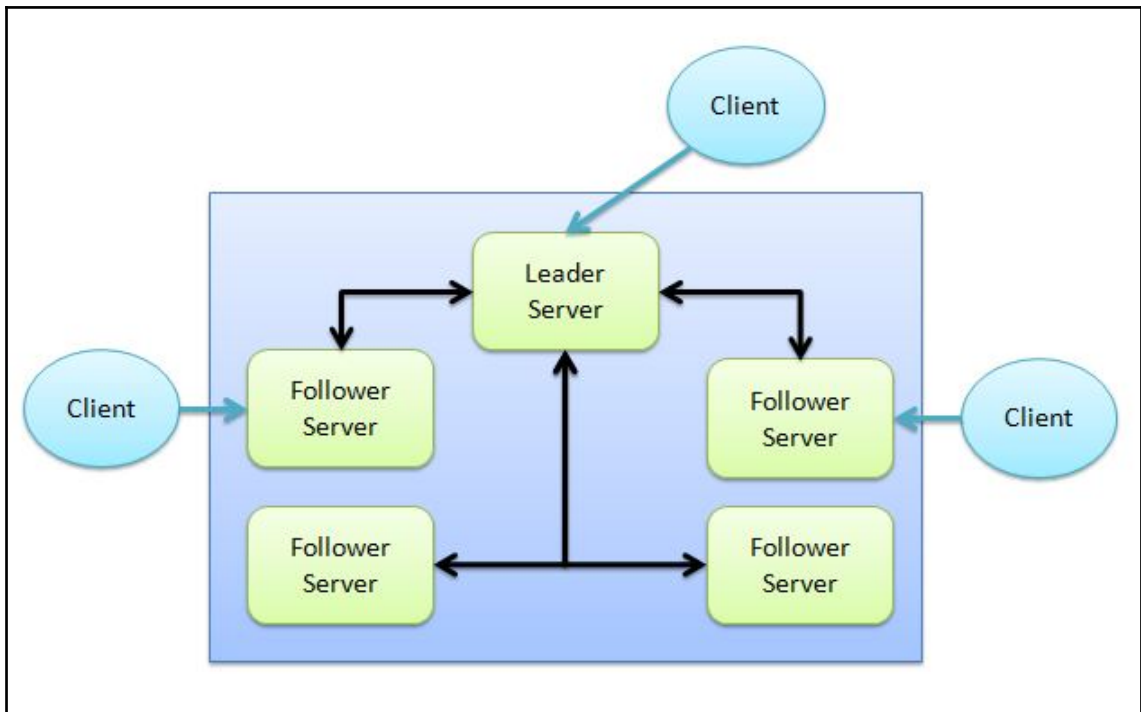
```
INFO mapred.LocalJobRunner: reduce > reduce
INFO mapred.Task: Task 'attempt_local482805000_0001_r_000000_0' done.
INFO mapred.LocalJobRunner: Finishing task: attempt_local482805000_0001_r_000000_0
INFO mapred.LocalJobRunner: reduce task executor complete.
INFO mapreduce.Job: map 100% reduce 100%
INFO mapreduce.Job: Job job_local482805000_0001 completed successfully
INFO mapreduce.Job: Counters: 35
em Counters      blocks
ILE: Number of bytes read=6866
ILE: Number of bytes written=653248
ILE: Number of read operations=0
ILE: Number of large read operations=0
ILE: Number of write operations=0
DFS: Number of bytes read=256
```

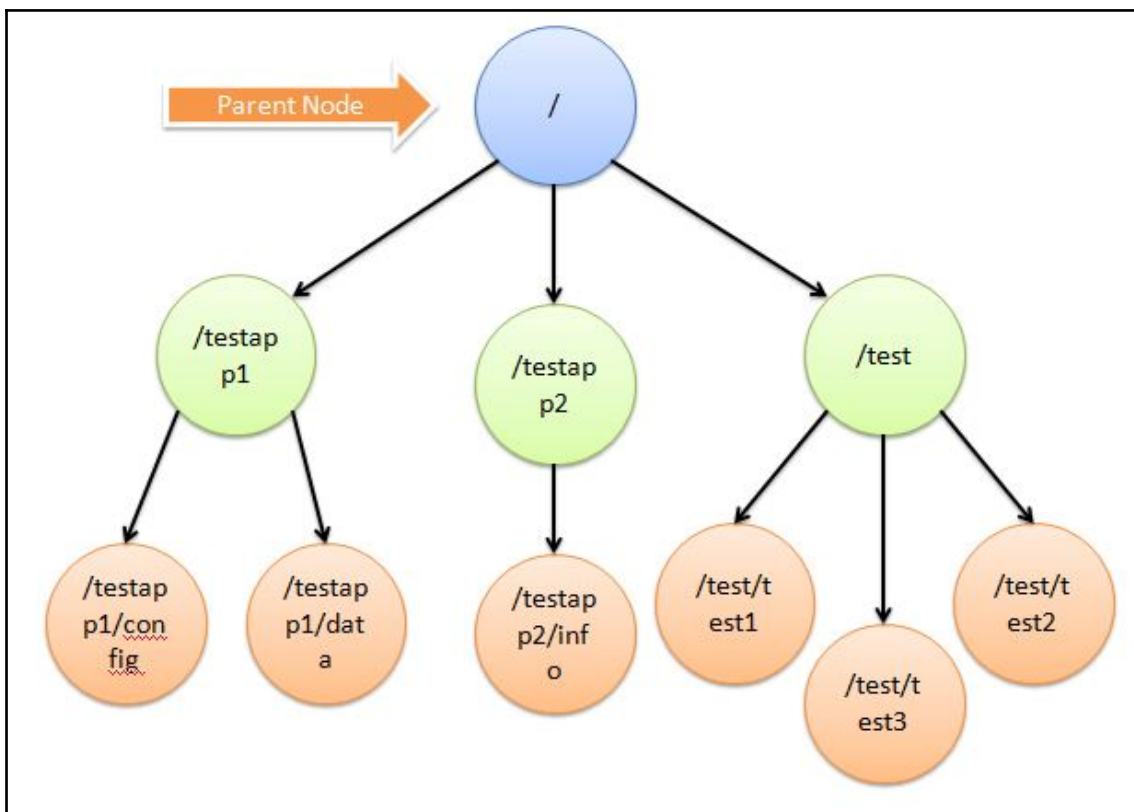
```
hadoopadmin@hadoopadmin-VirtualBox: ~
hadoopadmin@hadoopadmin-VirtualBox:~$ hdfs dfs -ls /utilityoutput
Found 2 items
-rw-r--r-- 1 hadoopadmin supergroup 0 2017-11-02 15:48 /utilityoutput/_SUCCESS
-rw-r--r-- 1 hadoopadmin supergroup 41 2017-11-02 15:48 /utilityoutput/part-r-00000
hadoopadmin@hadoopadmin-VirtualBox:~$
```

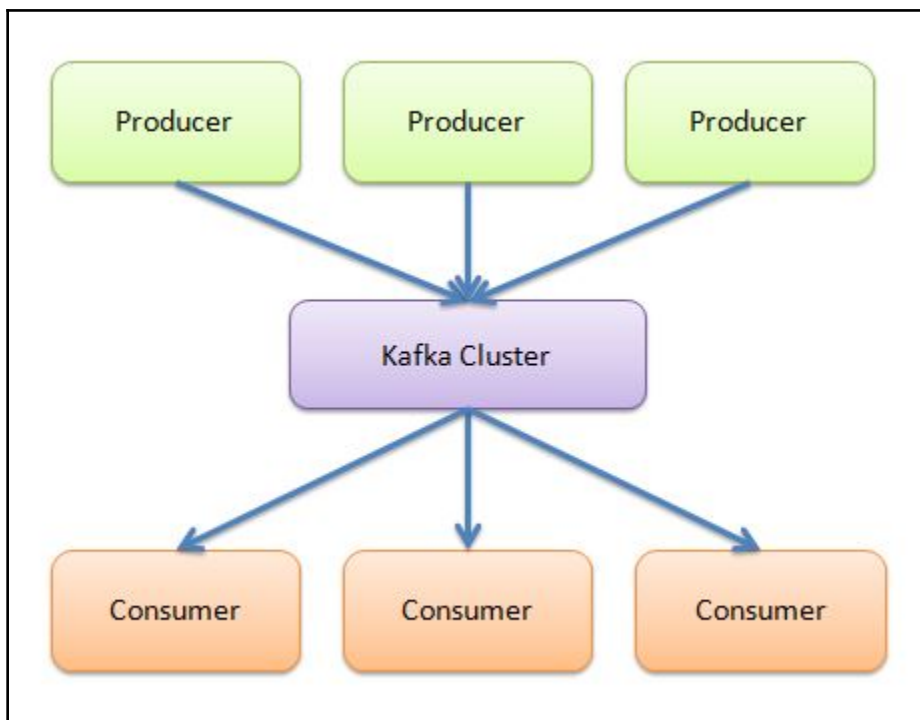
```
hadoopadmin@hadoopadmin-VirtualBox: ~  
hadoopadmin@hadoopadmin-VirtualBox:~$ hdfs dfs -cat /utilityoutput/part-r-00000  
electricity 4230  
gas 1550  
telephone 2500  
hadoopadmin@hadoopadmin-VirtualBox:~$
```

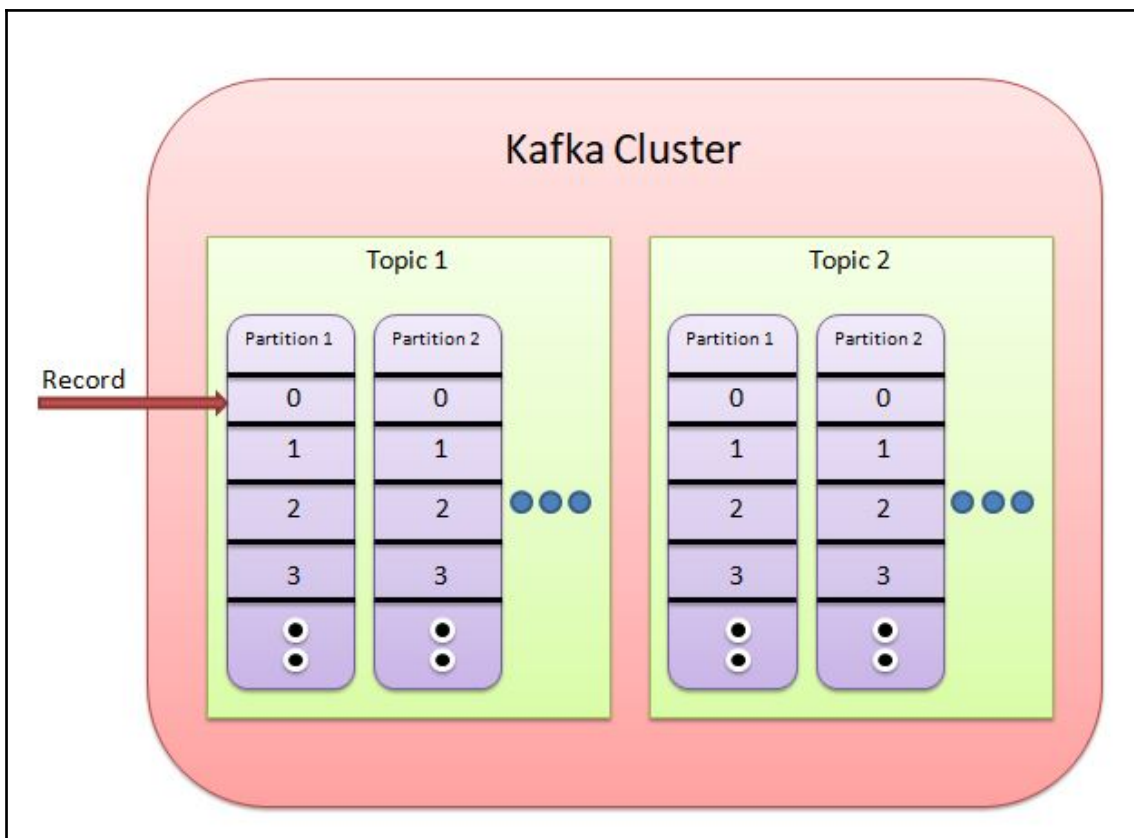


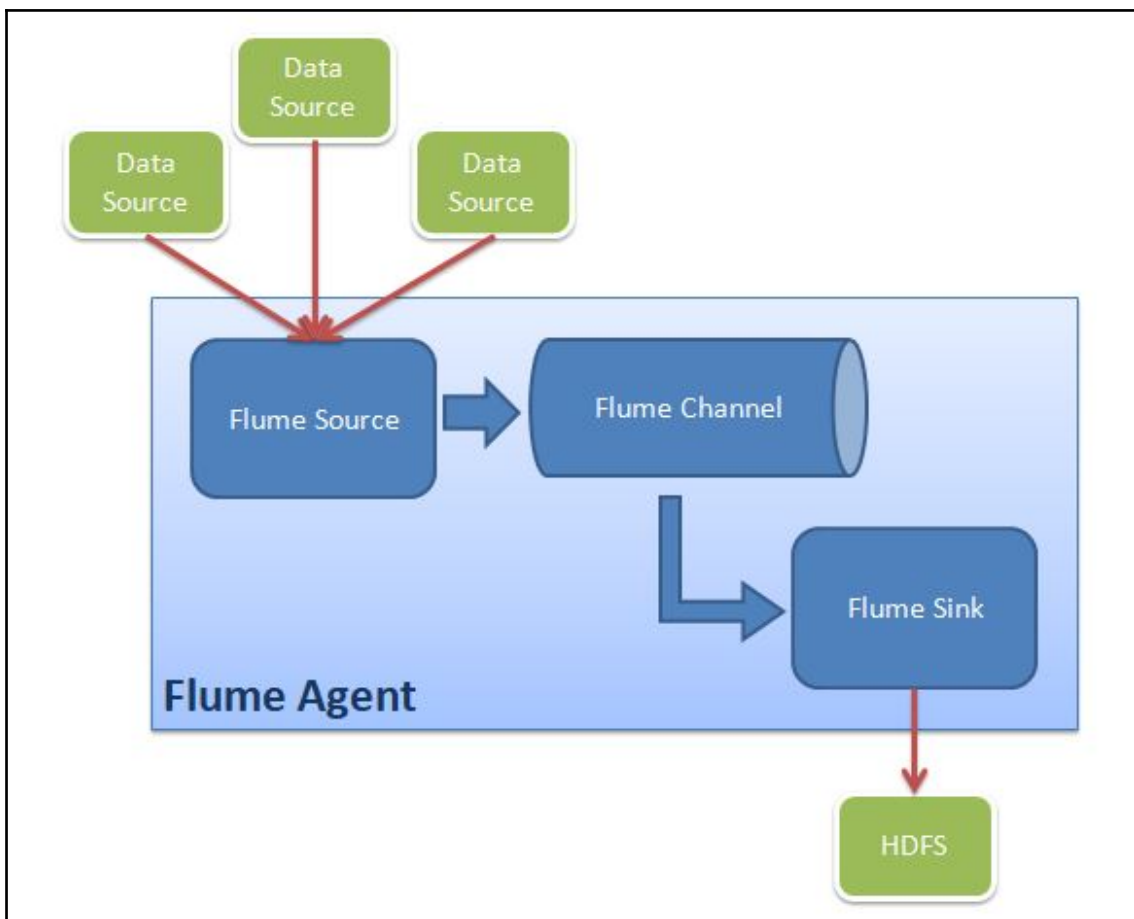


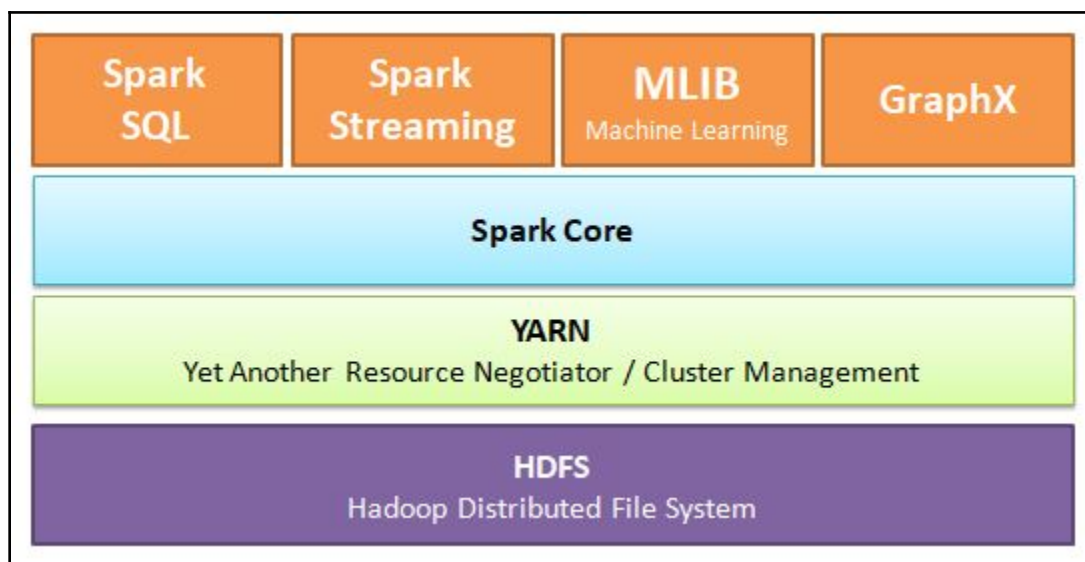




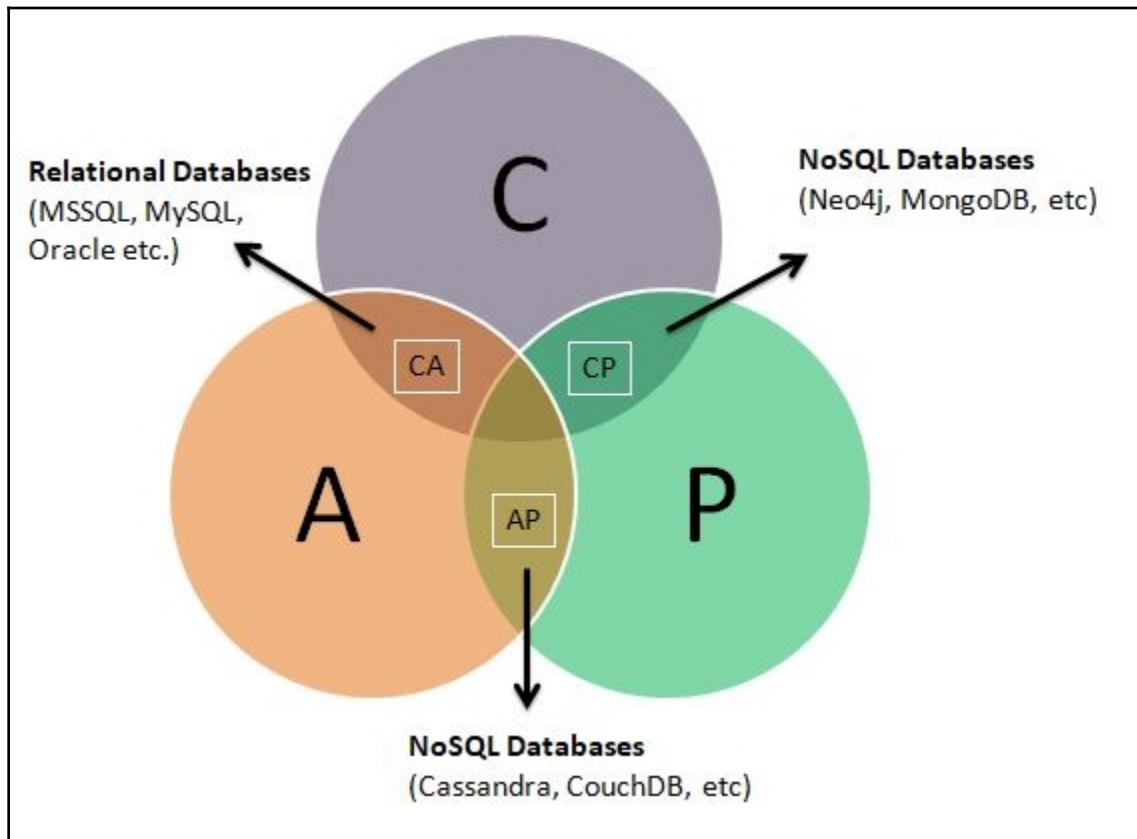


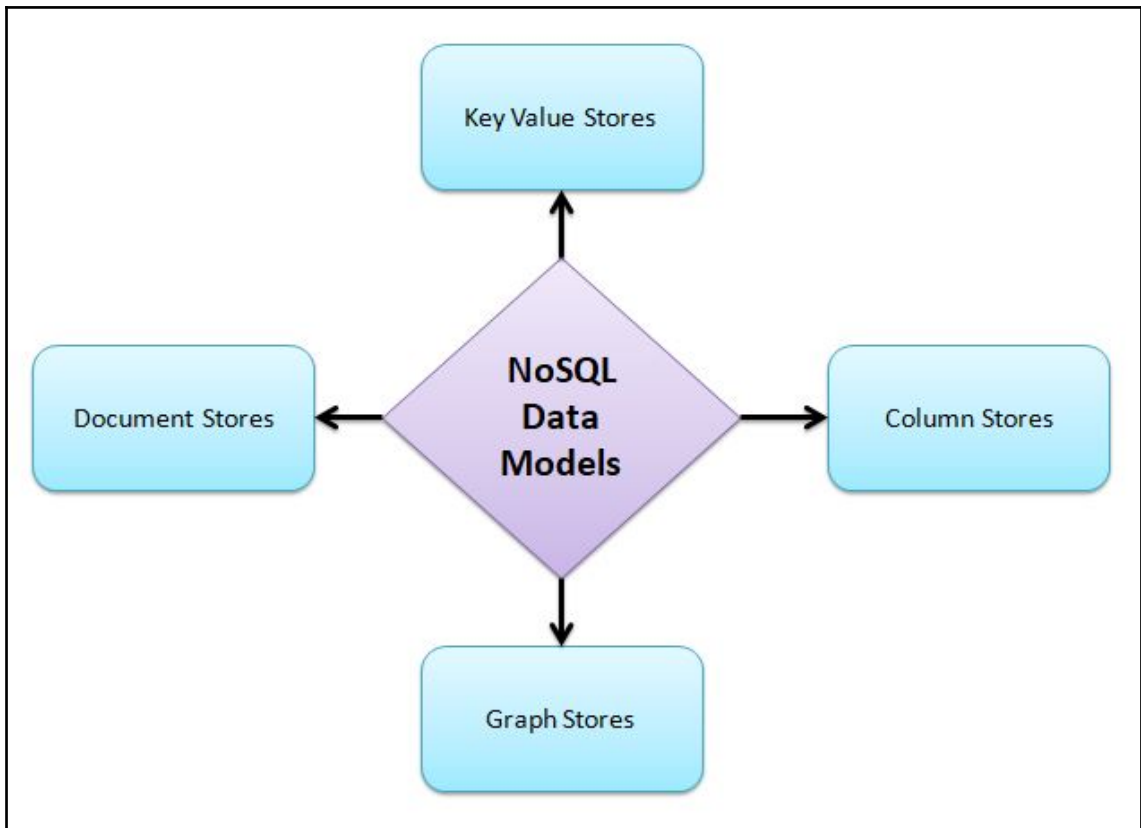






Chapter 4: NoSQL Database





```
hadoopadmin@hadoopadmin-VirtualBox: ~  
hadoopadmin@hadoopadmin-VirtualBox:~$ nodetool status  
Datacenter: datacenter1  
=====  
Status=Up/Down  
|| State=Normal/Leaving/Joining/Moving  
-- Address      Load        Tokens      Owns (effective)  Host ID                               Rack  
UN 127.0.0.1    103.67 KiB   256         100.0%           bbaa674e-59d6-4a70-a93d-f6659fa618bc rack1  
hadoopadmin@hadoopadmin-VirtualBox:~$
```

```
hadoopadmin@hadoopadmin-VirtualBox: ~  
hadoopadmin@hadoopadmin-VirtualBox:~$ cqlsh  
Connected to Test Cluster at 127.0.0.1:9042.  
[cqlsh 5.0.1 | Cassandra 3.11.1 | CQL spec 3.4.4 | Native protocol v4]  
Use HELP for help.  
cqlsh>
```

```

hadoopadmin@hadoopadmin-VirtualBox: ~
cqlsh> help

Documented shell commands:
=====
CAPTURE  CLS          COPY  DESCRIBE  EXPAND  LOGIN  SERIAL  SOURCE  UNICODE
CLEAR    CONSISTENCY  DESC  EXIT      HELP    PAGING  SHOW    TRACING

CQL help topics:
=====
AGGREGATES                CREATE_KEYSPACE          DROP_TRIGGER             TEXT
ALTER_KEYSPACE            CREATE_MATERIALIZED_VIEW DROP_TYPE                TIME
ALTER_MATERIALIZED_VIEW  CREATE_ROLE              DROP_USER                TIMESTAMP
ALTER_TABLE              CREATE_TABLE             FUNCTIONS                TRUNCATE
ALTER_TYPE               CREATE_TRIGGER           GRANT                    TYPES
ALTER_USER              CREATE_TYPE              INSERT                   UPDATE
APPLY                   CREATE_USER             INSERT_JSON              USE
ASCII                   DATE                    INT                      UUID
BATCH                   DELETE                  JSON
BEGIN                  DROP_AGGREGATE          KEYWORDS
BLOB                   DROP_COLUMNFAMILY      LIST_PERMISSIONS
BOOLEAN               DROP_FUNCTION          LIST_ROLES
COUNTER              DROP_INDEX             LIST_USERS
CREATE_AGGREGATE      DROP_KEYSPACE          PERMISSIONS
CREATE_COLUMNFAMILY  DROP_MATERIALIZED_VIEW REVOKE
CREATE_FUNCTION      DROP_ROLE              SELECT
CREATE_INDEX        DROP_TABLE             SELECT_JSON

cqlsh>

```

```

CQL
x +
file:///home/hadoopadmin/apache-cassandra-3.11.1/doc/cql3/CQL.html#cre...
CREATE KEYSPACE

Syntax:
<create-keyspace-stmt> ::= CREATE KEYSPACE (IF NOT EXISTS)? <identifier> WITH <properties>

Sample:
CREATE KEYSPACE Excelsior
  WITH replication = {'class': 'SimpleStrategy', 'replication_factor' : 3};

CREATE KEYSPACE Excalibur
  WITH replication = {'class': 'NetworkTopologyStrategy', 'DC1' : 1, 'DC2' : 3}
  AND durable_writes = false;

The CREATE KEYSPACE statement creates a new top-level keyspace. A keyspace is a namespace that defines a replication strategy and some options for a set of tables. Valid keyspace names are identifiers composed exclusively of alphanumerical characters and whose length is lesser or equal to 32. Note that as identifiers, keyspace names are case insensitive: use a quoted identifier for case sensitive keyspace names.

```

```
hadoopadmin@hadoopadmin-VirtualBox: ~
cqlsh> select * from system_schema.keyspaces;
```

keyspace_name	durable_writes	replication
system_auth	True	{'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication_factor': '1'}
system_schema	True	{'class': 'org.apache.cassandra.locator.LocalStrategy'}
system_distributed	True	{'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication_factor': '3'}
system	True	{'class': 'org.apache.cassandra.locator.LocalStrategy'}
testkeyspace	True	{'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication_factor': '3'}
system_traces	True	{'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication_factor': '2'}

```
(6 rows)
cqlsh>
```

```
hadoopadmin@hadoopadmin-VirtualBox: ~
cqlsh> Alter Keyspace testKeyspace With replication = {'class': 'SimpleStrategy', 'replication_factor': 2};
cqlsh> select * from system_schema.keyspaces;
```

keyspace_name	durable_writes	replication
system_auth	True	{'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication_factor': '1'}
system_schema	True	{'class': 'org.apache.cassandra.locator.LocalStrategy'}
system_distributed	True	{'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication_factor': '3'}
system	True	{'class': 'org.apache.cassandra.locator.LocalStrategy'}
testkeyspace	True	{'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication_factor': '2'}
system_traces	True	{'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication_factor': '2'}

```
(6 rows)
cqlsh>
```

```
hadoopadmin@hadoopadmin-VirtualBox: ~
cqlsh> drop keyspace testkeyspace;
cqlsh> select * from system_schema.keyspaces;
```

keyspace_name	durable_writes	replication
system_auth	True	{'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication_factor': '1'}
system_schema	True	{'class': 'org.apache.cassandra.locator.LocalStrategy'}
system_distributed	True	{'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication_factor': '3'}
system	True	{'class': 'org.apache.cassandra.locator.LocalStrategy'}
system_traces	True	{'class': 'org.apache.cassandra.locator.SimpleStrategy', 'replication_factor': '2'}

```
(5 rows)
cqlsh>
```

```
hadoopadmin@hadoopadmin-VirtualBox: ~
cqlsh> create keyspace StudentPortal with replication = {'class': 'SimpleStrategy', 'replication_factor': 2};
cqlsh> use StudentPortal;
cqlsh:studentportal> Create Table StudentInfo (
... std_id int PRIMARY KEY,
... std_name varchar,
... std_address text,
... std_phone varchar);
cqlsh:studentportal> describe tables;

studentinfo
cqlsh:studentportal>
```

```
hadoopadmin@hadoopadmin-VirtualBox: ~
cqlsh:studentportal> Alter table studentinfo
... add class_name text;
cqlsh:studentportal> describe table studentinfo;

CREATE TABLE studentportal.studentinfo (
  std_id int PRIMARY KEY,
  class_name text,
  std_address text,
  std_name text,
  std_phone text
) WITH bloom_filter_fp_chance = 0.01
AND caching = {'keys': 'ALL', 'rows per partition': 'NONE'}
```

```
hadoopadmin@hadoopadmin-VirtualBox: ~
cqlsh:studentportal> Insert into studentinfo
... (std_id, std_name, std_address, std_phone)
... values
... (1, 'John Doe', 'On Earth', '1-234-567890');
cqlsh:studentportal> select * from studentinfo;

std_id | std_address | std_name | std_phone
-----+-----+-----+-----
1 | On Earth | John Doe | 1-234-567890

(1 rows)
cqlsh:studentportal> █
```



```

hadoopadmin@hadoopadmin-VirtualBox: ~
cqlsh:studentportal> update studentinfo
... SET std_address = 'On Moon',
... std_phone = '0-987-654321'
... where std_id = 1;
cqlsh:studentportal> select * from studentinfo;

std_id | std_address | std_name | std_phone
-----+-----+-----+-----
      1 |    On Moon | John Doe | 0-987-654321

(1 rows)
cqlsh:studentportal>

```

```

hadoopadmin@hadoopadmin-VirtualBox: ~
cqlsh:studentportal> delete from studentinfo
... where std_id = 1;
cqlsh:studentportal> select * from studentinfo;

std_id | std_address | std_name | std_phone
-----+-----+-----+-----

(0 rows)
cqlsh:studentportal> █

```

```

2017-11-26T21:13:44.427+0500 I FTDC [initandlisten] Initializing full-time diagnostic data capture with directory '/var/lib/mongodb/diagnostic.data'
2017-11-26T21:13:44.445+0500 I INDEX [initandlisten] build index on: admin.system.version properties: { v: 2, key: { version: 1 }, name: "incompatible_with_version_32", ns: "admin.system.version" }
2017-11-26T21:13:44.445+0500 I INDEX [initandlisten] building index using bulk method; build may temporarily use up to 500 megabytes of RAM
2017-11-26T21:13:44.446+0500 I INDEX [initandlisten] build index done. scanned 0 total records. 0 secs
2017-11-26T21:13:44.447+0500 I COMMAND [initandlisten] setting featureCompatibilityVersion to 3.4
2017-11-26T21:13:44.447+0500 I NETWORK [thread1] waiting for connections on port 27017
hadoopadmin@hadoopadmin-VirtualBox:~$

```

```

hadoopadmin@hadoopadmin-VirtualBox: ~
hadoopadmin@hadoopadmin-VirtualBox:~$ mongo
MongoDB shell version v3.4.10
connecting to: mongodb://127.0.0.1:27017
MongoDB server version: 3.4.10
Welcome to the MongoDB shell.
For interactive help, type "help".
For more comprehensive documentation, see
http://docs.mongodb.org/
Questions? Try the support group
http://groups.google.com/group/mongodb-user
Server has startup warnings:
2017-11-26T21:13:44.255+0500 I STORAGE [initandlisten]
2017-11-26T21:13:44.255+0500 I STORAGE [initandlisten] ** WARNING: Using the XFS filesystem is strongly recommended with the WiredTiger stor
age engine
2017-11-26T21:13:44.309+0500 I CONTROL [initandlisten] ** See http://dochub.mongodb.org/core/prodnotes-filesystem
2017-11-26T21:13:44.309+0500 I CONTROL [initandlisten] ** WARNING: Access control is not enabled for the database.
2017-11-26T21:13:44.309+0500 I CONTROL [initandlisten] ** Read and write access to data and configuration is unrestricted.
2017-11-26T21:13:44.309+0500 I CONTROL [initandlisten]
2017-11-26T21:13:44.309+0500 I CONTROL [initandlisten] ** WARNING: /sys/kernel/mm/transparent_hugepage/enabled is 'always'.
2017-11-26T21:13:44.309+0500 I CONTROL [initandlisten] ** We suggest setting it to 'never'
2017-11-26T21:13:44.309+0500 I CONTROL [initandlisten]
>

```

```

hadoopadmin@hadoopadmin-VirtualBox: ~
> db.help()
DB methods:
db.adminCommand(nameOrDocument) - switches to 'admin' db, and runs command [ just calls db.runCommand(...) ]
db.auth(username, password)
db.cloneDatabase(fromhost)
db.commandHelp(name) returns the help for the command
db.copyDatabase(fromdb, todb, fromhost)
db.createCollection(name, { size : ..., capped : ..., max : ... } )
db.createView(name, viewOn, [ { $operator: {...}}, ... ], { viewOptions } )
db.createUser(userDocument)
db.currentOp() displays currently executing operations in the db
db.dropDatabase()
db.eval() - deprecated
db.fsyncLock() flush data to disk and lock server for backups
db.fsyncUnlock() unlocks server following a db.fsyncLock()
db.getCollection(cname) same as db['cname'] or db.cname
db.getCollectionInfos([filter]) - returns a list that contains the names and options of the db's collections
db.getCollectionNames()
db.getLastErrorMessage() - just returns the err msg string
db.getLastErrorMessageObj() - return full status object
db.getLogComponents()
db.getMongo() get the server connection object
db.getMongo().setSlaveOk() allow queries on a replication slave server
db.getName()
db.getPrevError()
db.getProfilingLevel() - deprecated

```

```

hadoopadmin@hadoopadmin-VirtualBox: ~
> use schoolportal;
switched to db schoolportal
> db.dropDatabase();
{ "ok" : 1 }
>

```

```
hadoopadmin@hadoopadmin-VirtualBox: ~
> db.createCollection("collectionName");
{ "ok" : 1 }
> show collections;
collectionName
> db.collectionName.drop();
true
> show collections;
>
```

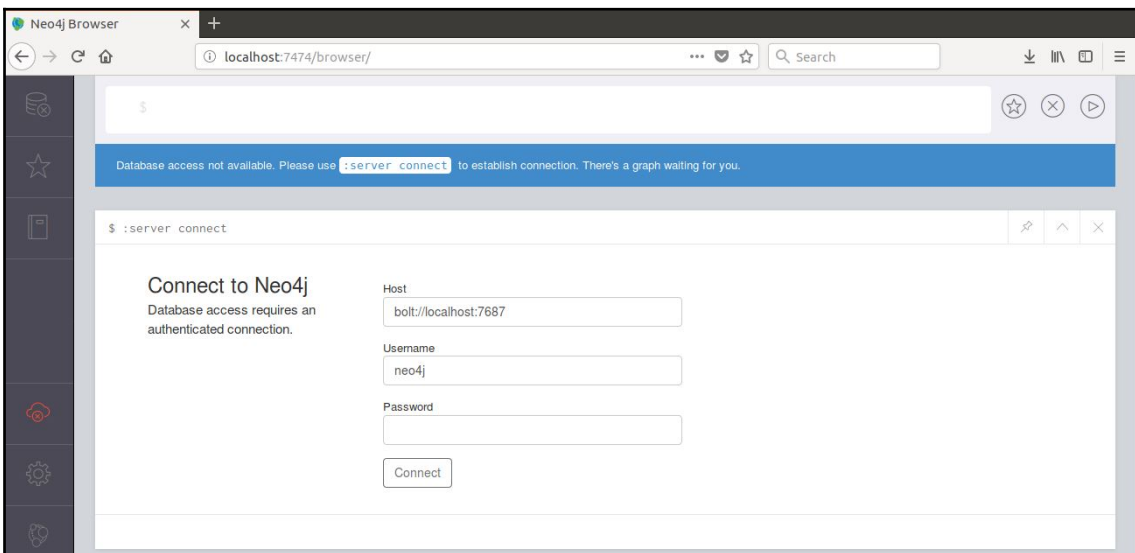
```
hadoopadmin@hadoopadmin-VirtualBox: ~
> db.employees.find()
{ "_id" : ObjectId("5a1e7169753935a62eeac82f"), "first_name" : "John", "last_name" : "Doe" }
{ "_id" : ObjectId("5a1e7979753935a62eeac830"), "first_name" : "Abc", "last_name" : "xyz" }
{ "_id" : ObjectId("5a1e7979753935a62eeac831"), "first_name" : "ijk", "last_name" : "xyz" }
>
```

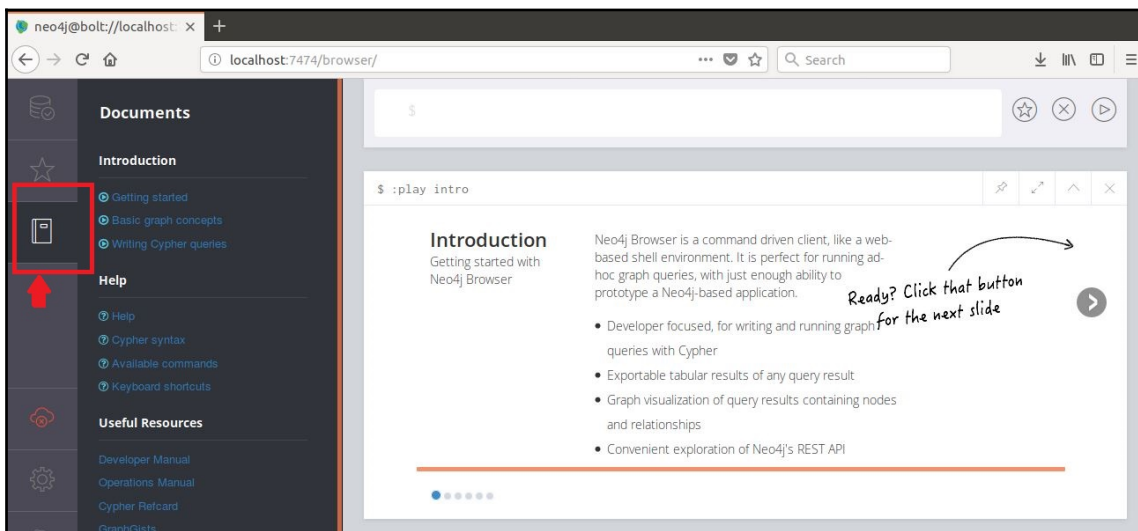
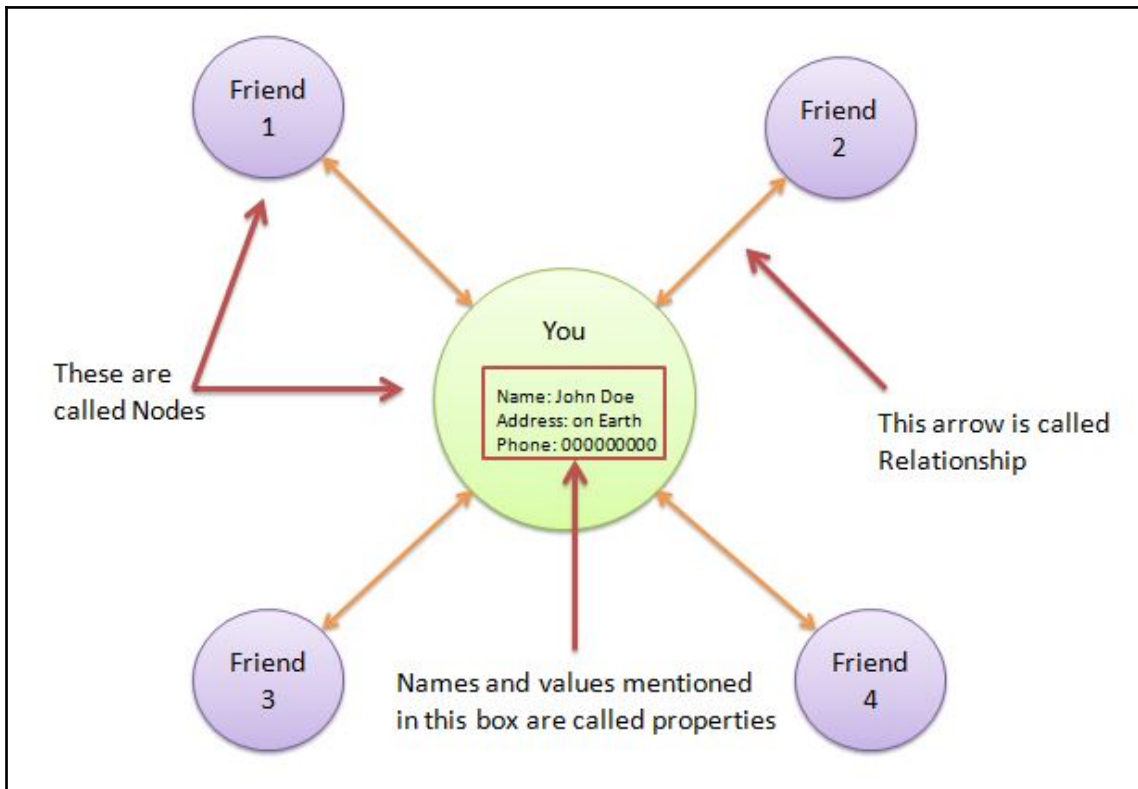
```
hadoopadmin@hadoopadmin-VirtualBox: ~
> db.employees.update({first_name:"Abc"}, {$set: {phone:"123456789"}})
WriteResult({ "nMatched" : 1, "nUpserted" : 0, "nModified" : 0 })
> db.employees.find()
{ "_id" : ObjectId("5a1e7169753935a62eeac82f"), "first_name" : "John", "last_name" : "Doe" }
{ "_id" : ObjectId("5a1e7979753935a62eeac830"), "first_name" : "Abc", "last_name" : "xyz", "phone" : "123456789" }
{ "_id" : ObjectId("5a1e7979753935a62eeac831"), "first_name" : "ijk", "last_name" : "xyz" }
>
```

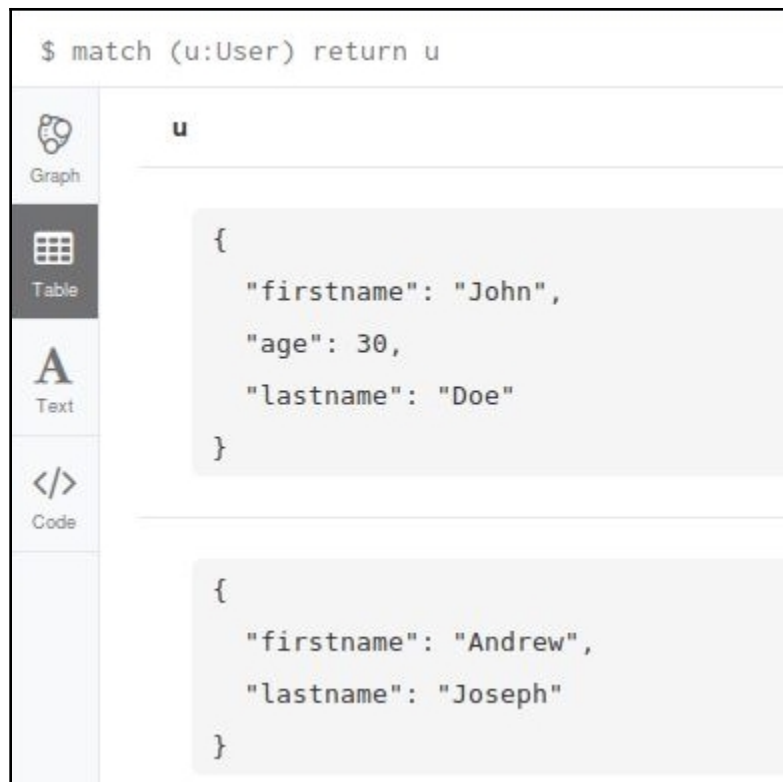
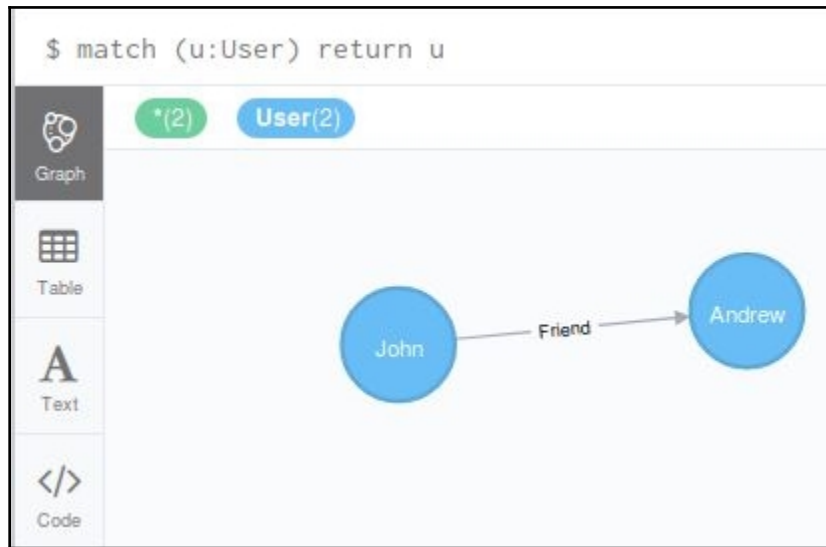
```
hadoopadmin@hadoopadmin-VirtualBox: ~
> db.employees.deleteOne({first_name: "Abc"})
{ "acknowledged" : true, "deletedCount" : 1 }
> db.employees.find()
{ "_id" : ObjectId("5a1e7169753935a62eeac82f"), "first_name" : "John", "last_name" : "Doe" }
{ "_id" : ObjectId("5a1e7979753935a62eeac831"), "first_name" : "ijk", "last_name" : "xyz" }
>
```

```
hadoopadmin@hadoopadmin-VirtualBox: ~
neo4j.service - Neo4j Graph Database
   Loaded: loaded (/lib/systemd/system/neo4j.service; disabled; vendor preset: enabled)
   Active: active (running) since Wed 2017-11-29 19:28:00 PKT; 2min 35s ago
     Main PID: 16686 (java)
       Tasks: 31 (limit: 4915)
    CGroup: /system.slice/neo4j.service
            └─16686 /usr/bin/java -cp /var/lib/neo4j/plugins:/etc/neo4j:/usr/share/neo4j/lib/*:/var/lib/neo4j/plugins/* -server -XX:+UseG1GC

Nov 29 19:28:00 hadoopadmin-VirtualBox neo4j[16686]: data: /var/lib/neo4j/data
Nov 29 19:28:00 hadoopadmin-VirtualBox neo4j[16686]: certificates: /var/lib/neo4j/certificates
Nov 29 19:28:00 hadoopadmin-VirtualBox neo4j[16686]: run: /var/run/neo4j
Nov 29 19:28:00 hadoopadmin-VirtualBox neo4j[16686]: Starting Neo4j.
Nov 29 19:28:02 hadoopadmin-VirtualBox neo4j[16686]: 2017-11-29 14:28:02.487+0000 WARN dbms.directories.certificates is deprecated.
Nov 29 19:28:02 hadoopadmin-VirtualBox neo4j[16686]: 2017-11-29 14:28:02.604+0000 INFO ===== Neo4j 3.3.0 =====
Nov 29 19:28:02 hadoopadmin-VirtualBox neo4j[16686]: 2017-11-29 14:28:02.764+0000 INFO Starting..
Nov 29 19:28:05 hadoopadmin-VirtualBox neo4j[16686]: 2017-11-29 14:28:05.843+0000 INFO Bolt enabled on 127.0.0.1:7687.
Nov 29 19:28:12 hadoopadmin-VirtualBox neo4j[16686]: 2017-11-29 14:28:12.384+0000 INFO Started.
Nov 29 19:28:14 hadoopadmin-VirtualBox neo4j[16686]: 2017-11-29 14:28:14.532+0000 INFO Remote interface available at http://localhost:7474/
```



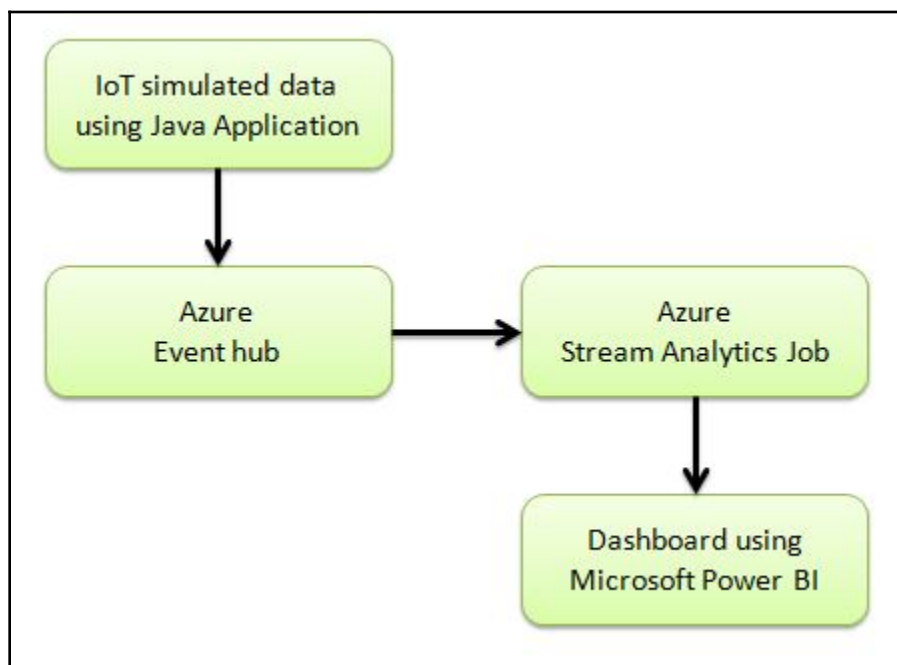




```
$ match (u:User) return u.firstname, u.lastname, u.age
```

u.firstname	u.lastname	u.age
"John"	"Doe"	30
"Andrew"	"Joseph"	null

Chapter 5: Off-the-Shelf Tools



https://signup.azure.com/signup?offer=ms-azr-0044p&appid=102&redirectURL=https:%2F%2Fazure.microsoft.com%2Fen-us%2Fget-started%2Fwelcome

Microsoft Azure fahadalkhtar2018@hotmail.com Sign out

Azure free account sign up

Start with a \$200 credit for 30 days, and keep going for free

1 About you

* Country/Region ⓘ
Pakistan

* First name

* Last name

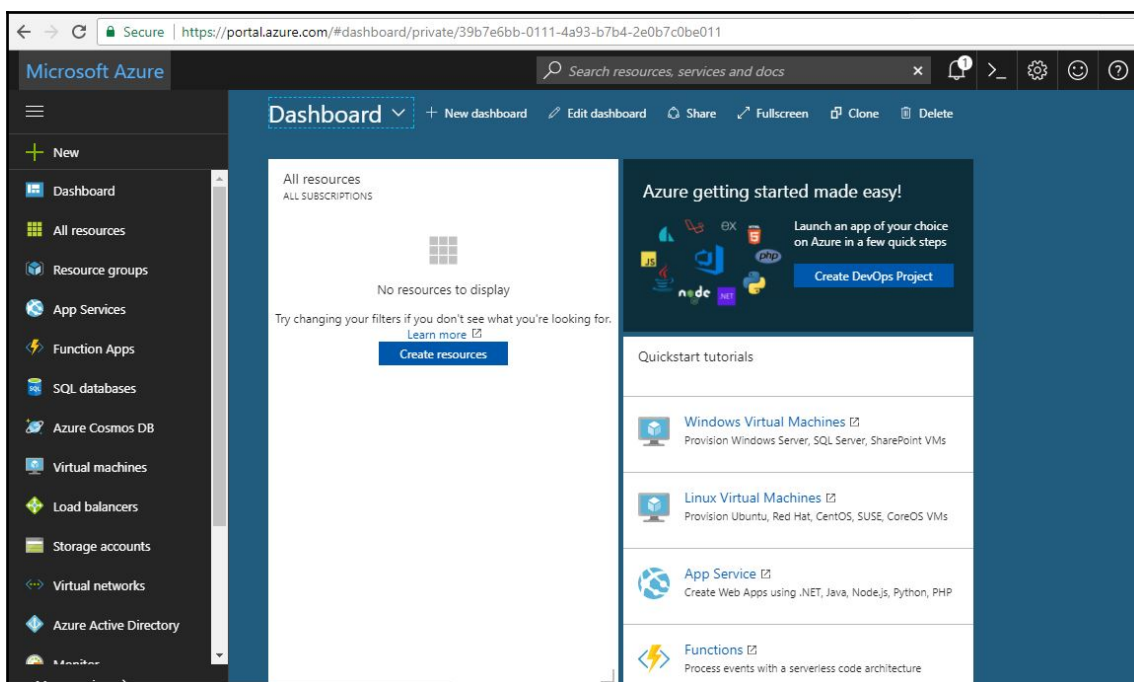
* Email address ⓘ

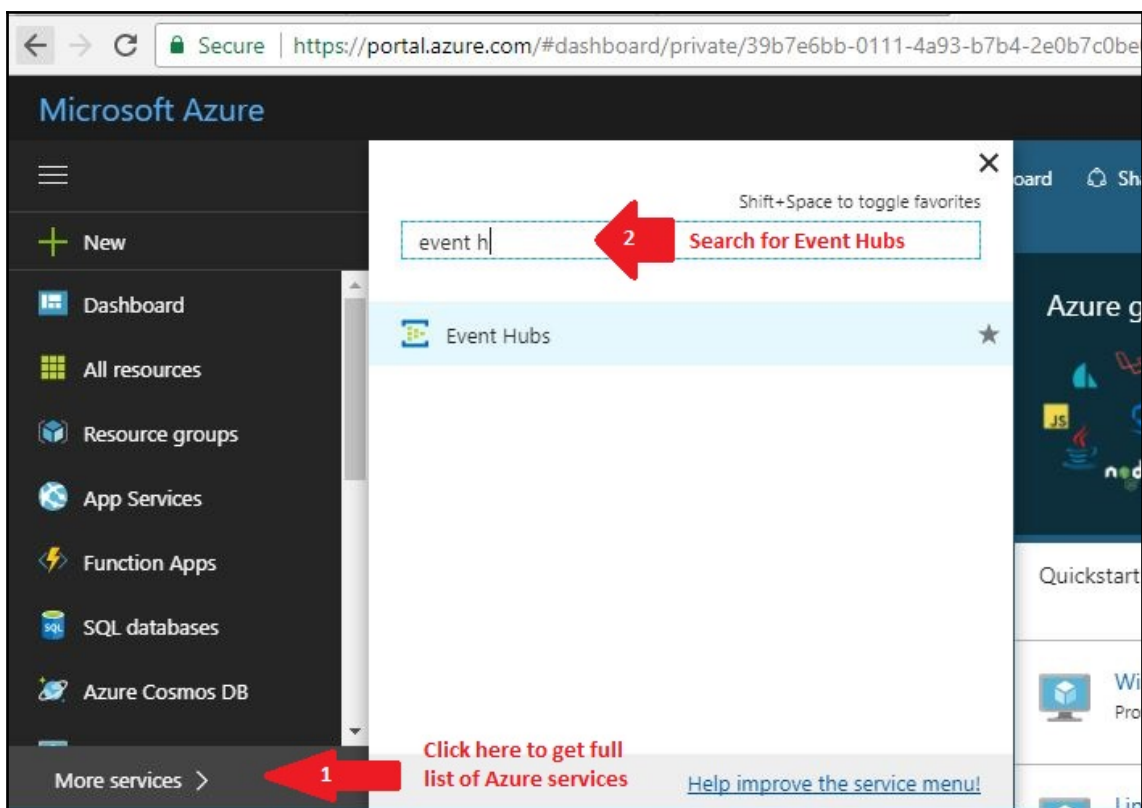
* Phone
Example: 342-1234567

Next

2 Identity verification by phone

3 Identity verification by card





Event Hubs

fahadakhtarhotmail (Default Directory)

+ Add

Assign Tags

More

Filter by name...

0 items

NAME

No Event Hubs to display

Try changing your filters if you don't see what you're looking for.

Create Event Hubs

Create namespace

Event Hubs

* Name

StreamApp

servicebus.windows.net

* Pricing tier

Standard

* Subscription

Free Trial

* Resource group

Create new

Use existing

streamapp_Resources

* Location

Australia East

Throughput Units

1

☐ Enable auto-inflate?

☐ Pin to dashboard

Create

Automation options

+ Add
Assign Tags
Columns
Refresh

Subscriptions: Free Trial

1 items

☐	NAME ↑	TYPE ↑	RESOURCE GROUP ↑	LOCATION ↑	SUBSCRIPTION
☐	StreamApp	Event Hub	streamapp_Resources	Australia East	Free Trial

StreamApp Event Hub

Click here to create an Event Hub

Search (Ctrl+/)

+ Event Hub Delete

Overview

Access control (IAM)

Tags

Diagnose and solve problems

SETTINGS

Shared access policies

Scale

Metrics (preview)

Properties

Locks

Automation script

Essentials

Resource group (change)
streamapp_Resources

Status
Active

Location
Australia East

Subscription name (change)
Free Trial

Subscription ID
45dd8a45-5e82-4b6e-be52-850ceccb0040

Connection Strings
Connection Strings

Throughput Units
1

Click here to get connection related details

NAMESPACE CONTENTS 0 EVENT-HUBS

PRICING TIER STANDARD

NAMESPACE STATUS ACTIVE

THROUGHPUT UNITS 1 UNIT

NAME	STATUS	MESSAGE RETENTION	PARTITION COUNT
StreamAppEvents	Active	1	2

Event Hubs

Create Event Hub

StreamApp

*

Name

StreamAppEvents

✓

Partition Count ⓘ

2

Message Retention ⓘ

1

Capture

On

Off

Time window (minutes)

5

Size window (MB)

300

Capture Provider

Azure Storage

▼

Create

SAS Policies
Event Hubs

+ Add

Search to filter items...

POLICY	CLAIMS
RootManageSharedAccessKey	Manage, Send, Listen

SAS Policy: RootManageSharedAccessKey X

Save Discard Delete More

☒ Manage

☒ Send

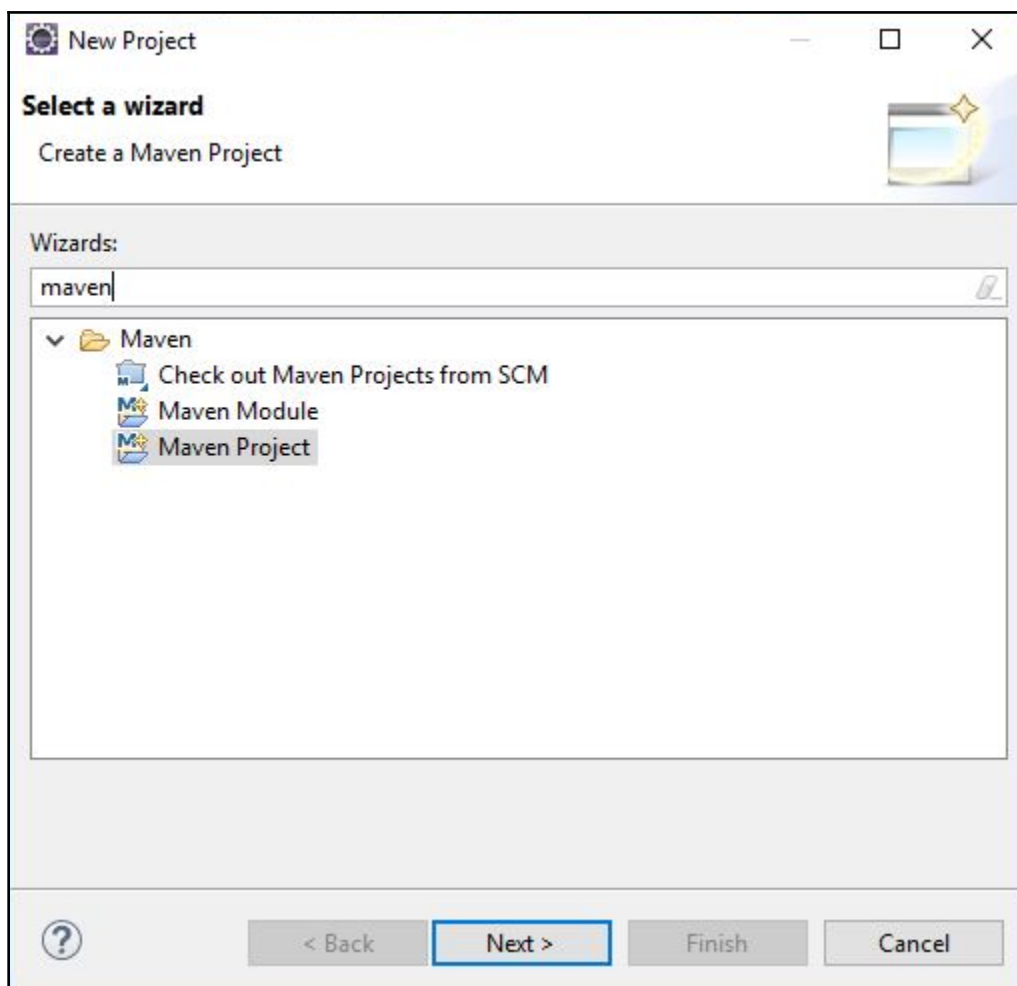
☒ Listen

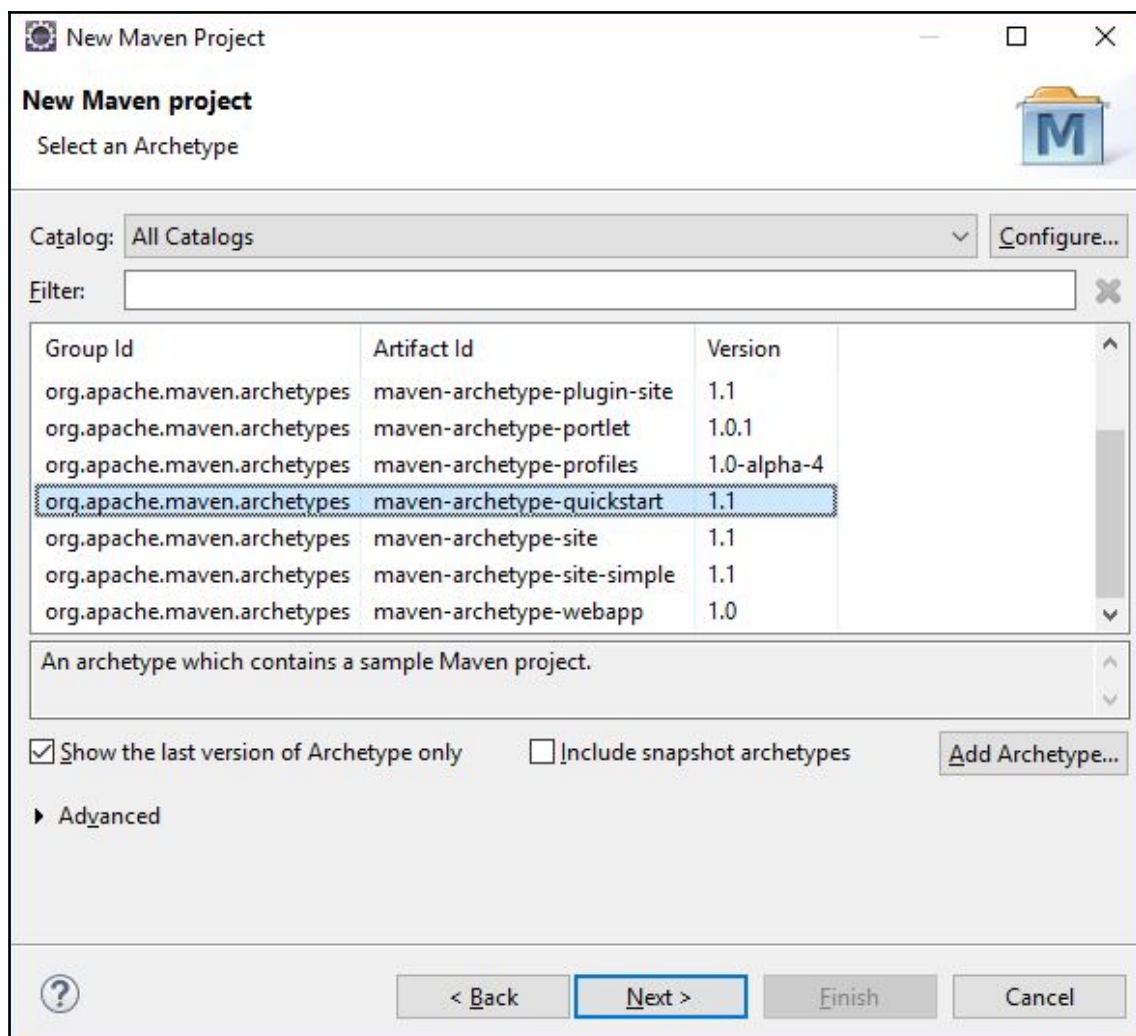
Primary key
78I2pEzXvq//lkEuLHjpbTQhO/XXtBdlOqKis7vuzi4=

Secondary key
9v2KfcYsS27x6DU4XRSmlhuQcEUc9vMgXxAllosPDkg=

Connection string—primary key
Endpoint=sbi://streamapp.servicebus.windows.net/S...

Connection string—secondary key
Endpoint=sbi://streamapp.servicebus.windows.net/S...





New Maven Project

New Maven project

Specify Archetype parameters

Group Id:

Artifact Id:

Version:

Package:

Properties available from archetype:

Name	Value

[Add...](#)

[Remove](#)

[Advanced](#)

[?< Back](#) [Next >](#) **Finish** [Cancel](#)

```

Problems @ Javadoc Declaration Console
SendStream [Java Application] C:\Program Files\Java\jre1.8.0_121\bin\javaw.exe (Dec 8, 2017, 12:19:48 PM)
Message Sent --> [{timestamp: 2017-12-08 12:19:58, temperature:48, humidity:54}]
Message Sent --> [{timestamp: 2017-12-08 12:20:05, temperature:20, humidity:29}]
Message Sent --> [{timestamp: 2017-12-08 12:20:17, temperature:26, humidity:20}]
Message Sent --> [{timestamp: 2017-12-08 12:20:30, temperature:51, humidity:36}]
Message Sent --> [{timestamp: 2017-12-08 12:20:52, temperature:50, humidity:51}]
Message Sent --> [{timestamp: 2017-12-08 12:21:01, temperature:33, humidity:36}]
Message Sent --> [{timestamp: 2017-12-08 12:21:17, temperature:44, humidity:34}]
Message Sent --> [{timestamp: 2017-12-08 12:21:24, temperature:20, humidity:56}]
Message Sent --> [{timestamp: 2017-12-08 12:21:30, temperature:44, humidity:23}]
Message Sent --> [{timestamp: 2017-12-08 12:21:38, temperature:25, humidity:32}]

```

Stream Analytics jobs

fahadakhtarhotmail (Default Directory)

+

 Add

Columns

More

0 items

NAME ↑↓



No Stream Analytics jobs to display

Try changing your filters if you don't see what you're looking for.

Create Stream Analytics jobs

New Stream Analytics Job

Job name

Subscription

Free Trial

Resource group

☐ Create new
 ☒ Use existing

streamapp_Resources

Location

Australia East

Hosting environment

Cloud

Edge

☒ Pin to dashboard

Create

Automation options

The screenshot shows the IoTData Stream Analytics job configuration interface. The left sidebar contains navigation options: Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, SETTINGS (Locks), and JOB TOPOLOGY (Inputs, Functions, Query, Outputs). The main panel displays job details and the Job Topology.

Job Details:

- Resource group (change): streamapp_Resources
- Status: Created
- Location: Australia East
- Subscription name (change): Free Trial
- Subscription ID: 45dd8a45-5e82-4b6e-be52-850ceccb0040
- Send feedback: UserVoice
- Created: Friday, December 8, 2017, 4:23:28 PM
- Started: -
- Last output: -
- Hosting environment: Cloud

Job Topology:

- Inputs:** 0. No results.
- Query:** <>. Second step to apply logic and get desired result.
- Outputs:** 0. No results. Third step to use the analyzed data in different output.

Three red arrows indicate the workflow steps:

1. First Step to define Input (points to the Inputs section).
2. Second step to apply logic and get desired result (points to the Query section).
3. Third step to use the analyzed data in different output (points to the Outputs section).

Inputs
IoTData

+ Add

NAME	SOURCE TYPE	SOURCE
Empty		

New input

* Input alias
streamappdata ✓

* Source Type ⓘ
Data stream ▾

* Source ⓘ
Event hub ▾

* Import option
Select event hub from your subscription... ▾

Service bus namespace
StreamApp ▾

Event hub name
streamappevents ▾

Event hub policy name
RootManageSharedAccessKey ▾

Event hub consumer group ⓘ

* Event serialization format ⓘ

Create

The screenshot shows the IoTData Query interface. At the top, there are buttons for Save, Discard, and Test. A red arrow labeled '3' points to the Test button with the text 'Click here to Test your query'. On the left, under 'Inputs (1)', 'streamappdata' is selected. A red arrow labeled '2' points to it with the text 'Click here to sample your data'. The main query window, labeled '1' with a red arrow and 'Query window to write your query', contains the following SQL query:

```
1 SELECT
2   timestamp, temperature, humidity
3 FROM
4   streamappdata
```

Below the query window, there is a note: 'Your query could be put in logs that are in a potentially different geography. Missing some language constructs? Let us know! (Powered by UserVoice - Privacy Policy)'. At the bottom, the 'output' section, labeled '4' with a red arrow and 'It displays the result of your query', shows the generated results:

Generated the Following:

- output with 181 rows.

Download results

TIMESTAMP	TEMPERATURE	HUMIDITY
"2017-12-08 04:33:53"	40	45

The screenshot shows the IoTData Outputs interface. On the left, there is a table with columns 'NAME' and 'SINK'. The table is empty, with 'Empty' in the NAME column. On the right, the 'New output' dialog is open. It has two sections: 'Output alias' and 'Sink'. The 'Output alias' section has a text input field with 'streamdataset' and a green checkmark. The 'Sink' section has a dropdown menu with 'Power BI' selected. Below these sections, there is a section titled 'Authorize Connection' with the text 'You'll need to authorize with Power BI to configure your output settings.' and a blue 'Authorize' button. At the bottom, there is a link 'Sign Up' with the text 'Don't have a Microsoft Power BI account yet?'.

Outputs

IoTData

+ Add

NAME	SINK
Empty	

New output

* Output alias
streamdataset ✓

* Sink
Power BI ✓

Group Workspace
IoTStreamingData ✓

* Dataset Name
IoTData ✓

⚠ If the dataset or table already exists in your Microsoft Power BI subscription, it will be overwritten.

* Table Name
tempAndHumData ✓

Create

IoTData

Stream Analytics job

Search (Ctrl+/)

- Overview
- Activity log
- Access control (IAM)
- Tags
- Diagnose and solve problems

SETTINGS

- Locks

Start Stop Delete

Created

Essentials

Resource group (change)
streamapp_Resources

Status
Created

Location
Australia East

Subscription name (change)
Free Trial

Subscription ID
45dd8a45-5e82-4b6e-be52-850ceccb0040

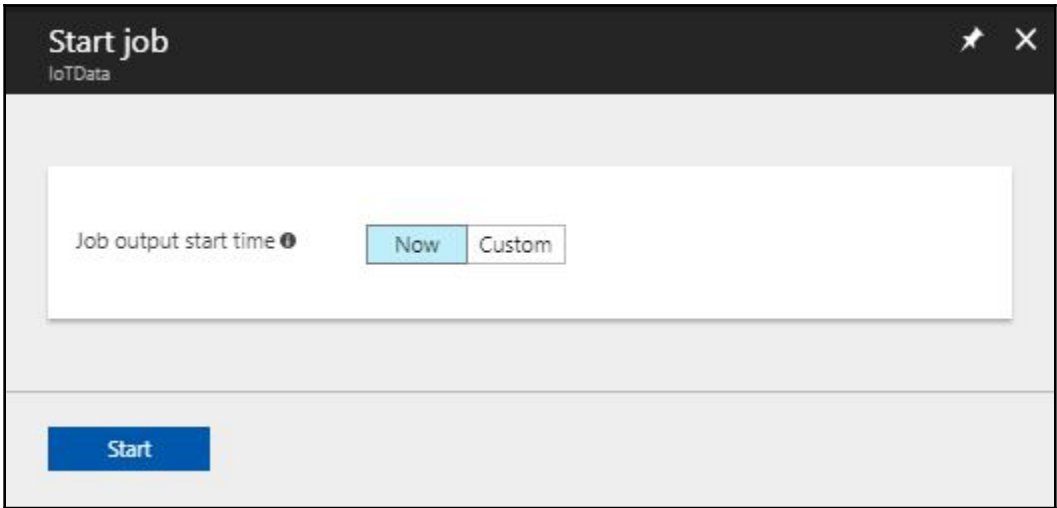
Send feedback
UserVoice

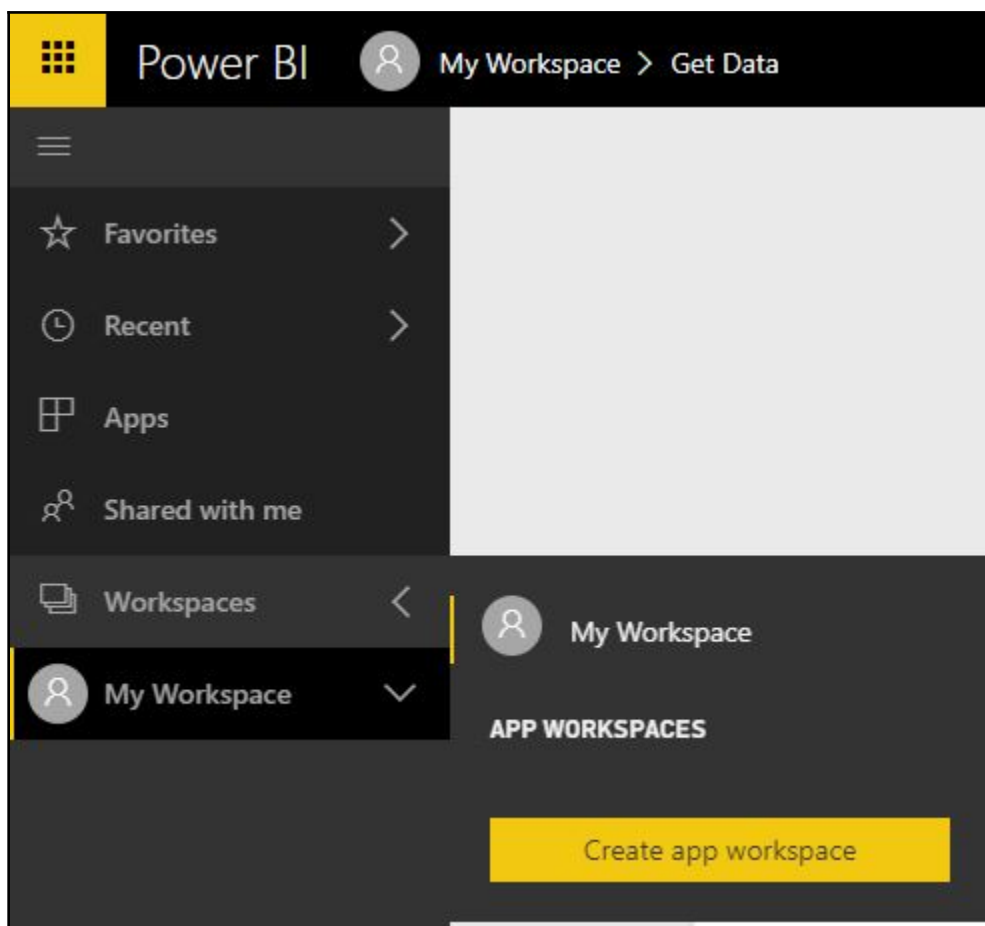
Created
Friday, December 8, 2017, 4:23:28 PM

Started
-

Last output
-

Hosting environment
Cloud





Create an app workspace

Name your workspace

Workspace ID

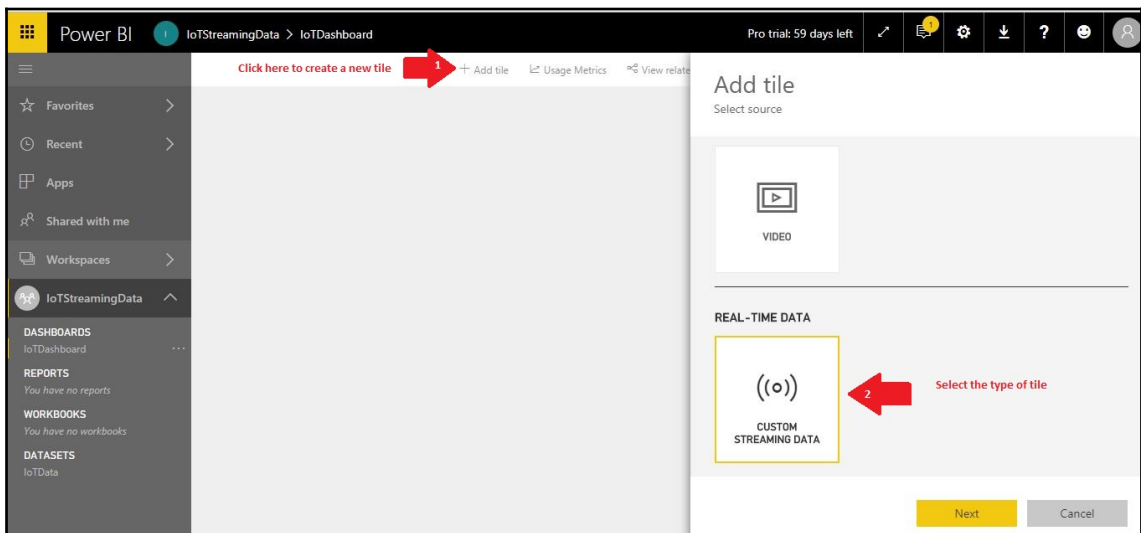
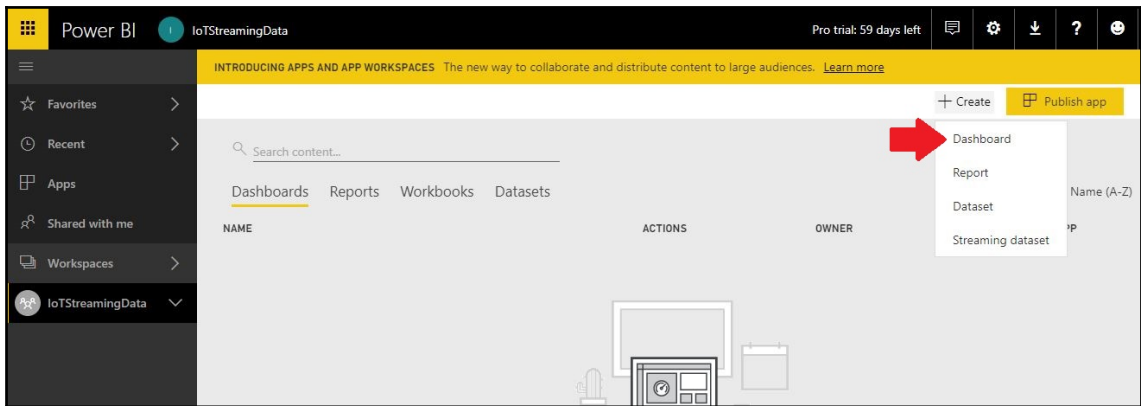
 Available

Private - Only approved members can see what's inside ▼

Members can only view Power BI content ▼

Add workspace members

Save Cancel



Add a custom streaming data tile

Choose a streaming dataset

 Add streaming dataset

YOUR DATASETS

IoTData

[Manage datasets](#)

Back

Next

Cancel

Visualization Type

Card ▼

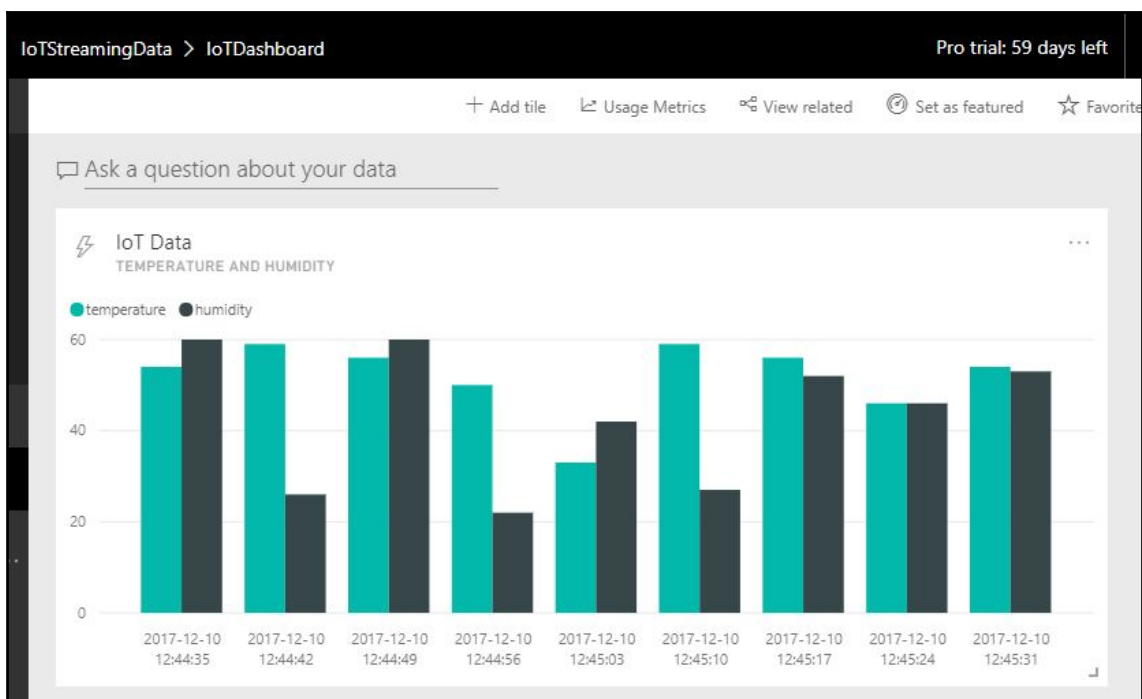
Card

Line chart

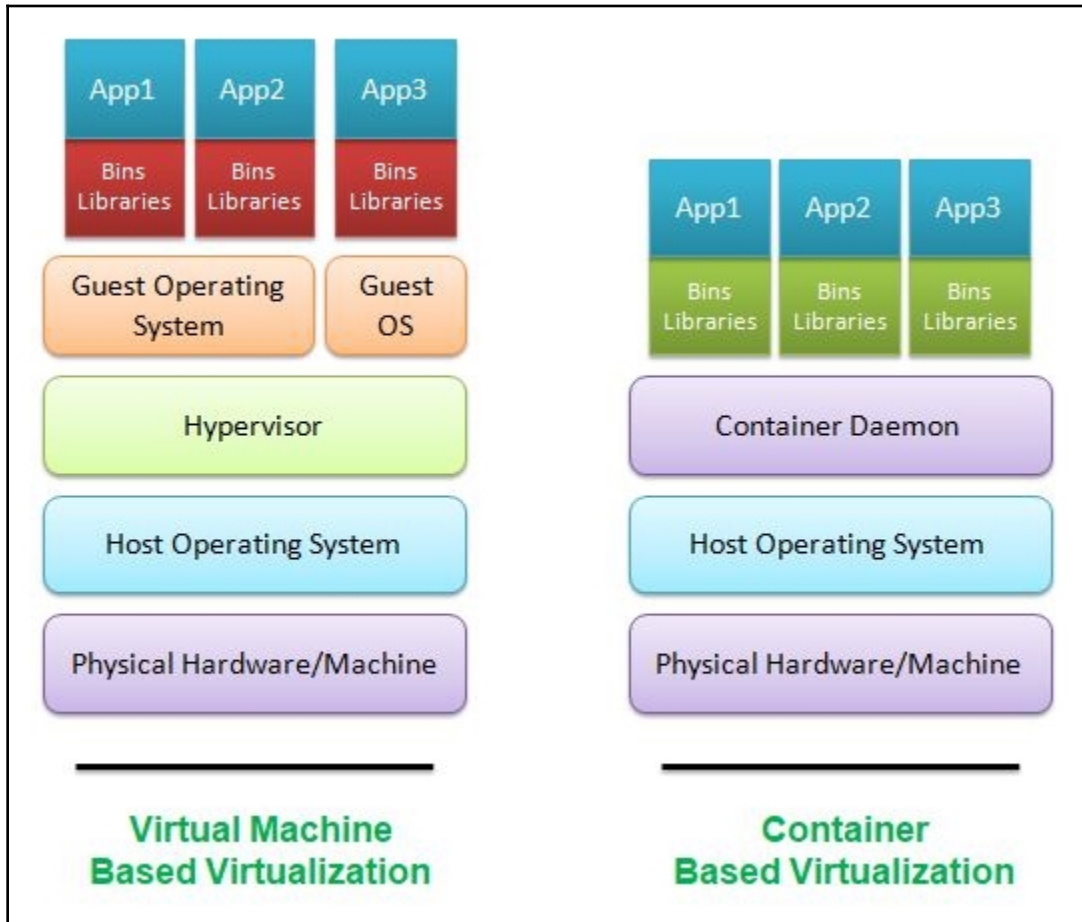
Clustered bar chart

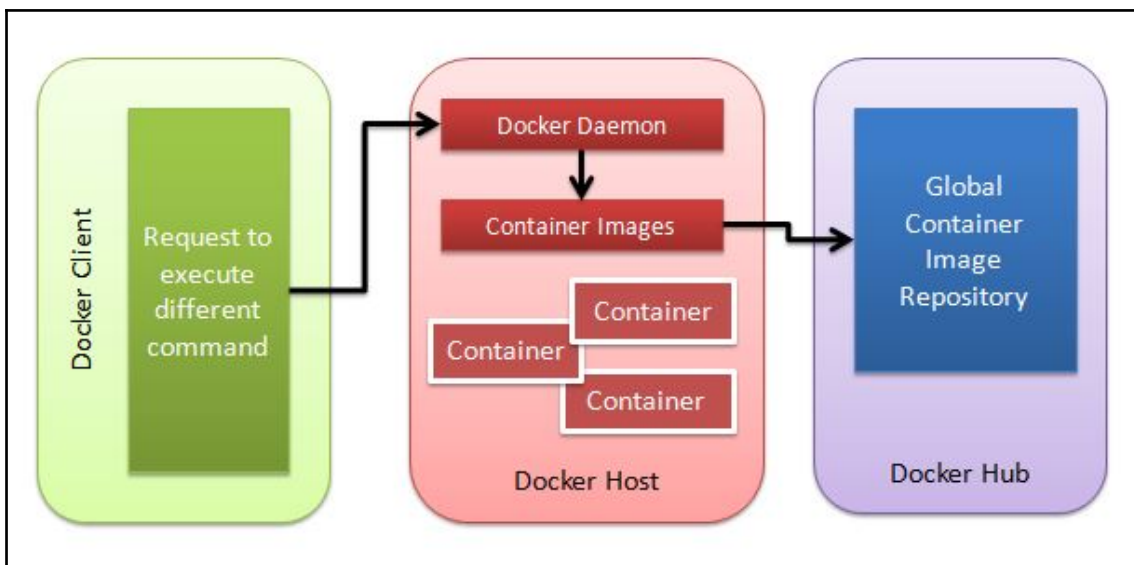
Clustered column chart

Gauge



Chapter 6: Containerization





```
hadoopadmin@hadoopadmin-VirtualBox: ~  
hadoopadmin@hadoopadmin-VirtualBox:~$ sudo docker run hello-world  
Unable to find image 'hello-world:latest' locally  
latest: Pulling from library/hello-world  
ca4f61b1923c: Pull complete  
Digest: sha256:be0cd392e45be79ffefffa6b05338b98ebb16c87b255f48e297ec7f98e123905c  
Status: Downloaded newer image for hello-world:latest  
  
Hello from Docker!  
This message shows that your installation appears to be working correctly.  
  
To generate this message, Docker took the following steps:  
1. The Docker client contacted the Docker daemon.  
2. The Docker daemon pulled the "hello-world" image from the Docker Hub.  
   (amd64)  
3. The Docker daemon created a new container from that image which runs the  
   executable that produces the output you are currently reading.  
4. The Docker daemon streamed that output to the Docker client, which sent it  
   to your terminal.  
  
To try something more ambitious, you can run an Ubuntu container with:  
$ docker run -it ubuntu bash  
  
Share images, automate workflows, and more with a free Docker ID:  
https://cloud.docker.com/  
  
For more examples and ideas, visit:  
https://docs.docker.com/engine/userguide/  
  
hadoopadmin@hadoopadmin-VirtualBox:~$
```

```

hadoopadmin@hadoopadmin-VirtualBox: ~
hadoopadmin@hadoopadmin-VirtualBox:~$ sudo docker info
Containers: 1
  Running: 0
  Paused: 0
  Stopped: 1
Images: 1
Server Version: 17.09.1-ce
Storage Driver: overlay2
  Backing Filesystem: extfs
  Supports d_type: true
  Native Overlay Diff: true
Logging Driver: json-file
Cgroup Driver: cgroupfs
Plugins:
  Volume: local
  Network: bridge host macvlan null overlay
  Log: awslogs fluentd gcplogs gelf journald json-file logent
Swarm: inactive
Runtimes: runc
Default Runtime: runc
Init Binaries: docker-init

```

```

hadoopadmin@hadoopadmin-VirtualBox: ~
hadoopadmin@hadoopadmin-VirtualBox:~$ sudo docker image ls
REPOSITORY          TAG                 IMAGE ID            CREATED             SIZE
hello-world          latest             f2a91732366c       3 weeks ago        1.85kB
hadoopadmin@hadoopadmin-VirtualBox:~$

```

```

hadoopadmin@hadoopadmin-VirtualBox: ~
hadoopadmin@hadoopadmin-VirtualBox:~$ sudo docker ps
CONTAINER ID        IMAGE               COMMAND             CREATED             STATUS              PORTS              NAMES
hadoopadmin@hadoopadmin-VirtualBox:~$ sudo docker container ls
CONTAINER ID        IMAGE               COMMAND             CREATED             STATUS              PORTS              NAMES
hadoopadmin@hadoopadmin-VirtualBox:~$

```

```

hadoopadmin@hadoopadmin-VirtualBox: ~
hadoopadmin@hadoopadmin-VirtualBox:~$ sudo docker ps -a
CONTAINER ID        IMAGE               COMMAND             CREATED             STATUS              PORTS              NAMES
02fa319e446c       nginx              "nginx -g 'daemon ..." 17 minutes ago      Exited (0) 2 minutes ago      zen_hawkin
g
da8eff7d5db5       hello-world        "/hello"           24 hours ago        Exited (0) 24 hours ago        adoring_sa
mnet
hadoopadmin@hadoopadmin-VirtualBox:~$

```

```
hadoopadmin@ubuntu: ~/dockerImage
hadoopadmin@ubuntu:~/dockerImage$ pwd
/home/hadoopadmin/dockerImage
hadoopadmin@ubuntu:~/dockerImage$ ls
Dockerfile index.html
hadoopadmin@ubuntu:~/dockerImage$
```

```
hadoopadmin@ubuntu: ~/dockerImage
hadoopadmin@ubuntu:~/dockerImage$ docker build -t webserver .
Sending build context to Docker daemon 3.072kB
Step 1/2 : FROM nginx:alpine
alpine: Pulling from library/nginx
128191993b8a: Pull complete
655cae3ea06e: Pull complete
dbc72c3fd216: Pull complete
f391a4589e37: Pull complete
Digest: sha256:34aa80bb22c79235d466ccbbfa3659ff81510ed21eddb1543c6847292010c4d
Status: Downloaded newer image for nginx:alpine
--> 22f5726c6dc0
Step 2/2 : COPY . /usr/share/nginx/html
--> ee6e31130682
Successfully built ee6e31130682
Successfully tagged webserver:latest
hadoopadmin@ubuntu:~/dockerImage$
```

```
hadoopadmin@ubuntu: ~/dockerImage
hadoopadmin@ubuntu:~/dockerImage$ docker images
```

REPOSITORY	TAG	IMAGE ID	CREATED	SIZE
webserver	latest	ee6e31130682	6 minutes ago	15.5MB
nginx	alpine	22f5726c6dc0	2 weeks ago	15.5MB

```
hadoopadmin@ubuntu:~/dockerImage$
```

```
hadoopadmin@ubuntu: ~/dockerImage
hadoopadmin@ubuntu:~/dockerImage$ docker run -d -p 80:80 webserver:latest
f47064d7c6cbc30afd7e322c931fee66c29f6b3c656c722a684c82f9bb1d8495
hadoopadmin@ubuntu:~/dockerImage$ docker ps
```

CONTAINER ID	IMAGE	COMMAND	CREATED	STATUS	PORTS	NAMES
f47064d7c6cb	webserver:latest	"nginx -g 'daemon ..."	2 minutes ago	Up About a minute	0.0.0.0:80->80/tcp	upbeat_mahavira

```
hadoopadmin@ubuntu:~/dockerImage$
```




```
hadoopadmin@ubuntu: ~/docker-Image
hadoopadmin@ubuntu:~/docker-Image$ docker images
REPOSITORY          TAG             IMAGE ID         CREATED          SIZE
smfahad/webserver    latest          ee6e31130682    45 hours ago    15.5MB
nginx                alpine          22f5726c6dc0    3 weeks ago     15.5MB
hadoopadmin@ubuntu:~/docker-Image$ docker save smfahad/webserver > webserver.tar
hadoopadmin@ubuntu:~/docker-Image$ ls
webserver.tar
hadoopadmin@ubuntu:~/docker-Image$
```

```
hadoopadmin@ubuntu: ~
hadoopadmin@ubuntu:~$ docker images
REPOSITORY          TAG             IMAGE ID         CREATED          SIZE
nginx                alpine          22f5726c6dc0    3 weeks ago     15.5MB
hadoopadmin@ubuntu:~$ docker load --input webserver.tar
f8896b0a4f8e: Loading layer [=====] 5.12kB/5.12kB
Loaded image: smfahad/webserver:latest
hadoopadmin@ubuntu:~$ docker images
REPOSITORY          TAG             IMAGE ID         CREATED          SIZE
smfahad/webserver    latest          ee6e31130682    45 hours ago    15.5MB
nginx                alpine          22f5726c6dc0    3 weeks ago     15.5MB
hadoopadmin@ubuntu:~$
```

```

hadoopadmin@hadoopadmin-VirtualBox: ~
hadoopadmin@hadoopadmin-VirtualBox:~$ sudo docker swarm init --listen-addr 10.0.2.15:2377
[sudo] password for hadoopadmin:
Swarm initialized: current node (j79cuhjdk470kfxea3ojcdawv) is now a manager.

To add a worker to this swarm, run the following command:

    docker swarm join --token SWMTKN-1-1o9xnsqxp5gg8nphkenrl47sid6en3cm8dfv0yhnajobm84d6-d2gjmnt9h
aegu5bo5p7ccwsio 10.0.2.15:2377

To add a manager to this swarm, run 'docker swarm join-token manager' and follow the instructions.

hadoopadmin@hadoopadmin-VirtualBox:~$

```

```

hadoopadmin@hadoopadmin-VirtualBox: ~
hadoopadmin@hadoopadmin-VirtualBox:~$ sudo docker node ls
ID                HOSTNAME          STATUS    AVAILABILITY    MANAGER STATUS
j79cuhjdk470kfxea3ojcdawv *  hadoopadmin-VirtualBox  Ready     Active           Leader
hadoopadmin@hadoopadmin-VirtualBox:~$

```

```

hadoopadmin@hadoopadmin-VirtualBox: ~
hadoopadmin@hadoopadmin-VirtualBox:~$ sudo docker service create --name NginxServer nginx
[sudo] password for hadoopadmin:
9551w7o5j125ffdmy1srd0qf
Since --detach=false was not specified, tasks will be created in the background.
In a future release, --detach=false will become the default.
hadoopadmin@hadoopadmin-VirtualBox:~$ sudo docker service ls
ID                NAME              MODE                REPLICAS            IMAGE              PORTS
9551w7o5j125      NginxServer       replicated          1/1                 nginx:latest
hadoopadmin@hadoopadmin-VirtualBox:~$

```

```

hadoopadmin@hadoopadmin-VirtualBox: ~
hadoopadmin@hadoopadmin-VirtualBox:~$ sudo docker service create --name NginxServer --replicas 5 nginx
zhlz5xlgotraooy50swsywdf8
Since --detach=false was not specified, tasks will be created in the background.
In a future release, --detach=false will become the default.
hadoopadmin@hadoopadmin-VirtualBox:~$ sudo docker service ls
ID                NAME              MODE                REPLICAS            IMAGE              PORTS
zhlz5xlgotra      NginxServer       replicated          5/5                 nginx:latest
hadoopadmin@hadoopadmin-VirtualBox:~$

```

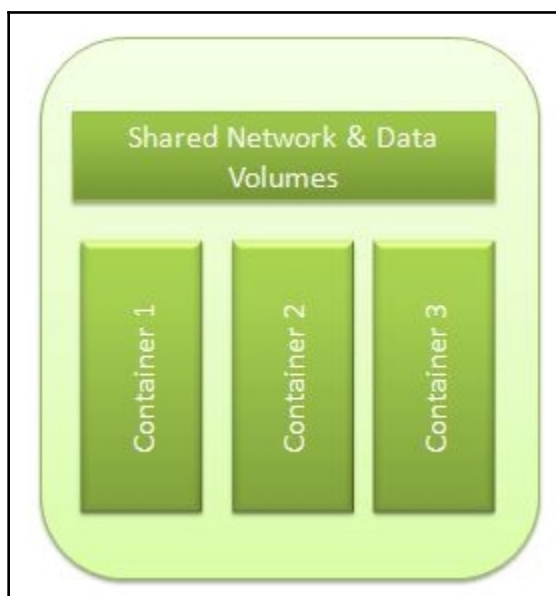
```

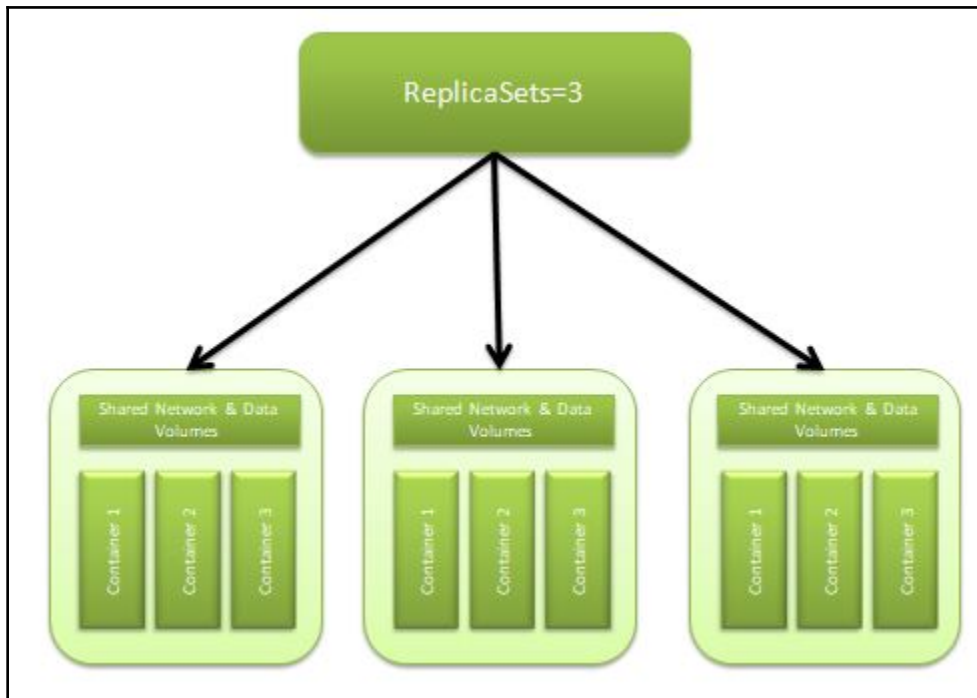
hadoopadmin@hadoopadmin-VirtualBox: ~
hadoopadmin@hadoopadmin-VirtualBox:~$ sudo docker service ps NginxServer
ID                NAME              IMAGE              NODE                DESIRED STATE     CURRENT STATE     ERROR
smfj3pzn7j7u      NginxServer.1     nginx:latest       hadoopadmin-VirtualBox  Running           Running 8 minutes ago
dyq0s33ncpnk      NginxServer.2     nginx:latest       hadoopadmin-VirtualBox  Running           Running 8 minutes ago
deeryr0phxar      NginxServer.3     nginx:latest       hadoopadmin-VirtualBox  Running           Running 8 minutes ago
u7wppjxptpcl      NginxServer.4     nginx:latest       hadoopadmin-VirtualBox  Running           Running 8 minutes ago
f61oyqz0bov16     NginxServer.5     nginx:latest       hadoopadmin-VirtualBox  Running           Running 8 minutes ago
hadoopadmin@hadoopadmin-VirtualBox:~$

```



```
hadoopadmin@hadoopadmin-VirtualBox: ~  
hadoopadmin@hadoopadmin-VirtualBox:~$ sudo docker service rm NginxServer  
NginxServer  
hadoopadmin@hadoopadmin-VirtualBox:~$ sudo docker service ls  
ID                NAME                MODE                REPLICAS            IMAGE                PORTS  
hadoopadmin@hadoopadmin-VirtualBox:~$ sudo docker service ps NginxServer  
no such service: NginxServer  
hadoopadmin@hadoopadmin-VirtualBox:~$
```





```
hadoopadmin@hadoopadmin-VirtualBox: ~  
hadoopadmin@hadoopadmin-VirtualBox:~$ kubectl version  
Client Version: version.Info{Major:"1", Minor:"9", GitVersion:"v1.9.0", GitCommit:"925c127ec6b946659ad0fd596fa959be43f0cc05", GitTreeState:"clean", BuildDate:"2017-12-15T21:07:38Z", GoVersion:"go1.9.2", Compiler:"gc", Platform:"linux/amd64"}  
The connection to the server localhost:8080 was refused - did you specify the right host or port?  
hadoopadmin@hadoopadmin-VirtualBox:~$ minikube version  
minikube version: v0.24.1  
hadoopadmin@hadoopadmin-VirtualBox:~$
```

```

hadoopadmin@ubuntu: ~
hadoopadmin@ubuntu:~$ minikube start
Starting local Kubernetes v1.8.0 cluster...
Starting VM...
Getting VM IP address...
Moving files into cluster...
Setting up certs...
Connecting to cluster...
Setting up kubeconfig...
Starting cluster components...
Kubectl is now configured to use the cluster.
Loading cached images from config file.
hadoopadmin@ubuntu:~$ kubectl cluster-info
Kubernetes master is running at https://192.168.99.100:8443

To further debug and diagnose cluster problems, use 'kubectl cluster-info dump'.
hadoopadmin@ubuntu:~$

```

```

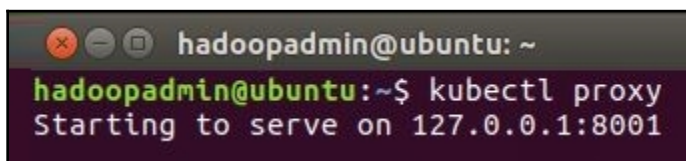
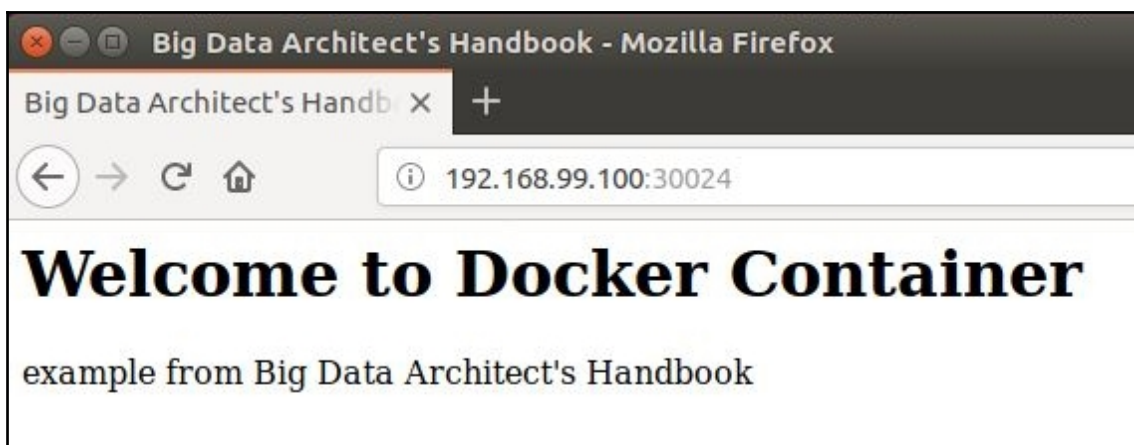
hadoopadmin@ubuntu: ~
hadoopadmin@ubuntu:~$ kubectl run webserver --image=smfahad/webserver --port=80
deployment "webserver" created
hadoopadmin@ubuntu:~$ kubectl get deployments
NAME          DESIRED   CURRENT   UP-TO-DATE   AVAILABLE   AGE
webserver     1         1         1            0           13s
hadoopadmin@ubuntu:~$

```

```

hadoopadmin@ubuntu: ~
hadoopadmin@ubuntu:~$ kubectl expose deployment webserver --port=80 --type=NodePort
service "webserver" exposed
hadoopadmin@ubuntu:~$ kubectl get services
NAME          TYPE          CLUSTER-IP   EXTERNAL-IP   PORT(S)          AGE
kubernetes    ClusterIP     10.96.0.1    <none>         443/TCP           3d
webserver     NodePort      10.96.75.99  <none>         80:30024/TCP     14s

```



Overview - Kubernetes Dashboard - Mozilla Firefox

Overview - Kubernetes X +

127.0.0.1:8001/api/v1/namespaces/kube-system/services/k... Search

kubernetes Search + CREATE

Overview

Cluster

- Namespaces
- Nodes
- Persistent Volumes
- Roles
- Storage Classes

Namespace

default

Overview

Workloads

- Cron Jobs
- Daemon Sets
- Deployments

Workloads

Workloads Statuses

100.00%

Deployments

100.00%

Pods

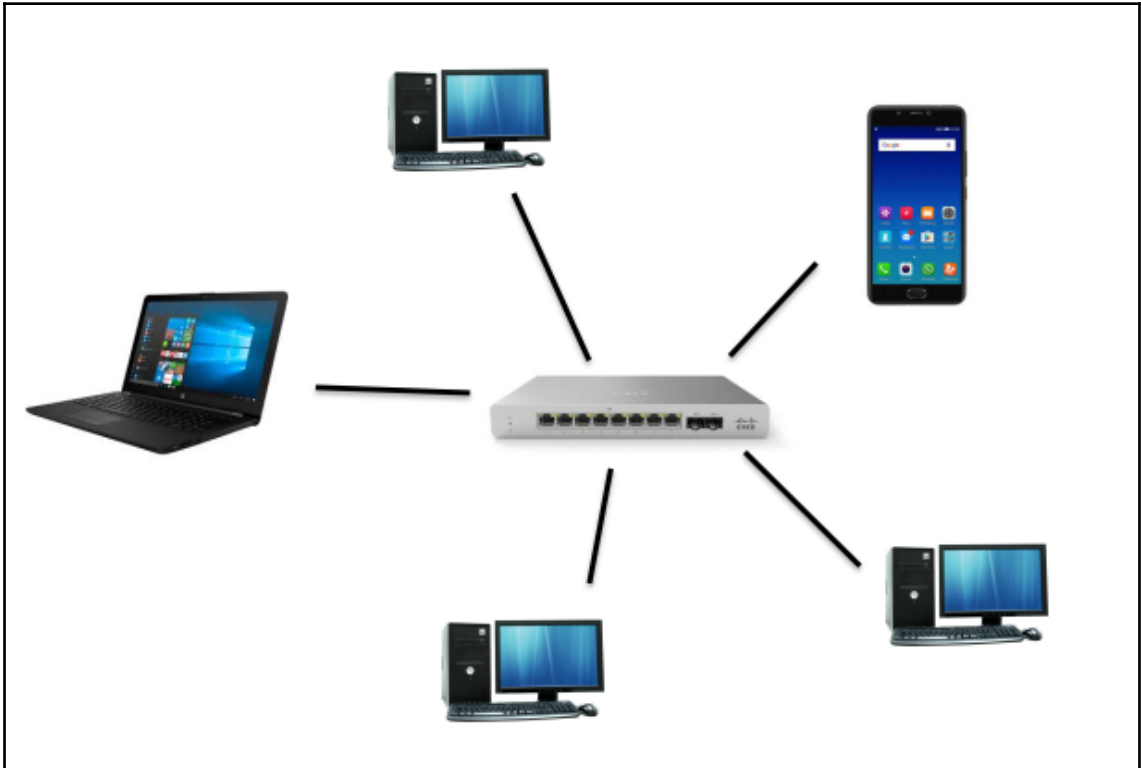
100.00%

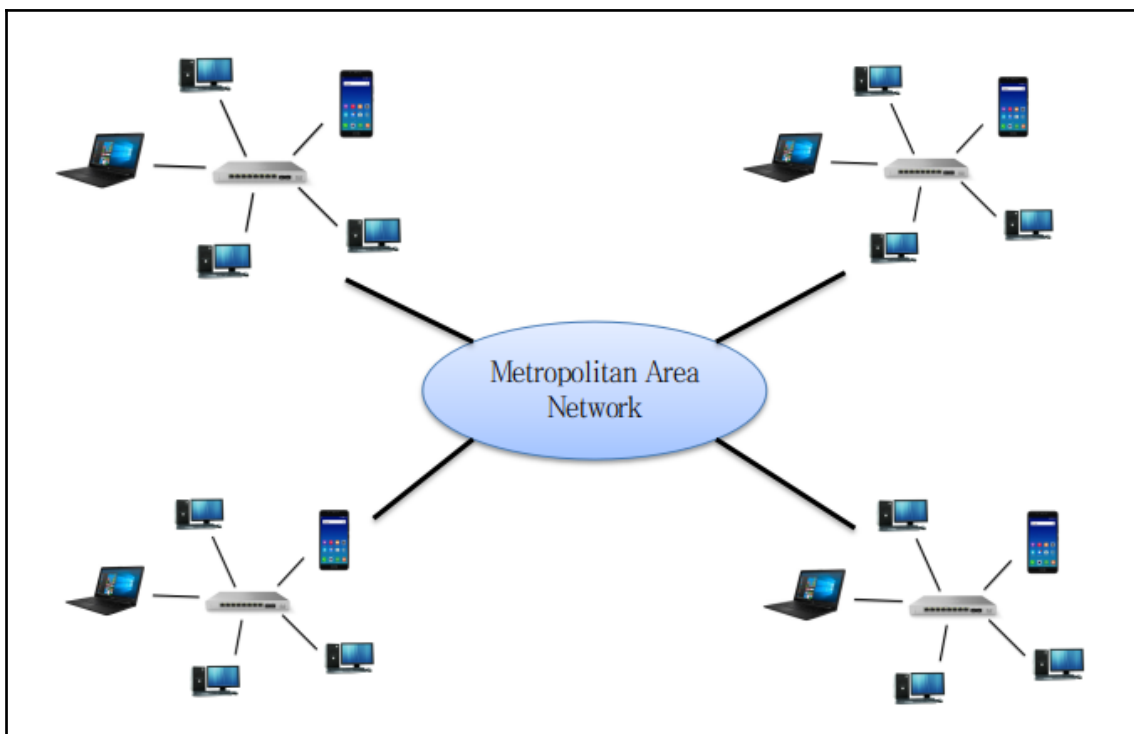
Replica Sets

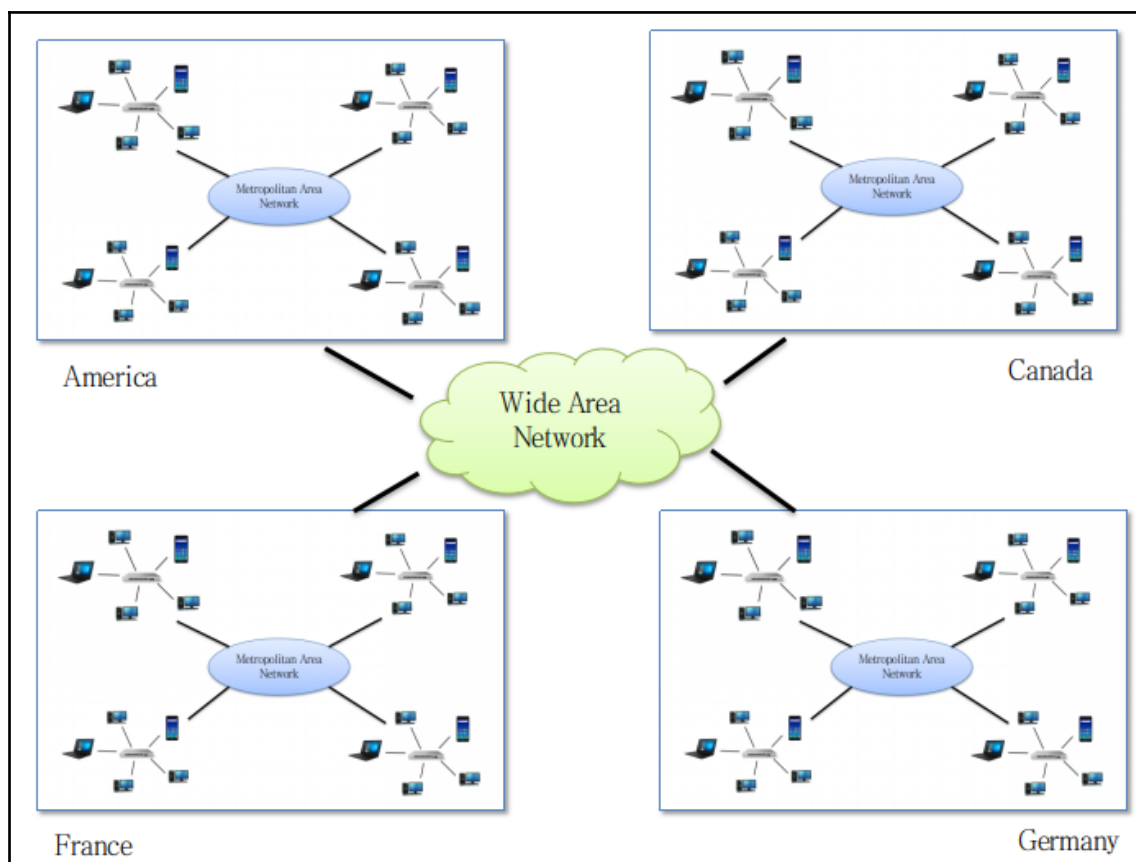
Deployments

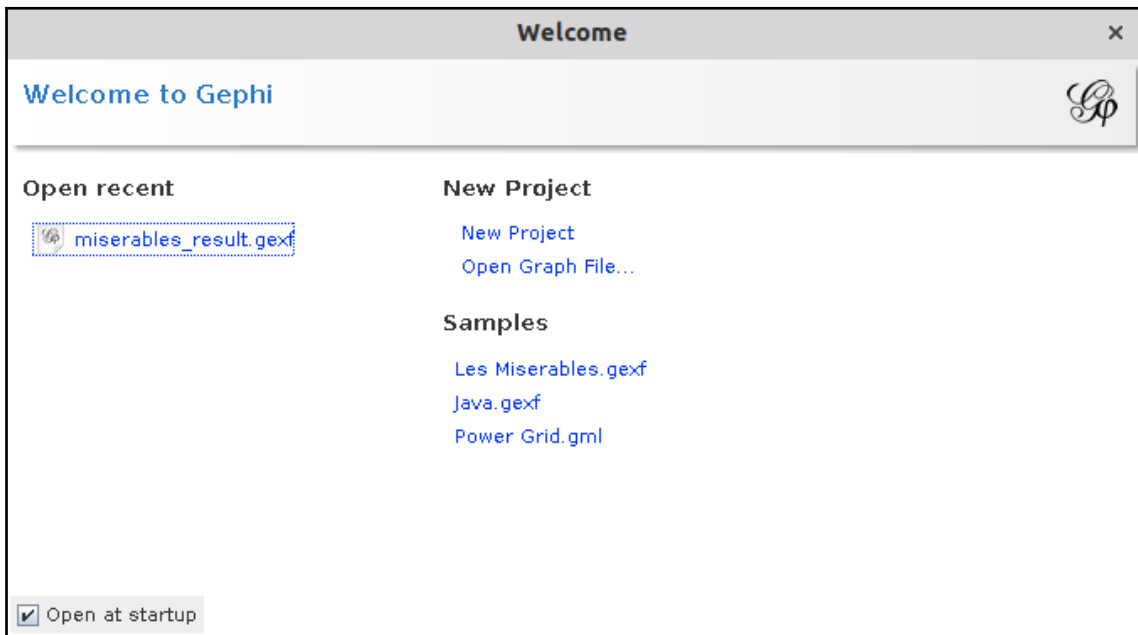
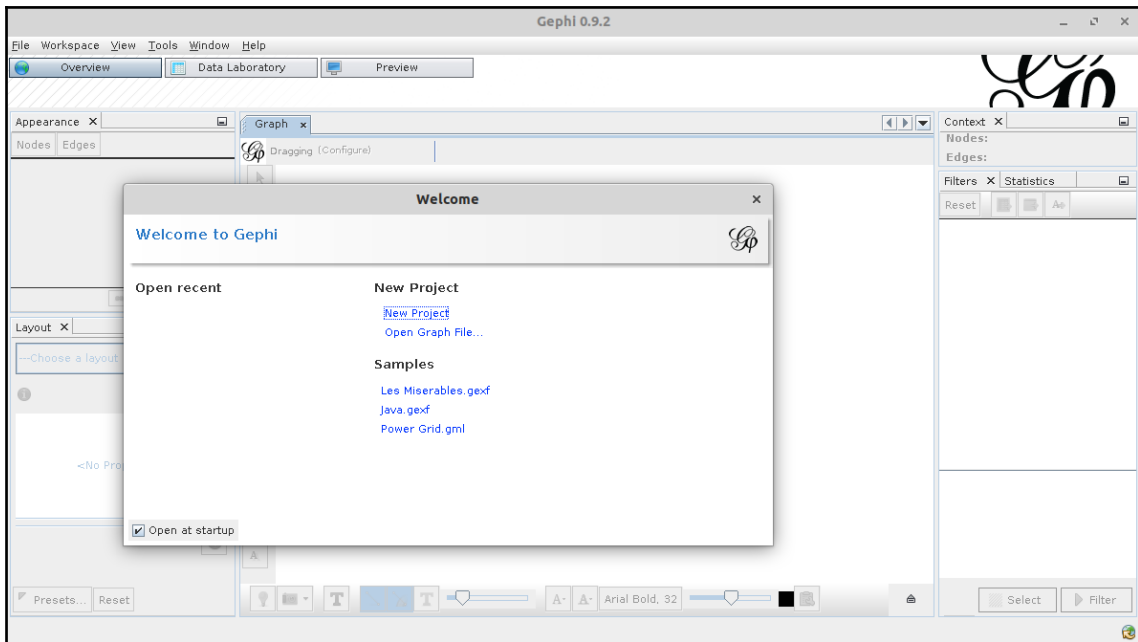
Name	Labels	Pods	Age	Images
✓ webserver	run: webserver	1 / 1	43 minutes	smfahad/webserver

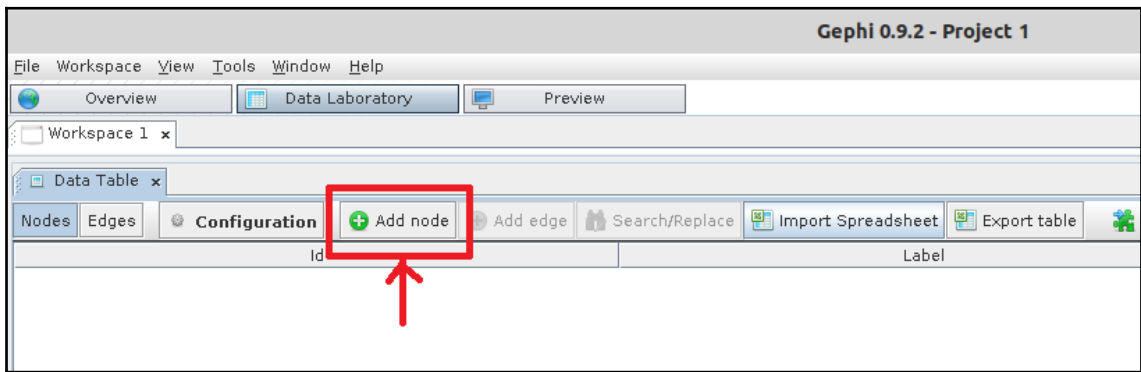
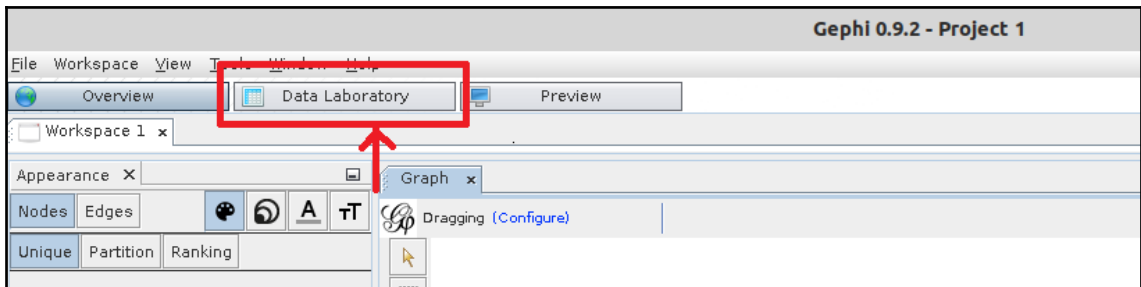
Chapter 7: Network Infrastructure

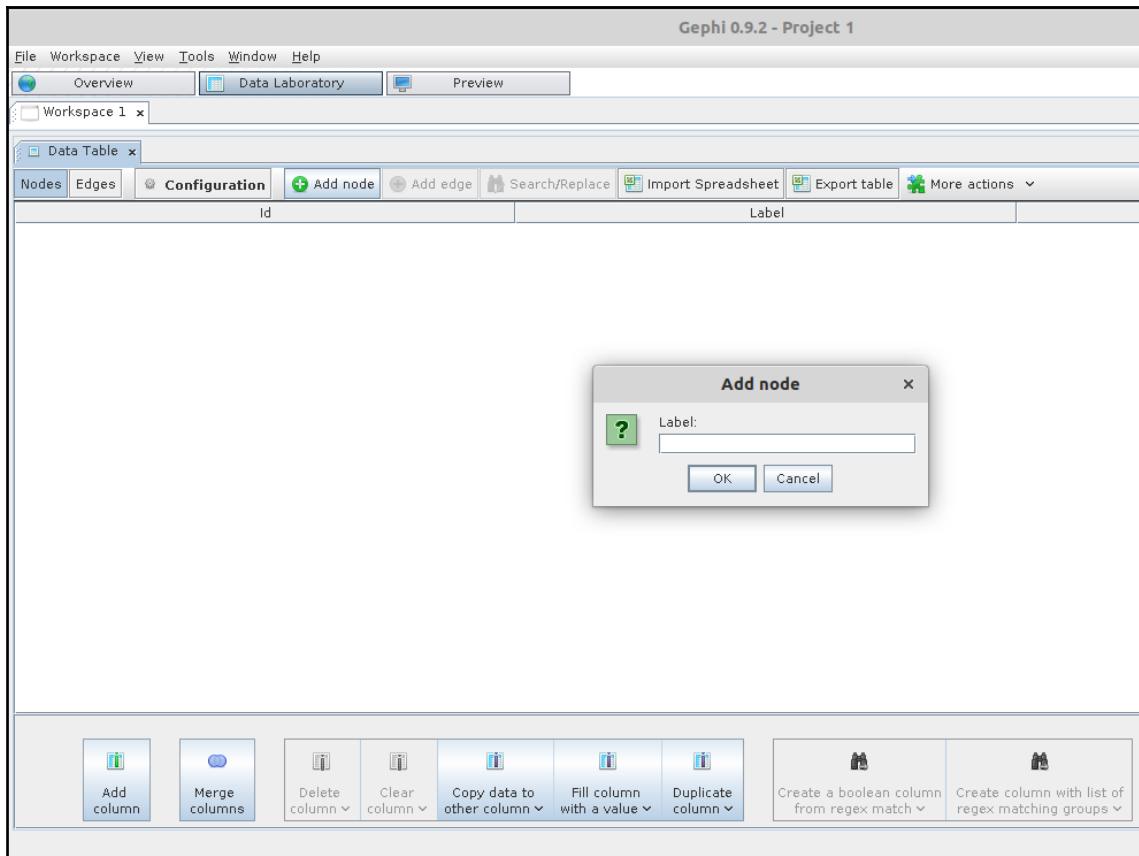






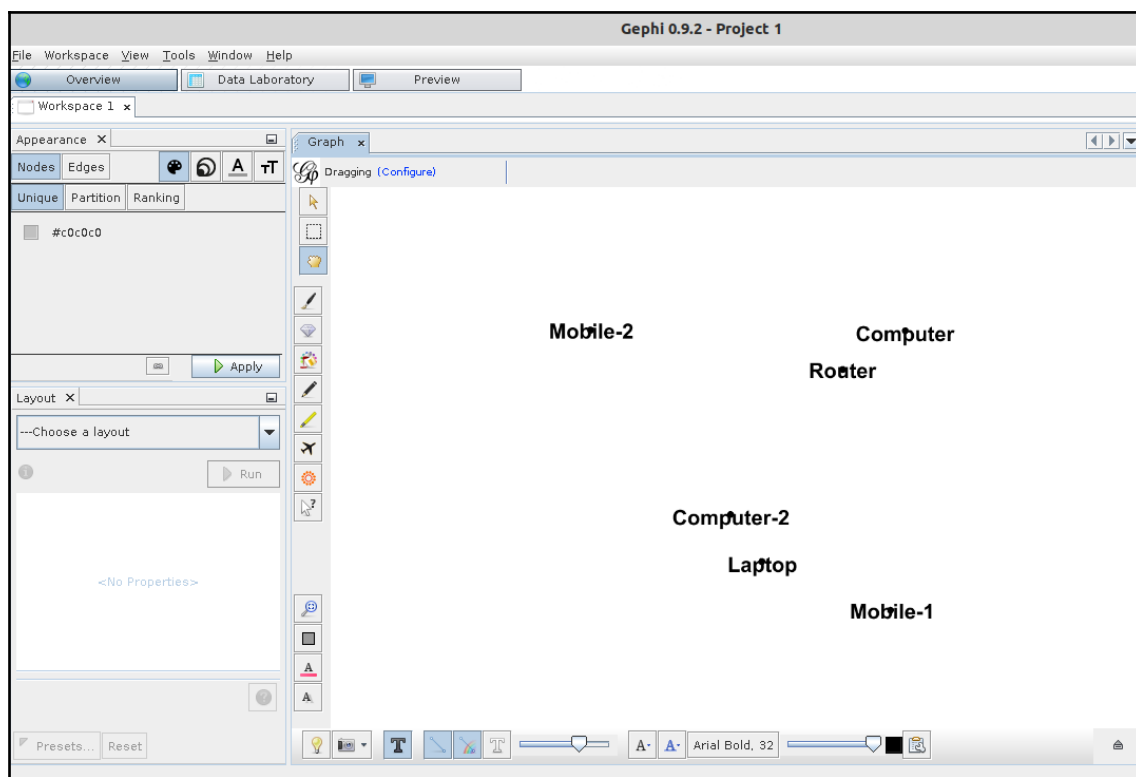


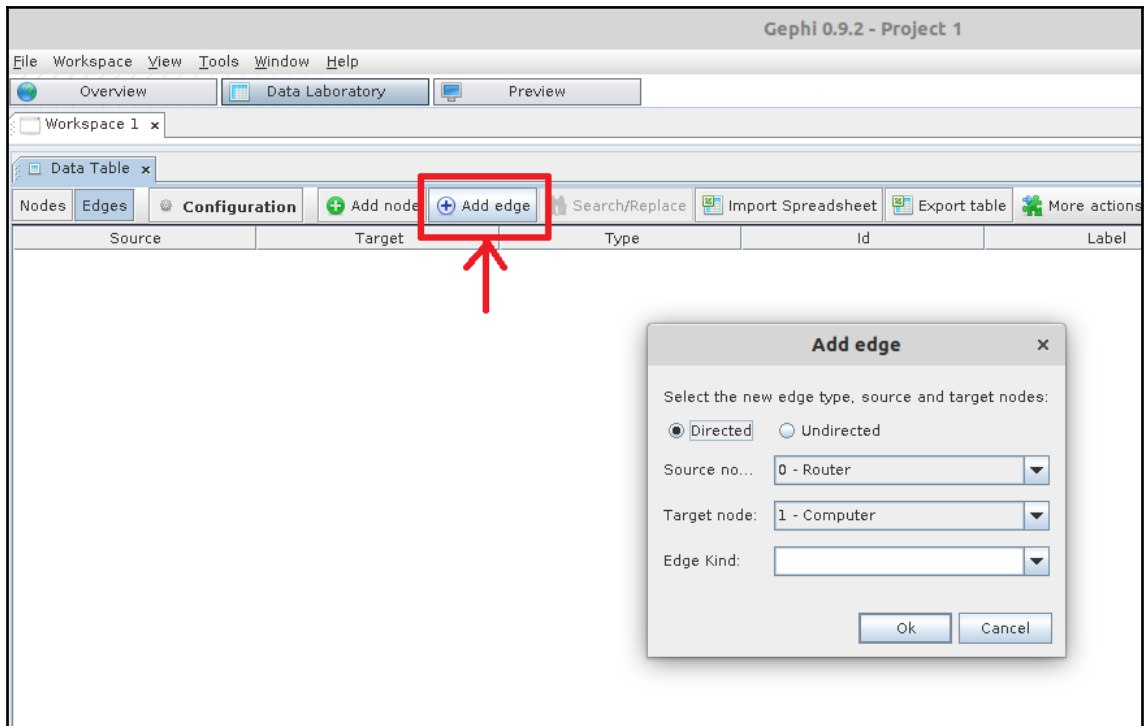




The screenshot shows the Gephi 0.9.2 - Project 1 interface. The 'Data Table' window is open, displaying a table with two columns: 'Id' and 'Label'. The table contains six rows of data. The 'Add node' dialog box is no longer visible. The main interface includes a menu bar (File, Workspace, View, Tools, Window, Help), a toolbar with buttons for Overview, Data Laboratory, and Preview, and a bottom toolbar with various data manipulation options like 'Add node', 'Add edge', 'Search/Replace', 'Import Spreadsheet', 'Export table', and 'More actions'.

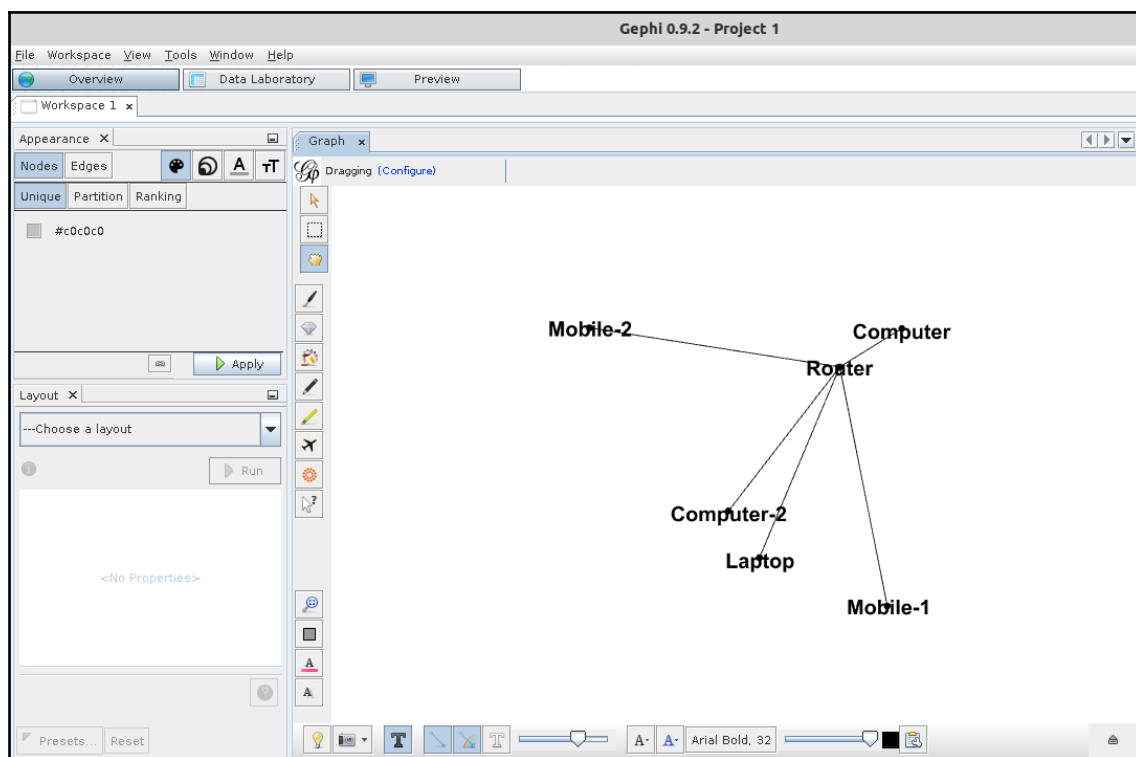
Id	Label
0	Router
1	Computer
2	Laptop
3	Mobile-1
4	Mobile-2
5	Computer-2



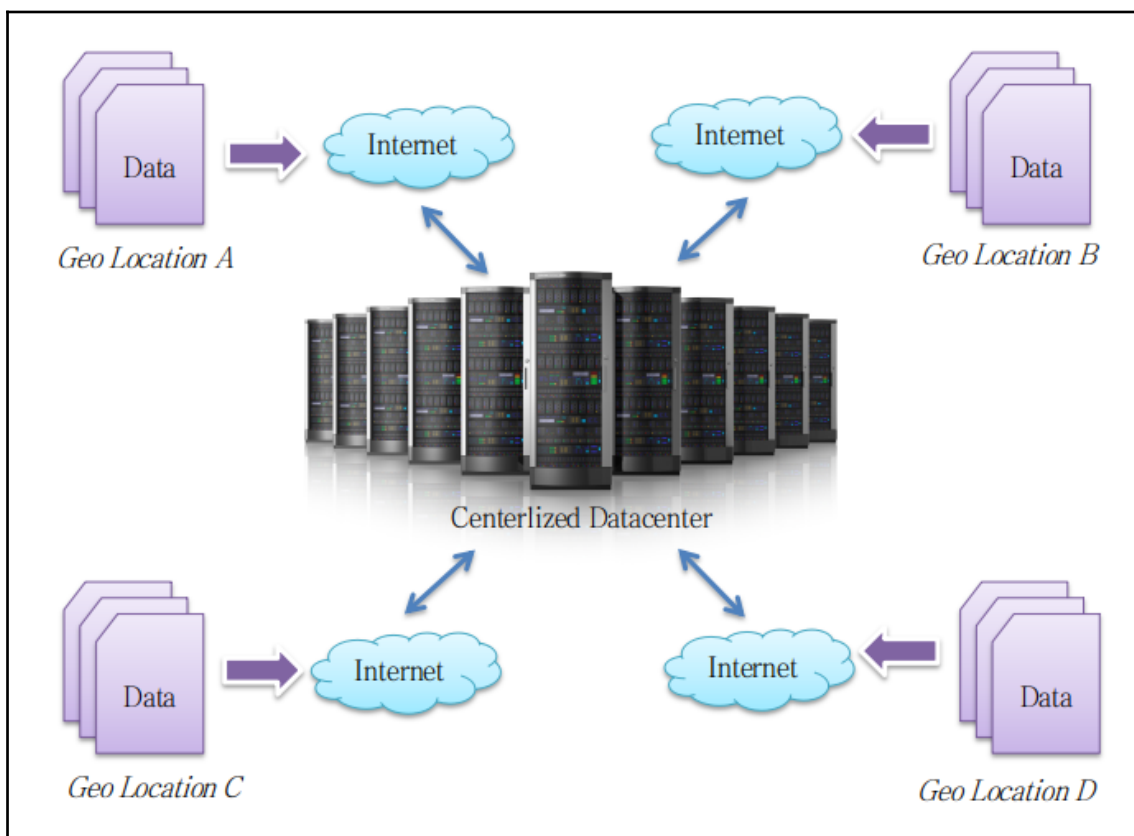


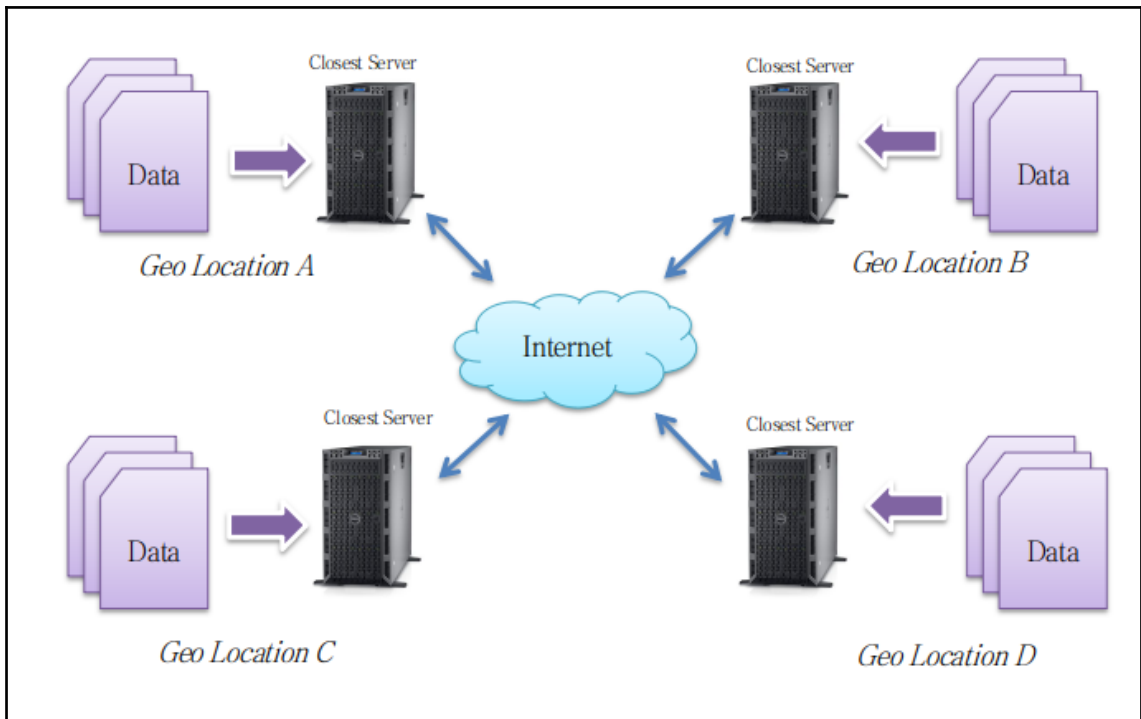
The screenshot shows the Gephi 0.9.2 - Project 1 interface with the 'Data Table' tab active. The table displays the following data:

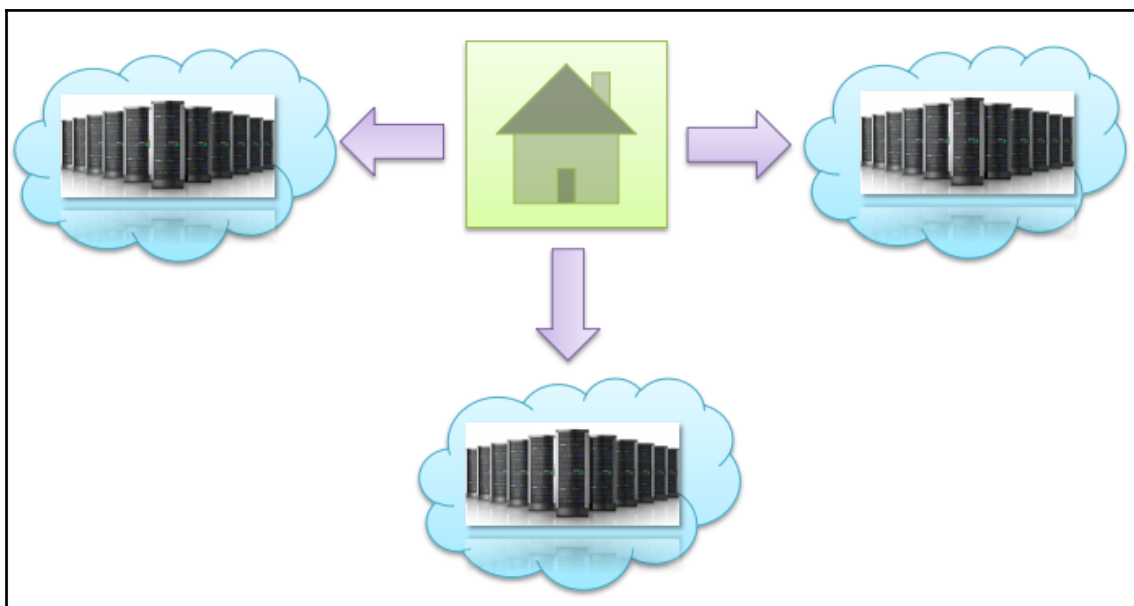
Source	Target	Type	Id	Label
0	1	Undirected	1	
0	2	Undirected	2	
0	3	Undirected	3	
0	4	Undirected	4	
0	5	Undirected	5	

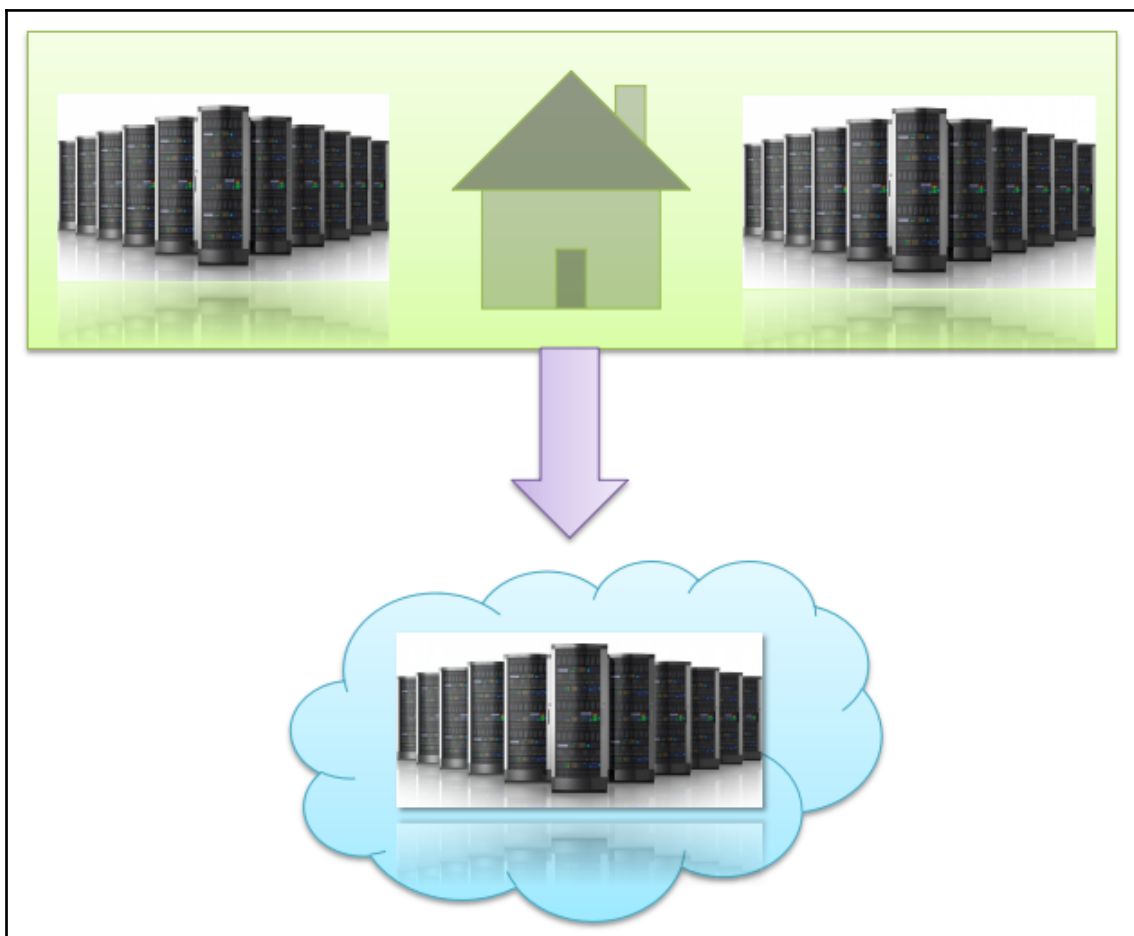


Chapter 8: Cloud Infrastructure



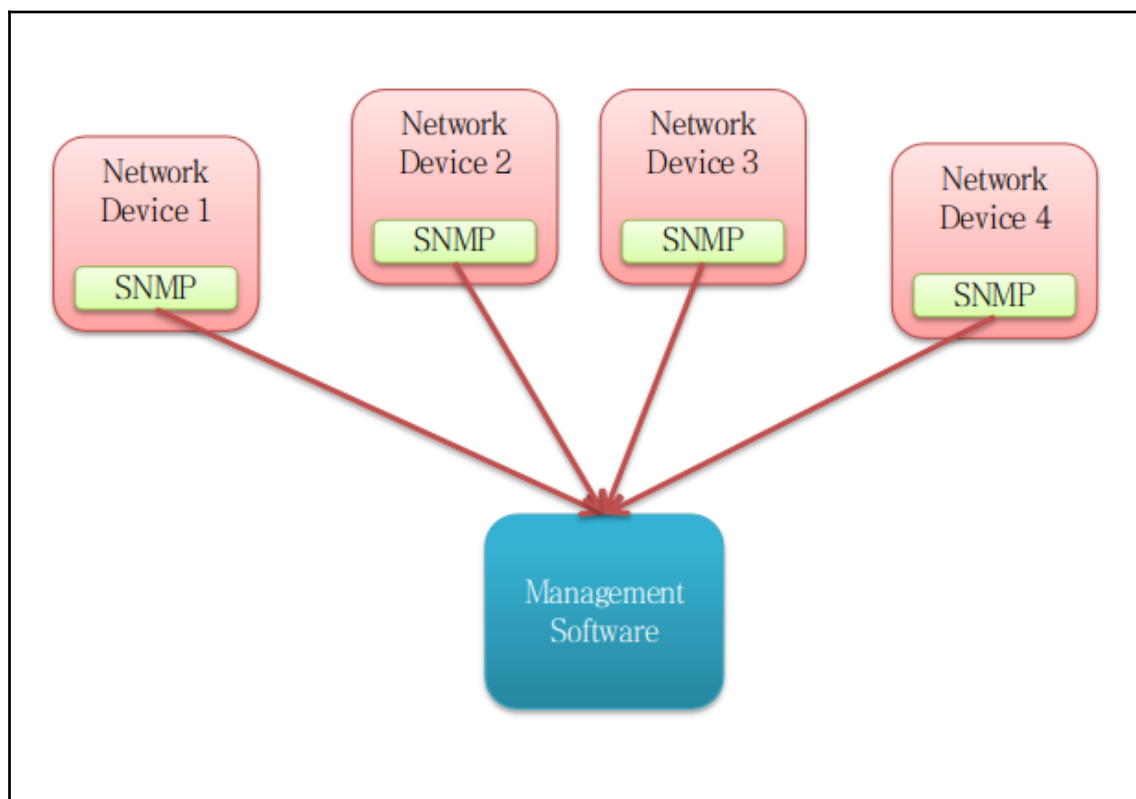


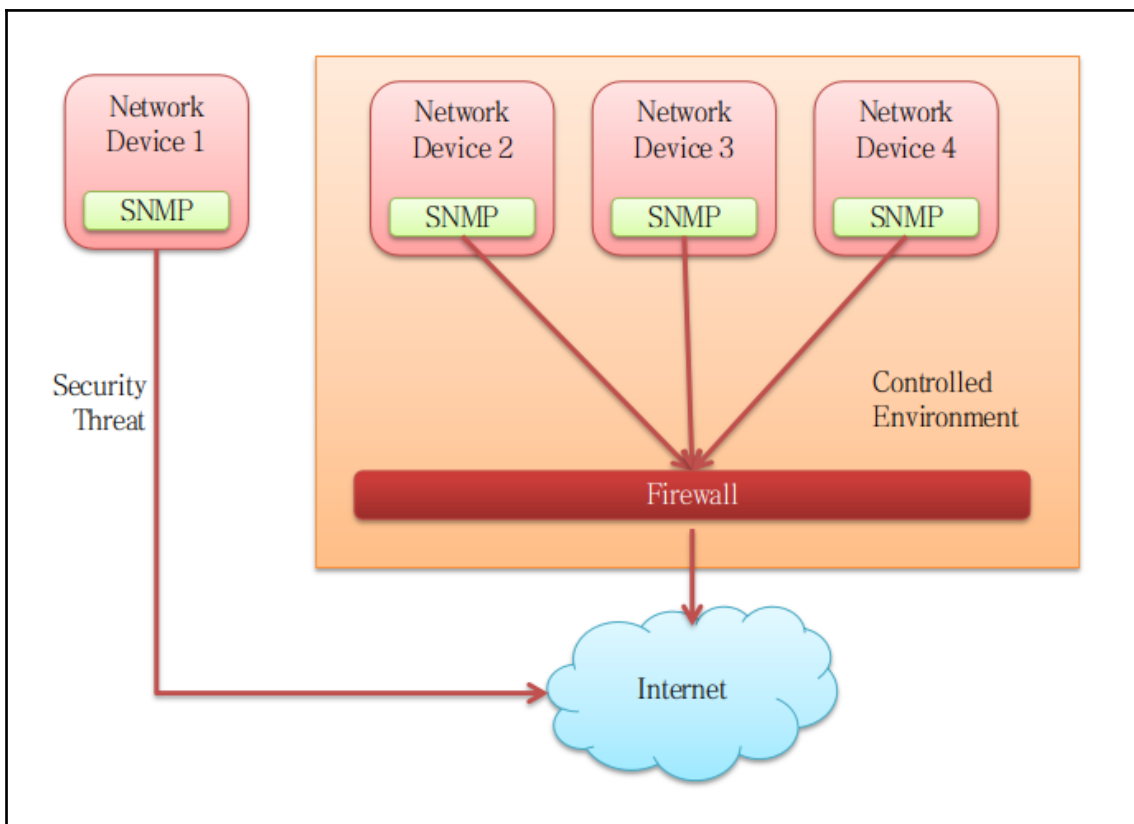


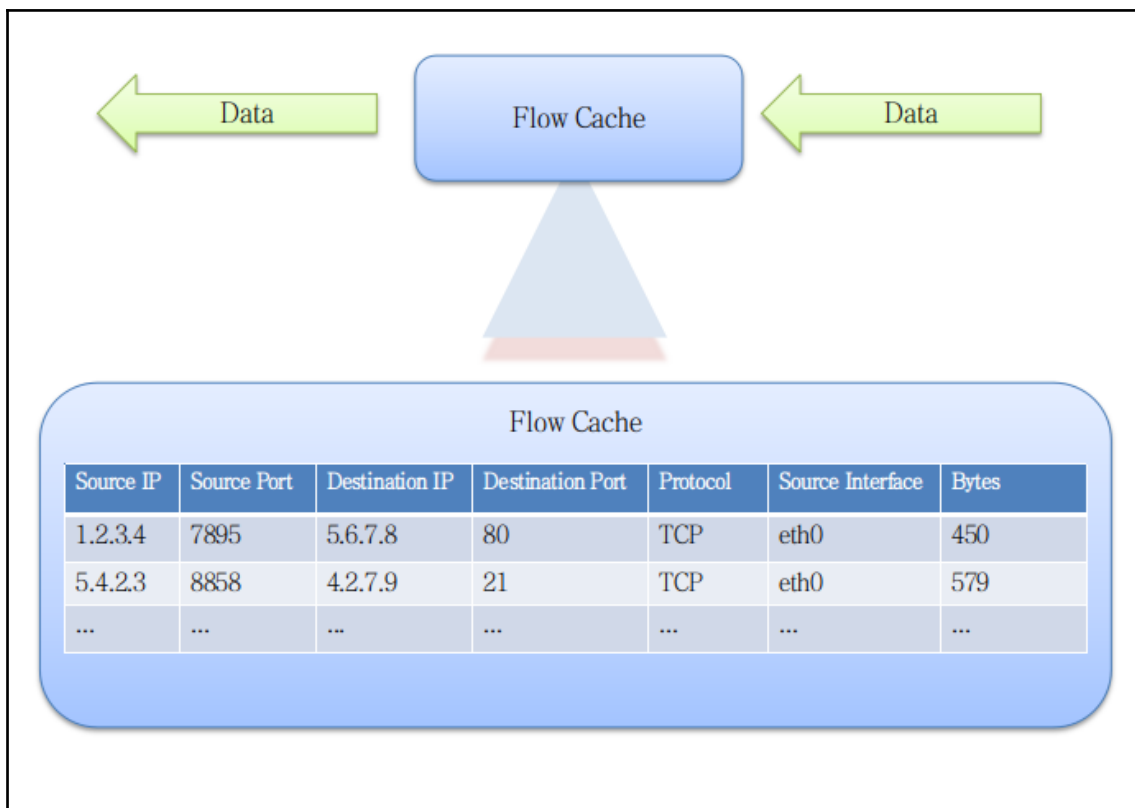


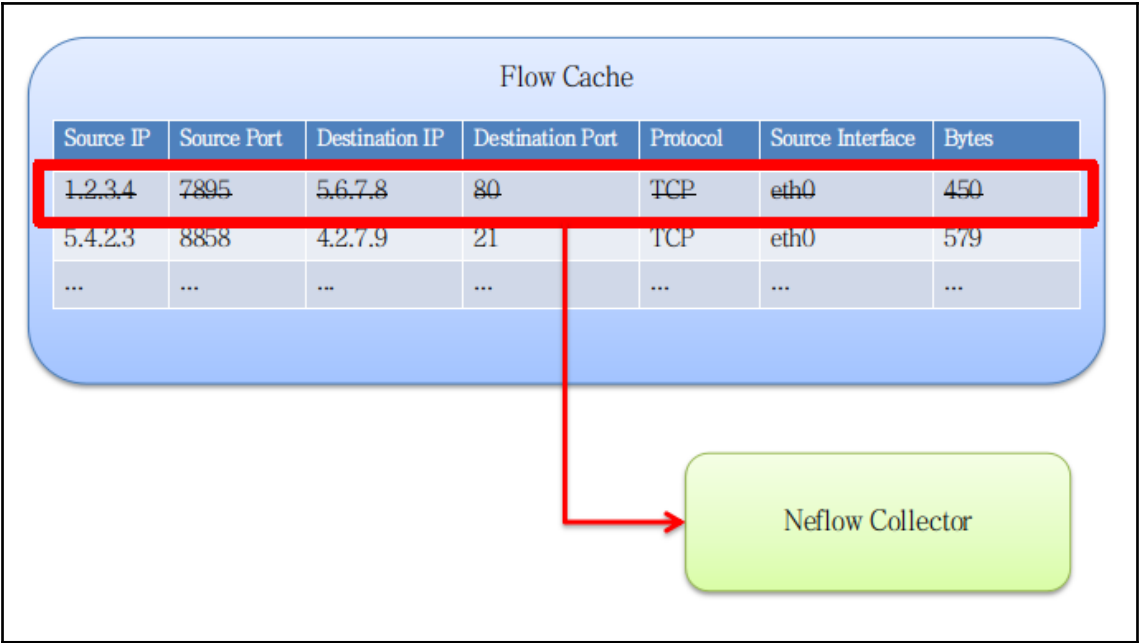


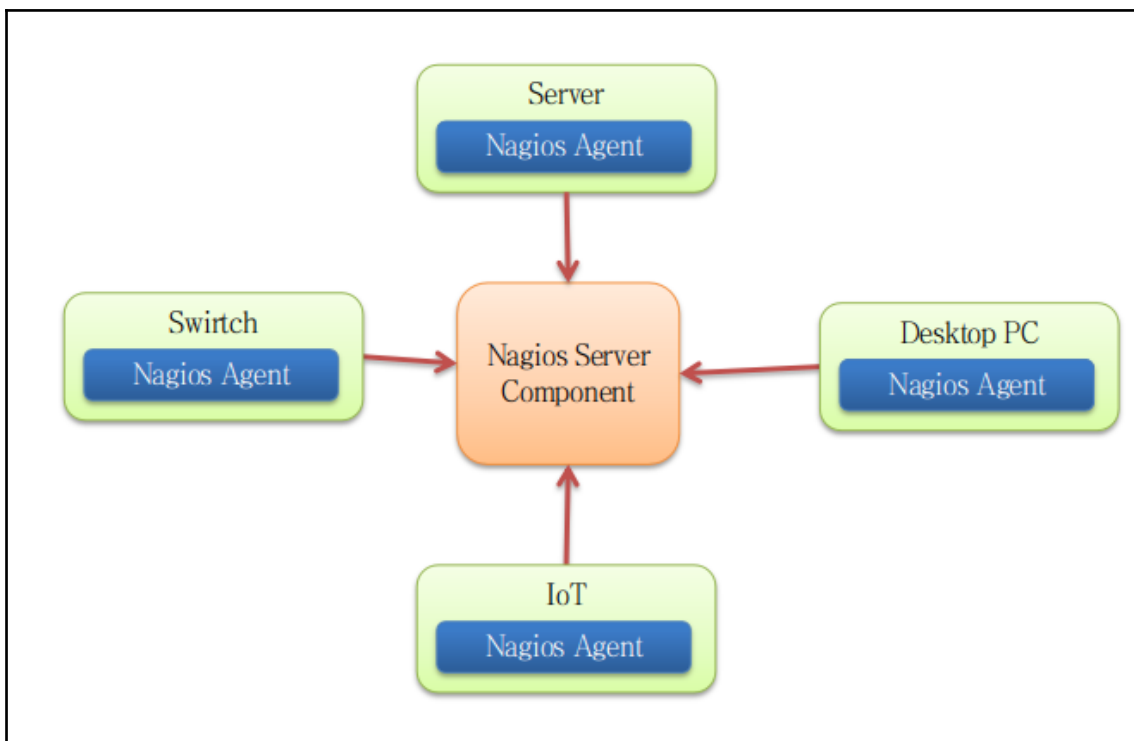
Chapter 9: Security and Monitoring

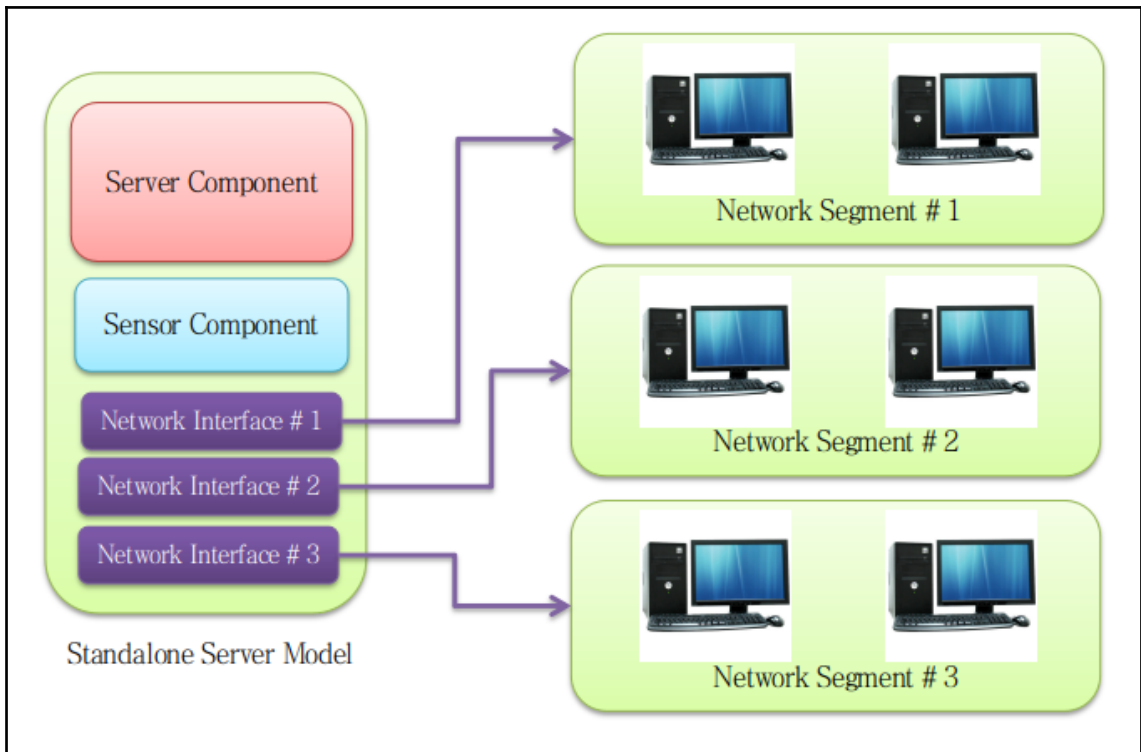


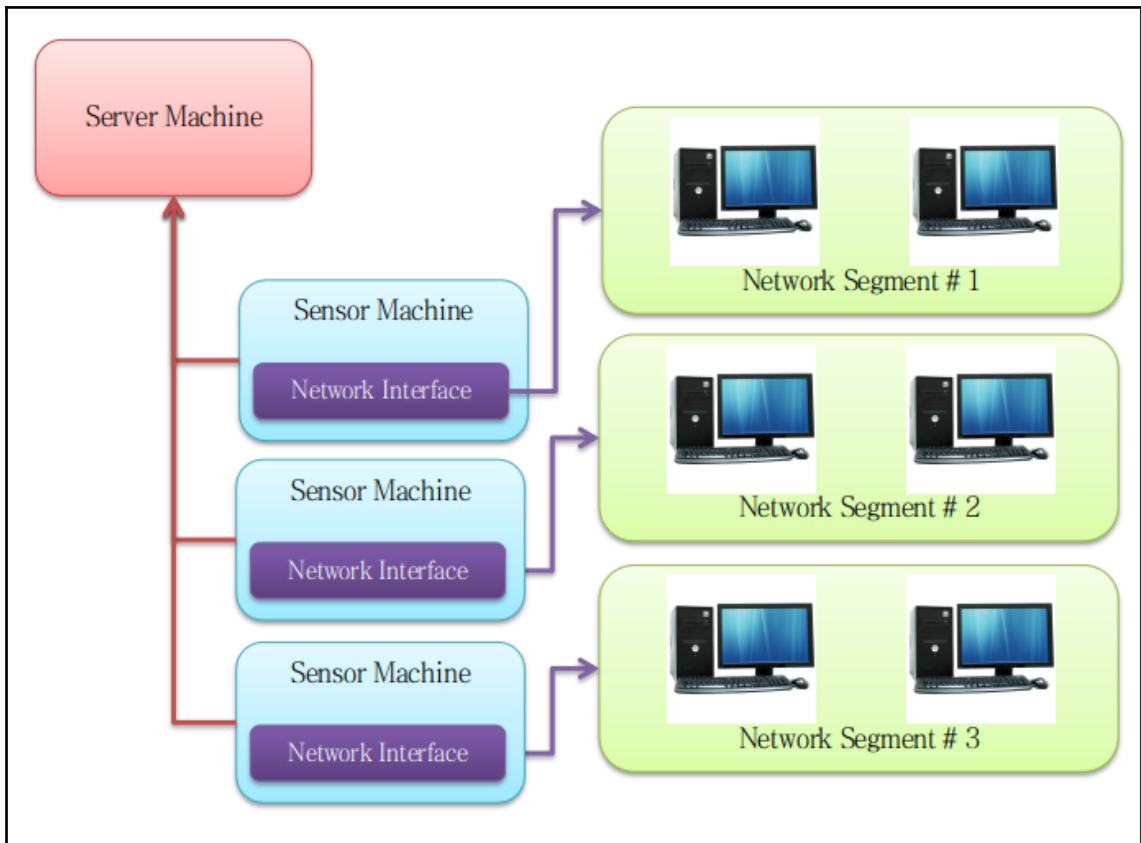


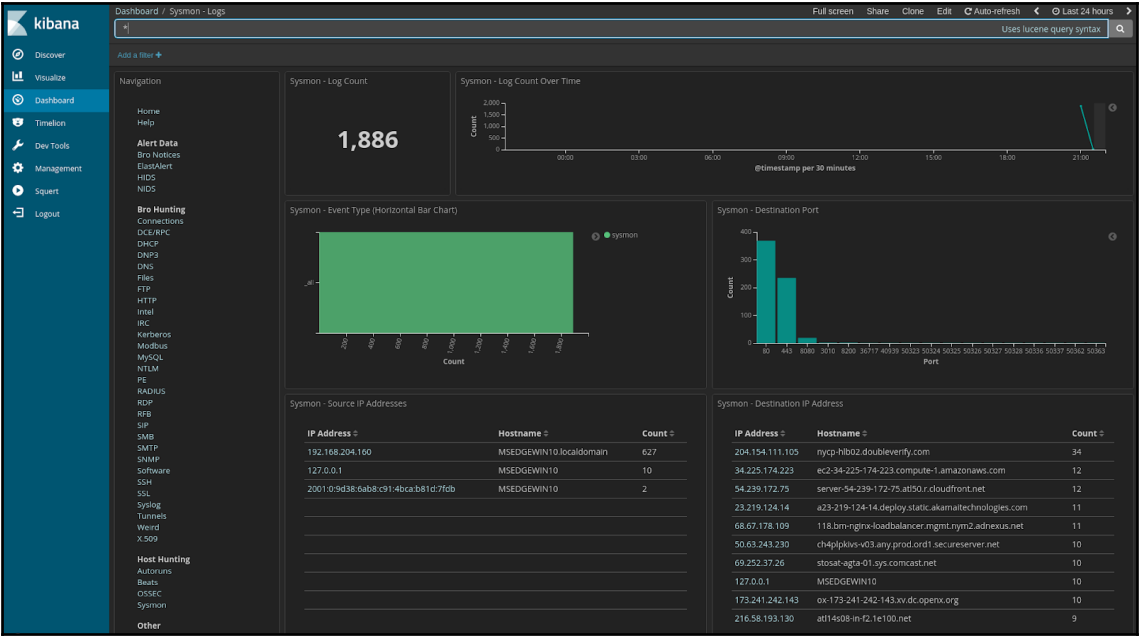




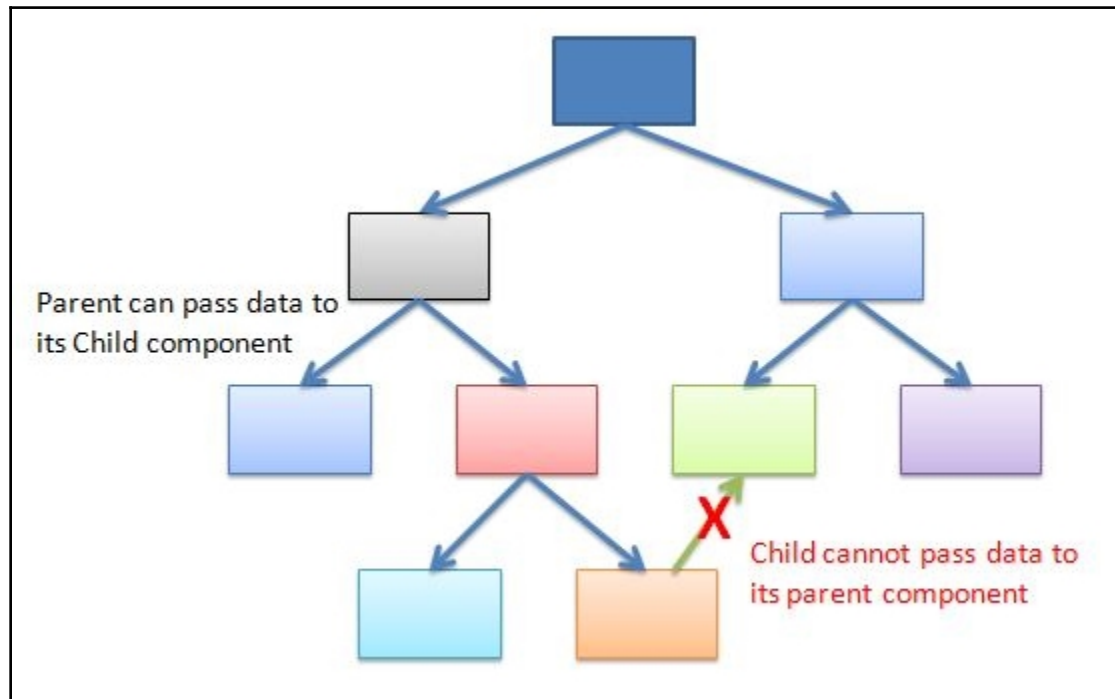
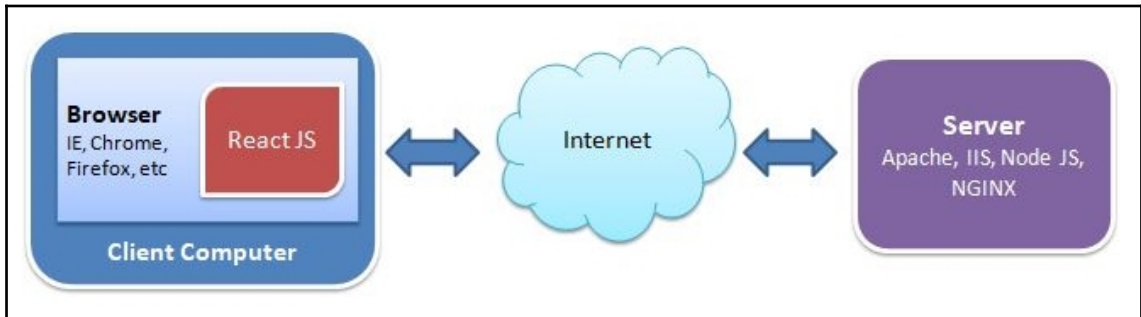








Chapter 10: Front End Architecture



```
fahadakhtar@hadoopadmin-VirtualBox: ~/react
File Edit View Search Terminal Help
├─ yargs-parser@4.2.1
├─ webpack-manifest-plugin@1.3.2
├─ fs-extra@0.30.0
├─ jsonfile@2.4.0
├─ klaw@1.3.1
└─ whatwg-fetch@2.0.3

Success! Created my-app at /home/fahadakhtar/react/my-app
Inside that directory, you can run several commands:

  npm start
    Starts the development server.

  npm run build
    Bundles the app into static files for production.

  npm test
    Starts the test runner.

  npm run eject
    Removes this tool and copies build dependencies, configuration files
    and scripts into the app directory. If you do this, you can't go back!

We suggest that you begin by typing:

  cd my-app
  npm start

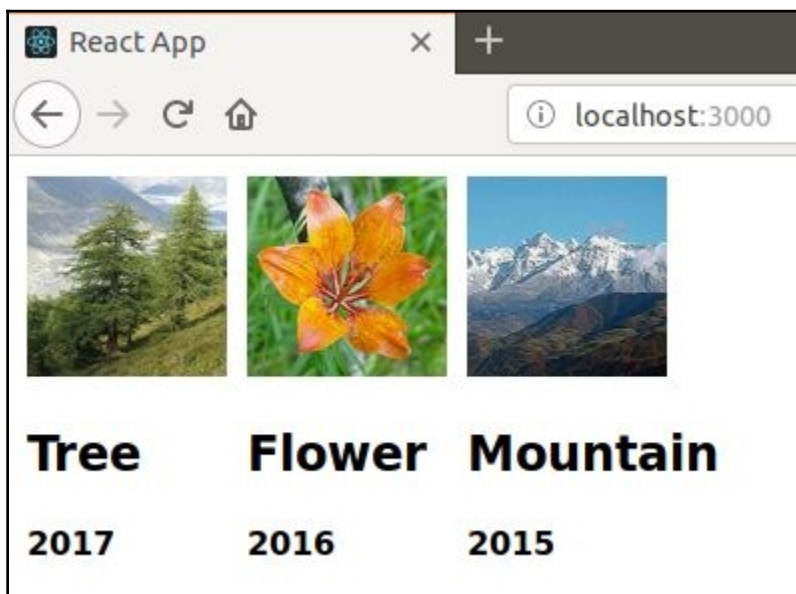
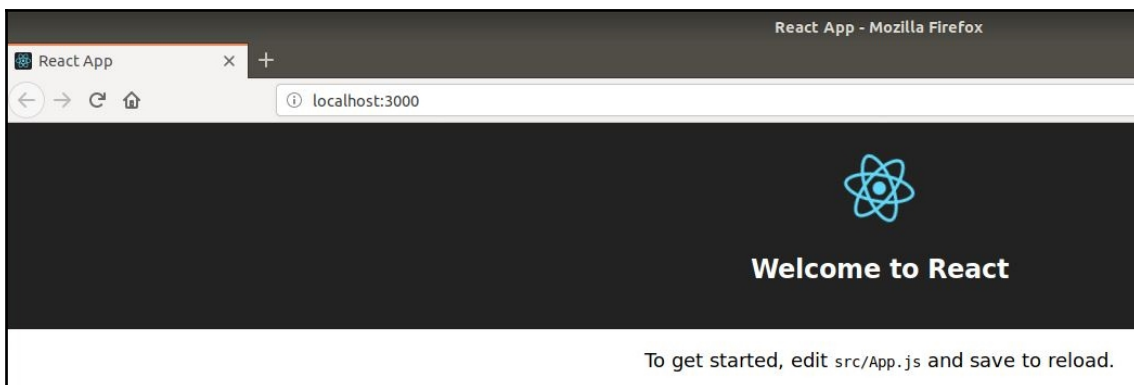
Happy hacking!
fahadakhtar@hadoopadmin-VirtualBox:~/react$
```

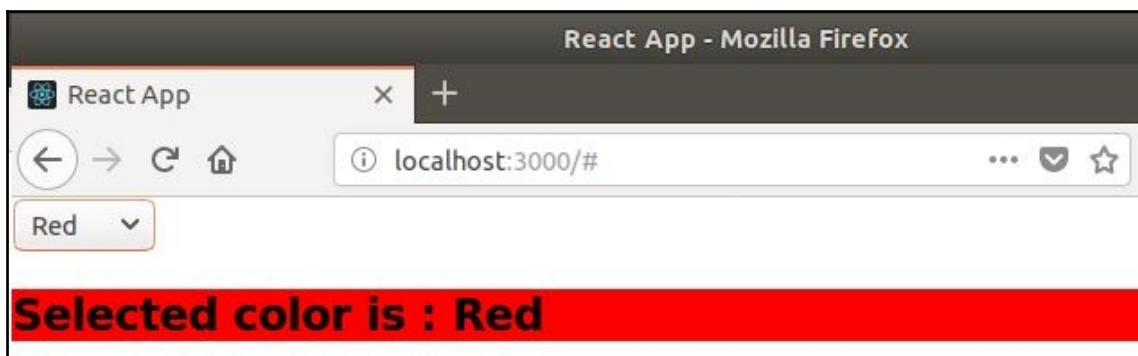
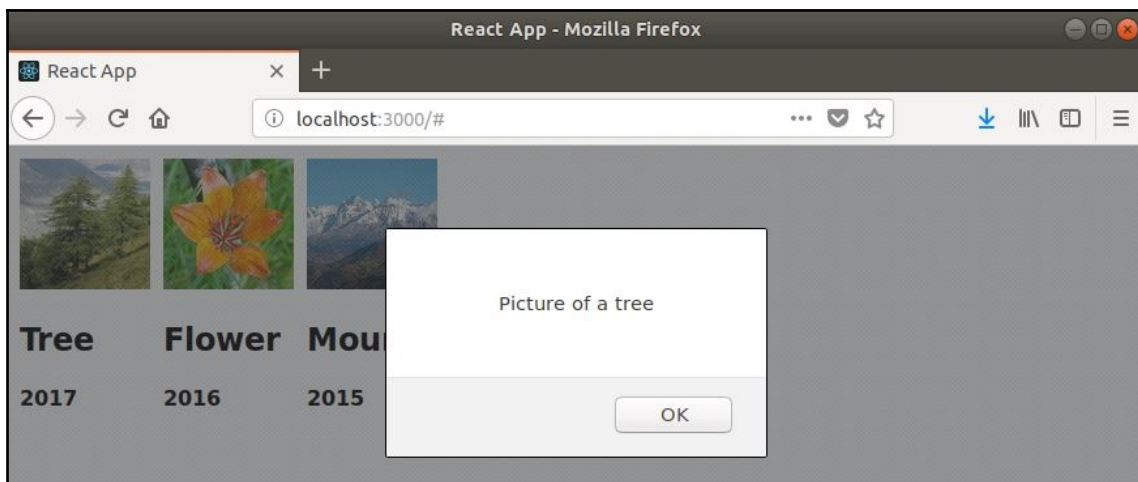
```
fahadakhtar@hadoopadmin-VirtualBox: ~/react/my-app
File Edit View Search Terminal Help
Compiled successfully!

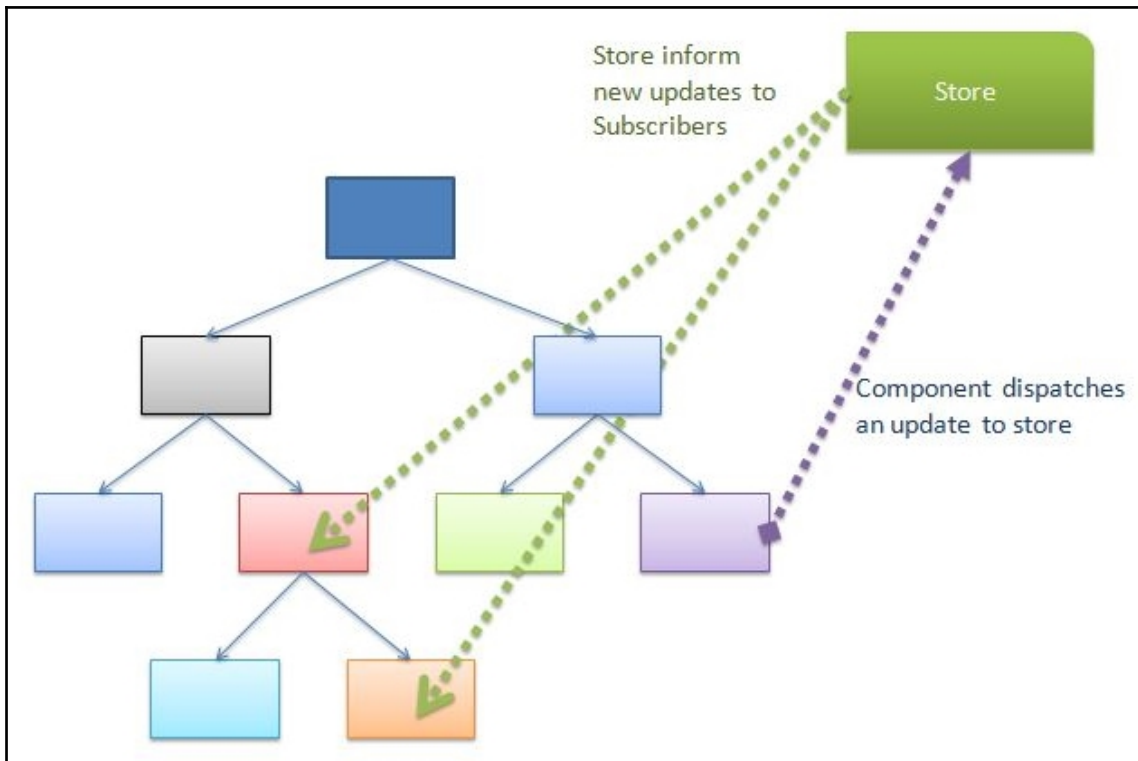
You can now view my-app in the browser.

  Local:            http://localhost:3000/
  On Your Network:  http://10.0.2.15:3000/

Note that the development build is not optimized.
To create a production build, use npm run build.
```



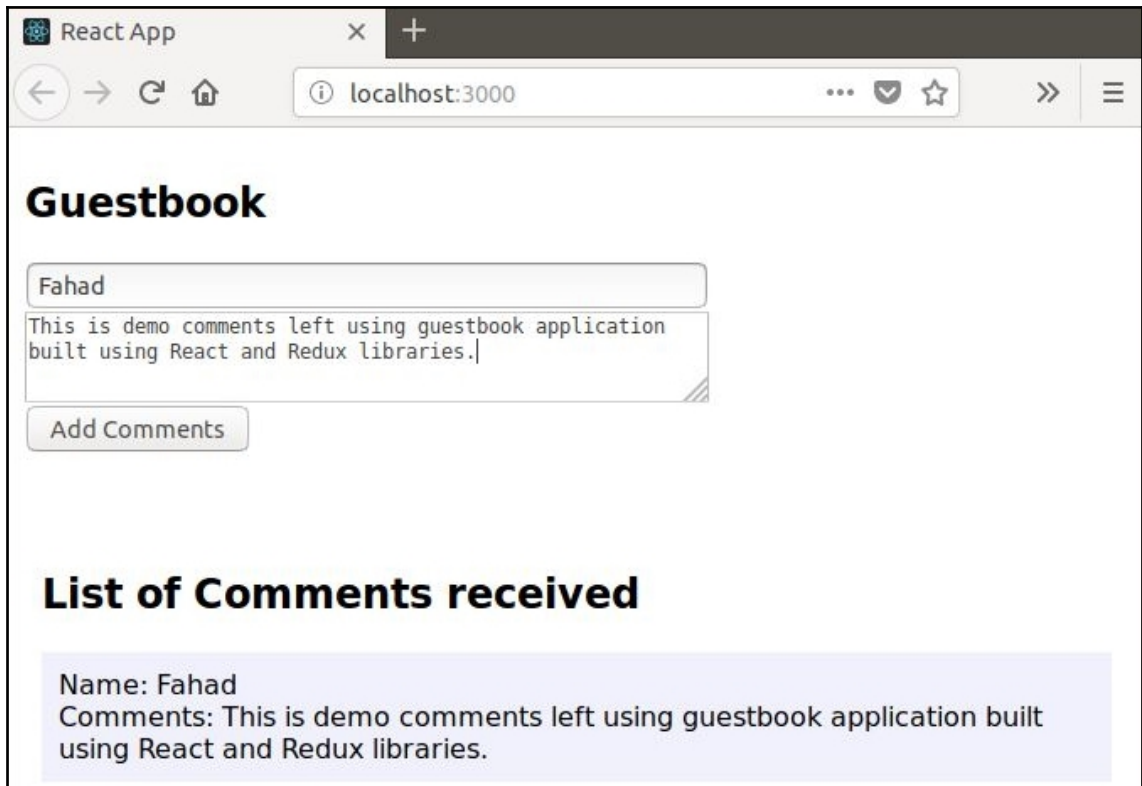




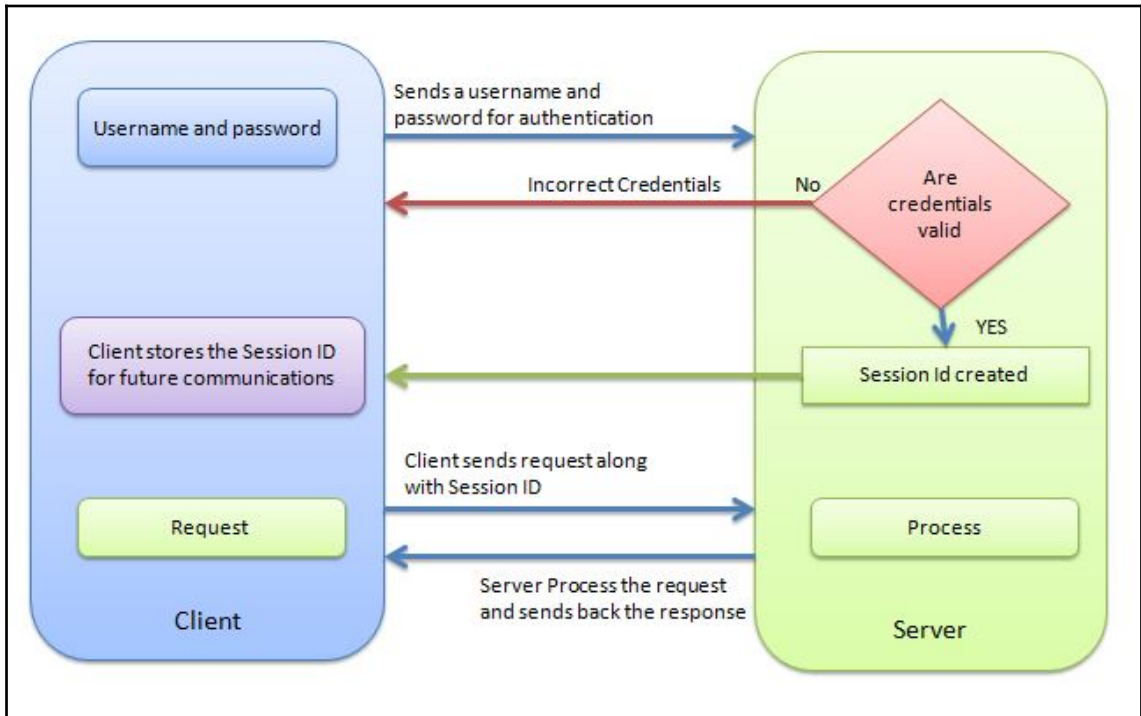
```
fahadakhtar@hadoopadmin-VirtualBox:~/react/guestbook$ npm install redux --save
questbook@0.1.0 /home/fahadakhtar/react/guestbook
├── redux@3.7.2
│   ├── lodash-es@4.17.4
│   └── symbol-observable@1.2.0
└─┘

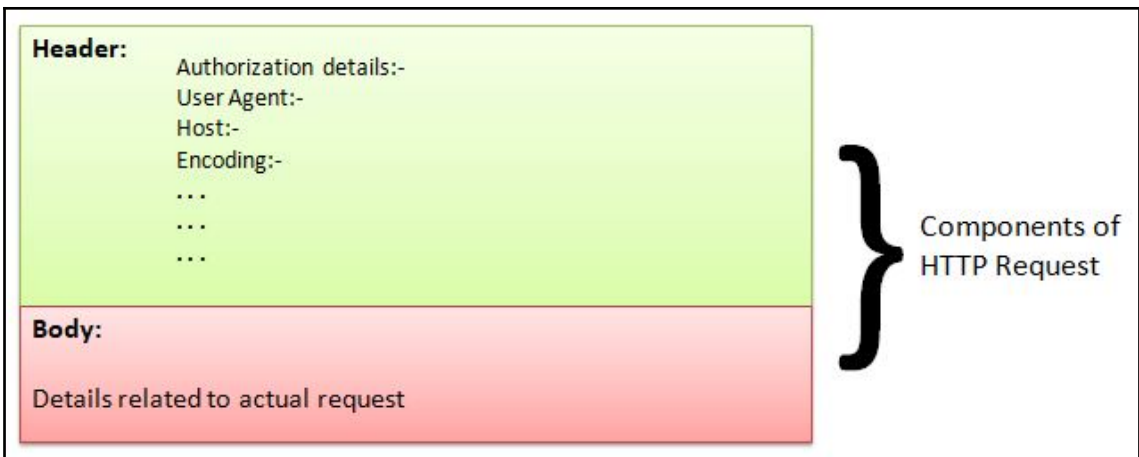
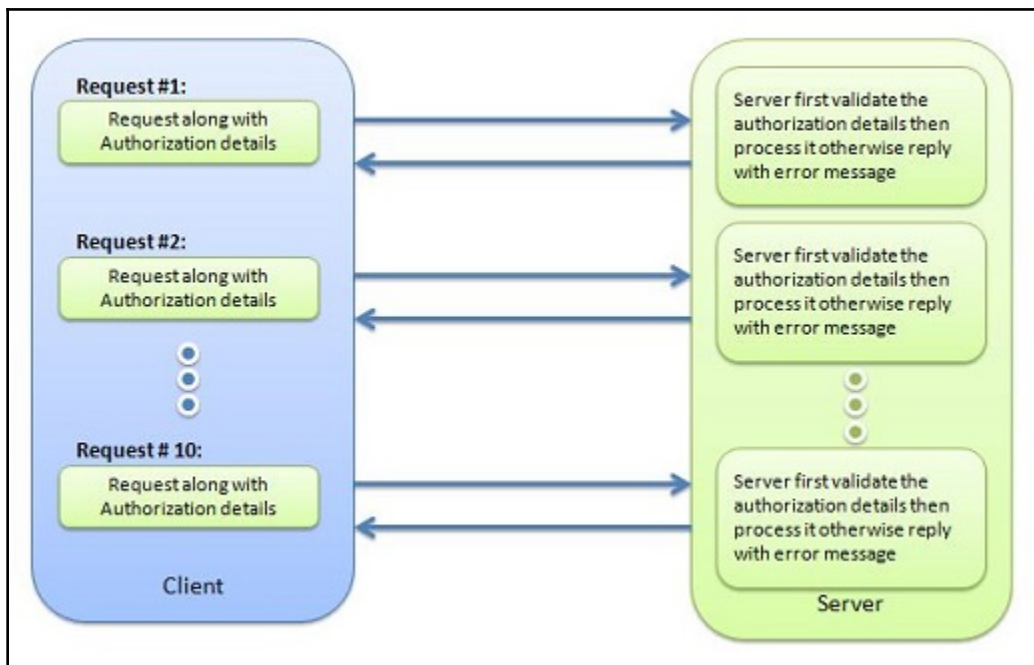
npm WARN optional Skipping failed optional dependency /chokidar/fsevents:
npm WARN notsup Not compatible with your operating system or architecture: fsevents@1.1.3
npm WARN optional Skipping failed optional dependency /react-scripts/fsevents:
npm WARN notsup Not compatible with your operating system or architecture: fsevents@1.1.2
fahadakhtar@hadoopadmin-VirtualBox:~/react/guestbook$ npm install react-redux --save
questbook@0.1.0 /home/fahadakhtar/react/guestbook
├── react-redux@5.0.6
└── hoist-non-react-statics@2.3.1

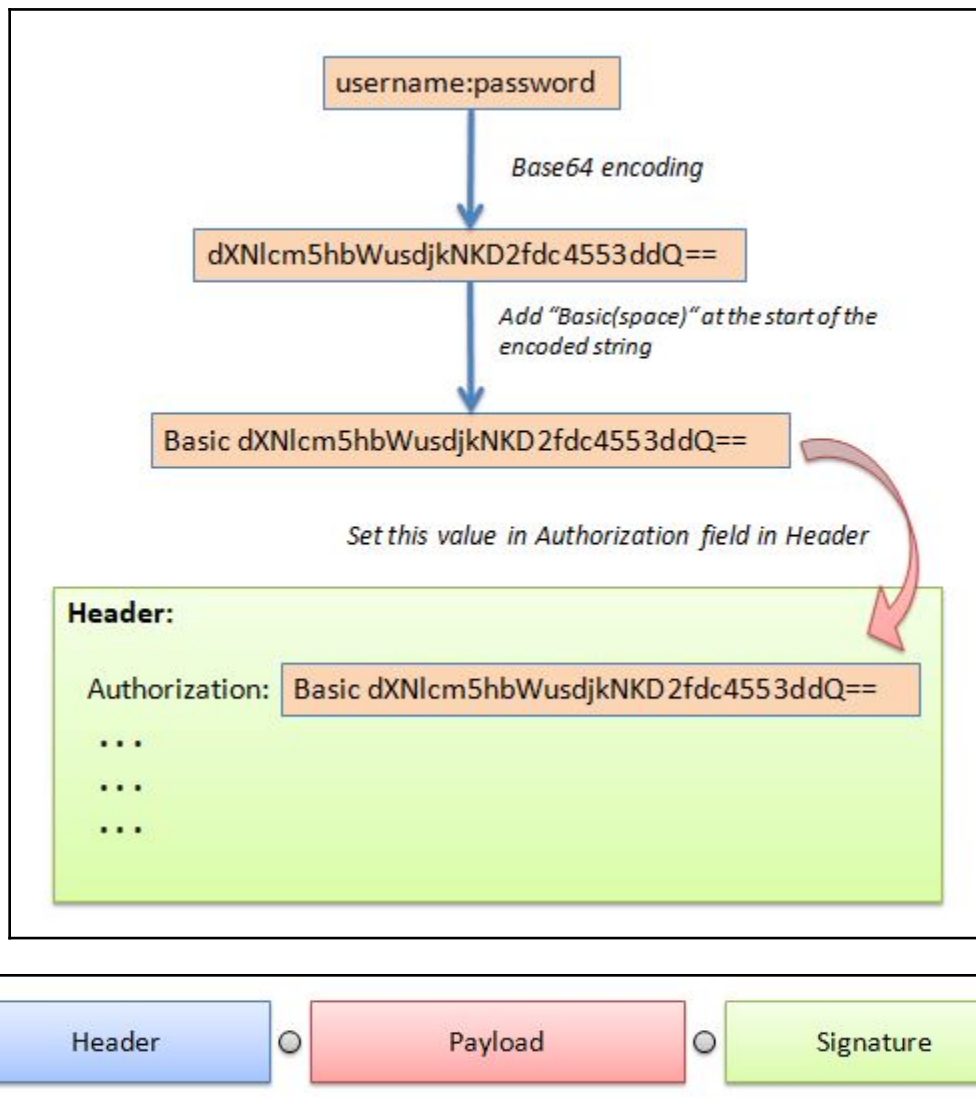
npm WARN optional Skipping failed optional dependency /chokidar/fsevents:
npm WARN notsup Not compatible with your operating system or architecture: fsevents@1.1.3
npm WARN optional Skipping failed optional dependency /react-scripts/fsevents:
npm WARN notsup Not compatible with your operating system or architecture: fsevents@1.1.2
fahadakhtar@hadoopadmin-VirtualBox:~/react/guestbook$
```

Chapter 11: Back End Architecture



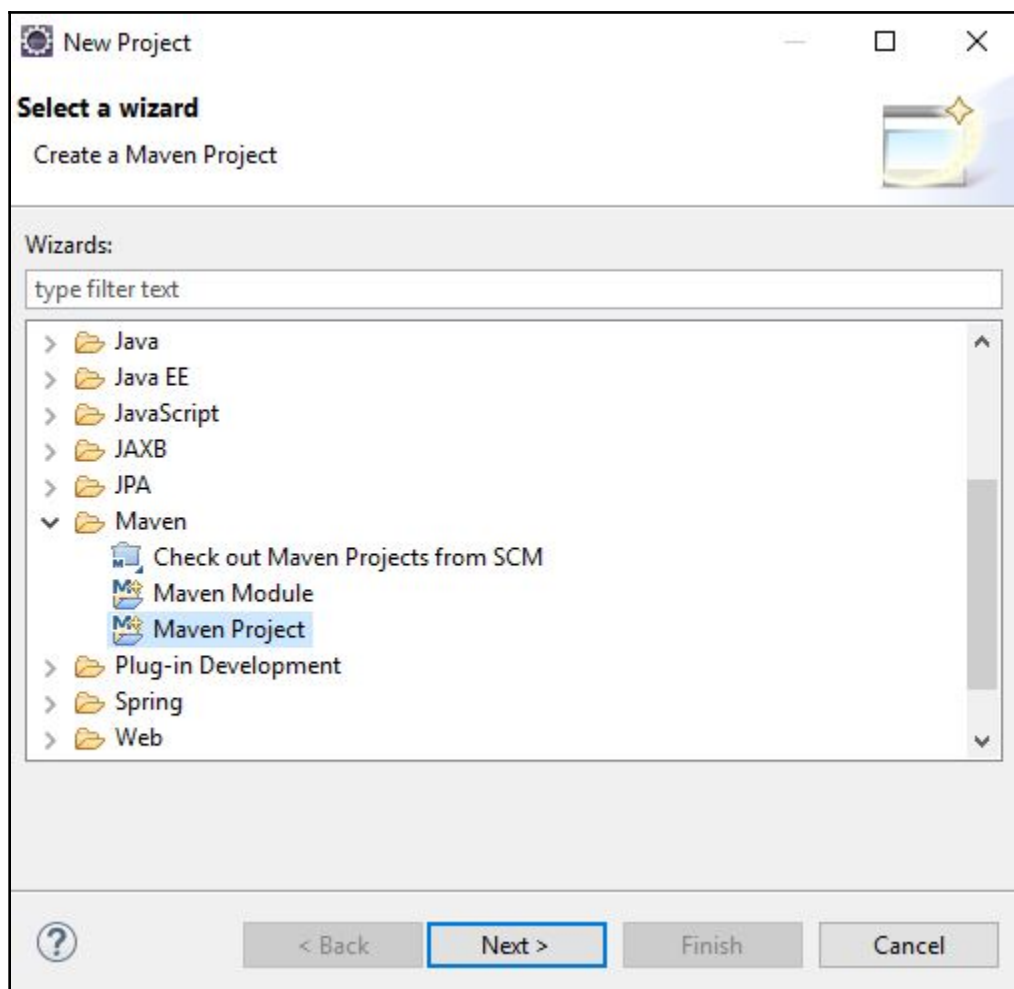


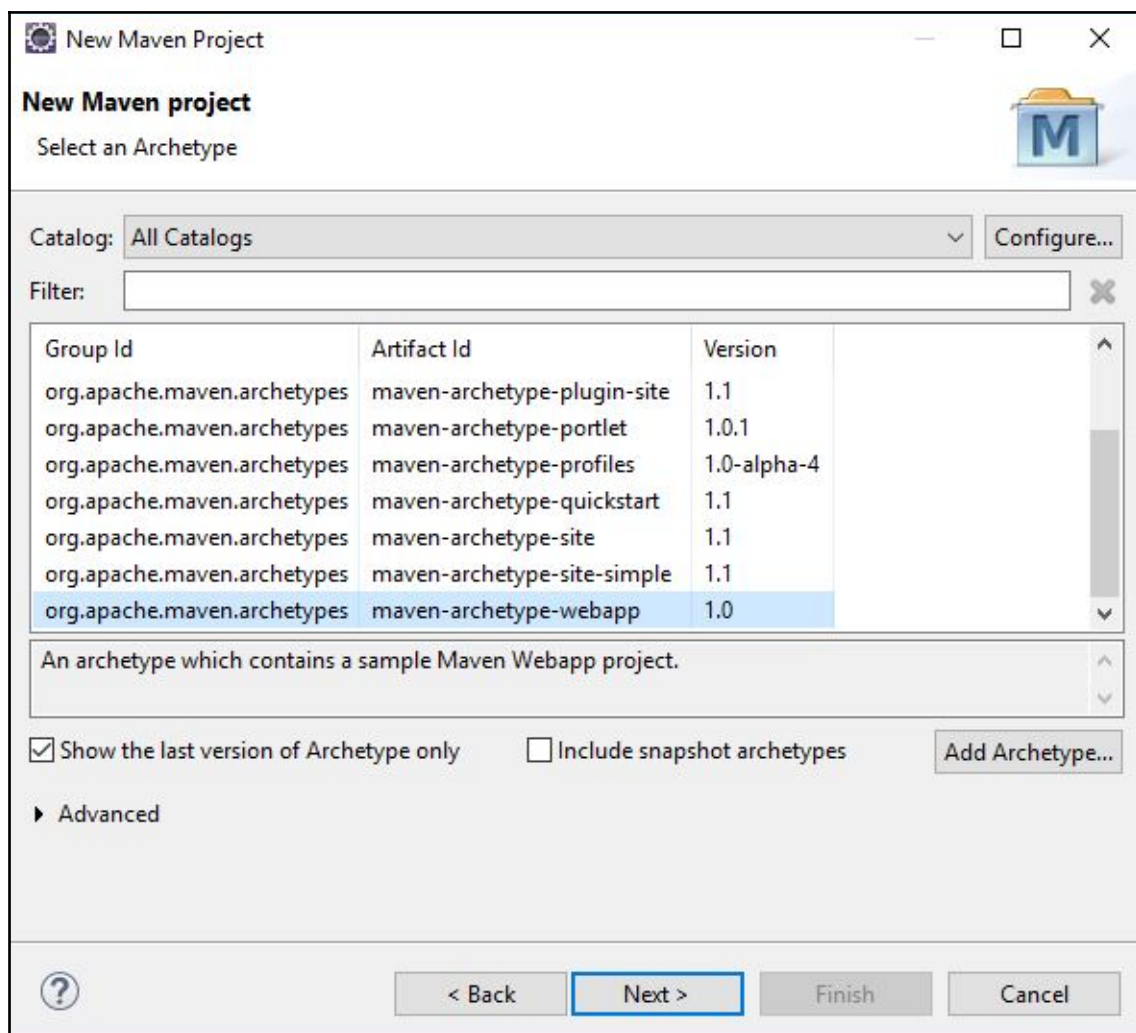


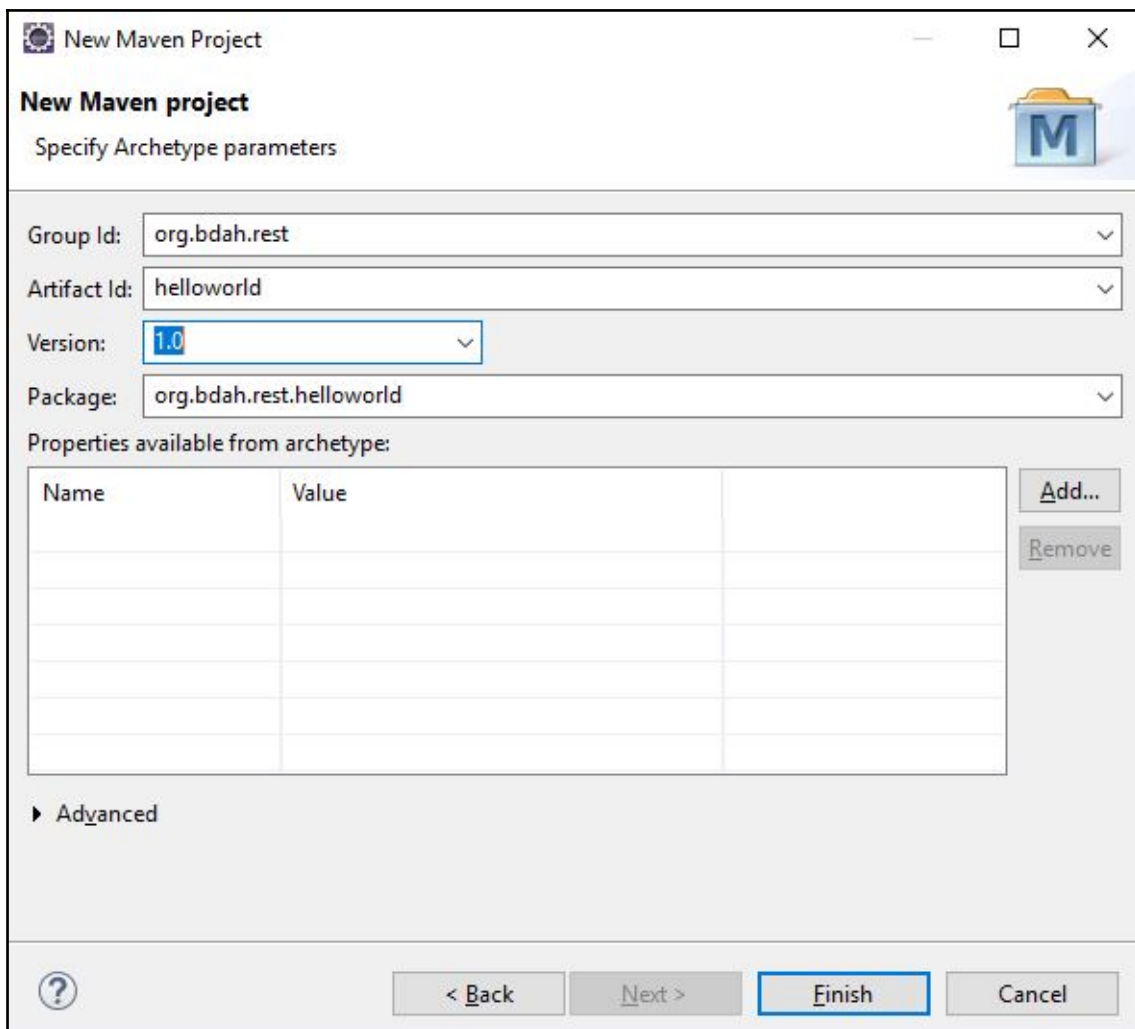
Signature

`hashAlgHs256 ( Base64-encode( header ) + "." + Base64-encode( payload ) , "secret-key" )`

`eyJhbGciOiJIUzI1NiIsInR5cCI6IkpXVCJ9.eyJ0b2tlbklEljoiMTIzNDU2Nzg5
MCIsInVzZXJuYW1lIjoiSm9obiBEb2UiLCJyb2xlIjoiYWRTaW4ifQ.jRaYcM54
ISbLaqd8bRH7bt9zk1J8b8RwCTCrKIA1hIk`







The image shows a 'New Maven Project' dialog box. At the top, it says 'New Maven project' and 'Specify Archetype parameters'. There is a Maven logo icon. Below this, there are four input fields: 'Group Id' with the value 'org.bdah.rest', 'Artifact Id' with the value 'helloworld', 'Version' with the value '1.0', and 'Package' with the value 'org.bdah.rest.helloworld'. Below these fields is a section titled 'Properties available from archetype:' which contains a table with two columns: 'Name' and 'Value'. The table is currently empty. To the right of the table are two buttons: 'Add...' and 'Remove'. Below the table is a section titled 'Advanced' with a right-pointing arrow. At the bottom of the dialog, there is a question mark icon, and four buttons: '< Back', 'Next >', 'Finish' (which is highlighted with a blue border), and 'Cancel'.

New Maven Project

New Maven project

Specify Archetype parameters

Group Id:

Artifact Id:

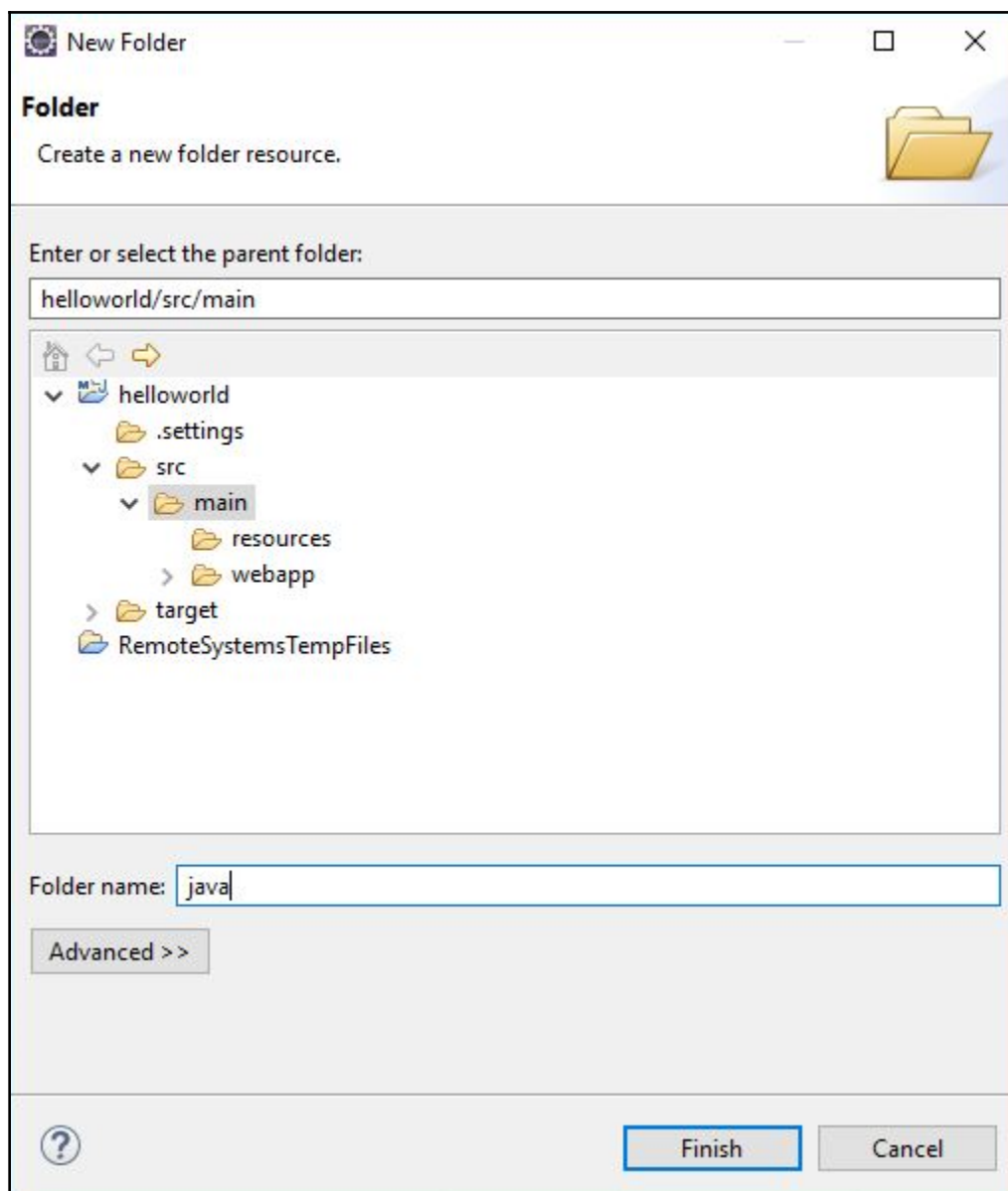
Version:

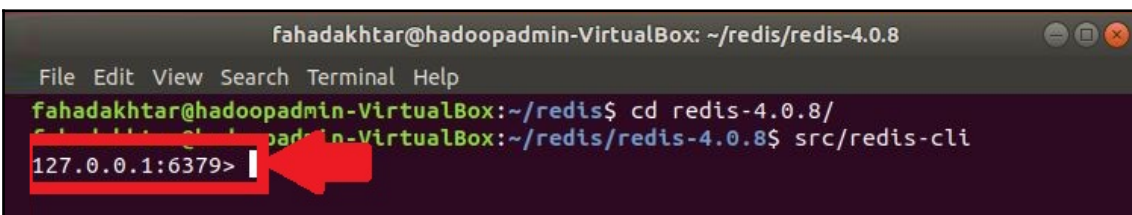
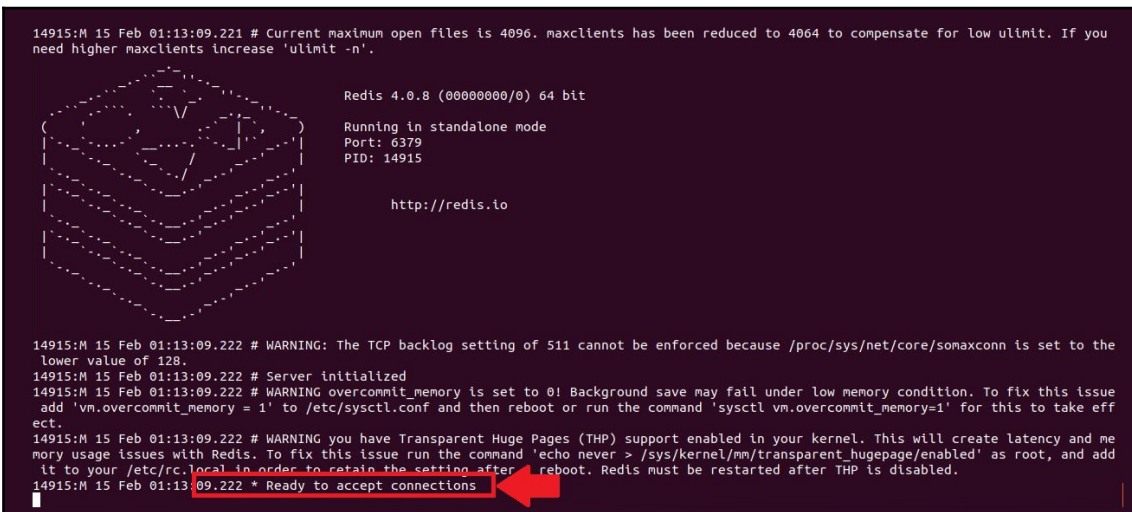
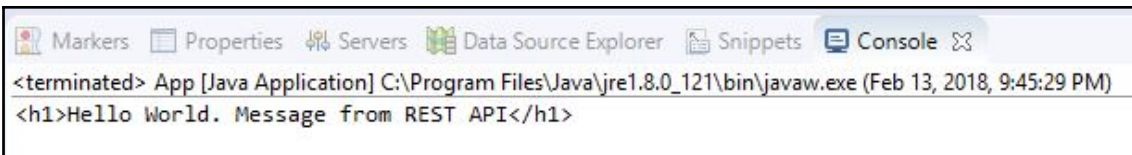
Package:

Properties available from archetype:

Name	Value

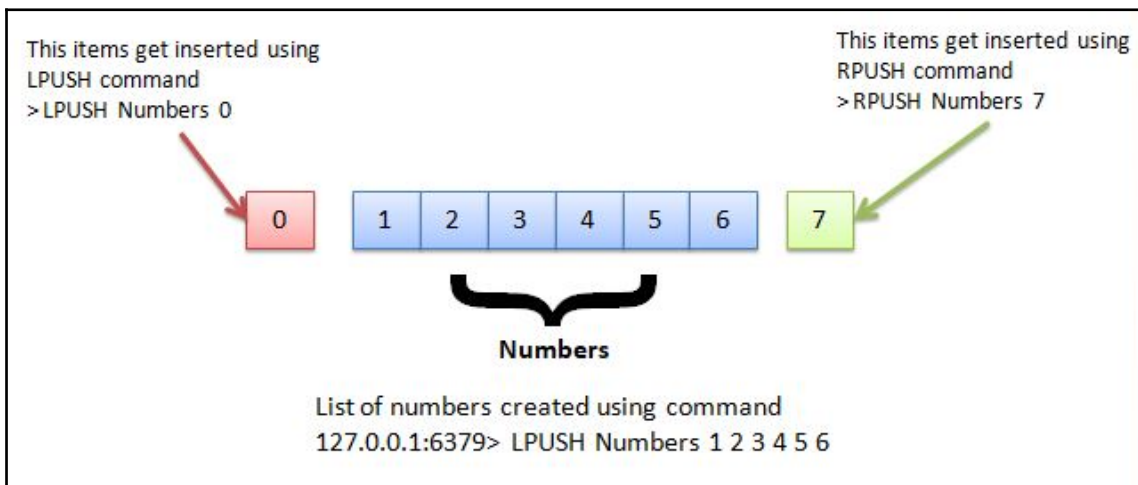
► Advanced





```
fahadakhtar@hadoopadmin-VirtualBox: ~/redis/redis-4.0.8
File Edit View Search Terminal Help
fahadakhtar@hadoopadmin-VirtualBox:~$ cd redis/redis-4.0.8/
fahadakhtar@hadoopadmin-VirtualBox:~/redis/redis-4.0.8$ src/redis-cli
127.0.0.1:6379> SET message "Hello World"
OK
127.0.0.1:6379> GET message
"Hello World"
127.0.0.1:6379>
```

```
fahadakhtar@hadoopadmin-VirtualBox: ~/redis/redis-4.0.8
File Edit View Search Terminal Help
127.0.0.1:6379> HMSET credentials username fahad password akhtar
OK
127.0.0.1:6379> HGETALL credentials
1) "username"
2) "fahad"
3) "password"
4) "akhtar"
127.0.0.1:6379> HGET credentials username
"fahad"
127.0.0.1:6379>
```



```
fahadakhtar@hadoopadmin-VirtualBox: ~/redis/redis-4.0.8
File Edit View Search Terminal Help
127.0.0.1:6379> sadd fruits Apple Mango
(integer) 2
127.0.0.1:6379> sadd fruits Banana
(integer) 1
127.0.0.1:6379> sadd fruits Mango
(integer) 0
127.0.0.1:6379> smembers fruits
1) "Banana"
2) "Apple"
3) "Mango"
127.0.0.1:6379> spop fruits
"Apple"
127.0.0.1:6379>
```

The image displays two Redis terminal windows side-by-side, illustrating the Pub/Sub (Publish/Subscribe) functionality.

Left Window (Subscriber):

```
fahadakhtar@hadoopadmin-VirtualBox: ~/redis/redis-4.0.8
File Edit View Search Terminal Help
Reading messages... (press Ctrl-C to quit)
1) "subscribe"
2) "LatestNews"
3) (integer) 1
1) "message"
2) "LatestNews"
3) "This is news bulletin"
```

Annotations for the left window:

- A red box highlights the output of the `smembers` command, showing the list of subscribers: `1) "message"`, `2) "LatestNews"`, and `3) "This is news bulletin"`.
- A red arrow points from the text "New Message received automatically" to the output of the `smembers` command.

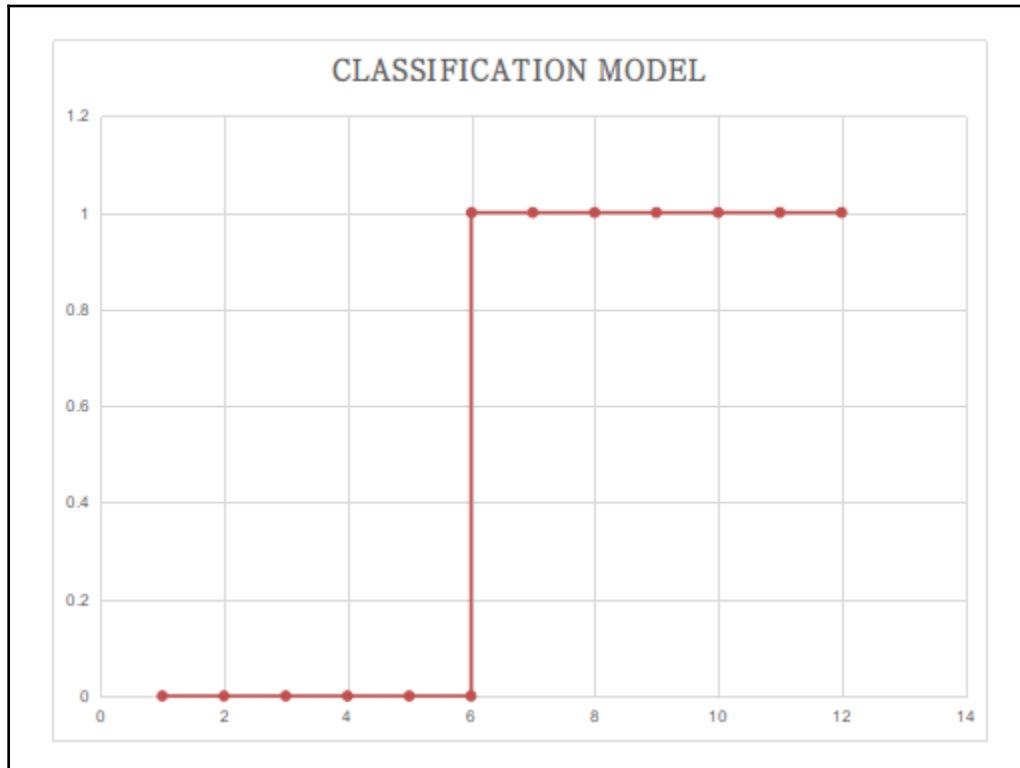
Right Window (Publisher):

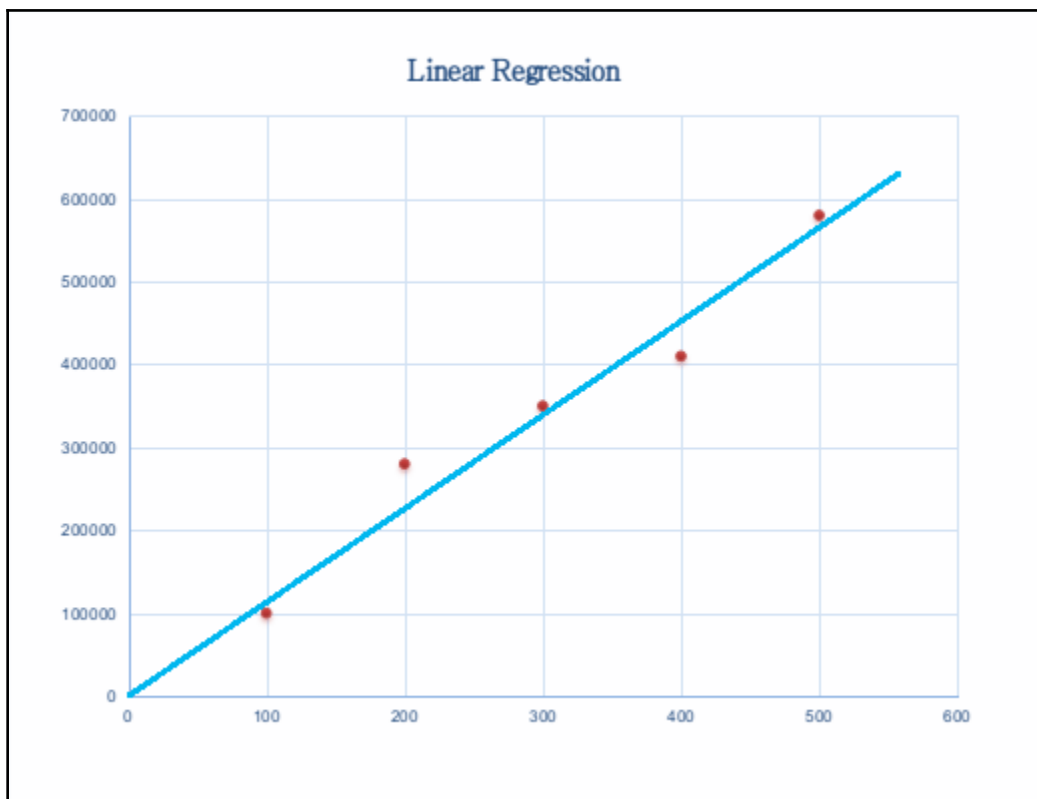
```
fahadakhtar@hadoopadmin-VirtualBox: ~/redis/redis-4.0.8
File Edit View Search Terminal Help
127.0.0.1:6379> clear
127.0.0.1:6379> PUBLISH LatestNews "This is news bulletin"
(integer) 2
127.0.0.1:6379>
```

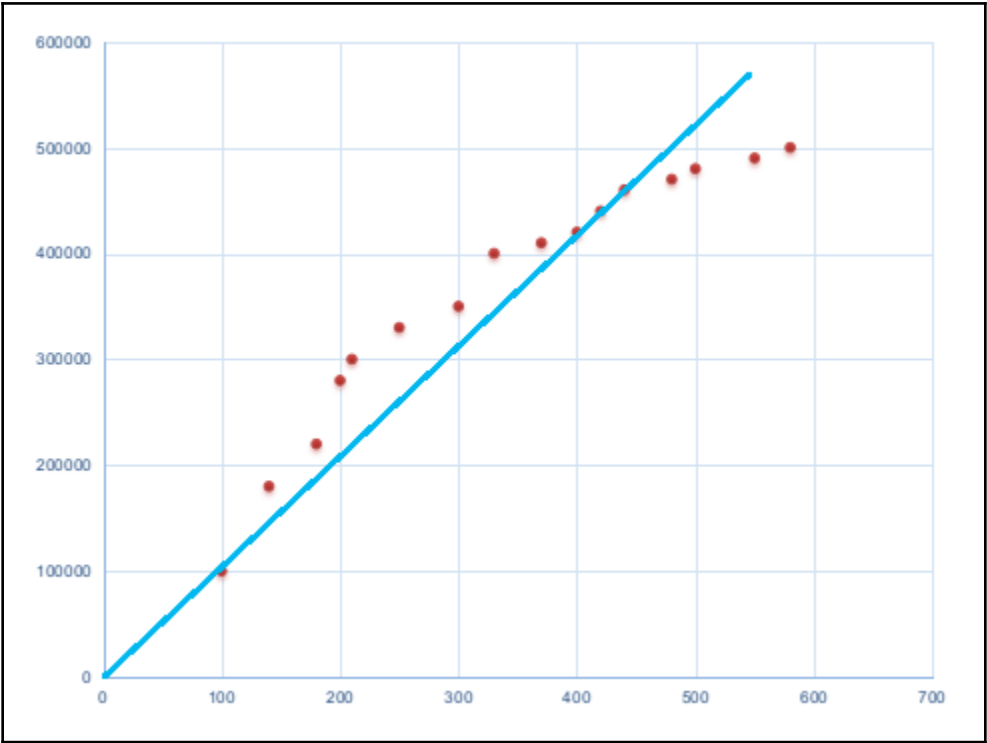
Annotations for the right window:

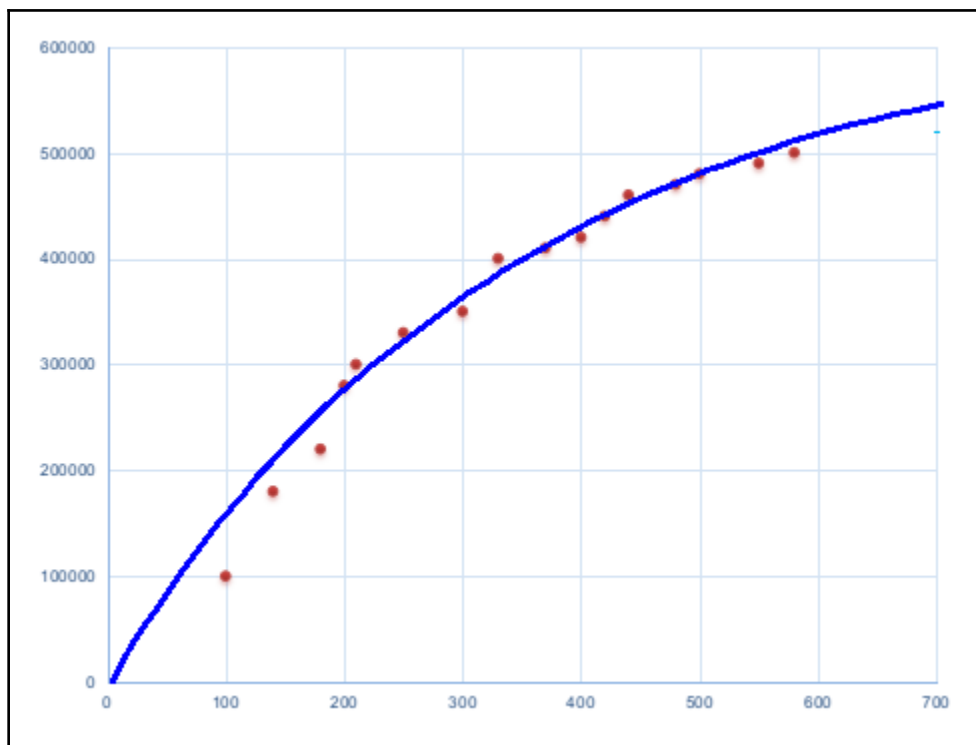
- A red box highlights the output of the `PUBLISH` command, showing the number of subscribers: `(integer) 2`.
- A red arrow points from the text "Number of Subscribers" to the output of the `PUBLISH` command.
- A red arrow points from the text "Publish a new message" to the `PUBLISH` command itself.

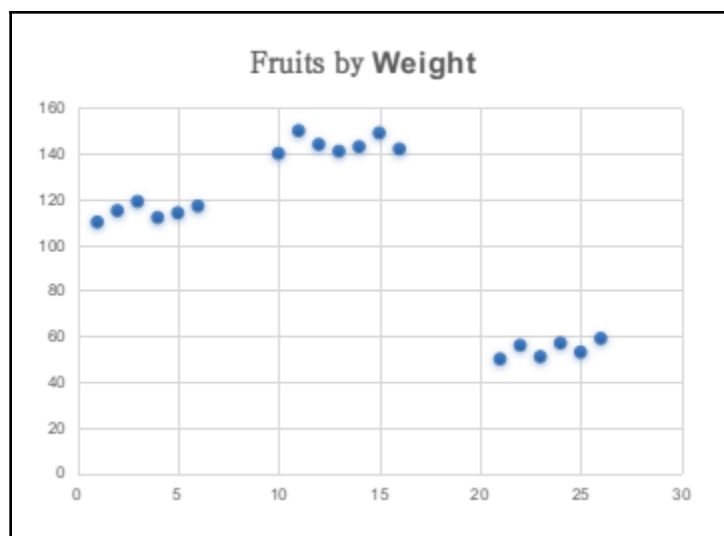
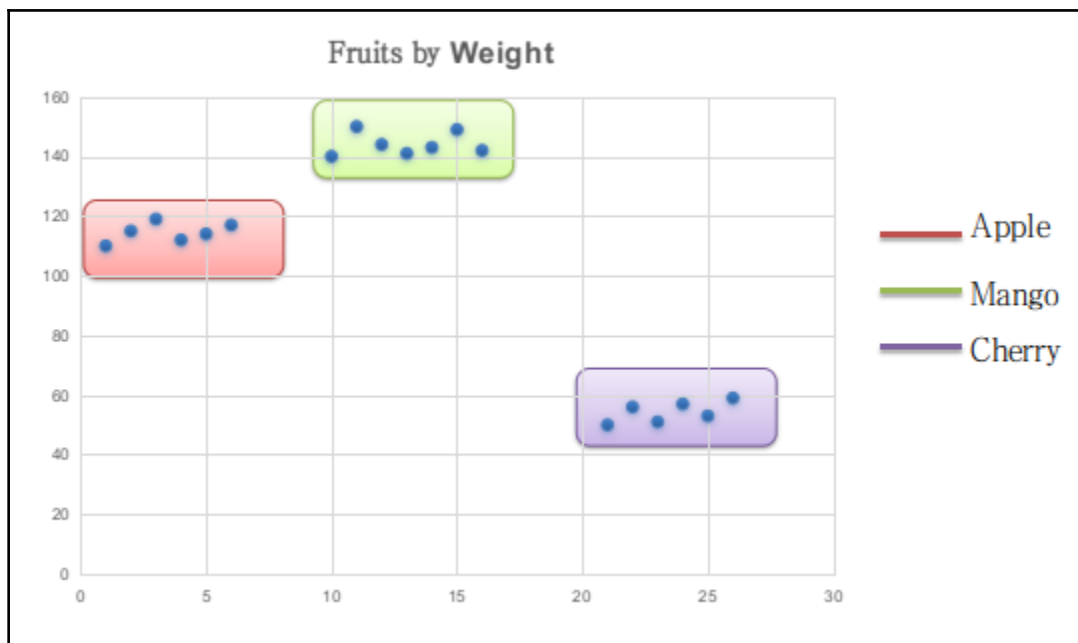
Chapter 12: Machine Learning

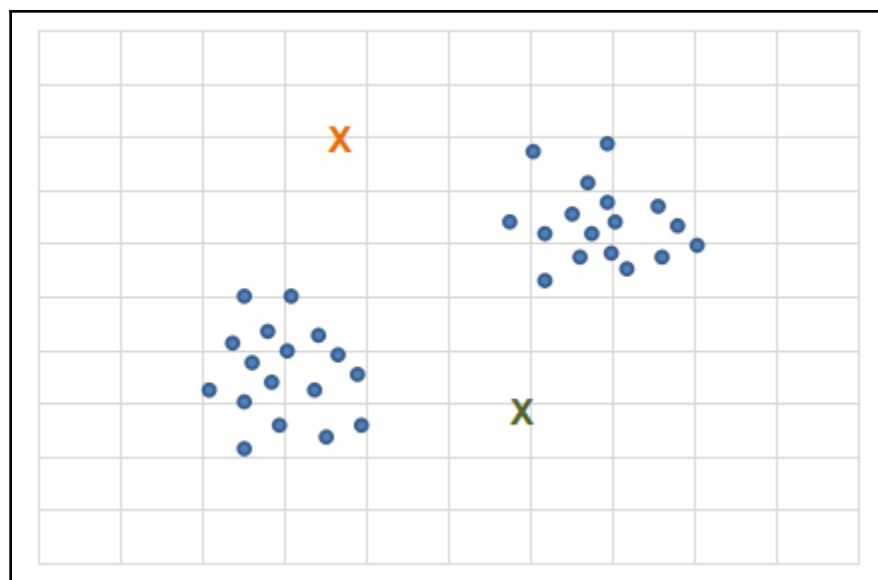
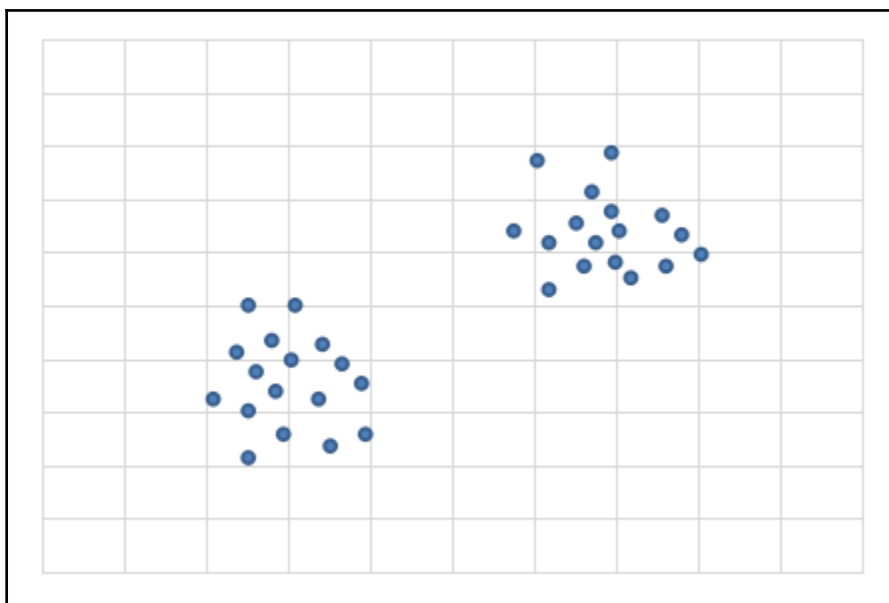


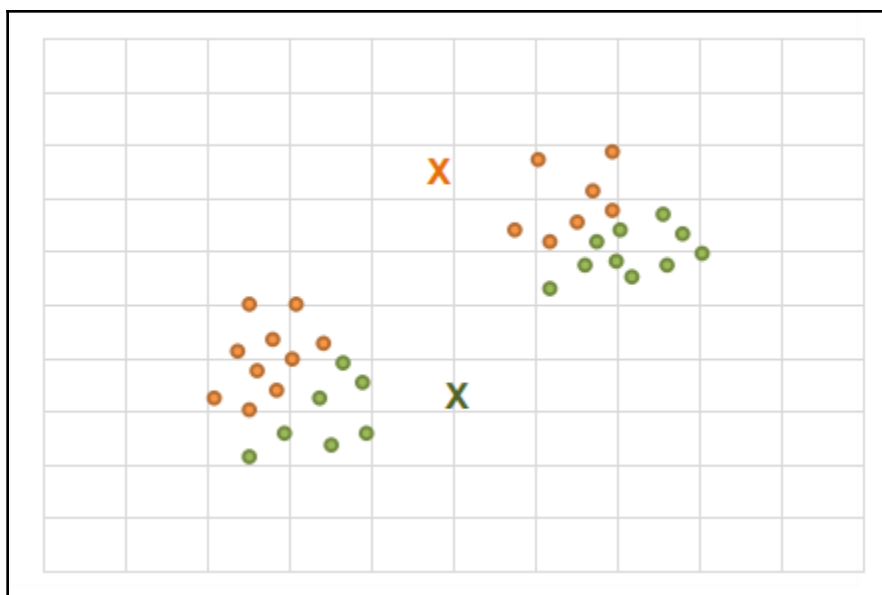
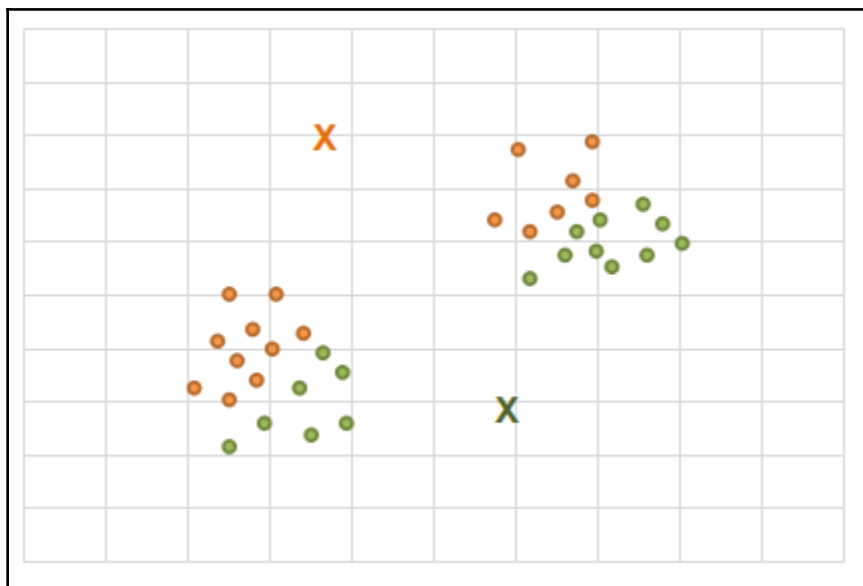


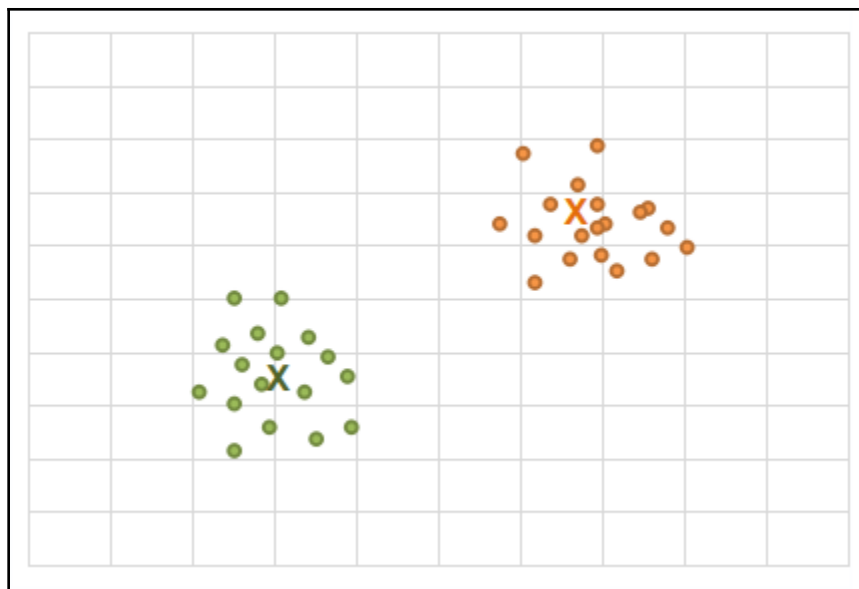
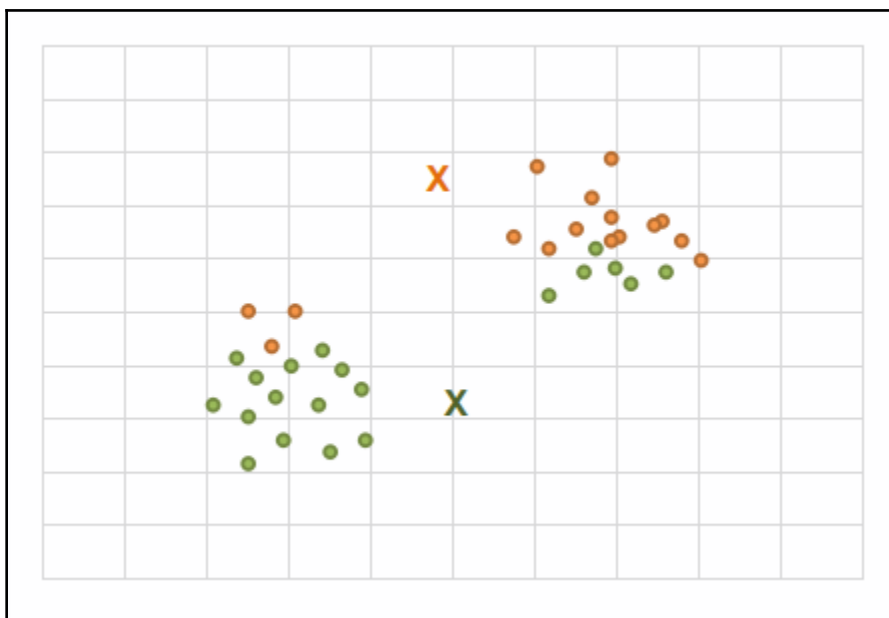


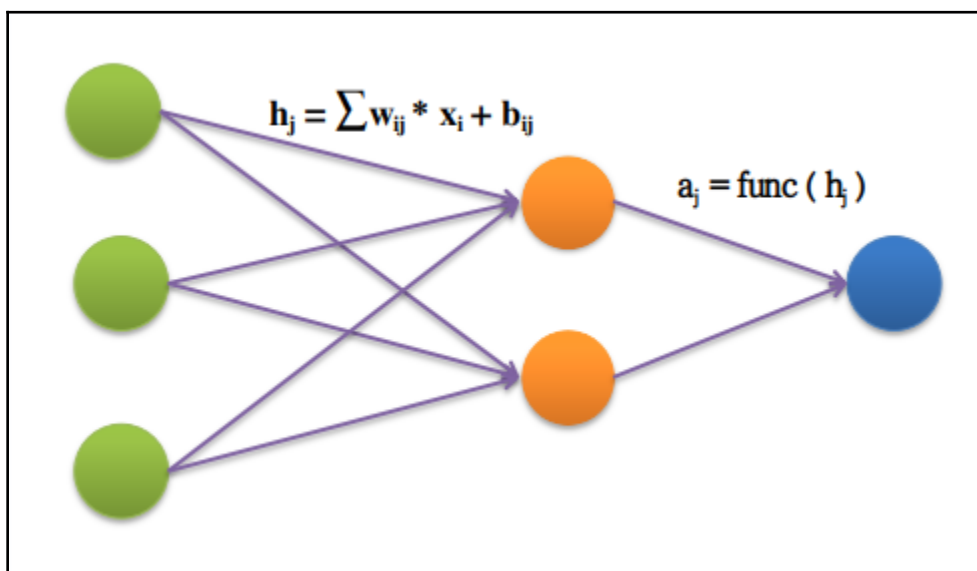
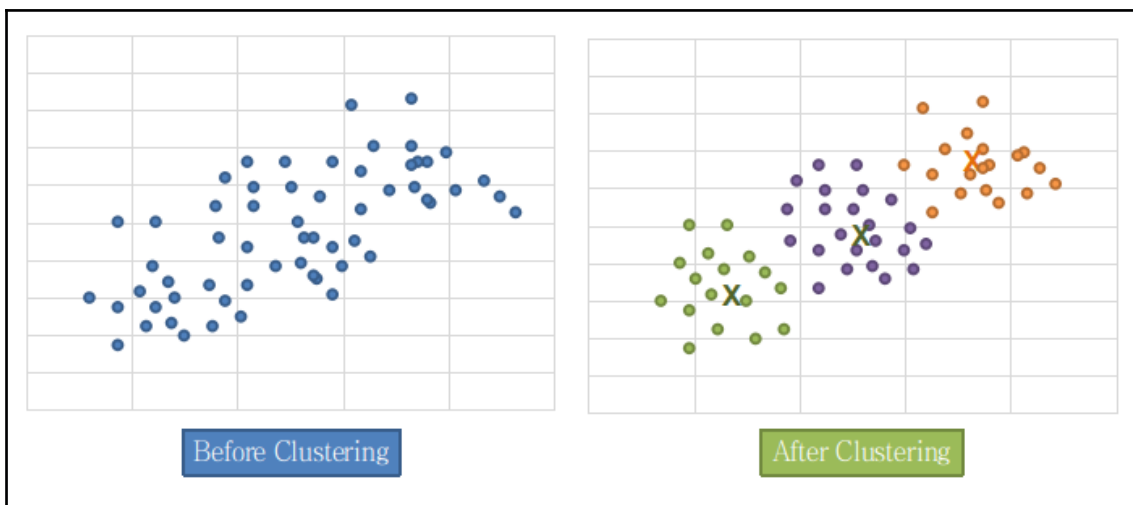


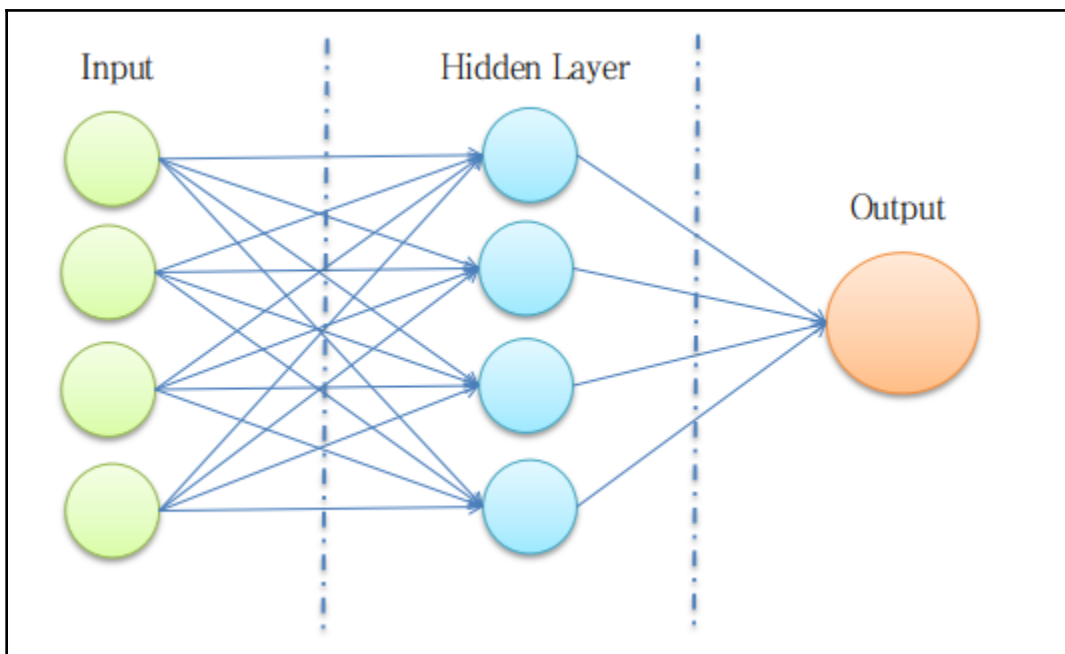


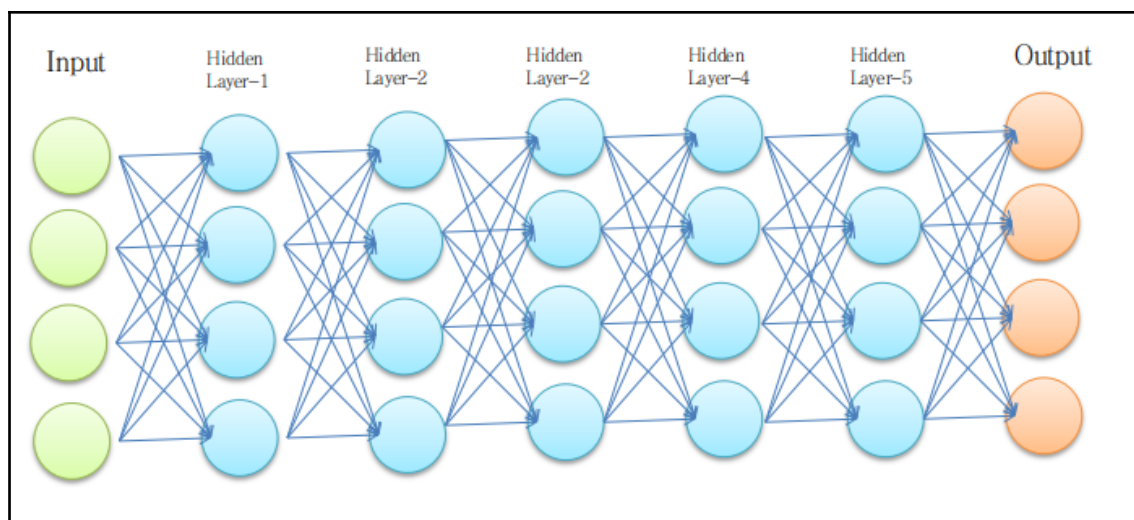
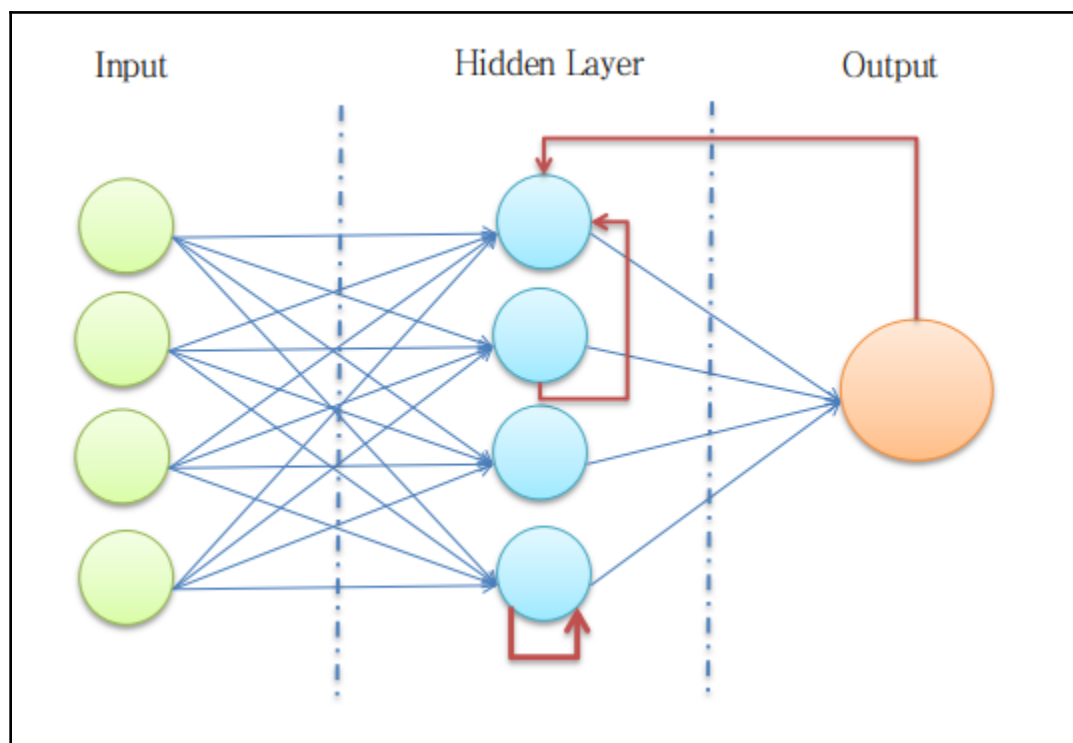


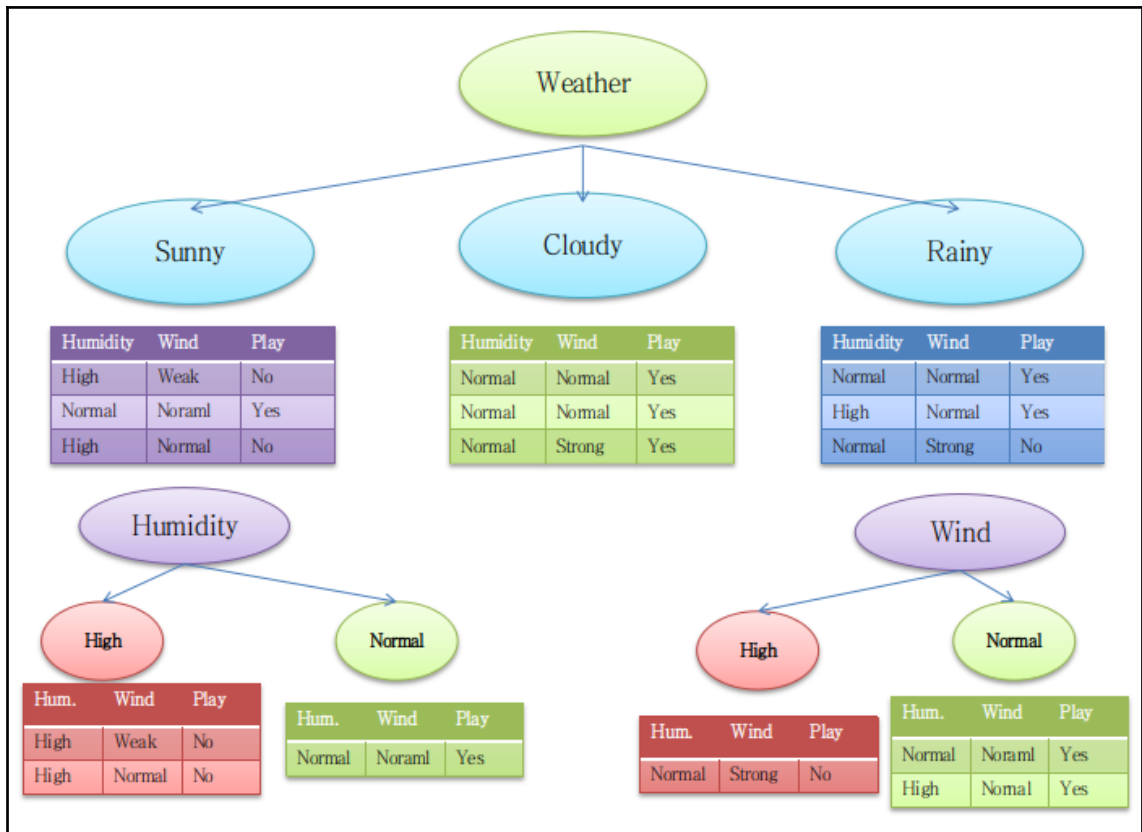




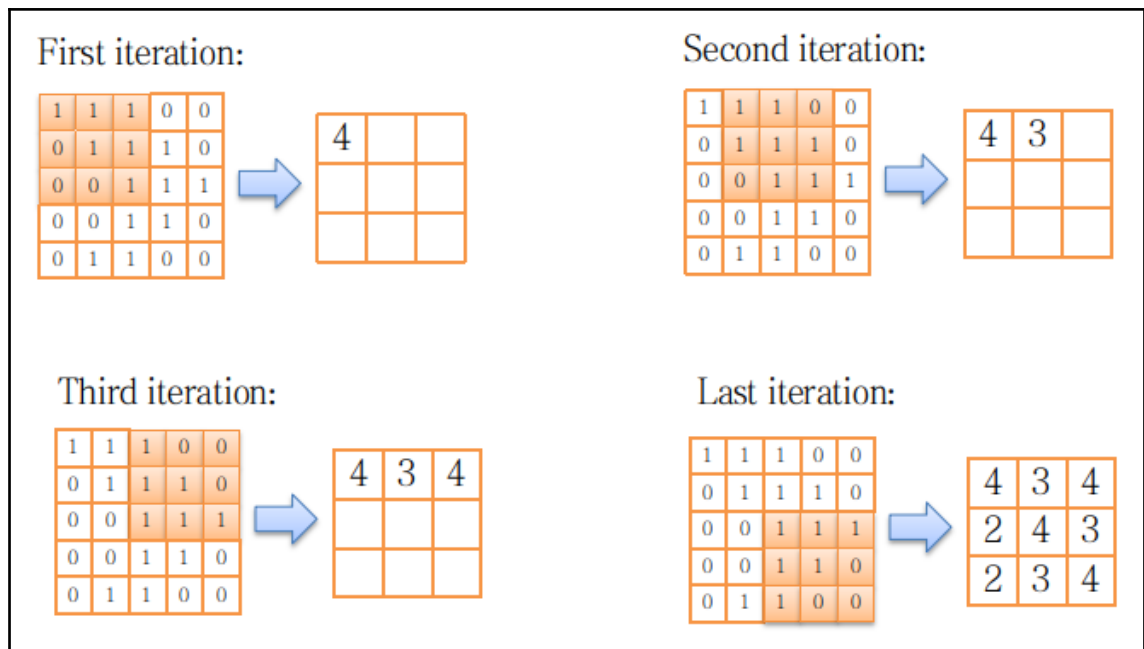
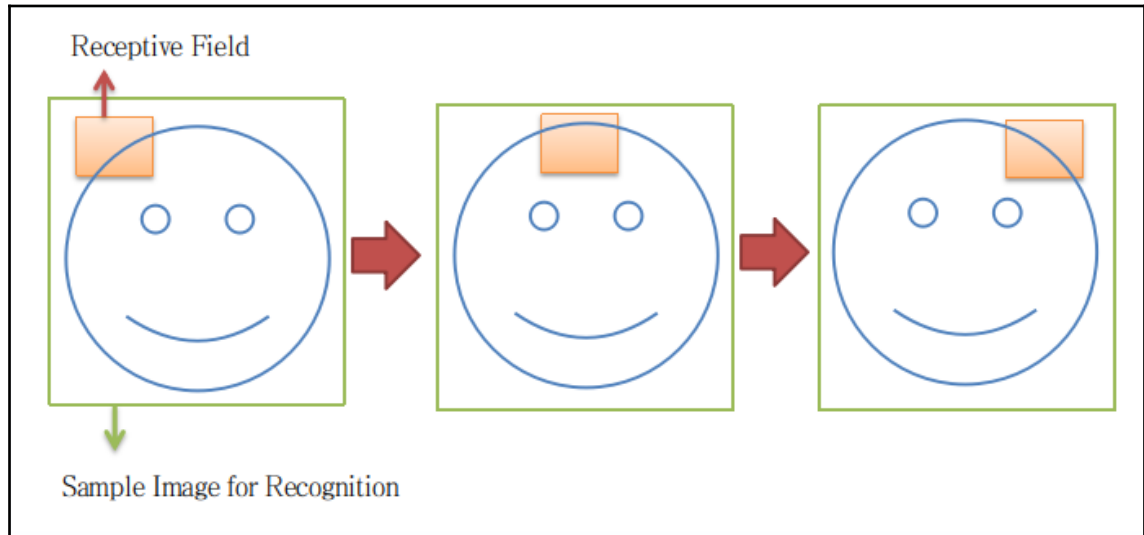


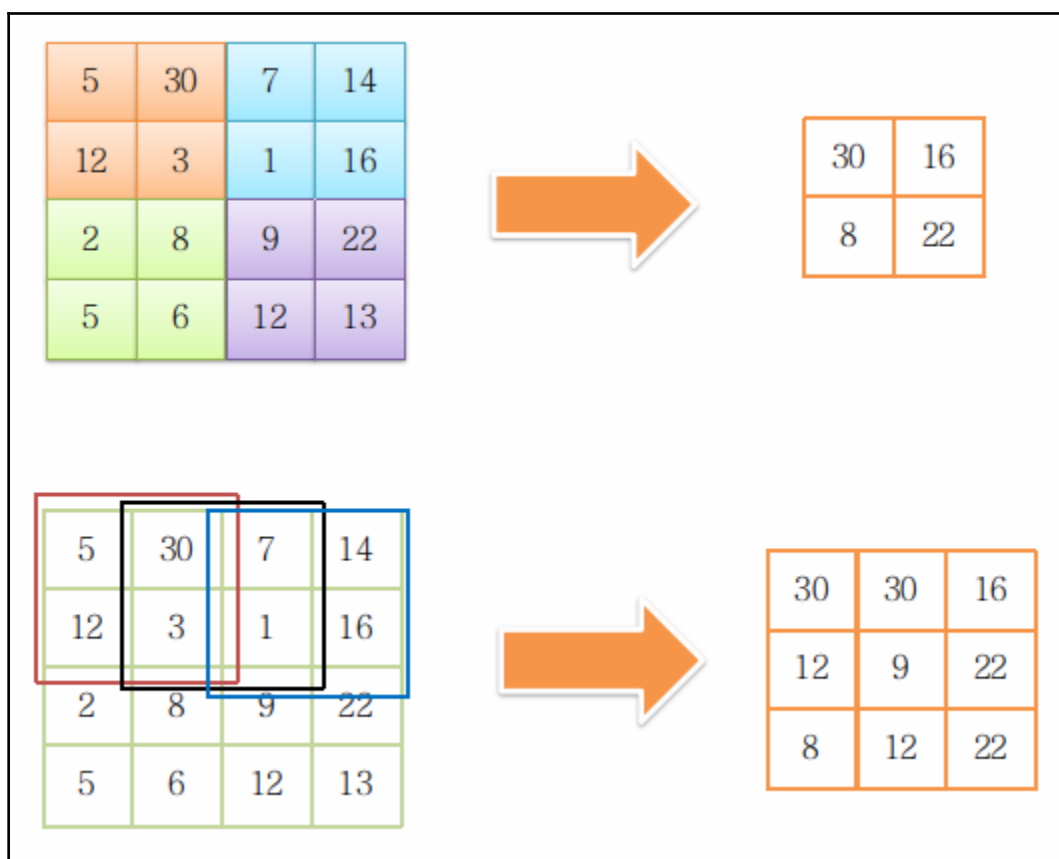
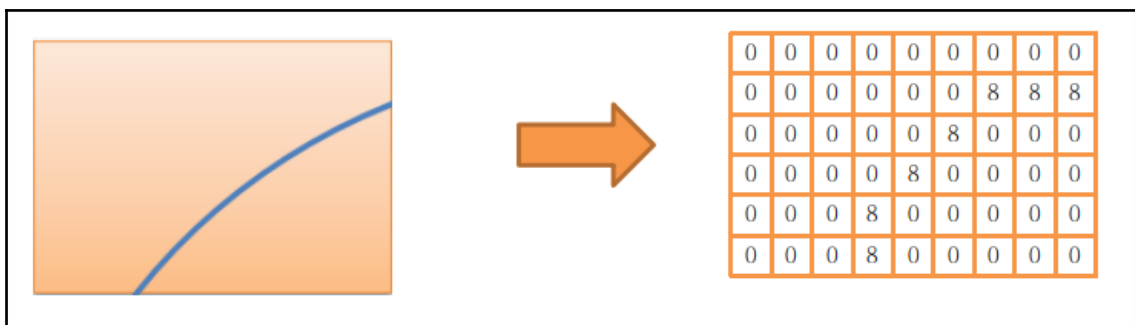






Chapter 13: Artificial Intelligence





```
fahad@faha

File Edit View Search Terminal Help
fahad@fahad-pc:~$ python --version
Python 2.7.14
fahad@fahad-pc:~$
```

```
fahad@fahad-pc: ~/tensorflow

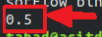
File Edit View Search Terminal Help
fahad@fahad-pc:~/tensorflow$ python3 tensor.py
2018-03-11 17:26:18.536395: I tensorflow/core/platform/cpu_feature_guard.cc:140] Your CPU supports ins
tructions that this TensorFlow binary was not compiled to use: AVX2 FMA
'Hello world. I have successfully installed TensorFlow!'
fahad@fahad-pc:~/tensorflow$
```



$$Result = (A * B) + (C/B) * A - C$$

```
fahad@asitdev-pc: ~/tensorflow

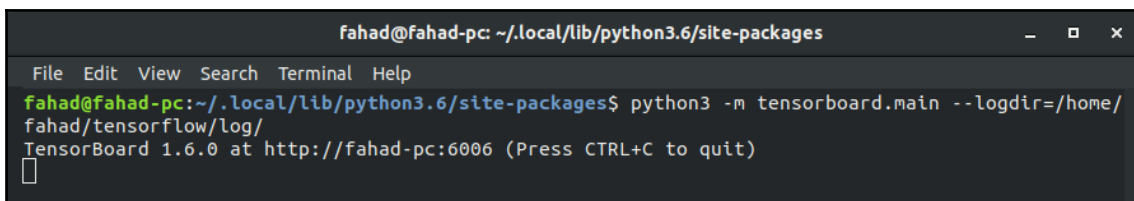
File Edit View Search Terminal Help
fahad@asitdev-pc:~$ cd tensorflow/
fahad@asitdev-pc:~/tensorflow$ python3 tensorboard.py
2018-03-17 23:00:31.999394: I tensorflow/core/platform/cpu_feature_guard.cc:140] Your CPU supports instructions that this Ten
sorFlow binary was not compiled to use: AVX2 FMA
0.5
fahad@asitdev-pc:~/tensorflow$
```



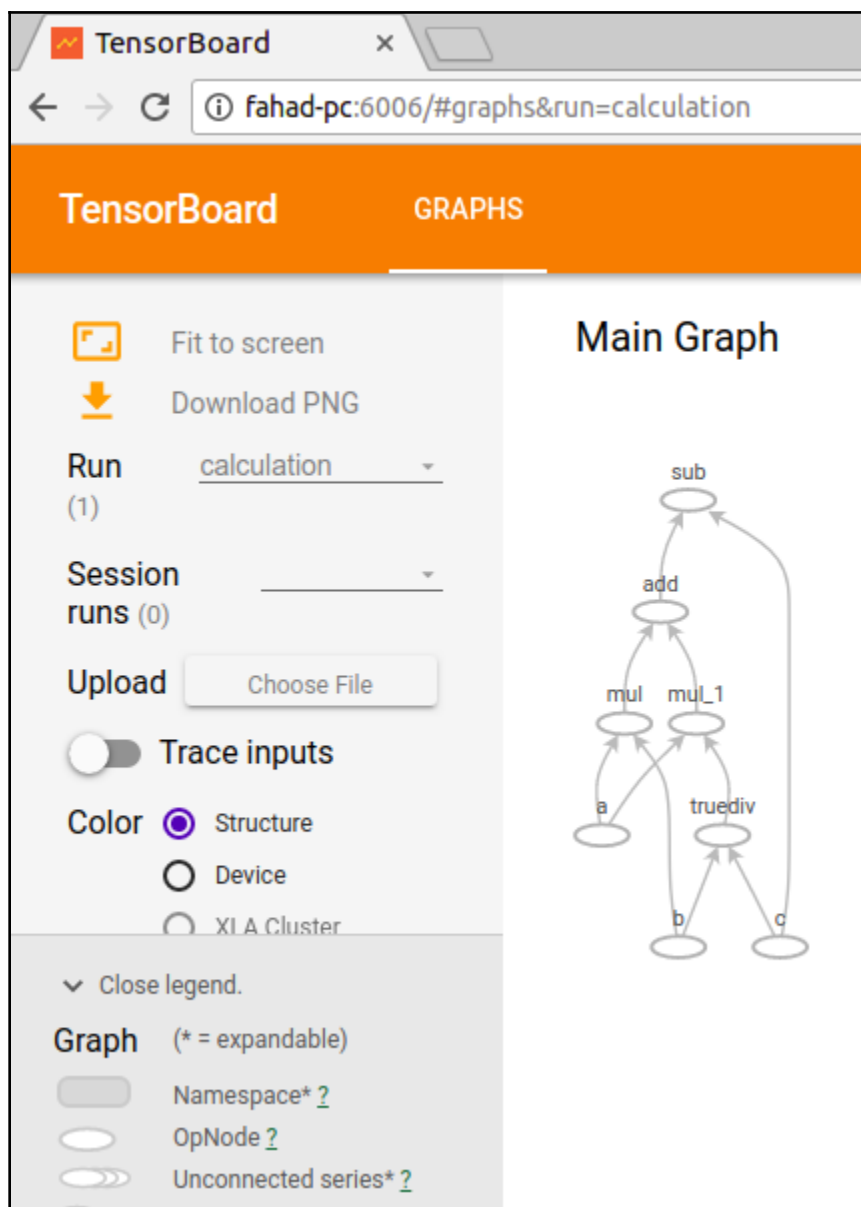
```
fahad@fahad-pc: ~

File Edit View Search Terminal Help
fahad@fahad-pc:~$ pip3 show tensorflow
Name: tensorflow
Version: 1.6.0
Summary: TensorFlow helps the tensors flow
Home-page: https://www.tensorflow.org/
Author: Google Inc.
Author-email: opensource@google.com
License: Apache 2.0
Location: /home/fahad/.local/lib/python3.6/site-packages
Requires: termcolor, six, wheel, tensorboard, numpy, protobuf, gast, astor, grpcio, absl-py
fahad@fahad-pc:~$
```



A terminal window with a dark background. The title bar reads 'fahad@fahad-pc: ~/.local/lib/python3.6/site-packages'. The menu bar includes 'File', 'Edit', 'View', 'Search', 'Terminal', and 'Help'. The prompt is 'fahad@fahad-pc:~/.local/lib/python3.6/site-packages\$'. The command entered is 'python3 -m tensorboard.main --logdir=/home/fahad/tensorflow/log/'. The output is 'TensorBoard 1.6.0 at http://fahad-pc:6006 (Press CTRL+C to quit)' followed by a cursor.

```
fahad@fahad-pc: ~/.local/lib/python3.6/site-packages
File Edit View Search Terminal Help
fahad@fahad-pc:~/.local/lib/python3.6/site-packages$ python3 -m tensorboard.main --logdir=/home/
fahad/tensorflow/log/
TensorBoard 1.6.0 at http://fahad-pc:6006 (Press CTRL+C to quit)
█
```



```

fahad@fahad-pc: ~/YOLO
File Edit View Search Terminal Help
fahad@fahad-pc:~/YOLO$ git clone https://github.com/pjreddie/darknet
Cloning into 'darknet'...
remote: Counting objects: 5453, done.
remote: Total 5453 (delta 0), reused 0 (delta 0), pack-reused 5453
Receiving objects: 100% (5453/5453), 4.42 MiB | 178.00 KiB/s, done.
Resolving deltas: 100% (3644/3644), done.
fahad@fahad-pc:~/YOLO$

```

```

fahad@fahad-pc: ~/YOLO/darknet
File Edit View Search Terminal Help
gcc -Iinclude/ -Isrc/ -Wall -Wno-unknown-pragmas -Wfatal-errors -fPIC -Ofast -c ./examples/segmenter.c -o obj/segmenter.o
gcc -Iinclude/ -Isrc/ -Wall -Wno-unknown-pragmas -Wfatal-errors -fPIC -Ofast -c ./examples/regressor.c -o obj/regressor.o
gcc -Iinclude/ -Isrc/ -Wall -Wno-unknown-pragmas -Wfatal-errors -fPIC -Ofast -c ./examples/classifier.c -o obj/classifier.o
gcc -Iinclude/ -Isrc/ -Wall -Wno-unknown-pragmas -Wfatal-errors -fPIC -Ofast -c ./examples/coco.c -o obj/coco.o
gcc -Iinclude/ -Isrc/ -Wall -Wno-unknown-pragmas -Wfatal-errors -fPIC -Ofast -c ./examples/yolo.c -o obj/yolo.o
gcc -Iinclude/ -Isrc/ -Wall -Wno-unknown-pragmas -Wfatal-errors -fPIC -Ofast -c ./examples/detector.c -o obj/detector.o
gcc -Iinclude/ -Isrc/ -Wall -Wno-unknown-pragmas -Wfatal-errors -fPIC -Ofast -c ./examples/nightmare.c -o obj/nightmare.o
gcc -Iinclude/ -Isrc/ -Wall -Wno-unknown-pragmas -Wfatal-errors -fPIC -Ofast -c ./examples/attention.c -o obj/attention.o
./examples/attention.c: In function 'network_loss_data':
./examples/attention.c:43:17: warning: unused variable 't' [-Wunused-variable]
    int t = max_index(y + b*test.y.cols, 1000);
        ^
./examples/attention.c: In function 'train_attention':
./examples/attention.c:179:15: warning: unused variable 'im' [-Wunused-variable]
    image im = float_to_image(64,64,3,resized.X.vals[0]);
        ^
./examples/attention.c: In function 'run_attention':
./examples/attention.c:452:11: warning: unused variable 'layer_s' [-Wunused-variable]
    char *layer_s = (argc > 7) ? argv[7]: 0;
        ^
gcc -Iinclude/ -Isrc/ -Wall -Wno-unknown-pragmas -Wfatal-errors -fPIC -Ofast -c ./examples/darknet.c -o obj/darknet.o
gcc -Iinclude/ -Isrc/ -Wall -Wno-unknown-pragmas -Wfatal-errors -fPIC -Ofast -c obj/captcha.o obj/lsd.o obj/super.o obj/art.o obj/tag.o obj/cifar.o obj/go.o obj/rnn.o obj/segmenter.o obj/regressor.o obj/classifier.o obj/coco.o obj/yolo.o obj/detector.o obj/nightmare.o obj/attention.o obj/darknet.o libdarknet.a -o darknet -lm -pthread libdarknet.a
fahad@fahad-pc:~/YOLO/darknet$

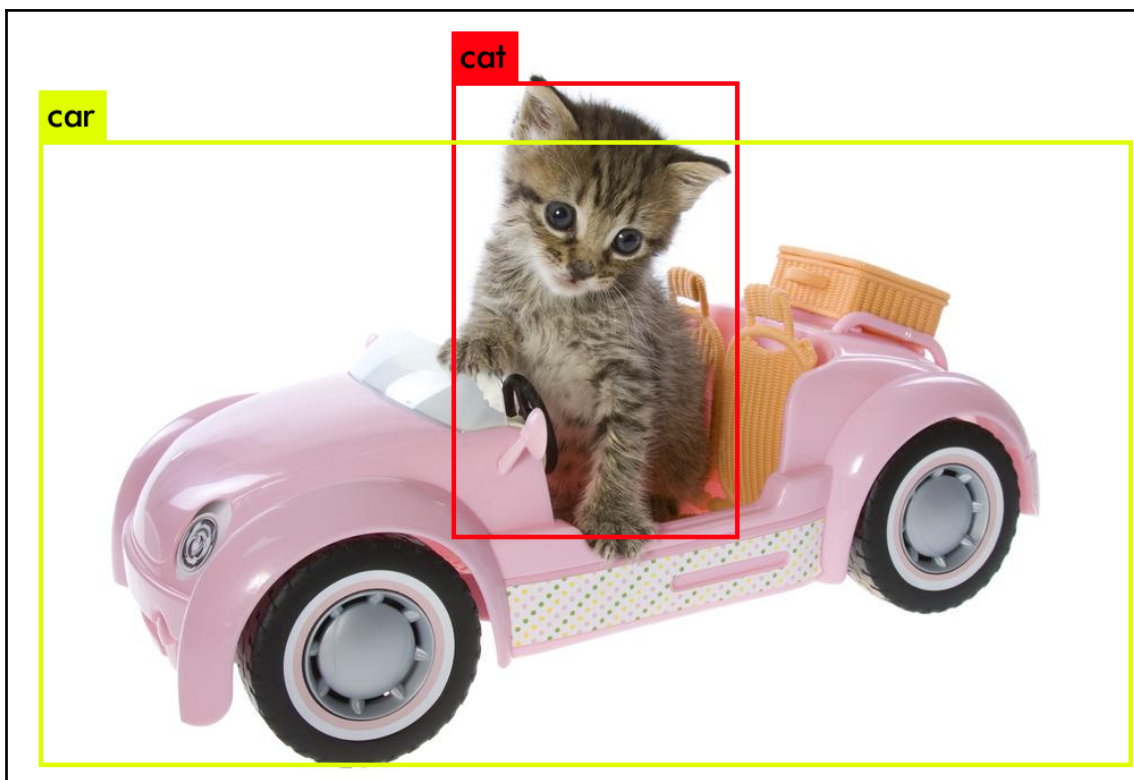
```



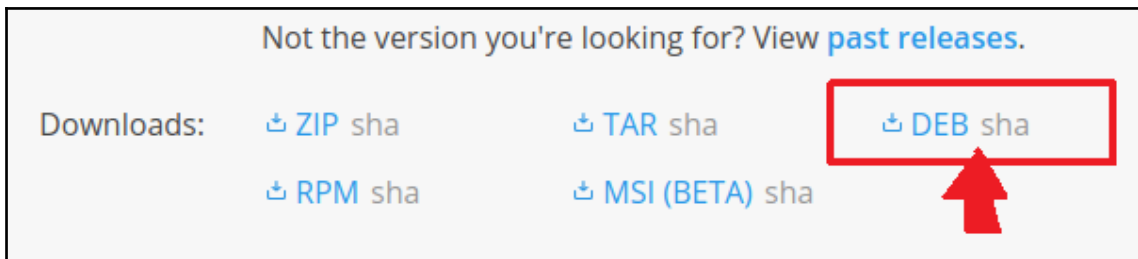
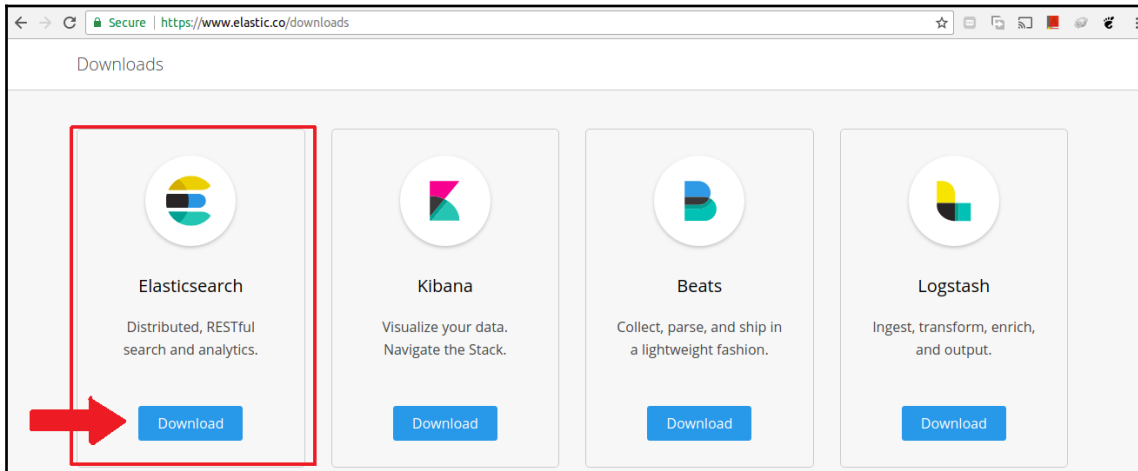
```

File Edit View Search Terminal Help
fahad@fahad-pc:~/YOLO/darknet$ ./darknet detect cfg/yolo.cfg yolo.weights data/sample.jpg
layer      filters  size      input           output
 0 conv     32  3 x 3 / 1  608 x 608 x 3  -> 608 x 608 x 32
 1 max       2  2 x 2 / 2  608 x 608 x 32  -> 304 x 304 x 32
 2 conv     64  3 x 3 / 1  304 x 304 x 32  -> 304 x 304 x 64
 3 max       2  2 x 2 / 2  304 x 304 x 64  -> 152 x 152 x 64
 4 conv    128  3 x 3 / 1  152 x 152 x 64  -> 152 x 152 x 128
 5 conv     64  1 x 1 / 1  152 x 152 x 128 -> 152 x 152 x 64
 6 conv    128  3 x 3 / 1  152 x 152 x 64  -> 152 x 152 x 128
 7 max       2  2 x 2 / 2  152 x 152 x 128 -> 76 x 76 x 128
 8 conv    256  3 x 3 / 1  76 x 76 x 128   -> 76 x 76 x 256
 9 conv    128  1 x 1 / 1  76 x 76 x 256   -> 76 x 76 x 128
10 conv    256  3 x 3 / 1  76 x 76 x 128   -> 76 x 76 x 256
11 max       2  2 x 2 / 2  76 x 76 x 256   -> 38 x 38 x 256
12 conv    512  3 x 3 / 1  38 x 38 x 256   -> 38 x 38 x 512
13 conv    256  1 x 1 / 1  38 x 38 x 512   -> 38 x 38 x 256
14 conv    512  3 x 3 / 1  38 x 38 x 256   -> 38 x 38 x 512
15 conv    256  1 x 1 / 1  38 x 38 x 512   -> 38 x 38 x 256
16 conv    512  3 x 3 / 1  38 x 38 x 256   -> 38 x 38 x 512
17 max       2  2 x 2 / 2  38 x 38 x 512   -> 19 x 19 x 512
18 conv   1024  3 x 3 / 1  19 x 19 x 512   -> 19 x 19 x1024
19 conv    512  1 x 1 / 1  19 x 19 x1024   -> 19 x 19 x 512
20 conv   1024  3 x 3 / 1  19 x 19 x 512   -> 19 x 19 x1024
21 conv    512  1 x 1 / 1  19 x 19 x1024   -> 19 x 19 x 512
22 conv   1024  3 x 3 / 1  19 x 19 x 512   -> 19 x 19 x1024
23 conv   1024  3 x 3 / 1  19 x 19 x1024   -> 19 x 19 x1024
24 conv   1024  3 x 3 / 1  19 x 19 x1024   -> 19 x 19 x1024
25 route    16
26 conv     64  1 x 1 / 1  38 x 38 x 512   -> 38 x 38 x 64
27 reorg           / 2  38 x 38 x 64   -> 19 x 19 x 256
28 route    27 24
29 conv   1024  3 x 3 / 1  19 x 19 x1280   -> 19 x 19 x1024
30 conv    425  1 x 1 / 1  19 x 19 x1024   -> 19 x 19 x 425
31 detection
mask_scale: Using default '1.000000'
Loading weights from yolo.weights...Done!
data/sample.jpg: Predicted in 10.643002 seconds.
cat: 84%
car: 48%
fahad@fahad-pc:~/YOLO/darknet$

```

Chapter 14: Elasticsearch



```
fahad@fahad-pc: ~/Downloads/elastic
File Edit View Search Terminal Help
fahad@fahad-pc:~/Downloads/elastic$ sudo gdebi elasticsearch-6.2.2.deb
Reading package lists... Done
Building dependency tree
Reading state information... Done
Reading state information... Done

Elasticsearch is a distributed RESTful search engine built for the cloud. Reference documentation can be found at https://www.elastic.co/guide/en/elasticsearch/reference/current/index.html and the 'Elasticsearch: The Definitive Guide' book can be found at https://www.elastic.co/guide/en/elasticsearch/guide/current/index.html
Elasticsearch subproject : distribution:deb
Do you want to install the software package? [y/N]:Y
Selecting previously unselected package elasticsearch.
(Reading database ... 285240 files and directories currently installed.)
Preparing to unpack elasticsearch-6.2.2.deb ...
Unpacking elasticsearch (6.2.2) ...
Setting up elasticsearch (6.2.2) ...
Processing triggers for ureadahead (0.100.0-20) ...
Processing triggers for systemd (234-2ubuntu12.1) ...
fahad@fahad-pc:~/Downloads/elastic$
```

```

fahad@fahad-pc: ~
File Edit View Search Terminal Help
fahad@fahad-pc:~$ sudo service elasticsearch start
[sudo] password for fahad:
fahad@fahad-pc:~$ sudo service elasticsearch status
● elasticsearch.service - Elasticsearch
   Loaded: loaded (/usr/lib/systemd/system/elasticsearch.service; disabled; vendor preset: enabled)
   Active: active (running) since Mon 2018-03-19 19:16:58 PKT; 3s ago
     Docs: http://www.elastic.co
   Main PID: 28839 (java)
    Tasks: 20 (limit: 4915)
   CGroup: /system.slice/elasticsearch.service
           └─28839 /usr/bin/java -Xms1g -Xmx1g -XX:+UseConcMarkSweepGC -XX:CMSInitiatingOccupancyFra

Mar 19 19:16:58 fahad-pc systemd[1]: Started Elasticsearch.
Mar 19 19:16:58 fahad-pc elasticsearch[28839]: Java HotSpot(TM) 64-Bit Server VM warning: Option Use
lines 1-11/11 (END)

```

```

← → ↻ ⓘ localhost:9200
{
  "name" : "OCL-tzT",
  "cluster_name" : "elasticsearch",
  "cluster_uuid" : "nPMaHUbKQ16iLZSYKE2aTQ",
  "version" : {
    "number" : "6.2.2",
    "build_hash" : "10bledd",
    "build_date" : "2018-02-16T19:01:30.685723Z",
    "build_snapshot" : false,
    "lucene_version" : "7.2.1",
    "minimum_wire_compatibility_version" : "5.6.0",
    "minimum_index_compatibility_version" : "5.0.0"
  },
  "tagline" : "You Know, for Search"
}

```

```

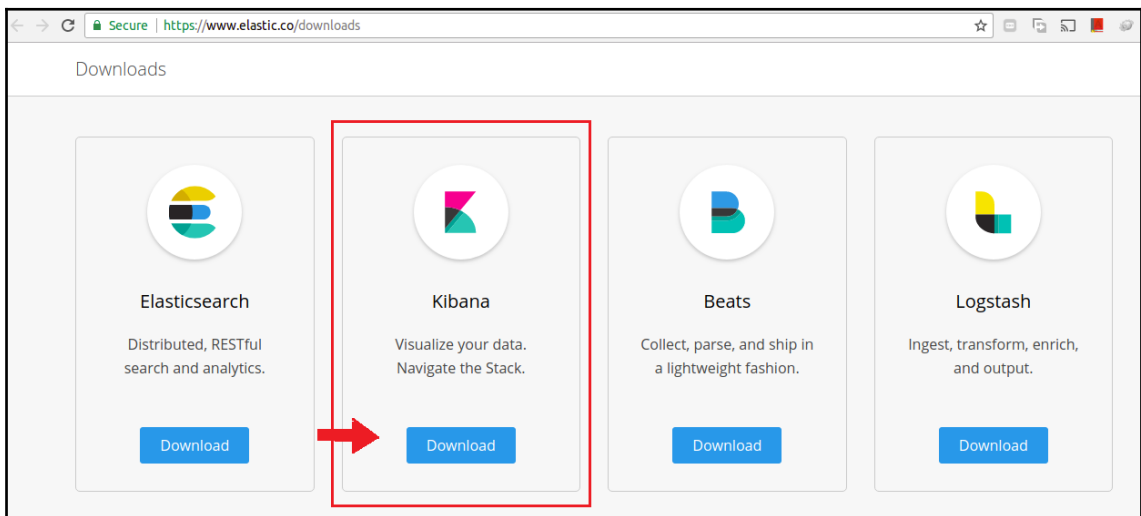
fahad@fahad-pc: ~/Downloads/elastic
File Edit View Search Terminal Help
fahad@fahad-pc:~/Downloads/elastic$ sudo /bin/systemctl daemon-reload
[sudo] password for fahad:
fahad@fahad-pc:~/Downloads/elastic$ sudo /bin/systemctl enable elasticsearch.service
Synchronizing state of elasticsearch.service with SysV service script with /lib/systemd/systemd-sysv-install.
Executing: /lib/systemd/systemd-sysv-install enable elasticsearch
Created symlink /etc/systemd/system/multi-user.target.wants/elasticsearch.service → /usr/lib/systemd/system/elasticsearch.ser
vice.
fahad@fahad-pc:~/Downloads/elastic$

```

```
fahad@fahad-pc: ~/Downloads/elastic

File Edit View Search Terminal Help

fahad@fahad-pc:~/Downloads/elastic$ sudo apt remove elasticsearch
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following packages will be REMOVED:
  elasticsearch
0 upgraded, 0 newly installed, 1 to remove and 0 not upgraded.
After this operation, 32.3 MB disk space will be freed.
Do you want to continue? [Y/n] Y
(Reading database ... 285375 files and directories currently installed.)
Removing elasticsearch (6.2.2) ...
Stopping elasticsearch service... OK
Deleting log directory... OK
fahad@fahad-pc:~/Downloads/elastic$
```



```

fahad@fahad-pc: ~/Downloads/elastic

File Edit View Search Terminal Help

fahad@fahad-pc:~/Downloads/elastic$ sudo gdebi kibana-6.2.2-amd64.deb
[sudo] password for fahad:
Reading package lists... Done
Building dependency tree
Reading state information... Done
Reading state information... Done

Explore and visualize your Elasticsearch data
Do you want to install the software package? [y/N]:Y
Selecting previously unselected package kibana.
(Reading database ... 285370 files and directories currently installed.)
Preparing to unpack kibana-6.2.2-amd64.deb ...
Unpacking kibana (6.2.2) ...
Setting up kibana (6.2.2) ...
Processing triggers for ureadahead (0.100.0-20) ...
Processing triggers for systemd (234-2ubuntu12.1) ...
fahad@fahad-pc:~/Downloads/elastic$

```

```

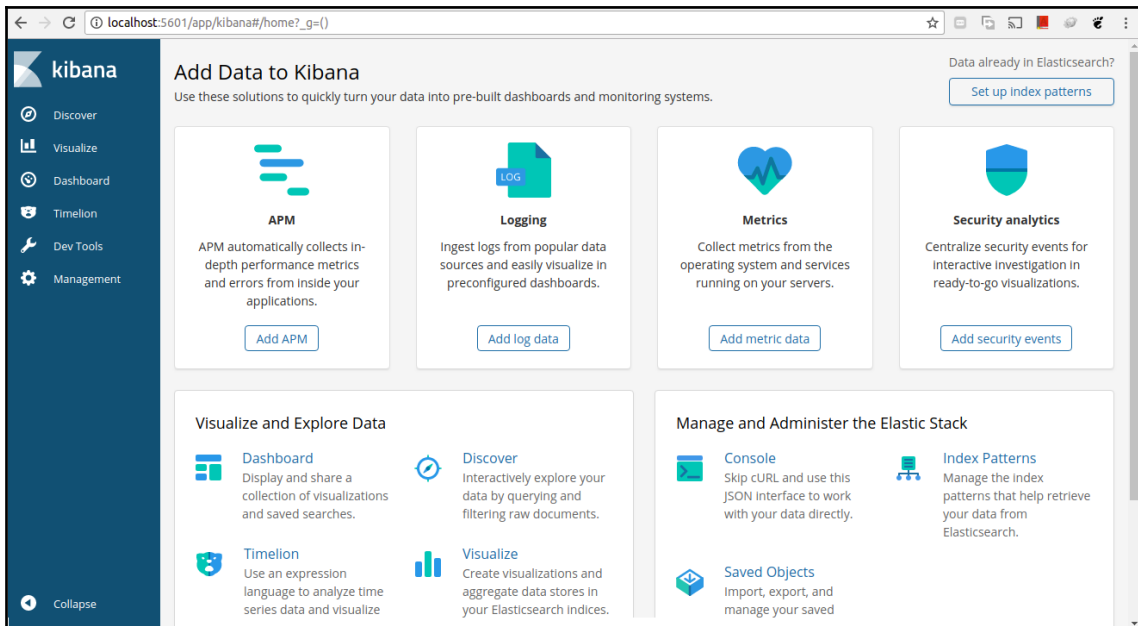
fahad@fahad-pc: ~/Downloads/elastic

File Edit View Search Terminal Help

fahad@fahad-pc:~/Downloads/elastic$ sudo service kibana start
fahad@fahad-pc:~/Downloads/elastic$ sudo service kibana status
● kibana.service - Kibana
   Loaded: loaded (/etc/systemd/system/kibana.service; disabled; vendor preset: enabled)
   Active: active (running) since Mon 2018-03-19 21:23:08 PKT; 3s ago
   Main PID: 12870 (node)
   Tasks: 10 (limit: 4915)
   CGroup: /system.slice/kibana.service
           └─12870 /usr/share/kibana/bin/../node/bin/node --no-warnings /usr/share/kibana/bin/../src/cli -c /etc/kibana/kiban

Mar 19 21:23:08 fahad-pc systemd[1]: Started Kibana.
Mar 19 21:23:10 fahad-pc kibana[12870]: {"type":"log","@timestamp":"2018-03-19T16:23:10Z","tags":["status","plugin:kibana@6.2
Mar 19 21:23:10 fahad-pc kibana[12870]: {"type":"log","@timestamp":"2018-03-19T16:23:10Z","tags":["status","plugin:elasticsea
Mar 19 21:23:11 fahad-pc kibana[12870]: {"type":"log","@timestamp":"2018-03-19T16:23:11Z","tags":["status","plugin:timelion@6
Mar 19 21:23:11 fahad-pc kibana[12870]: {"type":"log","@timestamp":"2018-03-19T16:23:11Z","tags":["status","plugin:console@6
Mar 19 21:23:11 fahad-pc kibana[12870]: {"type":"log","@timestamp":"2018-03-19T16:23:11Z","tags":["status","plugin:metrics@6
Mar 19 21:23:11 fahad-pc kibana[12870]: {"type":"log","@timestamp":"2018-03-19T16:23:11Z","tags":["listening","info"],"pid":1
Mar 19 21:23:11 fahad-pc kibana[12870]: {"type":"log","@timestamp":"2018-03-19T16:23:11Z","tags":["status","plugin:elasticsea
lines 1-16/16 (END)

```



```
fahad@fahad-pc: ~/Downloads/elastic
File Edit View Search Terminal Help
fahad@fahad-pc:~/Downloads/elastic$ sudo apt remove kibana
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following packages will be REMOVED:
  kibana
0 upgraded, 0 newly installed, 1 to remove and 0 not upgraded.
After this operation, 298 MB disk space will be freed.
Do you want to continue? [Y/n] y
(Reading database ... 339665 files and directories currently installed.)
Removing kibana (6.2.2) ...
Stopping kibana service... OK
fahad@fahad-pc:~/Downloads/elastic$
```

```

fahad@asitdev-pc: /usr/share/elasticsearch/bin
File Edit View Search Terminal Help
fahad@asitdev-pc:/usr/share/elasticsearch/bin$ sudo ./elasticsearch-plugin install x-pack
-> Downloading x-pack from elastic
[=====] 100%
@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
@ WARNING: plugin requires additional permissions @
@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
* java.io.FilePermission \\.\\pipe\\*#plus read,write
* java.lang.RuntimePermission accessClassInPackage.com.sun.activation.registries
* java.lang.RuntimePermission getClassLoader
* java.lang.RuntimePermission setContextClassLoader
* java.lang.RuntimePermission setFactory
* java.net.SocketPermission * connect,accept,resolve
* java.security.SecurityPermission createPolicy.JavaPolicy
* java.security.SecurityPermission getPolicy
* java.security.SecurityPermission putProviderProperty.BC
* java.security.SecurityPermission setPolicy
* java.util.PropertyPermission * read,write
See http://docs.oracle.com/javase/8/docs/technotes/guides/security/permissions.html
for descriptions of what these permissions allow and the associated risks.

Continue with installation? [y/N]y
@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
@ WARNING: plugin forks a native controller @
@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
This plugin launches a native controller that is not subject to the Java
security manager nor to system call filters.

Continue with installation? [y/N]y
Elasticsearch keystore is required by plugin [x-pack-security], creating...
-> Installed x-pack with: x-pack-deprecation,x-pack-ml,x-pack-logstash,x-pack-upgrade,x-pack-graph,x-pack-security,x-pack-mo
nitoring,x-pack-watcher,x-pack-core
fahad@asitdev-pc:/usr/share/elasticsearch/bin$

```

```

fahad@fahad-pc: /usr/share/elasticsearch/bin
File Edit View Search Terminal Help
fahad@fahad-pc:/usr/share/elasticsearch/bin$ sudo ./x-pack/setup-passwords auto
Initiating the setup of passwords for reserved users elastic,kibana,logstash_system.
The passwords will be randomly generated and printed to the console.
Please confirm that you would like to continue [y/N]y

Changed password for user kibana
PASSWORD kibana = sH35Peqdjysj5we2Kndj

Changed password for user logstash_system
PASSWORD logstash_system = GzwfhJHy1Qh9t0oW5YF3

Changed password for user elastic
PASSWORD elastic = 3gpYmmfaohlIb26Hlw1L

```

```
fahad@asitdev-pc: /usr/share/kibana/bin

File Edit View Search Terminal Help

fahad@asitdev-pc:/usr/share/kibana/bin$ sudo ./kibana-plugin install x-pack
Found previous install attempt. Deleting...
Attempting to transfer from x-pack
Attempting to transfer from https://artifacts.elastic.co/downloads/kibana-plugins/x-pack/x-pack-6.2.2.zip
Transferring 269704442 bytes.....
Transfer complete
Retrieving metadata from plugin archive
Extracting plugin archive
Extraction complete
Optimizing and caching browser bundles...
Plugin installation complete
fahad@asitdev-pc:/usr/share/kibana/bin$
```

```
fahad@fahad-pc: /etc/kibana

File Edit View Search Terminal Help

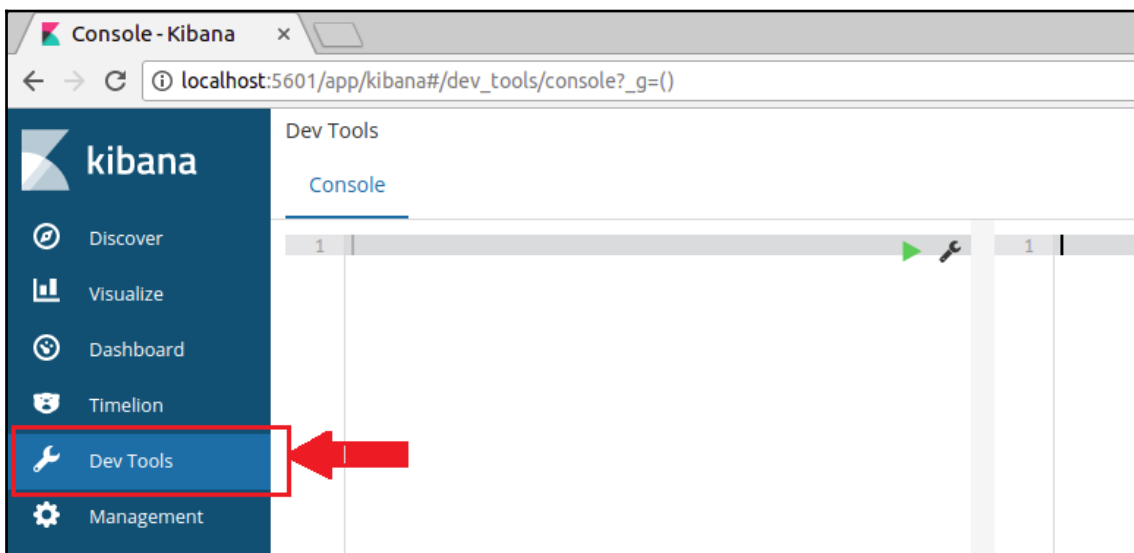
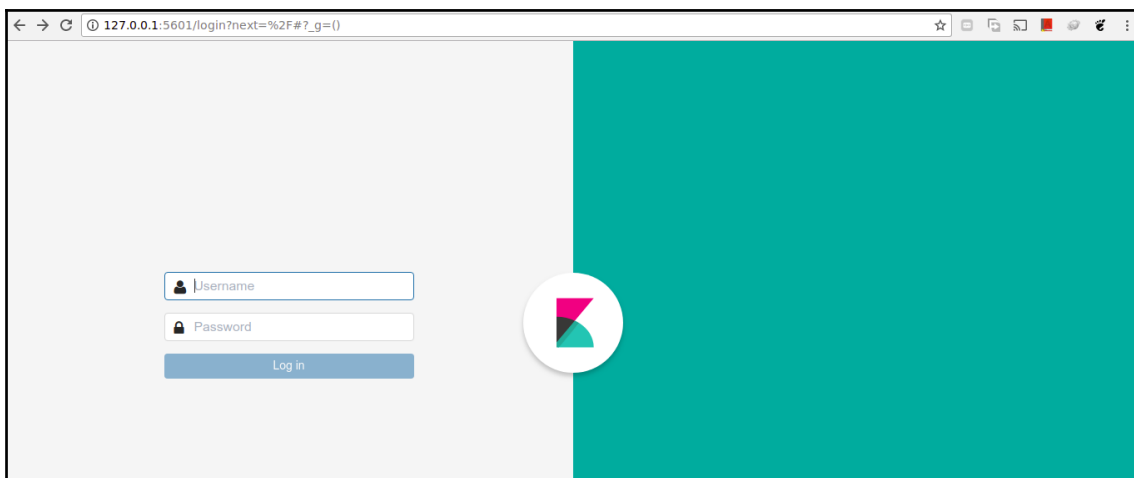
GNU nano 2.8.6 File: kibana.yml

# When this setting's value is true Kibana uses the hostname specified in the server.host
# setting. When the value of this setting is false, Kibana uses the hostname of the host
# that connects to this Kibana instance.
#elasticsearch.preserveHost: true

# Kibana uses an index in Elasticsearch to store saved searches, visualizations and
# dashboards. Kibana creates a new index if the index doesn't already exist.
#kibana.index: ".kibana"

# The default application to load.
#kibana.defaultAppId: "home"

# If your Elasticsearch is protected with basic authentication, these settings provide
# the username and password that the Kibana server uses to perform maintenance on the Kibana
# index at startup. Your Kibana users still need to authenticate with Elasticsearch, which
# is proxied through the Kibana server.
elasticsearch.username: "elastic"
elasticsearch.password: "3gpYmmfaohlB26Hlw1L"
```

```
Console
1 POST /student/detail
2 {
3   "name" : "Smith",
4   "age" : 10,
5   "phone" : "1-234-567890",
6   "address" : "some where on earth"
7 }
8

1 {
2   "_index": "student",
3   "_type": "detail",
4   "id": "4K2jv2IByjXVQnwMXRV8",
5   "_version": 1,
6   "result": "created",
7   "_shards": {
8     "total": 2,
9     "successful": 1,
10    "failed": 0
11  },
12   "_seq_no": 0,
13   "_primary_term": 1
14 }
```

```
Console
1 PUT /student/detail
2 {
3   "name" : "JohnDoe",
4   "age" : 12,
5   "phone" : "2-584-5222890",
6   "address" : "some where in the universe"
7 }
8
9

1 {
2   "error": "Incorrect HTTP method for uri [/student/detail] and method [PUT],
3   allowed: [POST]",
4   "status": 405
}
```

Chapter 15: Structured Data

```
fahad@hadoopadmin-VirtualBox: ~  
File Edit View Search Terminal Help  
fahad@hadoopadmin-VirtualBox:~$ sudo apt-get install mysql-server  
Reading package lists... Done  
Building dependency tree  
Reading state information... Done  
The following additional packages will be installed:  
  libevent-core-2.1-6 libhtml-template-perl mysql-client-5.7 mysql-client-core-5.7 mysql-server-5.7  
  mysql-server-core-5.7  
Suggested packages:  
  libipc-sharedcache-perl mailx tinyca  
The following NEW packages will be installed:  
  libevent-core-2.1-6 libhtml-template-perl mysql-client-5.7 mysql-client-core-5.7 mysql-server  
  mysql-server-5.7 mysql-server-core-5.7  
0 upgraded, 7 newly installed, 0 to remove and 0 not upgraded.  
Need to get 20.4 MB of archives.  
After this operation, 160 MB of additional disk space will be used.  
Do you want to continue? [Y/n]
```

```
fahad@hadoopadmin-VirtualBox: ~  
File Edit View Search Terminal Help  
Package configuration  
  
Configuring mysql-server-5.7  
While not mandatory, it is highly recommended that you set a password for the MySQL administrative "root" user.  
If this field is left blank, the password will not be changed.  
New password for the MySQL "root" user:  
_____  
<Ok>
```

```
Configuring mysql-server-5.7  
  
Repeat password for the MySQL "root" user:  
***** _____  
<Ok>
```

```
fahad@fahad-pc: ~  
File Edit View Search Terminal Help  
fahad@fahad-pc:~$ sudo service mysql status  
[sudo] password for fahad:  
● mysql.service - MySQL Community Server  
   Loaded: loaded (/lib/systemd/system/mysql.service; disabled; vendor preset: enabled)  
   Active: active (running) since Sat 2018-04-07 11:21:26 PKT; 1 day 7h ago  
 Main PID: 30350 (mysqld)  
    Tasks: 31 (limit: 4915)  
   CGroup: /system.slice/mysql.service  
           └─30350 /usr/sbin/mysqld  
  
Apr 07 11:21:24 fahad-pc systemd[1]: Starting MySQL Community Server...  
Apr 07 11:21:26 fahad-pc systemd[1]: Started MySQL Community Server.  
fahad@fahad-pc:~$
```

```
fahad@fahad-pc: ~  
File Edit View Search Terminal Help  
fahad@fahad-pc:~$ mysql -u root -p  
Enter password:  
Welcome to the MySQL monitor.  Commands end with ; or \g.  
Your MySQL connection id is 7  
Server version: 5.7.21-0ubuntu0.17.10.1 (Ubuntu)  
  
Copyright (c) 2000, 2018, Oracle and/or its affiliates. All rights reserved.  
  
Oracle is a registered trademark of Oracle Corporation and/or its  
affiliates. Other names may be trademarks of their respective  
owners.  
  
Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.  
  
mysql> show databases;  
+-----+  
| Database |  
+-----+  
| information schema |  
| employees |  
| mysql |  
| performance_schema |  
| sys |  
+-----+  
5 rows in set (0.00 sec)  
  
mysql>
```

Database that we will be working

```
fahad@fahad-pc: ~  
File Edit View Search Terminal Help  
mysql> show tables;  
+-----+  
| Tables_in_employees |  
+-----+  
| empInfo              |  
| empSalary            |  
+-----+  
2 rows in set (0.00 sec)
```

```
fahad@fahad-pc: ~  
File Edit View Search Terminal Help  
mysql> describe empInfo;  
+-----+-----+-----+-----+-----+-----+  
| Field      | Type      | Null | Key | Default | Extra |  
+-----+-----+-----+-----+-----+-----+  
| empId      | int(11)   | NO   | PRI | NULL    |       |  
| empName    | varchar(45) | YES  |     | NULL    |       |  
| phoneNo    | varchar(45) | YES  |     | NULL    |       |  
| address    | varchar(45) | YES  |     | NULL    |       |  
| designation | varchar(45) | YES  |     | NULL    |       |  
| department | varchar(45) | YES  |     | NULL    |       |  
+-----+-----+-----+-----+-----+-----+  
6 rows in set (0.00 sec)  
  
mysql>
```

```
fahad@fahad-pc: ~
File Edit View Search Terminal Help
mysql> describe empSalary;
+-----+-----+-----+-----+-----+-----+
| Field          | Type          | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+-----+
| salaryId       | int(11)       | NO   | PRI | NULL    |       |
| salaryMonth    | datetime     | YES  |     | NULL    |       |
| empId          | int(11)       | YES  |     | NULL    |       |
| createdAt      | varchar(45)   | YES  |     | NULL    |       |
+-----+-----+-----+-----+-----+-----+
4 rows in set (0.00 sec)

mysql>
```

```
fahad@fahad-pc: ~
File Edit View Search Terminal Help
mysql> select * from empInfo;
+-----+-----+-----+-----+-----+-----+
| empId | empName | phoneNo | address          | designation | department |
+-----+-----+-----+-----+-----+-----+
| 1     | John Smith | 123456789 | somewhere on earth | clerk       | Accounts   |
| 2     | John Doe  | 987987999 | somewhere on universe | engineer    | Development |
| 3     | Smith    | 5432123456 | somewhere         | engineer    | Development |
| 4     | Andrew   | 331222345 | under the hood    | manager     | Development |
| 5     | Troy     | 9102385731 | nowhere           | director    | Business   |
+-----+-----+-----+-----+-----+-----+
5 rows in set (0.00 sec)

mysql>
```

```
fahad@fahad-pc: ~  
File Edit View Search Terminal Help  
mysql> select * from empSalary;  
+-----+-----+-----+-----+  
| salaryId | salaryMonth | empId | salaryAmount |  
+-----+-----+-----+-----+  
|          1 | 2018-04      |      1 |          1025 |  
|          2 | 2018-04      |      2 |          2556 |  
|          3 | 2018-04      |      3 |          1500 |  
|          4 | 2018-04      |      4 |          5000 |  
|          5 | 2018-04      |      5 |          4780 |  
+-----+-----+-----+-----+  
5 rows in set (0.00 sec)  
  
mysql>
```

```
hadoopadmin@hadoopadmin-VirtualBox: ~/hbase-1.4.3/conf  
File Edit View Search Terminal Help  
GNU nano 2.8.6 File: hbase-env.sh  
  
# *  
# * Unless required by applicable law or agreed to in writing, software  
# * distributed under the License is distributed on an "AS IS" BASIS,  
# * WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.  
# * See the License for the specific language governing permissions and  
# * limitations under the License.  
# */  
  
# Set environment variables here.  
  
# This script sets variables multiple times over the course of starting an hbase process,  
# so try to keep things idempotent unless you want to take an even deeper look  
# into the startup scripts (bin/hbase, etc.)  
  
# The java implementation to use. Java 1.7+ required.  
export JAVA_HOME="/home/hadoopadmin/jdk1.8.0_144/"  
  
# Extra Java CLASSPATH elements. Optional.  
# export HBASE_CLASSPATH=
```

```

hadoopadmin@hadoopadmin-VirtualBox: ~/hbase-1.4.3/conf
File Edit View Search Terminal Help
GNU nano 2.8.6 File: hbase-env.sh Modi

# export HBASE_PID_DIR=/var/hadoop/pids

# Seconds to sleep between slave commands. Unset by default. This
# can be useful in large clusters, where, e.g., slave rsyncs can
# otherwise arrive faster than the master can service them.
# export HBASE_SLAVE_SLEEP=0.1

# Tell HBase whether it should manage it's own instance of Zookeeper or not.
export HBASE_MANAGES_ZK=true

# The default log rolling policy is RFA, where the log file is rolled as per the size defined for the
# RFA appender. Please refer to the log4j.properties file to see more details on this appender.
# In case one needs to do log rolling on a date change, one should set the environment property
# HBASE_ROOT_LOGGER to "<DESIRED_LOG_LEVEL>,DRFA".
# For example:

^G Get Help      ^O Write Out    ^W Where Is    ^K Cut Text     ^J Justify     ^C Cur Pos     M-U Undo
^X Exit         ^R Read File    ^_ Replace     ^U Uncut Text  ^T To Linter   ^_ Go To Line  M-E Redo

```

```

hadoopadmin@hadoopadmin-VirtualBox: ~/hbase-1.4.3/conf
File Edit View Search Terminal Help
GNU nano 2.8.6 File: hbase-site.xml

<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
/**
 *
 * Licensed to the Apache Software Foundation (ASF) under one
 * or more contributor license agreements. See the NOTICE file
 * distributed with this work for additional information
 * regarding copyright ownership. The ASF licenses this file
 * to you under the Apache License, Version 2.0 (the
 * "License"); you may not use this file except in compliance
 * with the License. You may obtain a copy of the License at
 *
 * http://www.apache.org/licenses/LICENSE-2.0
 *
 * Unless required by applicable law or agreed to in writing, software
 * distributed under the License is distributed on an "AS IS" BASIS,
 * WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
 * See the License for the specific language governing permissions and
 * limitations under the License.
 */
-->
<configuration>
  <property>
    <name>hbase.rootdir</name>
    <value>hdfs://localhost:9000/hbase</value>
  </property>
</configuration>

```


Namenode information

localhost:50070/dfshealth.html#tab-overview

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

Overview 'localhost:9000' (active)

Started:	Sun Apr 08 18:34:35 +0500 2018
Version:	2.8.1, r20fe5304904fc2f5a18053c389e43cd26f7a70fe
Compiled:	Fri Jun 02 11:14:00 +0500 2017 by vinodkv from branch-2.8.1-private
Cluster ID:	CID-ab35415f-b138-4ed7-9405-ce67233f5a2
Block Pool ID:	BP-1468514210-127.0.1.1-1523194438743

```
hadoopadmin@hadoopadmin-VirtualBox: ~/hbase-1.4.3/bin
File Edit View Search Terminal Help
hadoopadmin@hadoopadmin-VirtualBox:~$ cd ~/hbase-1.4.3/bin/
hadoopadmin@hadoopadmin-VirtualBox:~/hbase-1.4.3/bin$ ./start-hbase.sh
running master, logging to /home/hadoopadmin/hbase-1.4.3/bin/../logs/hbase-hadoopadmin-master-hadoopadmin-VirtualBox.out
Java HotSpot(TM) 64-Bit Server VM warning: ignoring option PermSize=128m; support was removed in 8.0
Java HotSpot(TM) 64-Bit Server VM warning: ignoring option MaxPermSize=128m; support was removed in 8.0
hadoopadmin@hadoopadmin-VirtualBox:~/hbase-1.4.3/bin$
```

```
hadoopadmin@hadoopadmin-VirtualBox: ~/hbase-1.4.3/bin
File Edit View Search Terminal Help
hadoopadmin@hadoopadmin-VirtualBox:~/hbase-1.4.3/bin$ ./hbase shell
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/hadoopadmin/hbase-1.4.3/lib/slf4j-log4j12-1.7.10.jar!/org/
SLF4J: Found binding in [jar:file:/home/hadoopadmin/hadoop-2.8.1/share/hadoop/common/lib/slf4j-log
icLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
HBase Shell
Use "help" to get list of supported commands.
Use "exit" to quit this interactive shell.
Version 1.4.3, r172373d1f02bbe0e3da37ec25efc97d0ec69fc96, Wed Mar 21 17:21:52 PDT 2018

hbase(main):001:0>
```

```

hadoopadmin@hadoopadmin-VirtualBox: ~/hbase-1.4.3/bin
File Edit View Search Terminal Help

hbase(main):001:0> create 'info','employee'
0 row(s) in 1.6750 seconds

hbase(main):002:0> create 'salary','employee'
0 row(s) in 1.2570 seconds

hbase(main):003:0>

```

```

hadoopadmin@hadoopadmin-VirtualBox: ~
File Edit View Search Terminal Help

hadoopadmin@hadoopadmin-VirtualBox:~$ wget http://www-eu.apache.org/dist/sqoop/1.4.7/sqoop-1.4.7.bin__hadoop-2.6.0.tar.gz
--2018-04-09 11:12:18-- http://www-eu.apache.org/dist/sqoop/1.4.7/sqoop-1.4.7.bin__hadoop-2.6.0.tar.gz
Resolving www-eu.apache.org (www-eu.apache.org)... 195.154.151.36, 2001:bc8:2142:300::
Connecting to www-eu.apache.org (www-eu.apache.org)|195.154.151.36|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 17953604 (17M) [application/x-gzip]
Saving to: 'sqoop-1.4.7.bin__hadoop-2.6.0.tar.gz'

sqoop-1.4.7.bin__hadoop-2.6.0.ta 100%[=====] 17.12M 472KB/s in 43s
2018-04-09 11:13:02 (410 KB/s) - 'sqoop-1.4.7.bin__hadoop-2.6.0.tar.gz' saved [17953604/17953604]

hadoopadmin@hadoopadmin-VirtualBox:~$

```

```

hadoopadmin@hadoopadmin-VirtualBox: ~/sqoop-1.4.7.bin__hadoop-2.6.0/conf
File Edit View Search Terminal Help
GNU nano 2.8.6 File: sqoop-env.sh

#Set path to where bin/hadoop is available
export HADOOP_COMMON_HOME="/home/hadoopadmin/hadoop-2.8.1"

#Set path to where hadoop-*-core.jar is available
export HADOOP_MAPRED_HOME="/home/hadoopadmin/hadoop-2.8.1"

#set the path to where bin/hbase is available
export HBASE_HOME="/home/hadoopadmin/hbase-1.4.3"
█

#Set the path to where bin/hive is available
#export HIVE_HOME=

#Set the path for where zookeeper config dir is
#export ZOO_CFG_DIR=

```

```

hadoopadmin@hadoopadmin-VirtualBox: ~
File Edit View Search Terminal Help
hadoopadmin@hadoopadmin-VirtualBox:~$ sqoop version
Warning: /home/hadoopadmin/sqoop-1.4.7.bin__hadoop-2.6.0../hcatalog does not exist! HCatalog jobs will fail.
Please set SHCAT_HOME to the root of your HCatalog installation.
Warning: /home/hadoopadmin/sqoop-1.4.7.bin__hadoop-2.6.0../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
Warning: /home/hadoopadmin/sqoop-1.4.7.bin__hadoop-2.6.0../zookeeper does not exist! Accumulo imports will fail.
Please set $ZOOKEEPER_HOME to the root of your Zookeeper installation.
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/hadoopadmin/hadoop-2.8.1/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/hadoopadmin/hbase-1.4.3/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
10/04/00 11:24:15 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
Sqoop 1.4.7
git commit id 2328971411f57f0cb683dfb79d19d4d19d185dd8
Compiled by maugli on Thu Dec 21 15:59:58 STD 2017
hadoopadmin@hadoopadmin-VirtualBox:~$

```

```

hadoopadmin@hadoopadmin-VirtualBox: ~/sqoopfiles
File Edit View Search Terminal Help
hadoopadmin@hadoopadmin-VirtualBox:~/sqoopfiles$ sqoop import --bindir ./ --connect jdbc:mysql://localhost/employees --username root -P --table empInfo --hbase-table info --column-family employee
Warning: /home/hadoopadmin/sqoop-1.4.7.bin__hadoop-2.6.0../hcatalog does not exist! HCatalog jobs will fail.
Please set SHCAT_HOME to the root of your HCatalog installation.
Warning: /home/hadoopadmin/sqoop-1.4.7.bin__hadoop-2.6.0../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
Warning: /home/hadoopadmin/sqoop-1.4.7.bin__hadoop-2.6.0../zookeeper does not exist! Accumulo imports will fail.
Please set $ZOOKEEPER_HOME to the root of your Zookeeper installation.
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/hadoopadmin/hadoop-2.8.1/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/hadoopadmin/hbase-1.4.3/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
18/04/09 16:13:08 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
Enter password:

```

```
hadoopadmin@hadoopadmin-VirtualBox: ~/sqoopfiles
File Edit View Search Terminal Help
18/04/09 16:11:44 INFO mapreduce.Job: Job job_local374151627_0001 completed successfully
18/04/09 16:11:44 INFO mapreduce.Job: Counters: 20
  File System Counters
    FILE: Number of bytes read=42777369
    FILE: Number of bytes written=43531494
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=0
    HDFS: Number of bytes written=0
    HDFS: Number of read operations=0
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=0
  Map-Reduce Framework
    Map input records=5
    Map output records=5
    Input split bytes=87
    Spilled Records=0
    Failed Shuffles=0
    Merged Map outputs=0
    GC time elapsed (ms)=4
    Total committed heap usage (bytes)=103702528
  File Input Format Counters
    Bytes Read=0
  File Output Format Counters
    Bytes Written=0
18/04/09 16:11:44 INFO mapreduce.ImportJobBase: Transferred 0 bytes in 5.6495 seconds (0 bytes/sec)
18/04/09 16:11:44 INFO mapreduce.ImportJobBase: Retrieved 5 records.
hadoopadmin@hadoopadmin-VirtualBox:~/sqoopfiles$
```

```

hadoopadmin@hadoopadmin-VirtualBox: ~/hbase-1.4.3/bin
File Edit View Search Terminal Help

hbase(main):001:0> scan 'info'
ROW                                COLUMN+CELL
1                                  column=employee:address, timestamp=1523272304551, value=somewhere on earth
1                                  column=employee:department, timestamp=1523272304551, value=Accounts
1                                  column=employee:designation, timestamp=1523272304551, value=clerk
1                                  column=employee:empName, timestamp=1523272304551, value=John Smith
1                                  column=employee:phoneNo, timestamp=1523272304551, value=123456789
2                                  column=employee:address, timestamp=1523272304551, value=somewhere on universe
2                                  column=employee:department, timestamp=1523272304551, value=Development
2                                  column=employee:designation, timestamp=1523272304551, value=engineer
2                                  column=employee:empName, timestamp=1523272304551, value=John Doe
2                                  column=employee:phoneNo, timestamp=1523272304551, value=987987999
3                                  column=employee:address, timestamp=1523272304551, value=somewhere
3                                  column=employee:department, timestamp=1523272304551, value=Development
3                                  column=employee:designation, timestamp=1523272304551, value=engineer
3                                  column=employee:empName, timestamp=1523272304551, value=Smith
3                                  column=employee:phoneNo, timestamp=1523272304551, value=5432123456
4                                  column=employee:address, timestamp=1523272304551, value=under the hood
4                                  column=employee:department, timestamp=1523272304551, value=Development
4                                  column=employee:designation, timestamp=1523272304551, value=manager
4                                  column=employee:empName, timestamp=1523272304551, value=Andrew
4                                  column=employee:phoneNo, timestamp=1523272304551, value=331222345
5                                  column=employee:address, timestamp=1523272304551, value=nowhere
5                                  column=employee:department, timestamp=1523272304551, value=Business
5                                  column=employee:designation, timestamp=1523272304551, value=director
5                                  column=employee:empName, timestamp=1523272304551, value=Troy
5                                  column=employee:phoneNo, timestamp=1523272304551, value=9102385731
5 row(s) in 0.7250 seconds

hbase(main):002:0> █

```

Chapter 16: UnStructured Data



Apache Flume™

Apache Flume

- How to Get Involved
- Download
- Documentation
- Releases
- Mailing lists
- Team
- Source Repository
- Apache License

Download

Apache Flume is distributed under the Apache License, version 2.0

The link in the Mirrors column should display a list of available mirrors with a default selection based on your inferred location. If different browser. The checksum and signature are links to the originals on the main distribution server.

Apache Flume binary (tar.gz)	apache-flume-1.8.0-bin.tar.gz	apache-flume-1.8.0-bin.tar.gz.md5	apache-flume-1.8.0-bin.tar.gz.sha1
Apache Flume source (tar.gz)	apache-flume-1.8.0-src.tar.gz	apache-flume-1.8.0-src.tar.gz.md5	apache-flume-1.8.0-src.tar.gz.sha1

It is essential that you verify the integrity of the downloaded files using the PGP or MD5 signatures. Please read [Verifying Apache](#) information on why you should verify our releases.

```
hadoopadmin@hadoopadmin-VirtualBox: ~/apache-flume-1.8.0-bin
File Edit View Search Terminal Help
hadoopadmin@hadoopadmin-VirtualBox:~$ cd $FLUME_HOME
hadoopadmin@hadoopadmin-VirtualBox:~/apache-flume-1.8.0-bin$ flume-ng agent -f flume-agent.conf -n agent
```

```

hadoopadmin@hadoopadmin-VirtualBox: ~/apache-flume-1.8.0-bin
File Edit View Search Terminal Help
tb/*:/home/hadoopadmin/hadoop-2.8.1/share/hadoop/mapreduce/*:/home/hadoopadmin/hadoop-2.8.1/contrb/capacity-scheduler/*.jar:/lib/* -Djava.lib
rary.path:/home/hadoopadmin/hadoop-2.8.1/lib/native org.apache.flume.node.Application -f flume-agent.conf -n agent
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/hadoopadmin/apache-flume-1.8.0-bin/lib/slf4j-log4j12-1.6.1.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/hadoopadmin/hadoop-2.8.1/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerB
inder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
18/04/20 11:18:48 INFO node.PollingPropertiesFileConfigurationProvider: Configuration provider starting
18/04/20 11:18:48 INFO node.PollingPropertiesFileConfigurationProvider: Reloading configuration file:flume-agent.conf
18/04/20 11:18:49 INFO conf.FlumeConfiguration: Processing:hdfsWriter
18/04/20 11:18:49 INFO conf.FlumeConfiguration: Processing:hdfsWriter
18/04/20 11:18:49 INFO conf.FlumeConfiguration: Processing:hdfsWriter
18/04/20 11:18:49 INFO conf.FlumeConfiguration: Processing:hdfsWriter
18/04/20 11:18:49 INFO conf.FlumeConfiguration: Processing:hdfsWriter
18/04/20 11:18:49 INFO conf.FlumeConfiguration: Added sinks: hdfsWriter Agent: agent
18/04/20 11:18:49 INFO conf.FlumeConfiguration: Post-validation flume configuration contains configuration for agents: [agent]
18/04/20 11:18:49 INFO node.AbstractConfigurationProvider: Creating channels
18/04/20 11:18:49 INFO channel.DefaultChannelFactory: Creating instance of channel memoryChannel type memory
18/04/20 11:18:49 INFO node.AbstractConfigurationProvider: Created channel memoryChannel
18/04/20 11:18:49 INFO source.DefaultSourceFactory: Creating instance of source netcatSource, type netcat
18/04/20 11:18:49 INFO sink.DefaultSinkFactory: Creating instance of sink: hdfsWriter, type: hdfs
18/04/20 11:18:49 INFO node.AbstractConfigurationProvider: Channel memoryChannel connected to [netcatSource, hdfsWriter]
18/04/20 11:18:49 INFO node.Application: Starting new configuration:{ sourceRunners:[netcatSource-EventDrivenSourceRunner: { source:org.apache.f
lume.source.NetcatSource[name:netcatSource,state:IDLE] } ] sinkRunners:[hdfsWriter=SinkRunner: { policy:org.apache.flume.sink.DefaultSinkProcesso
r$Sc6e4742 counterGroup:{ name:null counters:{} } } ] channels:{memoryChannel=org.apache.flume.channel.MemoryChannel{name: memoryChannel}} }
18/04/20 11:18:49 INFO node.Application: Starting channel memoryChannel
18/04/20 11:18:49 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: CHANNEL, name: memoryChannel: Successfully regis
tered new MBean.
18/04/20 11:18:49 INFO instrumentation.MonitoredCounterGroup: Component type: CHANNEL, name: memoryChannel started
18/04/20 11:18:49 INFO node.Application: Starting Sink hdfsWriter
18/04/20 11:18:49 INFO node.Application: Starting Source netcatSource
18/04/20 11:18:49 INFO source.NetcatSource: Source starting
18/04/20 11:18:49 INFO source.NetcatSource: Created serverSocket:sun.nio.ch.ServerSocketChannelImpl[/127.0.0.1:22222]
18/04/20 11:18:49 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SINK, name: hdfsWriter: Successfully registered
new MBean.
18/04/20 11:18:49 INFO instrumentation.MonitoredCounterGroup: Component type: SINK, name: hdfsWriter started

```

```

hadoopadmin@hadoopadmin-VirtualBox: ~/apache-flume-1.8.0-bin
File Edit View Search Terminal Help
hadoopadmin@hadoopadmin-VirtualBox:~/apache-flume-1.8.0-bin$ cat sample.log
Apr 19 19:00:24 hadoopadmin-VirtualBox kernel: [17863.963007] usb 2-1: USB disconnect, device number 2
Apr 19 19:00:24 hadoopadmin-VirtualBox kernel: [17864.573245] e1000: enp0s3 NIC Link is Down
Apr 19 19:00:24 hadoopadmin-VirtualBox kernel: [17864.573297] e1000 0000:00:03:00 enp0s3: Reset adapter
Apr 19 19:00:24 hadoopadmin-VirtualBox kernel: [17864.593244] usb 2-1: new full-speed USB device number 3 using ohci-pci
Apr 19 19:00:25 hadoopadmin-VirtualBox kernel: [17864.931662] usb 2-1: New USB device found, idVendor=80ee, idProduct=0021
Apr 19 19:00:25 hadoopadmin-VirtualBox kernel: [17864.931666] usb 2-1: New USB device strings: Mfr=1, Product=3, SerialNumber=0
Apr 19 19:00:25 hadoopadmin-VirtualBox kernel: [17864.931668] usb 2-1: Product: USB Tablet
Apr 19 19:00:25 hadoopadmin-VirtualBox kernel: [17864.931669] usb 2-1: Manufacturer: VirtualBox
Apr 19 19:00:25 hadoopadmin-VirtualBox kernel: [17864.943271] input: VirtualBox USB Tablet as /devices/pci0000:00/0000:00:06.0/usb
2/2-1/2-1.1.0/0003:80EE:0021.0002/input/input8
Apr 19 19:00:25 hadoopadmin-VirtualBox kernel: [17865.008570] hid-generic 0003:80EE:0021.0002: input,hidraw0: USB HID v1.10 Mouse
[VirtualBox USB Tablet] on usb-0000:00:06.0-1/input0
Apr 19 19:00:26 hadoopadmin-VirtualBox kernel: [17866.743875] e1000: enp0s3 NIC Link is Up 1000 Mbps Full Duplex, Flow Control: RX
hadoopadmin@hadoopadmin-VirtualBox:~/apache-flume-1.8.0-bin$

```

```

hadoopadmin@hadoopadmin-VirtualBox: ~/apache-flume-1.8.0-bin
File Edit View Search Terminal Help
hadoopadmin@hadoopadmin-VirtualBox:~/apache-flume-1.8.0-bin$ head sample.log | nc 127.0.0.1 22222
OK
OK
OK
OK
OK
OK
OK
OK
OK
OK

```

```

hadoopadmin@hadoopadmin-VirtualBox: ~/apache-flume-1.8.0-bin
File Edit View Search Terminal Help
source:org.apache.flume.source.NetcatSource{name:netcatSource,state:IDLE} }} sinkRunners:{hdfsWriter=SinkRunner: { policy:org
.apache.flume.sink.DefaultSinkProcessor@5ce64742 counterGroup:{ name:null counters:{} } }} channels:{memoryChannel=org.apache
.flume.channel.MemoryChannel{name: memoryChannel}} }
18/04/20 20:17:21 INFO node.Application: Starting Channel memoryChannel
18/04/20 20:17:21 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: CHANNEL, name: memoryChannel:
Successfully registered new MBean.
18/04/20 20:17:21 INFO instrumentation.MonitoredCounterGroup: Component type: CHANNEL, name: memoryChannel started
18/04/20 20:17:21 INFO node.Application: Starting Sink hdfsWriter
18/04/20 20:17:21 INFO node.Application: Starting Source netcatSource
18/04/20 20:17:21 INFO source.NetcatSource: Source starting
18/04/20 20:17:21 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SINK, name: hdfsWriter: Succe
ssfully registered new MBean.
18/04/20 20:17:21 INFO instrumentation.MonitoredCounterGroup: Component type: SINK, name: hdfsWriter started
18/04/20 20:17:21 INFO source.NetcatSource: Created server sun.nio.ch.ServerSocketChannelImpl[127.0.0.1:22222]
18/04/20 20:17:51 INFO hdfs.HDFSDataStream: Serializer = TEXT, UseRawLocalFileSystem = false
8/04/20 20:17:52 INFO hdfs.BucketWriter: Creating hdfs://localhost:9000/flumeData/FlumeData.1524237471655.tmp
8/04/20 20:17:54 INFO hdfs.BucketWriter: Closing hdfs://localhost:9000/flumeData/FlumeData.1524237471655.tmp
8/04/20 20:17:54 INFO hdfs.BucketWriter: Renaming hdfs://localhost:9000/flumeData/FlumeData.1524237471655.tmp to hdfs://loc
host:9000/flumeData/FlumeData.1524237471655
8/04/20 20:17:54 INFO hdfs.BucketWriter: Creating hdfs://localhost:9000/flumeData/FlumeData.1524237471656.tmp
8/04/20 20:18:24 INFO hdfs.BucketWriter: Closing hdfs://localhost:9000/flumeData/FlumeData.1524237471656.tmp
8/04/20 20:18:24 INFO hdfs.BucketWriter: Renaming hdfs://localhost:9000/flumeData/FlumeData.1524237471656.tmp to hdfs://loc
host:9000/flumeData/FlumeData.1524237471656
8/04/20 20:18:24 INFO hdfs.HDFSEventSink: Writer callback called.

```

```

hadoopadmin@hadoopadmin-VirtualBox: ~
File Edit View Search Terminal Help
hadoopadmin@hadoopadmin-VirtualBox:~$ hdfs dfs -ls /flumeData
Found 2 items
-rw-r--r-- 1 hadoopadmin supergroup 1035 2018-04-20 20:17 /flumeData/FlumeData.1524237471655
-rw-r--r-- 1 hadoopadmin supergroup 183 2018-04-20 20:18 /flumeData/FlumeData.1524237471656
hadoopadmin@hadoopadmin-VirtualBox:~$

```



```

hadoopadmin@hadoopadmin-VirtualBox: ~
File Edit View Search Terminal Help
hadoopadmin@hadoopadmin-VirtualBox:~$ hdfs dfs -cat /FlumeData/FlumeData.1524237471655
Apr 19 19:00:24 hadoopadmin-VirtualBox kernel: [17863.963007] usb 2-1: USB disconnect, device number 2
Apr 19 19:00:24 hadoopadmin-VirtualBox kernel: [17864.573245] e1000: enp0s3 NIC Link is Down
Apr 19 19:00:24 hadoopadmin-VirtualBox kernel: [17864.573297] e1000 0000:00:03:00 enp0s3: Reset adapter
Apr 19 19:00:24 hadoopadmin-VirtualBox kernel: [17864.593244] usb 2-1: new full-speed USB device number 3 using ohci-pci
Apr 19 19:00:25 hadoopadmin-VirtualBox kernel: [17864.931662] usb 2-1: New USB device found, idVendor=80ee, idProduct=0021
Apr 19 19:00:25 hadoopadmin-VirtualBox kernel: [17864.931666] usb 2-1: New USB device strings: Mfr=1, Product=3, SerialNumber=0
Apr 19 19:00:25 hadoopadmin-VirtualBox kernel: [17864.931668] usb 2-1: Product: USB Tablet
Apr 19 19:00:25 hadoopadmin-VirtualBox kernel: [17864.931669] usb 2-1: Manufacturer: VirtualBox
Apr 19 19:00:25 hadoopadmin-VirtualBox kernel: [17864.943271] input: VirtualBox USB Tablet as /devices/pci0000:00/0000:00:06.0/usb
2/2-1/2-1:1.0/0003:80EE:0021:0002/input/input8
hadoopadmin@hadoopadmin-VirtualBox:~$

```

```

hadoopadmin@hadoopadmin-VirtualBox: ~
File Edit View Search Terminal Help
hadoopadmin@hadoopadmin-VirtualBox:~$ hdfs dfs -cat /FlumeData/FlumeData.1524237471656
Apr 19 19:00:25 hadoopadmin-VirtualBox kernel: [17865.008570] hid-generic 0003:80EE:0021:0002: input,hidraw0: USB HID v1.10 Mouse
[VirtualBox USB Tablet] on usb-0000:00:06.0-1/input0
hadoopadmin@hadoopadmin-VirtualBox:~$

```

```

fahad@fahad-pc: ~
File Edit View Search Terminal Help
fahad@fahad-pc:~$ sudo apt-get install tesseract-ocr
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following additional packages will be installed:
  tesseract-ocr-eng tesseract-ocr-equ tesseract-ocr-osd
The following NEW packages will be installed:
  tesseract-ocr tesseract-ocr-eng tesseract-ocr-equ tesseract-ocr-osd
0 upgraded, 4 newly installed, 0 to remove and 0 not upgraded.
Need to get 12.5 MB of archives.
After this operation, 51.9 MB of additional disk space will be used.
Do you want to continue? [Y/n]

```

SCENE FROM "DAN'L DRUCE."

This interesting domestic drama, by Mr. W. S. Gilbert, has continued to engage the sympathies of a nightly sufficient audience at the Haymarket Theatre, where it has now been represented more than sixty times. Its subject and character were described by us, in the ordinary report of theatrical novelties, about two months ago. Our readers will probably not need to be reminded that the hero of the story, Dan'l Druce, the blacksmith, is a solitary recluse dwelling on the coast of Norfolk, where his lone cottage is visited by fugitives from party vengeance during the civil wars of the Commonwealth. His hoard of money is stolen; but a different sort of treasure, a helpless female infant; is left by some mysterious agency, and may be accepted, as in George Eliot's tale of "Silas Marner," for a Divine gift to the sad-hearted misanthrope, far better than riches. In this spirit, at least, he is content to receive the precious human charge; and so to those who would remove it from his home, Dan'l Druce here makes answer with the solemn exclamation, "Touch not the Lord's gift!" This character is well acted by Mr. Hermann Vezin.

```
hadoopadmin@hadoopadmin-VirtualBox: ~/ImageToText
File Edit View Search Terminal Help
hadoopadmin@hadoopadmin-VirtualBox:~/ImageToText$ python imgToText.py sample.png
=====
BigData Architect's Handbook Example: Convert Image into text
=====

Checking whether tesseract library is installed ...
Tesseract is not installed. Please install it first
If you are in Linux environment, please execute the following command in terminal in order to install it
sudo apt-get install tesseract-ocr
hadoopadmin@hadoopadmin-VirtualBox:~/ImageToText$ █
```

```
hadoopadmin@hadoopadmin-VirtualBox: ~/ImageToText
File Edit View Search Terminal Help
hadoopadmin@hadoopadmin-VirtualBox:~/ImageToText$ python imgToText.py sample.png
=====
BigData Architect's Handbook Example: Convert Image into text
=====

Checking whether tesseract library is installed ...
/usr/bin/tesseract
It is installed on the above mentioned path
-----

Working Directory      : /home/hadoopadmin/ImageToText
Image File Name        : /home/hadoopadmin/ImageToText/sample.png
Output Filename        : /home/hadoopadmin/ImageToText/sample-text

Tesseract Open Source OCR Engine v3.04.01 with Leptonica
Operation completed and the output text has been generated
hadoopadmin@hadoopadmin-VirtualBox:~/ImageToText$
```

```
hadoopadmin@hadoopadmin-VirtualBox: ~/ImageToText
File Edit View Search Terminal Help
hadoopadmin@hadoopadmin-VirtualBox:~/ImageToText$ cat sample-text.txt
SCENE FROM "DAN'L DEUCE."

This interesting domestic drama, by Mr. W. S. Gilbert,
has continued to engage the sympathies of a nightly
sufficient audience at the Haymarket Theatre, where it
has now been represented more than sixty times. Its
subject and character were described by us, in the
ordinary report of theatrical novelties, about two months
ago. Our readers will probably not need to be reminded
that the hero of the story, Dan'l Druce, the blacksmith,
is a so'itm-y recluse dwelling on the coast of Norfolk,
where his lone cottage is visited by fugitives from party
vzngeance during the civil wars of the Commonwealth.
l[is hoard of money is stolen; but a different sort of
treasure, a helpless female infant; is left by some mys-
terious agency, and may be accepted, as in George
Eliot's tale of "Silas Marner," for a Divine gift to the
sad-hearted misanthrope, far better than riches. In
this spirit, at least, he is content to receive the precious
human charge; and so to those who would remove it
from his home, Dan'l Druce here makes answer with
the solemn exclamation, "Touch not the Lord's gift!"
This character is well acted by Mr. Hermann Vezin.

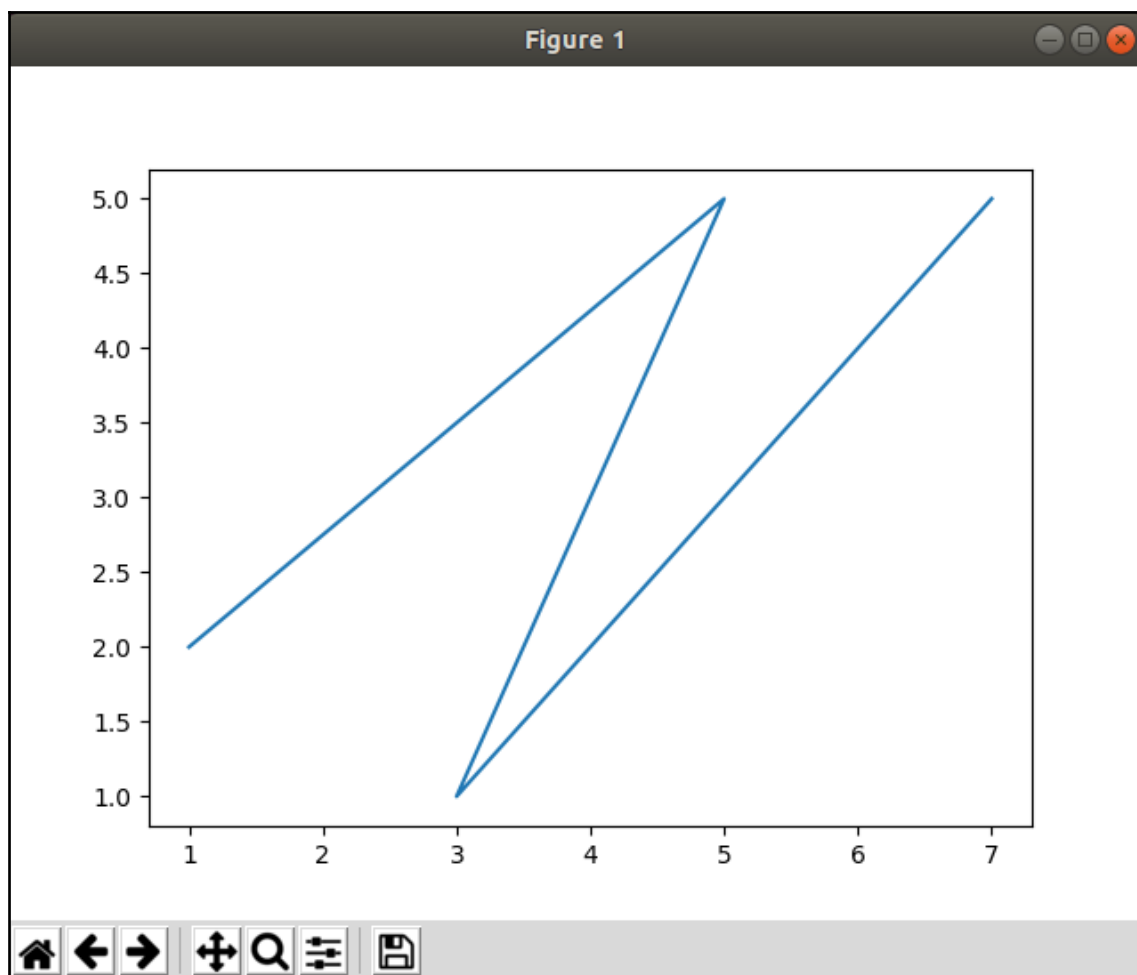
hadoopadmin@hadoopadmin-VirtualBox:~/ImageToText$
```

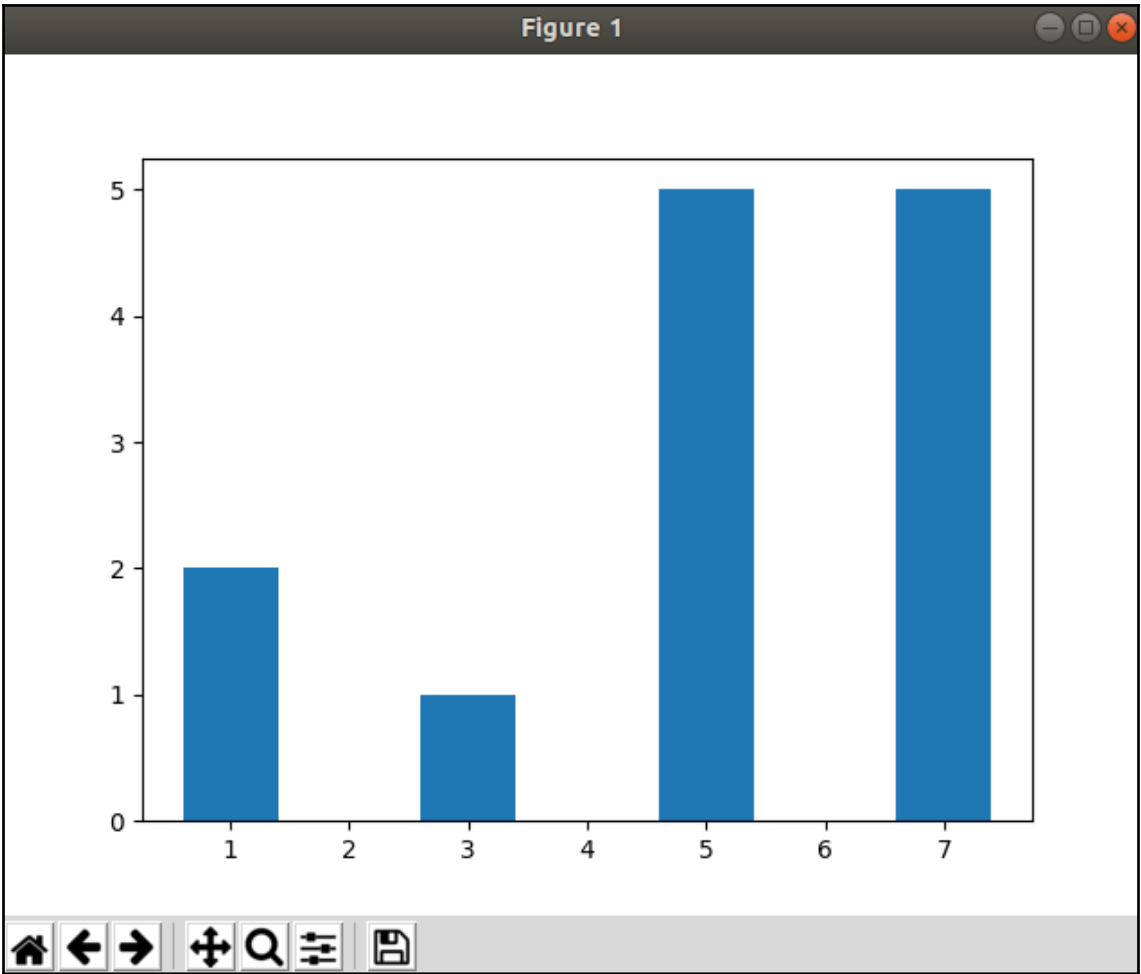
Chapter 17: Data Visualization

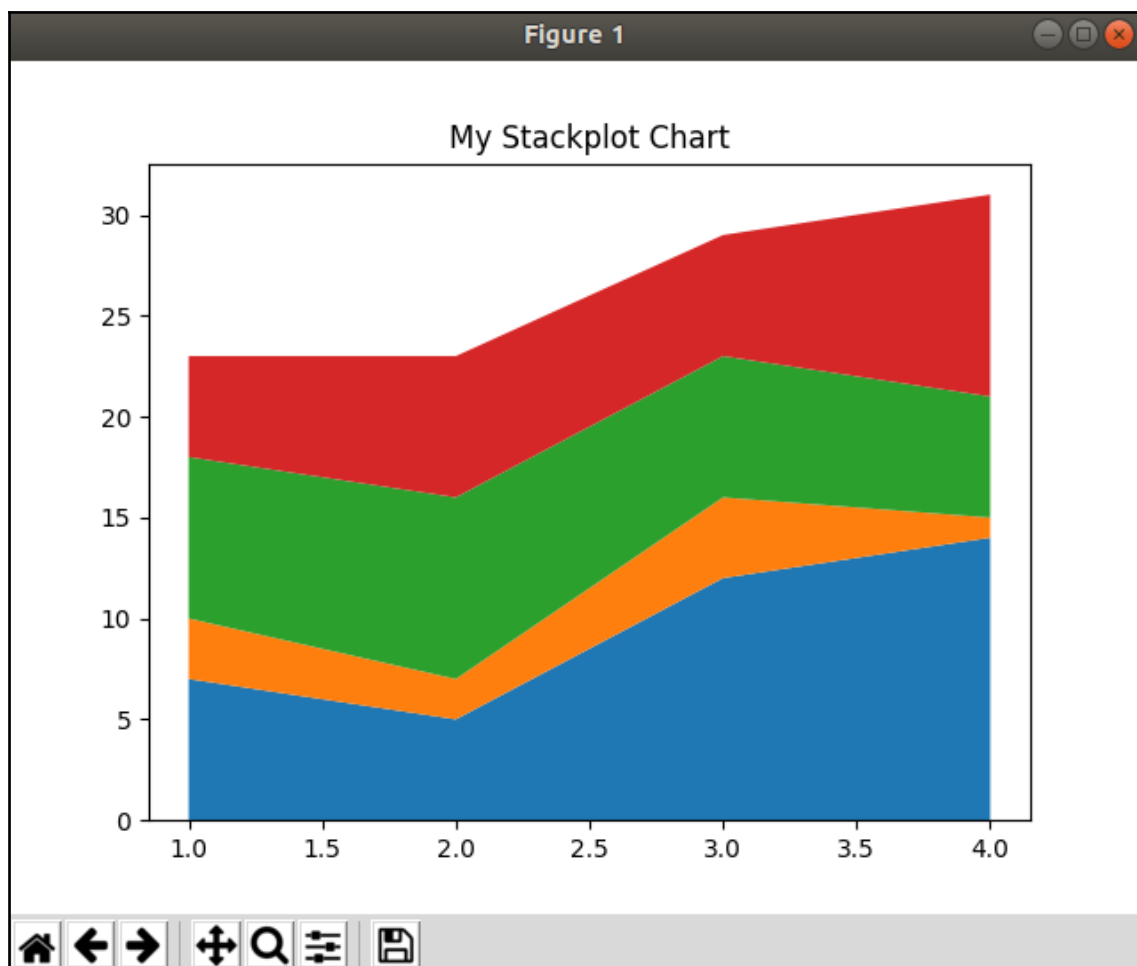
```

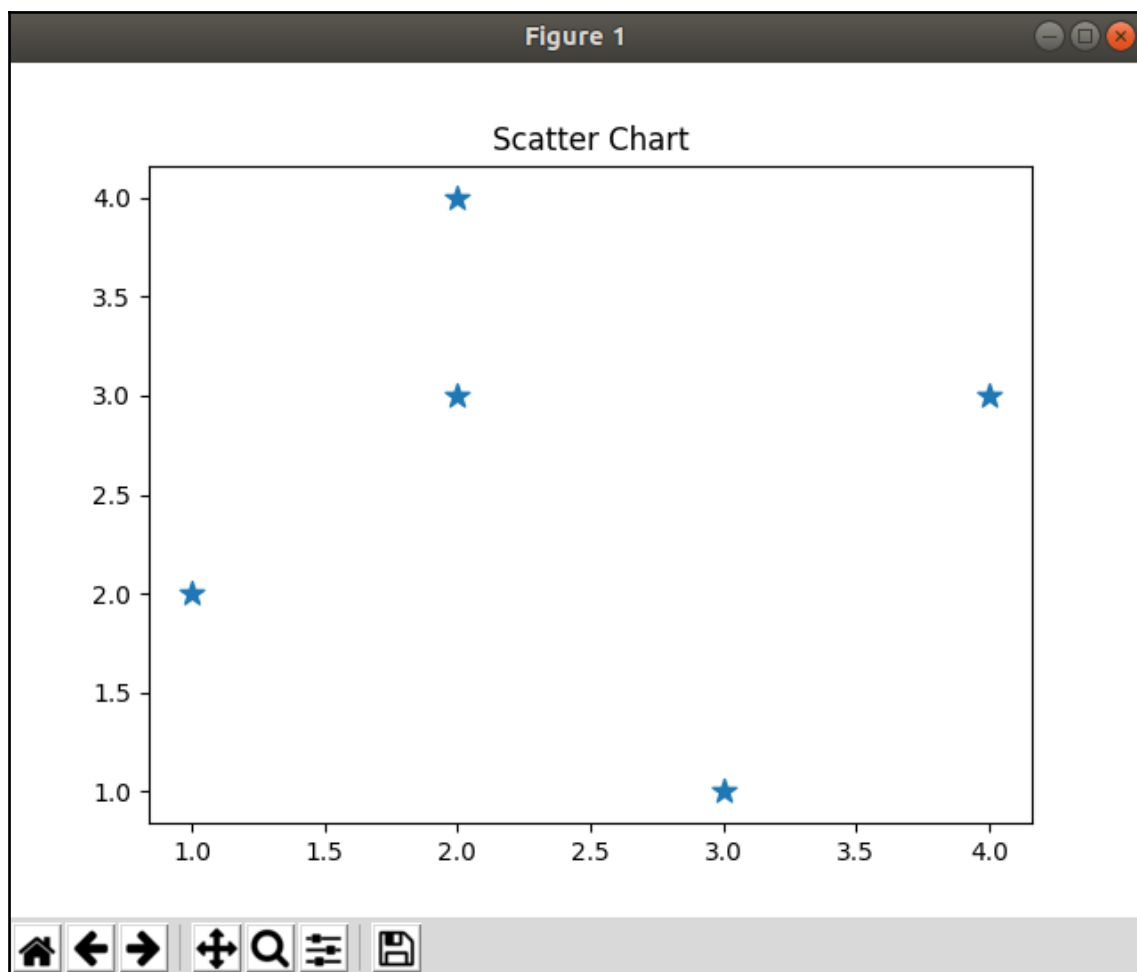
hadoopadmin@hadoopadmin-VirtualBox: ~
File Edit View Search Terminal Help
hadoopadmin@hadoopadmin-VirtualBox:~$ pip install matplotlib
Collecting matplotlib
  Downloading https://files.pythonhosted.org/packages/33/da/6409d32c46778ed20308e83bc501ff2ff8a9cbda5dd7b26362c5e99a6149/matplotlib-2.2.2-cp27-cp27mu-manylinux1_x86_64.whl (12.6MB)
    100% |#####| 12.6MB 69kB/s
Collecting cyclers>=0.10 (from matplotlib)
  Downloading https://files.pythonhosted.org/packages/f7/d2/e07d3ebb2bd7af696440ce7e754c59dd546ffe1bbe732c8ab6089c834e61/cyclers-0.10.0-py3-none-any.whl
Collecting numpy>=1.7.1 (from matplotlib)
  Downloading https://files.pythonhosted.org/packages/76/4d/418dda252cf92bad0ab82d6b2a856e7843b47a5c2f084aed34b14b67d64/numpy-1.14.2-cp27-cp27mu-manylinux1_x86_64.whl (12.1MB)
    100% |#####| 12.1MB 92kB/s
Collecting backports.functiontools-lru-cache (from matplotlib)
  Downloading https://files.pythonhosted.org/packages/03/8e/2424c0e65c4a066e28f539364deee49b6451f8fcd4f718fefa50cc3dcf48/backports.functiontools_lru_cache-1.5-py2.py3-none-any.whl
Collecting subprocess32 (from matplotlib)
  Downloading https://files.pythonhosted.org/packages/b8/2f/49e53b0d0e94611a2dc624a1ad24d41b6d94d0f1b0a078443407ea2214c2/subprocess32-3.2.7.tar.gz (54kB)
    100% |#####| 61kB 634kB/s
Collecting kiwisolver>=1.0.1 (from matplotlib)
  Downloading https://files.pythonhosted.org/packages/3a/62/a8c9bef3059d55ab38e41fe9c8a4fad773bfc04e47290bab84b1c18262e/kiwisolver-1.0.1-cp27-cp27mu-manylinux1_x86_64.whl (951kB)
    100% |#####| 952kB 231kB/s
Collecting pytz (from matplotlib)
  Downloading https://files.pythonhosted.org/packages/dc/83/15f7833b70d3e067ca91467ca245bae0f6fe56ddc7451aa0dc5606b120f2/pytz-2018.4-py3-none-any.whl (510kB)
    100% |#####| 512kB 292kB/s
Collecting six>=1.10 (from matplotlib)
  Downloading https://files.pythonhosted.org/packages/67/4b/141a581104b1f6397bfa78ac9d43d8ad29a7ca43ea90a2d863fe3056e86a/six-1.11.0-py2.py3-none-any.whl
Collecting python-dateutil>=2.1 (from matplotlib)
  Downloading https://files.pythonhosted.org/packages/0c/57/19f3a65bcfd5be570ee8c35a5398496e18a0ddcbc95393b2d17f86aaaf8/python_dateutil-2.7.2-py2.py3-none-any.whl (212kB)
    100% |#####| 215kB 303kB/s
Collecting pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 (from matplotlib)
  Downloading https://files.pythonhosted.org/packages/6a/8a/718fd7d3458f9fab8e67186b00abdd345b639976bc7fb3ae722e1b026a50/pyparsing-2.2.0-py3-none-any.whl (56kB)
    100% |#####| 61kB 397kB/s
Collecting setuptools (from kiwisolver>=1.0.1->matplotlib)
  Downloading https://files.pythonhosted.org/packages/20/d7/04a0b689d3035143e2ff288f4b9ee4bf6ed80585cc121c90bfd85a1a8c2e/setuptools-39.0.1-py2.py3-none-any.whl (569kB)
    100% |#####| 573kB 309kB/s
Building wheels for collected packages: subprocess32
  Running setup.py bdist_wheel for subprocess32 ... done
  Stored in directory: /home/hadoopadmin/.cache/pip/wheels/31/d3/4d/8151bf3fba7d27d0c9d03b78d4ff16097a8b62b298dc937c8d
Successfully built subprocess32
Installing collected packages: six, cyclers, numpy, backports.functiontools-lru-cache, subprocess32, setuptools, kiwisolver, pytz, python-dateutil, pyparsing, matplotlib
Successfully installed backports.functiontools-lru-cache-1.5 cyclers-0.10.0 kiwisolver-1.0.1 matplotlib-2.2.2 numpy-1.14.2 pyparsing-2.2.0 python-dateutil-2.7.2 pytz-2018.4 setuptoo
ls-39.0.1 six-1.11.0 subprocess32-3.2.7
hadoopadmin@hadoopadmin-VirtualBox:~$

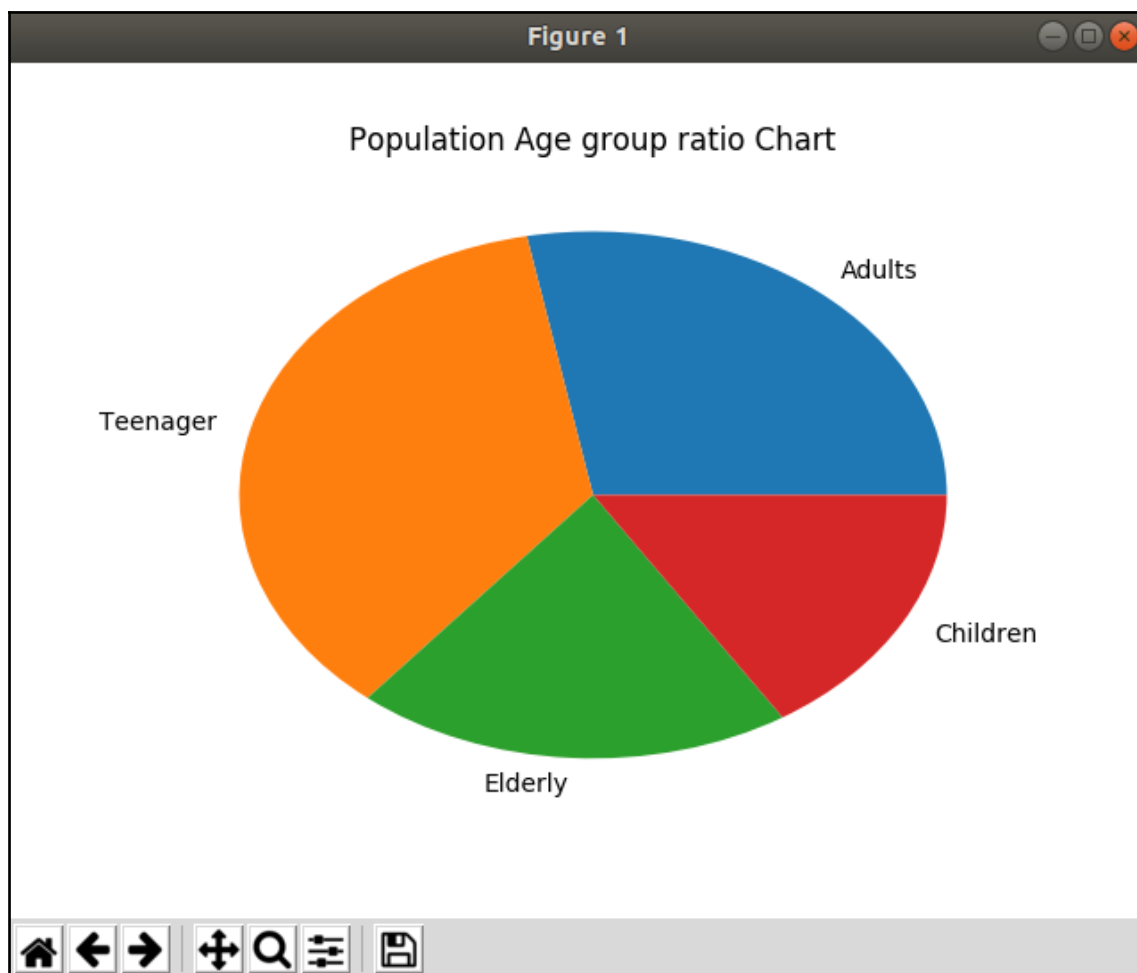
```

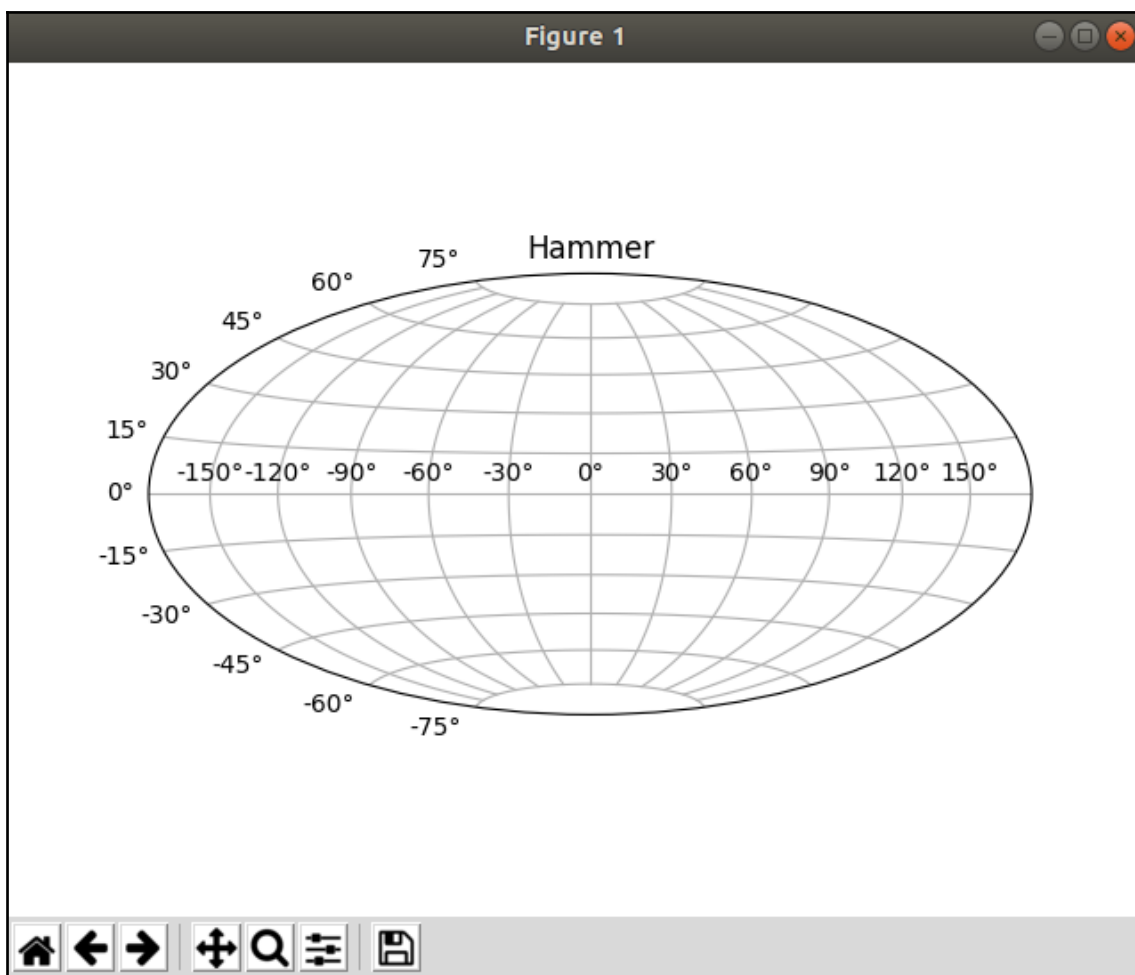


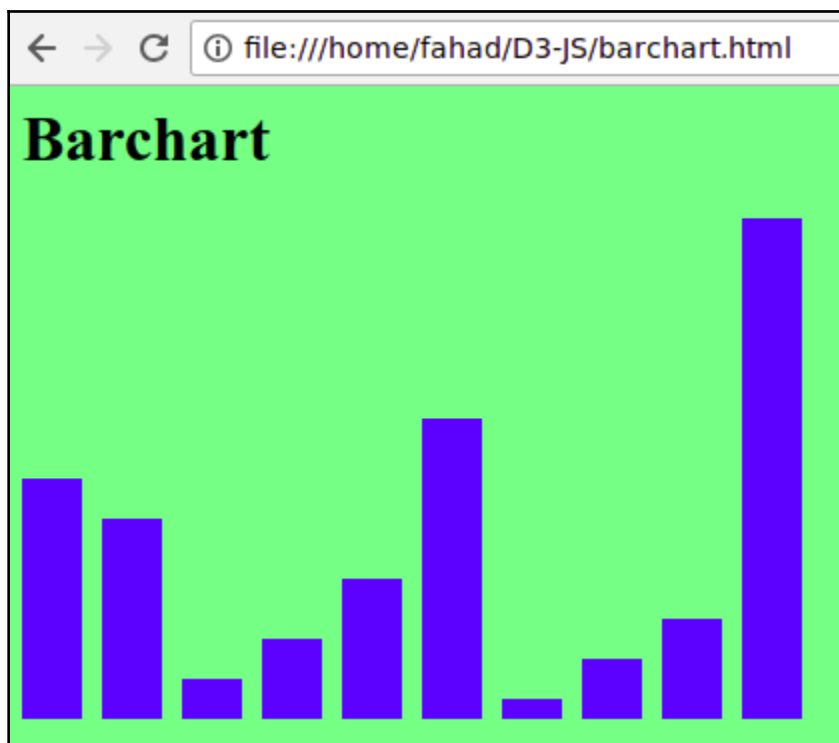












Chapter 18: Financial Trading system

Secure | <https://www.xm.com/register/account/demo?lang=en>

XM Homepage Member Login FAQ Contact About XM Regulation

LIVE CHAT ENGLISH

XM MEMBER OF TRADING POINT GROUP **URDU WEBINARS**

Demo Account Registration

Please complete the following form using English alphabet characters only.

Please note that you can only open up to five (5) demo accounts.

Personal Details

* First Name

* Last Name

* Country of Residence

* City / Town

* Phone

* E-mail

Why XM

- ▶ Regulated online broker
- ▶ Forex, metals, indices, energies
- ▶ Multiple online trading platforms
- ▶ Flexible trading conditions
- ▶ Instant deposits, same day withdrawals
- ▶ 24/5 support

Contact Us

Feel free to **contact us** by live chat, call back, email or freephone 24 hours a day, 5 days a week (Monday to Friday).

Risk Warning: Your capital is at risk. Leveraged products may not be suitable for everyone. Please consider our [Risk Disclosure](#).

Secure | https://www.xm.com/register/account/demo_success?lang=en

Demo Account Confirmation

Congratulations on opening an XM Demo Account

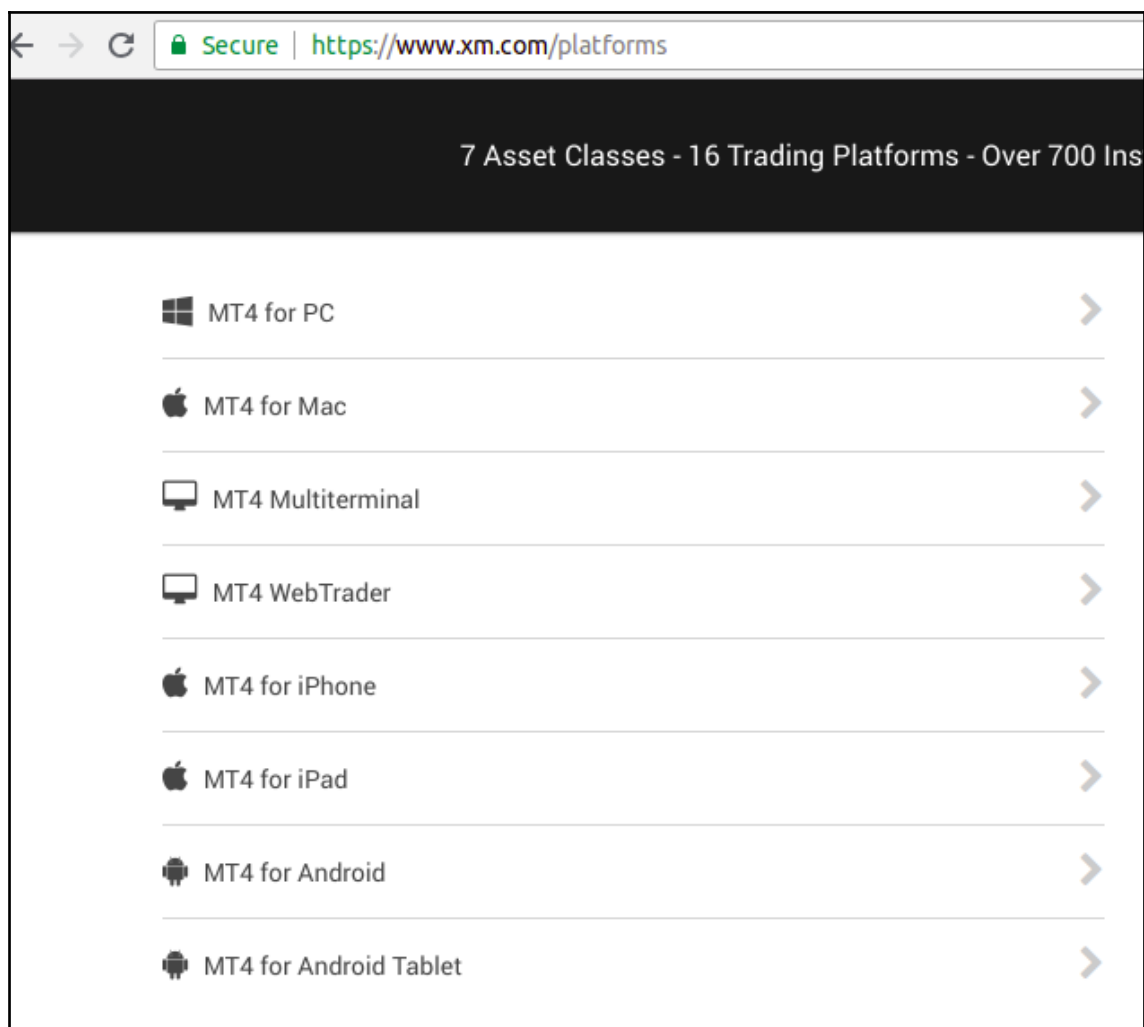
- ▶ To complete the registration process, please confirm your email by clicking on the link sent to the email address you have provided: me@smfahad.com
- ▶ You may need to check your spam folder if the message is not displayed in your inbox.

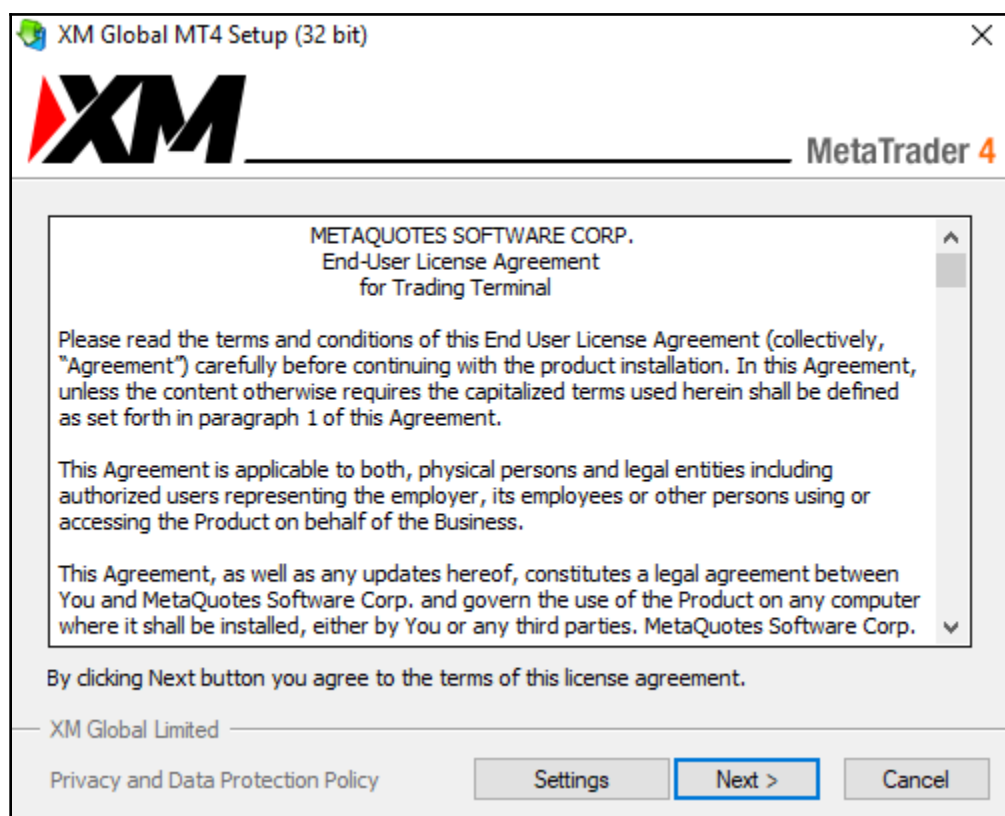
Demo Account Confirmation

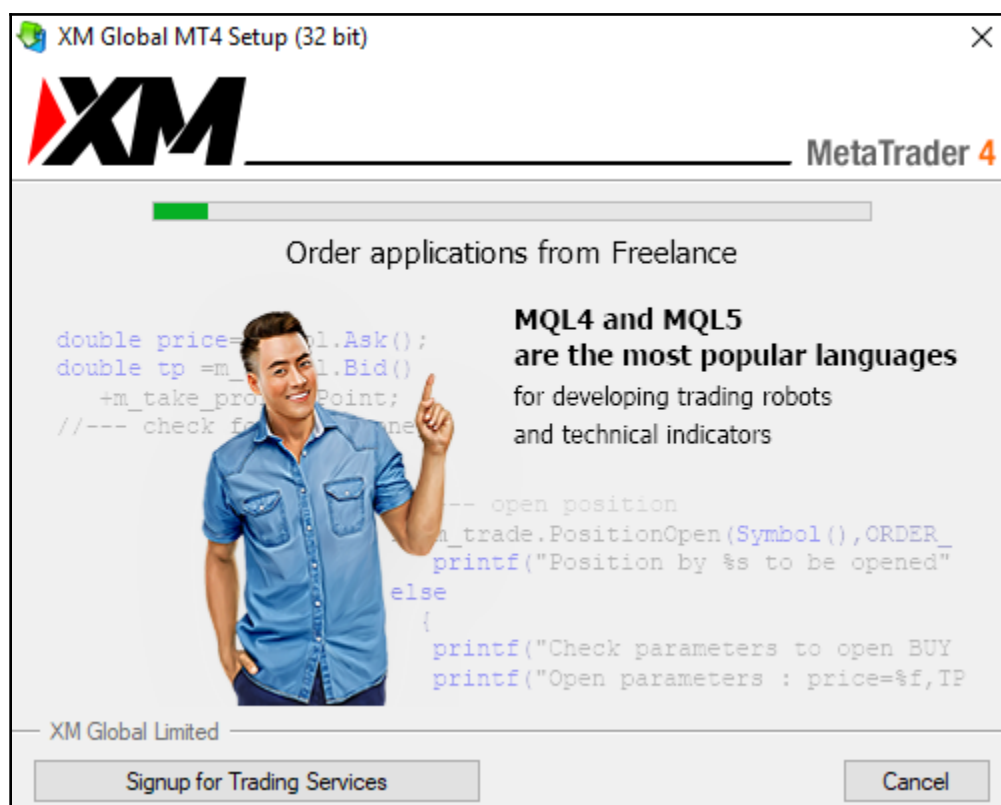


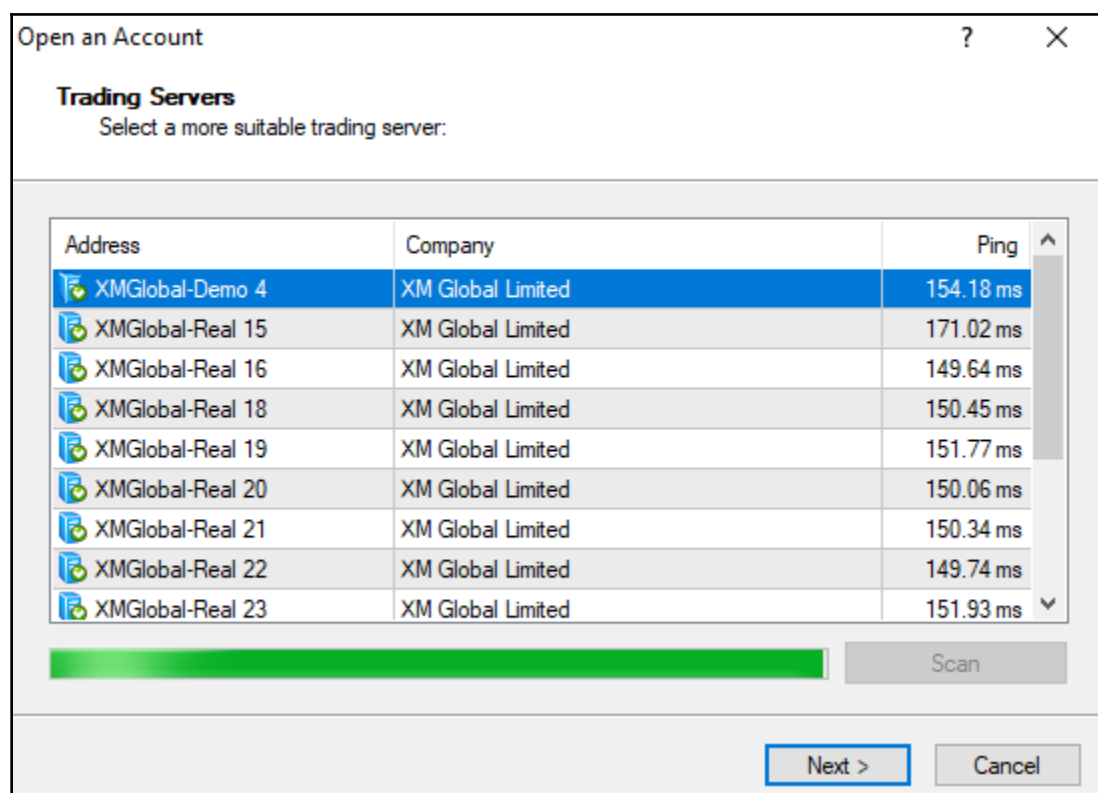
Congratulations on opening an XM Demo Account

- ▶ You have successfully opened a Demo Account that you can use for an unlimited period of time to practice trading and test your trading strategies risk free.
- ▶ Your MT4 ID is: **51384904**. Start trading on your demo account on the platform of your choice using the password created during your registration process and the MT4 ID given to you.
- ▶ Keep your login in a safe place. You can always request to have your password changed at any time by contacting our support team.









Open an Account

?

×

Select account type
To start working an account is required:

To start working, you need a valid trade account. If you don't have it, you can create a demo or real account by selecting the corresponding option below and pressing 'Next':

☒ Existing trade account

Login:

Password: ☒ Save password

Server: XMGlobal-Demo 4

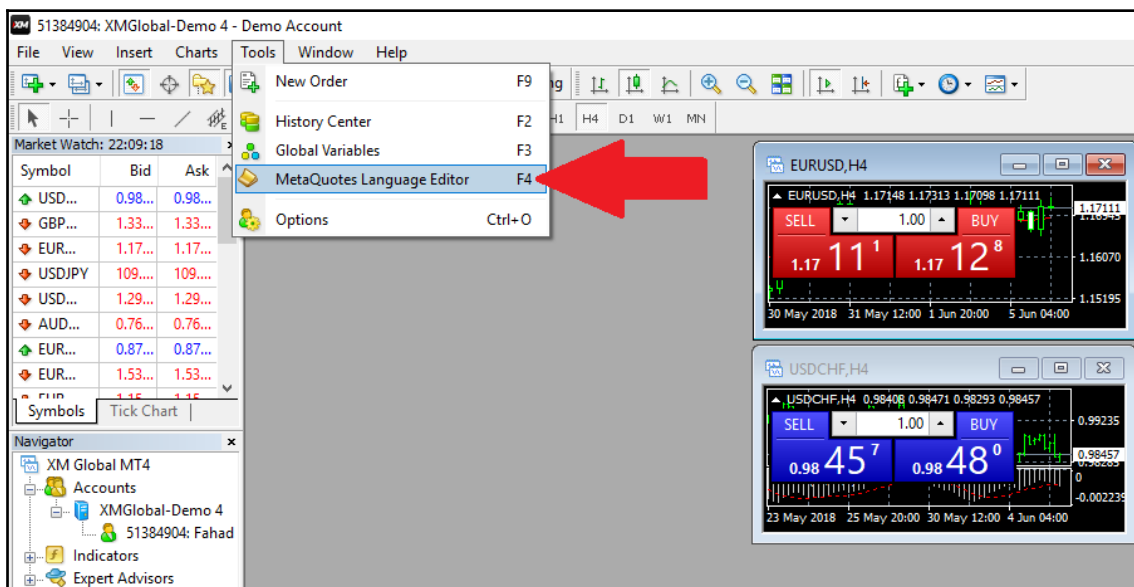
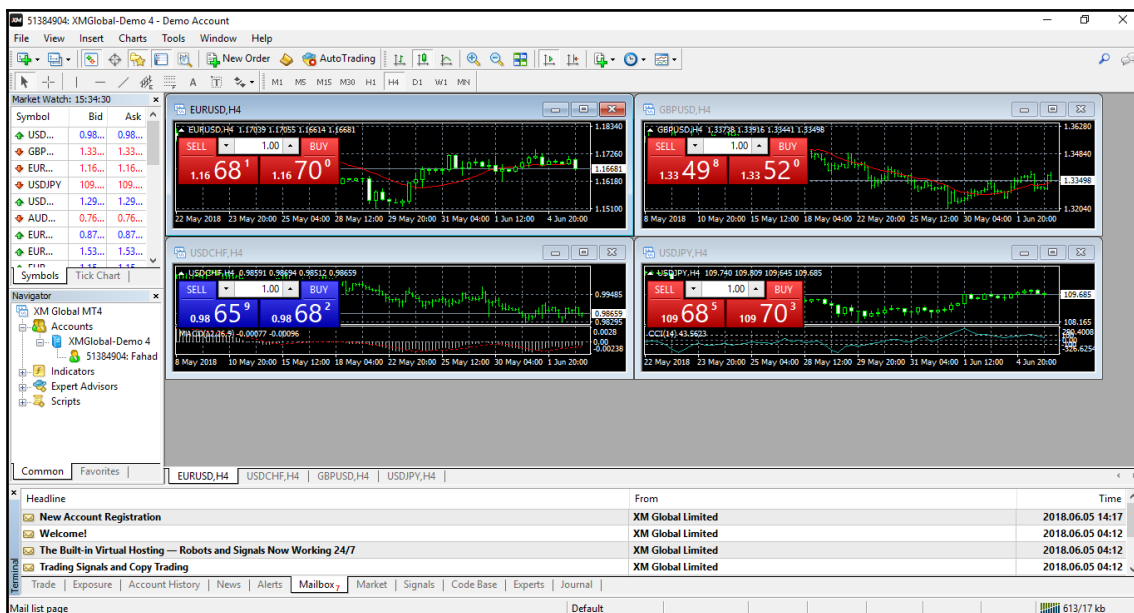
☐ New demo account

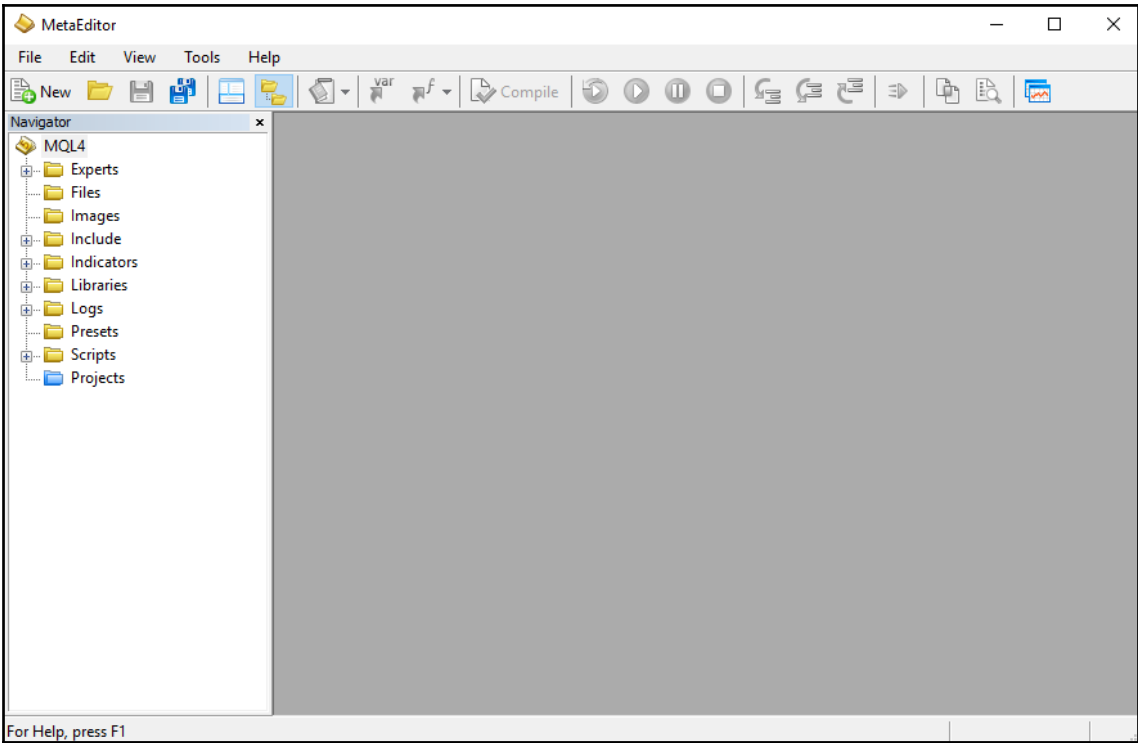
☐ New real account

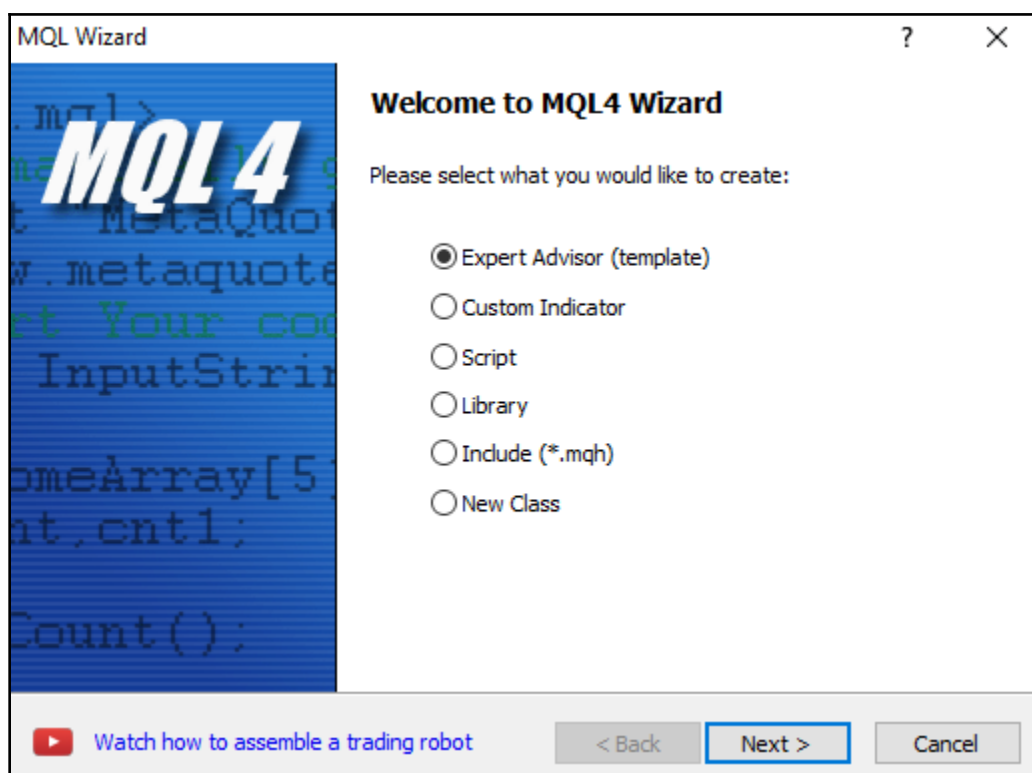
< Back

Finish

Cancel







MQL Wizard ? X

General properties of the Expert Advisor
Please specify general properties of the Expert Advisor.

Name:

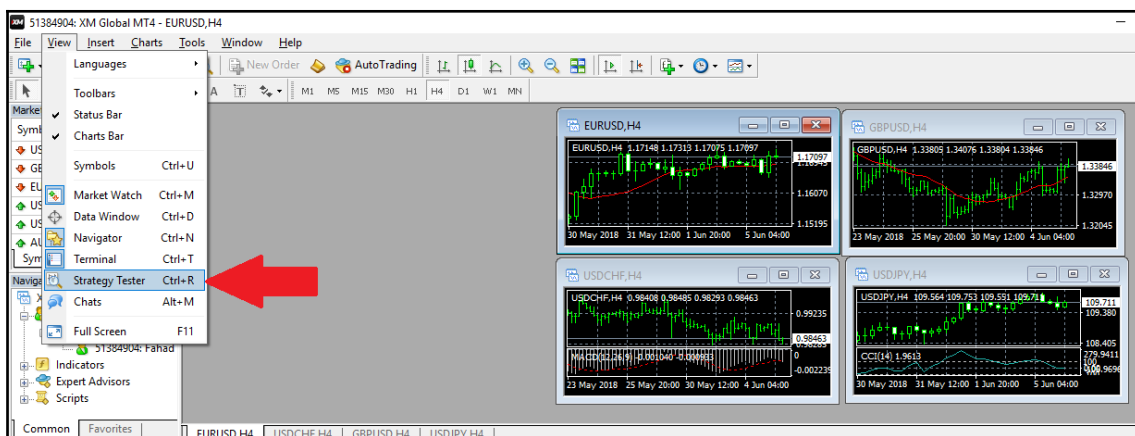
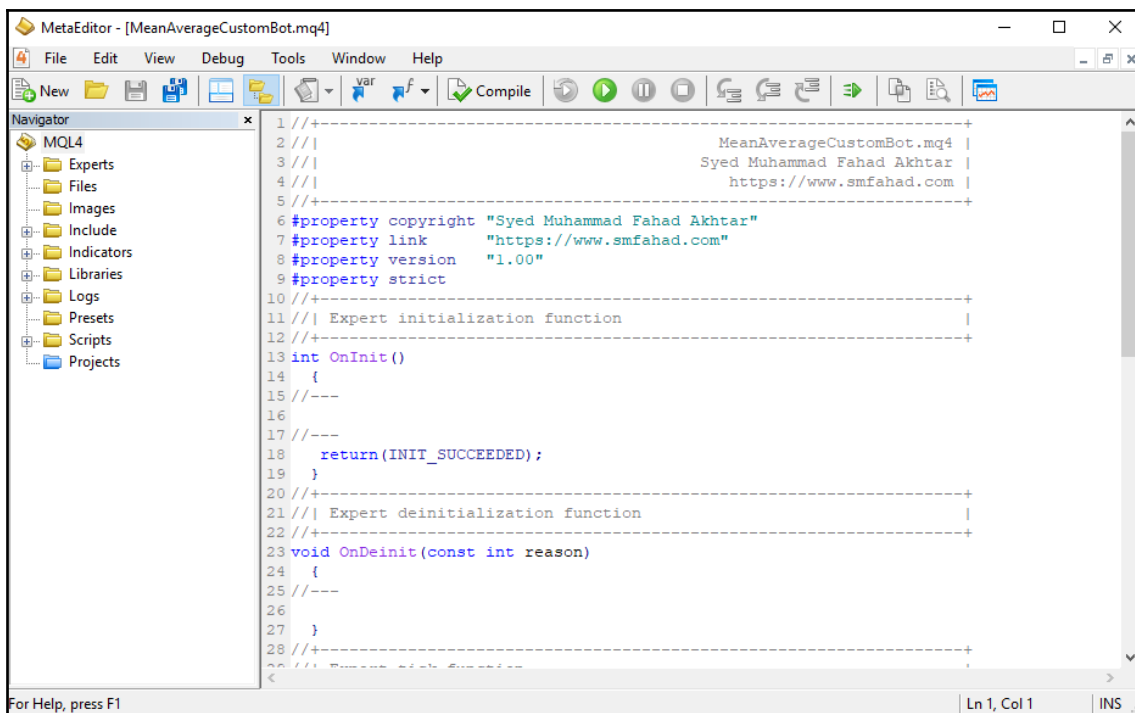
Author:

Link:

Parameters:

Name	Type	Initial value
------	------	---------------

 [Watch how to assemble a trading robot](#)



Expert Advisor **Moving Average ex4** Expert properties

Symbol: EURUSD, Euro vs US Dollar Period: H4 Symbol properties

Model: Every tick (the most precise method based on all available least timeframes to generate each tick) Spread: Current Open chart

Use date ☒ From: 2018.05.01 To: 2018.05.31 Optimization ☐ Modify expert

Visual mode ☐ Skip to: 2018.06.05

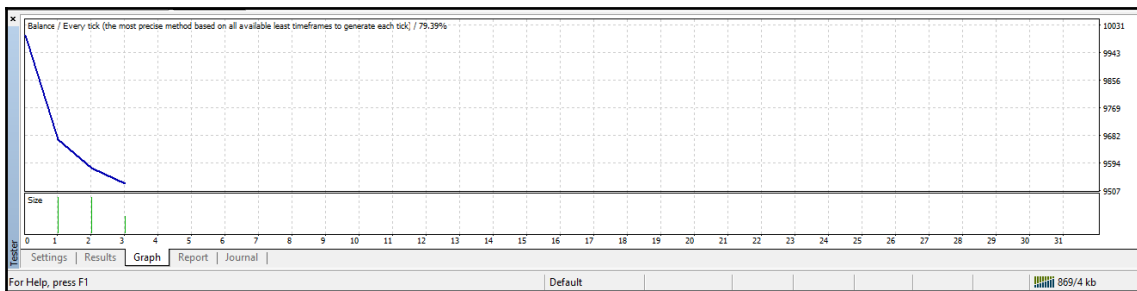
Start

Tester settings page Default 869/4 kb

#	Time	Type	Order	Size	Price	S / L	T / P	Profit	Balance
1	2018.05.10 16:00	buy	1	0.20	1.19282				
2	2018.05.22 08:00	close	1	0.20	1.17733			-331.22	9668.78
3	2018.05.22 12:00	buy	2	0.20	1.18091				
4	2018.05.23 08:00	close	2	0.20	1.17652			-89.94	9578.84
5	2018.05.28 08:00	buy	3	0.10	1.17253				
6	2018.05.28 12:00	close	3	0.10	1.16772			-48.10	9530.74

Tester Settings Results Graph Report Journal

For Help, press F1 Default 869/4 kb

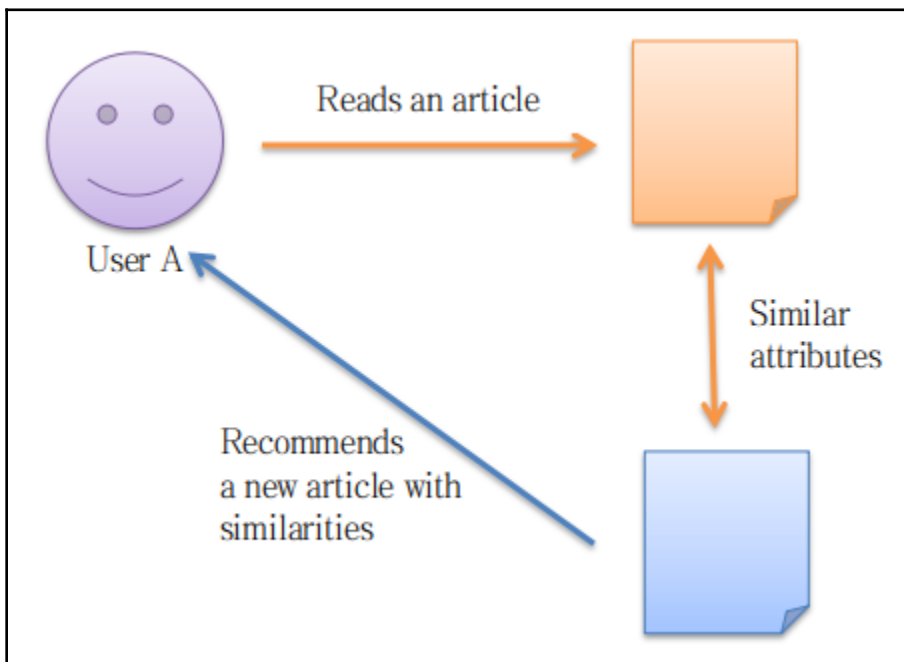
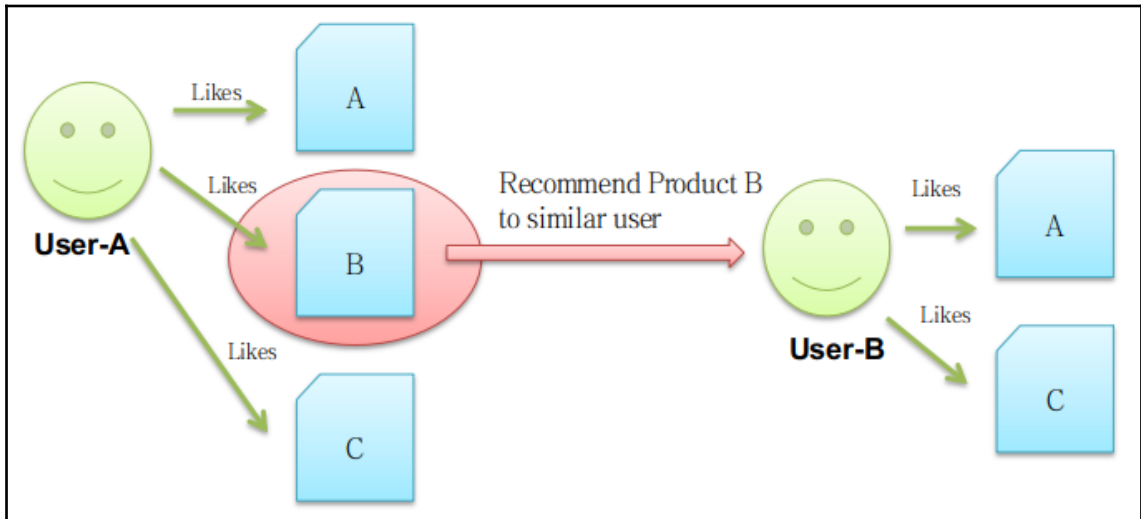


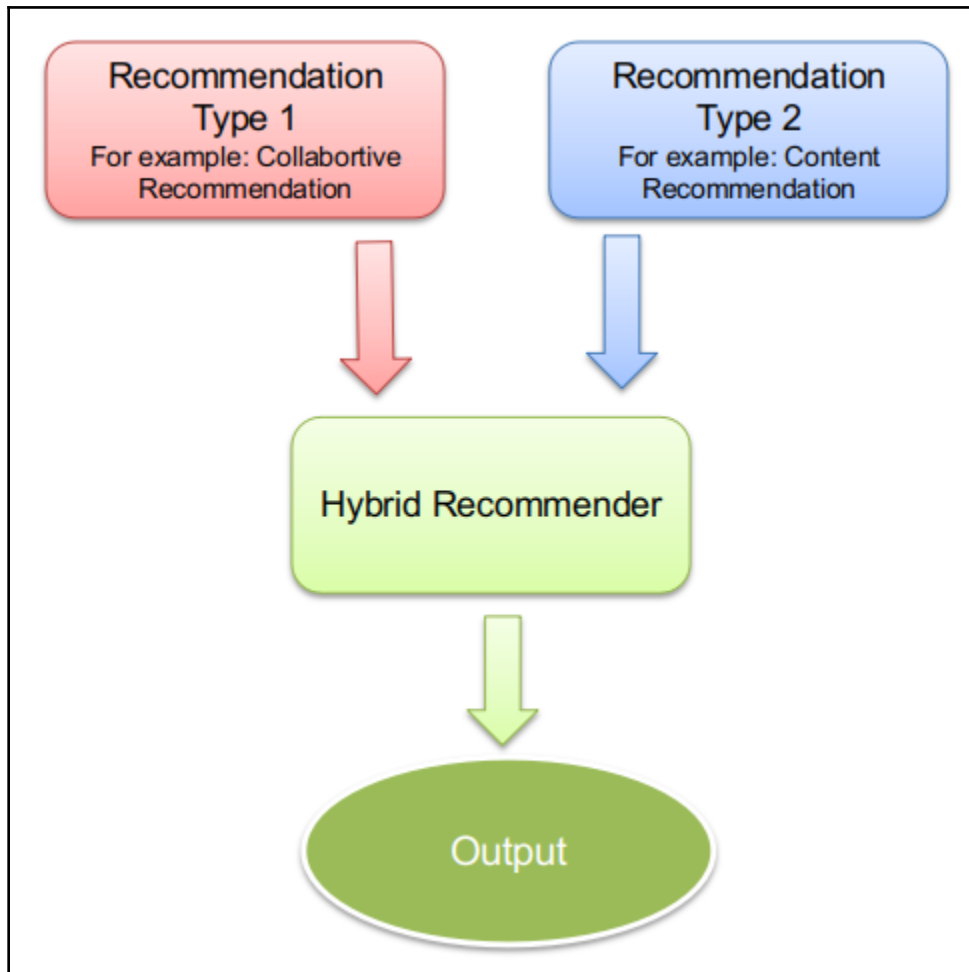
Bars in test	1133	Ticks modelled	1096639	Modelling quality	79.39%
Mismatched charts errors	0				
Initial deposit	10000.00			Spread	Current (17)
Total net profit	-469.26	Gross profit	0.00	Gross loss	-469.26
Profit factor	0.00	Expected payoff	-156.42		
Absolute drawdown	471.36	Maximal drawdown	601.28 (5.94%)	Relative drawdown	5.94% (601.28)
Total trades	3	Short positions (won %)	0 (0.00%)	Long positions (won %)	3 (0.00%)
		Profit trades (% of total)	0 (0.00%)	Loss trades (% of total)	3 (100.00%)
		Largest profit trade	0.00	loss trade	-331.22
		Average profit trade	0.00	loss trade	-156.42
		Maximum consecutive wins (profit in money)	0 (0.00)	consecutive losses (loss in money)	3 (-469.26)
		Maximal consecutive profit (count of wins)	0.00 (0)	consecutive loss (count of losses)	-469.26 (3)
		Average consecutive wins	0	consecutive losses	3

Tester Settings Results Graph Report Journal

For Help, press F1 Default 2018.05.31 16:00 O: 1.16718 H: 1.16976 L: 1.16402 C: 1.16697 V: 22063 869/4 kb

Chapter 19: Retail Recommendation System





```
print(books.shape)
print(list(books.columns))

(271360, 8)
['ISBN', 'Book-Title', 'Book-Author', 'Year-Of-Publication', 'Publisher', 'Image-URL-S', 'Image-URL-M', 'Image-URL-L']
```

```
print(users.shape)
print(list(users.columns))

(278858, 3)
['User-ID', 'Location', 'Age']
```

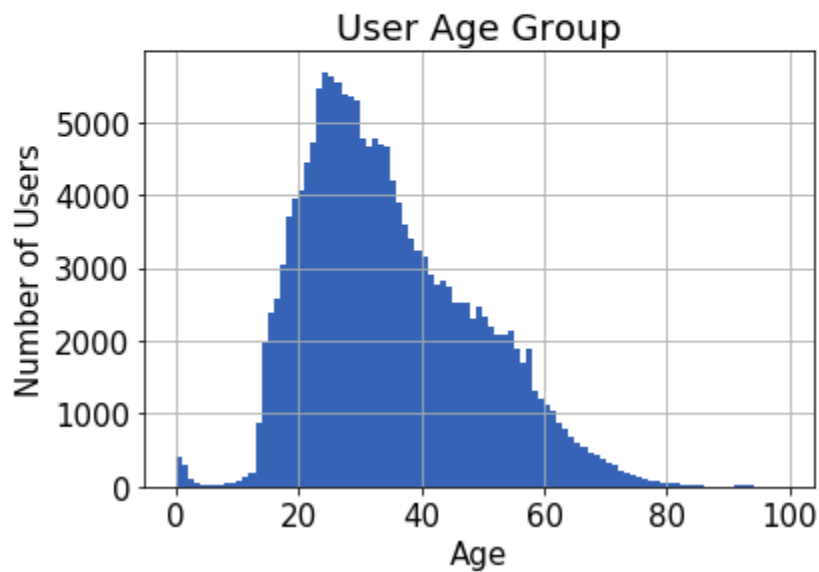
```
print(ratings.shape)
print(list(ratings.columns))

(1149780, 3)
['User-ID', 'ISBN', 'Book-Rating']
```

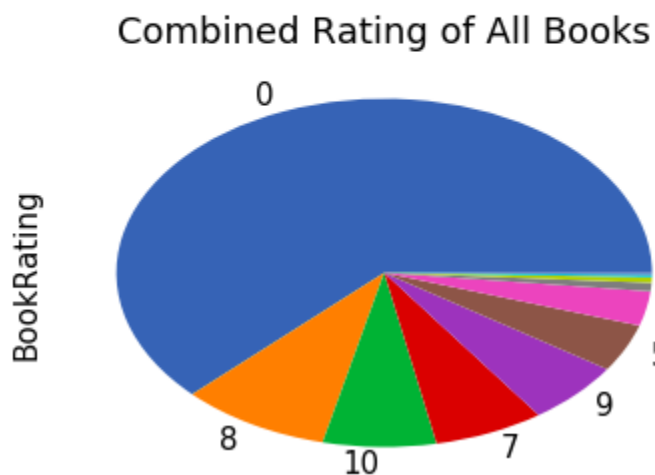
books.head()

	ISBN	Book-Title	Book-Author	Year-Of-Publication	Publisher	Image-URL-S	Image-URL-
0	0195153448	Classical Mythology	Mark P. O. Morford	2002	Oxford University Press	http://images.amazon.com/images/P/0195153448.0...	http://images.amazon.com/images/P/0195153448.
1	0002005018	Clara Callan	Richard Bruce Wright	2001	HarperFlamingo Canada	http://images.amazon.com/images/P/0002005018.0...	http://images.amazon.com/images/P/0002005018.
2	0060973129	Decision in Normandy	Carlo D'Este	1991	HarperPerennial	http://images.amazon.com/images/P/0060973129.0...	http://images.amazon.com/images/P/0060973129.
3	0374157065	Flu: The Story of the Great Influenza Pandemic...	Gina Bari Kolata	1999	Farrar Straus Giroux	http://images.amazon.com/images/P/0374157065.0...	http://images.amazon.com/images/P/0374157065.
4	0393045218	The Mummies of Urumchi	E. J. W. Barber	1999	W. W. Norton & Company	http://images.amazon.com/images/P/0393045218.0...	http://images.amazon.com/images/P/0393045218.

```
users.Age.hist(bins=range(0,100))  
  
plt.title('User Age Group')  
plt.xlabel('Age')  
plt.ylabel('Number of Users')  
plt.show()
```



```
ratings.BookRating.value_counts().plot(kind='pie')
plt.title('Combined Rating of All Books')
plt.show()
```



```
recommender_byRating = pd.DataFrame(ratings.groupby('ISBN')['Book-Rating'].count())
recommender_byRating.sort_values('Book-Rating', ascending=False).head()
```

Book-Rating	
ISBN	
0971880107	2502
0316666343	1295
0385504209	883
0060928336	732
0312195516	723

```
overall_rating = pd.merge(ratings, books, on='ISBN')

book Rated = (overall_rating.
               groupby(by = ['Book-Title']))['Book-Rating'].
               count().reset_index())

book Rated.head()
```

	Book-Title	Book-Rating
0	A Light in the Storm: The Civil War Diary of ...	4
1	Always Have Popsicles	1
2	Apple Magic (The Collector's series)	1
3	Ask Lily (Young Women of Faith: Lily Series, ...	1
4	Beyond IBM: Leadership Marketing and Finance ...	1

```

overall_ratingCount = overall_rating.merge(book_ratingCount, left_on = 'Book-Title', right_on = 'Book-Title', how =
bookUser = overall_ratingCount.merge(users, left_on = 'User-ID', right_on = 'User-ID', how = 'left')
demographic = bookUser[bookUser['Location'].str.contains("australia|spain")]
demographic.head()

```

	User-ID	ISBN	Book-Rating_x	Book-Title	Book-Author	Year-Of-Publication	Publisher	Image-URL-S	
62	276727	0446520802	0	The Notebook	Nicholas Sparks	1996	Warner Books	http://images.amazon.com/images/P/0446520802.0...	http://images.amazon.com/image
81	30711	0446520802	6	The Notebook	Nicholas Sparks	1996	Warner Books	http://images.amazon.com/images/P/0446520802.0...	http://images.amazon.com/image
759	15418	055356451X	0	Night Sins	TAMI HOAG	1995	Bantam	http://images.amazon.com/images/P/055356451X.0...	http://images.amazon.com/image
801	76355	055356451X	0	Night Sins	TAMI HOAG	1995	Bantam	http://images.amazon.com/images/P/055356451X.0...	http://images.amazon.com/image
818	115488	055356451X	0	Night Sins	TAMI HOAG	1995	Bantam	http://images.amazon.com/images/P/055356451X.0...	http://images.amazon.com/image