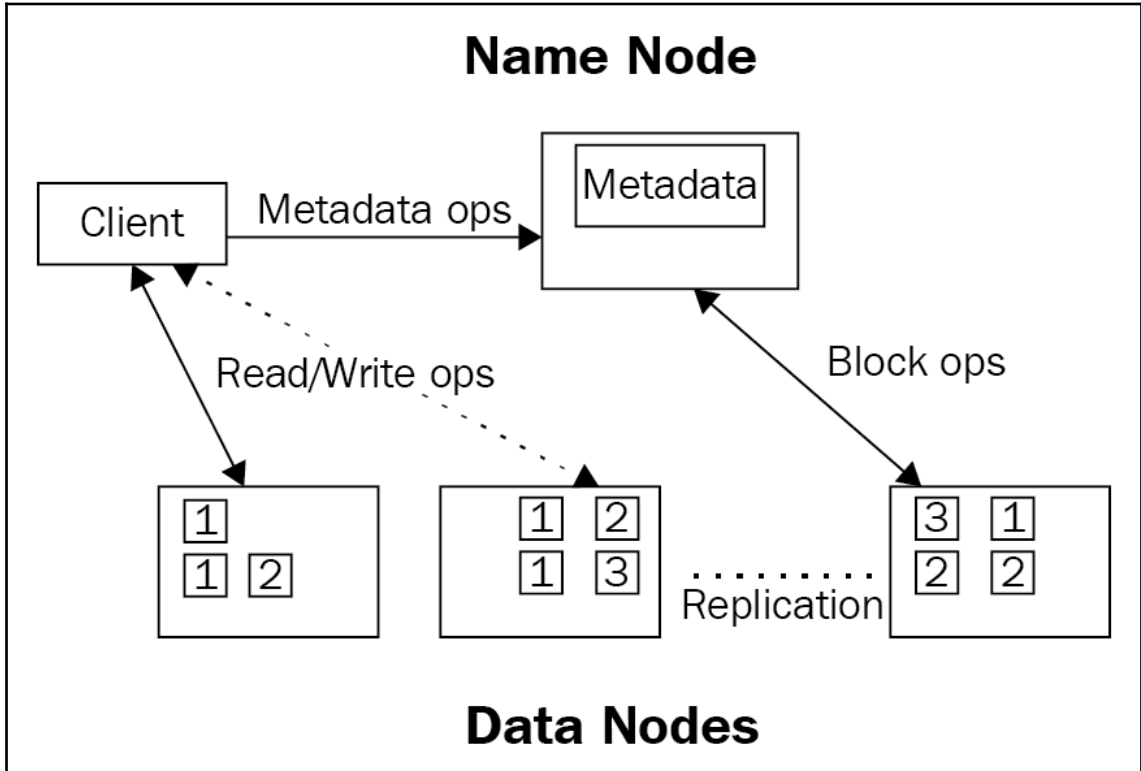
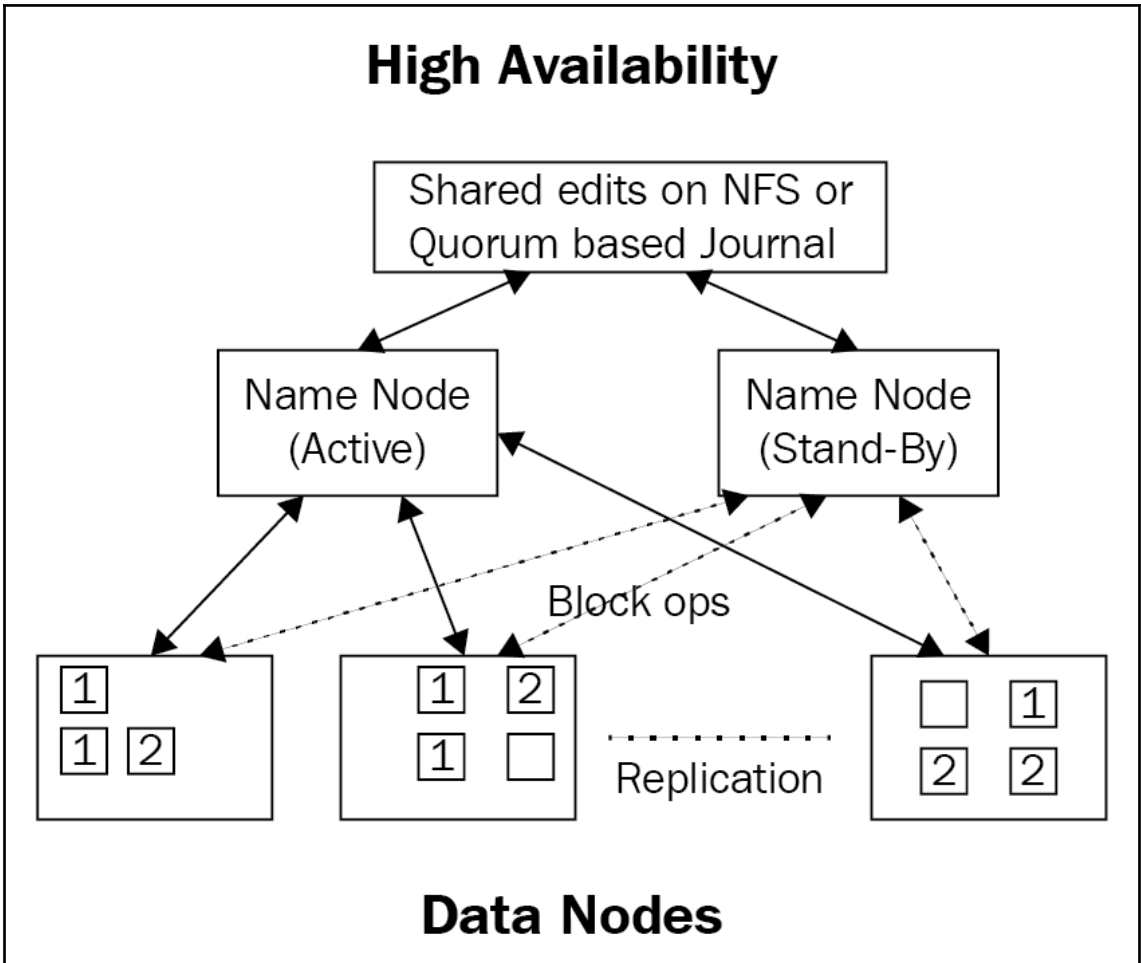
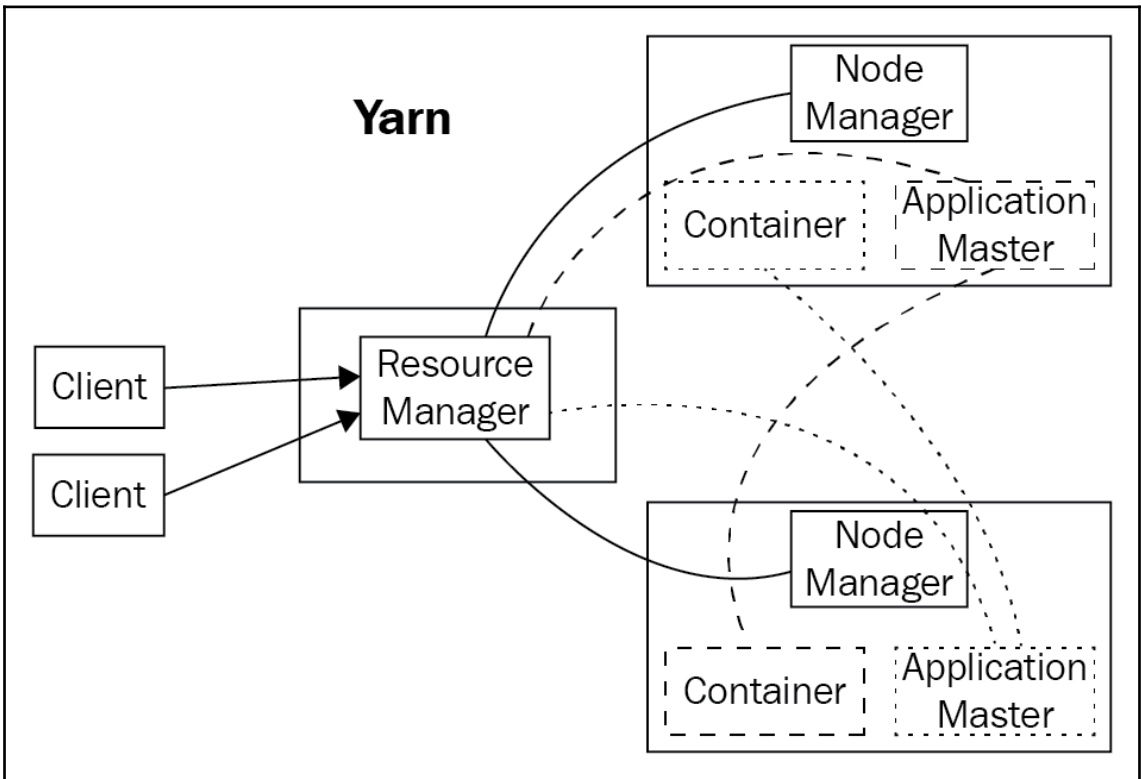
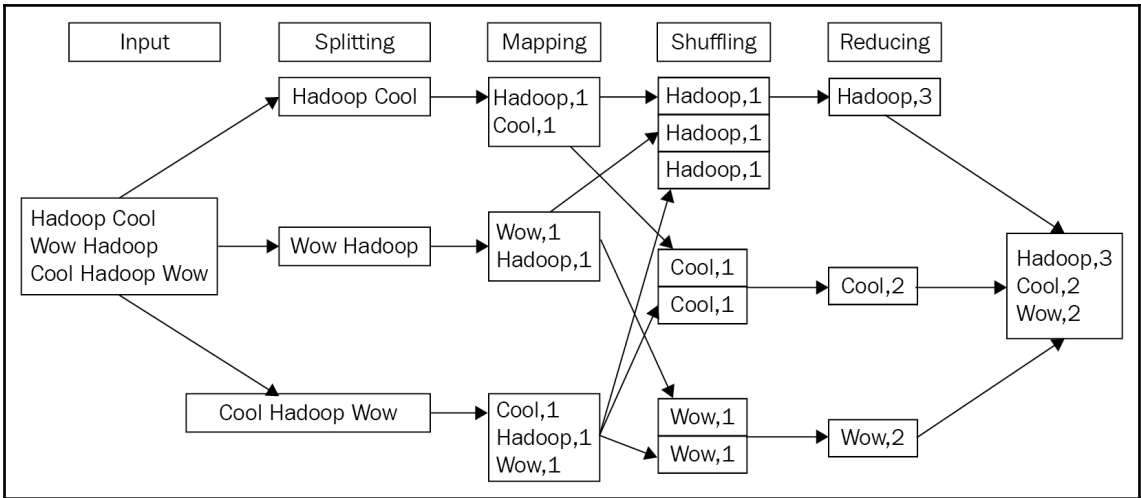


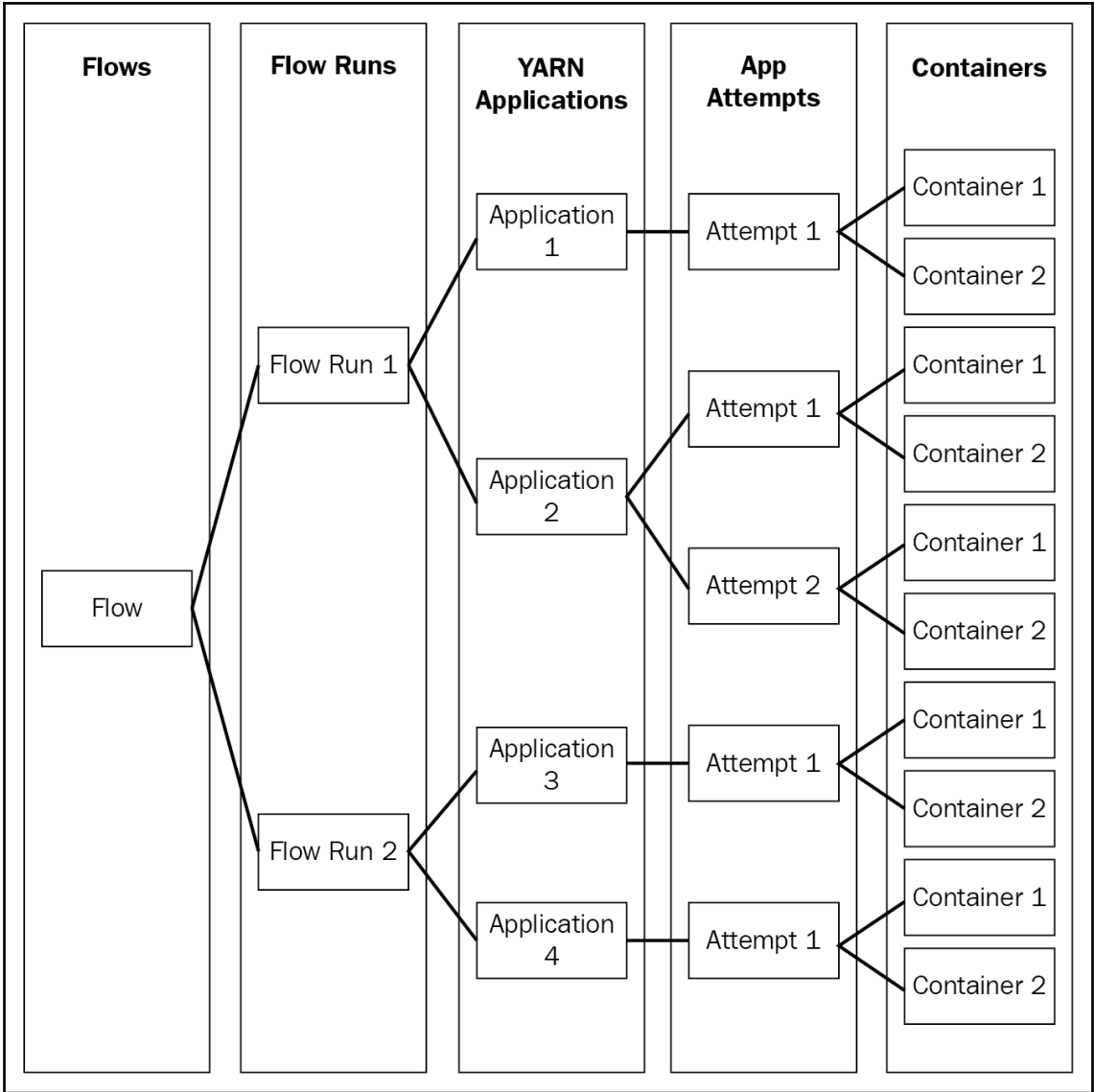
# Chapter 1: Introduction to Hadoop



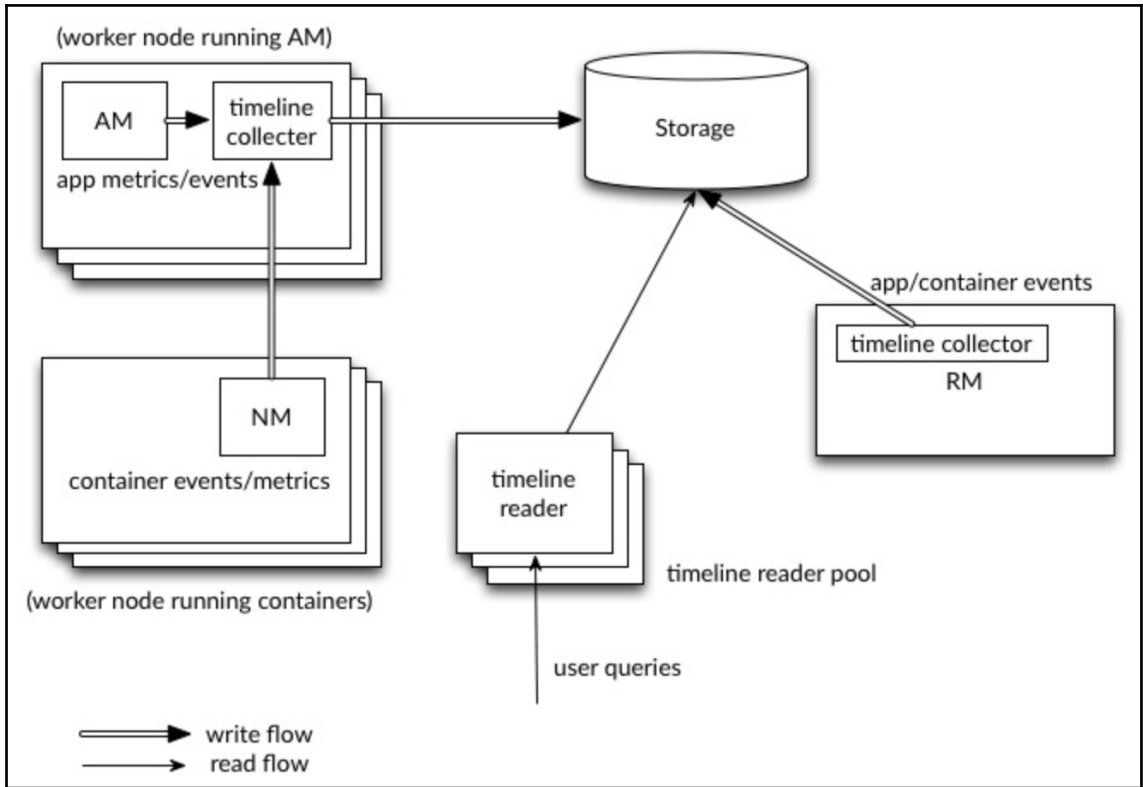
# High Availability












Oracle Corporation [US] | <https://www.java.com/en/download/>



Download Help

**All Java Downloads**

If you want to download Java for another computer or Operating System, click the link below.  
[All Java Downloads](#)

**Report an issue**

Why am I always redirected to this page when visiting a page with a Java app?  
 » [Learn more](#)  
 » [Report an issue](#)

## Free Java Download

Download Java for your desktop computer now!

**Version 8 Update 171**  
 Release date April 17, 2018

[Free Java Download](#)

» [What is Java?](#) » [Do I have Java?](#) » [Need Help?](#)

### Why download Java?

Java technology allows you to work and play in a secure computing environment. Upgrading to the latest Java version improves the security of your system, as older versions do not include the latest security updates.

Java allows you to play online games, chat with people around the world, calculate your mortgage interest, and view images in 3D, just to name a few.



Java software for your computer, or the Java Runtime Environment, is also referred to as the Java Runtime, Runtime Environment, Runtime, JRE, Java Virtual Machine, Virtual Machine, Java VM, JVM, VM, Java plug-in, Java plugin, Java add-on or Java download.

[Select Language](#) | [About Java](#) | [Support](#) | [Developers](#) | [Feedback](#)  
[Privacy](#) | [Cookie Preferences](#) | [Terms of Use](#) | [Trademarks](#) | [Disclaimer](#)

ORACLE

← → ↻ <https://apache.spinellcreations.com/hadoop/common/hadoop-3.1.0/>

# Index of /hadoop/common/hadoop-3.1.0

<a href="#">Name</a>	<a href="#">Last modified</a>	<a href="#">Size</a>	<a href="#">Description</a>
 <a href="#">Parent Directory</a>		-	
 <a href="#">hadoop-3.1.0-src.tar.gz</a>	05-Apr-2018 16:19	26M	
 <a href="#">hadoop-3.1.0.tar.gz</a>	05-Apr-2018 16:19	311M	

Apache/2.2.15 (CentOS) DAV/2 PHP/5.3.3 mod\_ssl/2.2.15 OpenSSL/1.0.0-fips Server at apache.spinellcreations.com Port 80

localhost:9870/dfshealth.html#tab-overview

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

## Overview 'localhost:9000' (active)

<b>Started:</b>	Sun Apr 22 21:49:18 -0400 2018
<b>Version:</b>	3.1.0, r16b70619a24cdc5d3b0fc4b58ca77238ccbe6d
<b>Compiled:</b>	Thu Mar 29 20:00:00 -0400 2018 by centos from branch-3.1.0
<b>Cluster ID:</b>	CID-6074d571-8638-49ab-9dcb-e24ed4866e50
<b>Block Pool ID:</b>	BP-770583072-10.0.0.103-1524448088168

## Summary

Security is off.  
 Safemode is off.

1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).

Heap Memory used 78.58 MB of 432 MB Heap Memory. Max Heap Memory is 3.56 GB.

Non Heap Memory used 46.55 MB of 47.84 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

<b>Configured Capacity:</b>	233.47 GB
<b>Configured Remote Capacity:</b>	0 B
<b>DFS Used:</b>	4 KB (0%)
<b>Non DFS Used:</b>	206.34 GB
<b>DFS Remaining:</b>	23.5 GB (10.07%)
<b>Block Pool Used:</b>	4 KB (0%)

localhost:9870/dfshealth.html#tab-datanode

Hadoop Overview **Datanodes** Datanode Volume Failures Snapshot Startup Progress Utilities

## Datanode Information

✔ In service 
 ● Down 
 ⊗ Decommissioned 
 ⊗ Decommissioned & dead 
 ⚡ In Maintenance & dead

Datanode usage histogram

Disk usage of each DataNode (%)

In operation

Show 25 entries Search:

Node	Http Address	Last contact	Last Block Report	Capacity	Blocks	Block pool used	Version
✔ 10.0.0.103-9866 (127.0.0.1:9866)	http://10.0.0.103:9864	0s	1m	233.47 GB	0	4 KB (0%)	3.1.0

Showing 1 to 1 of 1 entries Previous 1 Next

localhost:9870/logs/

## Directory: /logs/

<a href="#">SecurityAuth-sridharalla.audit</a>	0 bytes	Apr 22, 2018 9:31:12 PM
<a href="#">hadoop-sridharalla-datanode-Moogie.local.log</a>	31693 bytes	Apr 22, 2018 9:49:22 PM
<a href="#">hadoop-sridharalla-datanode-Moogie.local.out</a>	485 bytes	Apr 22, 2018 9:49:21 PM
<a href="#">hadoop-sridharalla-namenode-Moogie.local.log</a>	39439 bytes	Apr 22, 2018 9:50:24 PM
<a href="#">hadoop-sridharalla-namenode-Moogie.local.out</a>	485 bytes	Apr 22, 2018 9:49:19 PM
<a href="#">hadoop-sridharalla-secondarynamenode-Moogie.local.log</a>	33138 bytes	Apr 22, 2018 9:50:24 PM
<a href="#">hadoop-sridharalla-secondarynamenode-Moogie.local.out</a>	485 bytes	Apr 22, 2018 9:49:24 PM

---

← → ↻ ⓘ localhost:9870/jmx

```
{
  "beans" : [ {
    "name" : "Hadoop:service=NameNode,name=JvmMetrics",
    "modelerType" : "JvmMetrics",
    "tag.Context" : "jvm",
    "tag.ProcessName" : "NameNode",
    "tag.SessionId" : null,
    "tag.Hostname" : "Moogie.local",
    "MemNonHeapUsedM" : 47.575027,
    "MemNonHeapCommittedM" : 49.148438,
    "MemNonHeapMaxM" : -1.0,
    "MemHeapUsedM" : 103.684074,
    "MemHeapCommittedM" : 432.0,
    "MemHeapMaxM" : 3641.0,
    "MemMaxM" : 3641.0,
    "GcCount" : 8,
    "GcTimeMillis" : 247,
    "GcNumWarnThresholdExceeded" : 0,
    "GcNumInfoThresholdExceeded" : 0,
    "GcTotalExtraSleepTime" : 435,
    "ThreadsNew" : 0,
    "ThreadsRunnable" : 10,
    "ThreadsBlocked" : 0,
    "ThreadsWaiting" : 6,
    "ThreadsTimedWaiting" : 36,
    "ThreadsTerminated" : 0,
    "LogFatal" : 0,
    "LogError" : 0,
    "LogWarn" : 8,
    "LogInfo" : 130
  }, {
    "name" : "JMImplementation:type=MBeanServerDelegate",
    "modelerType" : "javax.management.MBeanServerDelegate",
    "MBeanServerId" : "Moogie.local_1524448157872",
    "SpecificationName" : "Java Management Extensions",
    "SpecificationVersion" : "1.4",
```

localhost:9870/conf

This XML file does not appear to have any style information associated with it. The document tree is shown below

```

<configuration>
  <property>
    <name>mapreduce.jobhistory.jhist.format</name>
    <value>binary</value>
    <final>false</final>
    <source>mapred-default.xml</source>
  </property>
  <property>
    <name>fs.s3a.retry.interval</name>
    <value>500ms</value>
    <final>false</final>
    <source>core-default.xml</source>
  </property>
  <property>
    <name>dfs.block.access.token.lifetime</name>
    <value>600</value>
    <final>false</final>
    <source>hdfs-default.xml</source>
  </property>
  <property>
    <name>mapreduce.job.heap.memory-mb.ratio</name>
    <value>0.8</value>
    <final>false</final>
    <source>mapred-default.xml</source>
  </property>
  <property>

```

localhost:9870/explorer.html#

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

## Browse Directory

/ Go

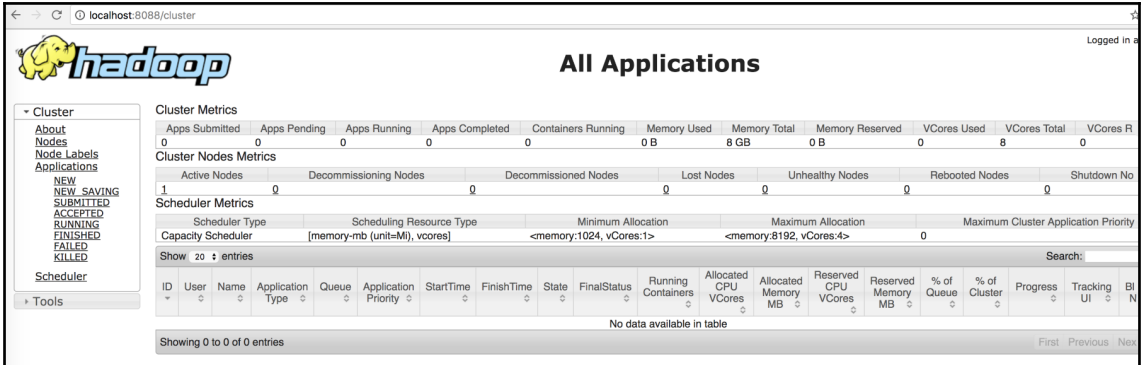
Show 25 entries Search:

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	sridharalla	supergroup	0 B	Apr 22 21:51	0	0 B	user

Showing 1 to 1 of 1 entries

Previous 1 Next

Hadoop, 2018.



localhost:8088/cluster

## All Applications

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores R
0	0	0	0	0	0 B	8 GB	0 B	0	8	0

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes	Shutdown No
1	0	0	0	0	0	0

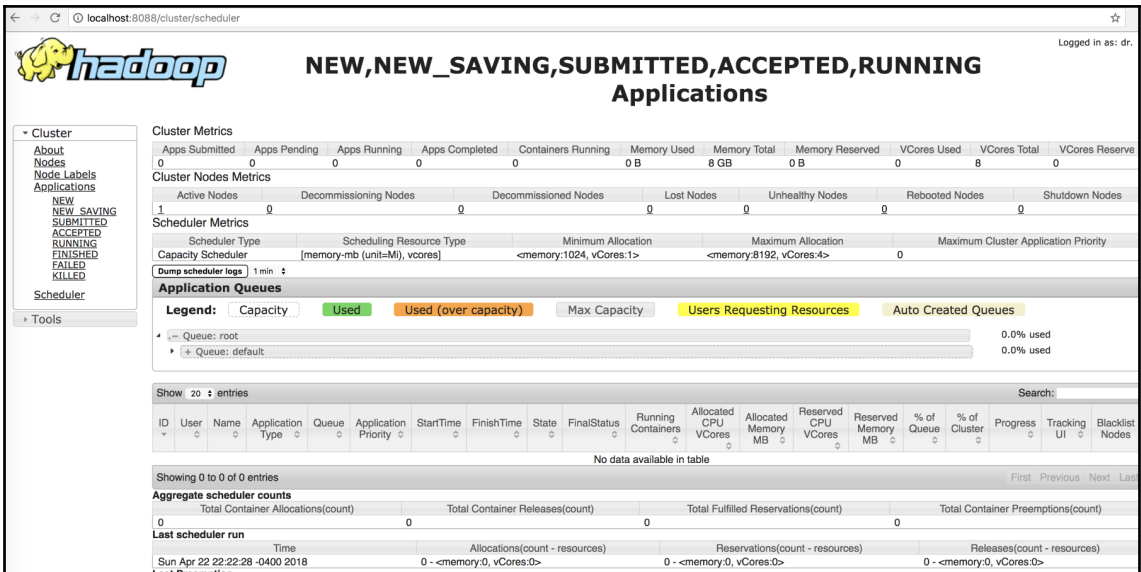
Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation	Maximum Cluster Application Priority
Capacity Scheduler	[memory-mb (unit=M), vcores]	<memory:1024, vCores:1>	<memory:8192, vCores:4>	0

Showing 0 to 0 of 0 entries

ID	User	Name	Application Type	Queue	Application Priority	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCores	Allocated Memory MB	Reserved CPU VCores	Reserved Memory MB	% of Queue	% of Cluster	Progress	Tracking UI	Bl N
No data available in table																			

Showing 0 to 0 of 0 entries



localhost:8088/cluster/scheduler

## NEW,NEW\_SAVING,SUBMITTED,ACCEPTED,RUNNING Applications

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserve
0	0	0	0	0	0 B	8 GB	0 B	0	8	0

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes	Shutdown Nodes
1	0	0	0	0	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation	Maximum Cluster Application Priority
Capacity Scheduler	[memory-mb (unit=M), vcores]	<memory:1024, vCores:1>	<memory:8192, vCores:4>	0

Dump scheduler logs 1 min

Application Queues

Legend: Capacity (Used, Used (over capacity), Max Capacity, Users Requesting Resources, Auto Created Queues)

- Queue: root 0.0% used
- Queue: default 0.0% used

Showing 0 to 0 of 0 entries

ID	User	Name	Application Type	Queue	Application Priority	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCores	Allocated Memory MB	Reserved CPU VCores	Reserved Memory MB	% of Queue	% of Cluster	Progress	Tracking UI	Blacklist Nodes
No data available in table																			

Showing 0 to 0 of 0 entries

Aggregate scheduler counts

Total Container Allocations(count)	Total Container Releases(count)	Total Fulfilled Reservations(count)	Total Container Preemptions(count)
0	0	0	0

Last scheduler run

Time	Allocations(count - resources)	Reservations(count - resources)	Releases(count - resources)
Sun Apr 22 22:22:28 -0400 2018	0 - <memory:0, vCores:0>	0 - <memory:0, vCores:0>	0 - <memory:0, vCores:0>

```

Moogie:hadoop-3.1.0 sridharalla$ ./bin/hdfs ec -listPolicies
2018-04-22 22:55:59,587 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Erasure Coding Policies:
ErasureCodingPolicy=[Name=RS-10-4-1024k, Schema=[ECSchema=[Codec=rs, numDataUnits=10, numParityUnits=4]], CellSize=1048576, Id=5], State=DISABLED
ErasureCodingPolicy=[Name=RS-3-2-1024k, Schema=[ECSchema=[Codec=rs, numDataUnits=3, numParityUnits=2]], CellSize=1048576, Id=2], State=DISABLED
ErasureCodingPolicy=[Name=RS-6-3-1024k, Schema=[ECSchema=[Codec=rs, numDataUnits=6, numParityUnits=3]], CellSize=1048576, Id=1], State=ENABLED
ErasureCodingPolicy=[Name=RS-LEGACY-6-3-1024k, Schema=[ECSchema=[Codec=rs-legacy, numDataUnits=6, numParityUnits=3]], CellSize=1048576, Id=3], State=DISABLED
ErasureCodingPolicy=[Name=XOR-2-1-1024k, Schema=[ECSchema=[Codec=xor, numDataUnits=2, numParityUnits=1]], CellSize=1048576, Id=4], State=DISABLED

```

```

Moogie:hadoop-3.1.0 sridharalla$ ./bin/hdfs dfs -copyFromLocal ~/Documents/OnlineRetail.csv /user/ec
2018-04-22 23:48:25,379 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes
where applicable
2018-04-22 23:48:26,284 WARN erasurecode.ErasureCodeNative: ISA-L support is not available in your platform... using builtin-java codec
here applicable
copyFromLocal: File /user/ec/OnlineRetail.csv._COPYING_ could only be written to 1 of the 6 required nodes for RS-6-3-1024k. There are 1
datanode(s) running and no node(s) are excluded in this operation.

```





```

Moogie:hadoop-3.1.0 sridharalla$ ./bin/hdfs diskbalancer -plan 10.0.0.103
2018-04-23 00:00:14,789 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes
where applicable
2018-04-23 00:00:15,691 INFO planner.GreedyPlanner: Starting plan for Node : 10.0.0.103:9867
2018-04-23 00:00:15,691 INFO planner.GreedyPlanner: Compute Plan for Node : 10.0.0.103:9867 took 16 ms
2018-04-23 00:00:15,692 INFO command.Command: No plan generated. DiskBalancing not needed for node: 10.0.0.103 threshold used: 10.0
No plan generated. DiskBalancing not needed for node: 10.0.0.103 threshold used: 10.0

```

← → ↻ [mirror.cogentco.com/pub/apache/hbase/1.2.6/](http://mirror.cogentco.com/pub/apache/hbase/1.2.6/)

## Index of /pub/apache/hbase/1.2.6

Name	Last modified	Size	Description
 <a href="#">Parent Directory</a>		-	
 <a href="#">1.2.5_1.2.6RC0_compat_report.html</a>	2017-05-29 02:37	24K	HBase: rel/1.2.5/d7b05f7 to 1.2.6RC0/2f9b9e1 compatibility report
 <a href="#">hbase-1.2.6-bin.tar.gz</a>	2017-05-29 10:36	100M	
 <a href="#">hbase-1.2.6-src.tar.gz</a>	2017-05-29 10:36	15M	

```

Moogie:~ sridharalla$ cd hbase-1.2.6
Moogie:hbase-1.2.6 sridharalla$ ls
CHANGES.txt      LICENSE.txt      README.txt      conf            hbase-webapps
LEGAL             NOTICE.txt     bin             docs            lib

```



localhost:16010/master-status

APACHE HBASE Home Table Details Local Logs Log Level Debug Dump Metrics Dump HBase Configuration

## Master 10.0.0.103

### Region Servers

Base Stats Memory Requests Storefiles Compactions

ServerName	Start time	Version	Requests Per Second
10.0.0.103,57540,1524457767491	Mon Apr 23 00:29:27 EDT 2018	1.2.6	0
Total:1			0

### Backup Masters

ServerName	Port	Start Time
Total:0		

### Tables

User Tables System Tables Snapshots

localhost:16010/master-status

APACHE HBASE Home Table Details Local Logs Log Level Debug Dump Metrics Dump HBase Configuration

Show All Monitored Tasks **Show non-RPC Tasks** Show All RPC Handler Tasks Show Active RPC Calls Show Client Operations View as JSON

No tasks currently running on this node.

## Software Attributes

Attribute Name	Value	Description
HBase Version	1.2.6, revision=Unknown	HBase version and revision
HBase Compiled	Mon May 29 02:25:32 CDT 2017, busbey	When HBase version was compiled and by whom
HBase Source Checksum	7e8ce83a648e252758e9dae1fbe779c9	HBase source MD5 checksum
Hadoop Version	2.5.1, revision=2e18d179e4a8065b6a9f29cf2de9451891265cce	Hadoop version and revision
Hadoop Compiled	2014-09-05T23:05Z, kasha	When Hadoop version was compiled and by whom
Hadoop Source Checksum	6424fcab95bfff8337780a181ad7c78	Hadoop source MD5 checksum
ZooKeeper Client Version	3.4.6, revision=1569965	ZooKeeper client version and revision
ZooKeeper Client Compiled	02/20/2014 09:09 GMT	When ZooKeeper client version was compiled
Zookeeper Quorum	localhost:2181	Addresses of all registered ZK servers. For more, see <a href="#">zk</a>
Zookeeper Base Path	/hbase	Root node of this cluster in ZK.
HBase Root Directory	hdfs://localhost:9000/hbase	Location of HBase home directory
HMaster Start Time	Mon Apr 23 00:29:26 EDT 2018	Date stamp of when this HMaster was started
HMaster Active Time	Mon Apr 23 00:29:27 EDT 2018	Date stamp of when this HMaster became active
HBase Cluster ID	8cacc1fe-fc23-4be1-b0df-a0c9f8b87be4	Unique identifier generated for each HBase cluster
Load average	2.00	Average number of regions per regionserver. Naive com

localhost:16010/master-status

APACHE HBASE Home Table Details Local Logs Log Level Debug Dump Metrics Dump HBase Configuration

## Tables

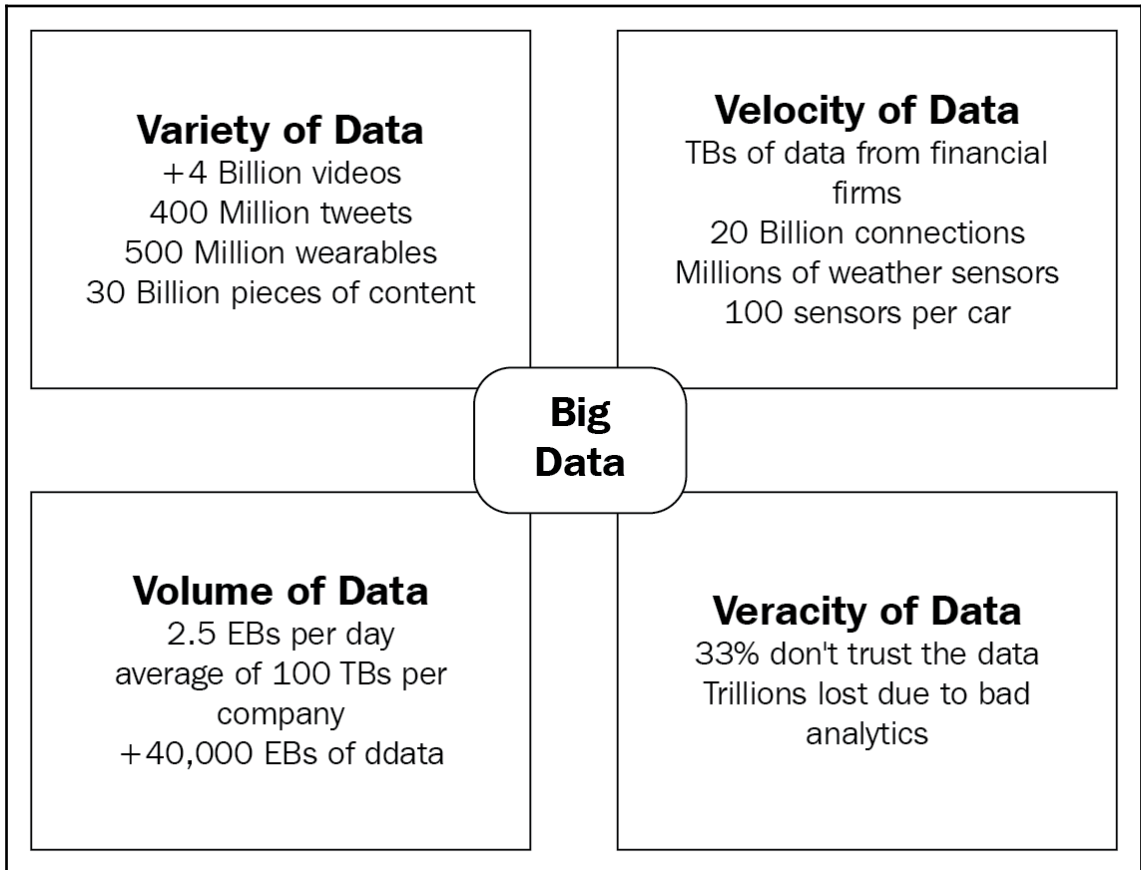
User Tables System Tables Snapshots

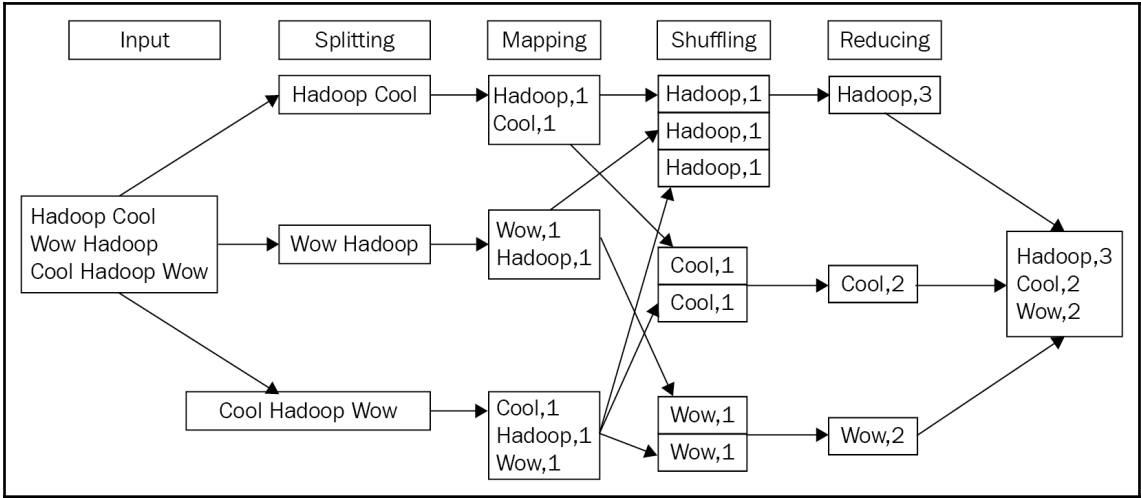
5 table(s) in set. [\[Details\]](#)

Namespace	Table Name	Online Regions	Offline Regions	Failed Regions	Split Regions	Other Regions	Description
default	<a href="#">prod.timelineservice.app_flow</a>	30	0	0	0	0	'prod.timelineservice.app_flow', {TABLE_ATTRIBUTES => {METADATA => {'KeyPrefixRegionSplitPolicy.prefix_length' => '4', 'SPLIT_POLICY' => 'org.apache.hadoop.hbase.regionserver.KeyPrefixRegionSplitPolicy'}, {NAME => 'm', BLOOMFILTER => 'ROWCOL'}}
default	<a href="#">prod.timelineservice.application</a>	30	0	0	0	0	'prod.timelineservice.application', {TABLE_ATTRIBUTES => {METADATA => {'KeyPrefixRegionSplitPolicy.prefix_length' => '4', 'SPLIT_POLICY' => 'org.apache.hadoop.hbase.regionserver.KeyPrefixRegionSplitPolicy'}, {NAME => 'c', BLOOMFILTER => 'ROWCOL'}, {NAME => 'l', BLOOMFILTER => 'ROWCOL'}, {NAME => 'm', VERSIONS => '10000', TTL => '2592000 SECONDS (30 DAYS)', MIN_VERSIONS => '1'}}
default	<a href="#">prod.timelineservice.entity</a>	30	0	0	0	0	'prod.timelineservice.entity', {TABLE_ATTRIBUTES => {METADATA => {'KeyPrefixRegionSplitPolicy.prefix_length' => '4', 'SPLIT_POLICY' => 'org.apache.hadoop.hbase.regionserver.KeyPrefixRegionSplitPolicy'}, {NAME => 'c', BLOOMFILTER => 'ROWCOL'}, {NAME => 'l', BLOOMFILTER => 'ROWCOL'}, {NAME => 'm', VERSIONS => '10000', TTL => '2592000 SECONDS (30 DAYS)', MIN_VERSIONS => '1'}}
default	<a href="#">prod.timelineservice.flowactivity</a>	1	0	0	0	0	'prod.timelineservice.flowactivity', {NAME => 'l', BLOOMFILTER => 'ROWCOL', VERSIONS => '2147483647', MIN_VERSIONS => '1'}
default	<a href="#">prod.timelineservice.subapplication</a>	30	0	0	0	0	'prod.timelineservice.subapplication', {TABLE_ATTRIBUTES => {METADATA => {'KeyPrefixRegionSplitPolicy.prefix_length' => '4', 'SPLIT_POLICY' => 'org.apache.hadoop.hbase.regionserver.KeyPrefixRegionSplitPolicy'}, {NAME => 'c', BLOOMFILTER => 'ROWCOL'}, {NAME => 'l', BLOOMFILTER => 'ROWCOL'}, {NAME => 'm', VERSIONS => '10000', TTL => '2592000 SECONDS (30 DAYS)', MIN_VERSIONS => '1'}}

---

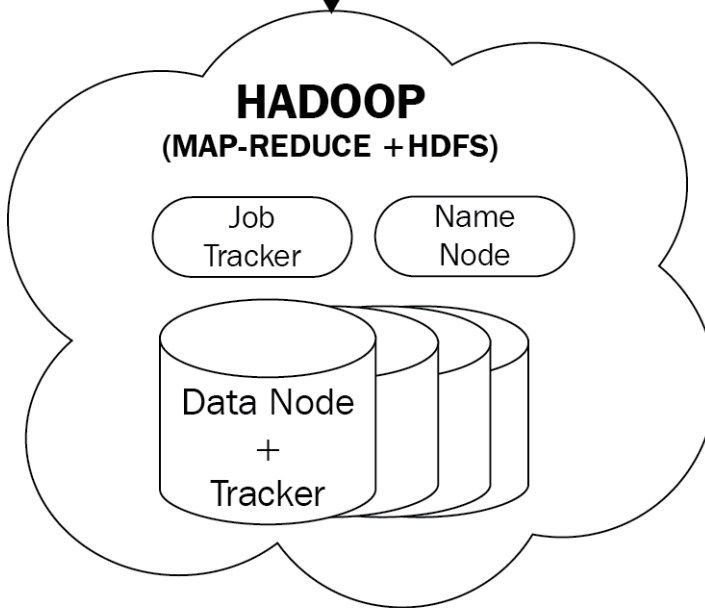
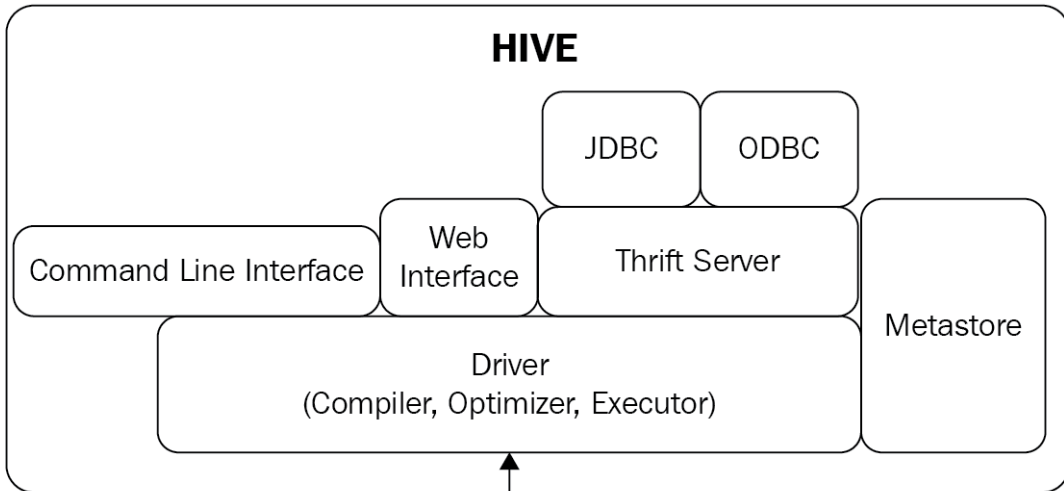
## Chapter 2: Overview of Big Data Analytics






---

# Hive Architecture



Secure | https://hive.apache.org/downloads.html



**GENERAL**  
[Home](#)  
[Downloads](#)  
[License](#)  
[Privacy Policy](#)

**DOCUMENTATION**  
[Language Manual](#)  
[Javadoc](#)  
[Wiki](#)

**COMMUNITY**  
[Becoming a Committer](#)  
[Edit Website](#)  
[How to Contribute](#)  
[Resources for contributors](#)  
[Issue Tracking](#)  
[Mailing Lists](#)  
[People](#)

**DEVELOPMENT**  
[Builds](#)  
[Design Docs](#)  
[FAQ](#)  
[Hive JIRA](#)  
[Precommit Patch Testing](#)  
[Version Control](#)

**PMC**  
[Builds](#)

## DOWNLOADS

Releases may be downloaded from Apache mirrors:  
[Download a release now!](#)

On the mirror, all recent releases are available, but are not guaranteed to be stable. For stable releases, look in the stable directory.

### News

3 April 2018 : release 2.3.3 available  
 This release works with Hadoop 2.x.y You can look at the complete [JIRA change log for this release](#).




18 November 2017 : release 2.3.2 available  
 This release works with Hadoop 2.x.y You can look at the complete [JIRA change log for this release](#).

24 October 2017 : release 2.3.1 available  
 This release works with Hadoop 2.x.y You can look at the complete [JIRA change log for this release](#).

25 July 2017 : release 2.2.0 available  
 This release works with Hadoop 2.x.y You can look at the complete [JIRA change log for this release](#).

mirror.cc.columbia.edu/pub/software/apache/hive/hive-2.3.3/


# Index of /pub/software/apache/hive/hive-2.3.3

<a href="#">Name</a>	<a href="#">Last modified</a>	<a href="#">Size</a>	<a href="#">Description</a>
 <a href="#">Parent Directory</a>		-	
 <a href="#">apache-hive-2.3.3-bin.tar.gz</a>	02-Apr-2018 15:00	221M	
 <a href="#">apache-hive-2.3.3-src.tar.gz</a>	02-Apr-2018 15:00	21M	

*Apache/2.2.22 (Fedora) Server at mirror.cc.columbia.edu Port 80*

← → ↻ https://db.apache.org/derby/derby\_downloads.html

apache > db >

**Apache Derby**  **The Apache DB Project** <http://db.apache.org>

Home Quick Start **Download** Community Documentation Resources

Last Published: 05/03/2018 14:28

Download

Overview

The Apache Software Foundation

Search the site with Google

Search

### Apache Derby: Downloads

- For Java 8 and Higher
  - For Java 6 and Higher
  - For Java 1.4 and Higher
  - For Java 1.3 and Higher
  - Deprecated Releases
  - Change History

**For Java 8 and Higher**

- [10.14.2.0](#) (May 3, 2018 / SVN 1828579)
- [10.13.1.1](#) (October 25, 2016 / SVN 1766613)

**For Java 6 and Higher**

- [10.12.1.1](#) (October 11, 2015 / SVN 1704137)
- [10.11.1.1](#) (August 26, 2014 / SVN 1616546)

**For Java 1.4 and Higher**

- [10.10.2.0](#) (April 15, 2014 / SVN 1582446)
- [10.10.1.1](#) (April 15, 2013 / SVN 1458268)
- [10.9.1.0](#) (June 25, 2012 / SVN 1344872)
- [10.8.3.0](#) (November 16, 2012 / SVN 1405108)
- [10.8.2.2](#) (October 24, 2011 / SVN 1181258)
- [10.8.1.2](#) (April 29, 2011 / SVN 1095077)
- [10.7.1.1](#) (December 14, 2010 / SVN 1040133)
- [10.6.2.1](#) (Oct 5, 2010 / SVN 999685)
- [10.6.1.0](#) (May 19, 2010 / SVN 938214)
- [10.5.3.0](#) (August 21, 2009 / SVN 802917)

```
Moogje:apache-hive-2.3.3-bin sridharalla$ hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/Users/sridharalla/apache-hive-2.3.3-bin/lib/log4j-slf4j-impl-2.6.2.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/Users/sridharalla/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/Users/sridharalla/hadoop-3.1.0/share/hadoop/common/lib/slf4j-log4j12-1.7.25.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.apache.logging.slf4j.Log4jLoggerFactory]

Logging initialized using configuration in jar:file:/Users/sridharalla/apache-hive-2.3.3-bin/lib/hive-common-2.3.3.jar!/hive-log4j2.properties Async: true
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
hive> |
```

```
hive> create database mydb;
OK
Time taken: 4.007 seconds
```

```
hive> use mydb;
OK
Time taken: 0.028 seconds
```

```
hive> create external table OnlineRetail (
> InvoiceNo string,
> StockCode string,
> Description string,
> Quantity integer,
> InvoiceDate string,
> UnitPrice float,
> CustomerID string,
> Country string
> ) ROW FORMAT DELIMITED
> FIELDS TERMINATED BY ','
> LOCATION '/user/normal';
OK
Time taken: 0.434 seconds
```

```
hive> select count(*) from OnlineRetail;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future
(i.e. spark, tez) or using Hive 1.X releases.
Query ID = sridharalla_20180423173731_d68999d5-5618-4170-a3a8-42be21851d51
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Job running in-process (local Hadoop)
2018-04-23 17:37:35,267 Stage-1 map = 100%,  reduce = 100%
Ended Job = job_local961179496_0001
MapReduce Jobs Launched:
Stage-Stage-1:  HDFS Read: 10714480 HDFS Write: 0 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
65500
Time taken: 3.518 seconds, Fetched: 1 row(s)
```



```

WHITE HANGING HEART T-LIGHT HOLDER      358
REGENCY CAKESTAND 3 TIER                 278
HEART OF WICKER SMALL                    224
HAND WARMER BABUSHKA DESIGN              213
SCOTTIE DOG HOT WATER BOTTLE            207
Time taken: 3.206 seconds, Fetched: 5 row(s)

```

```

hive> select * from OnlineRetail limit 5;
OK
InvoiceNo      StockCode      Description      NULL      InvoiceDate      NULL      CustomerID      Country
536365 85123A WHITE HANGING HEART T-LIGHT HOLDER      6      12/1/10 8:26      2.55      17850      United Kingdom
536365 71053 WHITE METAL LANTERN      6      12/1/10 8:26      3.39      17850      United Kingdom
536365 84406B CREAM CUPID HEARTS COAT HANGER      8      12/1/10 8:26      2.75      17850      United Kingdom
536365 84029G KNITTED UNION FLAG HOT WATER BOTTLE      6      12/1/10 8:26      3.39      17850      United Kingdom
Time taken: 5.25 seconds, Fetched: 5 row(s)

```

```

hive> select lower(description), quantity from OnlineRetail limit 5;
OK
description      NULL
white hanging heart t-light holder      6
white metal lantern      6
cream cupid hearts coat hanger      8
knitted union flag hot water bottle      6
Time taken: 0.154 seconds, Fetched: 5 row(s)

```

```

hive> select * from OnlineRetail where Description='WHITE METAL LANTERN' limit 5;
OK
536365 71053 WHITE METAL LANTERN      6      12/1/10 8:26      3.39      17850      United Kingdom
536373 71053 WHITE METAL LANTERN      6      12/1/10 9:02      3.39      17850      United Kingdom
536375 71053 WHITE METAL LANTERN      6      12/1/10 9:32      3.39      17850      United Kingdom
536396 71053 WHITE METAL LANTERN      6      12/1/10 10:51      3.39      17850      United Kingdom
536406 71053 WHITE METAL LANTERN      8      12/1/10 11:33      3.39      17850      United Kingdom
Time taken: 0.144 seconds, Fetched: 5 row(s)

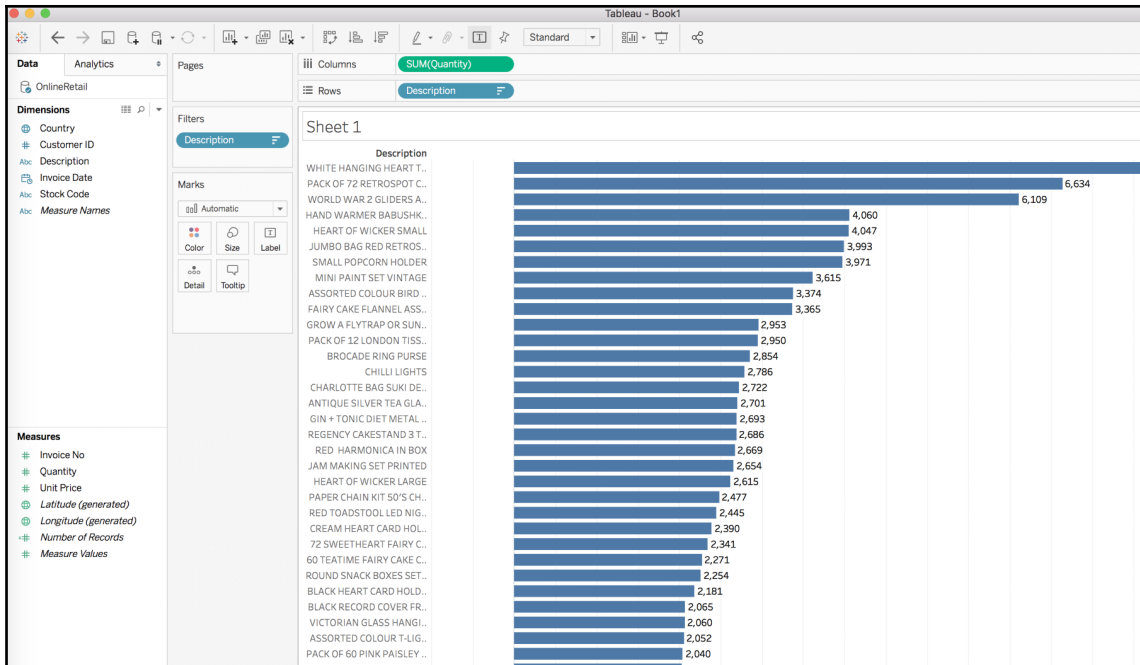
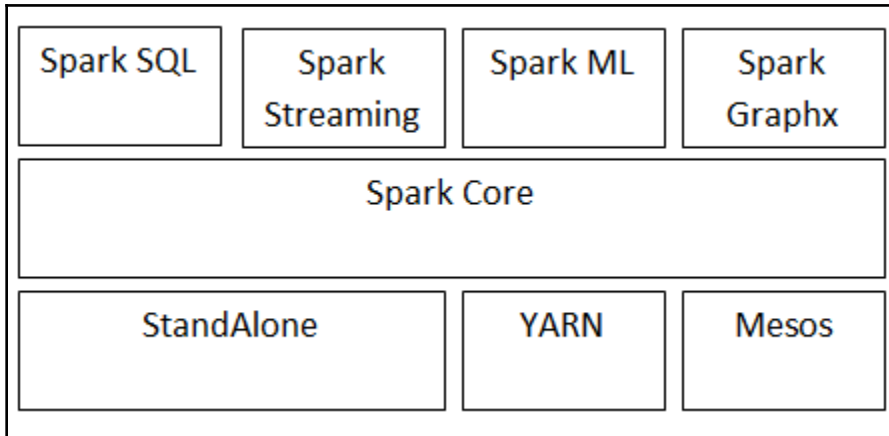
```

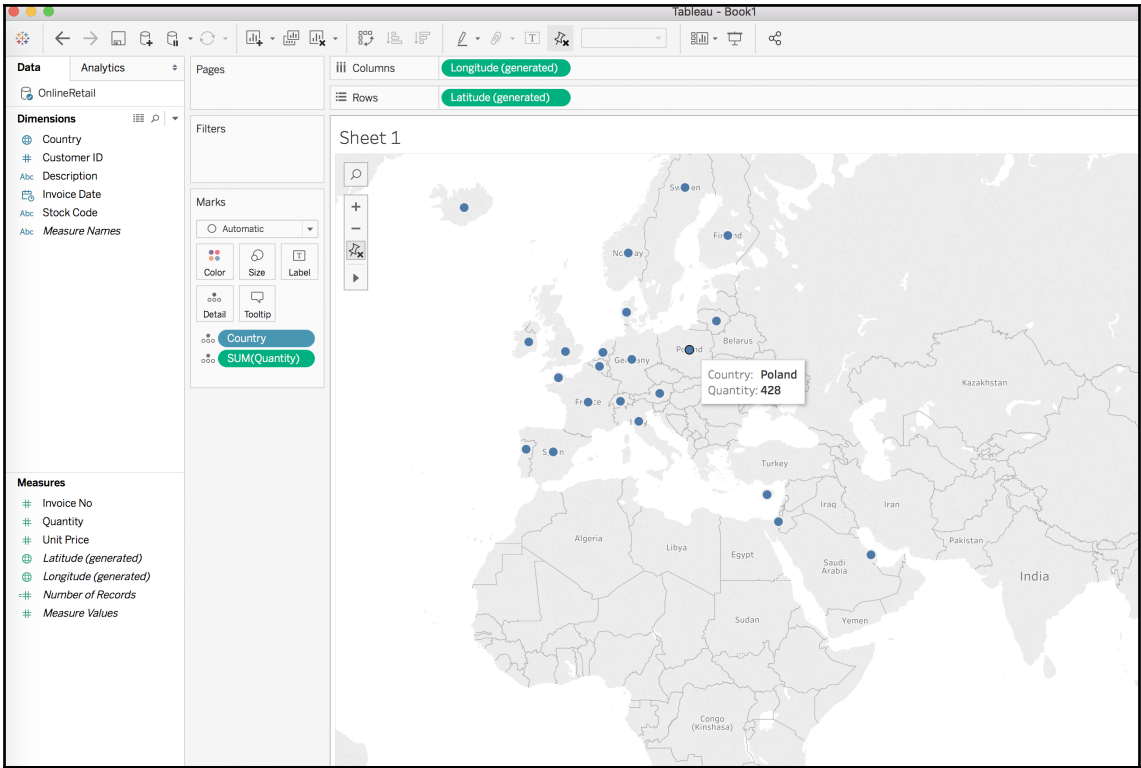
---

```
166
4 PURPLE FLOCK DINNER CANDLES 4
OVAL WALL MIRROR DIAMANTE 22
SET 2 TEA TOWELS I LOVE LONDON 102
"ACRYLIC HANGING JEWEL 1
Time taken: 1.6 seconds, Fetched: 5 row(s)
```

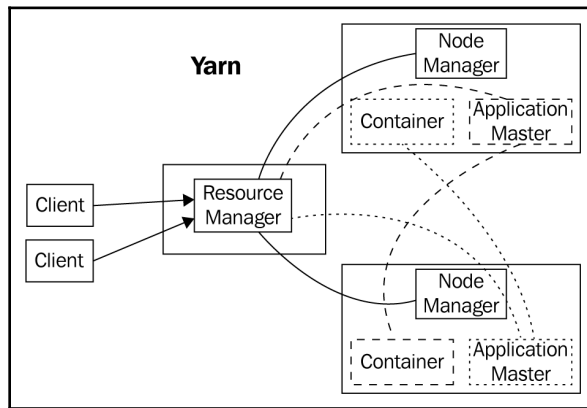
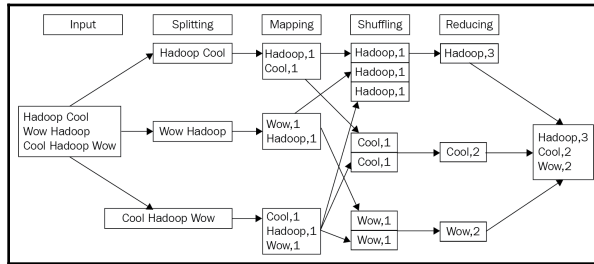
```
SET 2 TEA TOWELS I LOVE LONDON 102
"KEY FOB 110
6 RIBBONS RUSTIC CHARM 121
60 TEATIME FAIRY CAKE CASES 108
Time taken: 1.551 seconds, Fetched: 5 row(s)
```

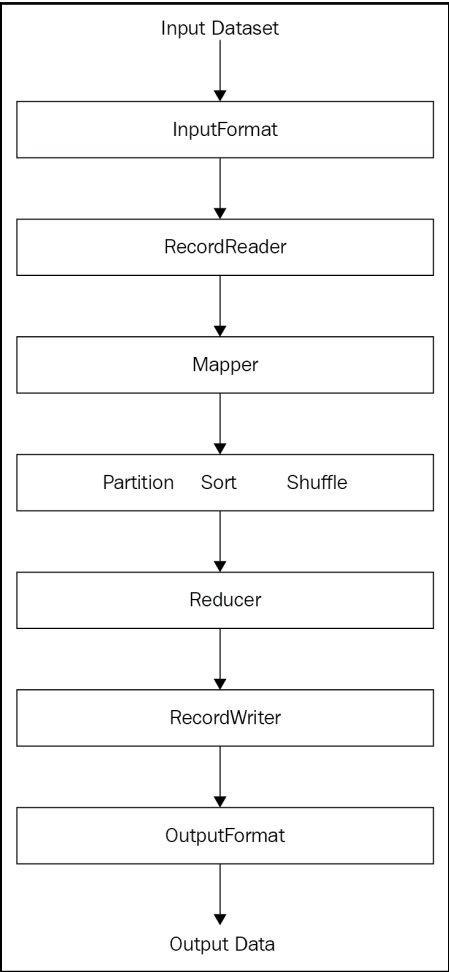
```
WHITE HANGING HEART T-LIGHT HOLDER 358
REGENCY CAKESTAND 3 TIER 278
HEART OF WICKER SMALL 224
HAND WARMER BABUSHKA DESIGN 213
SCOTTIE DOG HOT WATER BOTTLE 207
Time taken: 3.045 seconds, Fetched: 5 row(s)
```

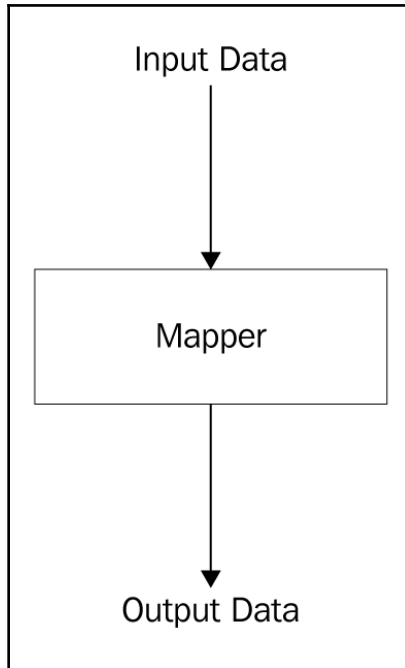




# Chapter 3: Big Data Processing with MapReduce










← → 127.0.0.1:9870/explorer.html#/user/normal

**Hadoop** Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

## Browse Directory

/user/normal Go!   

Show 25 entries Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	<input type="checkbox"/>
<input type="checkbox"/>	-rwxrwxrwx	<a href="#">sridharalla</a>	<a href="#">supergroup</a>	5.11 MB	May 20 17:49	1	128 MB	<a href="#">OnlineRetail.csv</a>	<input type="checkbox"/>
<input type="checkbox"/>	drwxrwxrwx	<a href="#">sridharalla</a>	<a href="#">supergroup</a>	0 B	May 20 18:05	0	0 B	<a href="#">avgtemperature</a>	<input type="checkbox"/>
<input type="checkbox"/>	-rwxrwxrwx	<a href="#">sridharalla</a>	<a href="#">supergroup</a>	81 B	May 20 17:49	1	128 MB	<a href="#">cities.csv</a>	<input type="checkbox"/>
<input type="checkbox"/>	drwxr-xr-x	<a href="#">sridharalla</a>	<a href="#">supergroup</a>	0 B	May 20 18:42	0	0 B	<a href="#">output</a>	<input type="checkbox"/>
<input type="checkbox"/>	-rwxrwxrwx	<a href="#">sridharalla</a>	<a href="#">supergroup</a>	452 B	May 20 17:49	1	128 MB	<a href="#">temperatures.csv</a>	<input type="checkbox"/>




Showing 1 to 5 of 5 entries Previous 1 Next

Hadoop, 2018.


← → C 127.0.0.1:9870/explorer.html#/user/normal/output

**Hadoop** Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities -

## Browse Directory

/user/normal/output Go!   

Show 25 entries Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	drwxr-xr-x	sridharalla	supergroup	0 B	May 20 18:42	0	0 B	SingleMapper	




Showing 1 to 1 of 1 entries Previous **1** Next

Hadoop, 2018.



← → C 127.0.0.1:9870/explorer.html#/user/normal/output/SingleMapper

**Hadoop** Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities -

## Browse Directory

/user/normal/output/SingleMapper Go!   

Show 25 entries Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	-rw-r--r--	sridharalla	supergroup	0 B	May 20 18:42	1	128 MB	_SUCCESS	
<input type="checkbox"/>	-rw-r--r--	sridharalla	supergroup	135 B	May 20 18:42	1	128 MB	part-r-00000	

Showing 1 to 2 of 2 entries Previous **1** Next

Hadoop, 2018.



← → C 127.0.0.1:9870/explorer.html#/user/normal/output/SingleMapper

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

## Browse Directory

/user/normal/output/SingleMapper

Show 25 entries Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	-rw-r--r--	<a href="#">sridharalla</a>	<a href="#">supergroup</a>	0 B	May 20 18:42	1	128 MB	<a href="#">_SUCCESS</a>	<input type="button" value="🗑️"/>
<input type="checkbox"/>	-rw-r--r--	<a href="#">sridharalla</a>	<a href="#">supergroup</a>	135 B	May 20 18:42	1	128 MB	<a href="#">part-r-00000</a>	<input type="button" value="🗑️"/>

Showing 1 to 2 of 2 entries

Hadoop, 2018.

← → C 127.0.0.1:9870/explorer.html#/user/normal/output/SingleMapper

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

## Browse Directory

/user/normal/output/SingleMapper

Show 25 entries Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	-rw-r--r--	<a href="#">sridharalla</a>	<a href="#">supergroup</a>	0 B	May 20 18:42	1	128 MB	<a href="#">_SUCCESS</a>	<input type="button" value="🗑️"/>
<input type="checkbox"/>	-rw-r--r--	<a href="#">sridharalla</a>	<a href="#">supergroup</a>	135 B	May 20 18:42	1	128 MB	<a href="#">part-r-00000</a>	<input type="button" value="🗑️"/>

Showing 1 to 2 of 2 entries

Hadoop, 2018.

**File information - part-r-00000**

[Download](#) [Head the file \(first 32K\)](#) [Tail the file \(last 32K\)](#)

**Block information -- Block 0**

Block ID: 1073741840

Block Pool ID: BP-146058116-10.0.0.103-1525565320513

Generation Stamp: 1016

Size: 135

Availability:

- 10.0.0.103

← → C 127.0.0.1:9870/explorer.html#/user/normal/output/SingleMapper

Hadoop Overview Datanodes Datanode Volumes Failures Snapshot Startup Progress Utilities

# Browse Director

/user/normal/output/SingleMapper

Show 25 entries

Permission	Owner
-rw-r--r--	sridharan
-rw-r--r--	sridharan

Showing 1 to 2 of 2 entries

Hadoop, 2018.

File information - part-r-00000

[Download](#) [Head the file \(first 32K\)](#) [Tail the file \(last 32K\)](#)

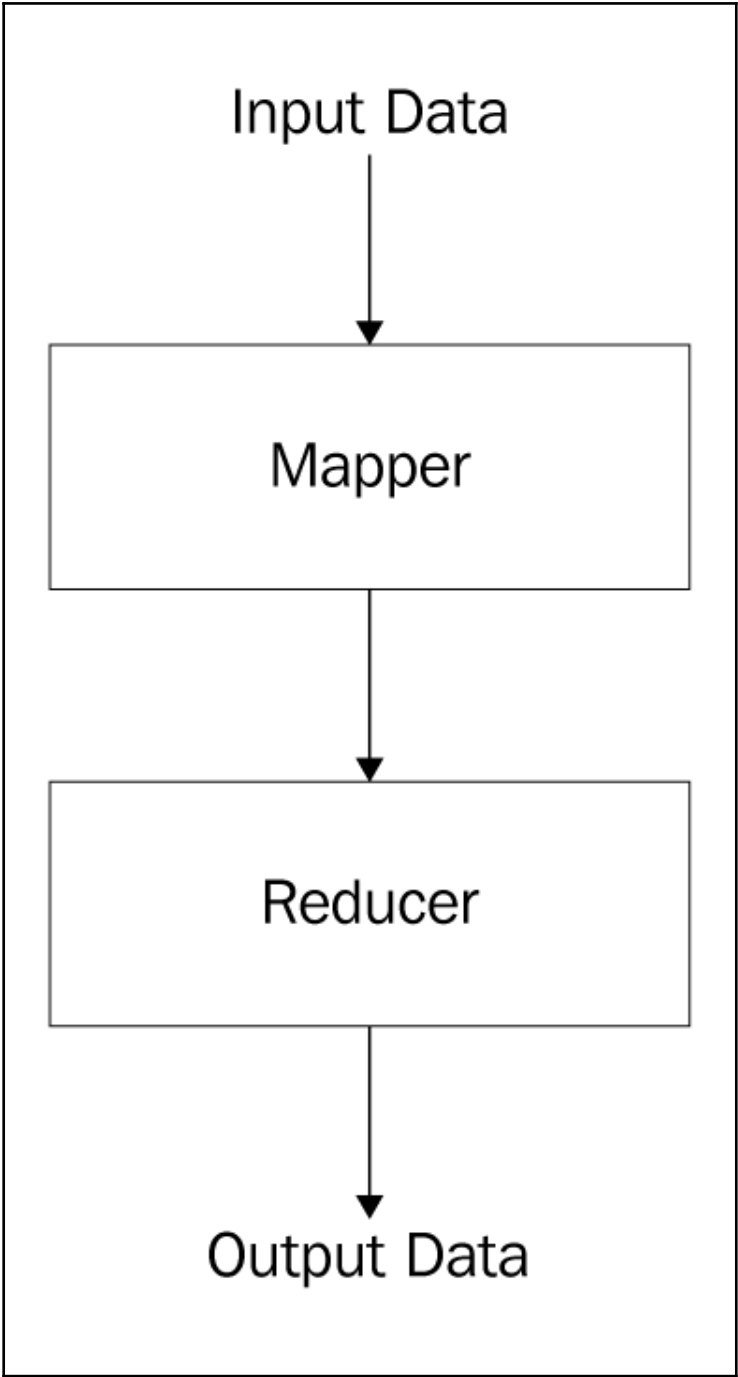
Block information -- Block 0 ▾

Block ID: 1073741840  
Block Pool ID: BP-146058116-10.0.0.103-1525565320513  
Generation Stamp: 1016  
Size: 135  
Availability:  
• 10.0.0.103

File contents

1	25
1	21
1	23
1	19
1	23
2	20
2	22
2	27




Close





← → 127.0.0.1:9870/explorer.html#/user/normal/output

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

## Browse Directory

/user/normal/output Go!   

Show 25 entries Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	drwxr-xr-x	sridharalla	supergroup	0 B	May 20 18:42	0	0 B	SingleMapper	
<input type="checkbox"/>	drwxr-xr-x	sridharalla	supergroup	0 B	May 20 18:43	0	0 B	SingleMapperReducer	




Showing 1 to 2 of 2 entries Previous **1** Next

Hadoop, 2018.



← → 127.0.0.1:9870/explorer.html#/user/normal/output/SingleMapperReducer

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

## Browse Directory

/user/normal/output/SingleMapperReducer Go!   

Show 25 entries Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	-rw-r--r--	sridharalla	supergroup	30 B	May 20 18:43	1	130 MB	_SUCCESS	
<input type="checkbox"/>	-rw-r--r--	sridharalla	supergroup	30 B	May 20 18:43	1	130 MB	part-r-00000	

Showing 1 to 2 of 2 entries Previous **1** Next

Hadoop, 2018.

**File information - part-r-00000** ✕

[Download](#) [Head the file \(first 32K\)](#) [Tail the file \(last 32K\)](#)

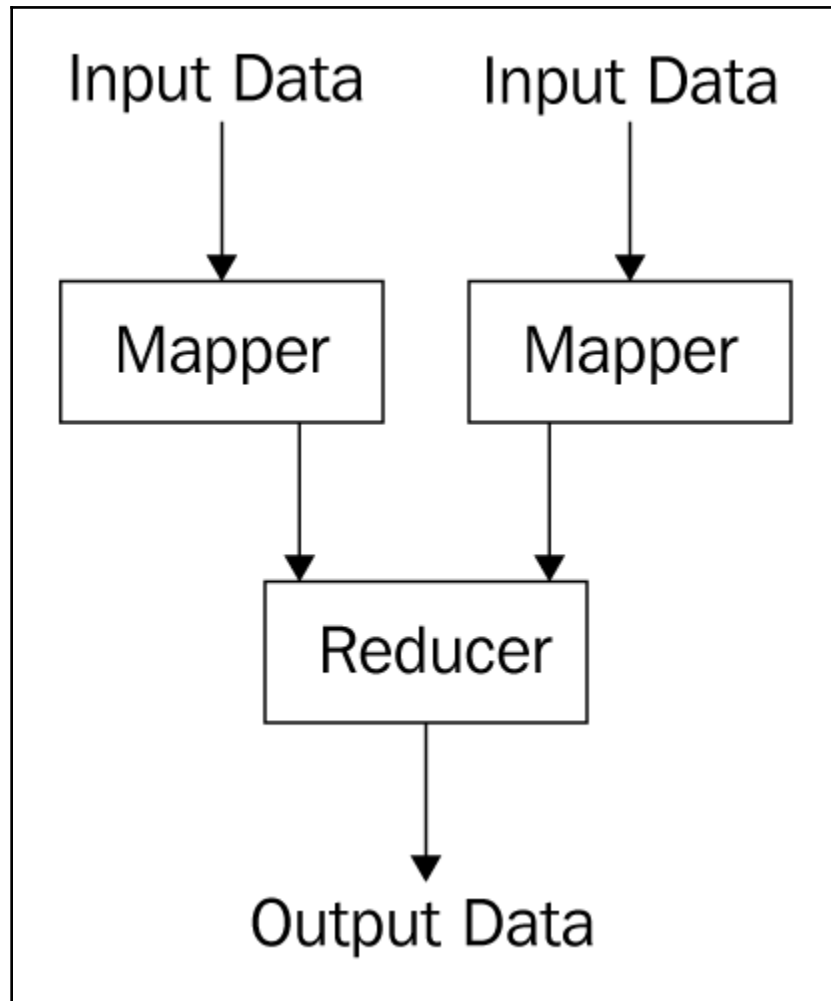
**Block information -- Block 0 ▾**

Block ID: 1073741841  
 Block Pool ID: BP-146058116-10.0.0.103-1525565320513  
 Generation Stamp: 1017  
 Size: 30  
 Availability:  
 • 10.0.0.103

**File contents**

```
1 22
2 23
3 23
4 23
5 22
6 22
```

[Close](#)



127.0.0.1:9870/explorer.html#/user/normal/output

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

## Browse Directory

/user/normal/output

Show 25 entries Search:

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	sridharalla	supergroup	0 B	May 20 18:45	0	0 B	MultipleMappersReducer
drwxr-xr-x	sridharalla	supergroup	0 B	May 20 18:42	0	0 B	SingleMapper
drwxr-xr-x	sridharalla	supergroup	0 B	May 20 18:44	0	0 B	SingleMapperCombinerReducer
drwxr-xr-x	sridharalla	supergroup	0 B	May 20 18:43	0	0 B	SingleMapperReducer

Showing 1 to 4 of 4 entries

Previous 1 Next

Hadoop, 2018.

127.0.0.1:9870/explorer.html#/user/normal/output/MultipleMappersReducer

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

## Browse Directory

/user/normal/output/MultipleMappersReducer

Show 25 entries Search:

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	sridharalla	supergroup	32K		1	128M	_SUCCESS
-rw-r--r--	sridharalla	supergroup	32K		1	128M	part-r-00000

Showing 1 to 2 of 2 entries

Previous 1 Next

Hadoop, 2018.

**File information - part-r-00000**

Download [Head the file \(first 32K\)](#) [Tail the file \(last 32K\)](#)

Block information -- Block 0

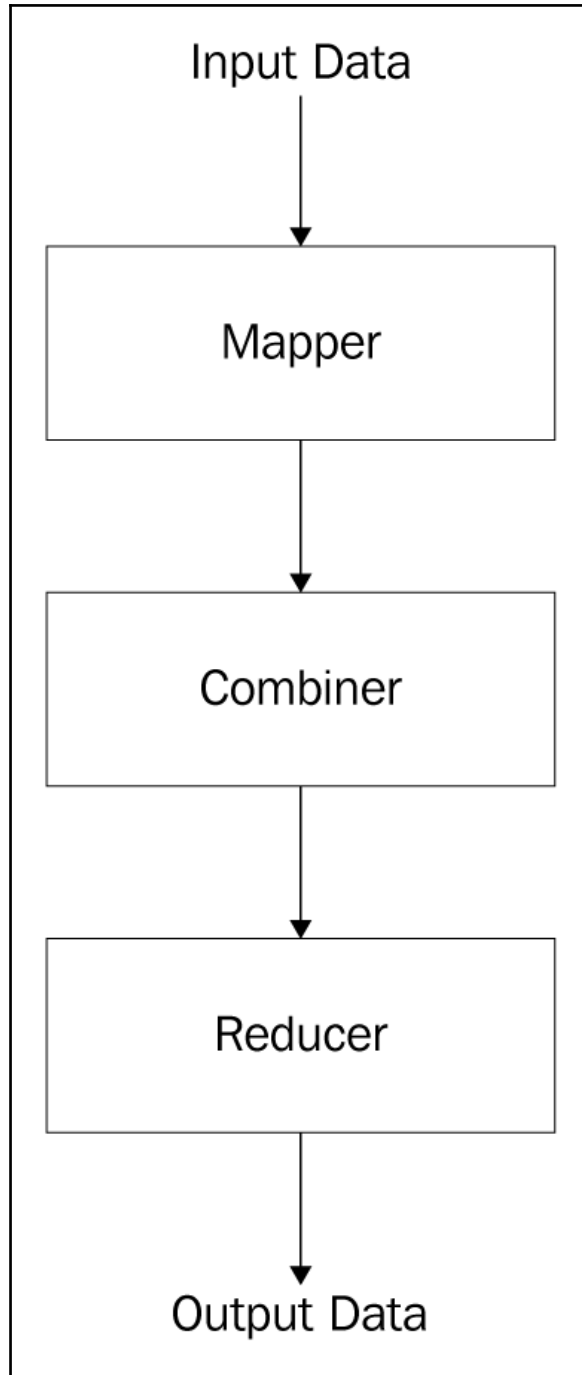
Block ID: 1073741844  
 Block Pool ID: BP-146058116-10.0.0.103-1525565320513  
 Generation Stamp: 1020  
 Size: 88  
 Availability:  
 • 10.0.0.103

File contents

```

Boston 22
New York 23
Chicago 23
Philadelphia 23
San Francisco 22
city-6 22
Las Vegas 0
      
```

Close



← → ↻ 127.0.0.1:9870/explorer.html#/user/normal/output

**Hadoop** Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

## Browse Directory

/user/normal/output

Show  entries Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	<input type="checkbox"/>
<input type="checkbox"/>	drwxr-xr-x	sridharalla	supergroup	0 B	May 20 18:42	0	0 B	SingleMapper	<input type="checkbox"/>
<input type="checkbox"/>	drwxr-xr-x	sridharalla	supergroup	0 B	May 20 18:44	0	0 B	SingleMapperCombinerReducer	<input type="checkbox"/>
<input type="checkbox"/>	drwxr-xr-x	sridharalla	supergroup	0 B	May 20 18:43	0	0 B	SingleMapperReducer	<input type="checkbox"/>

Showing 1 to 3 of 3 entries

Hadoop, 2018.

← → ↻ 127.0.0.1:9870/explorer.html#/user/normal/output/SingleMapperCombinerReducer

**Hadoop** Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

## Browse Directory

/user/normal/output/SingleMapperCombinerReducer

Show  entries Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	<input type="checkbox"/>
<input type="checkbox"/>	-rw-r--r--	sridharalla	supergroup	30 B	May 20 18:44	1	131072 B	part-r-00000	<input type="checkbox"/>
<input type="checkbox"/>	-rw-r--r--	sridharalla	supergroup	30 B	May 20 18:44	1	131072 B	part-r-00001	<input type="checkbox"/>

Showing 1 to 2 of 2 entries

Hadoop, 2018.

**File information - part-r-00000** ✕

[Download](#) [Head the file \(first 32K\)](#) [Tail the file \(last 32K\)](#)

**Block information -- Block 0 ▾**

Block ID: 1073741842  
 Block Pool ID: BP-146058116-10.0.0.103-1525565320513  
 Generation Stamp: 1018  
 Size: 30  
 Availability:  
 • 10.0.0.103

**File contents**

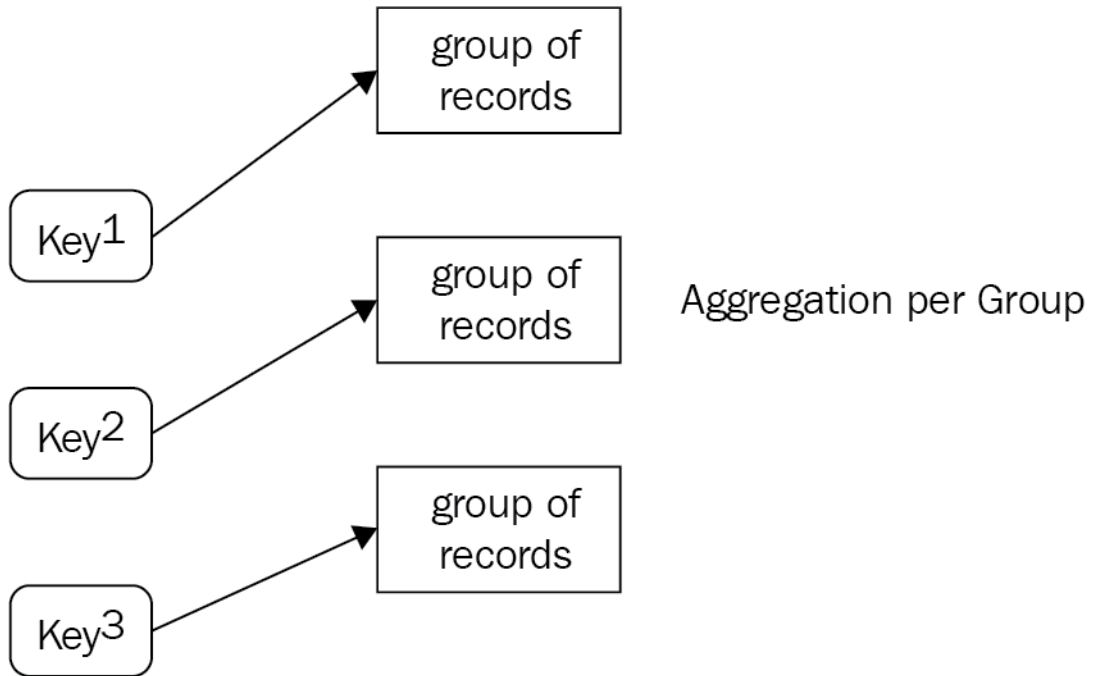
```

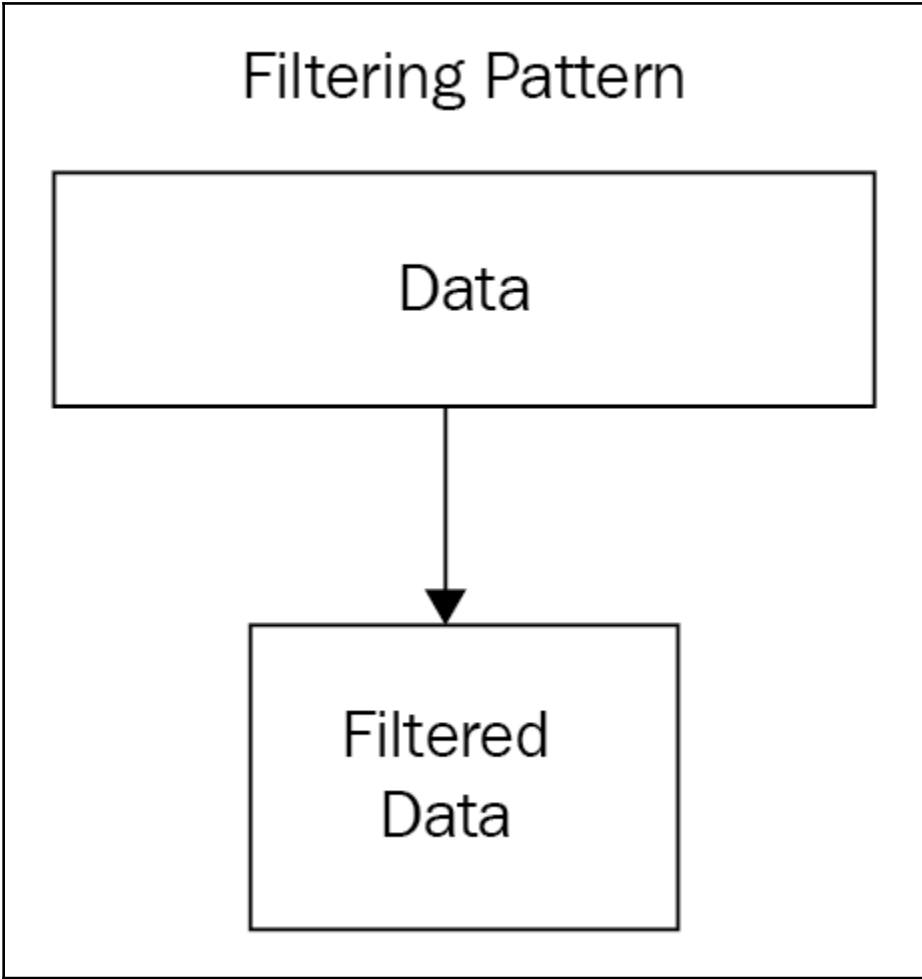
1 22
2 23
3 23
4 23
5 22
6 22
    
```

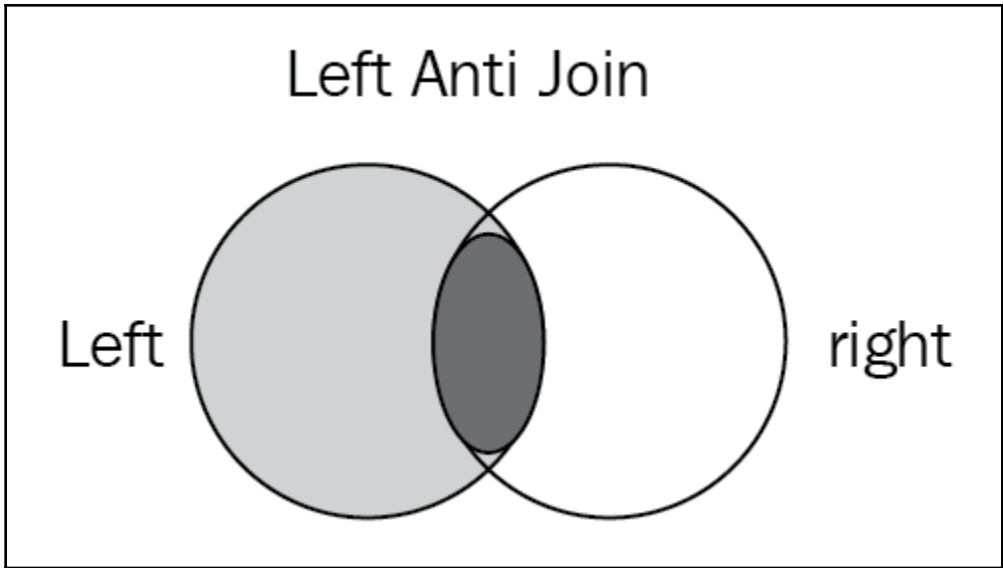
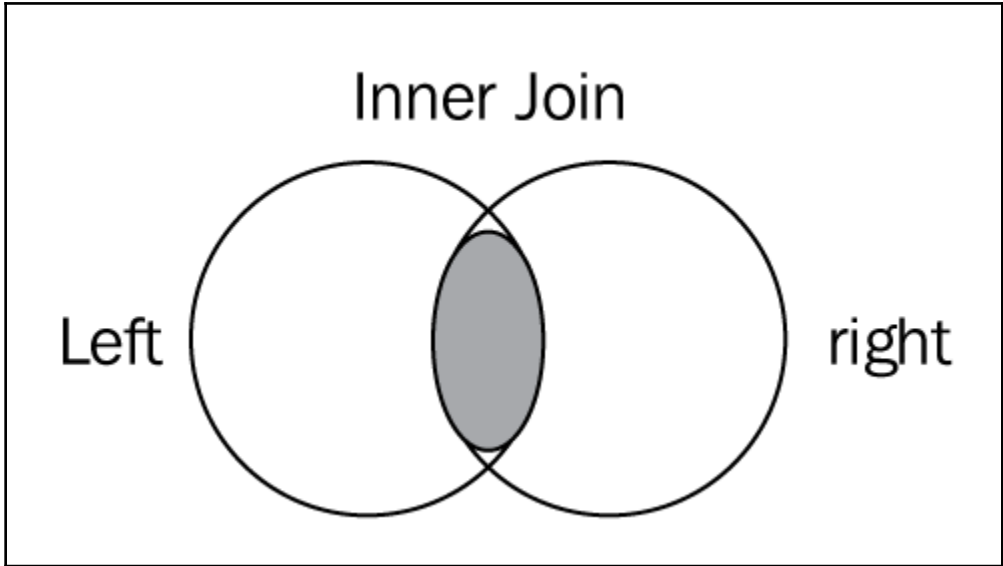


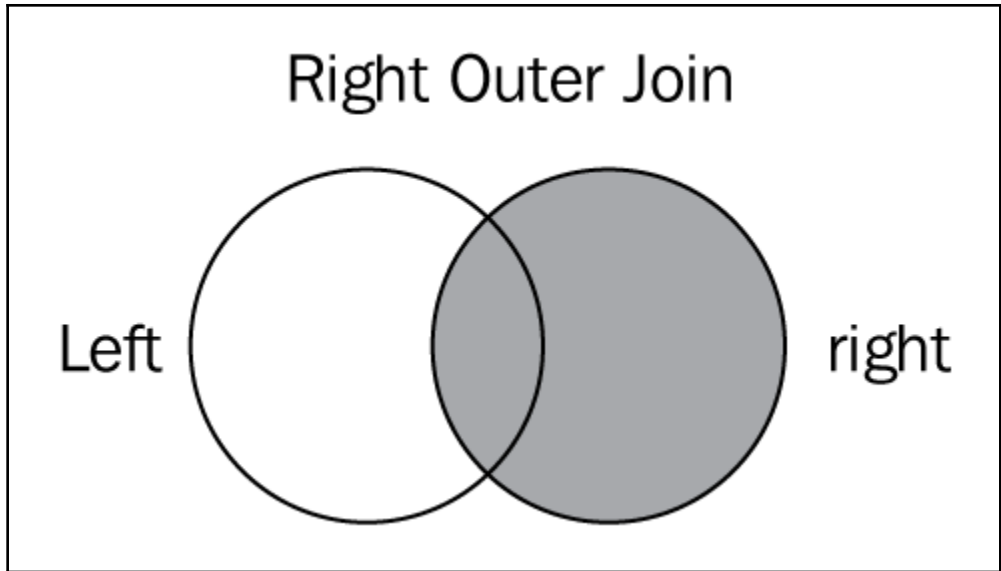
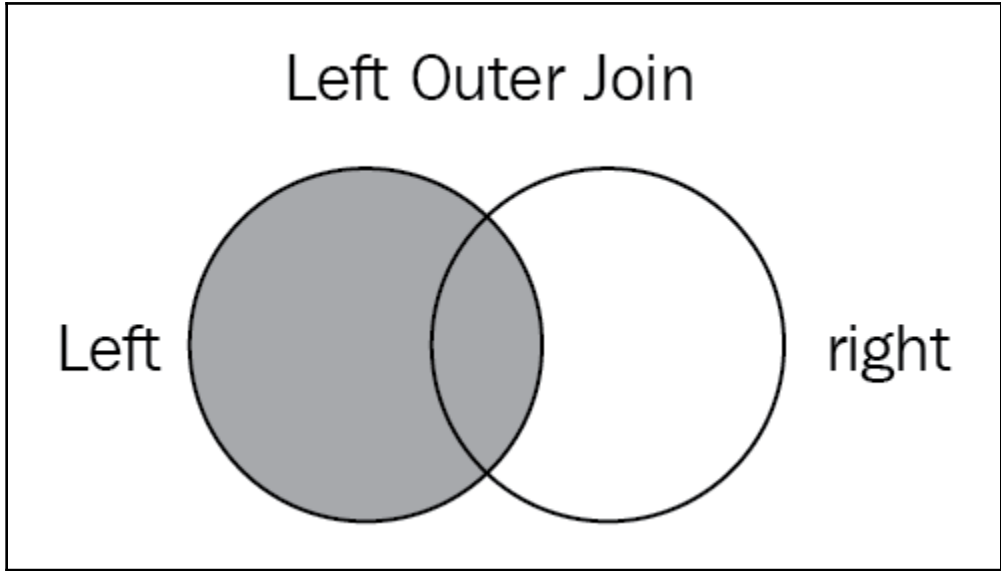
---

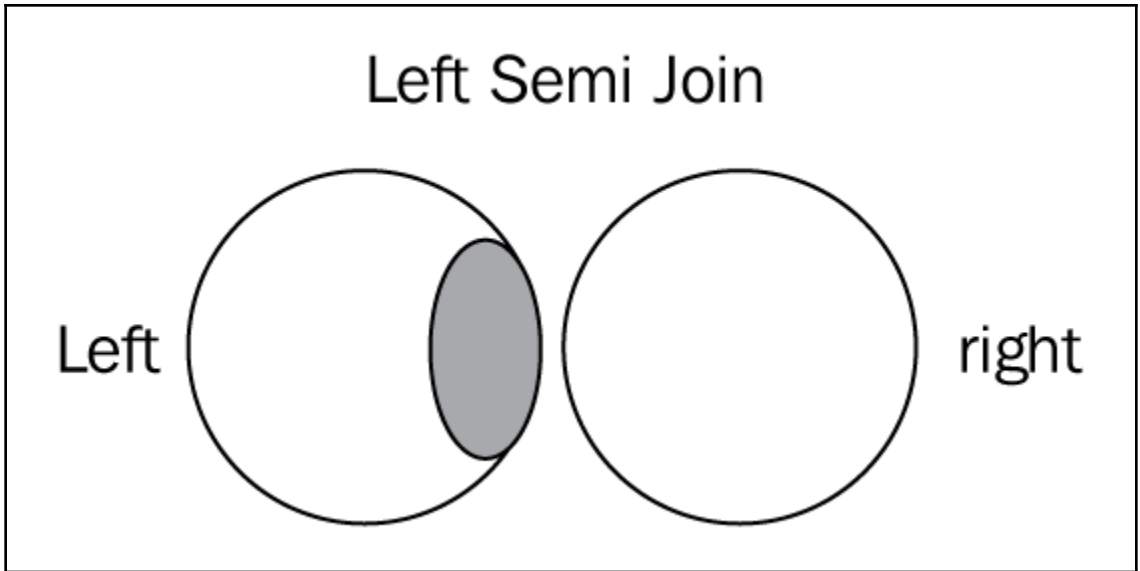
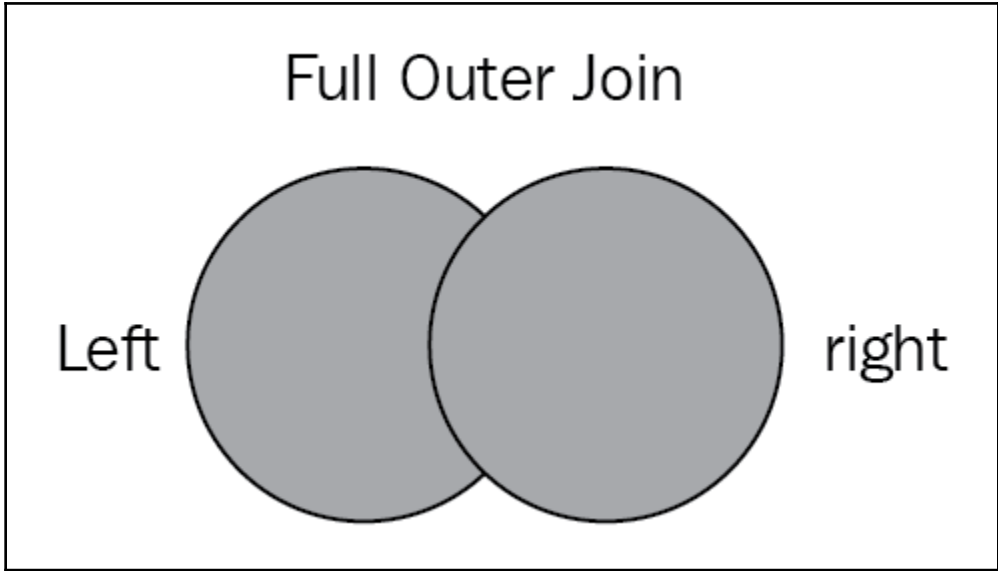
## Summarization Pattern

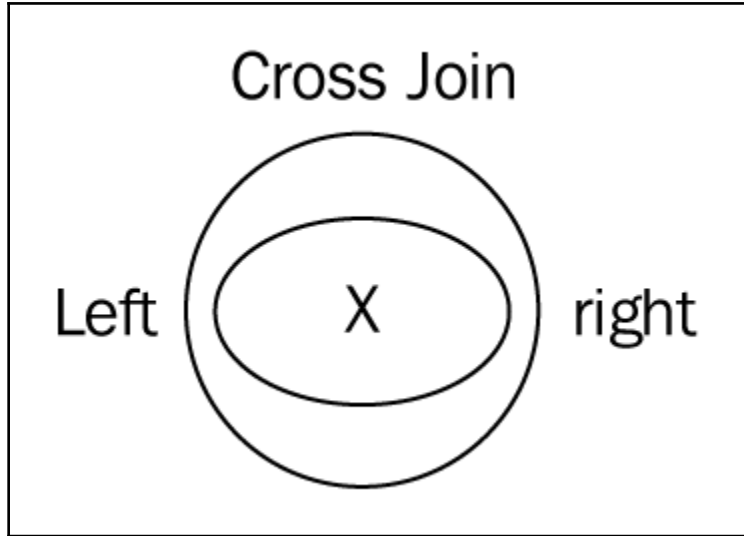




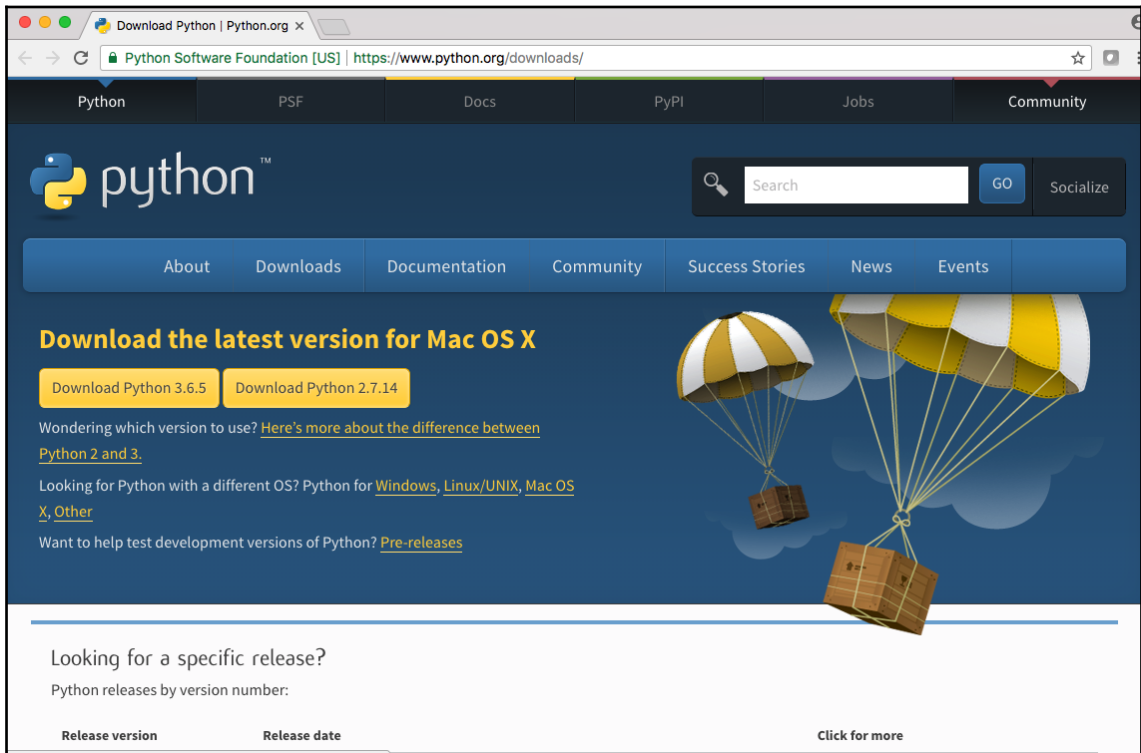








# Chapter 4: Scientific Computing and Big Data Analysis with Python and Hadoop



The screenshot shows the Python.org website's download page. The browser's address bar displays "Python Software Foundation [US] | https://www.python.org/downloads/". The navigation menu includes "Python", "PSF", "Docs", "PyPI", "Jobs", and "Community". The main content area features the Python logo, a search bar, and a "Socialize" button. Below this is a secondary navigation bar with links for "About", "Downloads", "Documentation", "Community", "Success Stories", "News", and "Events". The primary heading is "Download the latest version for Mac OS X". Two prominent buttons offer "Download Python 3.6.5" and "Download Python 2.7.14". Below these buttons, there are three paragraphs of text: the first discusses version selection with a link to "Here's more about the difference between Python 2 and 3"; the second provides links for "Windows, Linux/UNIX, Mac OS X, Other"; and the third links to "Pre-releases". An illustration of two parachutes carrying boxes is positioned to the right of the text. At the bottom, a section titled "Looking for a specific release?" includes the text "Python releases by version number:" and a table with headers "Release version", "Release date", and "Click for more".

Download Python 3.6.5    Download Python 2.7.14

Wondering which version to use? [Here's more about the difference between Python 2 and 3.](#)

Looking for Python with a different OS? Python for [Windows](#), [Linux/UNIX](#), [Mac OS X](#), [Other](#)

Want to help test development versions of Python? [Pre-releases](#)

Looking for a specific release?  
Python releases by version number:

Release version	Release date	Click for more
-----------------	--------------	----------------

The image shows a browser window displaying the Python Software Foundation website. The browser's address bar shows the URL <https://www.python.org/downloads/release/python-365/>. The website's navigation bar includes links for Python, PSF, Docs, PyPI, Jobs, and Community. Below this is the Python logo and a search bar with a 'GO' button and a 'Socialize' link. A secondary navigation bar contains links for About, Downloads, Documentation, Community, Success Stories, News, and Events. The main content area features the heading 'Python 3.6.5' and the release date '2018-03-28'. A paragraph states that Python 3.6.5 is the fifth maintenance release of Python 3.6. A list of new features is provided, each linked to a PEP document.

Python Release Python 3.6.5 | x

Python Software Foundation [US] | <https://www.python.org/downloads/release/python-365/>

Python PSF Docs PyPI Jobs Community

python™ Search GO Socialize

About Downloads Documentation Community Success Stories News Events

## Python 3.6.5

**Release Date:** 2018-03-28

Python 3.6.5 is the fifth maintenance release of Python 3.6. The Python 3.6 series contains many new features and optimizations.

Among the new major new features in Python 3.6 are:

- [PEP 468](#), Preserving Keyword Argument Order
- [PEP 487](#), Simpler customization of class creation
- [PEP 495](#), Local Time Disambiguation
- [PEP 498](#), Literal String Formatting
- [PEP 506](#), Adding A Secrets Module To The Standard Library
- [PEP 509](#), Add a private version to dict



Python Release Python 3.6.5 | x

Python Software Foundation [US] | <https://www.python.org/downloads/release/python-365/>

...changes in the 3.6 installer-supplied Python, particularly with regard to SSL certificate validation.

[Full Changelog](#)

## Files

Version	Operating System	Description	MD5 Sum	File Size	PGP
<a href="#">Gzipped source tarball</a>	Source release		ab25d24b1f8cc4990ade979f6dc37883	22994617	<a href="#">SIG</a>
<a href="#">XZ compressed source tarball</a>	Source release		9f49654a4d6f733ff3284ab9d227e9fd	17049912	<a href="#">SIG</a>
<a href="#">macOS 64-bit/32-bit installer</a>	Mac OS X	for Mac OS X 10.6 and later	bf319337bc68b52fc7d227dca5b6f2f6	28093627	<a href="#">SIG</a>
<a href="#">macOS 64-bit installer</a>	Mac OS X	for OS X 10.9 and later	37d891988b6aeedd7f03a70171a8420d	26987706	<a href="#">SIG</a>
<a href="#">Windows help file</a>	Windows		be70202d483c0b7291a666ec66539784	8065193	<a href="#">SIG</a>
<a href="#">Windows x86-64 embeddable zip file</a>	Windows	for AMD64/EM64T/x64	04cc4f6f6a14ba74f6ae1a8b685ec471	7190516	<a href="#">SIG</a>
<a href="#">Windows x86-64 executable installer</a>	Windows	for AMD64/EM64T/x64	9e96c934f5d16399f860812b4ac7002b	31776112	<a href="#">SIG</a>
<a href="#">Windows x86-64 web-based installer</a>	Windows	for AMD64/EM64T/x64	640736a3894022d30f7babff77391d6b	1320112	<a href="#">SIG</a>
<a href="#">Windows x86 embeddable zip file</a>	Windows		b0b099a4fa479fb37880c15f2b2f4f34	6429369	<a href="#">SIG</a>
<a href="#">Windows x86 executable installer</a>	Windows		2bb6ad2ecca6088171ef923bca483f02	30735232	<a href="#">SIG</a>
<a href="#">Windows x86 web-based installer</a>	Windows		596667cb91a9fb20e6f4f153f3a213a5	1294096	<a href="#">SIG</a>

Downloads | Anaconda

Secure | <https://www.anaconda.com/download/#macos>

Documentation Blog Contact Anaconda Cloud

What is Anaconda? Products Support Community Resources About [Download](#)

Don't Miss AnacondaCon Apr 8-11 Austin TX!

# Download Anaconda Distribution

Version 5.1 | Release Date: February 15, 2018

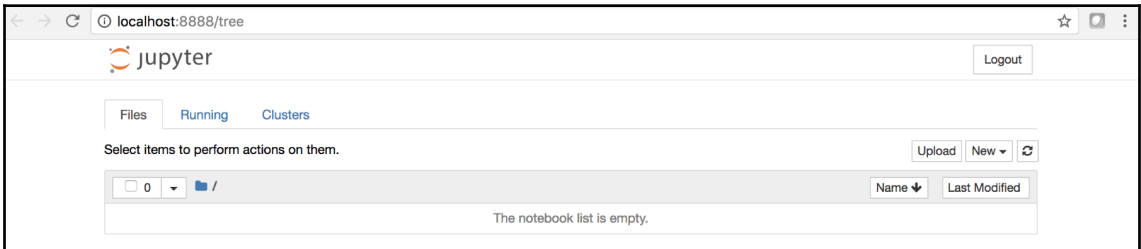
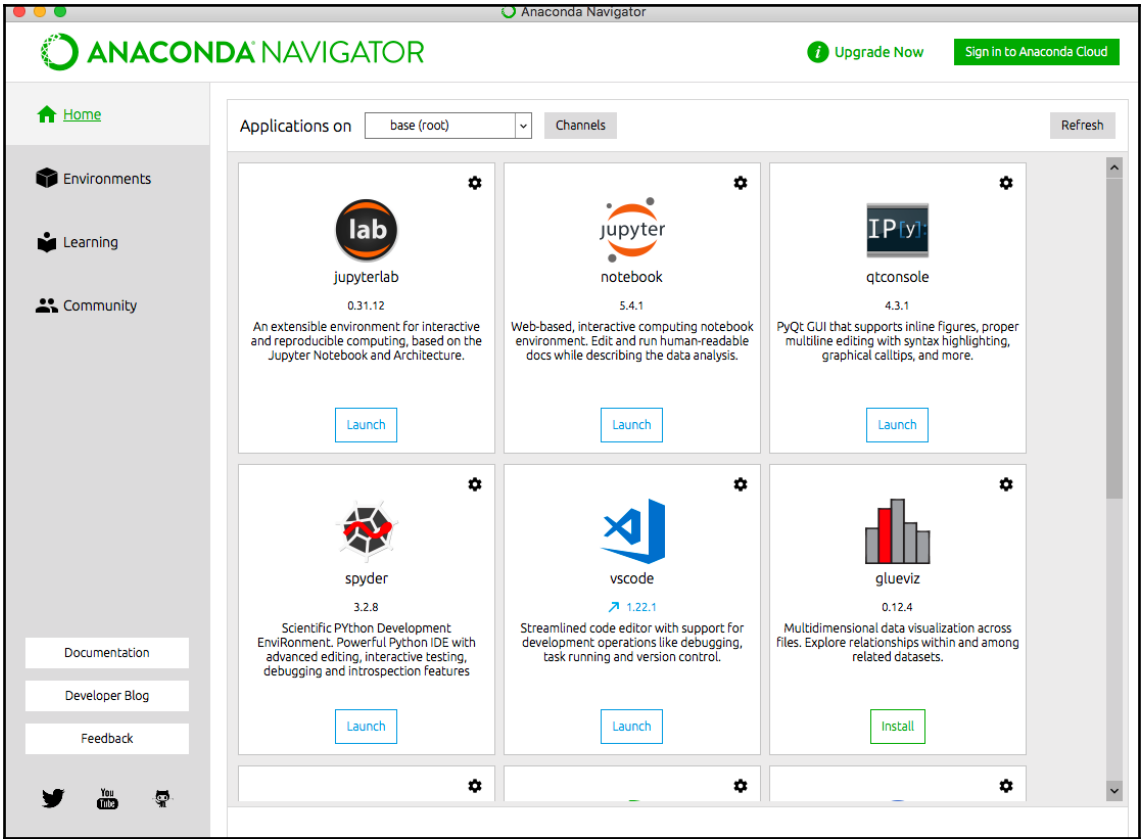
Download For:

**High-Performance Distribution**  
Easily install 1,000+ [data science packages](#)

**Package Management**  
Manage packages, dependencies and environments with [conda](#)

**Portal to Data Science**  
Uncover insights in your data and create interactive visualizations

Download for Your Preferred Platform



The screenshot shows a web browser window with the URL `https://conda.io/docs/user-guide/install/index.html`. The page title is "Conda" and the breadcrumb is "Docs » User guide » Installation". There is a search bar and a link to "Edit on GitHub". The left sidebar contains a navigation menu with categories like "User guide", "Installation", "Configuration", "Tutorials", "Tasks", "Cheat sheet", "Troubleshooting", "Help and support", "Release notes", and "Command reference". The main content area is titled "Installation" and includes a list of links: "System requirements", "Regular installation", "Installing in silent mode", and "Installing conda on a system that has other Python installations or packages". Below the links, there are two paragraphs of text. The first paragraph states that the fastest way to obtain conda is to install Miniconda, a mini version of Anaconda that includes only conda and its dependencies. The second paragraph recommends installing Anaconda for the local user, which does not require administrator permissions and is the most robust type of installation. Below the text, there is a section titled "System requirements" with a list of requirements: 32- or 64-bit computer, 400 MB disk space for Miniconda, 3 GB disk space for Anaconda, and Windows, macOS or Linux.

Secure | <https://conda.io/docs/user-guide/install/index.html>

Conda

Search docs

Docs » User guide » Installation [Edit on GitHub](#)

## Installation

- [System requirements](#)
- [Regular installation](#)
- [Installing in silent mode](#)
- [Installing conda on a system that has other Python installations or packages](#)

The fastest way to **obtain** conda is to install **Miniconda**, a mini version of **Anaconda** that includes only conda and its dependencies. If you prefer to have conda plus over 720 open source packages, install **Anaconda**.

We recommend you install Anaconda for the local user, which does not require administrator permissions and is the most robust type of installation. You can also install Anaconda system wide, which does require administrator permissions.

For information on using our graphical installers for Windows or macOS, see the instructions for [installing Anaconda](#).

### System requirements




- 32- or 64-bit computer.
- For Miniconda—400 MB disk space.
- For Anaconda—Minimum 3 GB disk space to download and install.
- Windows, macOS or Linux.

Secure | <https://conda.io/miniconda.html> ☆

Conda « <no title> Contents

# Miniconda

*Fork me on GitHub*

	 Windows	 Mac OS X	 Linux
<b>Python 3.6</b>	<a href="#">64-bit (exe installer)</a> <a href="#">32-bit (exe installer)</a>	<a href="#">64-bit (bash installer)</a>	<a href="#">64-bit (bash installer)</a> <a href="#">32-bit (bash installer)</a>
<b>Python 2.7</b>	<a href="#">64-bit (exe installer)</a> <a href="#">32-bit (exe installer)</a>	<a href="#">64-bit (bash installer)</a>	<a href="#">64-bit (bash installer)</a> <a href="#">32-bit (bash installer)</a>

Other resources:

- [Miniconda with Python 3.6 for Power8](#)
- [Miniconda with Python 2.7 for Power8](#)
- [Miniconda Docker images](#)
- [Installation instructions](#)
- [MD5 sums for the installers](#)
- [conda change log](#)

These Miniconda installers contain the conda package manager and Python. Once Miniconda is installed, you can use the conda command to install any other packages and create environments, etc. For example:

```
$ conda install numpy
```

```

sridhars-MacBook-Pro-2:~ sridharalla$ conda list
# packages in environment at /anaconda2:
#
_ipyw_jlab_nb_ext_conf    0.1.0                py27h172cb35_0
affine                    2.2.0                py_0               conda-forge
alabaster                 0.7.10              py27_1             conda-forge
anaconda                  custom              py27h2cfa9e9_0
anaconda-client          1.6.14              py_0               conda-forge
anaconda-navigator       1.8.1                py27_0
anaconda-project         0.8.2                py27_0             conda-forge
appnope                   0.1.0                py27_0             conda-forge
appscript                 1.0.1                py27_0             conda-forge
asn1crypto                0.24.0              py27_0             conda-forge
astroid                   1.6.2                py27_0             conda-forge
astropy                   2.0.5                py27_1             conda-forge
attrs                     17.4.0               py_0               conda-forge
babel                     2.5.3                py27_0             conda-forge
backports                 1.0                  py27_1             conda-forge
backports.functools_lru_cache 1.5                  py27_0             conda-forge
backports.shutil_get_terminal_size 1.0.0              py_3               conda-forge
backports_abc             0.5                  py27_0             conda-forge
beautifulsoup4           4.6.0                py27_0             conda-forge
bitarray                 0.8.1                py27_0             conda-forge
bkcharts                  0.2                  py27_0             conda-forge
blas                      1.1                  openblas           conda-forge
blaze                     0.11.3               py27_0             conda-forge
bleach                    2.1.3                py_0               conda-forge
blinker                   1.4                  py_0               conda-forge
hakeh                     0.12.14              nv27_0             conda-forge

```

```

sridhars-MacBook-Pro-2:~ sridharalla$ conda install scikit-learn
Fetching package metadata .....
Solving package specifications: .

Package plan for installation in environment /anaconda2:

The following packages will be UPDATED:

conda: 4.3.34-py27_0 conda-forge --> 4.5.1-py27_0 conda-forge

Proceed [y]/n)? y

conda-4.5.1-py 100% |#####| Time: 0:00:00 1.73 MB/s

```

---

- jupyter

The following packages will be downloaded:

package	build	
libxml2-2.9.8	hf84eae3_0	2.0 MB
backports_abc-0.5	py27h7b3c97b_0	12 KB
jsonschema-2.6.0	py27h7ed5aa4_0	61 KB
markupsafe-1.0	py27h97b2822_1	24 KB
notebook-5.4.1	py27_0	6.6 MB
libpng-1.6.34	hb9fc6fc_0	334 KB
bleach-2.1.3	py27_0	32 KB
ca-certificates-2018.03.07	0	124 KB
decorator-4.3.0	py27_0	15 KB
xz-5.2.3	h55aa19d_2	357 KB
gstreamer-1.14.0	hb453b48_1	3.8 MB
pyzmq-17.0.0	py27h14c3975_1	440 KB
jupyter_console-5.2.0	py27hc6bee7e_1	34 KB
icu-58.2	h9c2bf20_1	22.5 MB
ipython-5.6.0	py27_0	1020 KB
pandocfilters-1.4.2	py27h428e1e5_1	12 KB
tornado-5.0.1	py27_1	617 KB
prompt_toolkit-1.0.15	py27h1b593e1_0	333 KB
sip-4.19.8	py27hf484d3e_0	291 KB
ptyprocess-0.5.2	py27h4ccb14c_0	22 KB
widgetsnbextension-3.2.0	py27_0	1.7 MB
openssl-1.0.2o	h20670df_0	3.4 MB
simplegeneric-0.8.1	py27_2	9 KB
freetype-2.8	hab7d2ae_1	804 KB
pickleshare-0.7.4	py27h09770e1_0	11 KB
pygments-2.2.0	py27h4a8b6f5_0	1.3 MB
libsodium-1.0.16	h1bed415_0	302 KB
gst-plugins-base-1.14.0	hb8d80ab_1	6.3 MB
zeromq-4.2.5	h439df22_0	567 KB

```
[root@4b726275a804 /]# conda install pandas
Solving environment: done
```

```
## Package Plan ##
```

```
environment location: /root/miniconda2
```

```
added / updated specs:
```

```
- pandas
```

```
The following packages will be downloaded:
```

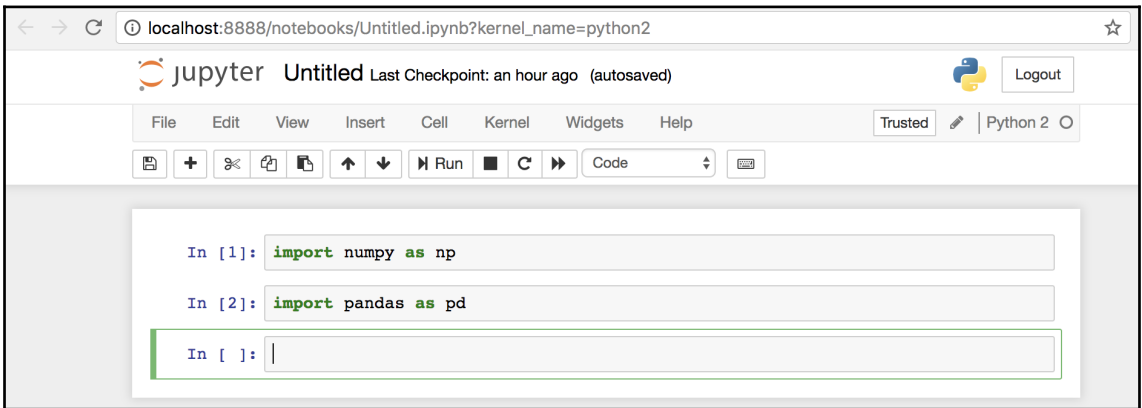
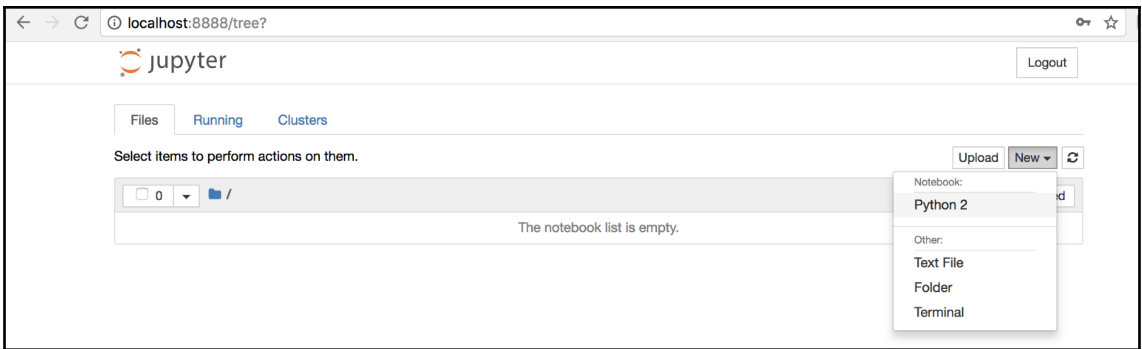
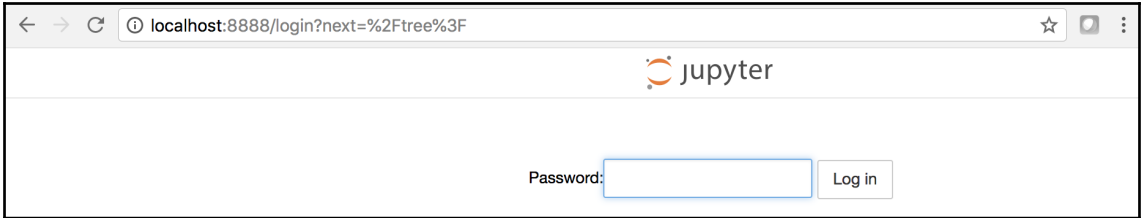
package	build	
-----	-----	
intel-openmp-2018.0.0	8	620 KB
mkl_fft-1.0.1	py27h3010b51_0	137 KB
pytz-2018.4	py27_0	211 KB
mkl-2018.0.2	1	205.2 MB
pandas-0.22.0	py27hf484d3e_0	10.5 MB
mkl_random-1.0.1	py27h629b387_0	361 KB
libgfortran-ng-7.2.0	hdf63c60_3	1.2 MB
numpy-1.14.2	py27hdbf6ddf_1	4.1 MB
-----	-----	
	Total:	222.3 MB

```
The following NEW packages will be INSTALLED:
```

```
intel-openmp: 2018.0.0-8
libgfortran-ng: 7.2.0-hdf63c60_3
mkl: 2018.0.2-1
mkl_fft: 1.0.1-py27h3010b51_0
mkl_random: 1.0.1-py27h629b387_0
numpy: 1.14.2-py27hdbf6ddf_1
pandas: 0.22.0-py27hf484d3e_0
pytz: 2018.4-py27_0
```



```
[root@4b726275a804 /]# jupyter notebook --allow-root --no-browser --ip=* --port=8888
[W 02:43:01.935 NotebookApp] WARNING: The notebook server is listening on all IP addresses and not using encryption. This is not recommended.
[I 02:43:01.955 NotebookApp] Serving notebooks from local directory: /
[I 02:43:01.955 NotebookApp] 0 active kernels
[I 02:43:01.955 NotebookApp] The Jupyter Notebook is running at:
[I 02:43:01.955 NotebookApp] http://[all ip addresses on your system]:8888/
[I 02:43:01.955 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
```



```
In [11]: df.head(3)
```

```
Out[11]:
```

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	12/1/10 8:26	2.55	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	12/1/10 8:26	3.39	17850.0	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/1/10 8:26	2.75	17850.0	United Kingdom

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
1	536365	71053	WHITE METAL LANTERN	6	12/1/10 8:26	3.39	17850.0	United Kingdom
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	12/1/10 8:26	3.39	17850.0	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	12/1/10 8:26	3.39	17850.0	United Kingdom

```
In [45]: df.loc[df.UnitPrice > 3.0, ['Description']] = 'Miscellaneous'
```

```
df.head(3)
```

```
Out[45]:
```

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	1	12/1/10 8:26	2.55	17850.0	United Kingdom
1	536365	71053	Miscellaneous	1	12/1/10 8:26	3.39	17850.0	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	1	12/1/10 8:26	2.75	17850.0	United Kingdom

```
In [55]: df.loc[2]
```

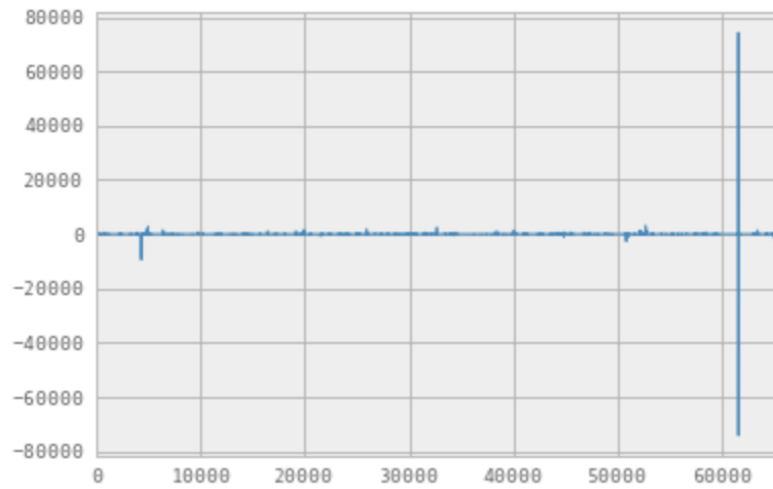
```
Out[55]:
```

InvoiceNo	536365
StockCode	84406B
Description	CREAM CUPID HEARTS COAT HANGER
Quantity	8
InvoiceDate	12/1/10 8:26
UnitPrice	2.75
CustomerID	17850
Country	United Kingdom
Name:	2, dtype: object

---

```
In [64]: df['Quantity'].plot()
```

```
Out[64]: <matplotlib.axes._subplots.AxesSubplot at 0x11a3c4b50>
```



```
In [35]: df3 = df[0:10].append(df[200:210])
df3
```

Out[35]:

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	12/1/10 8:26	2.55	17850.0	United Kingdom
1	536365	71053	Miscellaneous	6	12/1/10 8:26	3.39	17850.0	United States
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/1/10 8:26	2.75	17850.0	United Kingdom
3	536365	84029G	Miscellaneous	6	12/1/10 8:26	3.39	17850.0	United States
4	536365	84029E	Miscellaneous	6	12/1/10 8:26	3.39	17850.0	United States
5	536365	22752	Miscellaneous	2	12/1/10 8:26	7.65	17850.0	United States
6	536365	21730	Miscellaneous	6	12/1/10 8:26	4.25	17850.0	United States
7	536366	22633	HAND WARMER UNION JACK	6	12/1/10 8:28	1.85	17850.0	United Kingdom
8	536366	22632	HAND WARMER RED POLKA DOT	6	12/1/10 8:28	1.85	17850.0	United Kingdom
9	536367	84879	ASSORTED COLOUR BIRD ORNAMENT	32	12/1/10 8:34	1.69	13047.0	United Kingdom
200	536389	35004C	Miscellaneous	6	12/1/10 10:03	5.45	12431.0	United States
201	536389	35004G	Miscellaneous	4	12/1/10 10:03	6.35	12431.0	United States
202	536389	85014B	Miscellaneous	6	12/1/10 10:03	5.95	12431.0	United States
203	536389	85014A	Miscellaneous	3	12/1/10 10:03	5.95	12431.0	United States
204	536389	22193	Miscellaneous	2	12/1/10 10:03	8.50	12431.0	United States
205	536389	22726	Miscellaneous	4	12/1/10 10:03	3.75	12431.0	United States
206	536389	22727	Miscellaneous	4	12/1/10 10:03	3.75	12431.0	United States
207	536389	22192	Miscellaneous	2	12/1/10 10:03	8.50	12431.0	United States
208	536389	22191	Miscellaneous	2	12/1/10 10:03	8.50	12431.0	United States
209	536389	22195	LARGE HEART MEASURING SPOONS	24	12/1/10 10:03	1.65	12431.0	Australia

```
In [40]: df4 = pd.DataFrame(df, columns=['StockCode', 'Quantity', 'InvoiceDate', 'UnitPrice'])
df4.head(3)
```

Out[40]:

	StockCode	Quantity	InvoiceDate	UnitPrice
0	85123A	6	12/1/10 8:26	2.55
1	71053	6	12/1/10 8:26	3.39
2	84406B	8	12/1/10 8:26	2.75

```
In [166]: d1 = df[0:10]
          d2 = df[10:20]

          d3 = pd.concat([d1, d2])
          d3
```

Out[166]:

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	12/1/10 8:26	2.55	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	12/1/10 8:26	3.39	17850.0	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/1/10 8:26	2.75	17850.0	United Kingdom
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	12/1/10 8:26	3.39	17850.0	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	12/1/10 8:26	3.39	17850.0	United Kingdom
5	536365	22752	SET 7 BABUSHKA NESTING BOXES	2	12/1/10 8:26	7.65	17850.0	United Kingdom
6	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	6	12/1/10 8:26	4.25	17850.0	United Kingdom
7	536366	22633	HAND WARMER UNION JACK	6	12/1/10 8:28	1.85	17850.0	United Kingdom
8	536366	22632	HAND WARMER RED POLKA DOT	6	12/1/10 8:28	1.85	17850.0	United Kingdom
9	536367	84879	ASSORTED COLOUR BIRD ORNAMENT	32	12/1/10 8:34	1.69	13047.0	United Kingdom
10	536367	22745	POPPY'S PLAYHOUSE BEDROOM	6	12/1/10 8:34	2.10	13047.0	United Kingdom
11	536367	22748	POPPY'S PLAYHOUSE KITCHEN	6	12/1/10 8:34	2.10	13047.0	United Kingdom
12	536367	22749	FELTCRAFT PRINCESS CHARLOTTE DOLL	8	12/1/10 8:34	3.75	13047.0	United Kingdom
13	536367	22310	IVORY KNITTED MUG COSY	6	12/1/10 8:34	1.65	13047.0	United Kingdom
14	536367	84969	BOX OF 6 ASSORTED COLOUR TEASPOONS	6	12/1/10 8:34	4.25	13047.0	United Kingdom
15	536367	22623	BOX OF VINTAGE JIGSAW BLOCKS	3	12/1/10 8:34	4.95	13047.0	United Kingdom
16	536367	22622	BOX OF VINTAGE ALPHABET BLOCKS	2	12/1/10 8:34	9.95	13047.0	United Kingdom
17	536367	21754	HOME BUILDING BLOCK WORD	3	12/1/10 8:34	5.95	13047.0	United Kingdom
18	536367	21755	LOVE BUILDING BLOCK WORD	3	12/1/10 8:34	5.95	13047.0	United Kingdom
19	536367	21777	RECIPE BOX WITH METAL HEART	4	12/1/10 8:34	7.95	13047.0	United Kingdom

```
In [179]: d1 = df[0:10]
d2 = df[10:20]

d3 = pd.concat([d1, d2], keys=['d1', 'd2'])
d3
```

Out[179]:

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
<b>d1</b>	0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	12/1/10 8:26	2.55	17850.0 United Kingdom
	1	536365	71053	WHITE METAL LANTERN	6	12/1/10 8:26	3.39	17850.0 United Kingdom
	2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/1/10 8:26	2.75	17850.0 United Kingdom
	3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	12/1/10 8:26	3.39	17850.0 United Kingdom
	4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	12/1/10 8:26	3.39	17850.0 United Kingdom
	5	536365	22752	SET 7 BABUSHKA NESTING BOXES	2	12/1/10 8:26	7.65	17850.0 United Kingdom
	6	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	6	12/1/10 8:26	4.25	17850.0 United Kingdom
	7	536366	22633	HAND WARMER UNION JACK	6	12/1/10 8:28	1.85	17850.0 United Kingdom
	8	536366	22632	HAND WARMER RED POLKA DOT	6	12/1/10 8:28	1.85	17850.0 United Kingdom
	9	536367	84879	ASSORTED COLOUR BIRD ORNAMENT	32	12/1/10 8:34	1.69	13047.0 United Kingdom
<b>d2</b>	10	536367	22745	POPPY'S PLAYHOUSE BEDROOM	6	12/1/10 8:34	2.10	13047.0 United Kingdom
	11	536367	22748	POPPY'S PLAYHOUSE KITCHEN	6	12/1/10 8:34	2.10	13047.0 United Kingdom
	12	536367	22749	FELTCRAFT PRINCESS CHARLOTTE DOLL	8	12/1/10 8:34	3.75	13047.0 United Kingdom
	13	536367	22310	IVORY KNITTED MUG COSY	6	12/1/10 8:34	1.65	13047.0 United Kingdom
	14	536367	84969	BOX OF 6 ASSORTED COLOUR TEASPOONS	6	12/1/10 8:34	4.25	13047.0 United Kingdom
	15	536367	22623	BOX OF VINTAGE JIGSAW BLOCKS	3	12/1/10 8:34	4.95	13047.0 United Kingdom
	16	536367	22622	BOX OF VINTAGE ALPHABET BLOCKS	2	12/1/10 8:34	9.95	13047.0 United Kingdom
	17	536367	21754	HOME BUILDING BLOCK WORD	3	12/1/10 8:34	5.95	13047.0 United Kingdom
	18	536367	21755	LOVE BUILDING BLOCK WORD	3	12/1/10 8:34	5.95	13047.0 United Kingdom
	19	536367	21777	RECIPE BOX WITH METAL HEART	4	12/1/10 8:34	7.95	13047.0 United Kingdom

```
In [183]: d4 = pd.DataFrame(df, columns=['InvoiceNo', 'StockCode', 'Description'])[0:10]
d5 = pd.DataFrame(df, columns=['StockCode', 'Description', 'Quantity'])[0:10]

pd.concat([d4, d5])
```

Out[183]:

	Description	InvoiceNo	Quantity	StockCode
0	WHITE HANGING HEART T-LIGHT HOLDER	536365	NaN	85123A
1	WHITE METAL LANTERN	536365	NaN	71053
2	CREAM CUPID HEARTS COAT HANGER	536365	NaN	84406B
3	KNITTED UNION FLAG HOT WATER BOTTLE	536365	NaN	84029G
4	RED WOOLLY HOTTIE WHITE HEART.	536365	NaN	84029E
5	SET 7 BABUSHKA NESTING BOXES	536365	NaN	22752
6	GLASS STAR FROSTED T-LIGHT HOLDER	536365	NaN	21730
7	HAND WARMER UNION JACK	536366	NaN	22633
8	HAND WARMER RED POLKA DOT	536366	NaN	22632
9	ASSORTED COLOUR BIRD ORNAMENT	536367	NaN	84879
0	WHITE HANGING HEART T-LIGHT HOLDER	NaN	6.0	85123A
1	WHITE METAL LANTERN	NaN	6.0	71053
2	CREAM CUPID HEARTS COAT HANGER	NaN	8.0	84406B
3	KNITTED UNION FLAG HOT WATER BOTTLE	NaN	6.0	84029G
4	RED WOOLLY HOTTIE WHITE HEART.	NaN	6.0	84029E
5	SET 7 BABUSHKA NESTING BOXES	NaN	2.0	22752
6	GLASS STAR FROSTED T-LIGHT HOLDER	NaN	6.0	21730
7	HAND WARMER UNION JACK	NaN	6.0	22633
8	HAND WARMER RED POLKA DOT	NaN	6.0	22632
9	ASSORTED COLOUR BIRD ORNAMENT	NaN	32.0	84879

```
In [184]: d4 = pd.DataFrame(df, columns=['InvoiceNo', 'StockCode', 'Description'])[0:10]
d5 = pd.DataFrame(df, columns=['StockCode', 'Description', 'Quantity'])[0:10]

pd.concat([d4, d5], join = 'inner')
```

Out[184]:

	StockCode	Description
0	85123A	WHITE HANGING HEART T-LIGHT HOLDER
1	71053	WHITE METAL LANTERN
2	84406B	CREAM CUPID HEARTS COAT HANGER
3	84029G	KNITTED UNION FLAG HOT WATER BOTTLE
4	84029E	RED WOOLLY HOTTIE WHITE HEART.
5	22752	SET 7 BABUSHKA NESTING BOXES
6	21730	GLASS STAR FROSTED T-LIGHT HOLDER
7	22633	HAND WARMER UNION JACK
8	22632	HAND WARMER RED POLKA DOT
9	84879	ASSORTED COLOUR BIRD ORNAMENT
0	85123A	WHITE HANGING HEART T-LIGHT HOLDER
1	71053	WHITE METAL LANTERN
2	84406B	CREAM CUPID HEARTS COAT HANGER
3	84029G	KNITTED UNION FLAG HOT WATER BOTTLE
4	84029E	RED WOOLLY HOTTIE WHITE HEART.
5	22752	SET 7 BABUSHKA NESTING BOXES
6	21730	GLASS STAR FROSTED T-LIGHT HOLDER
7	22633	HAND WARMER UNION JACK
8	22632	HAND WARMER RED POLKA DOT
9	84879	ASSORTED COLOUR BIRD ORNAMENT



```
In [222]: pd.merge(d4, d5, how='left')
```

Out[222]:

	InvoiceNo	StockCode	Description	Quantity
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	NaN
1	536365	71053	WHITE METAL LANTERN	NaN
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	NaN
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	NaN
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	NaN
5	536365	22752	SET 7 BABUSHKA NESTING BOXES	NaN
6	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	NaN
7	536366	22633	HAND WARMER UNION JACK	NaN
8	536366	22632	HAND WARMER RED POLKA DOT	NaN
9	536367	84879	ASSORTED COLOUR BIRD ORNAMENT	NaN
10	536367	22745	POPPY'S PLAYHOUSE BEDROOM	6.0

```
In [223]: pd.merge(d4, d5, how='right')
```

Out[223]:

	InvoiceNo	StockCode	Description	Quantity
0	536367	22745	POPPY'S PLAYHOUSE BEDROOM	6
1	NaN	22748	POPPY'S PLAYHOUSE KITCHEN	6
2	NaN	22749	FELTCRAFT PRINCESS CHARLOTTE DOLL	8
3	NaN	22310	IVORY KNITTED MUG COSY	6
4	NaN	84969	BOX OF 6 ASSORTED COLOUR TEASPOONS	6
5	NaN	22623	BOX OF VINTAGE JIGSAW BLOCKS	3
6	NaN	22622	BOX OF VINTAGE ALPHABET BLOCKS	2
7	NaN	21754	HOME BUILDING BLOCK WORD	3
8	NaN	21755	LOVE BUILDING BLOCK WORD	3
9	NaN	21777	RECIPE BOX WITH METAL HEART	4

```
In [224]: pd.merge(d4, d5, how='inner')
```

Out[224]:

	InvoiceNo	StockCode	Description	Quantity
0	536367	22745	POPPY'S PLAYHOUSE BEDROOM	6

```
In [225]: pd.merge(d4, d5, how='outer')
```

```
Out[225]:
```

	InvoiceNo	StockCode	Description	Quantity
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	NaN
1	536365	71053	WHITE METAL LANTERN	NaN
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	NaN
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	NaN
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	NaN
5	536365	22752	SET 7 BABUSHKA NESTING BOXES	NaN
6	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	NaN
7	536366	22633	HAND WARMER UNION JACK	NaN
8	536366	22632	HAND WARMER RED POLKA DOT	NaN
9	536367	84879	ASSORTED COLOUR BIRD ORNAMENT	NaN
10	536367	22745	POPPY'S PLAYHOUSE BEDROOM	6.0
11	NaN	22748	POPPY'S PLAYHOUSE KITCHEN	6.0
12	NaN	22749	FELTCRAFT PRINCESS CHARLOTTE DOLL	8.0
13	NaN	22310	IVORY KNITTED MUG COSY	6.0
14	NaN	84969	BOX OF 6 ASSORTED COLOUR TEASPOONS	6.0
15	NaN	22623	BOX OF VINTAGE JIGSAW BLOCKS	3.0
16	NaN	22622	BOX OF VINTAGE ALPHABET BLOCKS	2.0
17	NaN	21754	HOME BUILDING BLOCK WORD	3.0
18	NaN	21755	LOVE BUILDING BLOCK WORD	3.0
19	NaN	21777	RECIPE BOX WITH METAL HEART	4.0

```
In [227]: pd.merge(d4, d5, on='StockCode', how='left')
```

Out[227]:

	InvoiceNo	StockCode	Description_x	Description_y	Quantity
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	NaN	NaN
1	536365	71053	WHITE METAL LANTERN	NaN	NaN
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	NaN	NaN
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	NaN	NaN
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	NaN	NaN
5	536365	22752	SET 7 BABUSHKA NESTING BOXES	NaN	NaN
6	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	NaN	NaN
7	536366	22633	HAND WARMER UNION JACK	NaN	NaN
8	536366	22632	HAND WARMER RED POLKA DOT	NaN	NaN
9	536367	84879	ASSORTED COLOUR BIRD ORNAMENT	NaN	NaN
10	536367	22745	POPPY'S PLAYHOUSE BEDROOM	POPPY'S PLAYHOUSE BEDROOM	6.0

```
In [228]: pd.merge(d4, d5, on='Description', how='left')
```

Out[228]:

	InvoiceNo	StockCode_x	Description	StockCode_y	Quantity
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	NaN	NaN
1	536365	71053	WHITE METAL LANTERN	NaN	NaN
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	NaN	NaN
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	NaN	NaN
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	NaN	NaN
5	536365	22752	SET 7 BABUSHKA NESTING BOXES	NaN	NaN
6	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	NaN	NaN
7	536366	22633	HAND WARMER UNION JACK	NaN	NaN
8	536366	22632	HAND WARMER RED POLKA DOT	NaN	NaN
9	536367	84879	ASSORTED COLOUR BIRD ORNAMENT	NaN	NaN
10	536367	22745	POPPY'S PLAYHOUSE BEDROOM	22745	6.0

```
In [229]: pd.merge(d4, d5, on=['StockCode', 'Description'], how='left')
```

```
Out[229]:
```

	InvoiceNo	StockCode	Description	Quantity
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	NaN
1	536365	71053	WHITE METAL LANTERN	NaN
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	NaN
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	NaN
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	NaN
5	536365	22752	SET 7 BABUSHKA NESTING BOXES	NaN
6	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	NaN
7	536366	22633	HAND WARMER UNION JACK	NaN
8	536366	22632	HAND WARMER RED POLKA DOT	NaN
9	536367	84879	ASSORTED COLOUR BIRD ORNAMENT	NaN
10	536367	22745	POPPY'S PLAYHOUSE BEDROOM	6.0

```
In [232]: pd.merge(d4, d5, on=['StockCode', 'Description'], how='left')
```

```
Out[232]:
```

	InvoiceNo	StockCode	Description	UnitPrice_x	Quantity	UnitPrice_y
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	2.55	NaN	NaN
1	536365	71053	WHITE METAL LANTERN	3.39	NaN	NaN
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	2.75	NaN	NaN
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	3.39	NaN	NaN
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	3.39	NaN	NaN
5	536365	22752	SET 7 BABUSHKA NESTING BOXES	7.65	NaN	NaN
6	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	4.25	NaN	NaN
7	536366	22633	HAND WARMER UNION JACK	1.85	NaN	NaN
8	536366	22632	HAND WARMER RED POLKA DOT	1.85	NaN	NaN
9	536367	84879	ASSORTED COLOUR BIRD ORNAMENT	1.69	NaN	NaN
10	536367	22745	POPPY'S PLAYHOUSE BEDROOM	2.10	6.0	2.1

```
In [237]: pd.merge(d4, d5, left_index = True, right_index=True, how='outer')
```

```
Out[237]:
```

	InvoiceNo	StockCode_x	Description_x	UnitPrice_x	StockCode_y	Description_y	Quantity	UnitPrice_y
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	2.55	NaN	NaN	NaN	NaN
1	536365	71053	WHITE METAL LANTERN	3.39	NaN	NaN	NaN	NaN
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	2.75	NaN	NaN	NaN	NaN
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	3.39	NaN	NaN	NaN	NaN
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	3.39	NaN	NaN	NaN	NaN
5	536365	22752	SET 7 BABUSHKA NESTING BOXES	7.65	NaN	NaN	NaN	NaN
6	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	4.25	NaN	NaN	NaN	NaN
7	536366	22633	HAND WARMER UNION JACK	1.85	NaN	NaN	NaN	NaN
8	536366	22632	HAND WARMER RED POLKA DOT	1.85	NaN	NaN	NaN	NaN
9	536367	84879	ASSORTED COLOUR BIRD ORNAMENT	1.69	NaN	NaN	NaN	NaN
10	536367	22745	POPPY'S PLAYHOUSE BEDROOM	2.10	22745	POPPY'S PLAYHOUSE BEDROOM	6.0	2.10
11	NaN	NaN	NaN	NaN	22748	POPPY'S PLAYHOUSE KITCHEN	6.0	2.10
12	NaN	NaN	NaN	NaN	22749	FELTCRAFT PRINCESS CHARLOTTE DOLL	8.0	3.75
13	NaN	NaN	NaN	NaN	22310	IVORY KNITTED MUG COSY	6.0	1.65
14	NaN	NaN	NaN	NaN	84969	BOX OF 6 ASSORTED COLOUR TEASPOONS	6.0	4.25
15	NaN	NaN	NaN	NaN	22623	BOX OF VINTAGE JIGSAW BLOCKS	3.0	4.95
16	NaN	NaN	NaN	NaN	22622	BOX OF VINTAGE ALPHABET BLOCKS	2.0	9.95
17	NaN	NaN	NaN	NaN	21754	HOME BUILDING BLOCK WORD	3.0	5.95
18	NaN	NaN	NaN	NaN	21755	LOVE BUILDING BLOCK WORD	3.0	5.95
19	NaN	NaN	NaN	NaN	21777	RECIPE BOX WITH METAL HEART	4.0	7.95

```
In [242]: d4 = pd.DataFrame(df, columns=['StockCode', 'UnitPrice'])[0:11]
          d5 = pd.DataFrame(df, columns=['Description', 'Quantity'])[10:20]
          d4.join(d5)
```

Out [242]:

	StockCode	UnitPrice	Description	Quantity
0	85123A	2.55	NaN	NaN
1	71053	3.39	NaN	NaN
2	84406B	2.75	NaN	NaN
3	84029G	3.39	NaN	NaN
4	84029E	3.39	NaN	NaN
5	22752	7.65	NaN	NaN
6	21730	4.25	NaN	NaN
7	22633	1.85	NaN	NaN
8	22632	1.85	NaN	NaN
9	84879	1.69	NaN	NaN
10	22745	2.10	POPPY'S PLAYHOUSE BEDROOM	6.0

```
In [243]: d4 = pd.DataFrame(df, columns=['StockCode', 'UnitPrice'])[0:11]
          d5 = pd.DataFrame(df, columns=['Description', 'Quantity'])[10:20]
          d5.join(d4)
```

Out[243]:

	Description	Quantity	StockCode	UnitPrice
10	POPPY'S PLAYHOUSE BEDROOM	6	22745	2.1
11	POPPY'S PLAYHOUSE KITCHEN	6	NaN	NaN
12	FELTCRAFT PRINCESS CHARLOTTE DOLL	8	NaN	NaN
13	IVORY KNITTED MUG COSY	6	NaN	NaN
14	BOX OF 6 ASSORTED COLOUR TEASPOONS	6	NaN	NaN
15	BOX OF VINTAGE JIGSAW BLOCKS	3	NaN	NaN
16	BOX OF VINTAGE ALPHABET BLOCKS	2	NaN	NaN
17	HOME BUILDING BLOCK WORD	3	NaN	NaN
18	LOVE BUILDING BLOCK WORD	3	NaN	NaN
19	RECIPE BOX WITH METAL HEART	4	NaN	NaN



```
In [59]: d6 = pd.DataFrame.copy(df)[0:5]
d7 = pd.DataFrame.copy(df)[2:8]

d6.loc[3, ['Quantity']] = 110
d6.loc[4, ['Quantity']] = 110

d7.loc[3, ['Quantity']] = 210
d7.loc[4, ['Quantity']] = 210
pd.concat([d6, d7], keys=['d6', 'd7'])
```

Out[59]:

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	
d6	0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	12/1/10 8:26	2.55	17850.0	United Kingdom
	1	536365	71053	WHITE METAL LANTERN	6	12/1/10 8:26	3.39	17850.0	United Kingdom
	2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/1/10 8:26	2.75	17850.0	United Kingdom
	3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	110	12/1/10 8:26	3.39	17850.0	United Kingdom
	4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	110	12/1/10 8:26	3.39	17850.0	United Kingdom
d7	2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/1/10 8:26	2.75	17850.0	United Kingdom
	3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	210	12/1/10 8:26	3.39	17850.0	United Kingdom
	4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	210	12/1/10 8:26	3.39	17850.0	United Kingdom
	5	536365	22752	SET 7 BABUSHKA NESTING BOXES	2	12/1/10 8:26	7.65	17850.0	United Kingdom
	6	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	6	12/1/10 8:26	4.25	17850.0	United Kingdom
	7	536366	22633	HAND WARMER UNION JACK	6	12/1/10 8:28	1.85	17850.0	United Kingdom

```
In [61]: d6.combine_first(d7)
```

Out[61]:

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	
	0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6.0	12/1/10 8:26	2.55	17850.0	United Kingdom
	1	536365	71053	WHITE METAL LANTERN	6.0	12/1/10 8:26	3.39	17850.0	United Kingdom
	2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8.0	12/1/10 8:26	2.75	17850.0	United Kingdom
	3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	110.0	12/1/10 8:26	3.39	17850.0	United Kingdom
	4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	110.0	12/1/10 8:26	3.39	17850.0	United Kingdom
	5	536365	22752	SET 7 BABUSHKA NESTING BOXES	2.0	12/1/10 8:26	7.65	17850.0	United Kingdom
	6	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	6.0	12/1/10 8:26	4.25	17850.0	United Kingdom
	7	536366	22633	HAND WARMER UNION JACK	6.0	12/1/10 8:28	1.85	17850.0	United Kingdom

```
In [63]: d7.combine_first(d6)
```

```
Out[63]:
```

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6.0	12/1/10 8:26	2.55	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6.0	12/1/10 8:26	3.39	17850.0	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8.0	12/1/10 8:26	2.75	17850.0	United Kingdom
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	210.0	12/1/10 8:26	3.39	17850.0	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	210.0	12/1/10 8:26	3.39	17850.0	United Kingdom
5	536365	22752	SET 7 BABUSHKA NESTING BOXES	2.0	12/1/10 8:26	7.65	17850.0	United Kingdom
6	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	6.0	12/1/10 8:26	4.25	17850.0	United Kingdom
7	536366	22633	HAND WARMER UNION JACK	6.0	12/1/10 8:28	1.85	17850.0	United Kingdom

```
In [8]: pd.value_counts(df['Country'])
```

```
Out[8]:
```

United Kingdom	61186
Germany	982
France	967
EIRE	504
Spain	355
Portugal	212
Netherlands	186
Switzerland	175
Norway	147
Australia	142
Belgium	142
Italy	112
Cyprus	99
Japan	69
Sweden	41
Lithuania	35
Poland	33
Iceland	31
Denmark	20
Channel Islands	17
Finland	17
Israel	16
Austria	9
Bahrain	2

Name: Country, dtype: int64

```
In [48]: d1 = pd.DataFrame(df, columns = ['InvoiceNo', 'StockCode', 'Description'])[0:100]
d2 = pd.DataFrame(df, columns = ['Description', 'InvoiceDate', 'Quantity'])[0:100]

pd.merge(d1, d2)
```

Out[48]:

	InvoiceNo	StockCode	Description	InvoiceDate	Quantity
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	12/1/10 8:26	6
1	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	12/1/10 9:02	6
2	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	12/1/10 9:32	6
3	536373	85123A	WHITE HANGING HEART T-LIGHT HOLDER	12/1/10 8:26	6
4	536373	85123A	WHITE HANGING HEART T-LIGHT HOLDER	12/1/10 9:02	6
5	536373	85123A	WHITE HANGING HEART T-LIGHT HOLDER	12/1/10 9:32	6
6	536375	85123A	WHITE HANGING HEART T-LIGHT HOLDER	12/1/10 8:26	6
7	536375	85123A	WHITE HANGING HEART T-LIGHT HOLDER	12/1/10 9:02	6
8	536375	85123A	WHITE HANGING HEART T-LIGHT HOLDER	12/1/10 9:32	6
9	536365	71053	WHITE METAL LANTERN	12/1/10 8:26	6
10	536365	71053	WHITE METAL LANTERN	12/1/10 9:02	6
11	536365	71053	WHITE METAL LANTERN	12/1/10 9:32	6
12	536373	71053	WHITE METAL LANTERN	12/1/10 8:26	6
13	536373	71053	WHITE METAL LANTERN	12/1/10 9:02	6
14	536373	71053	WHITE METAL LANTERN	12/1/10 9:32	6
15	536375	71053	WHITE METAL LANTERN	12/1/10 8:26	6
16	536375	71053	WHITE METAL LANTERN	12/1/10 9:02	6
17	536375	71053	WHITE METAL LANTERN	12/1/10 9:32	6
18	536365	84406B	CREAM CUPID HEARTS COAT HANGER	12/1/10 8:26	8
19	536365	84406B	CREAM CUPID HEARTS COAT HANGER	12/1/10 9:02	8
20	536365	84406B	CREAM CUPID HEARTS COAT HANGER	12/1/10 9:32	8
21	536373	84406B	CREAM CUPID HEARTS COAT HANGER	12/1/10 8:26	8

158	536378	22388	JUMBO BAG PINK POLKADOT	12/1/10 9:37	10
159	536378	85099C	JUMBO BAG BAROQUE BLACK WHITE	12/1/10 9:37	10
160	536378	21033	JUMBO BAG CHARLIE AND LOLA TOYS	12/1/10 9:37	10
161	536378	20723	STRAWBERRY CHARLOTTE BAG	12/1/10 9:37	10
162	536378	84997B	RED 3 PIECE RETROSPOT CUTLERY SET	12/1/10 9:37	12
163	536378	84997C	BLUE 3 PIECE POLKADOT CUTLERY SET	12/1/10 9:37	6
164	536378	21094	SET/6 RED SPOTTY PAPER PLATES	12/1/10 9:37	12
165	536378	20725	LUNCH BAG RED RETROSPOT	12/1/10 9:37	10
166	536378	21559	STRAWBERRY LUNCH BOX WITH CUTLERY	12/1/10 9:37	6
167	536378	22352	LUNCH BOX WITH CUTLERY RETROSPOT	12/1/10 9:37	6
168	536378	21212	PACK OF 72 RETROSPOT CAKE CASES	12/1/10 9:37	120
169	536378	21975	PACK OF 60 DINOSAUR CAKE CASES	12/1/10 9:37	24
170	536378	21977	PACK OF 60 PINK PAISLEY CAKE CASES	12/1/10 9:37	24
171	536378	84991	60 TEATIME FAIRY CAKE CASES	12/1/10 9:37	24

172 rows x 5 columns

```
In [62]: d1 = pd.DataFrame(df, columns = ['InvoiceNo', 'StockCode', 'Description'])[0:100]
d2 = pd.DataFrame(df, columns = ['Description', 'InvoiceDate', 'Quantity'])[0:100]

pd.merge(d1, d2).drop_duplicates(['StockCode'])
```

Out[62]:

	InvoiceNo	StockCode	Description	InvoiceDate	Quantity
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	12/1/10 8:26	6
9	536365	71053	WHITE METAL LANTERN	12/1/10 8:26	6
18	536365	84406B	CREAM CUPID HEARTS COAT HANGER	12/1/10 8:26	8
27	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	12/1/10 8:26	6
36	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	12/1/10 8:26	6
45	536365	22752	SET 7 BABUSHKA NESTING BOXES	12/1/10 8:26	2
54	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	12/1/10 8:26	6
63	536366	22633	HAND WARMER UNION JACK	12/1/10 8:28	6
72	536366	22632	HAND WARMER RED POLKA DOT	12/1/10 8:28	6
81	536367	84879	ASSORTED COLOUR BIRD ORNAMENT	12/1/10 8:34	32
82	536367	22745	POPPY'S PLAYHOUSE BEDROOM	12/1/10 8:34	6
83	536367	22748	POPPY'S PLAYHOUSE KITCHEN	12/1/10 8:34	6
84	536367	22749	FELTCRAFT PRINCESS CHARLOTTE DOLL	12/1/10 8:34	8
85	536367	22310	IVORY KNITTED MUG COSY	12/1/10 8:34	6
86	536367	84969	BOX OF 6 ASSORTED COLOUR TEASPOONS	12/1/10 8:34	6
87	536367	22623	BOX OF VINTAGE JIGSAW BLOCKS	12/1/10 8:34	3
88	536367	22622	BOX OF VINTAGE ALPHABET BLOCKS	12/1/10 8:34	2

165	536378	20725	LUNCH BAG RED RETROSPOT	12/1/10 9:37	10
166	536378	21559	STRAWBERRY LUNCH BOX WITH CUTLERY	12/1/10 9:37	6
167	536378	22352	LUNCH BOX WITH CUTLERY RETROSPOT	12/1/10 9:37	6
168	536378	21212	PACK OF 72 RETROSPOT CAKE CASES	12/1/10 9:37	120
169	536378	21975	PACK OF 60 DINOSAUR CAKE CASES	12/1/10 9:37	24
170	536378	21977	PACK OF 60 PINK PAISLEY CAKE CASES	12/1/10 9:37	24
171	536378	84991	60 TEATIME FAIRY CAKE CASES	12/1/10 9:37	24

73 rows × 5 columns

```
In [66]: d3.reset_index()
```

```
Out[66]:
```

	index	InvoiceNo	StockCode	Description	InvoiceDate	Quantity
0	0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	12/1/10 8:26	6
1	9	536365	71053	WHITE METAL LANTERN	12/1/10 8:26	6
2	18	536365	84406B	CREAM CUPID HEARTS COAT HANGER	12/1/10 8:26	8
3	27	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	12/1/10 8:26	6
4	36	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	12/1/10 8:26	6
5	45	536365	22752	SET 7 BABUSHKA NESTING BOXES	12/1/10 8:26	2
6	54	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	12/1/10 8:26	6
7	63	536366	22633	HAND WARMER UNION JACK	12/1/10 8:28	6
8	72	536366	22632	HAND WARMER RED POLKA DOT	12/1/10 8:28	6
9	81	536367	84879	ASSORTED COLOUR BIRD ORNAMENT	12/1/10 8:34	32
10	82	536367	22745	POPPY'S PLAYHOUSE BEDROOM	12/1/10 8:34	6
11	83	536367	22748	POPPY'S PLAYHOUSE KITCHEN	12/1/10 8:34	6
12	84	536367	22749	FELTCRAFT PRINCESS CHARLOTTE DOLL	12/1/10 8:34	8
13	85	536367	22310	IVORY KNITTED MUG COSY	12/1/10 8:34	6
14	86	536367	84969	BOX OF 6 ASSORTED COLOUR TEASPOONS	12/1/10 8:34	6

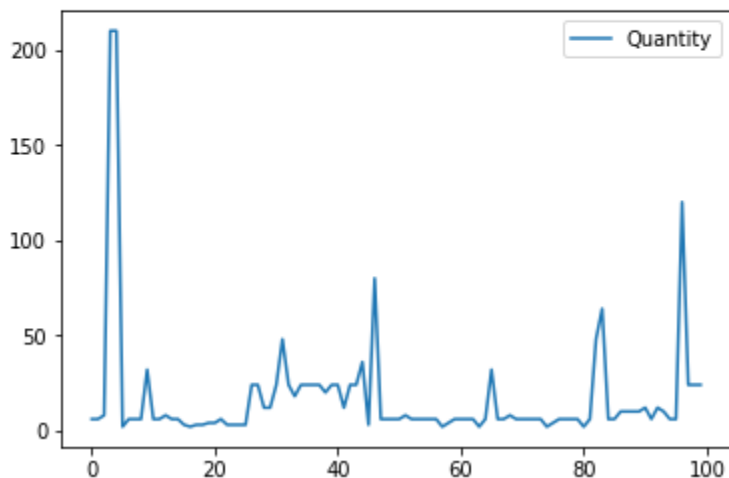
```
In [67]: d3.reset_index(drop=True)
```

```
Out[67]:
```

	InvoiceNo	StockCode	Description	InvoiceDate	Quantity
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	12/1/10 8:26	6
1	536365	71053	WHITE METAL LANTERN	12/1/10 8:26	6
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	12/1/10 8:26	8
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	12/1/10 8:26	6
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	12/1/10 8:26	6
5	536365	22752	SET 7 BABUSHKA NESTING BOXES	12/1/10 8:26	2
6	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	12/1/10 8:26	6
7	536366	22633	HAND WARMER UNION JACK	12/1/10 8:28	6
8	536366	22632	HAND WARMER RED POLKA DOT	12/1/10 8:28	6
9	536367	84879	ASSORTED COLOUR BIRD ORNAMENT	12/1/10 8:34	32
10	536367	22745	POPPY'S PLAYHOUSE BEDROOM	12/1/10 8:34	6
11	536367	22748	POPPY'S PLAYHOUSE KITCHEN	12/1/10 8:34	6
12	536367	22749	FELTCRAFT PRINCESS CHARLOTTE DOLL	12/1/10 8:34	8
13	536367	22310	IVORY KNITTED MUG COSY	12/1/10 8:34	6
14	536367	84969	BOX OF 6 ASSORTED COLOUR TEASPOONS	12/1/10 8:34	6
15	536367	22623	BOX OF VINTAGE JIGSAW BLOCKS	12/1/10 8:34	3
16	536367	22622	BOX OF VINTAGE ALPHABET BLOCKS	12/1/10 8:34	2
17	536367	21754	HOME BUILDING BLOCK WORD	12/1/10 8:34	3
18	536367	21755	LOVE BUILDING BLOCK WORD	12/1/10 8:34	3

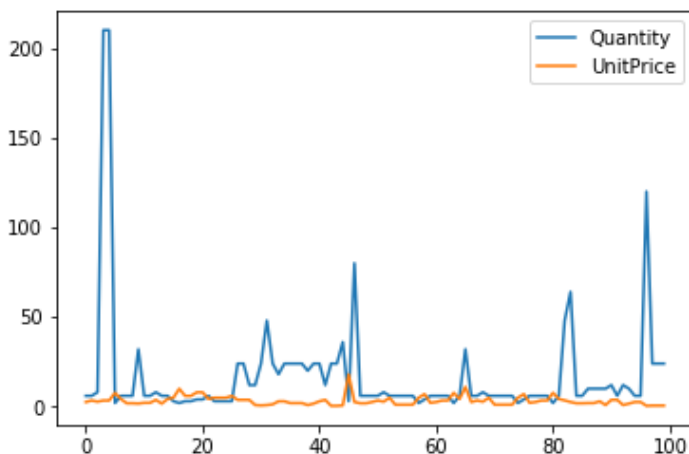
```
In [76]: d8 = pd.DataFrame(df, columns=['Quantity'])[0:100]
         d8.plot()
```

```
Out[76]: <matplotlib.axes._subplots.AxesSubplot at 0x1b39e3aeb00>
```



```
In [82]: d8 = pd.DataFrame(df, columns=['Quantity', 'UnitPrice'])[0:100]
         d8.plot()
```

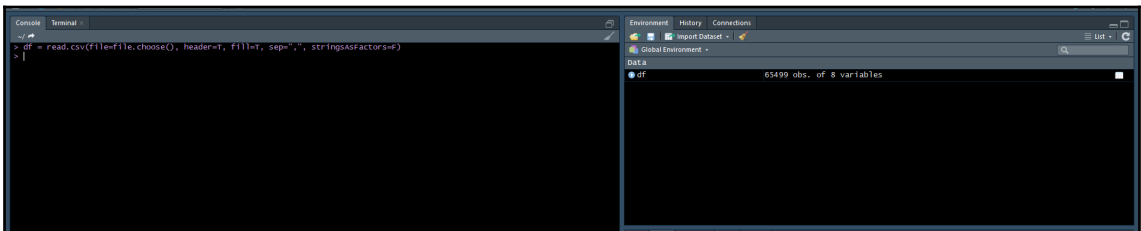
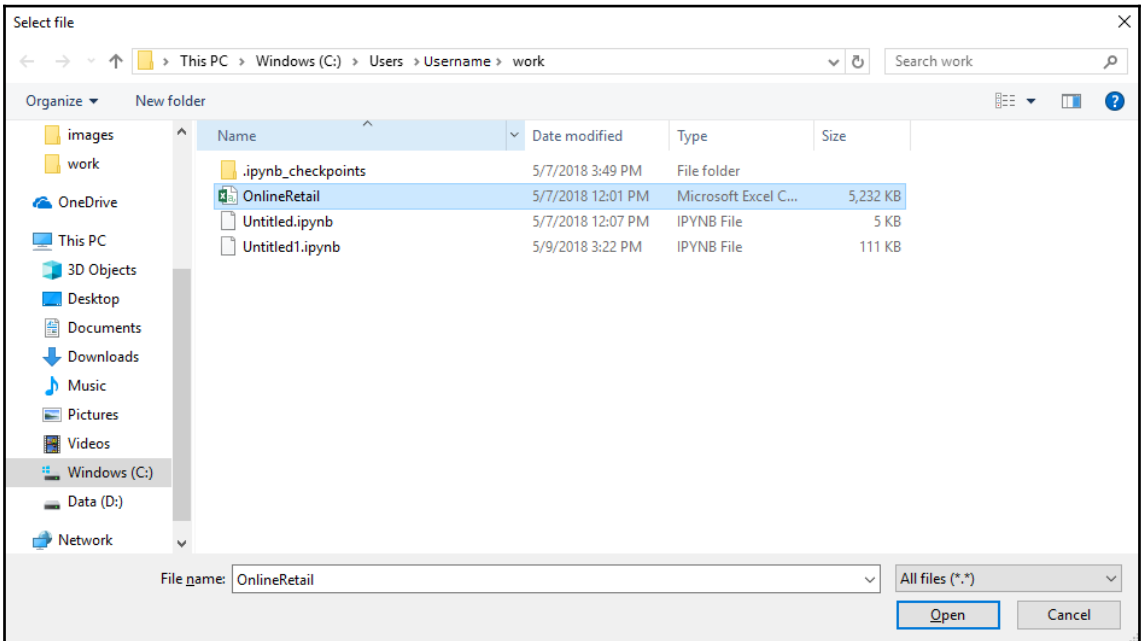
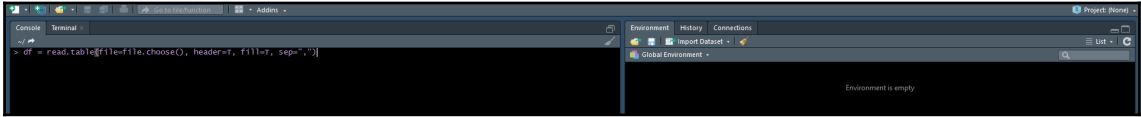
```
Out[82]: <matplotlib.axes._subplots.AxesSubplot at 0x1b39f21ef98>
```





---

# Chapter 5: Statistical Big Data Computing with R and Hadoop



The screenshot displays the RStudio interface with the following components:

- Environment Pane:** Shows a data frame 'df' with 65499 observations and 8 variables.
- Console:** Contains the R code:
 

```
df = read.csv(file=file.choose(), header=T, fill=T, sep=";", stringsAsFactors=F)
> view(df)
```
- Data Table:** A table with 20 rows and 8 columns:
 

InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
1	85123A	WHITE HANGING HEAT TIGHT HOLDER	6	12/1/10 8:26	2.55	17850	United Kingdom
2	536365	WHITE METAL LANTERN	6	12/1/10 8:26	3.39	17850	United Kingdom
3	536365	CREAM CURD HEARTS COAT HANGER	8	12/1/10 8:26	2.75	17850	United Kingdom
4	536365	KNITTED UNION FLAG HOT WATER BOTTLE	6	12/1/10 8:26	3.39	17850	United Kingdom
5	536365	RED WOOLLY HOTTIE WHITE HEART	6	12/1/10 8:26	3.39	17850	United Kingdom
6	536365	SET 7 BABUSHKA NESTING BOXES	2	12/1/10 8:26	7.65	17850	United Kingdom
7	536365	GLASS STAR FROSTED TIGHT HOLDER	6	12/1/10 8:26	4.25	17850	United Kingdom
8	536366	HAND WARMER UNION JACK	6	12/1/10 8:28	1.85	17850	United Kingdom
9	536366	HAND WARMER RED POLKA DOT	6	12/1/10 8:28	1.85	17850	United Kingdom
10	536367	ASSORTED COLOUR BIRD ORNAMENT	32	12/1/10 8:34	1.69	13047	United Kingdom
11	536367	POPPY'S PLAYHOUSE BEDROOM	6	12/1/10 8:34	2.10	13047	United Kingdom
12	536367	POPPY'S PLAYHOUSE KITCHEN	6	12/1/10 8:34	2.10	13047	United Kingdom
13	536367	FELICITY PRINCESS CHARLOTTE DOLL	8	12/1/10 8:34	3.75	13047	United Kingdom
14	536367	TWO KNITTED RUG DOGS	6	12/1/10 8:34	2.65	13047	United Kingdom
15	536367	BOX OF 6 ASSORTED COLOUR TEASPOONS	6	12/1/10 8:34	4.25	13047	United Kingdom
16	536367	BOX OF VINTAGE JIGSAW BLOCKS	3	12/1/10 8:34	4.95	13047	United Kingdom
17	536367	BOX OF VINTAGE ALPHABET BLOCKS	2	12/1/10 8:34	9.95	13047	United Kingdom
18	536367	HOME BUILDING BLOCK WORD	3	12/1/10 8:34	5.95	13047	United Kingdom
19	536367	LOVE BUILDING BLOCK WORD	3	12/1/10 8:34	5.95	13047	United Kingdom
20	536367	RECFE BOX WITH METAL HEART	4	12/1/10 8:34	7.95	13047	United Kingdom

InvoiceNo	StockCode	Description	
1	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER
2	536365	71053	WHITE METAL LANTERN
3	536365	84406B	CREAM CUPID HEARTS COAT HANGER
4	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE
5	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.
6	536365	22752	SET 7 BABUSHKA NESTING BOXES
7	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER
8	536366	22633	HAND WARMER UNION JACK
9	536366	22632	HAND WARMER RED POLKA DOT
10	536367	84879	ASSORTED COLOUR BIRD ORNAMENT
11	536367	22745	POPPYS PLAYHOUSE BEDROOM ,6,12/1/10 8:34,2....
12	536367	22749	FELTCRAFT PRINCESS CHARLOTTE DOLL
13	536367	22310	IVORY KNITTED MUG COSY
14	536367	84969	BOX OF 6 ASSORTED COLOUR TEASPOONS
15	536367	22623	BOX OF VINTAGE JIGSAW BLOCKS
16	536367	22622	BOX OF VINTAGE ALPHABET BLOCKS
17	536367	21754	HOME BUILDING BLOCK WORD
18	536367	21755	LOVE BUILDING BLOCK WORD
19	536367	21777	RECIPE BOX WTH METAL HEART
20	536367	48187	DOORMAT NEW ENGLAND

Showing 1 to 20 of 27,080 entries

	InvoiceNo	StockCode	Description
1	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER
2	536365	71053	WHITE METAL LANTERN
3	536365	84406B	CREAM CUPID HEARTS COAT HANGER
4	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE
5	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.
6	536365	22752	SET 7 BABUSHKA NESTING BOXES
7	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER
8	536366	22633	HAND WARMER UNION JACK
9	536366	22632	HAND WARMER RED POLKA DOT
10	536367	84879	ASSORTED COLOUR BIRD ORNAMENT

Showing 1 to 10 of 10 entries

```
> v1[1:5]
[1] "WHITE HANGING HEART T-LIGHT HOLDER" "WHITE METAL LANTERN" "CREAM CUPID HEARTS COAT HANGER"
[4] "KNITTED UNION FLAG HOT WATER BOTTLE" "RED WOOLLY HOTTIE WHITE HEART."
```

```
> v2 = df$Description
> v2[1:5]
[1] "WHITE HANGING HEART T-LIGHT HOLDER" "WHITE METAL LANTERN" "CREAM CUPID HEARTS COAT HANGER"
[4] "KNITTED UNION FLAG HOT WATER BOTTLE" "RED WOOLLY HOTTIE WHITE HEART."
> |
```

```
> d1[d1$stockCode == "85123A", ]
  InvoiceNo StockCode      Description
1     536365   85123A WHITE HANGING HEART T-LIGHT HOLDER
```

```
> d1 = df[1:10, c(1:8)]
> d1[2, c(1:8)]
  InvoiceNo StockCode      Description Quantity InvoiceDate UnitPrice CustomerID      Country
2     536365   71053 WHITE METAL LANTERN         6 12/1/10 8:26      3.39    17850 United Kingdom
```

```
> head(df)
  InvoiceNo StockCode      Description Quantity InvoiceDate UnitPrice CustomerID      Country
1     536365   85123A WHITE HANGING HEART T-LIGHT HOLDER         6 12/1/10 8:26      2.55    17850 United Kingdom
2     536365   71053           WHITE METAL LANTERN         6 12/1/10 8:26      3.39    17850 United Kingdom
3     536365   84406B      CREAM CUPID HEARTS COAT HANGER         8 12/1/10 8:26      2.75    17850 United Kingdom
4     536365   84029G KNITTED UNION FLAG HOT WATER BOTTLE         6 12/1/10 8:26      3.39    17850 United Kingdom
5     536365   84029E      RED WOOLLY HOTTIE WHITE HEART.         6 12/1/10 8:26      3.39    17850 United Kingdom
6     536365   22752           SET 7 BABUSHKA NESTING BOXES         2 12/1/10 8:26      7.65    17850 United Kingdom
> |
```

```
> head(df, 10)
  InvoiceNo StockCode Description Quantity InvoiceDate UnitPrice CustomerID Country
1 536365 85123A WHITE HANGING HEART T-LIGHT HOLDER 6 12/1/10 8:26 2.55 17850 United Kingdom
2 536365 71053 WHITE METAL LANTERN 6 12/1/10 8:26 3.39 17850 United Kingdom
3 536365 84406B CREAM CUPID HEARTS COAT HANGER 8 12/1/10 8:26 2.75 17850 United Kingdom
4 536365 84029G KNITTED UNION FLAG HOT WATER BOTTLE 6 12/1/10 8:26 3.39 17850 United Kingdom
5 536365 84029E RED WOOLLY HOTTIE WHITE HEART. 6 12/1/10 8:26 3.39 17850 United Kingdom
6 536365 22752 SET 7 BABUSHKA NESTING BOXES 2 12/1/10 8:26 7.65 17850 United Kingdom
7 536365 21730 GLASS STAR FROSTED T-LIGHT HOLDER 6 12/1/10 8:26 4.25 17850 United Kingdom
8 536366 22633 HAND WARMER UNION JACK 6 12/1/10 8:28 1.85 17850 United Kingdom
9 536366 22632 HAND WARMER RED POLKA DOT 6 12/1/10 8:28 1.85 17850 United Kingdom
10 536367 84879 ASSORTED COLOUR BIRD ORNAMENT 32 12/1/10 8:34 1.69 13047 United Kingdom
>
```

```
> head(d1, -2)
  InvoiceNo StockCode Description Quantity InvoiceDate UnitPrice CustomerID Country
1 536365 85123A WHITE HANGING HEART T-LIGHT HOLDER 6 12/1/10 8:26 2.55 17850 United Kingdom
2 536365 71053 WHITE METAL LANTERN 6 12/1/10 8:26 3.39 17850 United Kingdom
3 536365 84406B CREAM CUPID HEARTS COAT HANGER 8 12/1/10 8:26 2.75 17850 United Kingdom
4 536365 84029G KNITTED UNION FLAG HOT WATER BOTTLE 6 12/1/10 8:26 3.39 17850 United Kingdom
5 536365 84029E RED WOOLLY HOTTIE WHITE HEART. 6 12/1/10 8:26 3.39 17850 United Kingdom
6 536365 22752 SET 7 BABUSHKA NESTING BOXES 2 12/1/10 8:26 7.65 17850 United Kingdom
7 536365 21730 GLASS STAR FROSTED T-LIGHT HOLDER 6 12/1/10 8:26 4.25 17850 United Kingdom
8 536366 22633 HAND WARMER UNION JACK 6 12/1/10 8:28 1.85 17850 United Kingdom
>
```

```
> d1 = df[0:10, c(1:8)]
> tail(d1, 4)
  InvoiceNo StockCode Description Quantity InvoiceDate UnitPrice CustomerID Country
7 536365 21730 GLASS STAR FROSTED T-LIGHT HOLDER 6 12/1/10 8:26 4.25 17850 United Kingdom
8 536366 22633 HAND WARMER UNION JACK 6 12/1/10 8:28 1.85 17850 United Kingdom
9 536366 22632 HAND WARMER RED POLKA DOT 6 12/1/10 8:28 1.85 17850 United Kingdom
10 536367 84879 ASSORTED COLOUR BIRD ORNAMENT 32 12/1/10 8:34 1.69 13047 United Kingdom
>
```

```
> tail(d1, -2)
  InvoiceNo StockCode Description Quantity InvoiceDate UnitPrice CustomerID Country
3 536365 84406B CREAM CUPID HEARTS COAT HANGER 8 12/1/10 8:26 2.75 17850 United Kingdom
4 536365 84029G KNITTED UNION FLAG HOT WATER BOTTLE 6 12/1/10 8:26 3.39 17850 United Kingdom
5 536365 84029E RED WOOLLY HOTTIE WHITE HEART. 6 12/1/10 8:26 3.39 17850 United Kingdom
6 536365 22752 SET 7 BABUSHKA NESTING BOXES 2 12/1/10 8:26 7.65 17850 United Kingdom
7 536365 21730 GLASS STAR FROSTED T-LIGHT HOLDER 6 12/1/10 8:26 4.25 17850 United Kingdom
8 536366 22633 HAND WARMER UNION JACK 6 12/1/10 8:28 1.85 17850 United Kingdom
9 536366 22632 HAND WARMER RED POLKA DOT 6 12/1/10 8:28 1.85 17850 United Kingdom
10 536367 84879 ASSORTED COLOUR BIRD ORNAMENT 32 12/1/10 8:34 1.69 13047 United Kingdom
>
```

```
> quantile(df$UnitPrice)
 0%    25%    50%    75%   100%
 0.00  1.25  2.51  4.24 16888.02
>
```

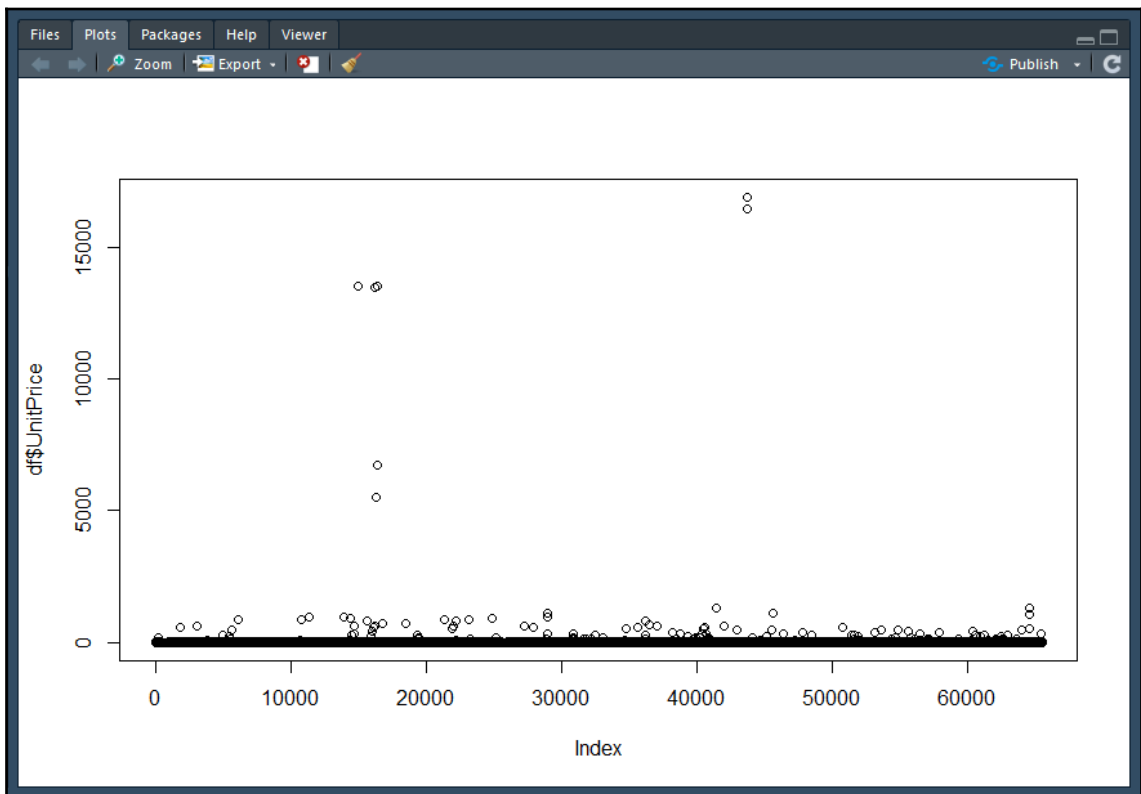
```
> quantile(df$UnitPrice, c(0, .1, .5, .9))
 0% 10% 50% 90%
 0.00 0.83 2.51 7.95
>
```

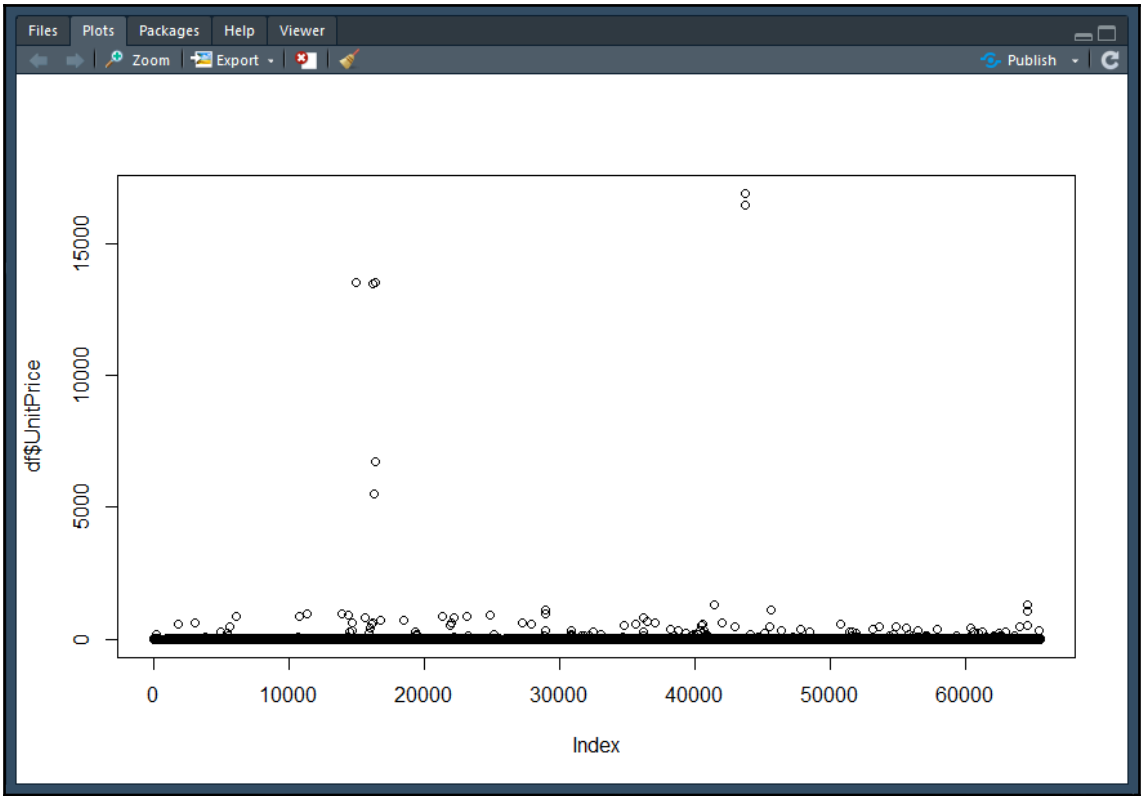
```
> sd(df$UnitPrice)
[1] 145.796
> |
```

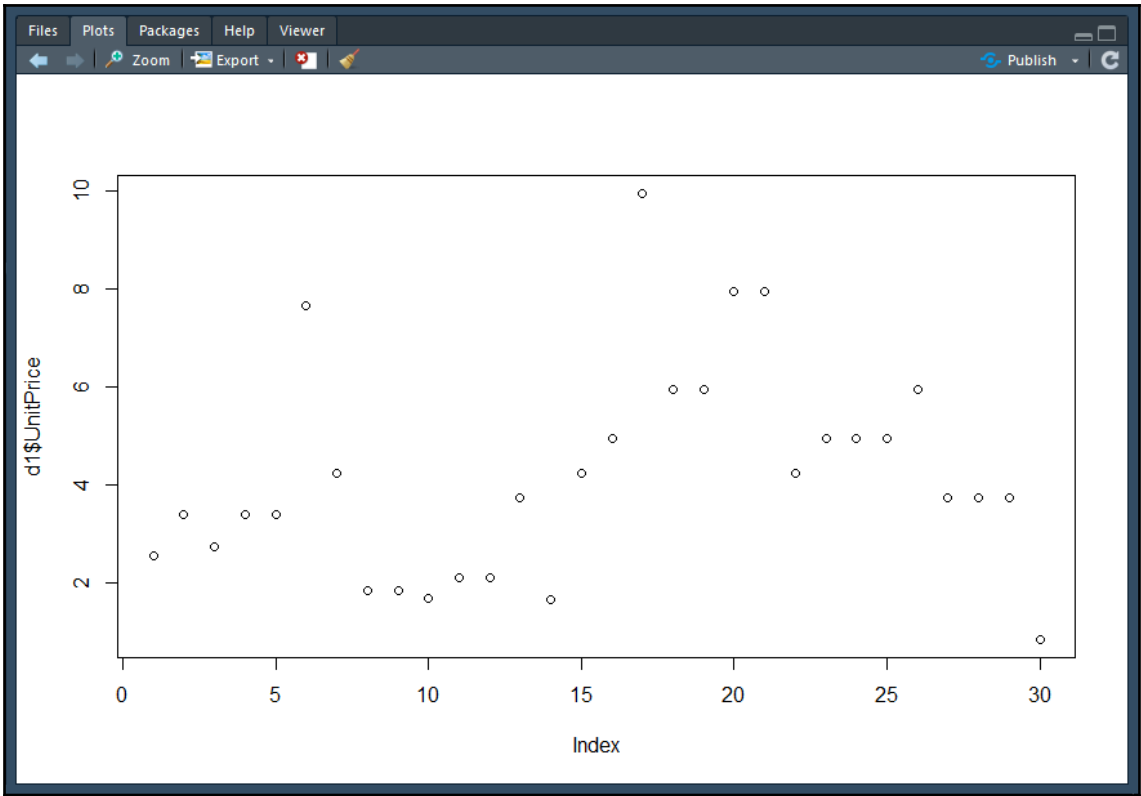
```
> var(df$UnitPrice)
[1] 21256.46
> |
```

```
> range(df$UnitPrice)
[1] 0.00 16888.02
> |
```

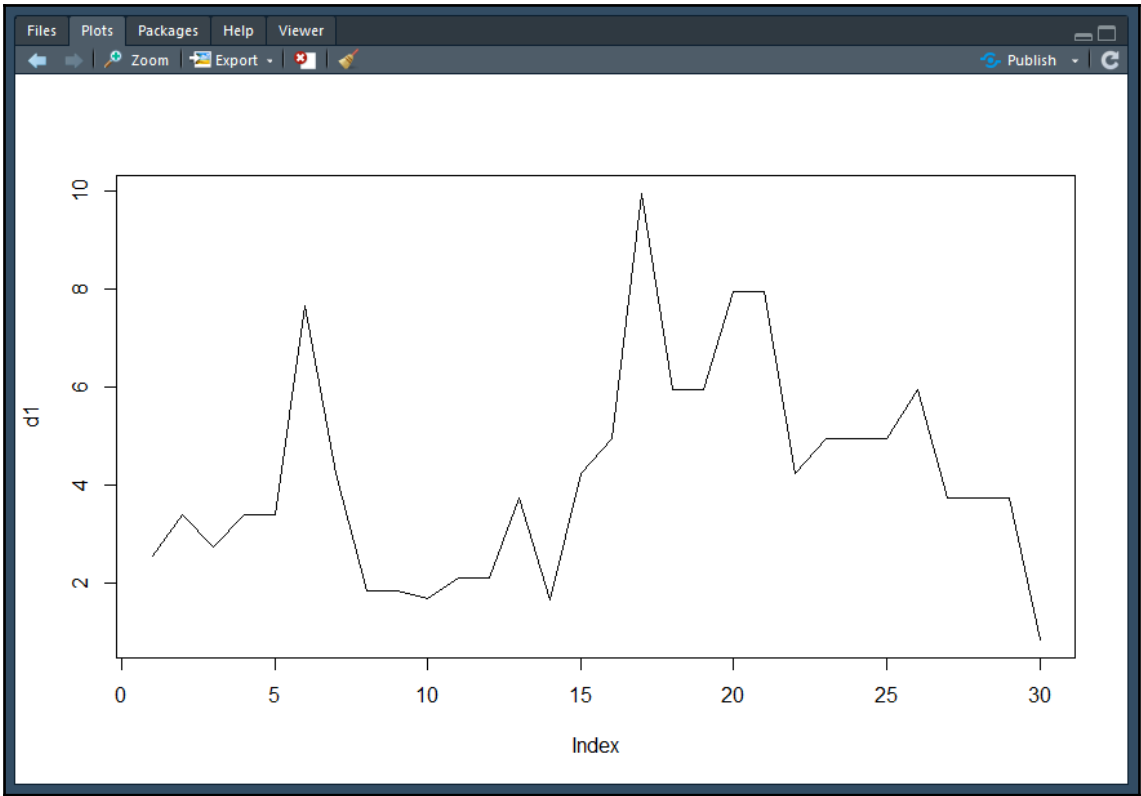
```
> fivenum(df$UnitPrice)
[1] 0.00 1.25 2.51 4.24 16888.02
> |
```

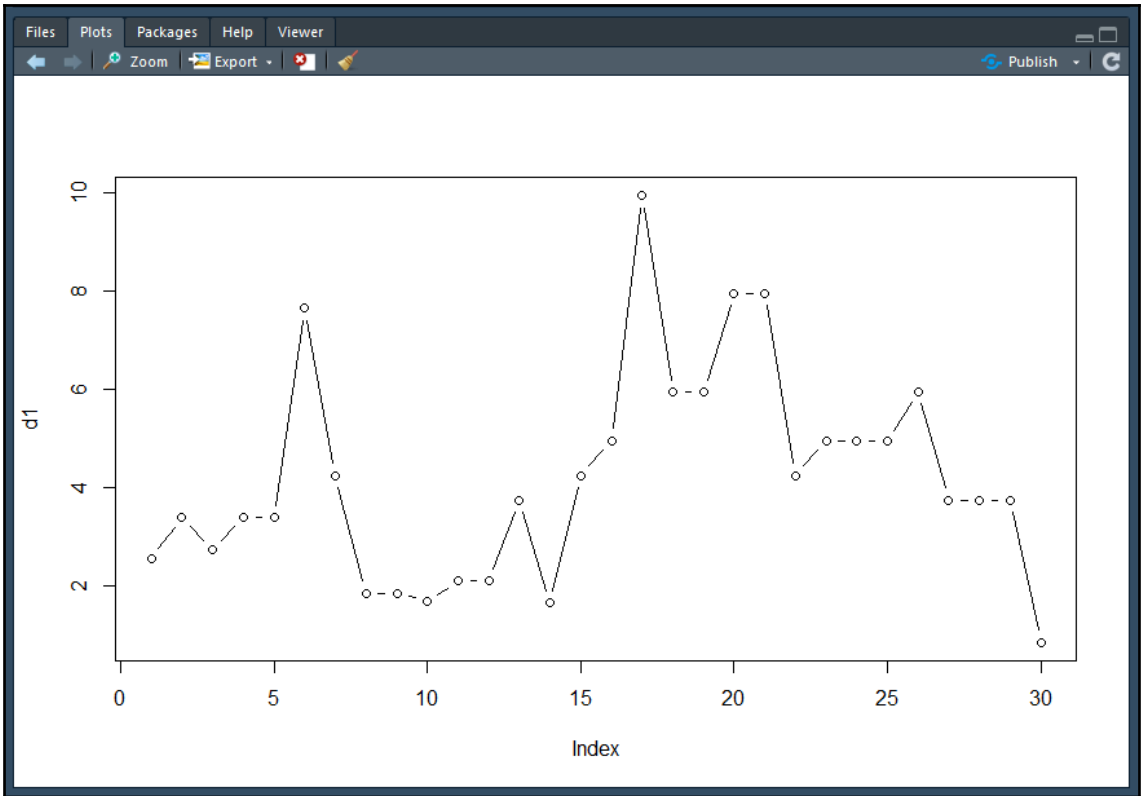


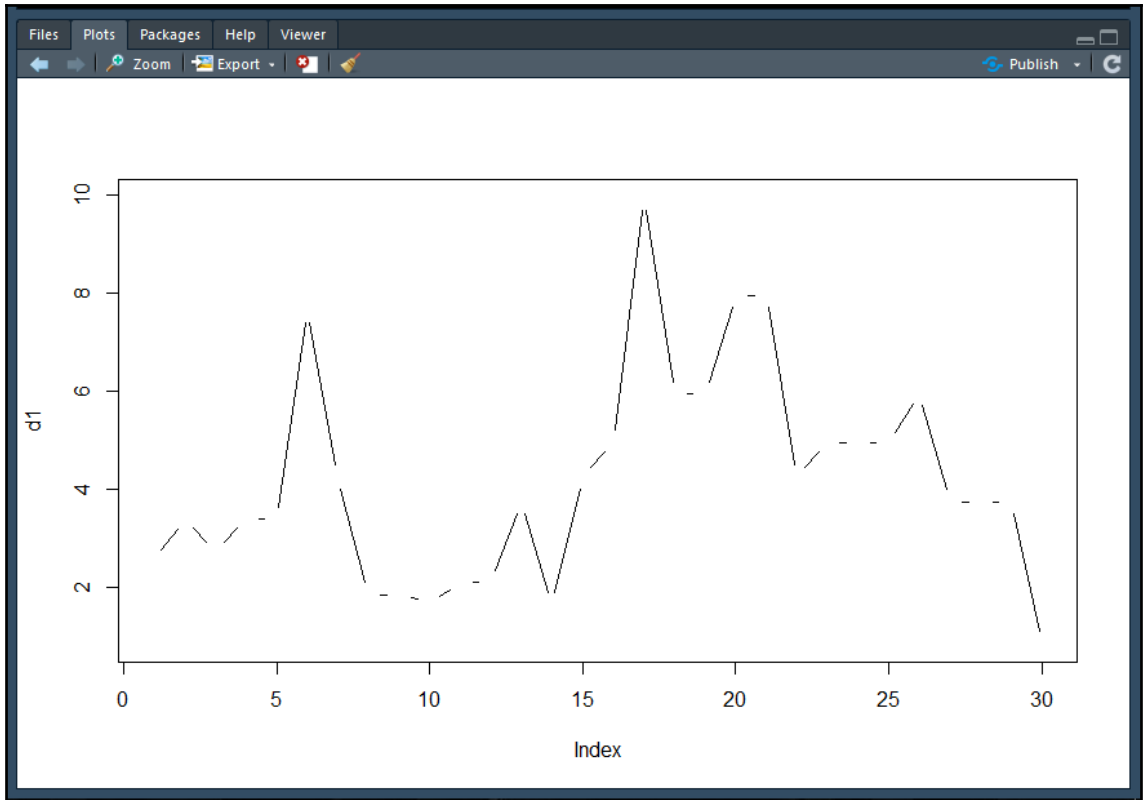


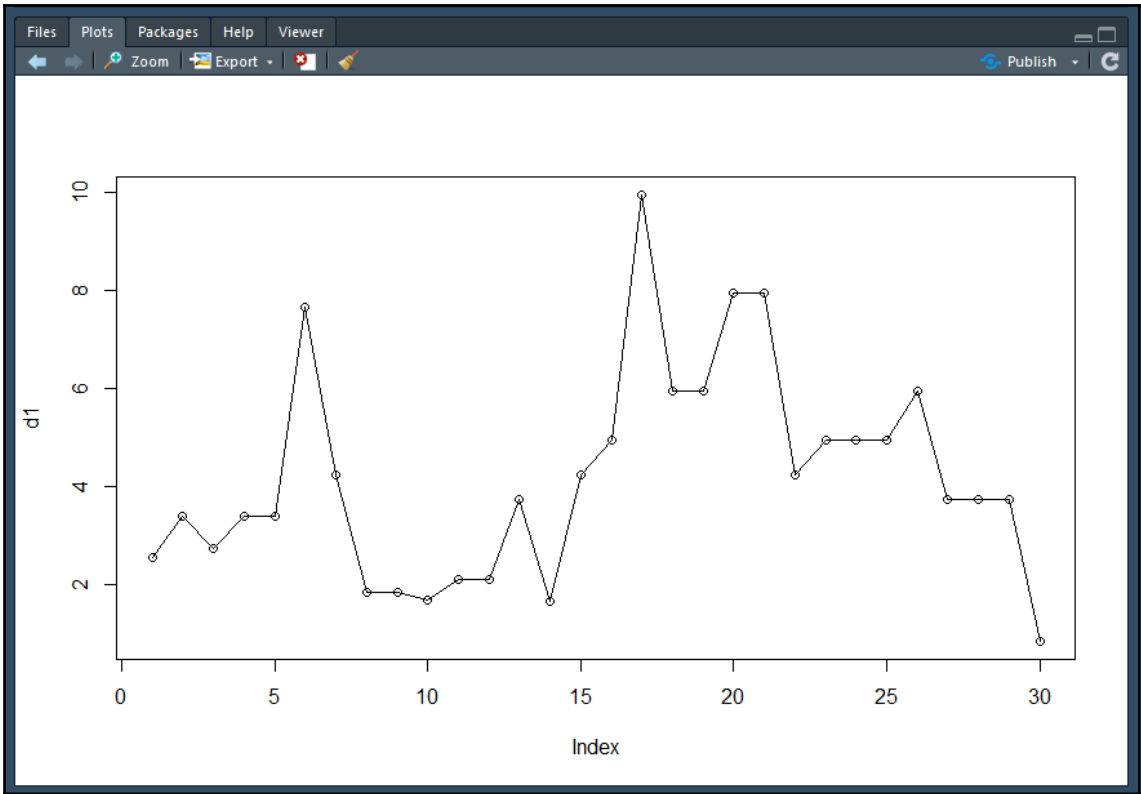


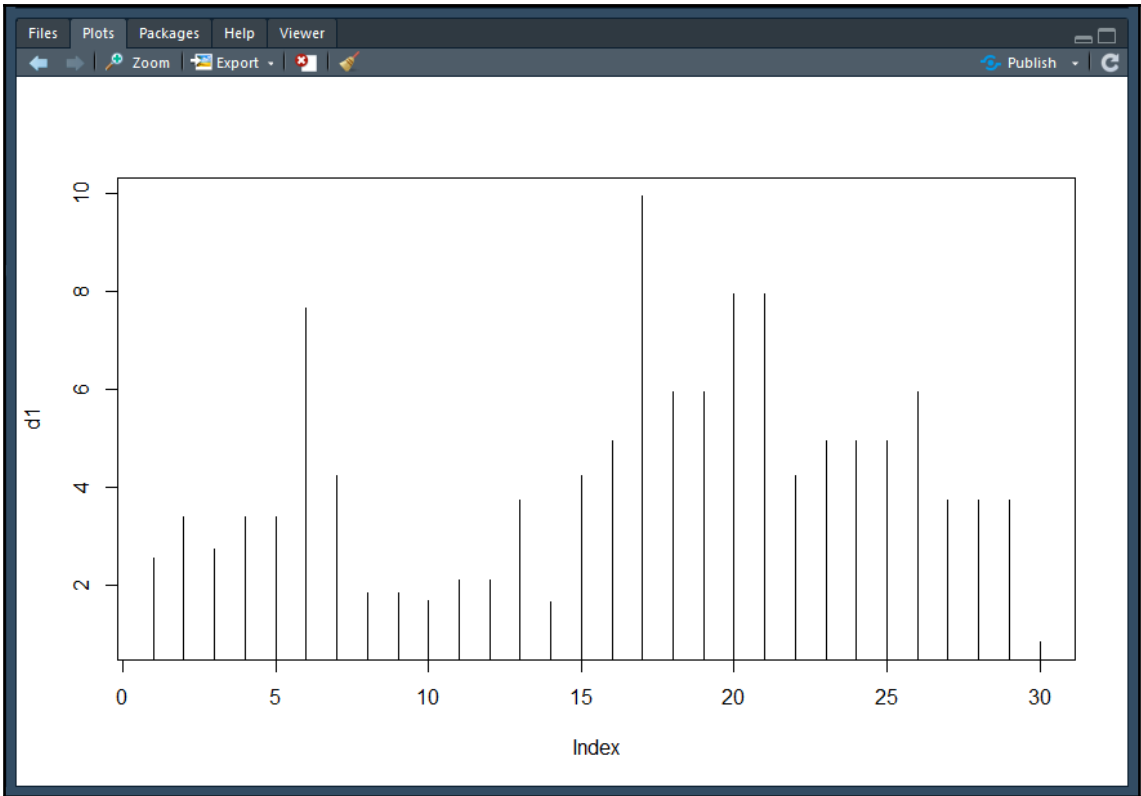


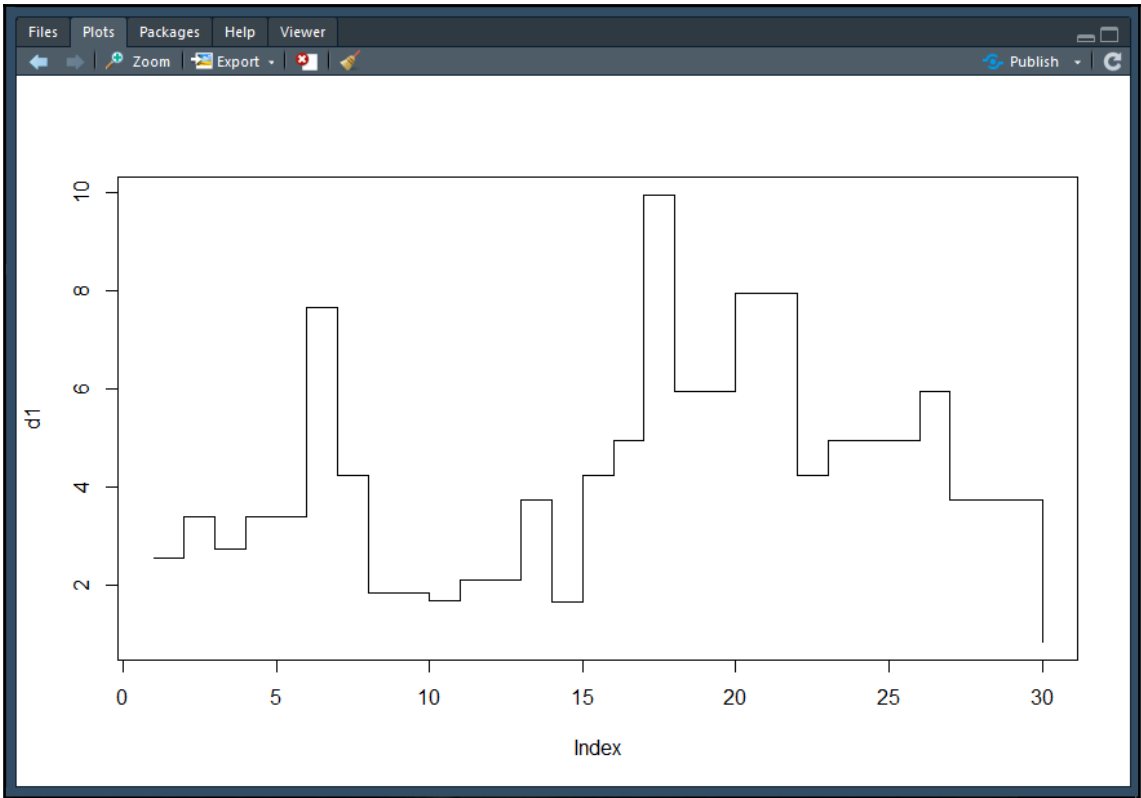


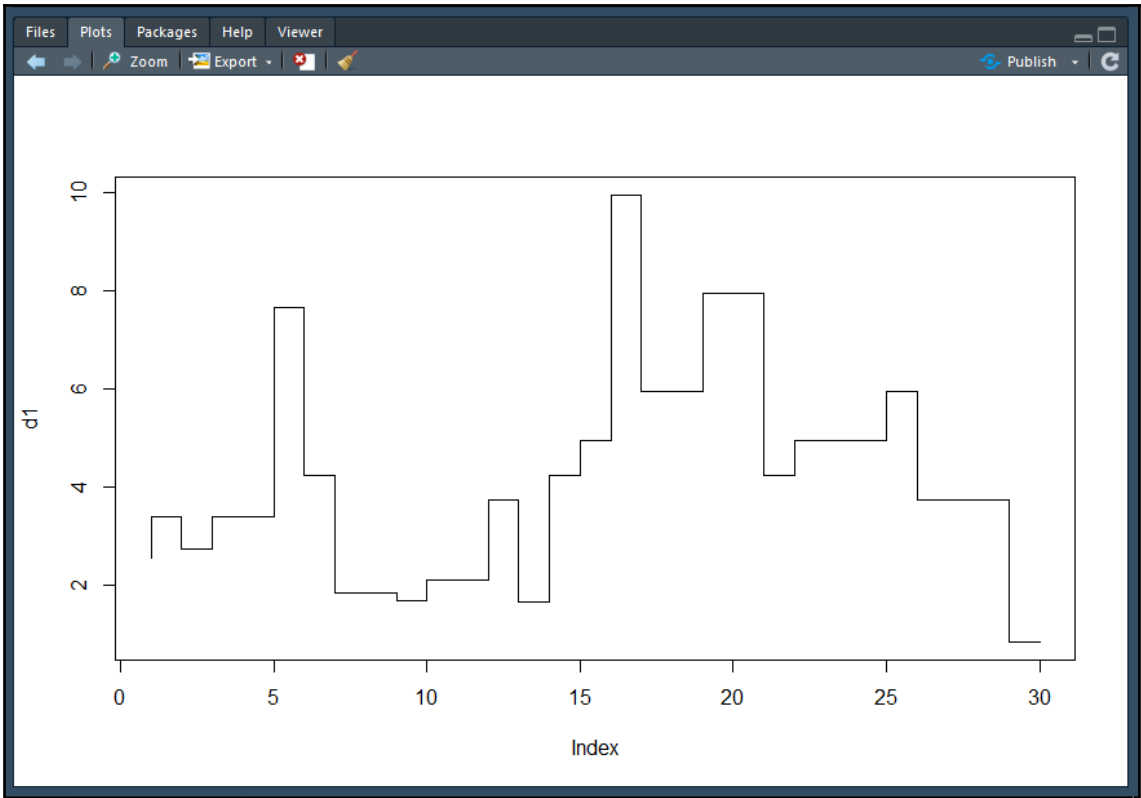


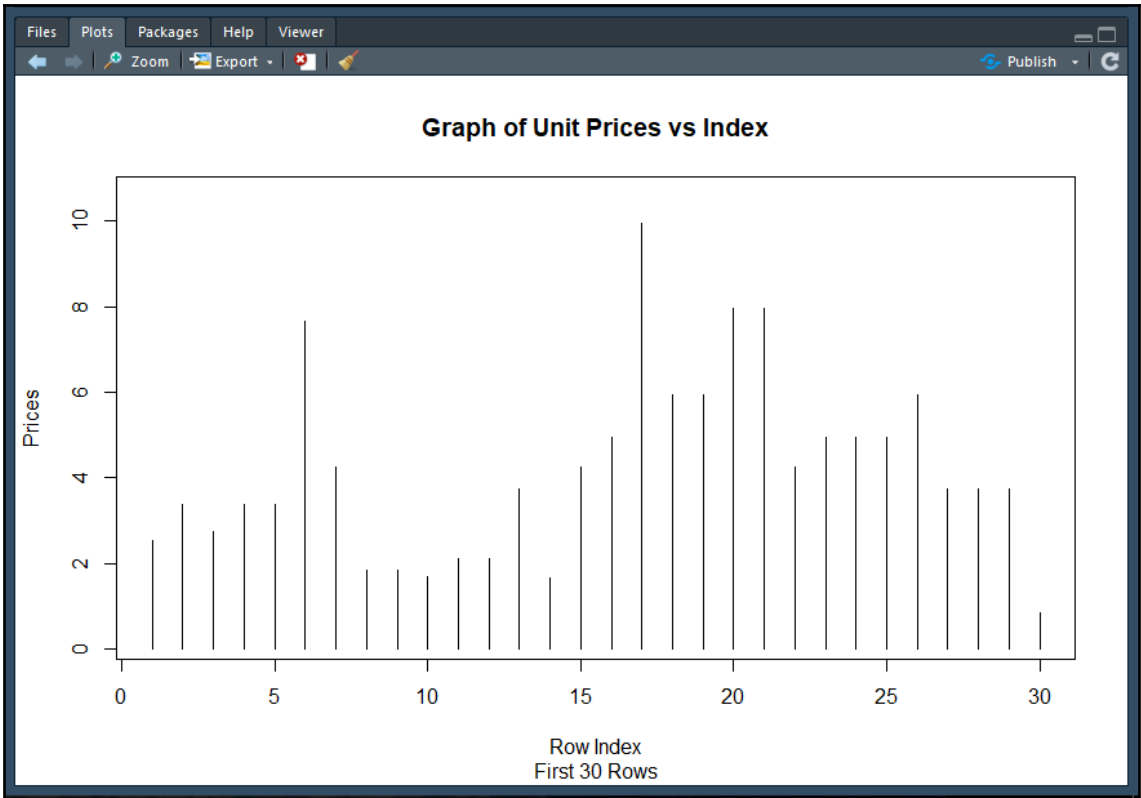












InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	
1	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	12/1/10 8:26	2.55	17850	United Kingdom
2	536365	71053	WHITE METAL LANTERN	6	12/1/10 8:26	3.39	17850	United Kingdom
3	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/1/10 8:26	2.75	17850	United Kingdom
4	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	12/1/10 8:26	3.39	17850	United Kingdom
5	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	12/1/10 8:26	3.39	17850	United Kingdom
6	536365	22752	SET 7 BABUSHKA NESTING BOXES	2	12/1/10 8:26	7.65	17850	United Kingdom
7	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	6	12/1/10 8:26	4.25	17850	United Kingdom
8	536366	22633	HAND WARMER UNION JACK	6	12/1/10 8:28	1.85	17850	United Kingdom
9	536366	22632	HAND WARMER RED POLKA DOT	6	12/1/10 8:28	1.85	17850	United Kingdom
10	536367	84879	ASSORTED COLOUR BIRD ORNAMENT	32	12/1/10 8:34	1.69	13047	United Kingdom



InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	
21	536368	22960	JAM MAKING SET WITH JARS	6	12/1/10 8:34	4.25	13047	United Kingdom
22	536368	22913	RED COAT RACK PARIS FASHION	3	12/1/10 8:34	4.95	13047	United Kingdom
23	536368	22912	YELLOW COAT RACK PARIS FASHION	3	12/1/10 8:34	4.95	13047	United Kingdom
24	536368	22914	BLUE COAT RACK PARIS FASHION	3	12/1/10 8:34	4.95	13047	United Kingdom
25	536369	21756	BATH BUILDING BLOCK WORD	3	12/1/10 8:35	5.95	13047	United Kingdom
26	536370	22728	ALARM CLOCK BAKELIKE PINK	24	12/1/10 8:45	3.75	12583	France
27	536370	22727	ALARM CLOCK BAKELIKE RED	24	12/1/10 8:45	3.75	12583	France
28	536370	22726	ALARM CLOCK BAKELIKE GREEN	12	12/1/10 8:45	3.75	12583	France
29	536370	21724	PANDA AND BUNNIES STICKER SHEET	12	12/1/10 8:45	0.85	12583	France
30	536370	21883	STARS GIFT TAPE	24	12/1/10 8:45	0.65	12583	France

InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	
1	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	12/1/10 8:26	2.55	17850	United Kingdom
2	536365	71053	WHITE METAL LANTERN	6	12/1/10 8:26	3.39	17850	United Kingdom
3	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/1/10 8:26	2.75	17850	United Kingdom
4	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	12/1/10 8:26	3.39	17850	United Kingdom
5	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	12/1/10 8:26	3.39	17850	United Kingdom
6	536365	22752	SET 7 BABUSHKA NESTING BOXES	2	12/1/10 8:26	7.65	17850	United Kingdom
7	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	6	12/1/10 8:26	4.25	17850	United Kingdom
8	536366	22633	HAND WARMER UNION JACK	6	12/1/10 8:28	1.85	17850	United Kingdom
9	536366	22632	HAND WARMER RED POLKA DOT	6	12/1/10 8:28	1.85	17850	United Kingdom
10	536367	84879	ASSORTED COLOUR BIRD ORNAMENT	32	12/1/10 8:34	1.69	13047	United Kingdom
21	536368	22960	JAM MAKING SET WITH JARS	6	12/1/10 8:34	4.25	13047	United Kingdom
22	536368	22913	RED COAT RACK PARIS FASHION	3	12/1/10 8:34	4.95	13047	United Kingdom
23	536368	22912	YELLOW COAT RACK PARIS FASHION	3	12/1/10 8:34	4.95	13047	United Kingdom
24	536368	22914	BLUE COAT RACK PARIS FASHION	3	12/1/10 8:34	4.95	13047	United Kingdom
25	536369	21756	BATH BUILDING BLOCK WORD	3	12/1/10 8:35	5.95	13047	United Kingdom
26	536370	22728	ALARM CLOCK BAKELIKE PINK	24	12/1/10 8:45	3.75	12583	France
27	536370	22727	ALARM CLOCK BAKELIKE RED	24	12/1/10 8:45	3.75	12583	France
28	536370	22726	ALARM CLOCK BAKELIKE GREEN	12	12/1/10 8:45	3.75	12583	France
29	536370	21724	PANDA AND BUNNIES STICKER SHEET	12	12/1/10 8:45	0.85	12583	France
30	536370	21883	STARS GIFT TAPE	24	12/1/10 8:45	0.65	12583	France

InvoiceNo	StockCode	Description
1	536365	85123A WHITE HANGING HEART T-LIGHT HOLDER
2	536365	71053 WHITE METAL LANTERN
3	536365	84406B CREAM CUPID HEARTS COAT HANGER
4	536365	84029G KNITTED UNION FLAG HOT WATER BOTTLE
5	536365	84029E RED WOOLLY HOTTIE WHITE HEART.
6	536365	22752 SET 7 BABUSHKA NESTING BOXES
7	536365	21730 GLASS STAR FROSTED T-LIGHT HOLDER
8	536366	22633 HAND WARMER UNION JACK
9	536366	22632 HAND WARMER RED POLKA DOT
10	536367	84879 ASSORTED COLOUR BIRD ORNAMENT
11	536367	22745 POPPYS PLAYHOUSE BEDROOM .6,12/1/10 8:34,2....

Showing 1 to 11 of 11 entries

StockCode	Description	Quantity
11	22745 POPPYS PLAYHOUSE BEDROOM .6,12/1/10 8:34,2....	6
12	22749 FELTCRAFT PRINCESS CHARLOTTE DOLL	8
13	22310 IVORY KNITTED MUG COSY	6
14	84969 BOX OF 6 ASSORTED COLOUR TEASPOONS	6
15	22623 BOX OF VINTAGE JIGSAW BLOCKS	3
16	22622 BOX OF VINTAGE ALPHABET BLOCKS	2
17	21754 HOME BUILDING BLOCK WORD	3
18	21755 LOVE BUILDING BLOCK WORD	3
19	21777 RECIPE BOX WITH METAL HEART	4
20	48187 DOORMAT NEW ENGLAND	4

Showing 1 to 10 of 10 entries

StockCode	Description	InvoiceNo	Quantity
1 22745	POPPYS PLAYHOUSE BEDROOM ,6,12/1/10 8:34,2....	536367	6

Showing 1 to 1 of 1 entries

StockCode	Description	InvoiceNo	Quantity
1 21730	GLASS STAR FROSTED T-LIGHT HOLDER	536365	NA
2 21754	HOME BUILDING BLOCK WORD	NA	3
3 21755	LOVE BUILDING BLOCK WORD	NA	3
4 21777	RECIPE BOX WTH METAL HEART	NA	4
5 22310	IVORY KNITTED MUG COSY	NA	6
6 22622	BOX OF VINTAGE ALPHABET BLOCKS	NA	2
7 22623	BOX OF VINTAGE JIGSAW BLOCKS	NA	3
8 22632	HAND WARMER RED POLKA DOT	536366	NA
9 22633	HAND WARMER UNION JACK	536366	NA
10 22745	POPPYS PLAYHOUSE BEDROOM ,6,12/1/10 8:34,2....	536367	6
11 22749	FELTCRAFT PRINCESS CHARLOTTE DOLL	NA	8
12 22752	SET 7 BABUSHKA NESTING BOXES	536365	NA
13 48187	DOORMAT NEW ENGLAND	NA	4
14 71053	WHITE METAL LANTERN	536365	NA
15 84029E	RED WOOLLY HOTTIE WHITE HEART.	536365	NA
16 84029G	KNITTED UNION FLAG HOT WATER BOTTLE	536365	NA
17 84406B	CREAM CUPID HEARTS COAT HANGER	536365	NA
18 84879	ASSORTED COLOUR BIRD ORNAMENT	536367	NA
19 84969	BOX OF 6 ASSORTED COLOUR TEASPOONS	NA	6
20 85123A	WHITE HANGING HEART T-LIGHT HOLDER	536365	NA

Showing 1 to 20 of 20 entries

StockCode	Description	InvoiceNo	Quantity
1 21730	GLASS STAR FROSTED T-LIGHT HOLDER	536365	NA
2 22632	HAND WARMER RED POLKA DOT	536366	NA
3 22633	HAND WARMER UNION JACK	536366	NA
4 22745	POPPYS PLAYHOUSE BEDROOM .6.12/1/10 8:34.2....	536367	6
5 22752	SET 7 BABUSHKA NESTING BOXES	536365	NA
6 71053	WHITE METAL LANTERN	536365	NA
7 84029E	RED WOOLLY HOTTIE WHITE HEART.	536365	NA
8 84029G	KNITTED UNION FLAG HOT WATER BOTTLE	536365	NA
9 84406B	CREAM CUPID HEARTS COAT HANGER	536365	NA
10 84879	ASSORTED COLOUR BIRD ORNAMENT	536367	NA
11 85123A	WHITE HANGING HEART T-LIGHT HOLDER	536365	NA

Showing 1 to 11 of 11 entries

StockCode	Description	InvoiceNo	Quantity
1 21754	HOME BUILDING BLOCK WORD	NA	3
2 21755	LOVE BUILDING BLOCK WORD	NA	3
3 21777	RECIPE BOX WITH METAL HEART	NA	4
4 22310	IVORY KNITTED MUG COSY	NA	6
5 22622	BOX OF VINTAGE ALPHABET BLOCKS	NA	2
6 22623	BOX OF VINTAGE JIGSAW BLOCKS	NA	3
7 22745	POPPYS PLAYHOUSE BEDROOM .6.12/1/10 8:34.2....	536367	6
8 22749	FELTCRAFT PRINCESS CHARLOTTE DOLL	NA	8
9 48187	DOORMAT NEW ENGLAND	NA	4
10 84969	BOX OF 6 ASSORTED COLOUR TEASPOONS	NA	6

Showing 1 to 10 of 10 entries

InvoiceNo	StockCode.x	Description.x	StockCode.y	Description.y	Quantity	
1	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	22745	POPPYS PLAYHOUSE BEDROOM ,6.12/1/10 8:34.2....	6
2	536365	71053	WHITE METAL LANTERN	22745	POPPYS PLAYHOUSE BEDROOM ,6.12/1/10 8:34.2....	6
3	536365	84406B	CREAM CUPID HEARTS COAT HANGER	22745	POPPYS PLAYHOUSE BEDROOM ,6.12/1/10 8:34.2....	6
4	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	22745	POPPYS PLAYHOUSE BEDROOM ,6.12/1/10 8:34.2....	6
5	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	22745	POPPYS PLAYHOUSE BEDROOM ,6.12/1/10 8:34.2....	6
6	536365	22752	SET 7 BABUSHKA NESTING BOXES	22745	POPPYS PLAYHOUSE BEDROOM ,6.12/1/10 8:34.2....	6
7	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	22745	POPPYS PLAYHOUSE BEDROOM ,6.12/1/10 8:34.2....	6
8	536366	22633	HAND WARMER UNION JACK	22745	POPPYS PLAYHOUSE BEDROOM ,6.12/1/10 8:34.2....	6
9	536366	22632	HAND WARMER RED POLKA DOT	22745	POPPYS PLAYHOUSE BEDROOM ,6.12/1/10 8:34.2....	6
10	536367	84879	ASSORTED COLOUR BIRD ORNAMENT	22745	POPPYS PLAYHOUSE BEDROOM ,6.12/1/10 8:34.2....	6
11	536367	22745	POPPYS PLAYHOUSE BEDROOM ,6.12/1/10 8:34.2....	22745	POPPYS PLAYHOUSE BEDROOM ,6.12/1/10 8:34.2....	6
12	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	22749	FELTCRAFT PRINCESS CHARLOTTE DOLL	8
13	536365	71053	WHITE METAL LANTERN	22749	FELTCRAFT PRINCESS CHARLOTTE DOLL	8
14	536365	84406B	CREAM CUPID HEARTS COAT HANGER	22749	FELTCRAFT PRINCESS CHARLOTTE DOLL	8
15	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	22749	FELTCRAFT PRINCESS CHARLOTTE DOLL	8
16	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	22749	FELTCRAFT PRINCESS CHARLOTTE DOLL	8
17	536365	22752	SET 7 BABUSHKA NESTING BOXES	22749	FELTCRAFT PRINCESS CHARLOTTE DOLL	8
18	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	22749	FELTCRAFT PRINCESS CHARLOTTE DOLL	8
19	536366	22633	HAND WARMER UNION JACK	22749	FELTCRAFT PRINCESS CHARLOTTE DOLL	8

Showing 1 to 20 of 110 entries

StockCode	InvoiceNo	Description.x	Description.y	Quantity
1	21730	536365	GLASS STAR FROSTED T-LIGHT HOLDER	NA
2	21754	NA	HOME BUILDING BLOCK WORD	3
3	21755	NA	LOVE BUILDING BLOCK WORD	3
4	21777	NA	RECIPE BOX WITH METAL HEART	4
5	22310	NA	IVORY KNITTED MUG COSY	6
6	22622	NA	BOX OF VINTAGE ALPHABET BLOCKS	2
7	22623	NA	BOX OF VINTAGE JIGSAW BLOCKS	3
8	22632	536366	HAND WARMER RED POLKA DOT	NA
9	22633	536366	HAND WARMER UNION JACK	NA
10	22745	536367	POPPYS PLAYHOUSE BEDROOM ,6.12/1/10 8:34.2....	6
11	22749	NA	FELTCRAFT PRINCESS CHARLOTTE DOLL	8
12	22752	536365	SET 7 BABUSHKA NESTING BOXES	NA
13	48187	NA	DOORMAT NEW ENGLAND	4
14	71053	536365	WHITE METAL LANTERN	NA
15	84029E	536365	RED WOOLLY HOTTIE WHITE HEART.	NA
16	84029G	536365	KNITTED UNION FLAG HOT WATER BOTTLE	NA
17	84406B	536365	CREAM CUPID HEARTS COAT HANGER	NA
18	84879	536367	ASSORTED COLOUR BIRD ORNAMENT	NA
19	84969	NA	BOX OF 6 ASSORTED COLOUR TEASPOONS	6
20	85123A	536365	WHITE HANGING HEART T-LIGHT HOLDER	NA

Showing 1 to 20 of 20 entries

```
Environment History Connections
To Console To Source
lines(d4$UnitPrice, type="p")
lines(d4$UnitPrice)
lines(d4$UnitPrice)
lines(d4$UnitPrice)
plot(d4$UnitPrice)
lines(d4$UnitPrice)
plot()
plot(d4$StockCode, d4$UnitPrice)
view(d4)
df = read.table(file=file.choose(), header=T, fill=T, sep=",")
view(df)
df = read.table(file=file.choose(), header=T, fill=T, sep=",")
history()
df = read.table(file=file.choose(), header=T, fill=T, sep=",")
d1 = df[0:10, c(1:8)]
savehistory(file=file.choose())
savehistory(file="log.Rhistory")
savehistory(file="logname.Rhistory")
```

```
> colSums(is.na(df))
InvoiceNo      StockCode Description      Quantity InvoiceDate      UnitPrice      CustomerID      Country
0              0              0              0              0              0              0              0
```

```
> colSums(is.na(d4))
stockCode Description      InvoiceNo      Quantity
0              0              9              10
```

InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
1	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	12/1/10 8:26	2.55	17850 United Kingdom
2	536365	71053	WHITE METAL LANTERN	6	12/1/10 8:26	3.39	17850 United Kingdom
3	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/1/10 8:26	2.75	17850 United Kingdom
4	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	12/1/10 8:26	3.39	17850 United Kingdom
5	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	12/1/10 8:26	3.39	17850 United Kingdom
6	536365	22752	SET 7 BABUSHKA NESTING BOXES	2	12/1/10 8:26	7.65	17850 United Kingdom
7	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	6	12/1/10 8:26	4.25	17850 United Kingdom
8	536366	22633	HAND WARMER UNION JACK	6	12/1/10 8:28	1.85	17850 United Kingdom
9	536366	22632	HAND WARMER RED POLKA DOT	6	12/1/10 8:28	1.85	17850 United Kingdom
10	536367	84879	ASSORTED COLOUR BIRD ORNAMENT	32	12/1/10 8:34	1.69	13047 United Kingdom
11	536367	22745	POPPY'S PLAYHOUSE BEDROOM	6	12/1/10 8:34	2.10	13047 United Kingdom
12	536367	22748	POPPY'S PLAYHOUSE KITCHEN	6	12/1/10 8:34	2.10	13047 United Kingdom
13	536367	22749	FELTCRAFT PRINCESS CHARLOTTE DOLL	8	12/1/10 8:34	3.75	13047 United Kingdom
14	536367	22310	IVORY KNITTED MUG COSY	6	12/1/10 8:34	1.65	13047 United Kingdom
15	536367	84969	BOX OF 6 ASSORTED COLOUR TEASPOONS	6	12/1/10 8:34	4.25	13047 United Kingdom
16	536367	22623	BOX OF VINTAGE JIGSAW BLOCKS	3	12/1/10 8:34	4.95	13047 United Kingdom
17	536367	22622	BOX OF VINTAGE ALPHABET BLOCKS	2	12/1/10 8:34	9.95	13047 United Kingdom
18	536367	21754	HOME BUILDING BLOCK WORD	3	12/1/10 8:34	5.95	13047 United Kingdom
19	536367	21755	LOVE BUILDING BLOCK WORD	3	12/1/10 8:34	5.95	13047 United Kingdom
20	536367	21777	RECIPE BOX WITH METAL HEART	4	12/1/10 8:34	7.95	13047 United Kingdom

Showing 1 to 20 of 30 entries

```

Console Terminal
~/
> d1 = df[0:30, c(1:8)]
>

```

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
1	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	12/1/10 8:26	2.55	17850	United Kingdom
2	536365	71053	Miscellaneous	6	12/1/10 8:26	3.39	17850	United Kingdom
3	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/1/10 8:26	2.75	17850	United Kingdom
4	536365	84029G	Miscellaneous	6	12/1/10 8:26	3.39	17850	United Kingdom
5	536365	84029E	Miscellaneous	6	12/1/10 8:26	3.39	17850	United Kingdom
6	536365	22752	Miscellaneous	2	12/1/10 8:26	7.65	17850	United Kingdom
7	536365	21730	Miscellaneous	6	12/1/10 8:26	4.25	17850	United Kingdom
8	536366	22633	HAND WARMER UNION JACK	6	12/1/10 8:28	1.85	17850	United Kingdom
9	536366	22632	HAND WARMER RED POLKA DOT	6	12/1/10 8:28	1.85	17850	United Kingdom
10	536367	84879	ASSORTED COLOUR BIRD ORNAMENT	32	12/1/10 8:34	1.69	13047	United Kingdom
11	536367	22745	POPPY'S PLAYHOUSE BEDROOM	6	12/1/10 8:34	2.10	13047	United Kingdom
12	536367	22748	POPPY'S PLAYHOUSE KITCHEN	6	12/1/10 8:34	2.10	13047	United Kingdom
13	536367	22749	Miscellaneous	8	12/1/10 8:34	3.75	13047	United Kingdom
14	536367	22310	IVORY KNITTED MUG COSY	6	12/1/10 8:34	1.65	13047	United Kingdom
15	536367	84969	Miscellaneous	6	12/1/10 8:34	4.25	13047	United Kingdom
16	536367	22623	Miscellaneous	3	12/1/10 8:34	4.95	13047	United Kingdom
17	536367	22622	Miscellaneous	2	12/1/10 8:34	9.95	13047	United Kingdom
18	536367	21754	Miscellaneous	3	12/1/10 8:34	5.95	13047	United Kingdom
19	536367	21755	Miscellaneous	3	12/1/10 8:34	5.95	13047	United Kingdom
20	536367	21777	Miscellaneous	4	12/1/10 8:34	7.95	13047	United Kingdom

Showing 1 to 20 of 30 entries

```

Console Terminal
~/
> d1 = df[0:30, c(1:8)]
> d1[d1$UnitPrice > 3, "Description"] <- "Miscellaneous"
>

```

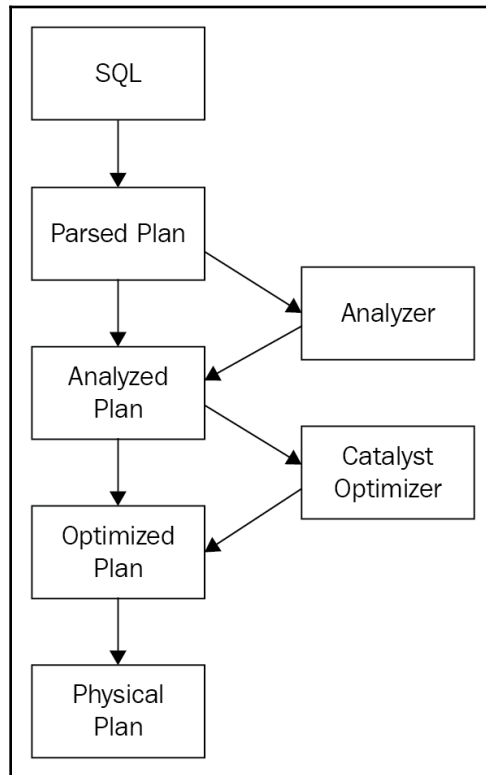
	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
1	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	12/1/10 8:26	2.55	17850	United States
2	536365	71053	Miscellaneous	6	12/1/10 8:26	3.39	17850	United States
3	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/1/10 8:26	2.75	17850	United States
4	536365	84029G	Miscellaneous	6	12/1/10 8:26	3.39	17850	United States
5	536365	84029E	Miscellaneous	6	12/1/10 8:26	3.39	17850	United States
6	536365	22752	Miscellaneous	2	12/1/10 8:26	7.65	17850	United States
7	536365	21730	Miscellaneous	6	12/1/10 8:26	4.25	17850	United States
8	536366	22633	HAND WARMER UNION JACK	6	12/1/10 8:28	1.85	17850	United Kingdom
9	536366	22632	HAND WARMER RED POLKA DOT	6	12/1/10 8:28	1.85	17850	United Kingdom
10	536367	84879	ASSORTED COLOUR BIRD ORNAMENT	32	12/1/10 8:34	1.69	13047	United Kingdom
11	536367	22745	POPPY'S PLAYHOUSE BEDROOM	6	12/1/10 8:34	2.10	13047	United Kingdom
12	536367	22748	POPPY'S PLAYHOUSE KITCHEN	6	12/1/10 8:34	2.10	13047	United Kingdom
13	536367	22749	Miscellaneous	8	12/1/10 8:34	3.75	13047	United Kingdom
14	536367	22310	IVORY KNITTED MUG COSY	6	12/1/10 8:34	1.65	13047	United Kingdom
15	536367	84969	Miscellaneous	6	12/1/10 8:34	4.25	13047	United Kingdom
16	536367	22623	Miscellaneous	3	12/1/10 8:34	4.95	13047	United Kingdom
17	536367	22622	Miscellaneous	2	12/1/10 8:34	9.95	13047	United Kingdom
18	536367	21754	Miscellaneous	3	12/1/10 8:34	5.95	13047	United Kingdom
19	536367	21755	Miscellaneous	3	12/1/10 8:34	5.95	13047	United Kingdom
20	536367	21777	Miscellaneous	4	12/1/10 8:34	7.95	13047	United Kingdom

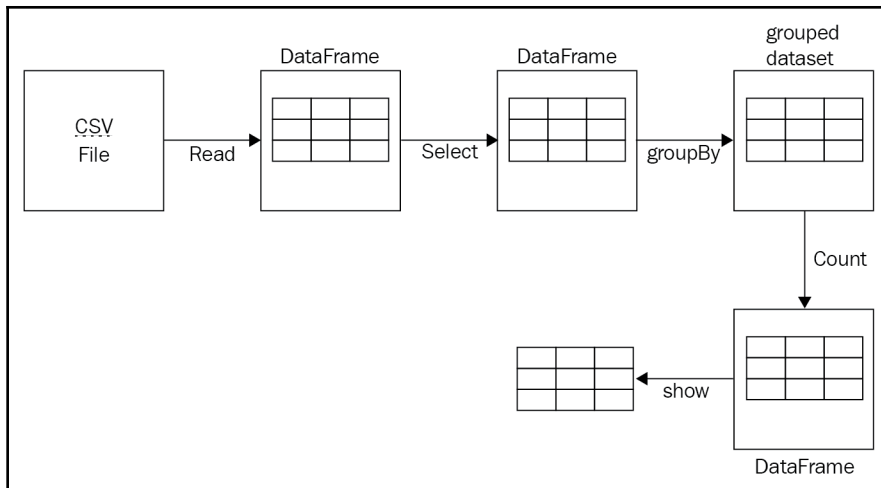
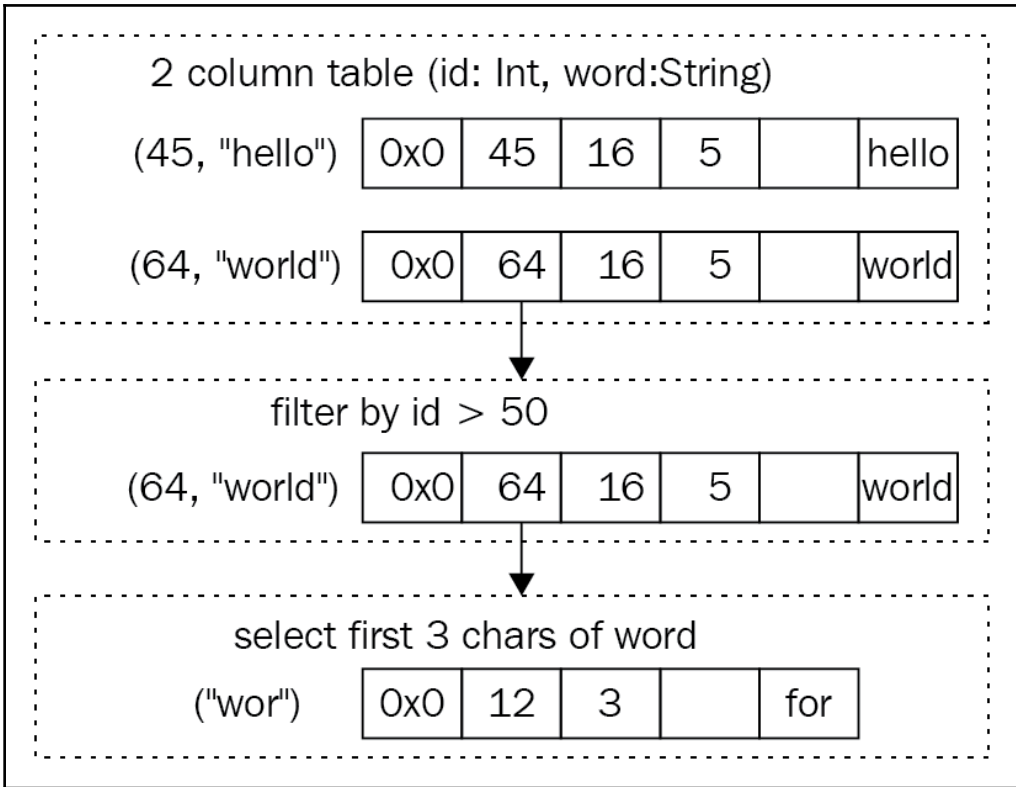
Showing 1 to 20 of 30 entries

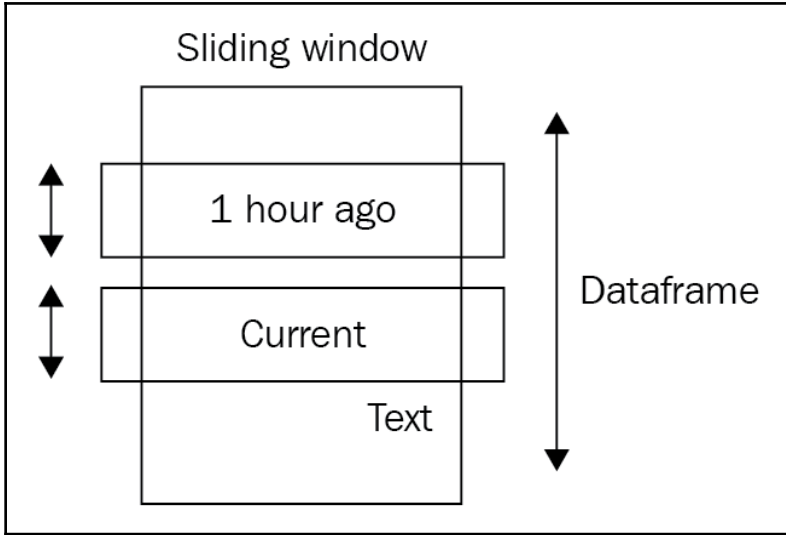


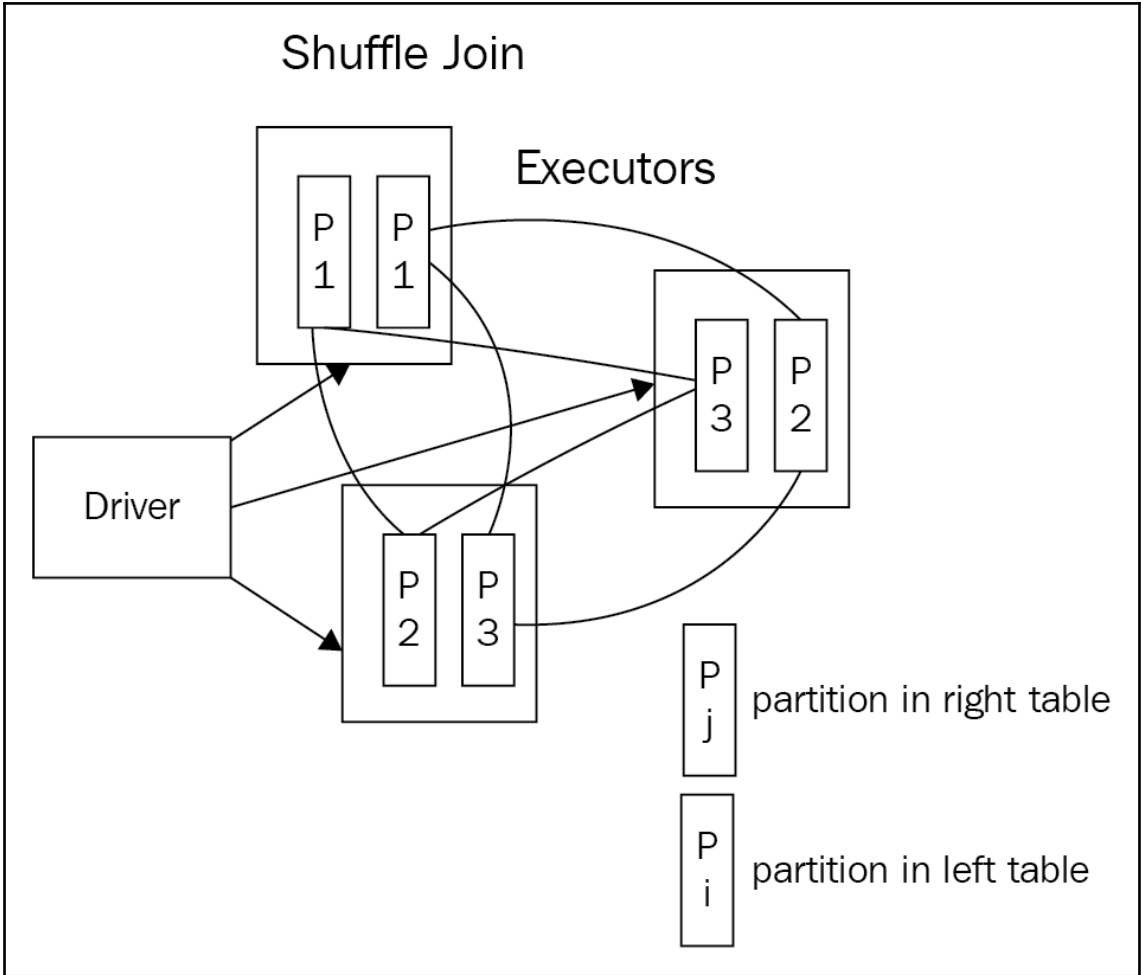
---

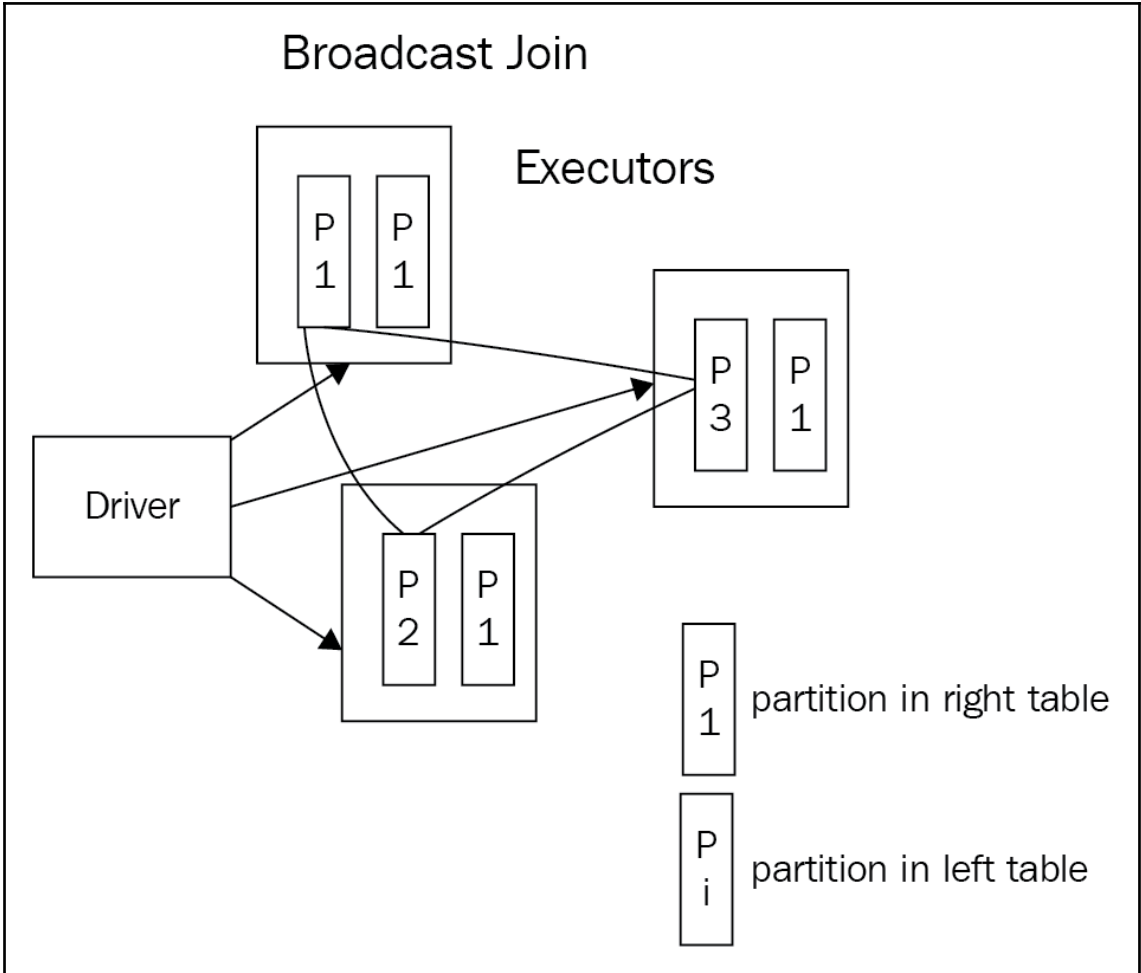
## Chapter 6: Batch Analytics with Apache Spark

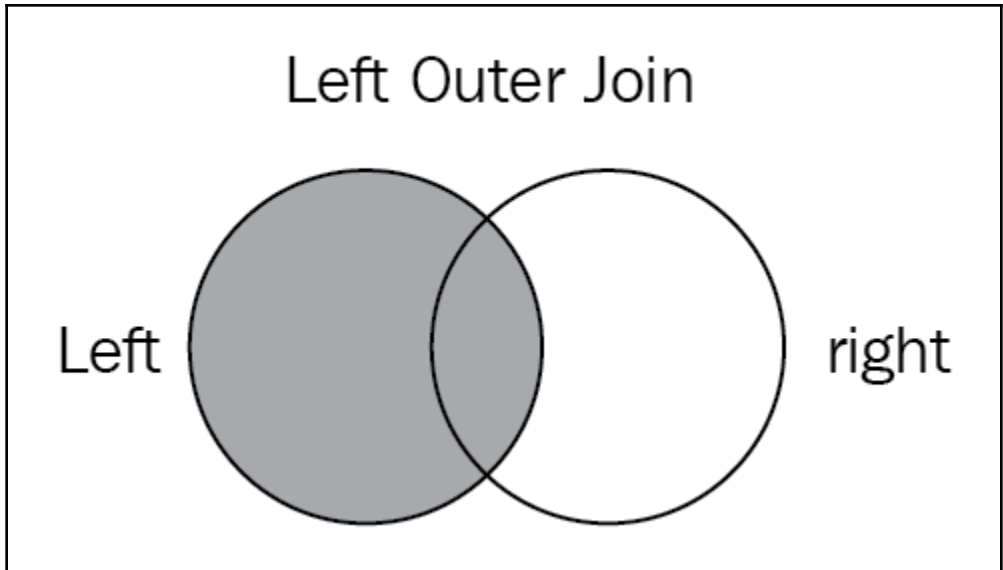
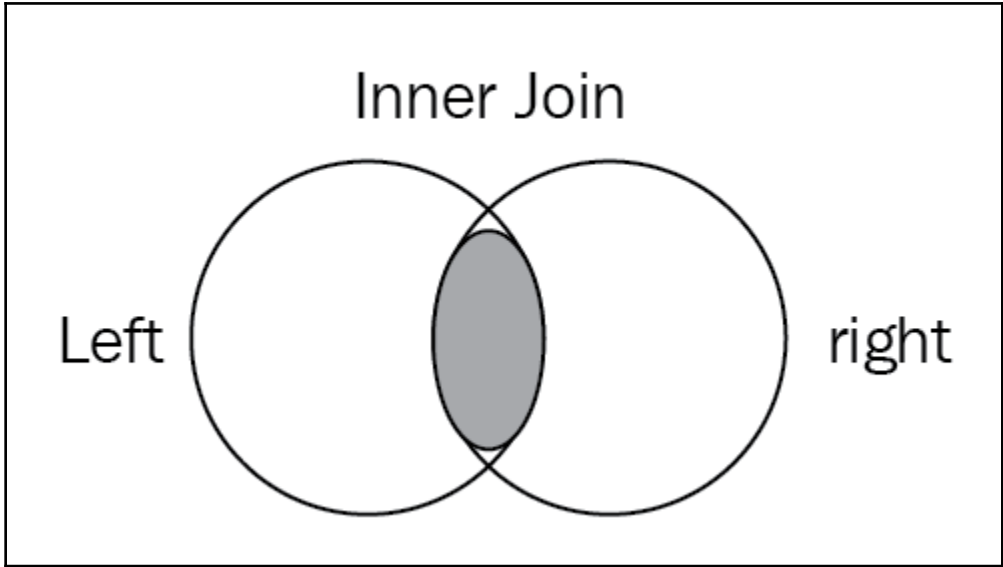


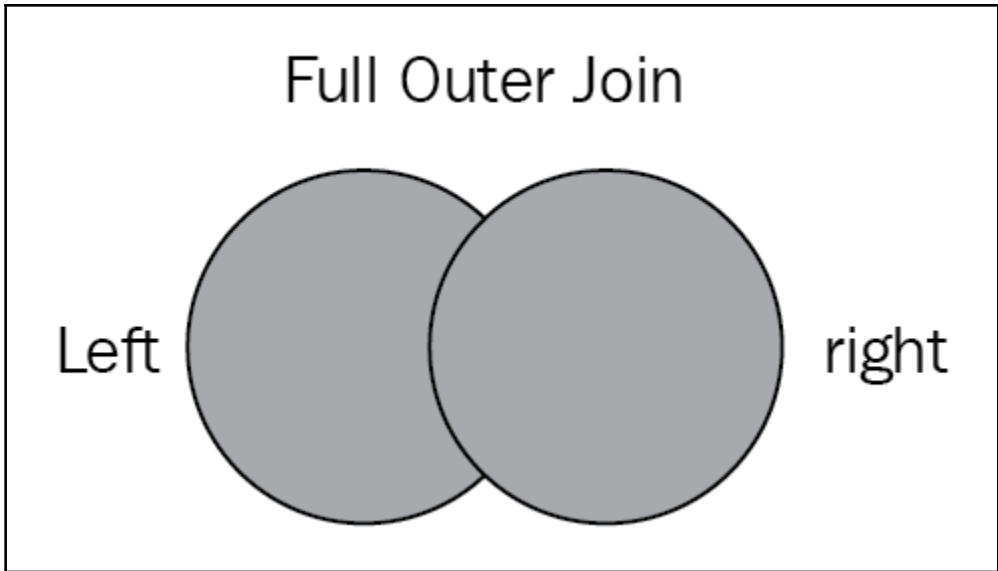
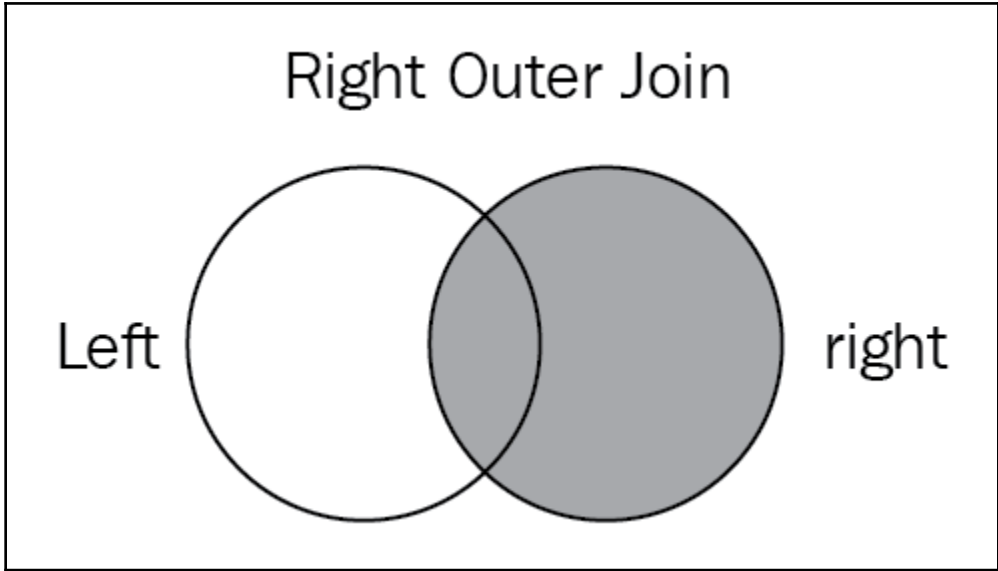


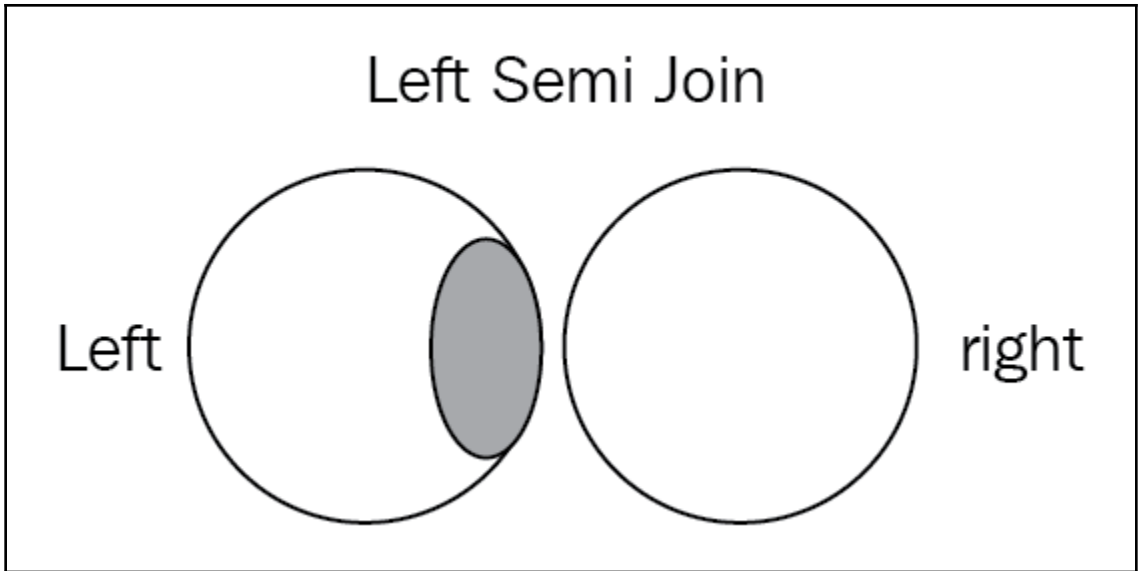
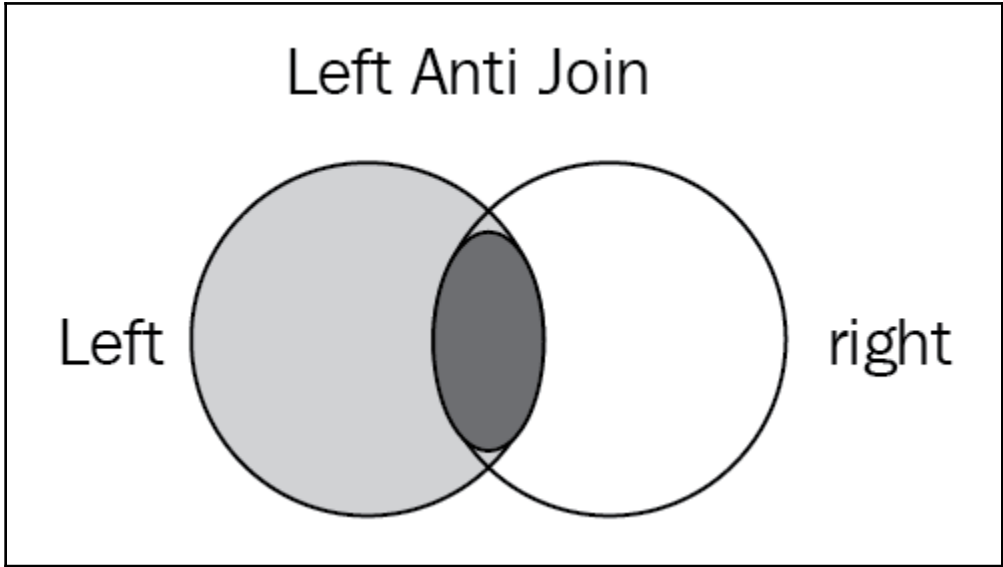




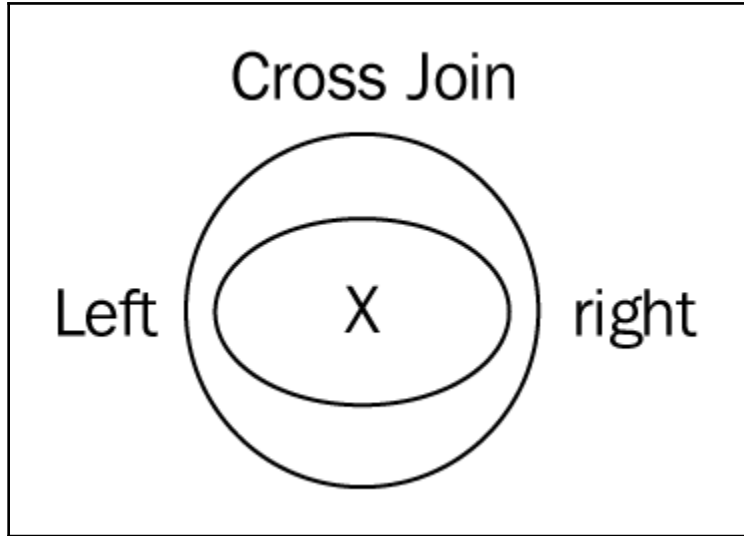






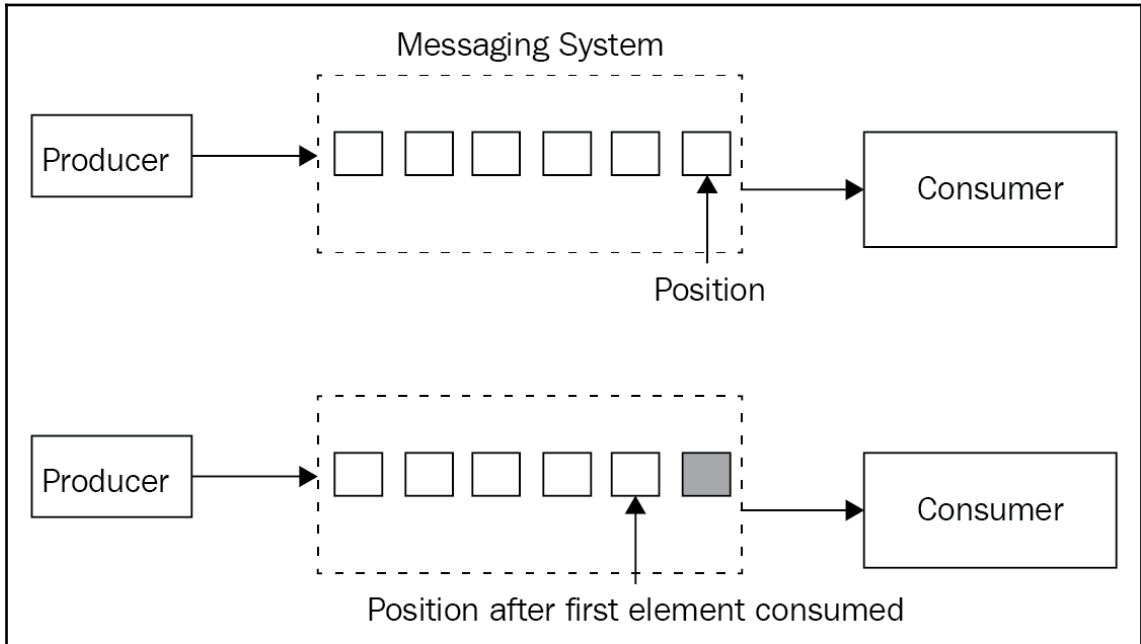


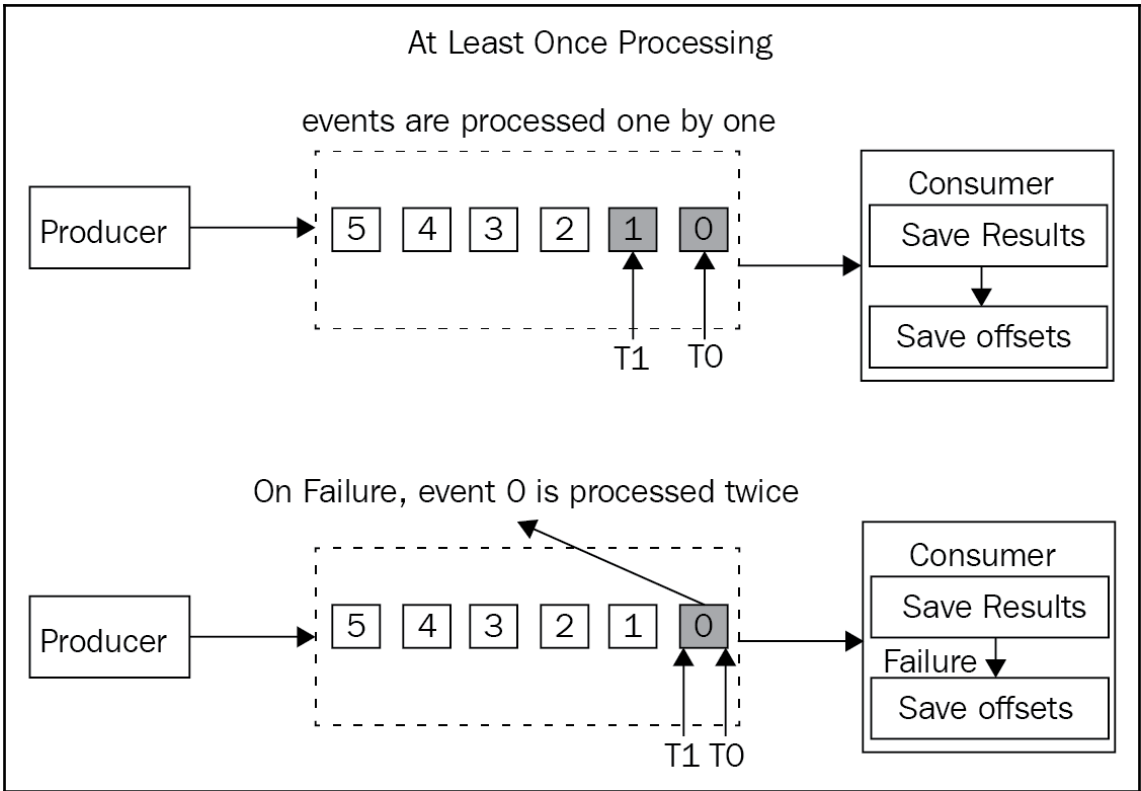


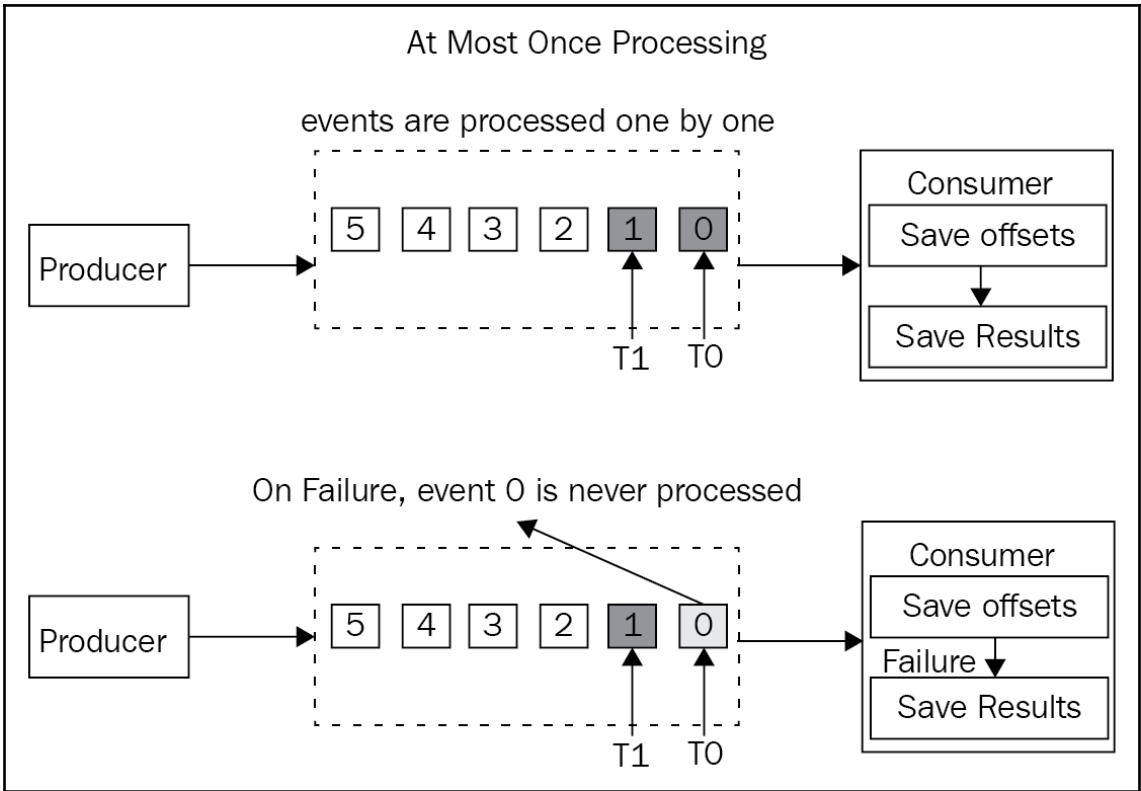


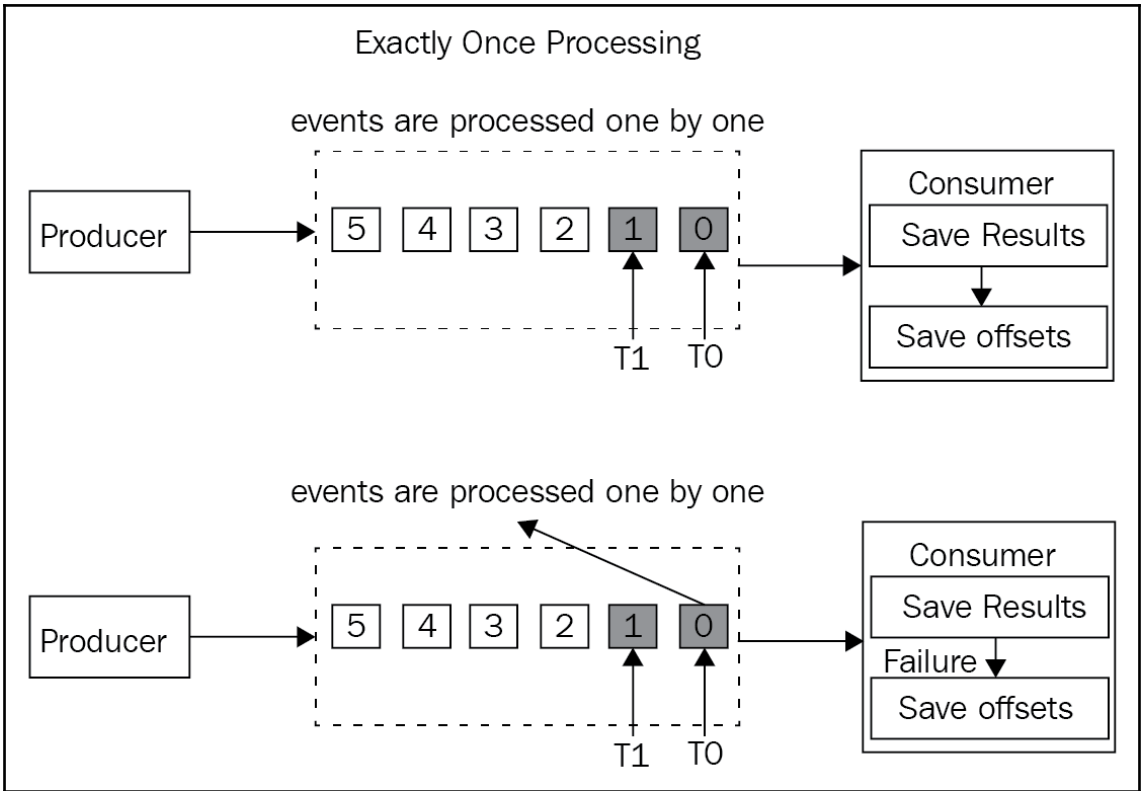
---

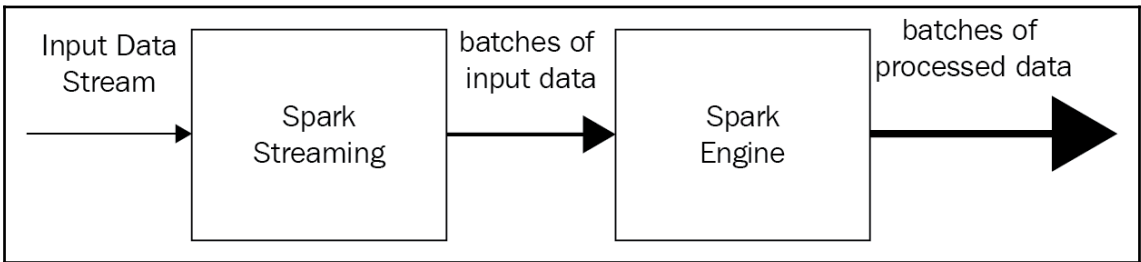
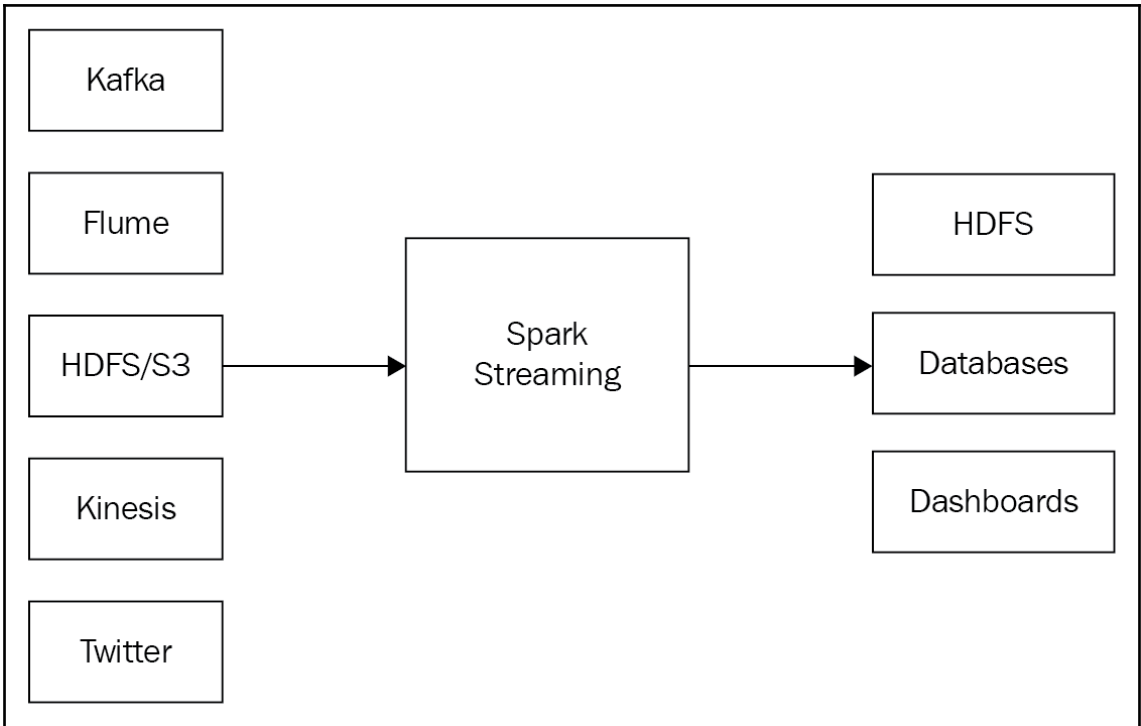
# Chapter 7: Real-Time Analytics with Apache Spark



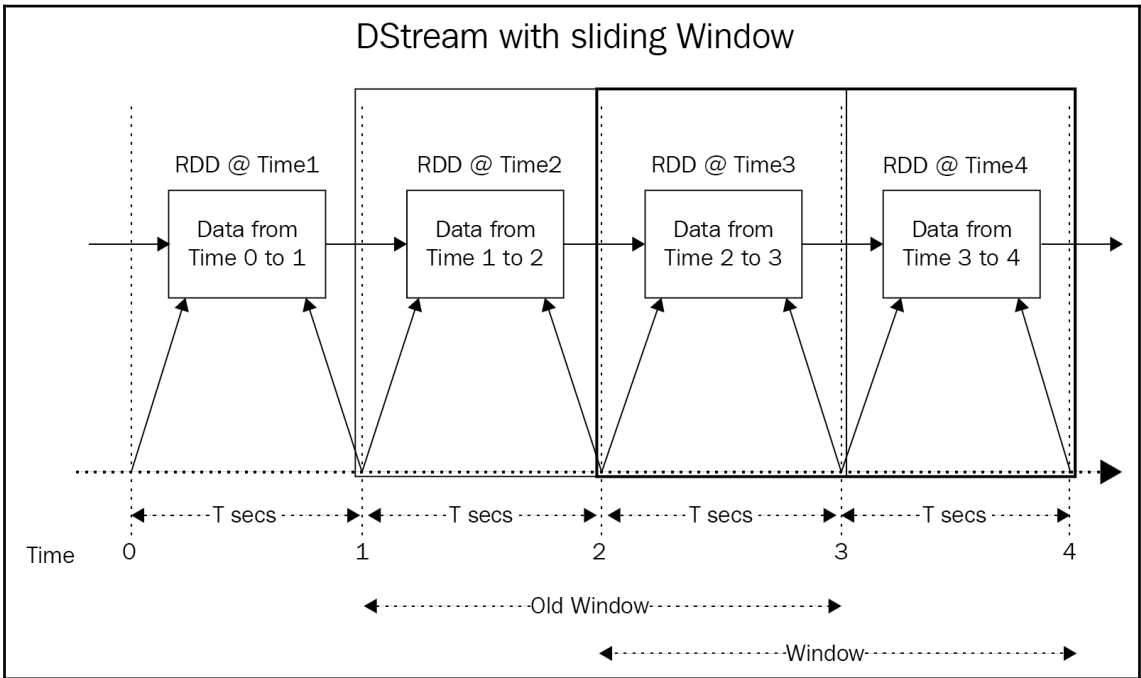
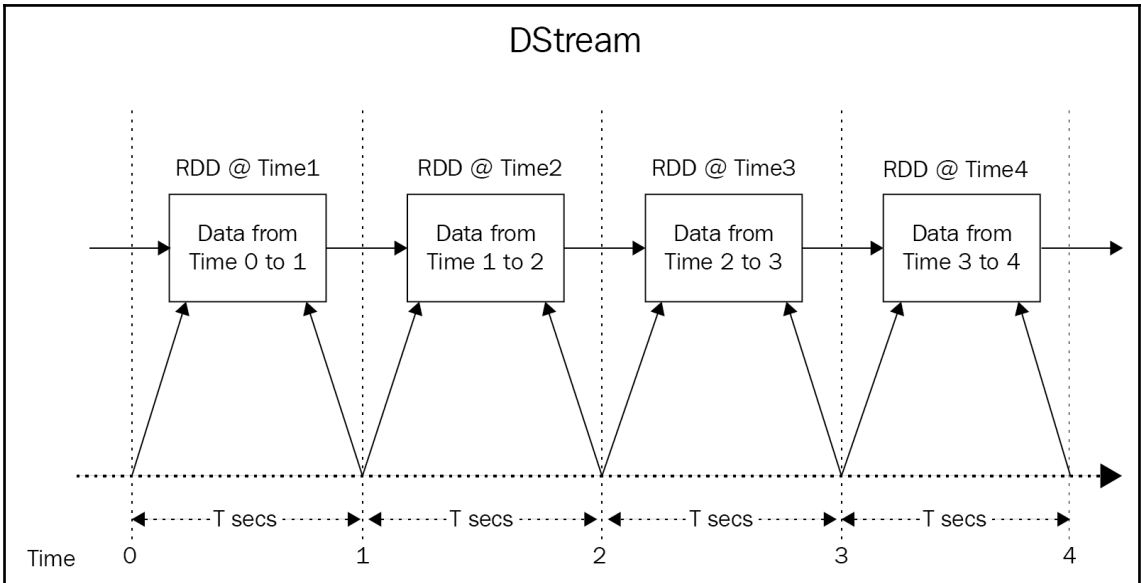








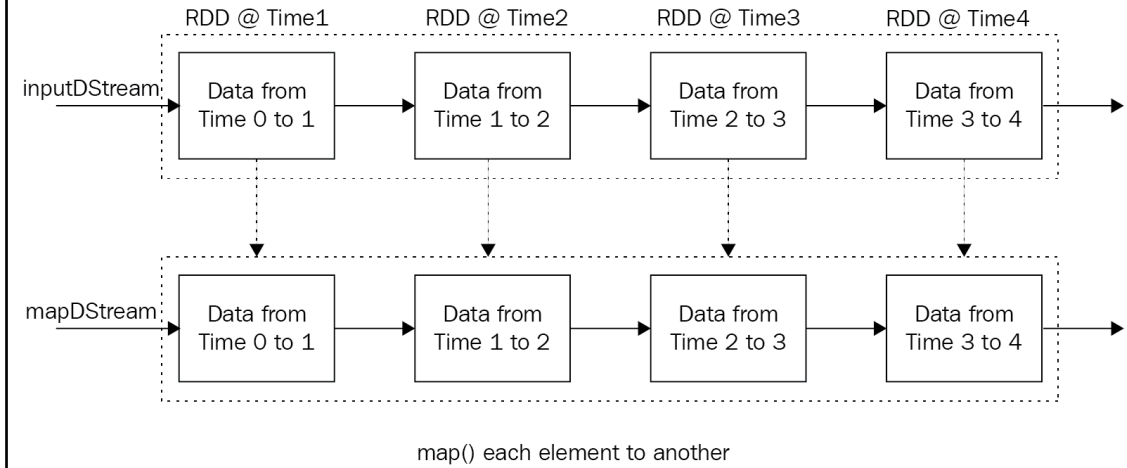


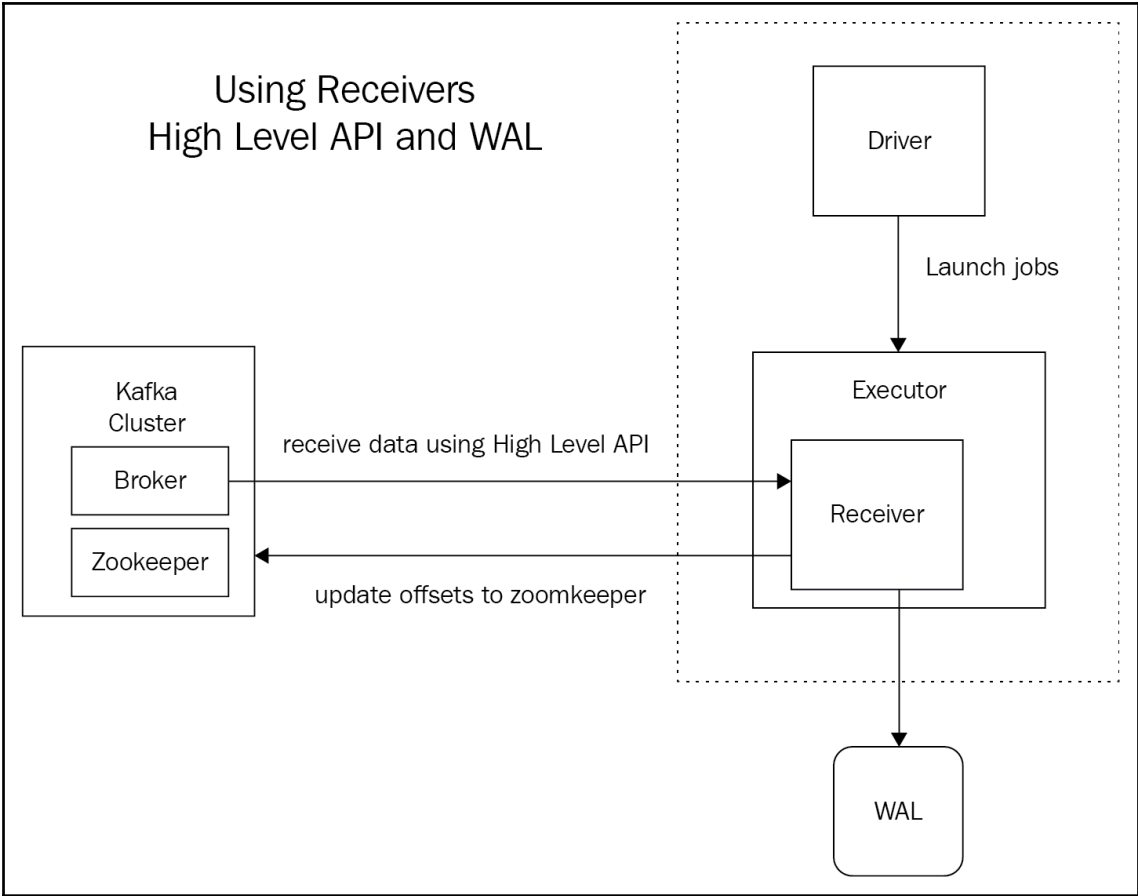


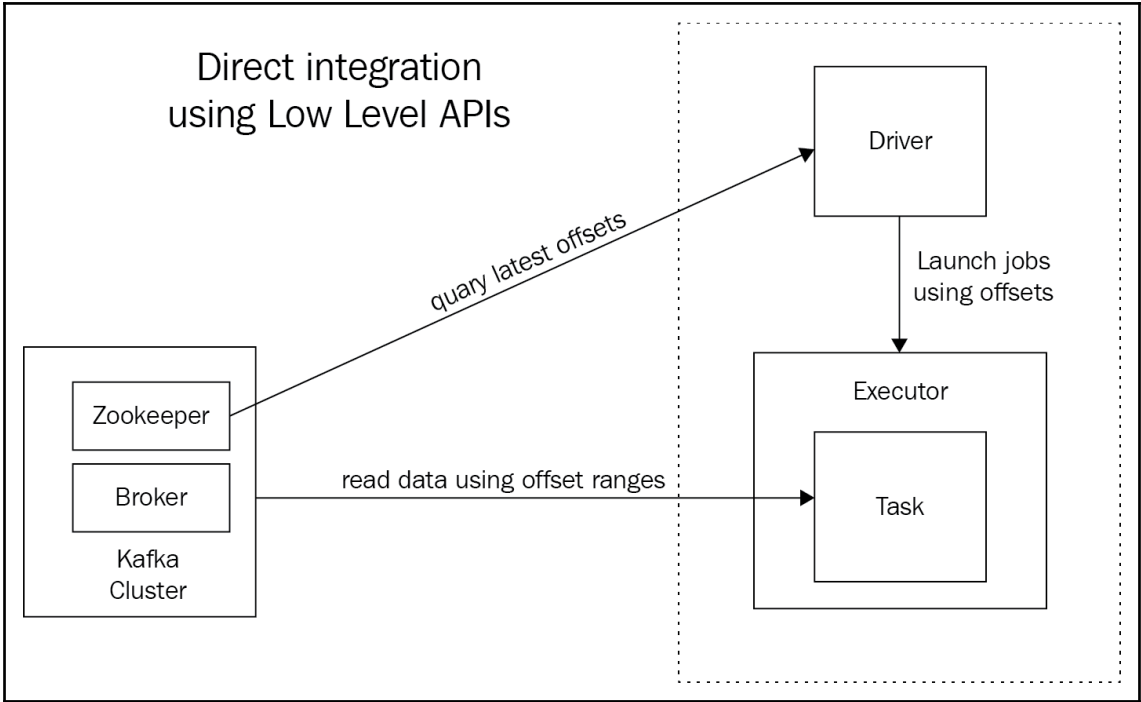


---

## Stateless transformations

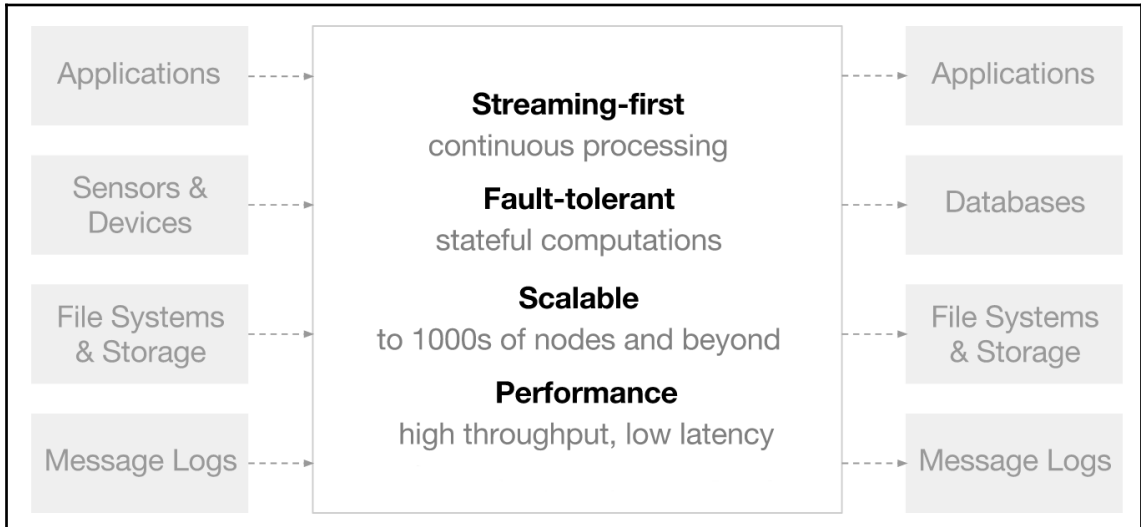






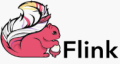
---

# Chapter 8: Batch Analytics with Apache Flink



<b>Deploy</b>	<b>Local</b> Single JVM		<b>Cluster</b> Standalone, YARN		<b>Cloud</b> GCE, EC2	
	<b>Runtime</b> Distributed Streaming Dataflow					
	<b>DataStream API</b> Stream Processing			<b>DataSet API</b> Batch Processing		
	<b>CEP</b> Event Processing	<b>Table</b> Relational		<b>FlinkML</b> Machine Learning	<b>Gelly</b> Graph Processing	<b>Table</b> Relational
<b>APIs &amp; Libraries</b>						

🏠
🔒 <https://flink.apache.org/downloads.html>
📄 ⋮ 🍷 ☆



[Download Flink](#)

- Home
- [Introduction to Flink](#)
- [Flink Use Cases](#)
- [Powered by Flink](#)
- [Ecosystem](#)
- [Community & Project Info](#)
- [How to Contribute](#)
- [Flink Blog](#)

---

- [Documentation ▾](#)
- [Quickstart ↗](#)
- [Flink on GitHub ↗](#)

---

- [Project FAQ](#)
- [@ApacheFlink ↗](#)

## Downloads

[Latest stable release \(v1.4.2\)](#)

- [Binaries](#)
- [Source](#)

[Verifying Hashes and Signatures](#)

[Maven Dependencies](#)

[Update Policy for old releases](#)

[All releases](#)

### Latest stable release (v1.4.2)

Apache Flink® 1.4.2 is our latest stable release.

An Apache Hadoop installation is [not required](#) to use Flink. If you plan to run Flink in YARN or process data stored in HDFS then select the version matching your installed Hadoop version.

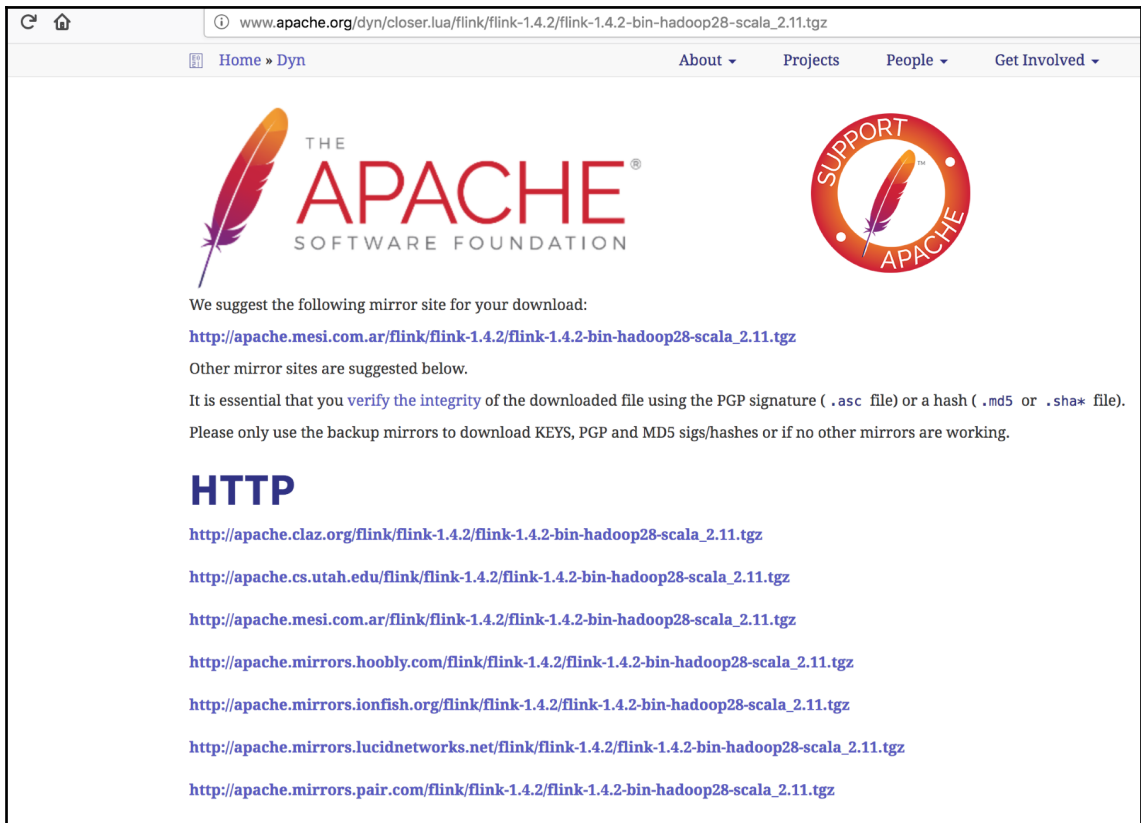
The binary releases marked with a Hadoop version come bundled with binaries for that Hadoop version, the binary release without bundled Hadoop can be used without Hadoop or with a Hadoop version that is installed in the environment, i.e. this version can pick up a Hadoop version from the classpath.

#### Binaries

	Scala 2.11
Without bundled Hadoop®	<a href="#">Download (asc, md5)</a>
Hadoop® 2.8	<a href="#">Download (asc, md5)</a>
Hadoop® 2.7	<a href="#">Download (asc, md5)</a>
Hadoop® 2.6	<a href="#">Download (asc, md5)</a>
Hadoop® 2.4	<a href="#">Download (asc, md5)</a>


#### Source

---



www.apache.org/dyn/closer.lua/flink/flink-1.4.2/flink-1.4.2-bin-hadoop28-scala\_2.11.tgz

Home » Dyn About Projects People Get Involved

We suggest the following mirror site for your download:  
[http://apache.mesi.com.ar/flink/flink-1.4.2/flink-1.4.2-bin-hadoop28-scala\\_2.11.tgz](http://apache.mesi.com.ar/flink/flink-1.4.2/flink-1.4.2-bin-hadoop28-scala_2.11.tgz)

Other mirror sites are suggested below.

It is essential that you [verify the integrity](#) of the downloaded file using the PGP signature ( .asc file) or a hash ( .md5 or .sha\* file). Please only use the backup mirrors to download KEYS, PGP and MD5 sigs/hashes or if no other mirrors are working.

## HTTP

[http://apache.claz.org/flink/flink-1.4.2/flink-1.4.2-bin-hadoop28-scala\\_2.11.tgz](http://apache.claz.org/flink/flink-1.4.2/flink-1.4.2-bin-hadoop28-scala_2.11.tgz)

[http://apache.cs.utah.edu/flink/flink-1.4.2/flink-1.4.2-bin-hadoop28-scala\\_2.11.tgz](http://apache.cs.utah.edu/flink/flink-1.4.2/flink-1.4.2-bin-hadoop28-scala_2.11.tgz)

[http://apache.mesi.com.ar/flink/flink-1.4.2/flink-1.4.2-bin-hadoop28-scala\\_2.11.tgz](http://apache.mesi.com.ar/flink/flink-1.4.2/flink-1.4.2-bin-hadoop28-scala_2.11.tgz)

[http://apache.mirrors.hoobly.com/flink/flink-1.4.2/flink-1.4.2-bin-hadoop28-scala\\_2.11.tgz](http://apache.mirrors.hoobly.com/flink/flink-1.4.2/flink-1.4.2-bin-hadoop28-scala_2.11.tgz)

[http://apache.mirrors.ionfish.org/flink/flink-1.4.2/flink-1.4.2-bin-hadoop28-scala\\_2.11.tgz](http://apache.mirrors.ionfish.org/flink/flink-1.4.2/flink-1.4.2-bin-hadoop28-scala_2.11.tgz)

[http://apache.mirrors.lucidnetworks.net/flink/flink-1.4.2/flink-1.4.2-bin-hadoop28-scala\\_2.11.tgz](http://apache.mirrors.lucidnetworks.net/flink/flink-1.4.2/flink-1.4.2-bin-hadoop28-scala_2.11.tgz)

[http://apache.mirrors.pair.com/flink/flink-1.4.2/flink-1.4.2-bin-hadoop28-scala\\_2.11.tgz](http://apache.mirrors.pair.com/flink/flink-1.4.2/flink-1.4.2-bin-hadoop28-scala_2.11.tgz)

```
Moogie:~ sridharalla$ tar -xvzf flink-1.4.2-bin-hadoop28-scala_2.11.tgz
x flink-1.4.2/
x flink-1.4.2/lib/
x flink-1.4.2/lib/flink-dist_2.11-1.4.2.jar
```

```
Moogie:flink-1.4.2 sridharalla$ ls
LICENSE      README.txt   conf         lib          opt          tools
NOTICE      bin         examples    log         resources
```

localhost:8081/#/overview

Overview Version: 1.4.2 Commit: 04e4c85

Task Managers: 1  
Task Slots: 1  
Available Task Slots: 1

Total Jobs	
Running	0
Finished	0
Canceled	0
Failed	0

**Running Jobs**

Start Time	End Time	Duration	Job Name	Job ID	Tasks	Status

**Completed Jobs**

Start Time	End Time	Duration	Job Name	Job ID	Tasks	Status

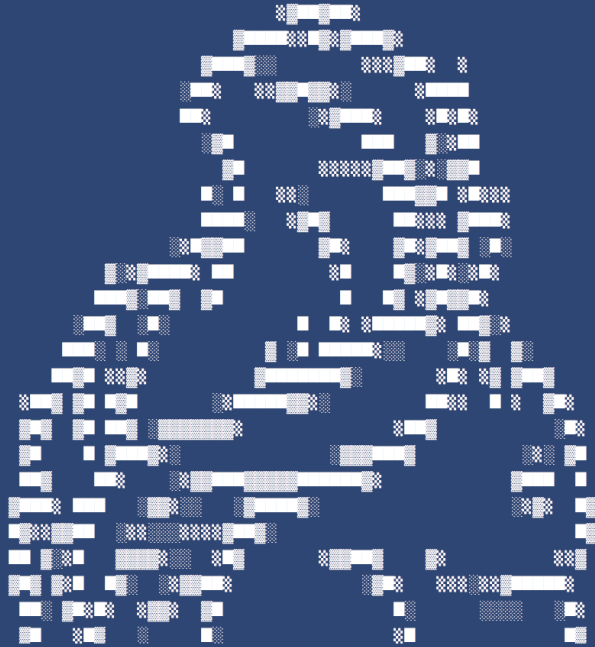
```
Moogie:flink-1.4.2 sridharalla$ tail log/flink-*-jobmanager-*.log
2018-04-27 13:38:50,400 INFO org.apache.flink.runtime.jobmanager.JobManager - Starting JobManager actor
2018-04-27 13:38:50,405 INFO org.apache.flink.runtime.blob.BlobServer - Created BLOB server storage directory /var/fo
rs/vv/5yxz8w1j3js11f94tpj4n2_h0000gn/T/blobStore-02f217cf-9d4d-4ce9-9351-769b543ddbce
2018-04-27 13:38:50,408 INFO org.apache.flink.runtime.blob.BlobServer - Started BLOB server at 0.0.0.0:57213 - max concu
rrent requests: 50 - max backlog: 1000
2018-04-27 13:38:50,492 INFO org.apache.flink.runtime.jobmanager.JobManager - Starting JobManager at akka.tcp://flink@localhos
t:6123/user/jobmanager,
2018-04-27 13:38:50,492 INFO org.apache.flink.runtime.jobmanager.MemoryArchivist - Started memory archivist akka://flink/user/archi
ve
2018-04-27 13:38:50,502 INFO org.apache.flink.runtime.jobmanager.JobManager - JobManager akka.tcp://flink@localhost:6123/user/
jobmanager was granted leadership with leader session ID Some(00000000-0000-0000-0000-000000000000).
2018-04-27 13:38:50,505 INFO org.apache.flink.runtime.clusterframework.standalone.StandaloneResourceManager - Trying to associate with JobMa
nager leader akka.tcp://flink@localhost:6123/user/jobmanager
2018-04-27 13:38:50,510 INFO org.apache.flink.runtime.clusterframework.standalone.StandaloneResourceManager - Resource Manager associating w
ith leading JobManager Actor[akka://flink/user/jobmanager#-1691988345] - leader session 00000000-0000-0000-0000-000000000000
2018-04-27 13:38:51,561 INFO org.apache.flink.runtime.clusterframework.standalone.StandaloneResourceManager - TaskManager 675697f1ebad9bc592
62c69d1fb85311 has started.
2018-04-27 13:38:51,563 INFO org.apache.flink.runtime.instance.InstanceManager - Registered TaskManager at 10.0.0.103 (akka.tcp://
flink@moogie.local:57215/user/taskmanager) as bba64099cb58b029ccfcb7a4ecd7720. Current number of registered hosts is 1. Current number of al
ive task slots is 1.
```



---

```
Moogie:flink-1.4.2 sridharalla$ ./bin/start-scala-shell.sh remote localhost 6123
Starting Flink Shell:
log4j:WARN No appenders could be found for logger (org.apache.flink.configuration.GlobalConfiguration).
log4j:WARN Please initialize the log4j system properly.
log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#noconfig for more info.

Connecting to Flink cluster (host: localhost, port: 6123).
```









Path, ID	Data Port	Last Heartbeat	All Slots	Free Slots	CPU Cores	Physical Memory	JVM Heap Size	Flink Managed Memory
akka.tcp://flink@mooglee.local:57215 /user/taskmanager 675697f1EBAD9bc59262c69D1fB85311	57217	2018-04-29, 19:03:57	1	1	8	16.0 GB	1.00 GB	641 MB

```

Task Manager | Last Heartbeat: 2018-04-29, 19:03:42 | akka.tcp://flink@mooglee.local:57215/user/taskmanager

Metrics | Logs | Stdout

Task Manager Logs

2018-04-27 13:38:49,151 INFO org.apache.flink.runtime.taskmanager.TaskManager
-----
2018-04-27 13:38:49,153 INFO org.apache.flink.runtime.taskmanager.TaskManager
ate:28.02.2018 @ 12:53:14 UTC)
2018-04-27 13:38:49,153 INFO org.apache.flink.runtime.taskmanager.TaskManager
- OS current user: sridharalla
2018-04-27 13:38:49,441 WARN org.apache.hadoop.util.NativeCodeLoader
- Unable to load native-hadoop library for your platform
m... using builtin-java classes where applicable
2018-04-27 13:38:49,501 INFO org.apache.flink.runtime.taskmanager.TaskManager
- Current Hadoop/Kerberos user: sridharalla
2018-04-27 13:38:49,501 INFO org.apache.flink.runtime.taskmanager.TaskManager
- JVM: Java HotSpot(TM) 64-Bit Server VM - Oracle Corp
ration - 1.9/25.172-b11
2018-04-27 13:38:49,501 INFO org.apache.flink.runtime.taskmanager.TaskManager
- Maximum heap size: 1024 MiBytes
2018-04-27 13:38:49,501 INFO org.apache.flink.runtime.taskmanager.TaskManager
- JAVA_HOME: (not set)
2018-04-27 13:38:49,505 INFO org.apache.flink.runtime.taskmanager.TaskManager
- Hadoop version: 2.8.0
2018-04-27 13:38:49,505 INFO org.apache.flink.runtime.taskmanager.TaskManager
- JVM Options:
2018-04-27 13:38:49,505 INFO org.apache.flink.runtime.taskmanager.TaskManager
- --XX:+UseG1GC
2018-04-27 13:38:49,505 INFO org.apache.flink.runtime.taskmanager.TaskManager
- -Xms1024M
2018-04-27 13:38:49,505 INFO org.apache.flink.runtime.taskmanager.TaskManager
- -Xmx1024M
2018-04-27 13:38:49,505 INFO org.apache.flink.runtime.taskmanager.TaskManager
- --XX:MaxDirectMemorySize=8388607T
2018-04-27 13:38:49,505 INFO org.apache.flink.runtime.taskmanager.TaskManager
- -Dlog.file=/Users/sridharalla/flink-1.4.2/log/fli
nk-sridharalla-taskmanager-0-Mooglee.local.log
2018-04-27 13:38:49,505 INFO org.apache.flink.runtime.taskmanager.TaskManager
- -Dlog4j.configuration=file:/Users/sridharalla/fli
nk-1.4.2/conf/log4j.properties
2018-04-27 13:38:49,505 INFO org.apache.flink.runtime.taskmanager.TaskManager
- -Dlogback.configurationFile=file:/Users/sridharal
la/flink-1.4.2/conf/logback.xml
2018-04-27 13:38:49,505 INFO org.apache.flink.runtime.taskmanager.TaskManager
- Program Arguments:
2018-04-27 13:38:49,506 INFO org.apache.flink.runtime.taskmanager.TaskManager
- --configDir
2018-04-27 13:38:49,506 INFO org.apache.flink.runtime.taskmanager.TaskManager
- /Users/sridharalla/flink-1.4.2/conf
2018-04-27 13:38:49,506 INFO org.apache.flink.runtime.taskmanager.TaskManager
- Classpath: /Users/sridharalla/flink-1.4.2/lib/flink-
python_2.11-1.4.2.jar:/Users/sridharalla/flink-1.4.2/lib/flink-shaded-hadoop2-uber-1.4.2.jar:/Users/sridharalla/flink-1.4.2/lib/log4j-1.2.17.jar:/U
sers/sridharalla/flink-1.4.2/lib/stf4j-log4j12-1.7.7.jar:/Users/sridharalla/flink-1.4.2/lib/flink-dist_2.11-1.4.2.jar:::
2018-04-27 13:38:49,506 INFO org.apache.flink.runtime.taskmanager.TaskManager
-----

```

← → ↻ 🏠 localhost:8081/#/taskmanager/6756971ebad9bc59262c69d1fb85311/metrics ... 📧 ☆

**Apache Flink Dashboard**

Task Manager Last Heartbeat: 2018-04-29, 19:03:32 akka.tcp://flink@moogie.local:57215/user/taskmanager

Metrics Logs Stdout

### Overview

Data Port	All Slots	Free Slots	CPU Cores	Physical Memory	JVM Heap Size	Flink Managed Memory
57217	1	1	8	16.0 GB	1.00 GB	641 MB

### Memory

#### JVM (Heap/Non-Heap)

Type	Committed	Used	Maximum
Heap	1.00 GB	861 MB	1.00 GB
Non-Heap	52.8 MB	51.6 MB	-1 B
Total	1.05 GB	912 MB	1.00 GB

#### Outside JVM

Type	Count	Used	Capacity
Direct	7	263 KB	263 KB
Mapped	0	0 B	0 B

### Network

#### Memory Segments

← → ↻ 🏠 localhost:8081/#/submit ... 📧 ☆

**Apache Flink Dashboard**

Submit new Job

#### Uploaded Jars

Name	Upload Time	Entry Class

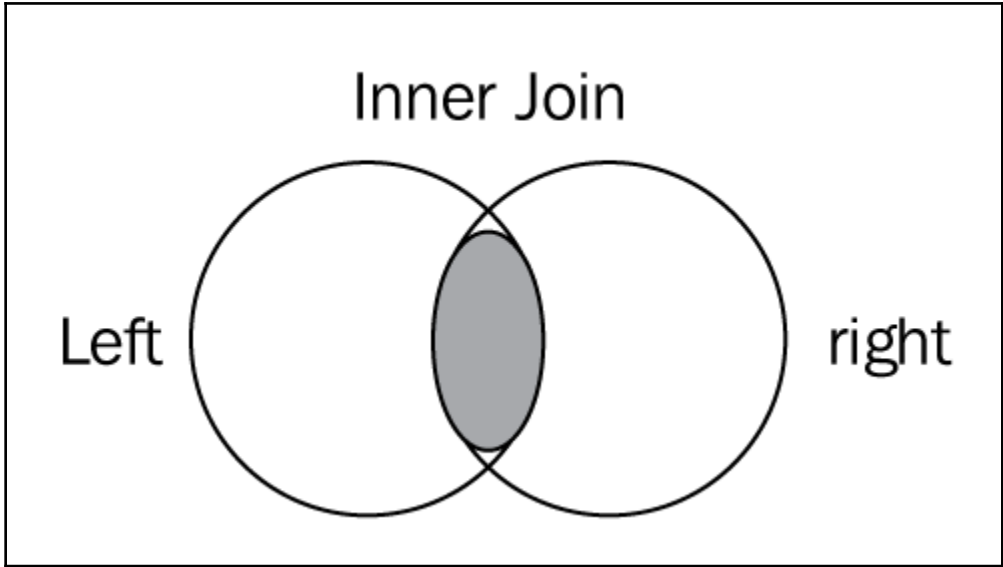
Add New +











localhost:8081/#/jobs/859cb2cd9f60371e70b6ef9450ea00c6

Apache Flink Dashboard

Flink Java Job at Tue May 22 17:08:18 EDT 2018  
2018-05-22, 17:08:21 - 2018-05-22, 17:08:22 305ms

Overview Timeline Exceptions Configuration

```

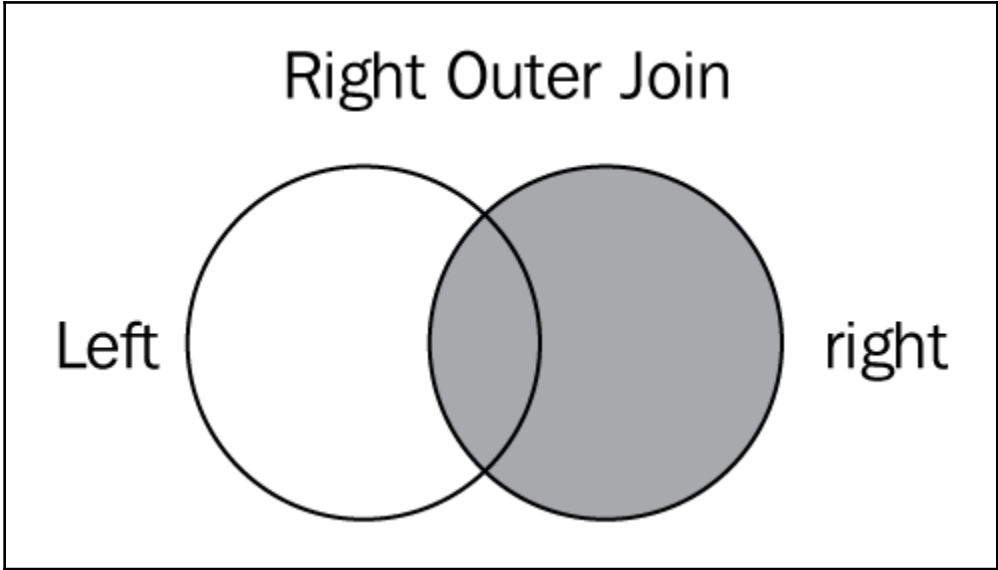
    graph LR
      DS[CHAIN DataSource at Sline19.Read] --> F1[Filter at Sline21]
      F1 --> M1[Map at Sline21]
      DS --> M2[Map at Sline17]
      M1 --> J[Join]
      M2 --> J
      J --> GR1[GroupReduce]
      GR1 --> GR2[GroupReduce]
      GR2 --> DSink[Data Sink]
  
```

Subtasks

Aggregate task statistics by TaskManager

Start Time	End Time	Duration	Name
2018-05-22, 17:08:21	2018-05-22, 17:08:21	36ms	CHAIN DataSource (at Sline19.Read) -> Filter (at Sline21) -> Map (at Sline21) -> Join (at org.apache.flink.optimizer.optimizer.Operator) -> Map (at Sline17) -> GroupReduce (GroupReduce) -> Data Sink (collect)





localhost:8081/#/jobs/6e24cddb31f06ae7aa7229c3bdd7d504

Flink Java Job at Tue May 22 17:06:31 EDT 2018 2018-05-22, 17:06:34 - 2018-05-22, 17:06:34 223ms

Overview Timeline Exceptions Configuration

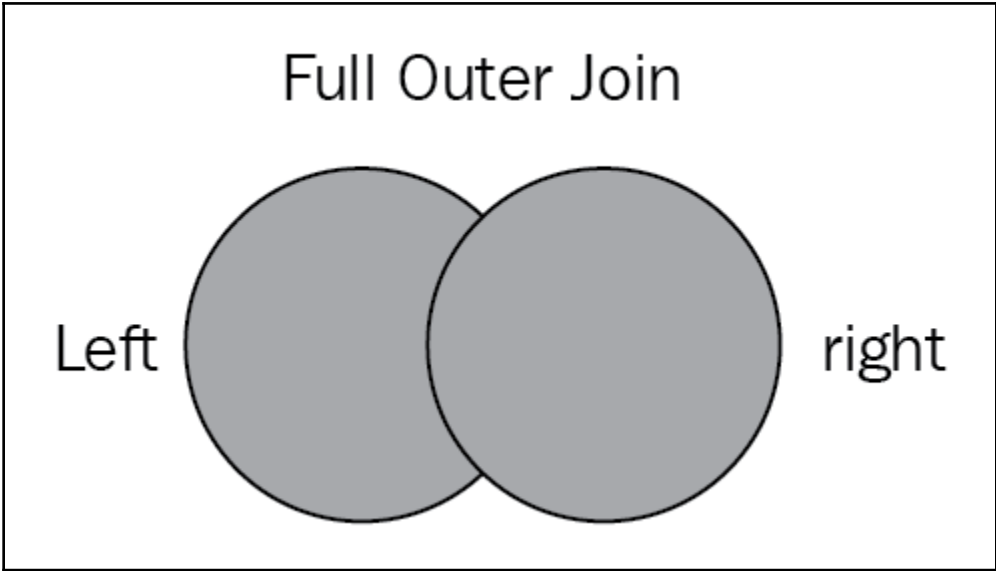
```

graph LR
    DS[Data Source at] --> F1[Filter]
    F1 --> M1[Map]
    F1 --> F2[Filter]
    F2 --> M2[Map]
    M1 --> OJ[Outer Join]
    M2 --> OJ
    OJ --> GR1[GroupReduce SUM]
    GR1 --> GR2[GroupReduce at]
    GR2 --> DSink[Data Sink outlet]
  
```

**Subtasks** Task Metrics Watermarks Accumulators Checkpoints

Aggregate task statistics by TaskManager

Start Time	End Time	Duration	Name
2018-05-22, 17:06:34	2018-05-22, 17:06:34	40ms	CHAIN DataSource (at) (Filter at) (Map at) (Map at)



localhost:8081/#/jobs/c7435ee774db425c1bd01b52a4bae7e3

**Apache Flink Dashboard**

Flink Java Job at Tue May 22 17:07:35 EDT 2018  
c7435ee774db425c1bd01b52a4bae7e3 2018-05-22, 17:07:38 - 2018-05-22, 17:07:38 307ms

Overview Timeline Exceptions Configuration

The job execution graph shows the following components:

- Data Source (at file:///...)**: Reads data from a file.
- Filter at [line19...]**: Filters the data based on a condition.
- Map at [line21...]**: Applies a map function to the filtered data.
- Outer Join -> Map -> GroupCombine**: Performs a full outer join with another dataset, followed by a map and group combine operation.
- GroupReduce Reduce (SUM1)**: Aggregates the data using a sum reducer.
- GroupReduce GroupReduce (GroupReduce at [line21...])**: Further aggregates the data.
- Data Sink (collect)**: Writes the final aggregated results to a sink.

Subtasks Task Metrics Watermarks Accumulators Checkpoints

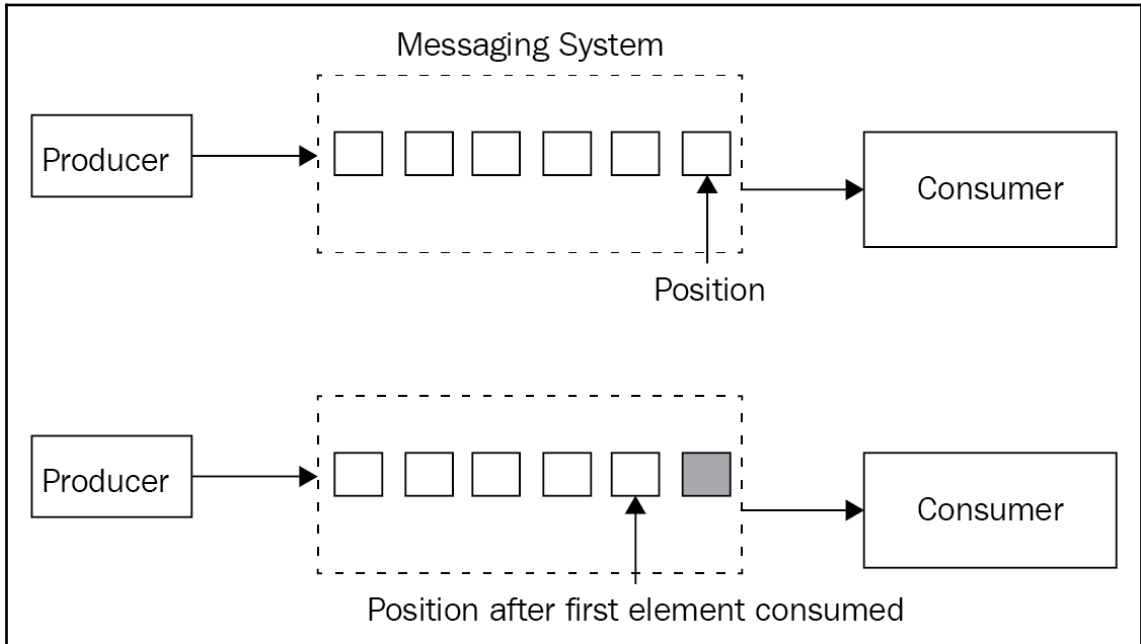
Aggregate task statistics by TaskManager

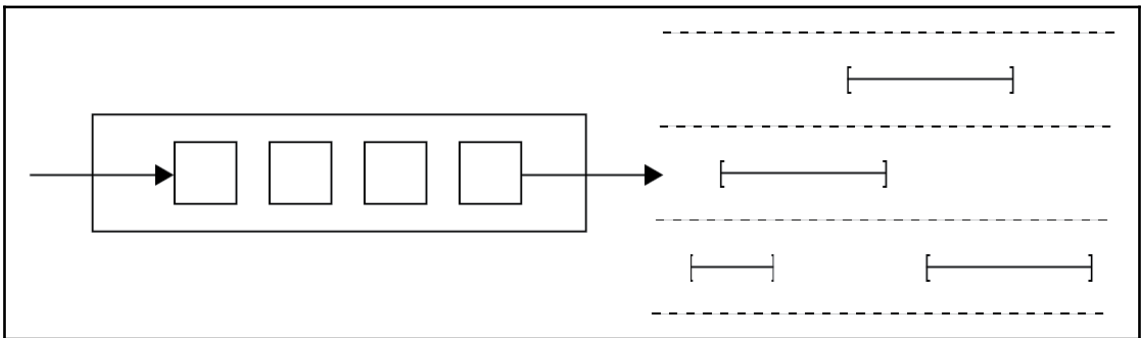
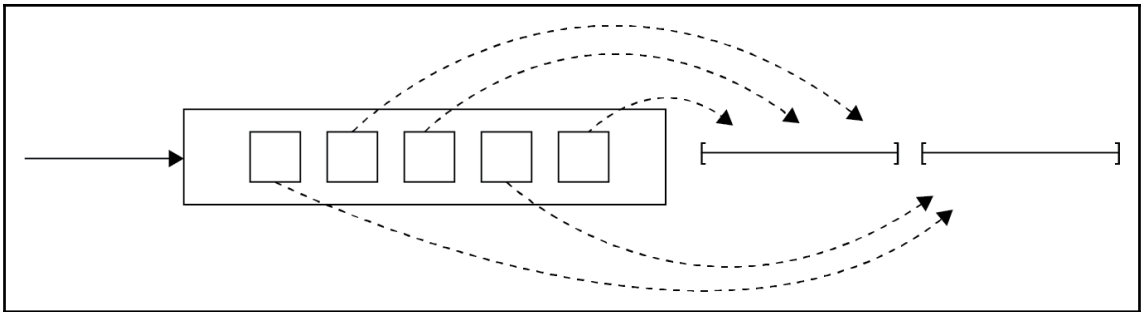
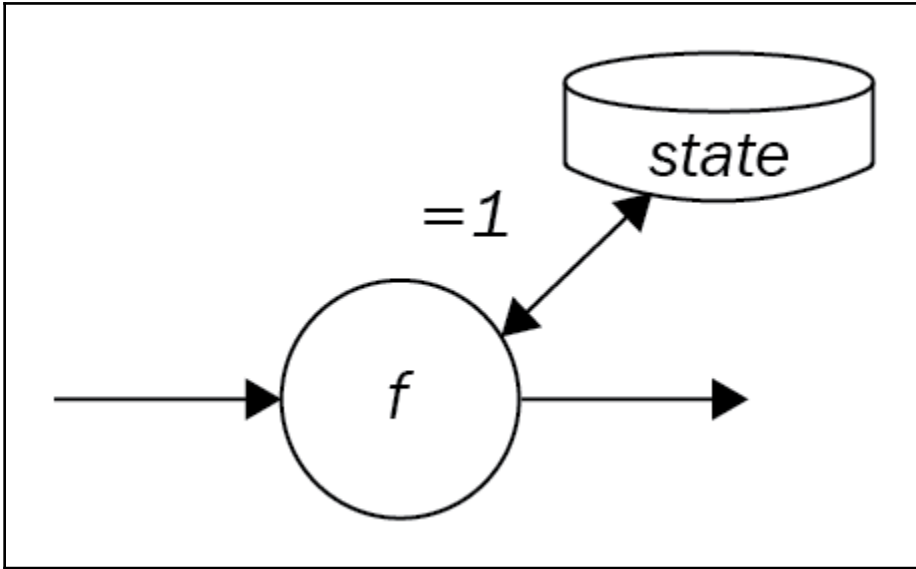
Start Time	End Time	Duration	Name
2018-05-22, 17:07:38	2018-05-22, 17:07:38	42ms	CHAIN DataSource (at [line19...]) (Filter at [line19...]) -> Map at [line21...] -> Map at [line21...] -> Map at [line21...]
2018-05-22, 17:07:38	2018-05-22, 17:07:38	42ms	CHAIN DataSource (at [line19...])

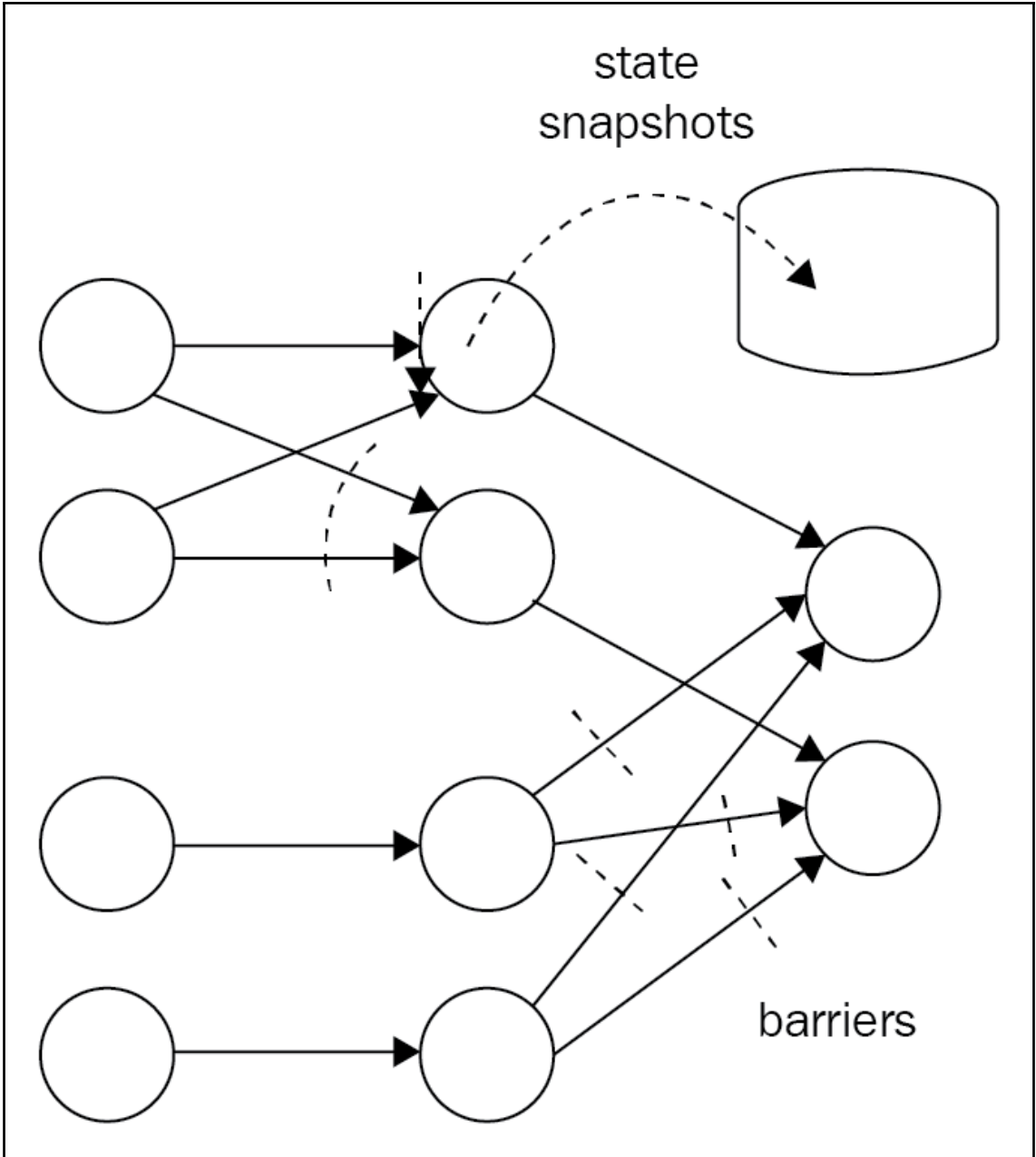


---

## Chapter 9: Stream Processing with Apache Flink









---

```
Moogie:flink-1.4.2 sridharalla$ ./bin/start-scala-shell.sh remote localhost 6123
Starting Flink Shell:
log4j:WARN No appenders could be found for logger (org.apache.flink.configuration.GlobalConfiguration).
log4j:WARN Please initialize the log4j system properly.
log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#noconfig for more info.

Connecting to Flink cluster (host: localhost, port: 6123).
```



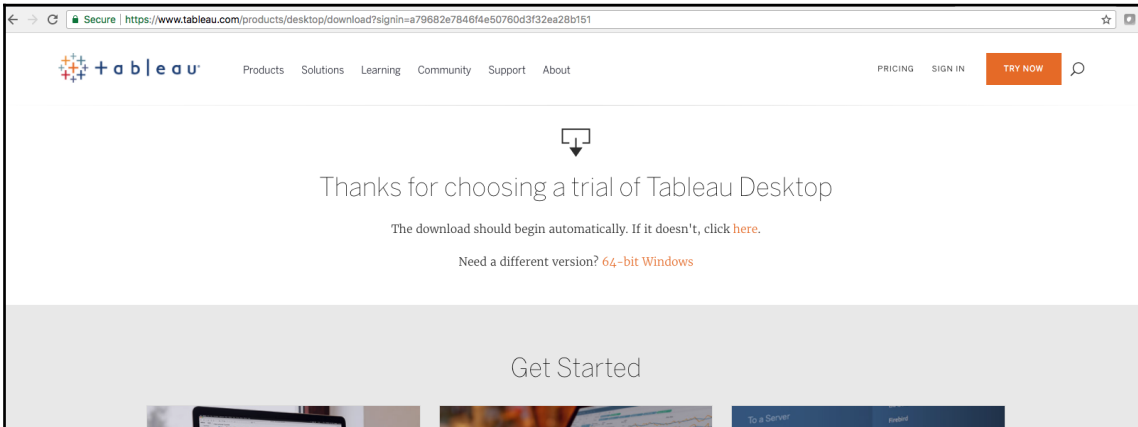
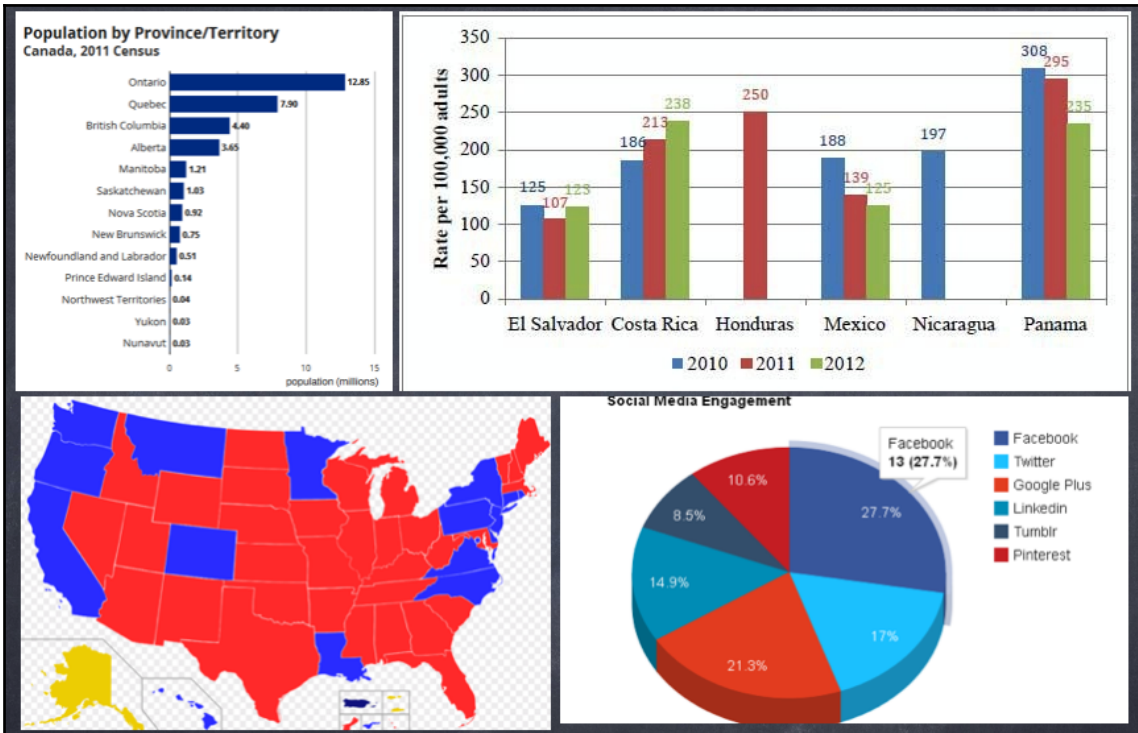
```
Moogie:~ sridharalla$ nc -l 9000
```

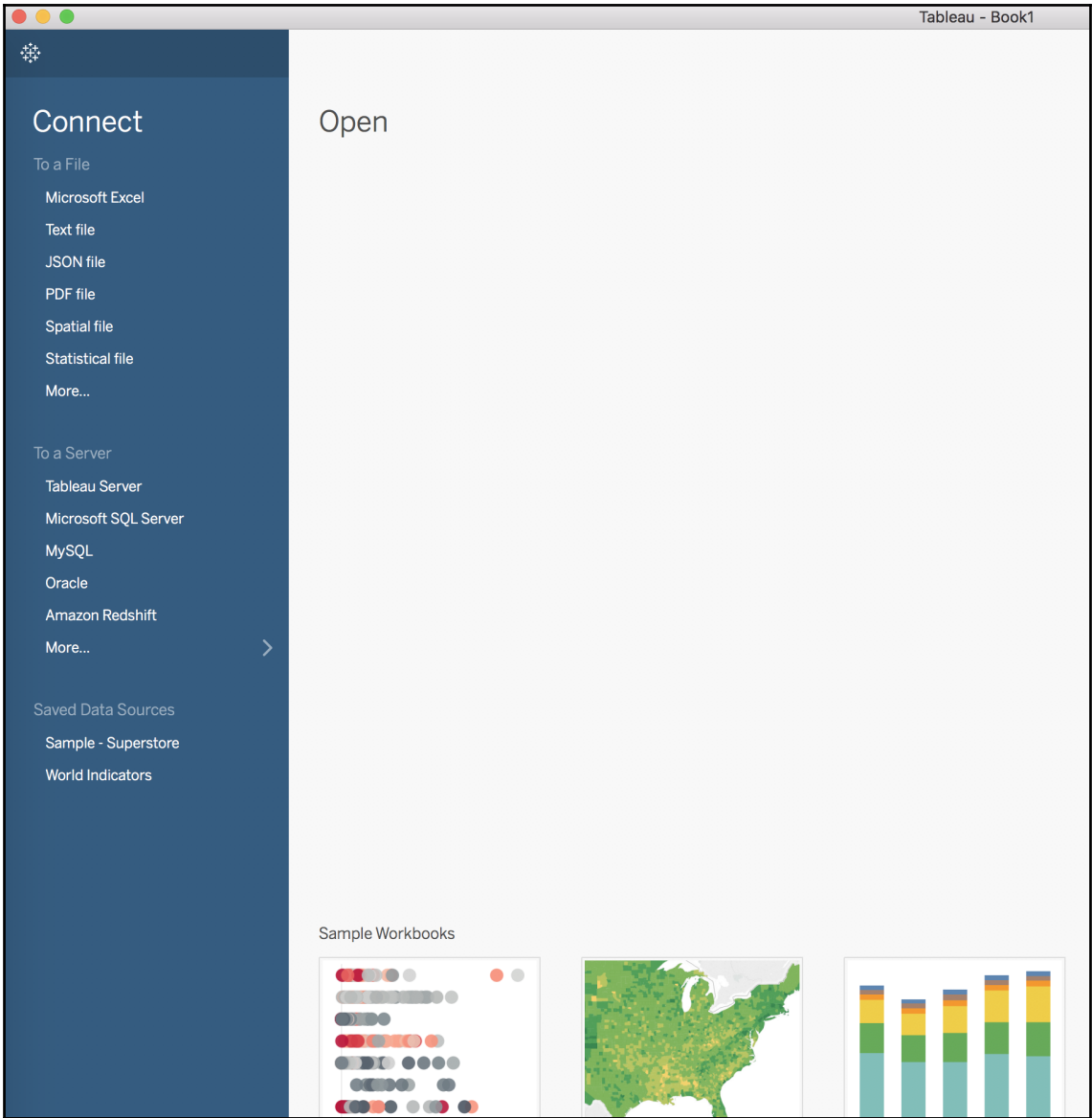


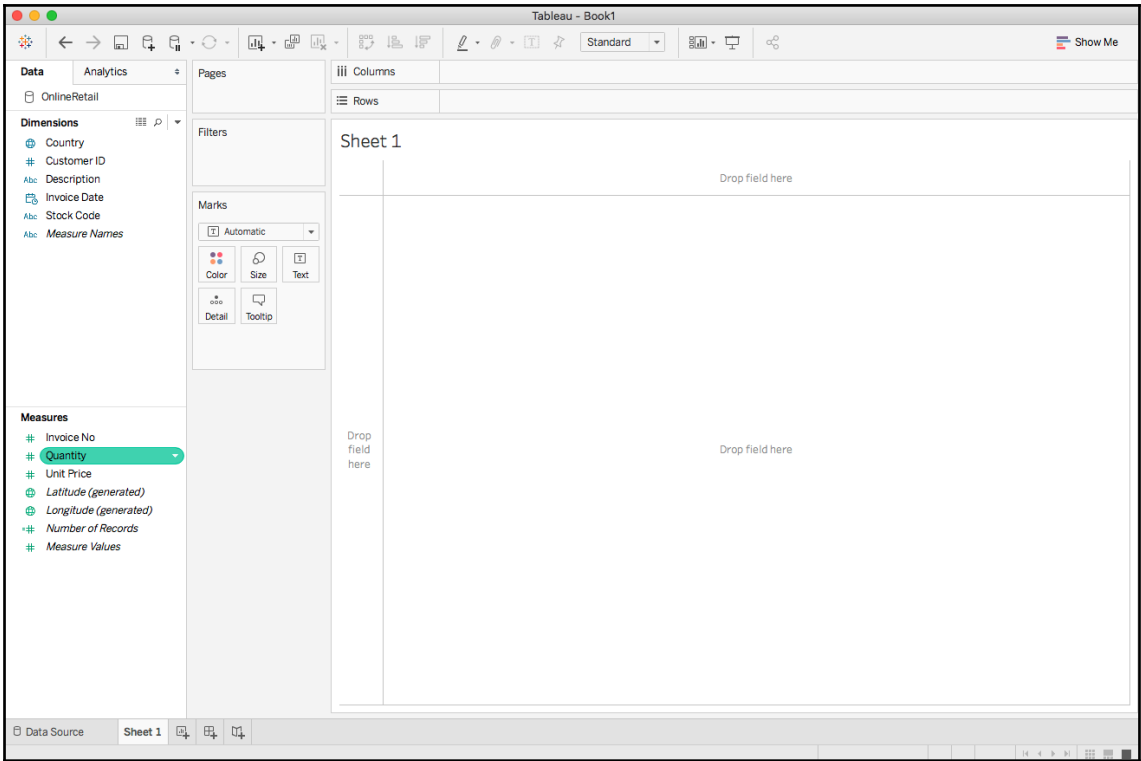


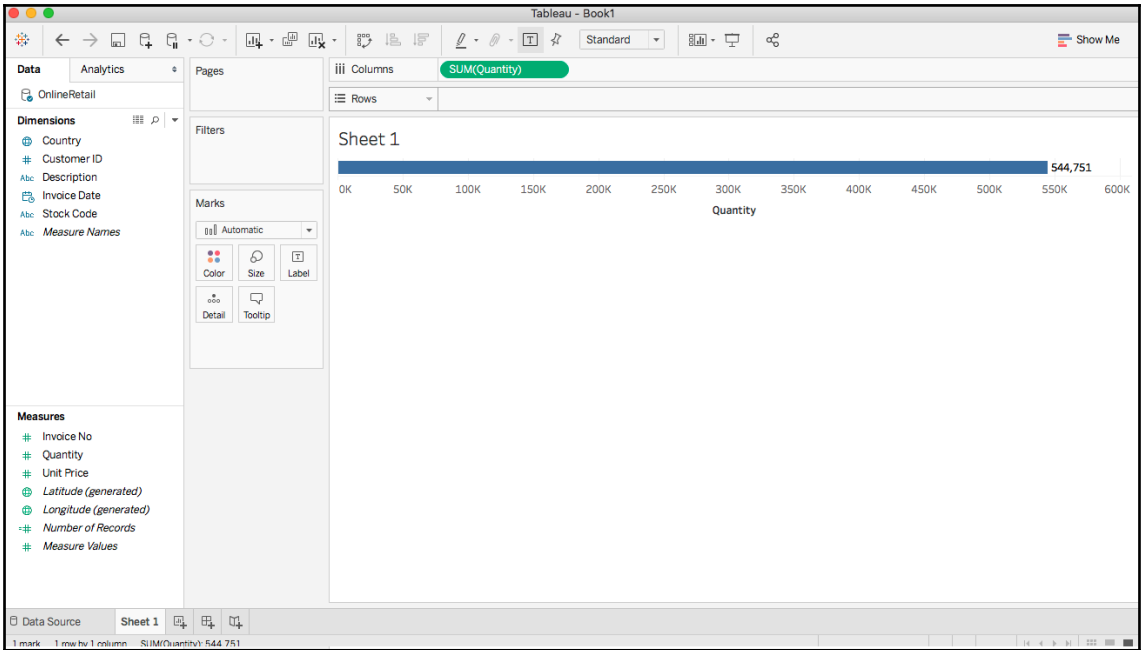














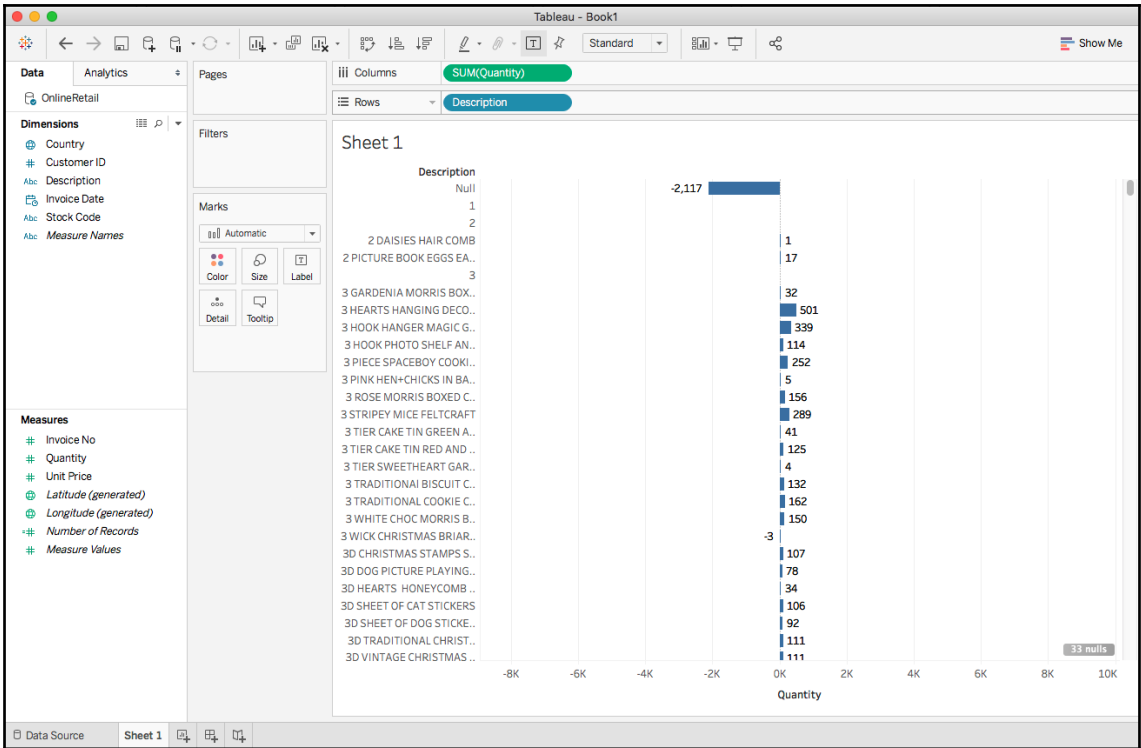


Tableau - Book1

Columns: SUM(Quantity) | Rows: Description

Sheet 1

Filter Field [Quantity]

How do you want to filter on [Quantity]?

- # All values
- # Sum
- # Average
- # Median
- # Count
- # Count (Distinct)
- # Minimum
- # Maximum
- # Standard deviation
- # Standard deviation (Population)
- # Variance
- # Variance (Population)
- # Attribute

Cancel Next >

33 nulls

The screenshot displays the Tableau Desktop interface. The main view is a bar chart titled 'Sheet 1' showing the sum of quantities for various product descriptions. The x-axis is labeled 'Quantity' and ranges from -8K to 10K. The y-axis is labeled 'Description' and ranges from 01 to 111. A single bar is visible for the 'Null' description, with a value of -2,117. A dialog box titled 'Filter [Quantity]' is open, showing the 'Range of values' tab. The range is set from -74,215 to 74,215. The 'Show:' dropdown is set to 'Only Relevant Values' and the 'Include Null Values' checkbox is unchecked. The status bar at the bottom indicates '2884 marks 2917 rows by 1 column SUM(Quantity): 544,751'.

Tableau - Book1

Columns: SUM(Quantity)

Rows: Description

Sheet 1

Filter [Quantity]

Range of values

Range of values: 0 to 74,215

Show: Only Relevant Values

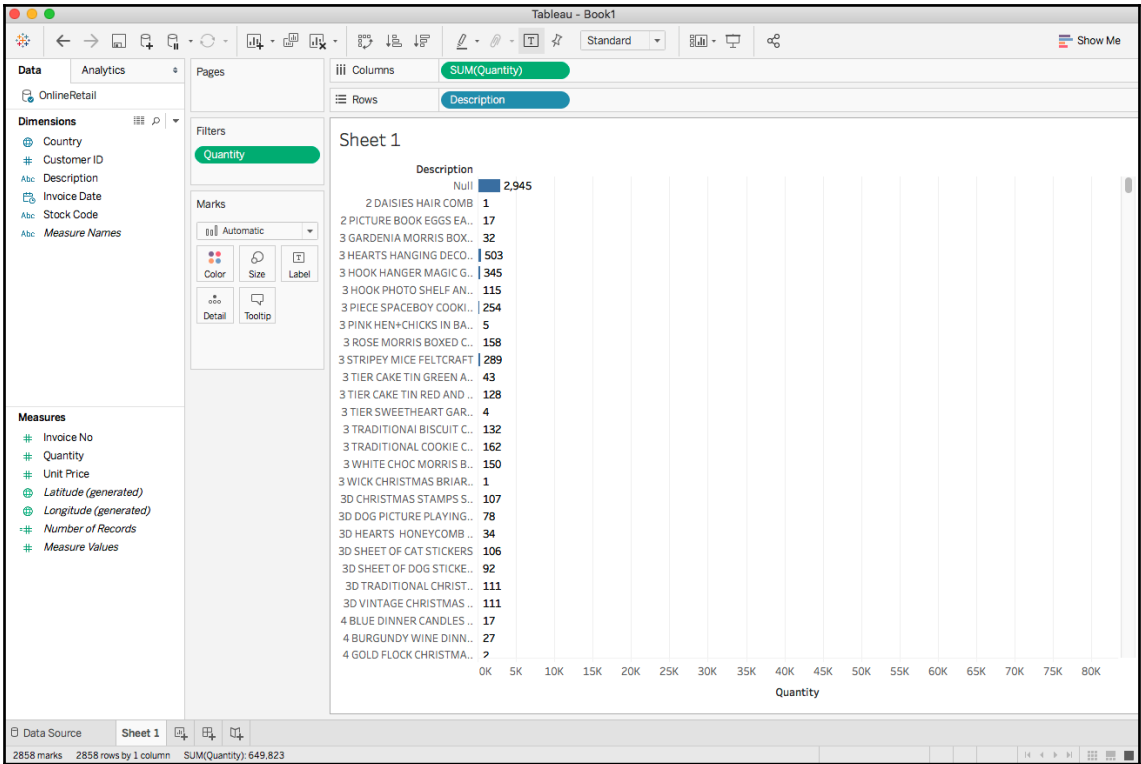
Include Null Values:

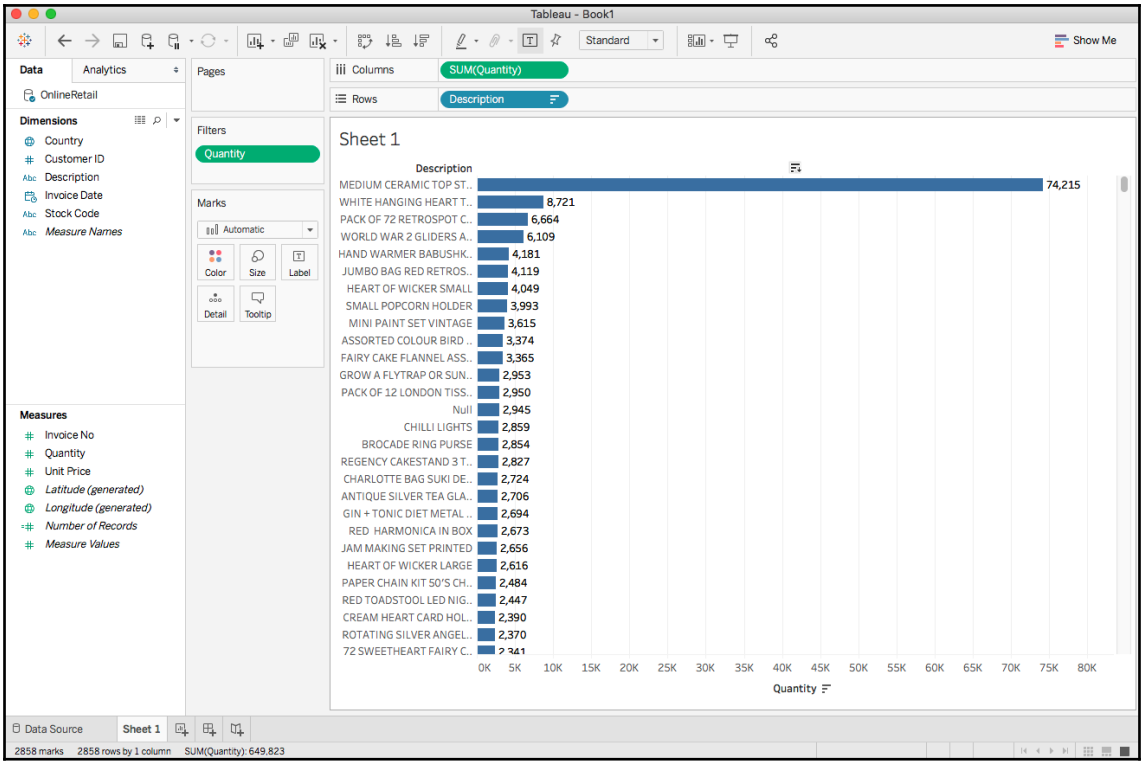
Reset Apply Cancel OK

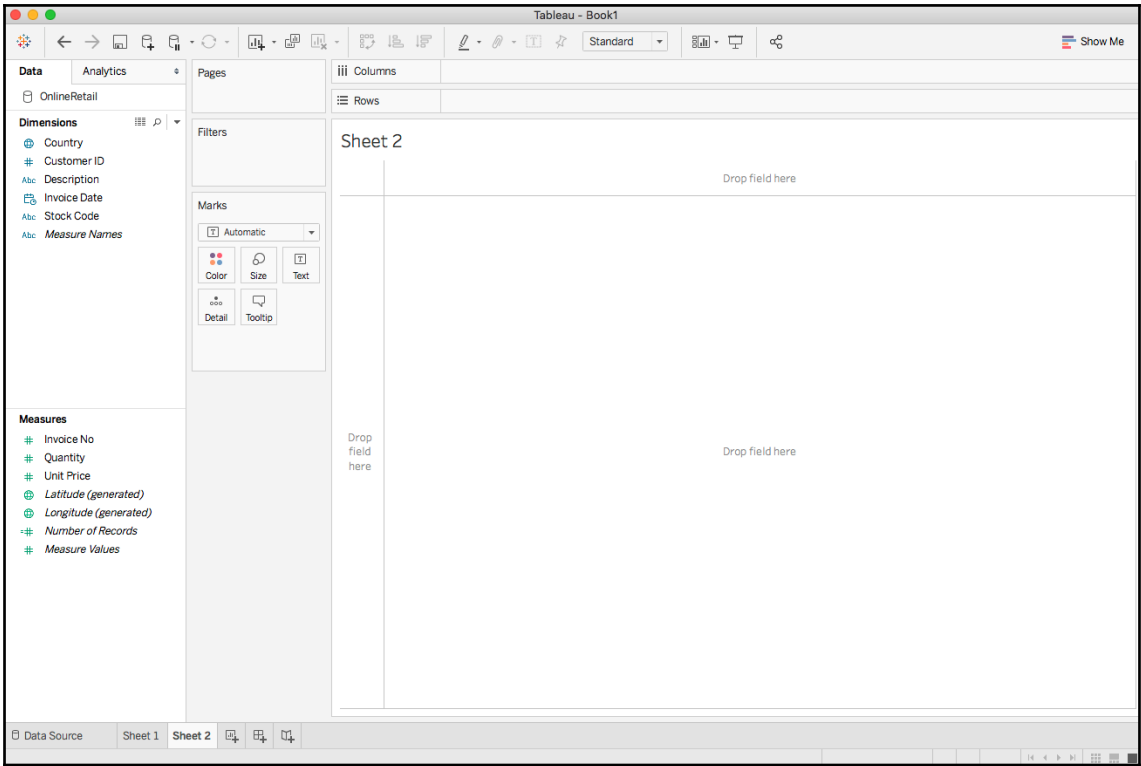
Quantity

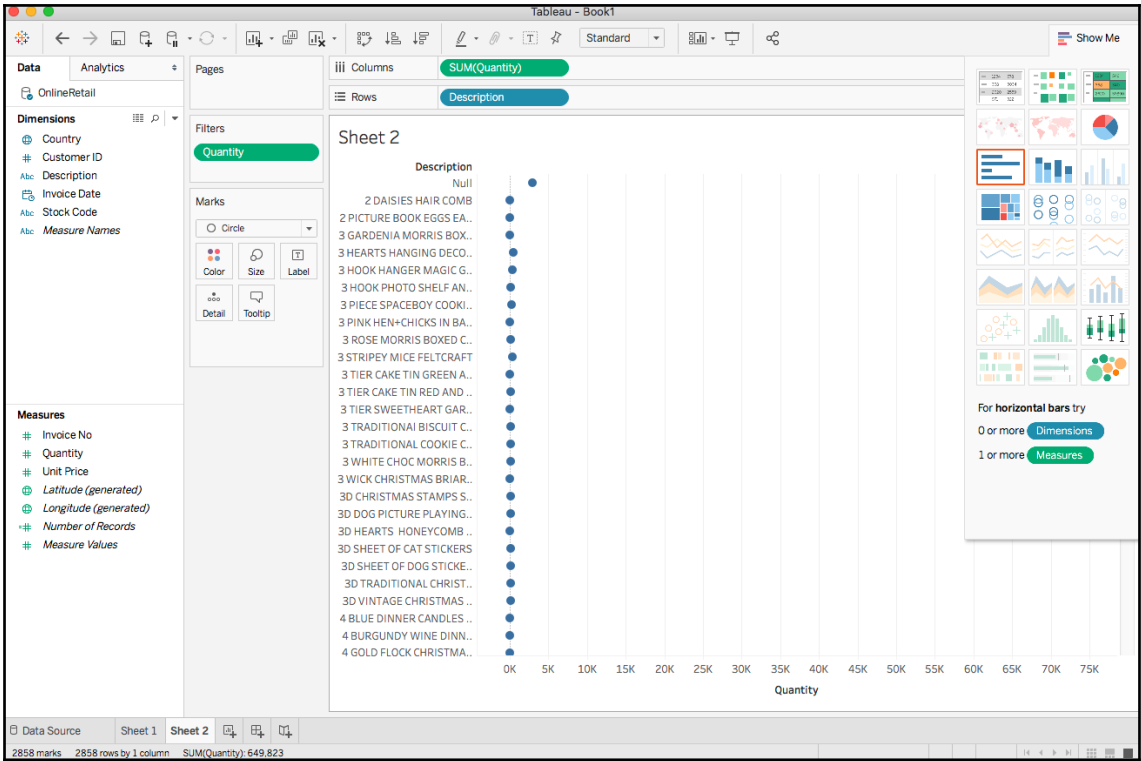
33 nulls

2884 marks 2917 rows by 1 column SUM(Quantity): 544,751

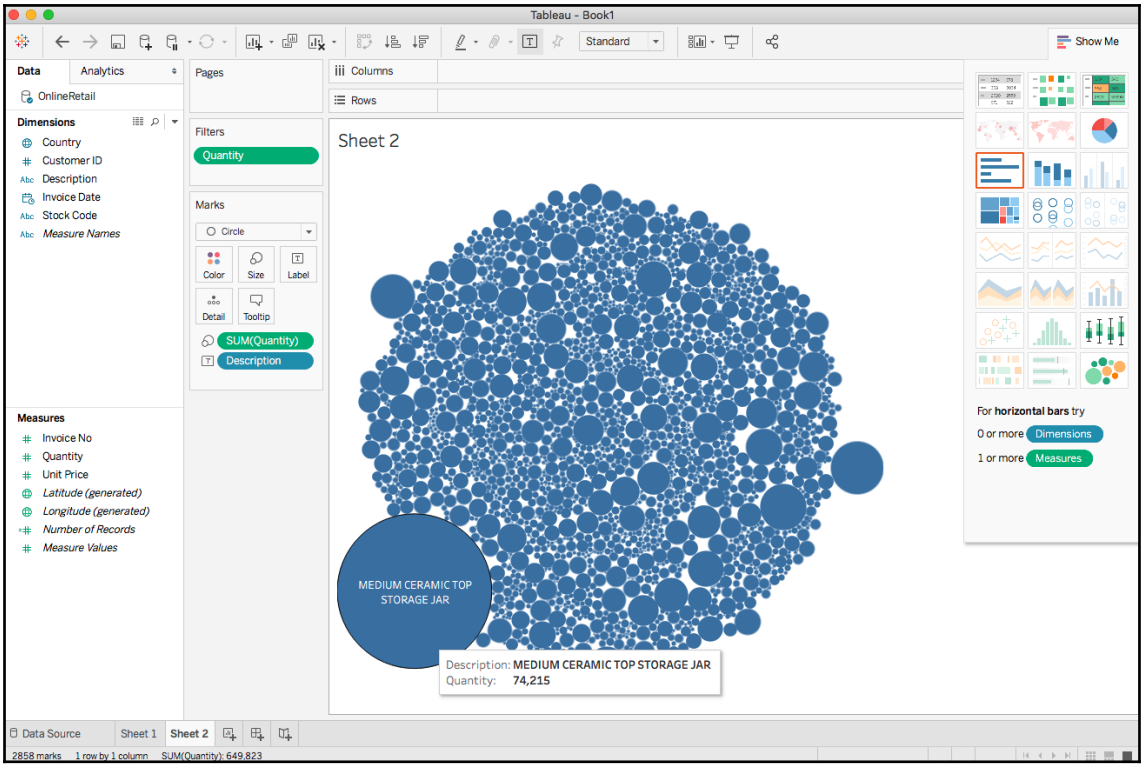


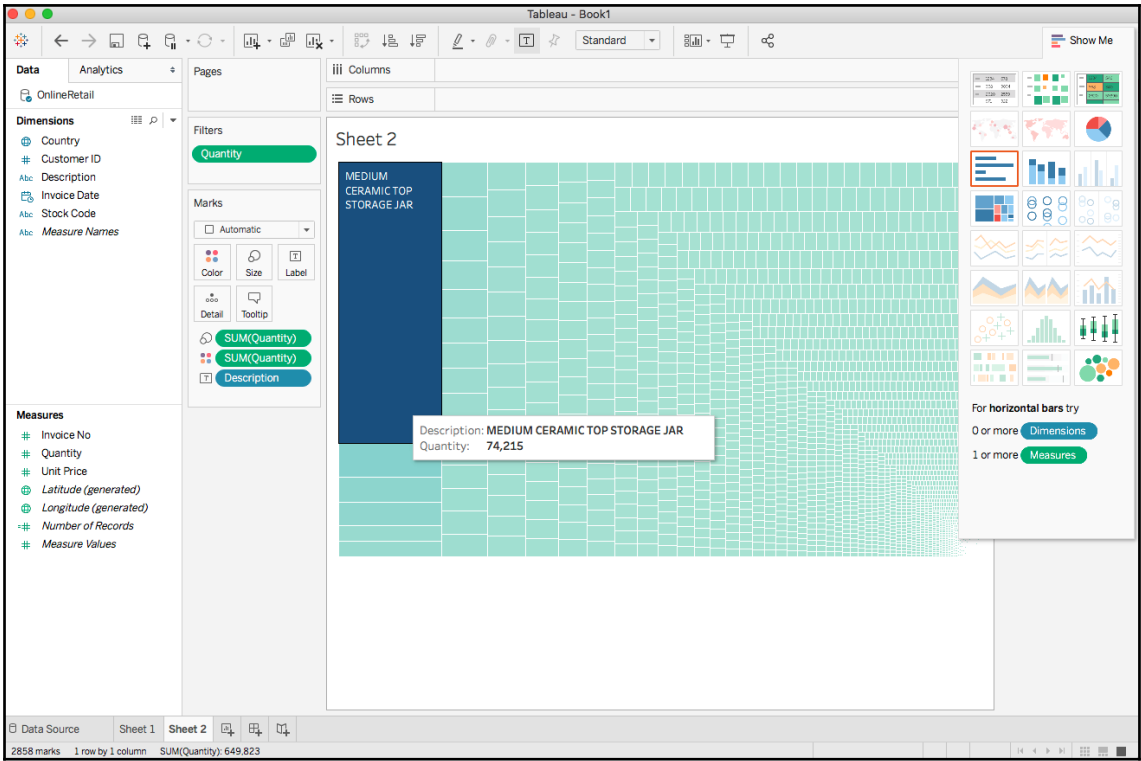


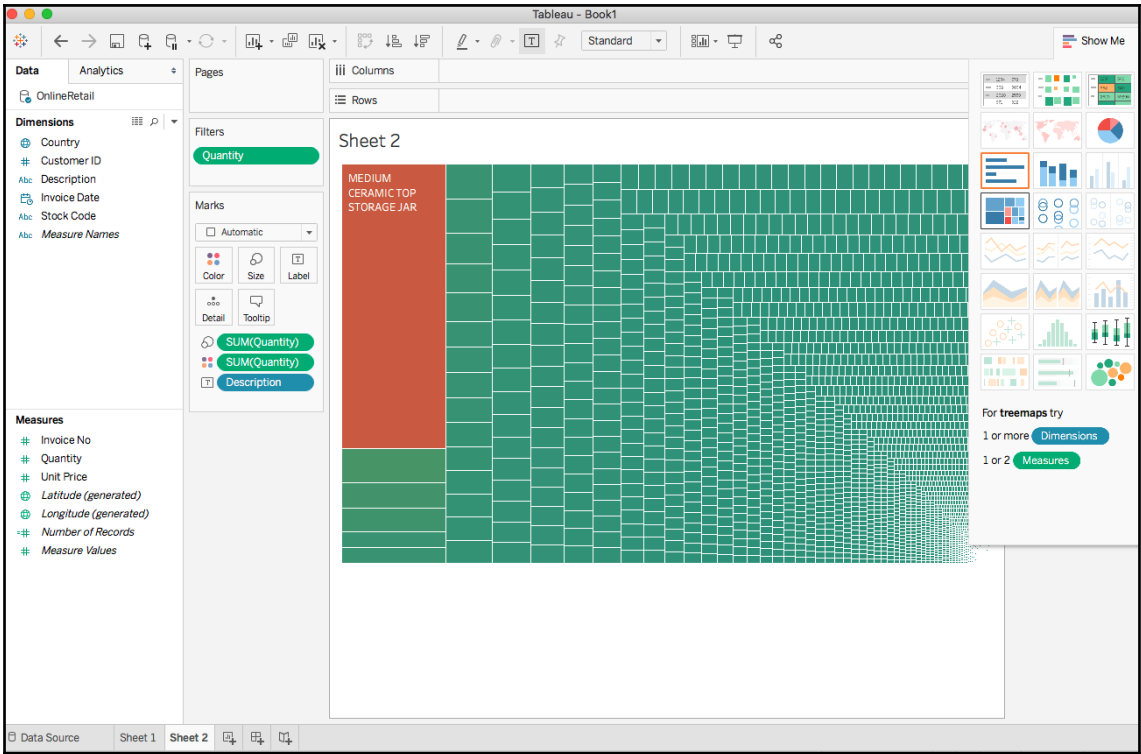


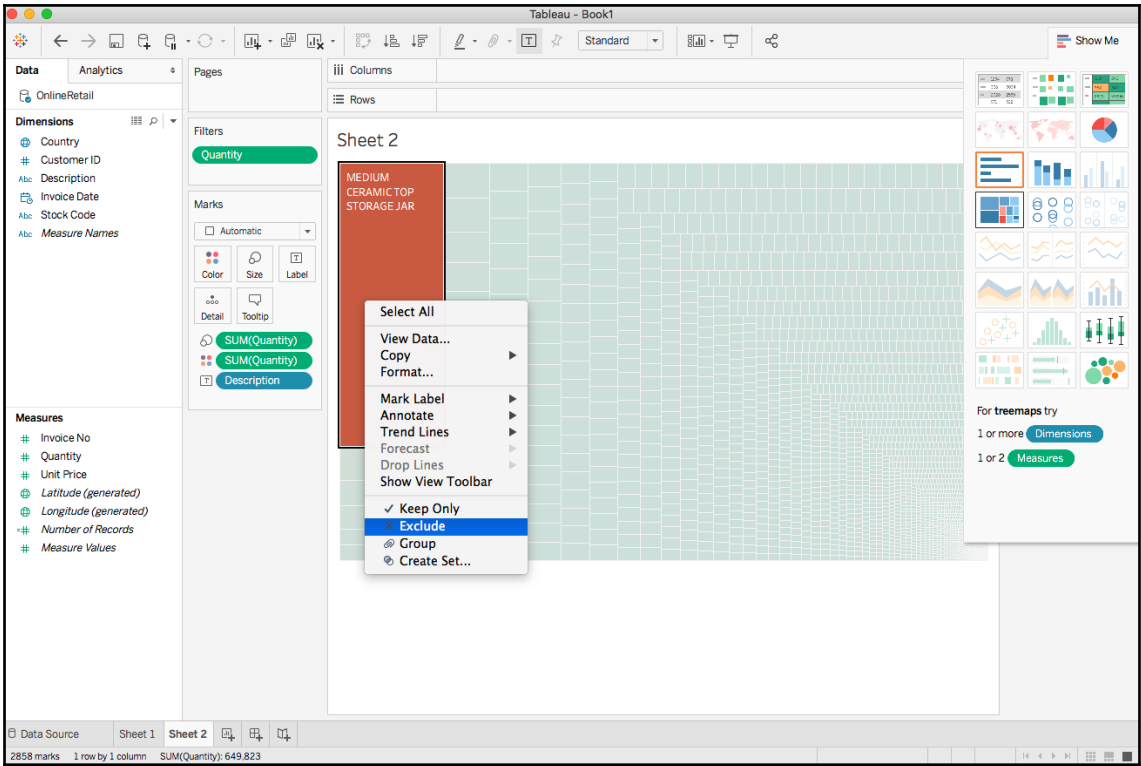


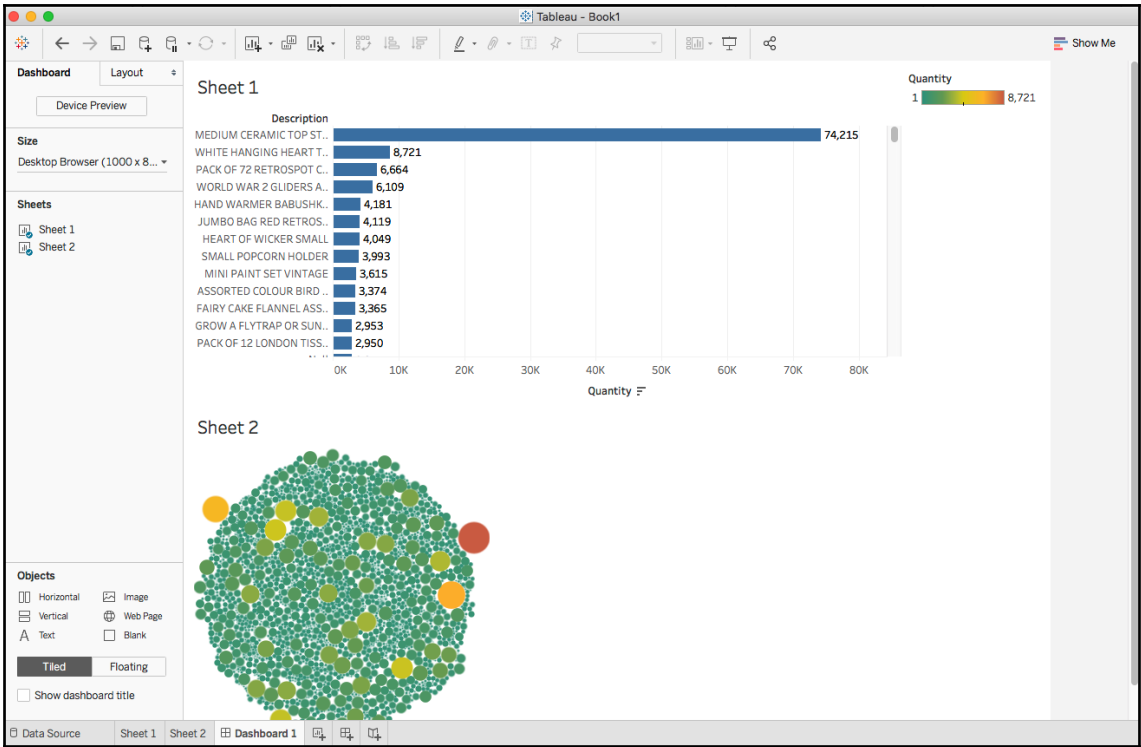


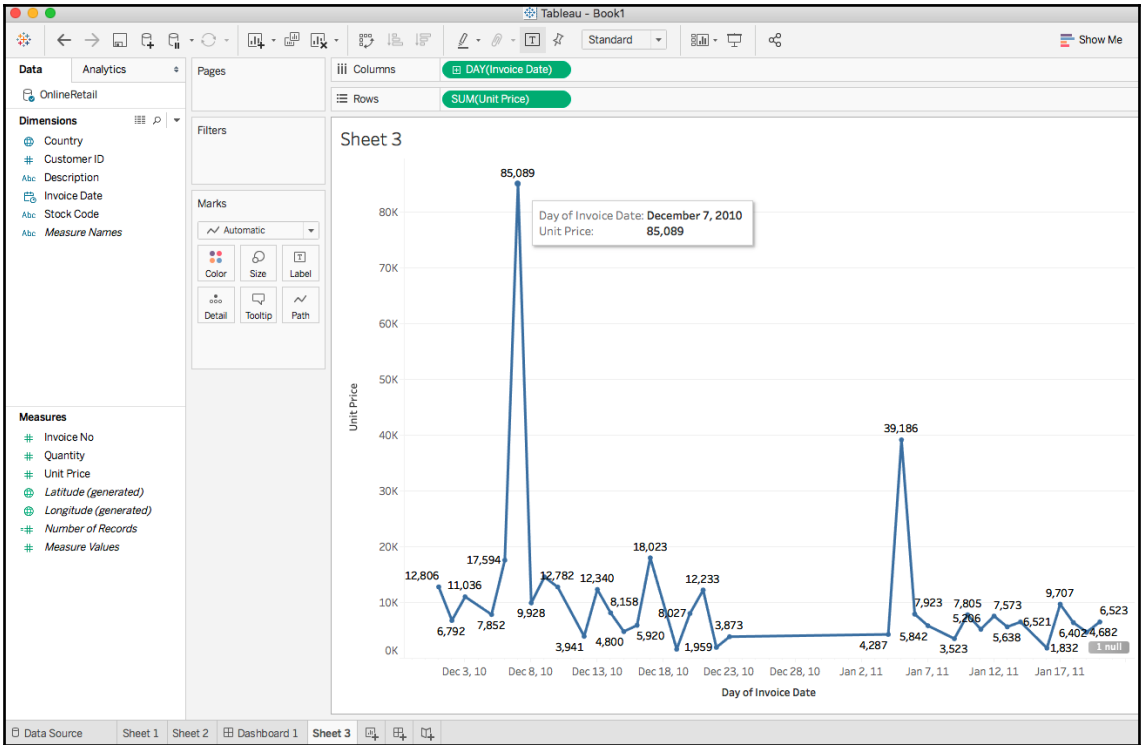


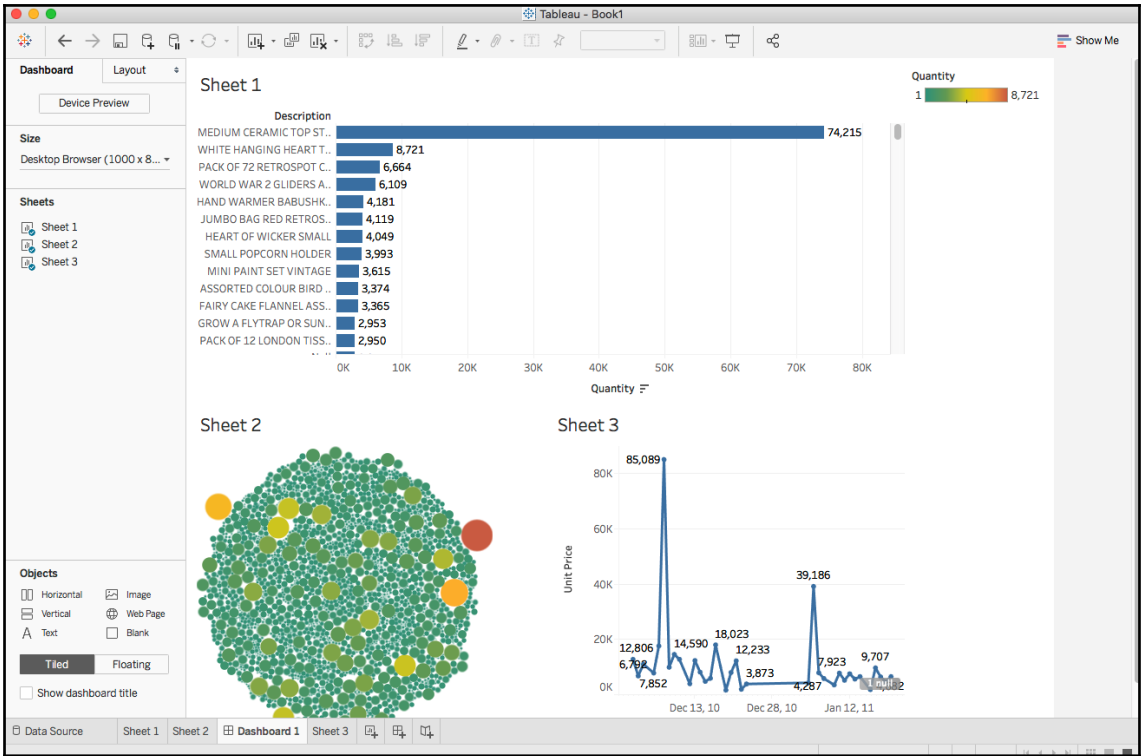


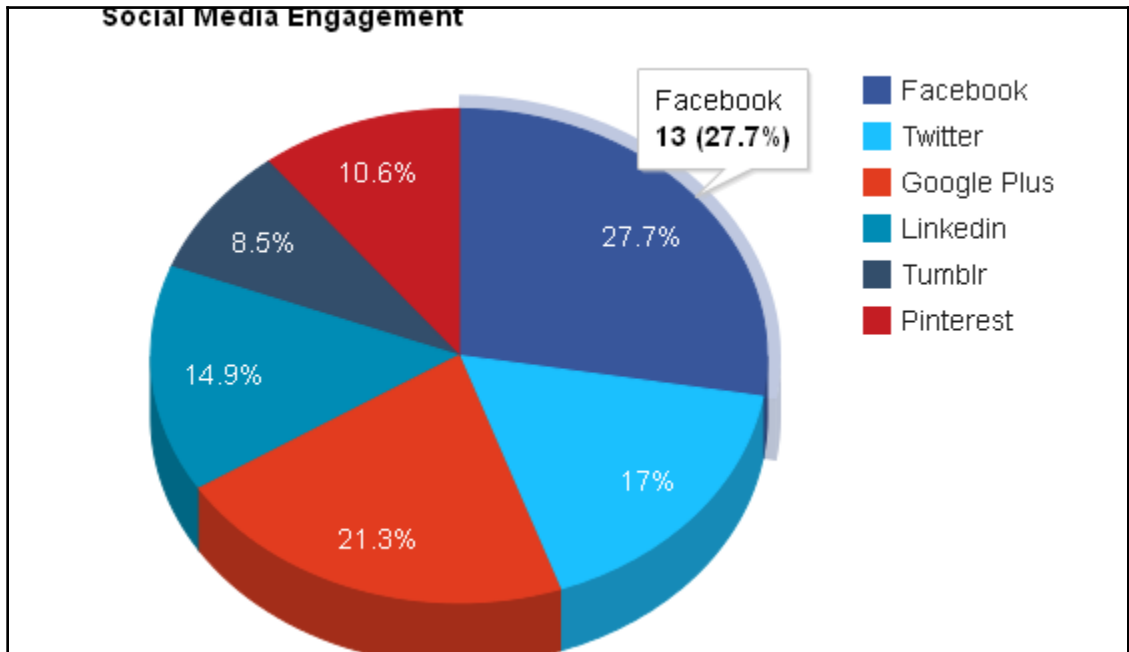
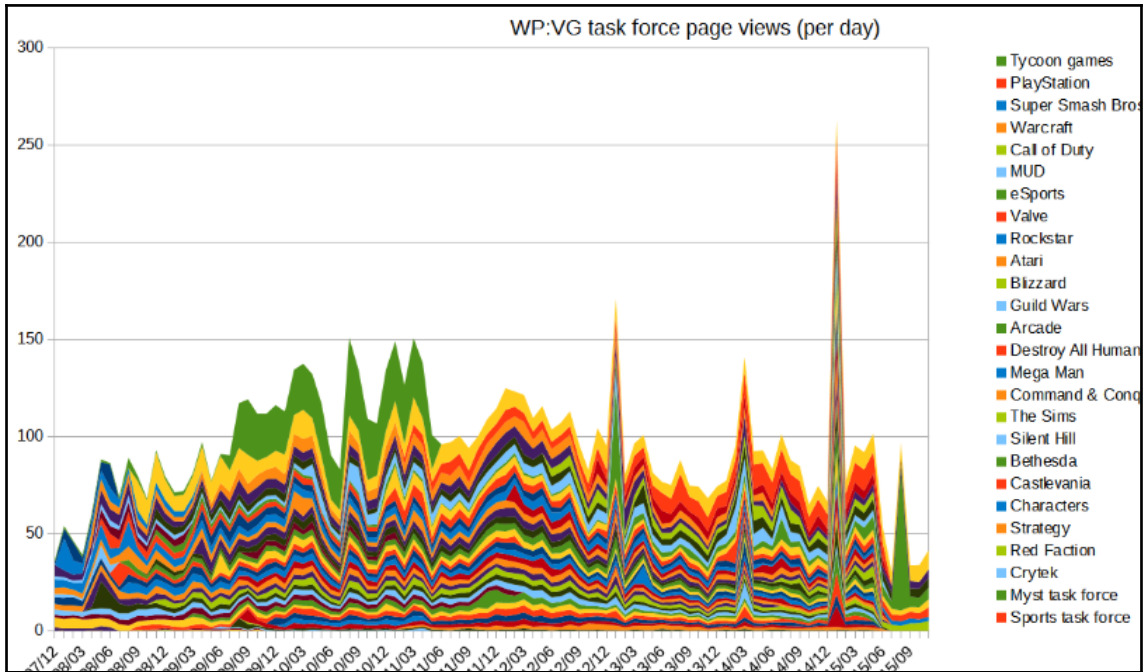




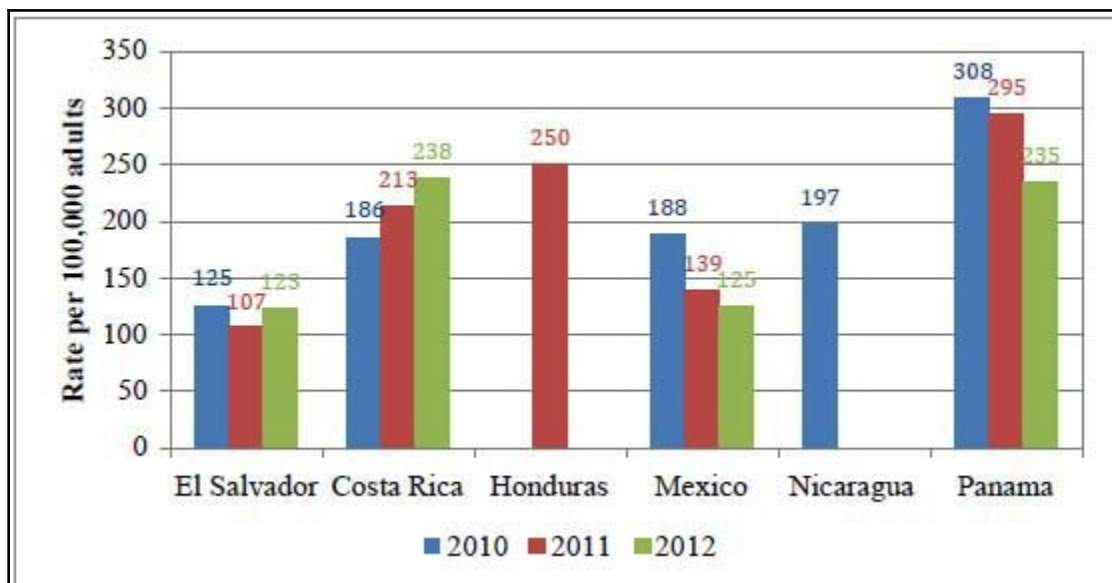






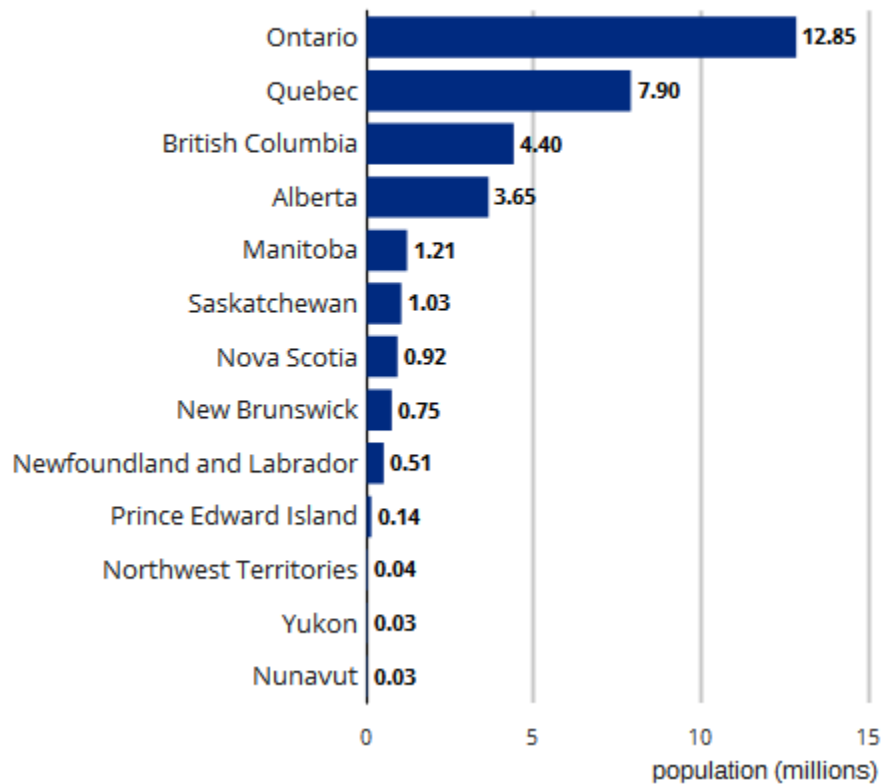






---

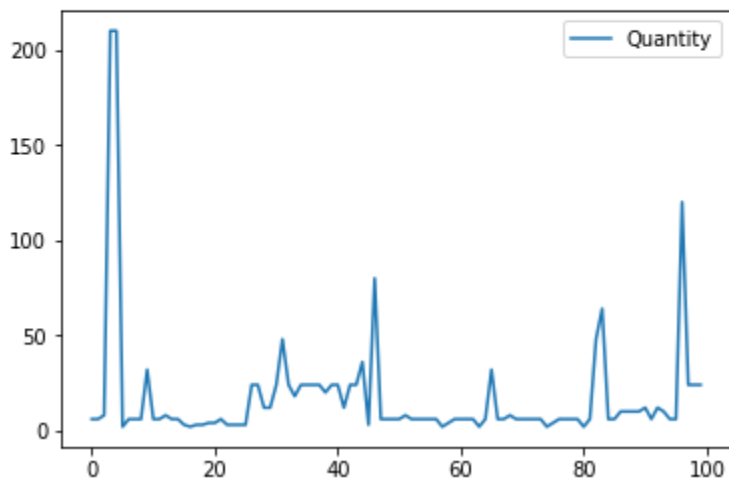
## Population by Province/Territory Canada, 2011 Census



Fastest Cars in the World				
Koenigsegg		Hennessey		Vector Avtech WX8 Hypercar
One	Agera R	Venom GT	Venom GT Spyder	Bugatti Veyron 16.4 Super Sport
Agera R Oman	Agera S	Porsche 9ff GT9-R	911 GT2: 414 km/h	SSC Ultimate Aero TT

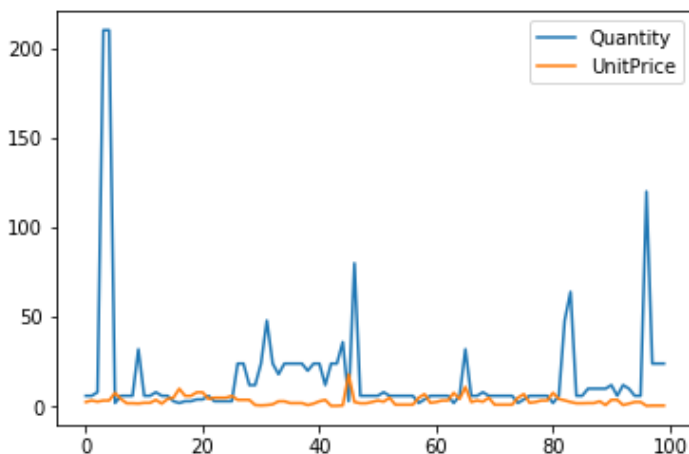
```
In [76]: d8 = pd.DataFrame(df, columns=['Quantity'])[0:100]
d8.plot()
```

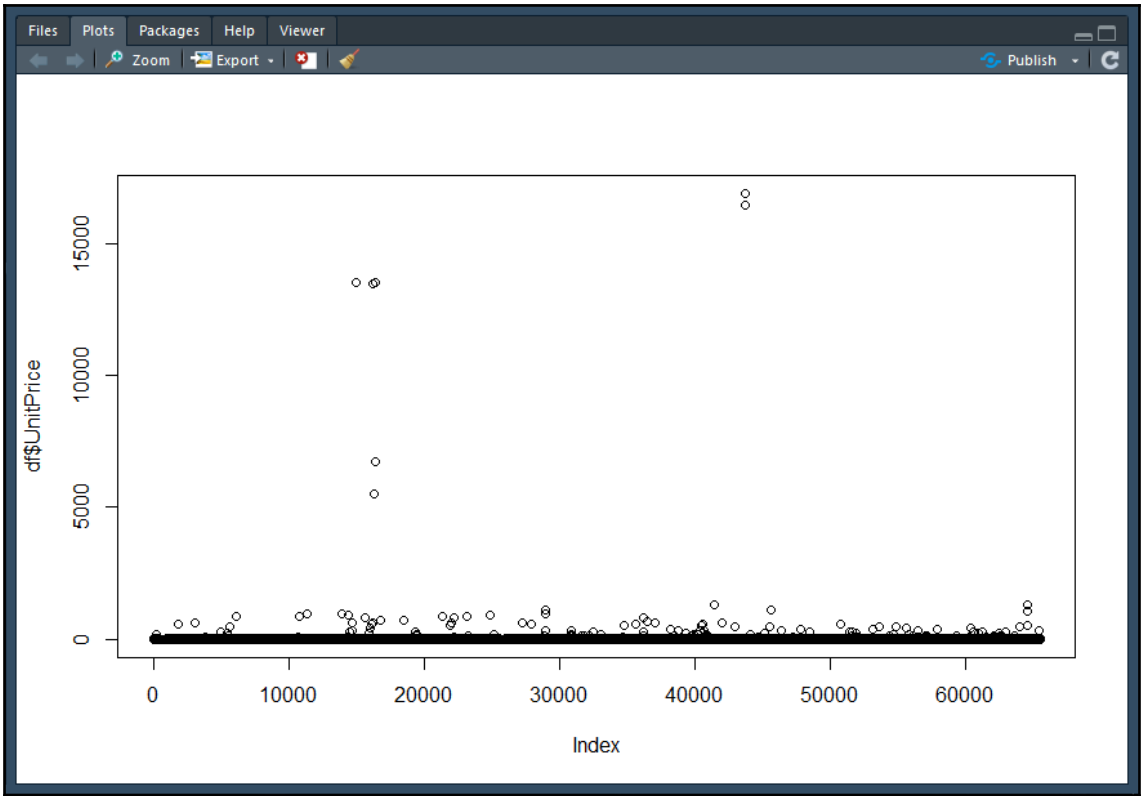
```
Out[76]: <matplotlib.axes._subplots.AxesSubplot at 0x1b39e3aeb00>
```

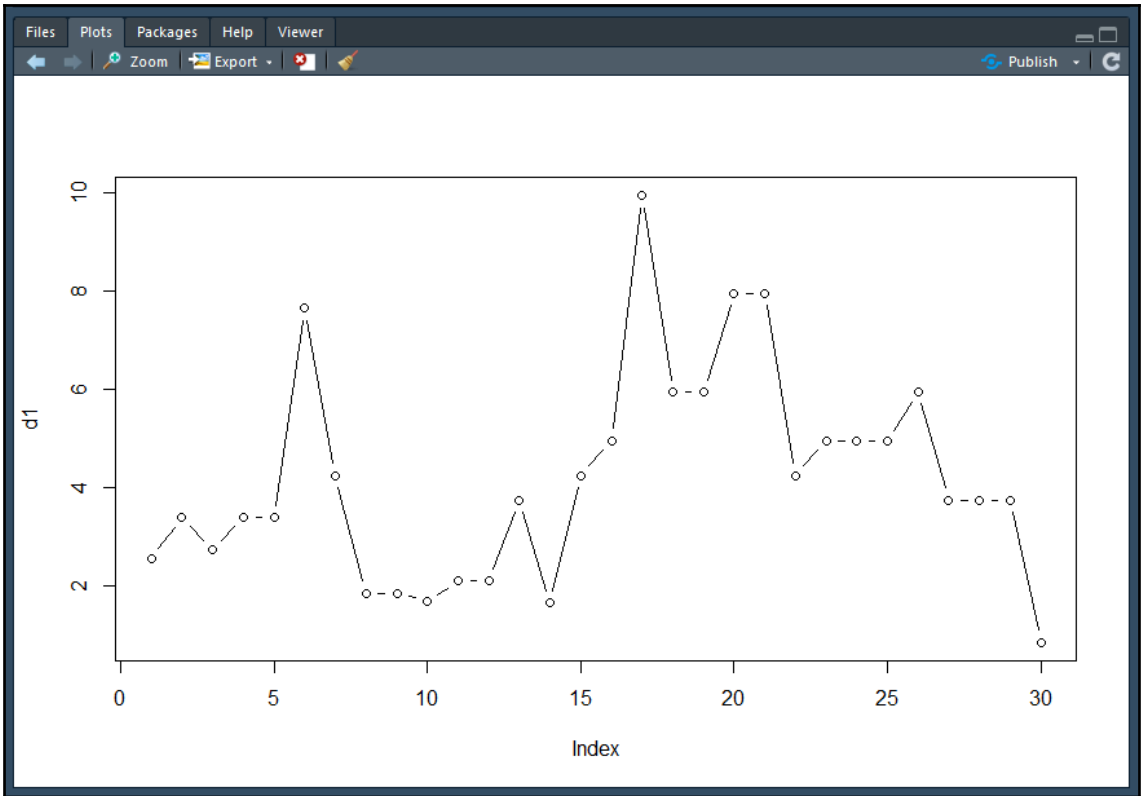


```
In [82]: d8 = pd.DataFrame(df, columns=['Quantity', 'UnitPrice'])[0:100]
d8.plot()
```

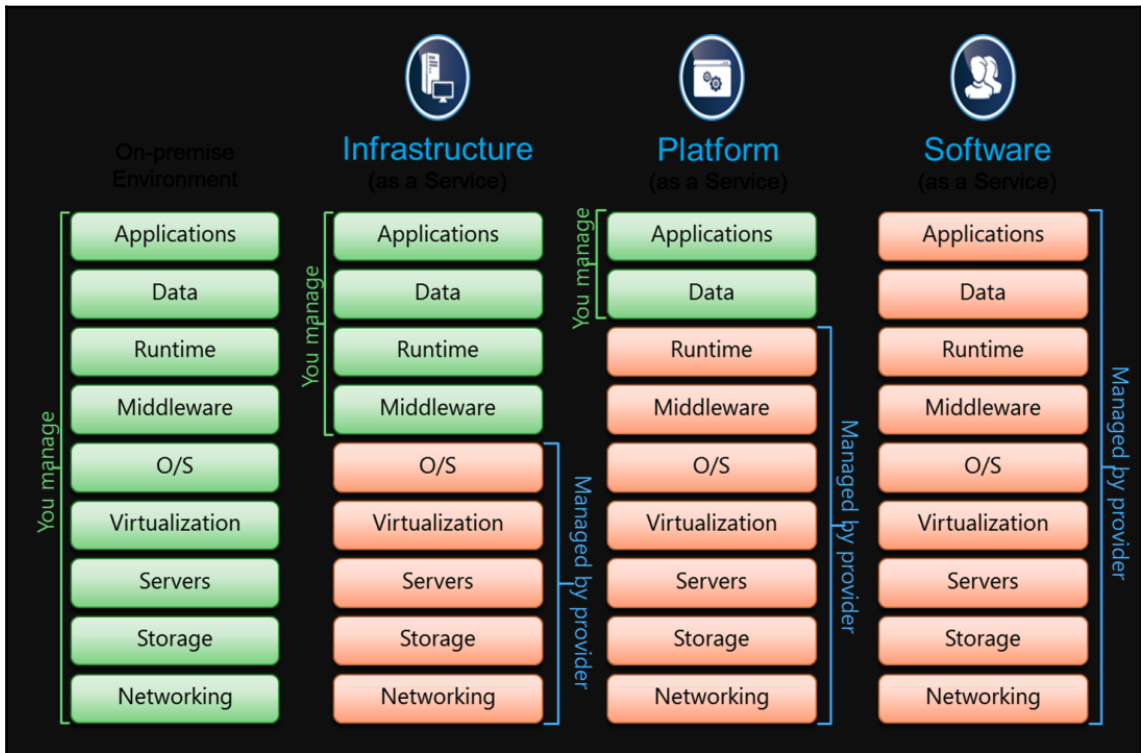
```
Out[82]: <matplotlib.axes._subplots.AxesSubplot at 0x1b39f21ef98>
```

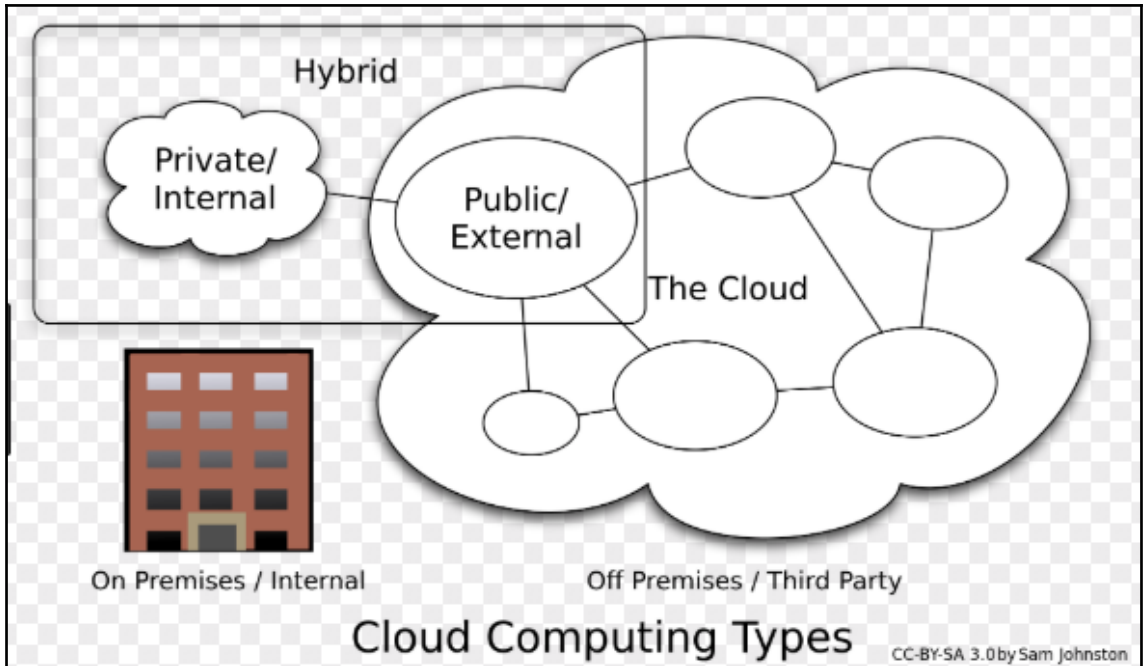






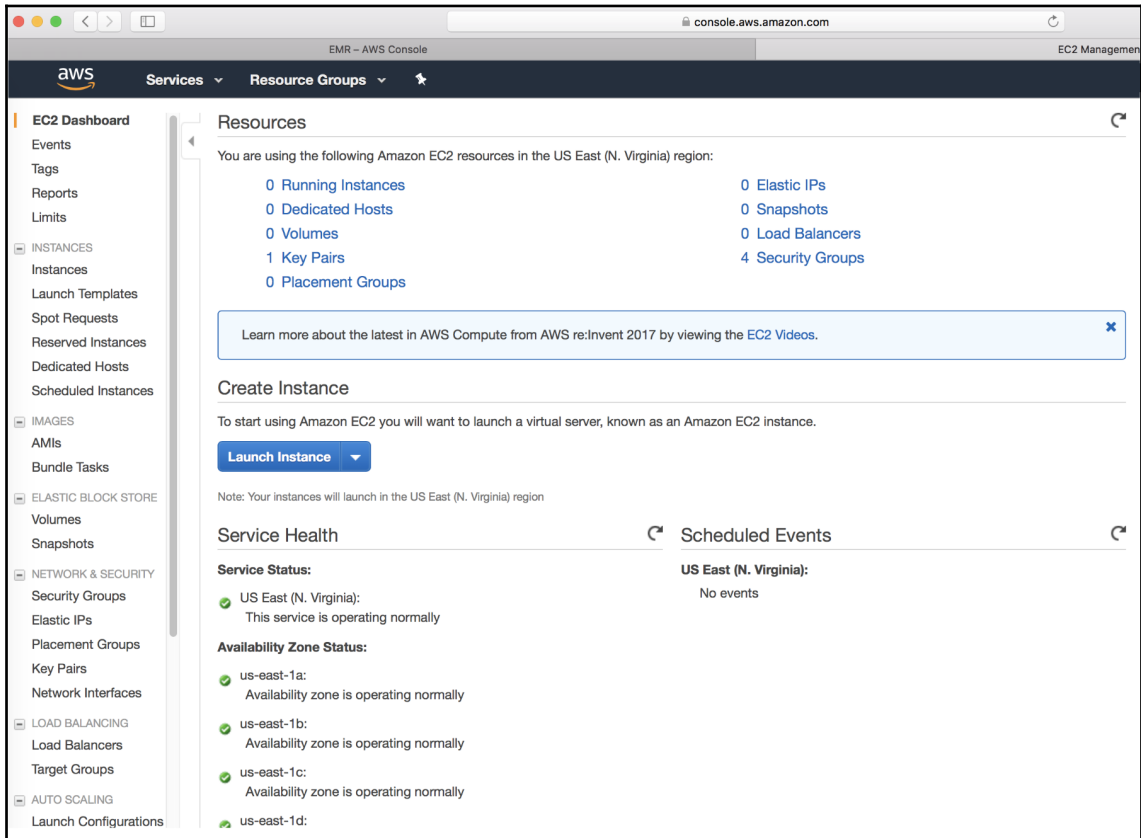
# Chapter 11: Introduction to Cloud Computing



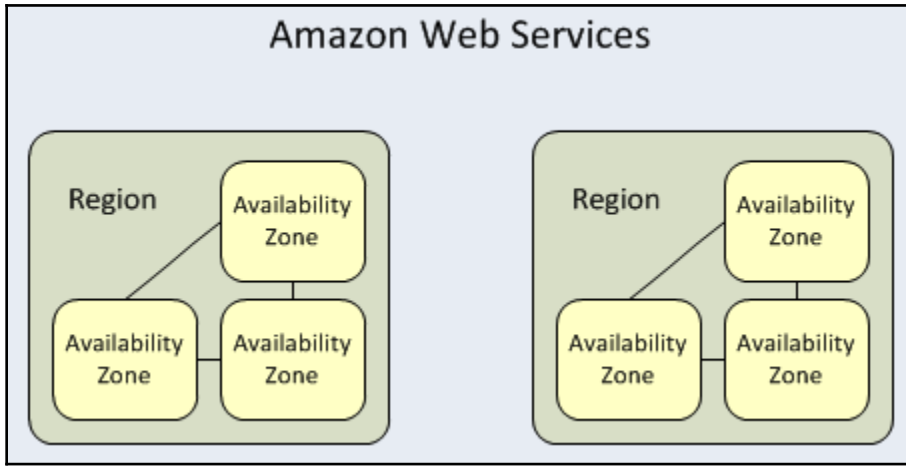




# Chapter 12: Using Amazon Web Services



# Amazon Web Services



The screenshot shows the AWS Management Console interface. At the top, there is a navigation bar with the AWS logo, 'Services', 'Resource Groups', and user information for 'SRIDHAR ALLA' in 'N. Virginia'. The main content area is divided into several sections:

- AWS services:** A search bar and a list of 'Recently visited services' including EC2, EMR, and Billing.
- Build a solution:** A section titled 'Get started with simple wizards and automated workflows.' containing six quick-start options: 'Launch a virtual machine' (With EC2, ~2-3 minutes), 'Build a web app' (With Elastic Beanstalk, ~6 minutes), 'Build using virtual servers' (With Lightsail, ~1-2 minutes), 'Connect an IoT device' (With AWS IoT, ~5 minutes), 'Start a development project' (With CodeStar, ~5 minutes), and 'Register a domain' (With Route 53, ~3 minutes).
- Learn to build:** A section titled 'Learn to deploy your solutions through step-by-step guides, labs, and videos.' with sub-sections for 'Websites', 'DevOps', and 'Backup and recovery'.
- Helpful tips:** A section with three tips: 'Manage your costs', 'Create an organization', and 'Explore AWS'.

console.aws.amazon.com

Services Resource Groups

SRIDHAR ALLA N. Virginia


## Welcome to Amazon Elastic MapReduce

Amazon Elastic MapReduce (Amazon EMR) is a web service that enables businesses, researchers, data analysts, and developers to easily and cost-effectively process vast amounts of data.

You do not appear to have any clusters. Create one now:

[Create cluster](#)


### How Elastic MapReduce Works



**Upload**

Upload your data and processing application to S3.


[Learn more](#)



**Create**

Configure and create your cluster by specifying data inputs, outputs, cluster size, security settings, etc.

[Learn more](#)



**Monitor**

Monitor the health and progress of your cluster. Retrieve the output in S3.

[Learn more](#)

#### Additional Information

More about Elastic MapReduce

- [EMR overview](#)
- [FAQs](#)
- [Pricing](#)

More Help Using Elastic MapReduce

- [Forum](#)
- [Documentation](#)
- [Developer Guide](#)
- [API Reference](#)
- [EMR on GitHub](#)
- [Help portal](#)

console.aws.amazon.com

Services Resource Groups

SRIDHAR ALLA N. Vir

## Create Cluster - Quick Options [Go to advanced options](#)

### General configuration

**Cluster name**

**Logging** ⓘ

S3 folder

**Launch mode**  Cluster ⓘ  Step execution ⓘ

### Software configuration

**Release**  ⓘ

**Applications**

- Core Hadoop: Hadoop 2.8.3 with Ganglia 3.7.2, Hive 2.3.2, Hue 4.1.0, Mahout 0.13.0, Pig 0.17.0, and Tez 0.8.4
- HBase: HBase 1.4.2 with Ganglia 3.7.2, Hadoop 2.8.3, Hive 2.3.2, Hue 4.1.0, Phoenix 4.13.0, and ZooKeeper 3.4.10
- Presto: Presto 0.194 with Hadoop 2.8.3 HDFS and Hive 2.3.2 Metastore
- Spark: Spark 2.3.0 on Hadoop 2.8.3 YARN with Ganglia 3.7.2 and Zeppelin 0.7.3

Use AWS Glue Data Catalog for table metadata ⓘ

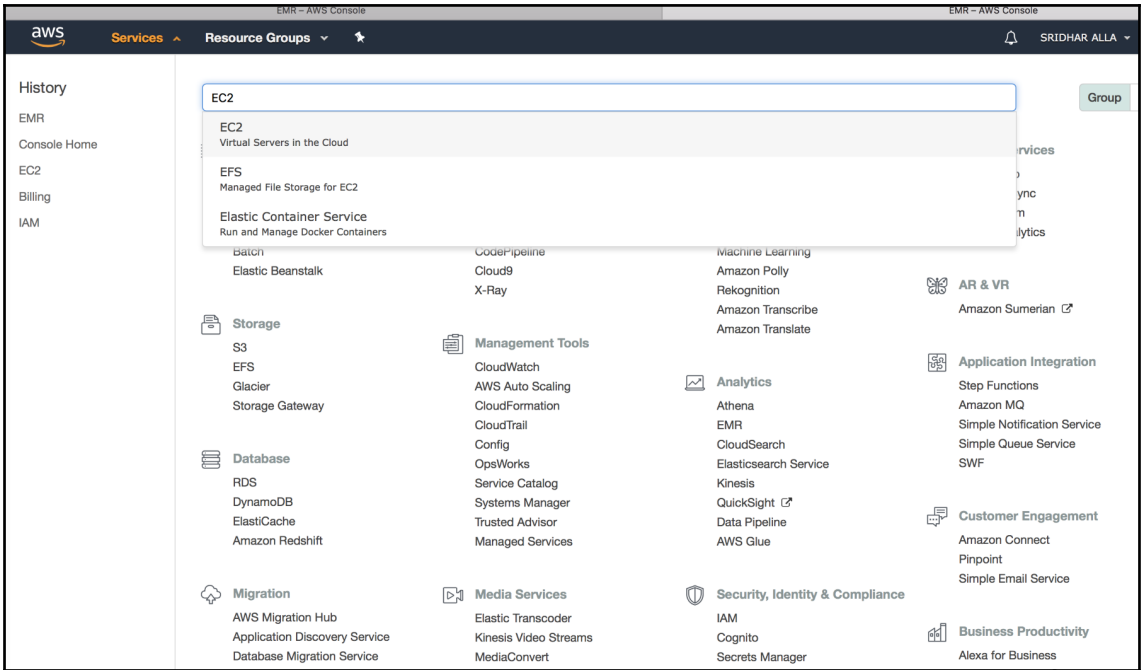
### Hardware configuration

**Instance type**  ⓘ

**Number of instances**  (1 master and 2 core nodes)

### Security and access

**EC2 key pair**  ⓘ [Learn how to create an EC2 key pair](#)



EMR - AWS Console console.aws.amazon.com EC2 Management

Services Resource Groups

### EC2 Dashboard

- Events
- Tags
- Reports
- Limits
- INSTANCES
  - Instances
  - Launch Templates
  - Spot Requests
  - Reserved Instances
  - Dedicated Hosts
  - Scheduled Instances
- IMAGES
  - AMIs
  - Bundle Tasks
- ELASTIC BLOCK STORE
  - Volumes
  - Snapshots
- NETWORK & SECURITY
  - Security Groups
  - Elastic IPs
  - Placement Groups
  - Key Pairs
  - Network Interfaces
- LOAD BALANCING
  - Load Balancers
  - Target Groups
- AUTO SCALING
  - Launch Configurations

## Resources

You are using the following Amazon EC2 resources in the US East (N. Virginia) region:

0 Running Instances	0 Elastic IPs
0 Dedicated Hosts	0 Snapshots
0 Volumes	0 Load Balancers
1 Key Pairs	4 Security Groups
0 Placement Groups	

Learn more about the latest in AWS Compute from AWS re:Invent 2017 by viewing the [EC2 Videos](#).

### Create Instance

To start using Amazon EC2 you will want to launch a virtual server, known as an Amazon EC2 instance.

[Launch Instance](#)

Note: Your instances will launch in the US East (N. Virginia) region

### Service Health

**Service Status:**

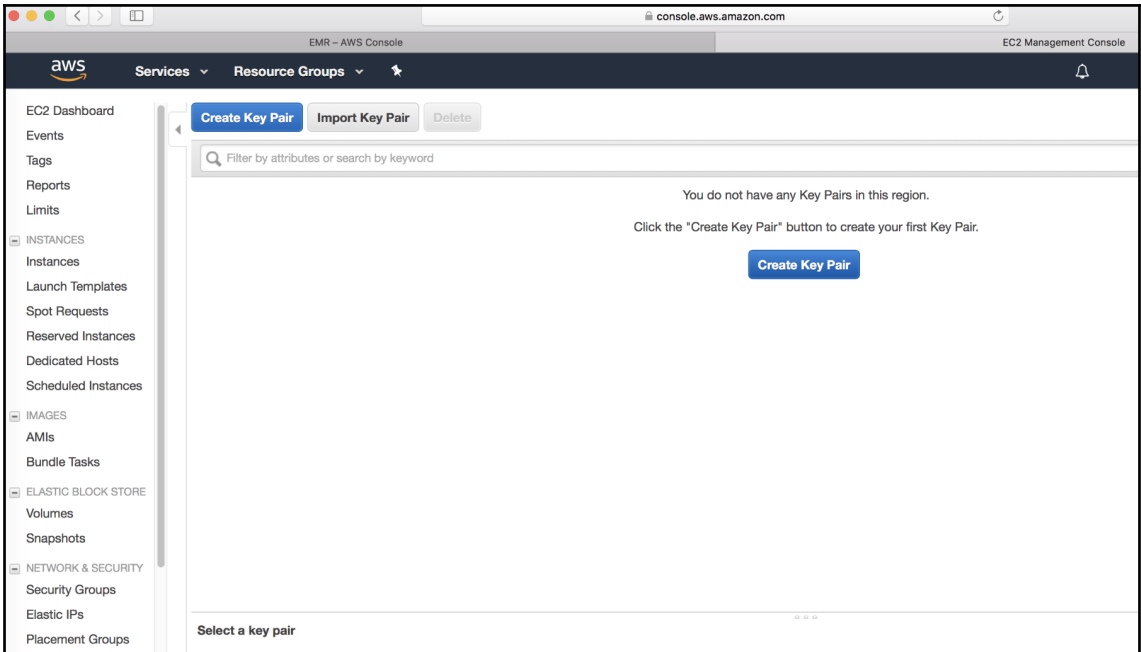
- US East (N. Virginia): This service is operating normally

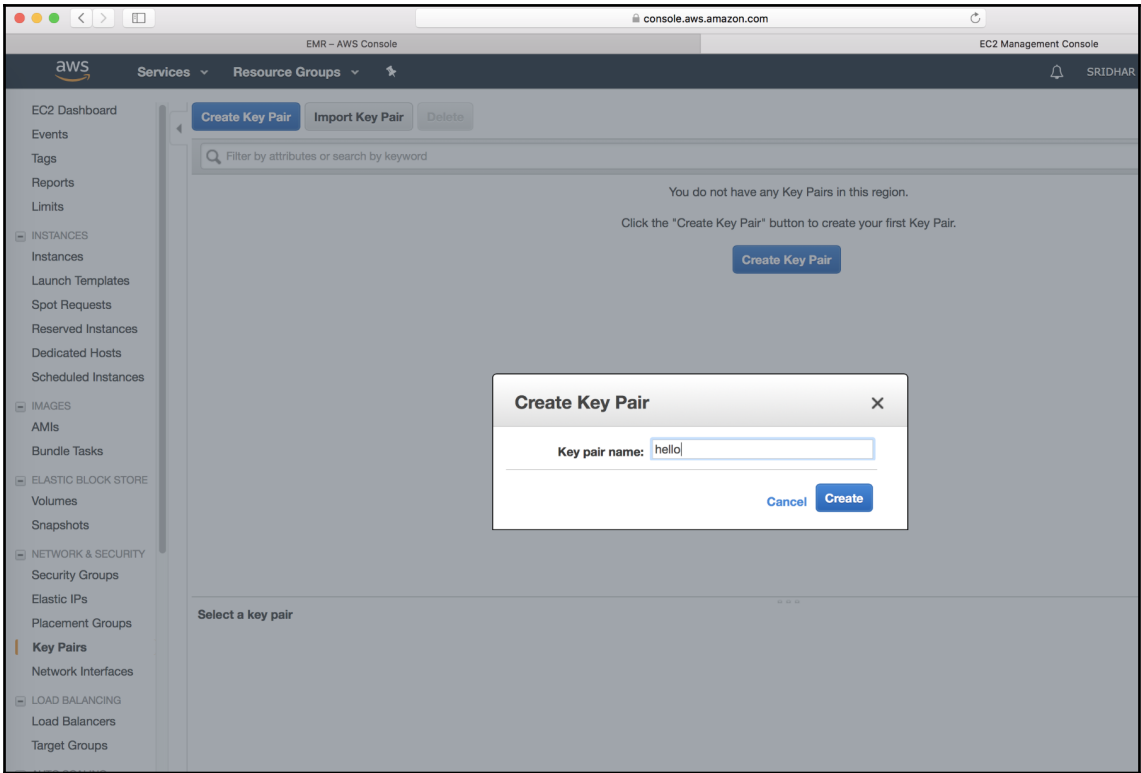
**Availability Zone Status:**

- us-east-1a: Availability zone is operating normally
- us-east-1b: Availability zone is operating normally
- us-east-1c: Availability zone is operating normally
- us-east-1d: Availability zone is operating normally

### Scheduled Events

**US East (N. Virginia):**  
No events





console.aws.amazon.com

EMR - AWS Console

EC2 Management Console

Services Resource Groups

EC2 Dashboard

Events

Tags

Reports

Limits

INSTANCES

Instances

Launch Templates

Spot Requests

Reserved Instances

Dedicated Hosts

Scheduled Instances

IMAGES

AMIs

Bundle Tasks

ELASTIC BLOCK STORE

Volumes

Snapshots

NETWORK & SECURITY

Security Groups

Elastic IPs

Placement Groups

Key Pairs

Network Interfaces

LOAD BALANCING

Load Balancers

Create Key Pair Import Key Pair Delete

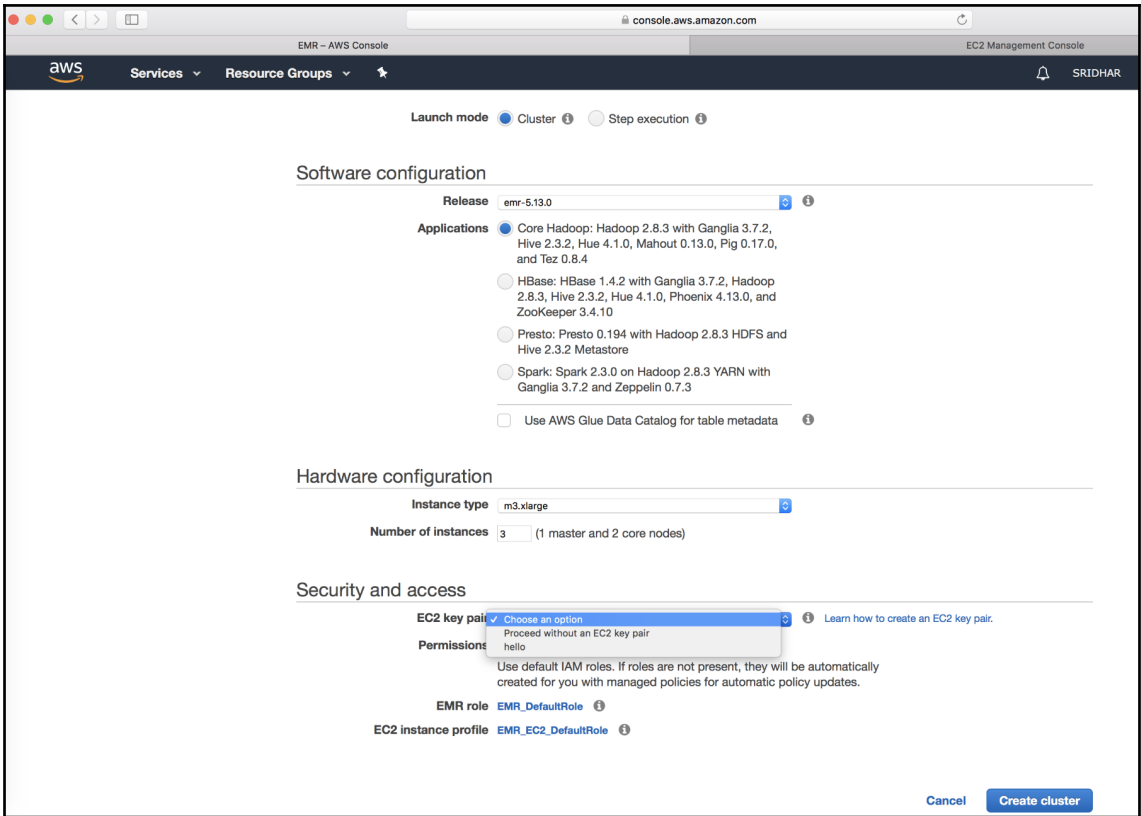
hello.pem.txt

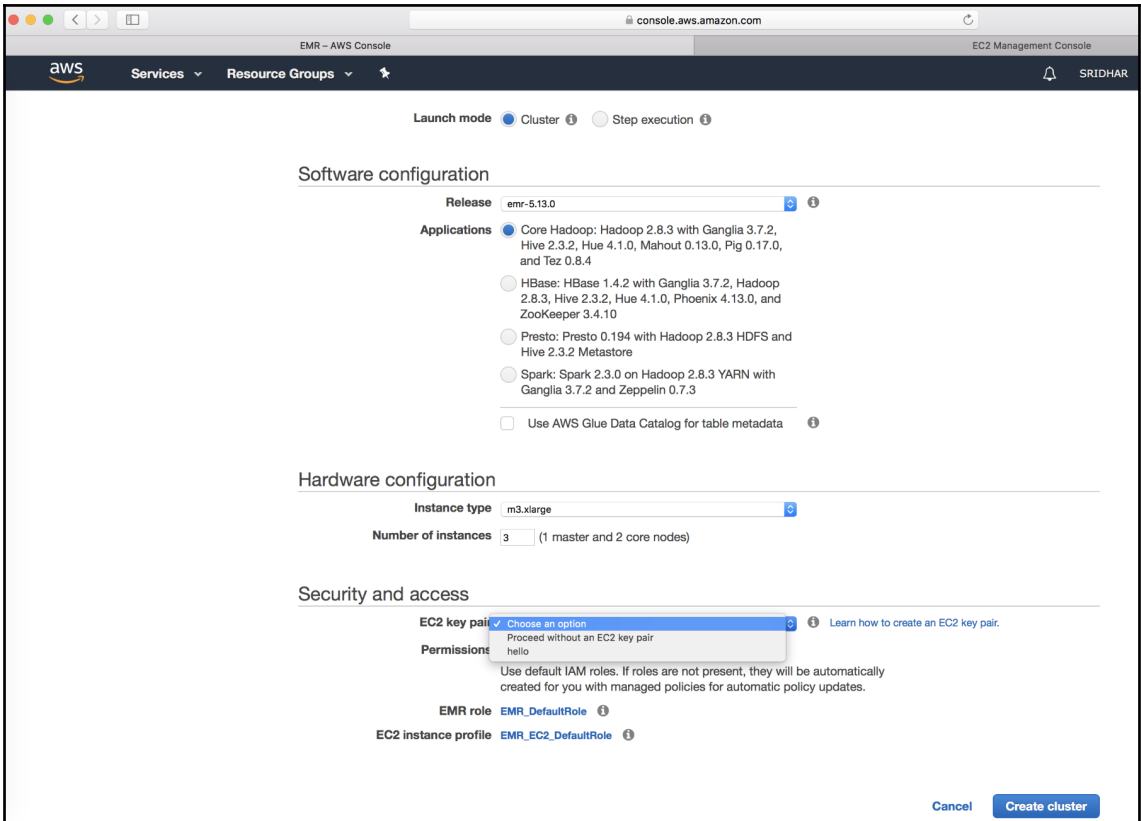
```
-----BEGIN RSA PRIVATE KEY-----
MIIEpAIBAAKCAQEAxMm6IpsZ54gS8TU0kuwhniyenMybtRtFz6y2F+oKwaN1o03qio19kJzCfV
+4fK08/DEB157nwy18fDu4xHgKs89vZ6FbCwLo5uE949xS1D2/GovfXR0r0+cMqyo5KH3/f
Aadrhmc987hnVx5FtdfAenwWbUvp18UNkLeq0vrgaY7188hnlnc1W70-3MP111Y1NC62E
1duiyDXW1HDguiekYrhnKGF1NoXMaqVLogSGOYJWkaM+1004ahRbh9U15Ibn+2X+1S+kJ/ThVn
dNS9b4f1b2FTHsLczaSkEayezzi1e6G3ha/Zf09Nby9kdT6krR01DA0ABaIBAD1Rcv9e6jUC
K/r85zpaM44d+w0w+i-nl30wNcbv1Xr5c7ZXB1d976JBeXGBRr3E3y1W710cmEyyvniE65
X+/1NC1mWZ2Hc10wuThfc160nUSXa5VrxSPdWTB1EHTav7585edX1j4f3wqFP9b306570kuB
Rzps5/rjUbcw02N16euCWH/gvZkhizT6Ponbyk056XC9wtmMOU1A/XWOM78GGAbBcTzjMRU0ZU
MgyC91wBpd1hbAqKfjP1Xag99008os/oFT8ahN87yusLe4M514e1nhQW9s9NHXuxz88MK+u
r0E0e+h+jZUL1dXfY76m58a5EcqYEA/KQRaUe011y1eQDF72K0f1REZC0f188931kPZ
M2gtK0wL1t31la5ujtc9vcdsDuE1b4FmsMdwNcM75tUHTFCa14JEvU3oDDFjB1CndHF8PIH
YMNcLnL5+3L4x63CFH2ZEK0A1vVvK+kr8AqjKocRGe504Y68CgYExas5km3c2189PdZ1Va6d
5A1Tnux+A2K18z3pJUZ2u59N2zdx+wk3D2w/gwY1NwUcn01MPfE/LtXX0E1pEZACRkQUKkK
A4263/N93m4dZeyoYCV1qKXU2f03pEcu+krn/LtqDrFM40LjdfhxZGCK5L5MDBYdKvH+GsuUsC
gYEAvm7FNqAD0/H2MNXJoIwbnPNT1q1zKxm1UeCE8UBzYnMavFn0/yjBUJvWxHwxe/Y9/gIw
Sk19JYcIno4FD83FvKA+99Fev42rwn1j1ko1Demu1P3n0/uKE1eV8NA5E0wLVN0HteK005
rgEw0e182f/ey94211h3pucq03unc1mY6c5xUgFb1vz0Du31rrf8m2140Nz+9591ZolkyC
d/b0C7+1hY0eJm/8KPs1Ice0ha01uidK4qx61cv1WSG0ca7AE4A+G1Hx2Pbx1No0evX0LeAZ7y
Umz1L1b2Tki0Hh0Duraq+Kc2NYPsn610bK801HHTFpLo0Kba0CCEHhJ0geV93K0bXV20vbcZBo
qKCYU75011Um/1BA1v1kv0FS11Cvz20h12/MDTJNP1caEnnU26YME39xPeJmt0meK02KcyCm
150fZwB9vR0tCu+QhZ8P1MKVrXF87Jatvzr01uz+XN11tL1L2L5pLcS/VD0wMecpIKYA=
-----END RSA PRIVATE KEY-----
```

Key Pair: hello

Key pair name	hello
Fingerprint	b7:1f24:7e:1a:5e:6e:b3:2b:31:3b:e2:5c:43:b2:08:c6:d6:36:67







EMR - AWS Console console.aws.amazon.com EC2 Management Console

Services Resource Groups SRIDHAR ALLA

Launch mode  Cluster  Step execution

### Software configuration

Release emr-5.13.0

Applications

- Core Hadoop: Hadoop 2.8.3 with Ganglia 3.7.2, Hive 2.3.2, Hue 4.1.0, Mahout 0.13.0, Pig 0.17.0, and Tez 0.8.4
- HBase: HBase 1.4.2 with Ganglia 3.7.2, Hadoop 2.8.3, Hive 2.3.2, Hue 4.1.0, Phoenix 4.13.0, and ZooKeeper 3.4.10
- Presto: Presto 0.194 with Hadoop 2.8.3 HDFS and Hive 2.3.2 Metastore
- Spark: Spark 2.3.0 on Hadoop 2.8.3 YARN with Ganglia 3.7.2 and Zeppelin 0.7.3
- Use AWS Glue Data Catalog for table metadata

### Hardware configuration

Instance type m3.xlarge

Number of instances 3 (1 master and 2 core nodes)

### Security and access

EC2 key pair hello

Permissions  Default  Custom

Use default IAM roles. If roles are not present, they will be automatically created for you with managed policies for automatic policy updates.

EMR role EMR\_DefaultRole

EC2 instance profile EMR\_EC2\_DefaultRole

Cancel Create cluster

Amazon EMR console showing cluster details for "My cluster" in a "Starting" state.

**Cluster: My cluster** Starting

Buttons: Clone, Terminate, AWS CLI export

Navigation tabs: Summary, Application history, Monitoring, Hardware, Events, Steps, Configurations, Bootstrap actions

Summary details:

- ID: j-2VGNZ5HWCBXM6
- Creation date: 2018-04-21 20:47 (UTC-4)
- Elapsed time: 0 seconds
- Auto-terminate: No
- Termination protection: Off [Change](#)

Configuration details:

- Release label: emr-5.13.0
- Hadoop distribution: Amazon 2.8.3
- Applications: Ganglia 3.7.2, Hive 2.3.2, Hue 4.1.0, Mahout 0.13.0, Pig 0.17.0, Tez 0.8.4
- Log URI: s3://aws-logs-180072566896-us-east-1/elasticmapreduce/
- EMRFS consistent view: Disabled
- Custom AMI ID: --

Network and hardware:

- Availability zone: --
- Subnet ID: subnet-b553e29f
- Master: Provisioning 1 m3.xlarge
- Core: Provisioning 2 m3.xlarge
- Task: --

Security and access:

- Key name: hello
- EC2 instance profile: EMR\_EC2\_DefaultRole
- EMR role: EMR\_DefaultRole
- Visible to all users: All [Change](#)
- Security groups for Master: --
- Security groups for Core & Task: --

Amazon EMR console showing cluster details for "My cluster" in a "Waiting" state.

**Cluster: My cluster** Waiting Cluster ready after last step completed.

Buttons: Clone, Terminate, AWS CLI export

Navigation tabs: Summary, Application history, Monitoring, Hardware, Events, Steps, Configurations, Bootstrap actions

Buttons: Add task instance group

Instance groups: 2 instance groups (all loaded)

ID	Status	Node type & name	Instance type	Instance count	Purchasing option
ig-2DZQGYXDH4GOP	Running	CORE Core Instance Group	m3.xlarge 8 vCore, 15 GiB memory, 80 SSD GB storage EBS Storage: none	2 Instances <a href="#">Resize</a>	On-demand
ig-1DEJ243T1LBCK	Running	MASTER Master Instance Group	m3.xlarge 8 vCore, 15 GiB memory, 80 SSD GB storage EBS Storage: none	1 Instances	On-demand

Clone Terminate AWS CLI export

Cluster: My cluster Waiting Cluster ready after last step completed.

Summary Application history Monitoring Hardware Events Steps Configurations Bootstrap actions

Time	Event description	Source ID	Source type	Event type	Severity	Full date & time
Apr 21 09:03 PM	Amazon EMR cluster j-2VGNZ5HWCBOXM6 (My cluster) finished running all pending steps at 2018-04-22 01:02 UTC.	j-2VGNZ5HWCBOXM6	Cluster	Cluster State Change	INFO	April 21, 2018 at 09:03:38 PM (UT)
Apr 21 09:03 PM	Step s-1KJJF52B8QYQU (Setup hadoop debugging) in Amazon EMR cluster j-2VGNZ5HWCBOXM6 (My cluster) completed execution at 2018-04-22 01:02 UTC. The step started running at 2018-04-22 01:02 UTC and took 0 minutes to complete.	s-1KJJF52B8QYQU	Step	Step State Change	INFO	April 21, 2018 at 09:03:36 PM (UT)
Apr 21 09:02 PM	Amazon EMR cluster j-2VGNZ5HWCBOXM6 (My cluster) began running steps at 2018-04-22 01:01 UTC.	j-2VGNZ5HWCBOXM6	Cluster	Cluster State Change	INFO	April 21, 2018 at 09:02:49 PM (UT)
Apr 21 09:02 PM	Step s-1KJJF52B8QYQU (Setup hadoop debugging) in Amazon EMR cluster j-2VGNZ5HWCBOXM6 (My cluster) started running at 2018-04-22 01:02 UTC.	s-1KJJF52B8QYQU	Step	Step State Change	INFO	April 21, 2018 at 09:02:49 PM (UT)
Apr 21 08:47 PM	Amazon EMR cluster j-2VGNZ5HWCBOXM6 (My cluster) was requested at 2018-04-22 00:47 UTC and is being created.	j-2VGNZ5HWCBOXM6	Cluster	Cluster State Change	INFO	April 21, 2018 at 08:47:27 PM (UT)

aws Services Resource Groups

EC2 Dashboard

Events Tags Reports Limits INSTANCES Instances Launch Templates Spot Requests Reserved Instances Dedicated Hosts Scheduled Instances IMAGES AMIs Bundle Tasks ELASTIC BLOCK STORE Volumes Snapshots NETWORK & SECURITY Security Groups Elastic IPs

Create Security Group Actions

search: sg-67b4db2e Add filter

Name	Group ID	Group Name	VPC ID	Description
<input type="checkbox"/>	sg-67b4db2e	ElasticMapReduce-master	vpc-dcbe84b8	Master group for Elastic MapReduce created on 2018-04-21T21:42:09.901Z
<input type="checkbox"/>	sg-fcb8d9b5	ElasticMapReduce-slave	vpc-dcbe84b8	Slave group for Elastic MapReduce created on 2018-04-21T22:11:07.797Z

Select a security group above

aws Services Resource Groups SRIDHAR ALLA N. Virginia

EC2 Dashboard Create Security Group Actions

Events  
Tags  
Reports  
Limits

INSTANCES  
Instances  
Launch Templates  
Spot Requests  
Reserved Instances  
Dedicated Hosts  
Scheduled Instances

IMAGES  
AMIs  
Bundle Tasks

ELASTIC BLOCK STORE  
Volumes  
Snapshots

NETWORK & SECURITY  
**Security Groups**  
Elastic IPs  
Placement Groups  
Key Pairs  
Network Interfaces

LOAD BALANCING

search: sg-67b4db2e Add filter

Name	Group ID	Group Name	VPC ID	Description
<input checked="" type="checkbox"/>	sg-67b4db2e	ElasticMapReduce-master	vpc-dcbe84b8	Master group for Elastic MapReduce created on 2018-04-21T21:42:09.901Z
<input type="checkbox"/>	sg-fcb6d9b5	ElasticMapReduce-slave	vpc-dcbe84b8	Slave group for Elastic MapReduce created on 2018-04-21T22:11:07.797Z


**Security Group: sg-67b4db2e**

Description Inbound Outbound Tags

Edit

Type	Protocol	Port Range	Source	Description
All TCP	TCP	0 - 65535	0.0.0.0/0	
All TCP	TCP	0 - 65535	:::0	
All TCP	TCP	0 - 65535	sg-fcb6d9b5 (ElasticMapReduce-slave)	

54.172.107.83:8088/cluster Logged in as: dr.who

 **All Applications**

Cluster

- About
- Nodes
- Node Labels
- Applications
- NEW
- SAVING
- SUBMITTED
- ACCEPTED
- RUNNING
- FINISHED
- FAILED
- KILLED
- Scheduler

Tools

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved
0	0	0	0	0	0 B	22.50 GB	0 B	0	16	0

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes	Shutdown Nodes
2	0	0	0	0	0	0

Scheduler Metrics


Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation	Maximum Cluster Application Priority
Capacity Scheduler	[MEMORY]	<memory:32, vCores:1>	<memory:11520, vCores:8>	0

Show 20 entries

ID	User	Name	Application Type	Queue	Application Priority	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCores	Allocated Memory MB	% of Queue	% of Cluster	Progress	Tracking UI	Blacklisted Nodes
No data available in table																	

Showing 0 to 0 of 0 entries First Previous Next Last

54.172.107.83:8088/cluster/scheduler Logged in as: dr.who



## NEW,NEW\_SAVING,SUBMITTED,ACCEPTED,RUNNING Applications

Cluster

- About
- Nodes
- Node Labels
- Applications
- NEW
- NEW\_SAVING
- SUBMITTED
- ACCEPTED
- RUNNING
- FINISHED
- FAILED
- KILLED

Scheduler

Tools

### Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved
0	0	0	0	0	0 B	22.50 GB	0 B	0	16	0

### Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes	Shutdown Nodes
2	0	0	0	0	0	0

### Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation	Maximum Cluster Application Priority
Capacity Scheduler	[MEMORY]	<memory:32, vCores:1>	<memory:11520, vCores:8>	0

[Dump scheduler logs](#) 1 min

### Application Queues

**Legend:** Capacity Used Used (over capacity) Max Capacity Users Requesting Resources

- Queue: root 0.0% used
- + Queue: default 0.0% used

Show 20 entries Search:

ID	User	Name	Application Type	Queue	Application Priority	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCores	Allocated Memory MB	% of Queue	% of Cluster	Progress	Tracking UI	Blacklisted Nodes
No data available in table																	

Showing 0 to 0 of 0 entries First Previous Next Last

## Overview 'ip-172-30-0-72.ec2.internal:8020' (active)

<b>Started:</b>	Sat Apr 21 20:55:44 -0400 2018
<b>Version:</b>	2.8.3-amzn-0, r300391394352b074b85b529e870816a72c6f314a
<b>Compiled:</b>	Wed Mar 21 19:38:00 -0400 2018 by ec2-user from (HEAD detached at 3003913943)
<b>Cluster ID:</b>	CID-530a38e7-16f8-4b40-9e61-bce472a2874d
<b>Block Pool ID:</b>	BP-1513736012-172.30.0.72-1524358539906

## Summary

Security is off.

Safemode is off.

1,125 files and directories, 1,088 blocks = 2,213 total filesystem object(s).

Heap Memory used 301.12 MB of 529 MB Heap Memory. Max Heap Memory is 1.51 GB.

Non Heap Memory used 60.06 MB of 61.31 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

<b>Configured Capacity:</b>	137.9 GB
<b>DFS Used:</b>	798.88 MB (0.57%)
<b>Non DFS Used:</b>	0 B
<b>DFS Remaining:</b>	137.12 GB (99.43%)
<b>Block Pool Used:</b>	798.88 MB (0.57%)
<b>DataNodes usages% (Min/Median/Max/stdDev):</b>	0.51% / 0.62% / 0.62% / 0.06%
<b>Live Nodes</b>	2 (Decommissioned: 0)



54.172.107.83:50070/dfshealth.html#tab-datanode

**Hadoop** Overview **Datanodes** Datanode Volume Failures Snapshot Startup Progress Utilities

## Datanode Information

✔ In service 
 ● Down 
 ✂ Decommissioned 
 ⊖ Decommissioned & dead

### Datanode usage histogram

Disk usage of each DataNode (%)

### In operation

Show  entries Search:

Node	Http Address	Last contact	Capacity	Blocks	Block pool used	Version
✔ ip-172-30-0-172.ec2.internal:50010 (172.30.0.172:50010)	<a href="http://ip-172-30-0-172.ec2.internal:50075">http://ip-172-30-0-172.ec2.internal:50075</a>	1s	68.95 GB <div style="width: 100%;"></div>	543	360.09 MB (0.51%)	2.8.3-amzn-0
✔ ip-172-30-0-83.ec2.internal:50010 (172.30.0.83:50010)	<a href="http://ip-172-30-0-83.ec2.internal:50075">http://ip-172-30-0-83.ec2.internal:50075</a>	1s	68.95 GB <div style="width: 100%;"></div>	545	438.79 MB (0.62%)	2.8.3-amzn-0

Showing 1 to 2 of 2 entries

Previous **1** Next

54.172.107.83:50070/explorer.html#

**Hadoop** Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

## Browse Directory

/

Show  entries Search:

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	hdfs	hadoop	0 B	Apr 21 20:55	0	0 B	apps
drwxr-xr-x	hdfs	hadoop	0 B	Apr 21 20:55	0	0 B	user
drwxr-xr-x	hdfs	hadoop	0 B	Apr 21 20:55	0	0 B	var
drwxrwxrwt	hdfs	hadoop	0 B	Apr 21 21:00	0	0 B	tmp

Showing 1 to 4 of 4 entries

Previous **1** Next

Hadoop, 2017.