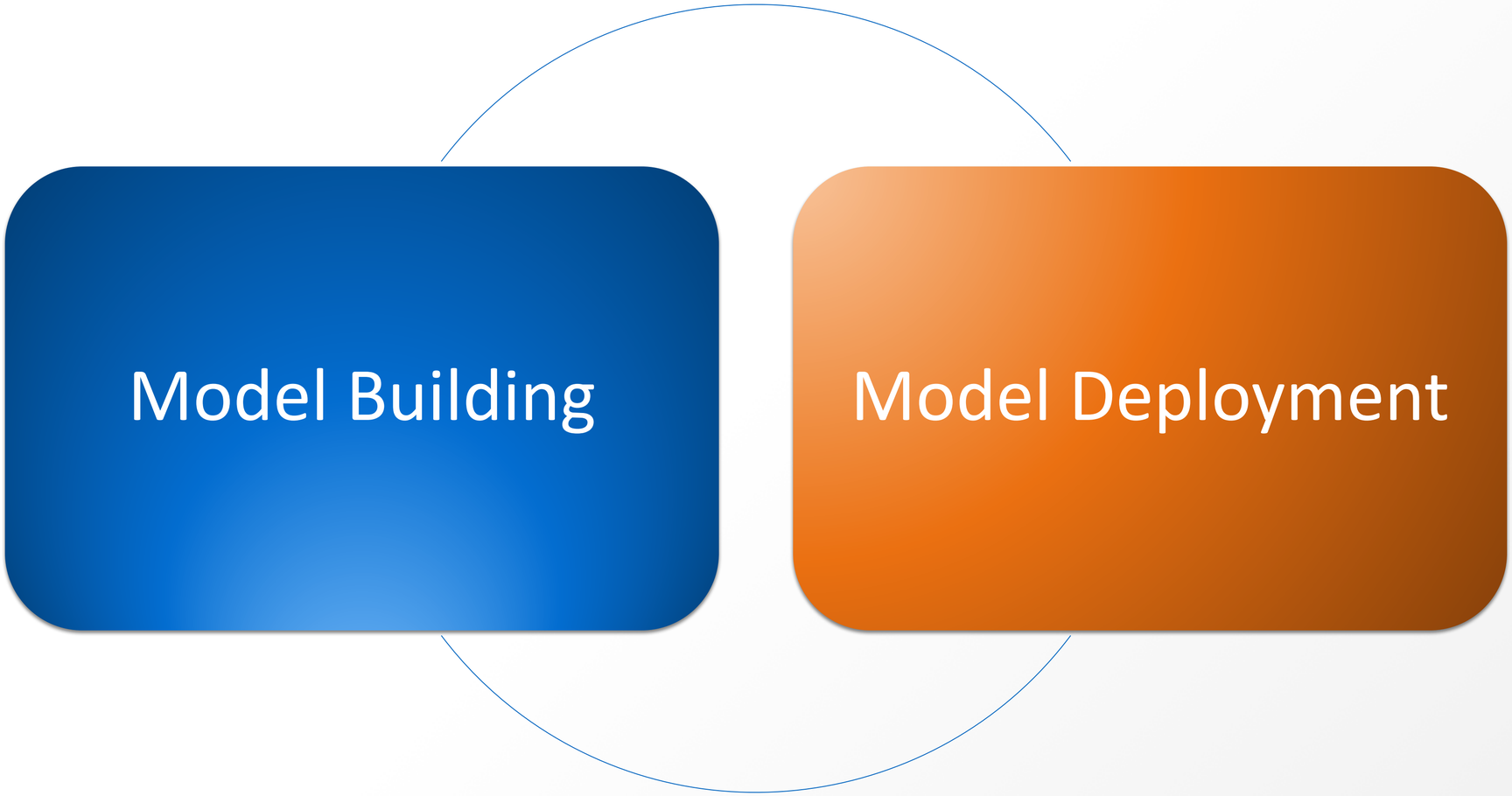
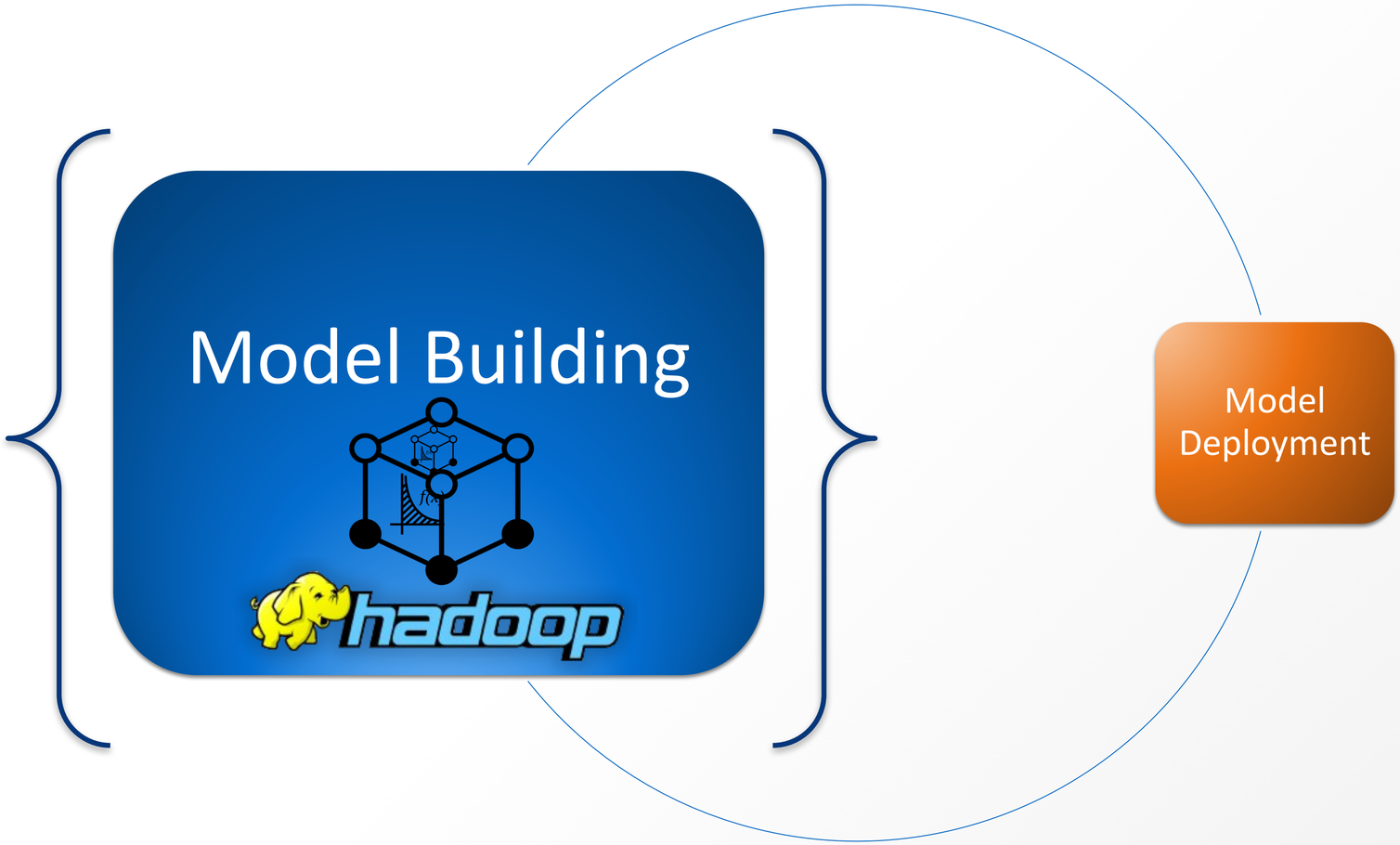
H2O in Hadoop Environment

avkash@h2o.ai

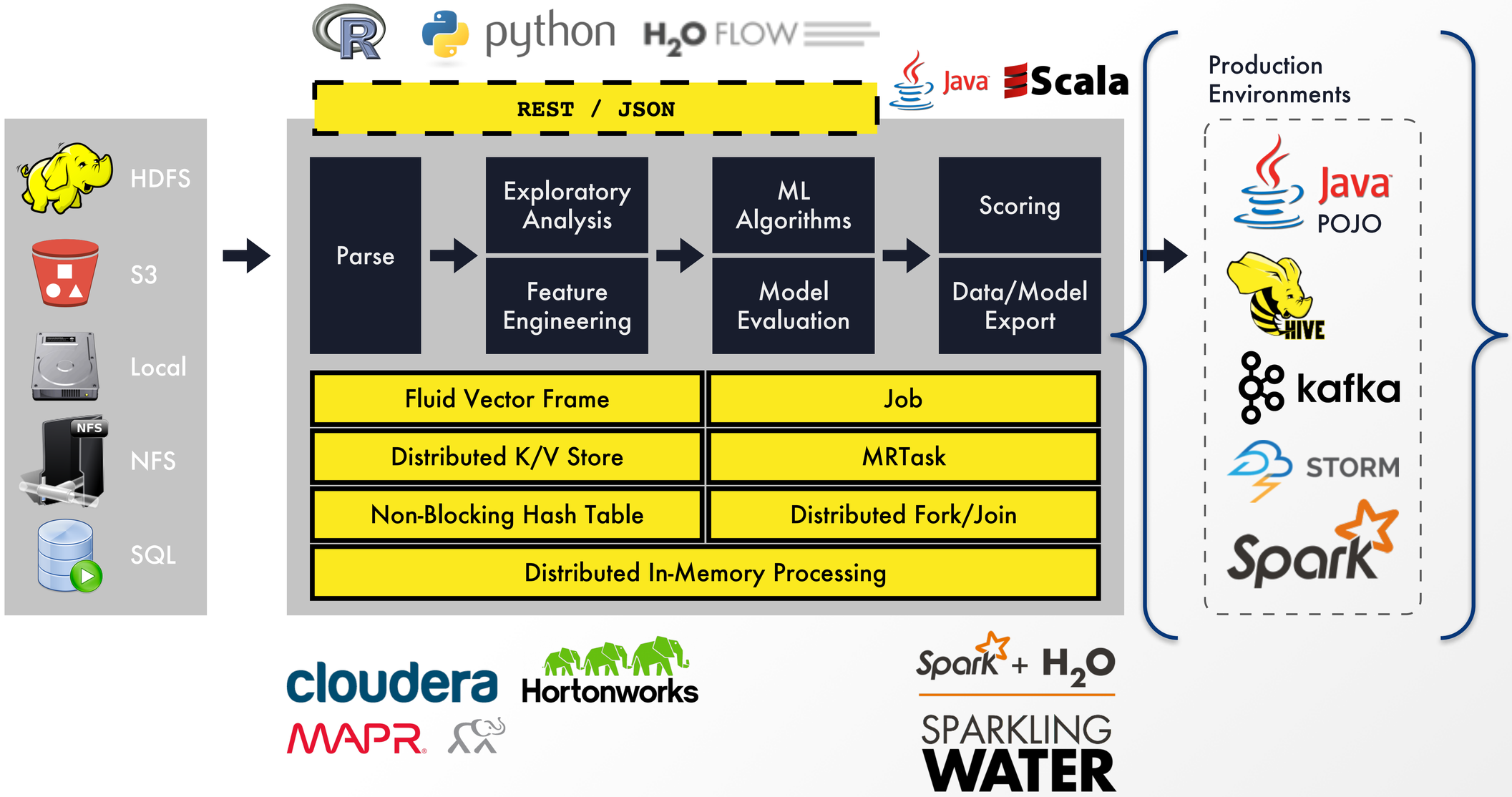
Machine Learning



Machine Learning in Hadoop with H2O



Core H2O Architecture



Get to know your Hadoop environment

- **$ hadoop version**
  - **Hadoop 2.6.0-cdh5.8.4Subversion**
  - http://github.com/cloudera/hadoop -r 4119786c06a223a58e9d0d0e0eb57491f7761c35Compiled by jenkins on 2017-02-06T22:27ZCompiled with protoc 2.5.0From source with checksum 4f6d6f3956e2822b5812ea0c1b37bc1This command was run using **/opt/cloudera/parcels/CDH-5.8.4-1.cdh5.8.4.p0.5/jars/hadoop-common-2.6.0-cdh5.8.4.jar**
- **Get proper H2O driver for Hadoop**
  - **$ wget wget https://h2o-release.s3.amazonaws.com/h2o/rel-weierstrass/2/h2o-3.14.0.2-cdh5.8.zip**

Make sure you have access to user folder

- **:~/h2o-3.14.0.2-cdh5.8$ hadoop fs -ls /user/avkash**

- Found 10 items
- drwx—— - avkash avkash 0 2017-09-15 17:00 /user/avkash/.Trash
- drwxr-xr-x - avkash avkash 0 2016-12-29 00:09 /user/avkash/.sparkStaging
- drwx—— - avkash avkash 0 2017-09-17 10:15 /user/avkash/.staging
- drwxr-xr-x - avkash avkash 0 2017-09-17 10:15 /user/avkash/001
- -rw-r–r– 3 avkash avkash 71560 2017-09-13 17:46 /user/avkash/ScoreData-1.0-SNAPSHOT.jar
- -rw-r–r– 3 avkash avkash 1463376 2017-09-13 16:56 /user/avkash/TTS-XRE-test1.parquet
- drwxr-xr-x - avkash hdfs 0 2017-06-19 12:31 /user/avkash/UDFtest
- -rw-r–r– 3 avkash avkash 7030878 2017-09-13 17:45 /user/avkash/h2o-genmodel.jar
- drwxr-xr-x - avkash avkash 0 2017-09-13 17:36 /user/avkash/prostate
- -rw-r–r– 3 avkash avkash 9254 2017-09-13 17:33 /user/avkash/prostate.csv

Know where you H2O driver is

- **:~/h2o-3.14.0.2-cdh5.8$ ll**

- total 96596drwxr-xr-x 5 avkash avkash 4096 Aug 21 22:32 ./
- drwxr-xr-x 5 avkash avkash 4096 Sep 17 10:16 ../
- drwxr-xr-x 3 avkash avkash 4096 Aug 21 22:32 bindings/
- -rw-r–r– 1 avkash avkash 98889185 Aug 21 22:32 h2odriver.jar
- drwxr-xr-x 2 avkash avkash 4096 Aug 21 22:32 python/
- drwxr-xr-x 2 avkash avkash 4096 Aug 21 22:32 R/
- -rw-r–r– 1 avkash avkash 1733 Aug 21 22:32 README.txt

Launching H2O Cluster

- **$ hadoop jar h2odriver.jar -nodes 2 -mapperXmx 4g -output /user/avkash/002**

```
avkash@mr-0xd7-precise1:~/h2o-3.14.0.2-cdh5.8$ hadoop jar h2odriver.jar -nodes 2 -mapperXmx 4g -output /user/avkash/002
Determining driver host interface for mapper->driver callback...
    [Possible callback IP address: 172.16.2.217]
    [Possible callback IP address: 127.0.0.1]
Using mapper->driver callback IP address and port: 172.16.2.217:46622
(You can override these with -driverif and -driverport/-driverportrange.)
Memory Settings:
    mapreduce.map.java.opts:     -Xms4g -Xmx4g -XX:PermSize=256m -verbose:gc -XX:+PrintGCDetails -XX:+PrintGCTimeStamps -Dlog4j.defaultInitOverride=true
    Extra memory percent:        10
    mapreduce.map.memory.mb:     4505
17/09/17 10:17:57 INFO client.RMProxy: Connecting to ResourceManager at mr-0xd1-precise1.0xdata.loc/172.16.2.211:8032
17/09/17 10:17:59 INFO mapreduce.JobSubmitter: number of splits:2
17/09/17 10:17:59 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1499294366934_0542
17/09/17 10:17:59 INFO impl.YarnClientImpl: Submitted application application_1499294366934_0542
17/09/17 10:17:59 INFO mapreduce.Job: The url to track the job: http://mr-0xd1-precise1.0xdata.loc:8088/proxy/application_1499294366934_0542/
Job name 'H2O_83587' submitted
JobTracker job ID is 'job_1499294366934_0542'
For YARN users, logs command is 'yarn logs -applicationId application_1499294366934_0542'
Waiting for H2O cluster to come up...
H2O node 172.16.2.212:54321 requested flatfile
H2O node 172.16.2.218:54321 requested flatfile
Sending flatfiles to nodes...
    [Sending flatfile to node 172.16.2.212:54321]
    [Sending flatfile to node 172.16.2.218:54321]
H2O node 172.16.2.218:54321 reports H2O cluster size 1
H2O node 172.16.2.212:54321 reports H2O cluster size 1
H2O node 172.16.2.218:54321 reports H2O cluster size 2
H2O node 172.16.2.212:54321 reports H2O cluster size 2
H2O cluster (2 nodes) is up
(Note: Use the -disown option to exit the driver after cluster formation)

Open H2O Flow in your web browser: http://172.16.2.212:54321

(Press Ctrl-C to kill the cluster)
Blocking until the H2O cluster shuts down...
```

We are launching 2 node H2O cluster on Hadoop with 4 GB memory in each node, total 8 GB

2 Node H2O Cluster

H2O Cluster is ready and open web browser with this IP Address

Blocking call so as long as this window is open, H2O will be up

H2O.ai

Open H2O UI from FLOW

```
avkash@mr-0xd7-precise1:~/h2o-3.14.0.2-cdh5.8$ hadoop jar h2odriver.jar -nodes 2 -mapperXmx 4g -output /user/avkash/002
Determining driver host interface for mapper->driver callback...
    [Possible callback IP address: 172.16.2.217]
    [Possible callback IP address: 127.0.0.1]
Using mapper->driver callback IP address and port: 172.16.2.217:46622
(You can override these with -driverif and -driverport/-driverportrange.)
Memory Settings:
    mapreduce.map.java.opts:     -Xms4g -Xmx4g -XX:PermSize=256m -verbose:gc -XX:+PrintGCDetails -XX:+PrintGCTimeStamps -Dlog4j.defaultInitOverride=true
    Extra memory percent:        10
    mapreduce.map.memory.mb:     4505
17/09/17 10:17:57 INFO client.RMProxy: Connecting to ResourceManager at mr-0xd1-precise1.0xdata.loc/172.16.2.211:8032
17/09/17 10:17:59 INFO mapreduce.JobSubmitter: number of splits:2
17/09/17 10:17:59 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1499294366934_0542
17/09/17 10:17:59 INFO impl.YarnClientImpl: Submitted application application_1499294366934_0542
17/09/17 10:17:59 INFO mapreduce.Job: The url to track the job: http://mr-0xd1-precise1.0xdata.loc:8088/proxy/application_1499294366934_0542/
Job name 'H2O_83587' submitted
JobTracker job ID is 'job_1499294366934_0542'
For YARN users, logs command is 'yarn logs -applicationId application_1499294366934_0542'
Waiting for H2O cluster to come up...
H2O node 172.16.2.212:54321 requested flatfile
H2O node 172.16.2.218:54321 requested flatfile
Sending flatfiles to nodes...
    [Sending flatfile to node 172.16.2.212:54321]
    [Sending flatfile to node 172.16.2.218:54321]
H2O node 172.16.2.218:54321 reports H2O cluster size 1
H2O node 172.16.2.212:54321 reports H2O cluster size 1
H2O node 172.16.2.218:54321 reports H2O cluster size 2
H2O node 172.16.2.212:54321 reports H2O cluster size 2
H2O cluster (2 nodes) is up
(Note: Use the -disown option to exit the driver after cluster formation)

Open H2O Flow in your web browser: http://172.16.2.212:54321

(Press Ctrl-C to kill the cluster)
Blocking until the H2O cluster shuts down...
```

172.16.2.212:54321/flow/index.html

H2O FLOW Flow Cell Data Model Score Admin Help

Untitled Flow

CS getCloud

H2O_83587

CLOUD STATUS

HEALTHY CONSENSUS LOCKED

Version	Started	Nodes (Used / All)
3.14.0.2	17 minutes ago	2 / 2

NODES

Name	Ping	Cores	Load	My CPU %	Sys CPU %	GFLOPS	Memory Bandwidth	Data (Used/Total)	Data (% Cached)	GC (Free / Total / Max)	Disk (Free / Max)	Disk (% Free)
172.16.2.212:54321	a few seconds ago	32	1.710	-1	-1	33.682	77.13 GB / s	- / NaN undefined	NaN%	3.05 GB / NaN undefined / 3.83 GB	1.60 TB / 1.88 TB	85%
172.16.2.218:54321	a few seconds ago	32	0.380	-1	-1	32.869	36.87 GB / s	- / NaN undefined	NaN%	3.83 GB / NaN undefined / 3.83 GB	1.54 TB / 1.88 TB	81%
TOTAL	-	64	2.090	-	-	66.551	114.00 GB / s	- / NaN undefined	NaN%	6.88 GB / NaN undefined / 7.67 GB	3.13 TB / 3.75 TB	83%

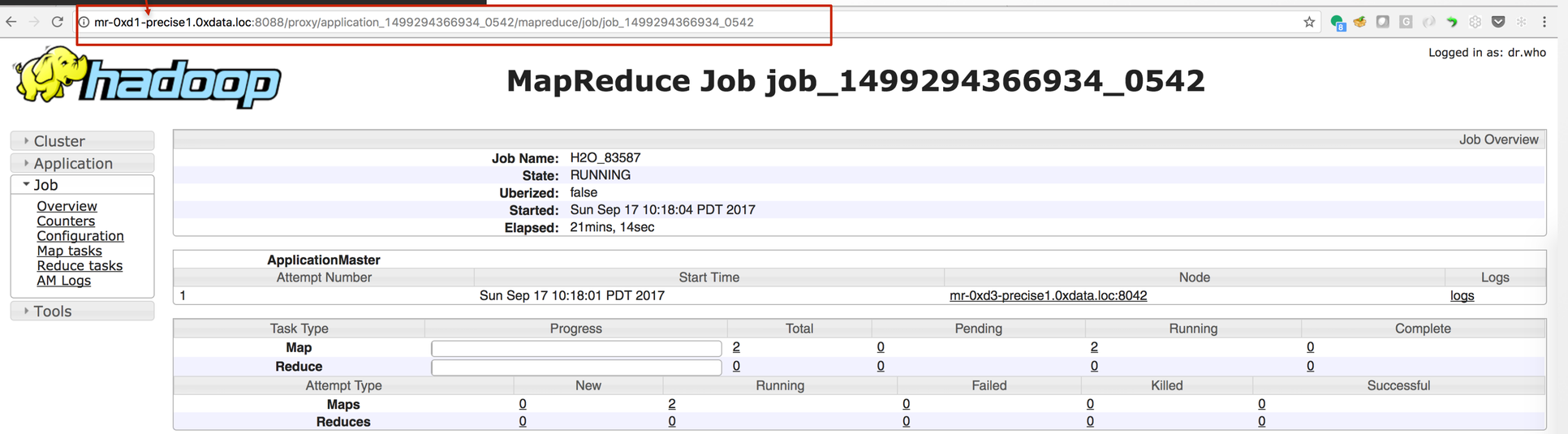
Refresh

H2O cluster as MapReduce Job in Hadoop

```
avkash@mr-0xd7-precise1:~/h2o-3.14.0.2-cdh5.8$ hadoop jar h2odriver.jar -nodes 2 -mapperXmx 4g -output /user/avkash/002
Determining driver host interface for mapper->driver callback...
    [Possible callback IP address: 172.16.2.217]
    [Possible callback IP address: 127.0.0.1]
Using mapper->driver callback IP address and port: 172.16.2.217:46622
(You can override these with -driverif and -driverport/-driverportrange.)
Memory Settings:
    mapreduce.map.java.opts:     -Xms4g -Xmx4g -XX:PermSize=256m -verbose:gc -XX:+PrintGCDetails -XX:+PrintGCTimeStamps -Dlog4j.defaultInitOverride=true
    Extra memory percent:        10
    mapreduce.map.memory.mb:     4505
17/09/17 10:17:57 INFO client.RMProxy: Connecting to ResourceManager at mr-0xd1-precise1.0xdata.loc/172.16.2.211:8032
17/09/17 10:17:59 INFO mapreduce.JobSubmitter: number of splits:2
17/09/17 10:17:59 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1499294366934_0542
17/09/17 10:17:59 INFO impl.YarnClientImpl: Submitted application application_1499294366934_0542
17/09/17 10:17:59 INFO mapreduce.Job: The url to track the job: http://mr-0xd1-precise1.0xdata.loc:8088/proxy/application_1499294366934_0542/
Job name 'H2O_83587' submitted
JobTracker job ID is 'job_1499294366934_0542'
For YARN users, logs command is 'yarn logs -applicationId application_1499294366934_0542'
Waiting for H2O cluster to come up...
H2O node 172.16.2.212:54321 requested flatfile
H2O node 172.16.2.218:54321 requested flatfile
Sending flatfiles to nodes...
    [Sending flatfile to node 172.16.2.212:54321]
    [Sending flatfile to node 172.16.2.218:54321]
H2O node 172.16.2.218:54321 reports H2O cluster size 1
H2O node 172.16.2.212:54321 reports H2O cluster size 1
H2O node 172.16.2.218:54321 reports H2O cluster size 2
H2O node 172.16.2.212:54321 reports H2O cluster size 2
H2O cluster (2 nodes) is up
(Note: Use the -disown option to exit the driver after cluster formation)

Open H2O Flow in your web browser: http://172.16.2.212:54321

(Press Ctrl-C to kill the cluster)
Blocking until the H2O cluster shuts down...
```



H2O.ai

Connecting H2O cluster from R/RStudio

```
avkash@mr-0xd7-precise1:~/h2o-3.14.0.2-cdh5.8$ hadoop jar h2odriver.jar -nodes 2 -mapperXmx 4g -output /user/avkash/002
Determining driver host interface for mapper->driver callback...
    [Possible callback IP address: 172.16.2.217]
    [Possible callback IP address: 127.0.0.1]
Using mapper->driver callback IP address and port: 172.16.2.217:46622
(You can override these with -driverif and -driverport/-driverportrange.)
Memory Settings:
    mapreduce.map.java.opts:     -Xms4g -Xmx4g -XX:PermSize=256m -verbose:gc -XX:+PrintGCDetails -XX:+PrintGCTimeStamps -Dlog4j.defaultInitOverride=true
    Extra memory percent:        10
    mapreduce.map.memory.mb:     4505
17/09/17 10:17:57 INFO client.RMProxy: Connecting to ResourceManager at mr-0xd1-precise1.0xdata.loc/172.16.2.211:8032
17/09/17 10:17:59 INFO mapreduce.JobSubmitter: number of splits:2
17/09/17 10:17:59 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1499294366934_0542
17/09/17 10:17:59 INFO impl.YarnClientImpl: Submitted application application_1499294366934_0542
17/09/17 10:17:59 INFO mapreduce.Job: The url to track the job: http://mr-0xd1-precise1.0xdata.loc:8088/proxy/application_1499294366934_0542/
Job name 'H2O_83587' submitted
JobTracker job ID is 'job_1499294366934_0542'
For YARN users, logs command is 'yarn logs -applicationId application_1499294366934_0542'
Waiting for H2O cluster to come up...
H2O node 172.16.2.212:54321 requested flatfile
H2O node 172.16.2.218:54321 requested flatfile
Sending flatfiles to nodes...
    [Sending flatfile to node 172.16.2.212:54321]
    [Sending flatfile to node 172.16.2.218:54321]
H2O node 172.16.2.218:54321 reports H2O cluster size 1
H2O node 172.16.2.212:54321 reports H2O cluster size 1
H2O node 172.16.2.218:54321 reports H2O cluster size 2
H2O node 172.16.2.212:54321 reports H2O cluster size 2
H2O cluster (2 nodes) is up
(Note: Use the -disown option to exit the driver after cluster formation)

Open H2O Flow in your web browser: http://172.16.2.212:54321

(Press Ctrl-C to kill the cluster)
Blocking until the H2O cluster shuts down...
```

> ip = “172.16.2.212” <= As String
> port = 54321 <= As Integer

Console ~/

```
> library(h2o)
> h2o.init(ip = "172.16.2.212", port = 54321)
Reading in config file: ./.h2oconfig
 Connection successful!

R is connected to the H2O cluster:
    H2O cluster uptime:         26 minutes 42 seconds
    H2O cluster version:        3.14.0.2
    H2O cluster version age:    26 days
    H2O cluster name:           H2O_83587
    H2O cluster total nodes:    2
    H2O cluster total memory:   5.92 GB
    H2O cluster total cores:    64
    H2O cluster allowed cores:  64
    H2O cluster healthy:        TRUE
    H2O Connection ip:          172.16.2.212
    H2O Connection port:        54321
    H2O Connection proxy:       NA
    H2O Internal Security:      FALSE
    H2O API Extensions:         AutoML, Algos, Core V3, Core V4
    R Version:                  R version 3.4.1 (2017-06-30)

> h2o.clusterStatus()
Version: 3.14.0.2
Cluster name: H2O_83587
Cluster size: 2
Cluster is locked

                                          h2o healthy   last_ping num_cpus sys_load mem_value_size   free_mem
1 mr-0xd2-precise1.0xdata.loc/172.16.2.212:54321    TRUE 1.50567e+12       32     0.74              0 3270657024
2 mr-0xd8-precise1.0xdata.loc/172.16.2.218:54321    TRUE 1.50567e+12       32      0.6              0 3089351680
    pojo_mem swap_mem    free_disk     max_disk   pid num_keys tcps_active open_fds rpcs_active
1  846051328        0 1.754946e+12 2.064111e+12 25251        0           0      183           0
2 1027356672        0 1.689253e+12 2.064111e+12  9155        0           0      183           0
```

Connecting H2O cluster from Python

```
avkash@mr-0xd7-precise1:~/h2o-3.14.0.2-cdh5.8$ hadoop jar h2odriver.jar -nodes 2 -mapperXmx 4g -output /user/avkash/002
Determining driver host interface for mapper->driver callback...
    [Possible callback IP address: 172.16.2.217]
    [Possible callback IP address: 127.0.0.1]
Using mapper->driver callback IP address and port: 172.16.2.217:46622
(You can override these with -driverif and -driverport/-driverportrange.)
Memory Settings:
    mapreduce.map.java.opts:     -Xms4g -Xmx4g -XX:PermSize=256m -verbose:gc -XX:+PrintGCDetails -XX:+PrintGCTimeStamps -Dlog4j.defaultInitOverride=true
    Extra memory percent:        10
    mapreduce.map.memory.mb:     4505
17/09/17 10:17:57 INFO client.RMProxy: Connecting to ResourceManager at mr-0xd1-precise1.0xdata.loc/172.16.2.211:8032
17/09/17 10:17:59 INFO mapreduce.JobSubmitter: number of splits:2
17/09/17 10:17:59 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1499294366934_0542
17/09/17 10:17:59 INFO impl.YarnClientImpl: Submitted application application_1499294366934_0542
17/09/17 10:17:59 INFO mapreduce.Job: The url to track the job: http://mr-0xd1-precise1.0xdata.loc:8088/proxy/application_1499294366934_0542/
Job name 'H2O_83587' submitted
JobTracker job ID is 'job_1499294366934_0542'
For YARN users, logs command is 'yarn logs -applicationId application_1499294366934_0542'
Waiting for H2O cluster to come up...
H2O node 172.16.2.212:54321 requested flatfile
H2O node 172.16.2.218:54321 requested flatfile
Sending flatfiles to nodes...
    [Sending flatfile to node 172.16.2.212:54321]
    [Sending flatfile to node 172.16.2.218:54321]
H2O node 172.16.2.218:54321 reports H2O cluster size 1
H2O node 172.16.2.212:54321 reports H2O cluster size 1
H2O node 172.16.2.218:54321 reports H2O cluster size 2
H2O node 172.16.2.212:54321 reports H2O cluster size 2
H2O cluster (2 nodes) is up
(Note: Use the -disown option to exit the driver after cluster formation)

Open H2O Flow in your web browser: http://172.16.2.212:54321

(Press Ctrl-C to kill the cluster)
Blocking until the H2O cluster shuts down...
```

> ip = “172.16.2.212” <= As String
> port = 54321 <= As Integer

```
In [4]: h2o.init(ip = '172.16.2.212', port = 54321, strict_version_check=False)
```

```
ERROR:h2o:Key init.version_check is not a valid config key
```

```
Warning: connecting to remote server but falling back to local... Did you mean to use `h2o.connect()`?
Checking whether there is an H2O instance running at http://172.16.2.212:54321. connected.
```

H2O cluster uptime:	30 mins 39 secs
H2O cluster version:	3.14.0.2
H2O cluster version age:	26 days
H2O cluster name:	H2O_83587
H2O cluster total nodes:	2
H2O cluster free memory:	5.923 Gb
H2O cluster total cores:	64
H2O cluster allowed cores:	64
H2O cluster status:	locked, healthy
H2O connection url:	http://172.16.2.212:54321
H2O connection proxy:	None
H2O internal security:	False
H2O API Extensions:	AutoML, Algos, Core V3, Core V4
Python version:	3.6.1 final

```
In [13]: h2o.cluster().show_status(True)
```

H2O cluster uptime:	31 mins 45 secs
H2O cluster version:	3.14.0.2
H2O cluster version age:	26 days
H2O cluster name:	H2O_83587
H2O cluster total nodes:	2
H2O cluster free memory:	5.923 Gb
H2O cluster total cores:	64
H2O cluster allowed cores:	64
H2O cluster status:	locked, healthy
H2O connection url:	http://172.16.2.212:54321
H2O connection proxy:	None
H2O internal security:	False
H2O API Extensions:	AutoML, Algos, Core V3, Core V4
Python version:	3.6.1 final

Nodes info:	Node 1	Node 2
h2o	mr-0xd2-precise1.0xdata.loc/172.16.2.212:54321	mr-0xd8-precise1.0xdata.loc/172.16.2.218:54321
healthy	True	True
last_ping	1505670594772.0000000	1505670595621.0000000
num_cpus	32	32
sys_load	0.16	0.81
mem_value_size	0	0
free_mem	3270657024.0000000	3089351680.0000000
pojo_mem	846051328	1027356672.0000000
swap_mem	0	0
free_disk	1754946076672.0000000	1689252790272.0000000
max_disk	2064111370240.0000000	2064111370240.0000000
pid	25251	9155
num_keys	0	0
tcps_active	0	0
open_fds	184	183
rpcs_active	0	0

H2O.ai

Importing Data into H2O

- File location on HDFS:
  - hdfs://mr-0xd1-precise1.0xdata.loc:8020/user/avkash/prostate.csv

H2O FLOW

```
importFiles
```

Import Files

Search: hdfs://mr-0xd1-precise1.0xdata.loc:8020/user/avkash/prostate.csv

Search Results: (All files added)

Selected Files: 1 file selected: Clear All

✖ hdfs://mr-0xd1-precise1.0xdata.loc:8020/user/avkash/prostate.csv

Actions: Import

```
CS
importFiles [ "hdfs://mr-0xd1-precise1.0xdata.loc:8020/user/avkash/prostate.csv" ]
```

1 / 1 files imported.

Files hdfs://mr-0xd1-precise1.0xdata.loc:8020/user/avkash/prostate.csv

Actions Parse these files...

R/RStudio

```
> df_r = h2o.importFile('hdfs://mr-0xd1-precise1.0xdata.loc:8020/user/avkash/prostate.csv')
  |=============================================================================================================| 100%
> df_r
  ID CAPSULE AGE RACE DPROS DCAPS  PSA  VOL GLEASON
1  1       0  65    1     2     1  1.4  0.0       6
2  2       0  72    1     3     2  6.7  0.0       7
3  3       0  70    1     1     2  4.9  0.0       6
4  4       0  76    2     2     1 51.2 20.0       7
5  5       0  69    1     1     1 12.3 55.9       6
6  6       1  71    1     3     2  3.3  0.0       8

[380 rows x 9 columns]
```

Python

```
In [14]: df_p = h2o.import_file('hdfs://mr-0xd1-precise1.0xdata.loc:8020/user/avkash/prostate.csv')
         Parse progress: |████████████████████████████████████████████████████████| 100%

In [15]: df_p
```

ID	CAPSULE	AGE	RACE	DPROS	DCAPS	PSA	VOL	GLEASON
1	0	65	1	2	1	1.4	0	6
2	0	72	1	3	2	6.7	0	7
3	0	70	1	1	2	4.9	0	6
4	0	76	2	2	1	51.2	20	7
5	0	69	1	1	1	12.3	55.9	6
6	1	71	1	3	2	3.3	0	8
7	0	68	2	4	2	31.9	0	7
8	0	61	2	4	2	66.7	27.2	7
9	0	69	1	1	1	3.9	24	7
10	0	68	2	1	2	13	0	6

H2O.ai

Saving MOJO/POJO

R

```
> gbm_model = h2o.getModel('gbm_model_test')
> gbm_model
Model Details:
==============

H2ORegressionModel: gbm
Model ID:  gbm_model_test
Model Summary:
  number_of_trees number_of_internal_trees model_size_in_bytes min_depth max_depth mean_depth min_leaves max_leaves
1              50                       50               12389         5         5    5.00000          8         24
  mean_leaves
1    14.76000


H2ORegressionMetrics: gbm
** Reported on training data. **

MSE:  0.06846676
RMSE:  0.2616615
MAE:  0.2066947
RMSLE:  0.1868665
Mean Residual Deviance :  0.06846676




> h2o.download_mojo(gbm_model, path='/tmp/', get_genmodel_jar = TRUE)
[1] "gbm_model_test.zip"
> h2o.download_pojo(gbm_model, path='/tmp/', get_genmodel_jar = TRUE)
Error in h2o.download_pojo(gbm_model, path = "/tmp/", get_genmodel_jar = TRUE) :
  unused argument (get_genmodel_jar = TRUE)
> h2o.download_pojo(gbm_model, path='/tmp/')
[1] "gbm_model_test.java"
```

Python

```
In [29]: gbm_model = h2o.get_model('gbm_model_test')
         #print(gbm_model)
         gbm_model.download_mojo(path='/tmp/', get_genmodel_jar=True)

Out[29]: '/tmp/gbm_model_test.zip'

In [30]: gbm_model.download_pojo(path='/tmp/', get_genmodel_jar=True)

Out[30]: '/tmp/gbm_model_test.java'
```

```
> h2o.download_pojo(gbm_model, path='hdfs://mr-0xd1-precise1.0xdata.loc:8020/user/avkash/')
Error in h2o.download_pojo(gbm_model, path = "hdfs://mr-0xd1-precise1.0xdata.loc:8020/user/avkash/") :
  'path',hdfs://mr-0xd1-precise1.0xdata.loc:8020/user/avkash/, to save pojo cannot be found.
```

Note: The path will be used for the local file system where R/Python is running

H2O.ai

avkash@h2o.ai