How to build a 7 node Raspberry Pi Hadoop Cluster

Introduction

Inspired by a desire to learn more about Hadoop and the fact I already owned a Raspberry Pi I wondered whether anyone had yet built a Hadoop cluster based on this hobby computers. I wasn't surprised to discover that people have already done this and the following instructions are the where I started:

Carsten Mönning: <u>http://scn.sap.com/community/bi-platform/blog/2015/04/25/a-hadoop-data-lab-project-on-raspberry-pi--part-14</u>

Jonas Widriksson: http://www.widriksson.com/raspberry-pi-hadoop-cluster/

Jonas's instructions are based on Hadoop version 1.0 and Carsten's is based on version 2.x

If, like me, you're interested in building with the newer version of Hadoop then follow Carsten's instructions but read through Jonas's too because he provides useful links for downloading the Raspian (Linux Operating System built specifically for the Raspberry Pi) distribution as well as commands and example files for testing your cluster.

The first stage is to build a single node cluster where your one node performs all tasks such as NameNode, Secondary NameNode and DataNode. Once you have this up-and-running you're reading to add a second node. This second node will be a dedicated DataNode from which you will clone all subsequent DataNodes.

Creating the second node is slightly more difficult which is why I decided to write this post in the hope that it will save others time and effort.



Hardware

For a seven node Raspberry Pi (Model 2 B+ used here) Hadoop cluster you will need the following hardware:

- 7x Raspberry Pi Model 2 B+ (ThePiHut.com)
- 1x 10GigaBit Switch (8 Ports TP-Link) (eBuyer.com)
- 1x USB Power Hub (7 Powered Ports by StarTech.com) (eBuyer.com)
- 7x 8GB Class 10 MicroSD cards NB mine came with the NOOBs Operating System preinstalled which shrinks the size to 3GB. You can expand this partition to use the whole disk from within the raspi-config utility but I used fdisk to delete the existing partition and create a new one (a reformat is required after doing this). (ThePiHut.com)
- 6x 19.05mm standoffs (ModMyPi.com)
- 7x Angled USB to MicroUSB cables (eBuyer.com)
- 7x Short Ethernet Cables (eBuyer.com)
- 7x 32GB Class 10 USB sticks (eBuyer.com)
- 1x Case to put it all in Optional

Building the Single Node Cluster

So, as stated previously, follow these instructions to create your NameNode/Single Node Cluster: http://scn.sap.com/community/bi-platform/blog/2015/04/25/a-hadoop-data-lab-project-on-raspberry-pi--part-14

Adding a new node to the cluster

Once your NameNode (192.168.0.110 node1) is up and running adjust the following files to include your planned/subsequent nodes. In order to edit the system owned files you'll need to switch user or 'sudo'. The simplest editor to use on this flavour of Linux is 'nano':

\$ sudo nano /etc/hosts

Add all the nodes you plan to have in your cluster:

192.168.0.110 node1 192.168.0.111 node2 192.168.0.112 node3 192.168.0.113 node4 192.168.0.114 node5 192.168.0.115 node6 192.168.0.116 node7

The Hadoop specific settings files, for versions 2.x can be found in the following directory:

/opt/hadoop/etc/hadoop

Two notable files are the master and slaves file. Ensure that node1 is the only line listed in your master file and all nodes (if you configure your node1 to still be a DataNode) in the slaves file.

For simplicity I have zipped up the files I had to update so you may copy them directly to your node1 and node2 Pi's or use them as reference:

http://www.nigelpond.com/uploads/node1-config-files.zip

http://www.nigelpond.com/uploads/node2-config-files.zip

NB: Because you'll be temporarily connecting to an IP address of an existing Pi it's probably worth having only the NameNode powered up when performing this step.



- Clone the NameNode MicroSD card and duplicate the image file on to a new MicroSD card. On Linux the commands to do this are as follows. Note that when mounting your MicroSD card your device name may differ from /dev/sdb: \$ sudo dd if=/dev/sdb of=node1.img
 - \$ sudo dd if=node1.img of=/dev/sdb

2. Format the USB storage device - giving it a label of 'hdfs'. Again I used Linux to format the device and used the EXT4 file system.

3. Put your cloned MicroSD card and USB storage stick into your new Raspberry Pi

- 4. Attach the Ethernet network cable & power cable
- 5. SSH into the Pi using the IP address of the cloned Pi using the hduser/password credentials:

\$ ssh hduser@192.168.0.110

6. Configure the Raspberry Pi using the raspi-config utility - you'll be prompted to reboot when you've finished making this set of changes:

\$ sudo raspi-config

- Change the node name (i.e. from node1 to node2)
- Check: enable SSH server
- Check: memory split should be 16M
- Overclocking to Pi2 (reboot)

8. After the reboot from exiting raspi-config wait a few seconds and then SSH back in again. Edit the /etc/network/interfaces file and change the IP address to the next available one:

\$ sudo nano /etc/network/interfaces

9. Check the /etc/hosts file contains ALL the IP addresses and node names of ALL your Pi's:

\$ cat /etc/hosts

10. Reboot the Pi

\$ sudo reboot

11. SSH back into the Pi using the new IP address:

\$ ssh hduser@192.168.0.111

12. From the new Pi node SSH into the NameNode - say 'yes' to question:

\$ ssh node1

13. On the NameNode SSH to the new node name - say 'yes' to question – and then return to node2:

\$ ssh node2

\$ exit

14. Creating the /hdfs data partition. Change directory into HDFS partition. Create the /tmp folder and change ownership.

\$ sudo mkdir -p /hdfs/tmp

\$ sudo chown -R hduser:hadoop /hdfs

\$ sudo chmod 750 /hdfs/tmp

15. Add an entry to the /etc/fstab file so that the /hdfs partition is mounted on boot

\$ sudo nano /etc/fstab

Make it match this but remember the /dev/sd* may differ:

B	hduser@node1: ~									
hduser@node1 ~ \$	cat /etc/	fstab							^	
proc	/proc	proc	defaults							
/dev/mmcblk0p1	/boot	vfat	defaults							
/dev/mmcblk0p2		ext4	defaults,n	noatime 0						
/dev/sda1	/hdfs	ext4	defaults							
# a swapfile is n	not a swap	partition,	so no using	swapon off	from here	on, use	dphys-swapfile swap[on off]	for that	:	
hduser@node1 ~ \$										
									\sim	

15. Make the required /opt/hadoop/etc/hadoop file changes - we're changing the purpose of this machine from NameNode + DataNode to just DataNode. Check the aforementioned node2-config-files.zip for reference.

16. On the NameNode (node1) Start DFS and YARN:

/opt/hadoop/sbin/start-dfs.sh

/opt/hadoop/sbin/start-yarn.sh

On the DataNode, node2, take a look in the /hdfs/tmp directory and you should see that the NameNode has created a whole bunch of sub-directories.

\$ Is -Irt /hdfs/tmp

Check to see that the DataNode tasks are running with the 'jps' command:



17. Browse to the NameNode web interface: http://192.168.0.110:8088/ and ensure you can see the new node. Here's a screenshot showing all 7 nodes included:

🗋 Joshua David Bartho 🗴	(3) Welco	ome! Link	e × 🔰 (7)) Twitter	× 🚺 🗛	Hadoop data	a lab pr ×	Raspberry	PI Hadoop	× 🗅 No	des of the clust	er × 📃			(<u> </u>	
← → C 🗋 192.168.0	0.110 :8088/cl	uster/node	es													★ ∞ ≡
👯 Apps 🔺 Bookmarks 🚦	Google 🗋 Ne	ew Tab 🗀	Imported F	From IE 🗀 Im	ported From I	•									🗀 Oth	er bookmarks
1 con															Logge	ed in as: dr.who
							Nor	les of t	he c	luste	r					
							1100	103 01 1		lusic	•					
	-															
✓ Cluster	Cluster Metr	ics														
About	Apps	Apps	Apps	Apps	Containers	Memory	Memory	Memory	VCores	VCores	VCores	Active I	Decommissione	Lost	Unhealthy	Rebooted
Applications	0	0	0	0	0	0 B	56 GB	0 B	0 Used	56	0	7 C	Nodes	0	Nodes 0	0
NEW																
SUBMITTED	Node Labels	Deute	. Node	State Node A	ddress Node	HTTP Addre	ss				Container	s Mem Use	d Mem Avail	VCores Used	VCores Ava	ail Version
RUNNING	÷	Rack	Ŷ	\$	\$		Last r Last r	health-update 👳	Heal	tn-report	Ŷ.		• •	\$		¢ ¢
FINISHED		/default-r	ack RUNNI	ING node/:t	0349 <u>node/</u>	:8042	10-Jur	-2015 09:44:51			0	0 B	8 GB	0	8	2.6.0
KILLED		/default-r	ack RUNN	ING node5:	19889 node5	-8042 -8042	10-Jur	2015 09:44:50			0	0.8	8 GB	0	8	2.6.0
		/default-r	ack RUNN	ING node6:	5418 node6	8042	10-Jur	-2015 09:44:50			0	0.8	8 GB	0	8	260
Scheduler		/default-r	ack RUNN	ING node3:5	9032 node3	:8042	10-Jur	-2015 09:44:51			0	0 8	8 GB	0	8	2.6.0
> Tools		/default-r	ack RUNN	ING node4:5	7745 node4	:8042	10-Jur	-2015 09:44:50			0	0 B	8 GB	0	8	2.6.0
. 10010		/default-r	ack RUNN	ING node1:5	2464 <u>node1</u>	:8042	10-Jur	-2015 09:44:54			0	0 B	8 GB	0	8	2.6.0
	Showing 1 to	7 of 7 entri	es											First	Previous 1	Next Last

18. Running MapReduce tasks - the instructions are identical to those found on Jonas's site except there are some subtle updates to the hdfs commands for version 2.x of Hadoop. The steps I used are:

\$ hadoop fs -copyFromLocal mediumfile.txt /mediumfile.txt

\$ yarn jar /opt/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.0.jar wordcount /mediumfile.txt /mediumfile-out

You can monitor the running tasks from the NameNode web interface. Here are some screenshots of my test runs:

)				All Ap	plicatio	ons			
Cluster Metric:	s									
Apps Submitted	Apps Apps Pending Rupping	Apps	s Conta	ainers Memory	Memory Memor	y VCores	VCores VCores	Active d Nodes	Decommissioned	Lost Unhe
7 0	4	3	4	8 GB	56 GB 0 B	4	56 0	<u>Z</u>	<u>0</u>	<u>Q</u> <u>Q</u>
Show 20 • er	ntries									Search:
application_143	33929364944_0007	hduser	word count	MAPREDUCE	default	Wed, 10 Jun 2015 10:04:06	FinishTime ≎ N/A	ACCEPTED	UNDEFINED	Progress V
application_143	33929364944_0006	hduser	word count	MAPREDUCE	default	Wed, 10 Jun 2015 10:04:03 GMT	N/A	ACCEPTED	UNDEFINED	
application_143	33929364944_0005	hduser	word count	MAPREDUCE	default	Wed, 10 Jun 2015 10:04:00 GMT	N/A	ACCEPTED	UNDEFINED	
application_143	33929364944_0004	hduser	word count	MAPREDUCE	default	Wed, 10 Jun 2015 10:03:57 GMT	N/A	ACCEPTED	UNDEFINED	
application_143	33929364944_0003	hduser	word count	MAPREDUCE	default	Wed, 10 Jun 2015 09:58:44 GMT	Wed, 10 Jun 2015 10:02:12 GMT	FINISHED	SUCCEEDED	
application_143	33929364944_0002	hduser	word count	MAPREDUCE	default	Wed, 10 Jun 2015 09:58:40 GMT	Wed, 10 Jun 2015 10:01:55 GMT	FINISHED	SUCCEEDED	
application_143	33929364944_0001	hduser	word count	MAPREDUCE	default	Wed, 10 Jun 2015 09:52:02 GMT	Wed, 10 Jun 2015 09:55:20 GMT	FINISHED	SUCCEEDED	
Showing 1 to 7	of 7 entries									
ter Metrics				NEW,I	NEW_SAVII	NG,SUBN	MITTED, AC	CEPTE	D,RUNNIN	IG Applic
Iuster Metrics Apps Submitted A 10 0	Ipps Pending Apps 1 3	Running 7	Apps Complete	NEW, I containers Rur 24	NEW_SAVII	MG,SUBN Memory Total 56 GB 0	MITTED, AC	CEPTE	Cores Total VCores	IG Applic
Iuster Metrics Apps Submitted A Application Queuee	Apps Pending Apps I 3	Running 7	Apps Complete	NEW, 1 ed Containers Run 24	NEW_SAVII	MG,SUBN Memory Total 56 GB 0	MITTED, AC	CEPTE	Cores Total VCores	IG Applic
Luster Metrics Apps Submitted A 10 0 Application Queuen Legend: Capa	pps Pending Apps I 3 s acity : Used	Running 7 7 Used	Apps Complete 7	NEW, I ad Containers Rur 24	NEW_SAVII	MG,SUBN Memory Total 56 GB 0	MITTED, AC	CEPTE	ED,RUNNIN //Cores Total VCoret 5 0	IG Applic
Iuster Metrics Apps Submitted A 10 0 Application Queuer Legend: Cape	Apps Pending Apps I 3 s acity USed	Running 7 Used	Apps Complete 7	NEW, I	NEW_SAVII	MG,SUBN Memory Total 56 GB 0	MITTED, AC	CEPTE	Cores Total VCores	IG Applic
Luster Metrics Apps Submitted A 10 0 Application Queue Legend: Capa Tea 	spps Pending Apps 1 3 S	Running 7 7 Uséd	Apps Complete 7 § (over capacit	NEW, I 24 9 Max Cap	NEW_SAVII	Memory Total 56 GB 0	MITTED, AC	CEPTE	ED,RUNNIN VCores Total VCores	IG Applic
App Submitted A App Submitted A Application Queue Legend: Capa	upps Pending Apps 3 3 Sacity : Used Queu Used Ci	Running 7 Used e State: F	Apps Complete 7 I (over capacit RUNNING 48.2%	NEW, I	NEW_SAVII	Memory Total 56 GB 0	MITTED, AC	CEPTE	Cores Total VCores	IG Applic
App Submitted A App Submitted A Application Queue Legend: Capa	sps Pending Apps 3 s actity : Used Used Cr Absolute Used Cr Absolute Sect	e State: F spacity: 4 apacity: 4 apacity: 4	Apps Completi (over capacit RUNNING 18.2% 18.2%	NEW, I	NEW_SAVII	NG, SUBN Memory Total 56 GB 0	MITTED, AC	CCEPTE	Cores Total VCores	IG Applic
Luter Metrics Apps Submitted A 10 00 Application Queue Legend: Capa Capatric Capa	ops Pending Apps 3 s acity Used Absolute Used C Absolute Used C Absolute Max C	e State: F spacity: 4 spacity: 4 spacity: 1 spacity: 1 spacity: 1	Apps Complete 7 1 (over capacit RUNNING 48.2% 48.2% 100.0%	NEW, I ad Containers Rur 24 Wax Cap	NEW_SAVII	Memory Total 56 OB 0	MITTED, AC	CCEPTE	Cores Total VCores	IG Applic
Uster Metrics Apps Submitted A o Submitted A o Repeated A construction Queues Legend: Capa	opps Pending Apps 3 s acity Used Absolute Used C Absolute Used C Absolute Used C Absolute Appli	e State: F pacity: 4 upacity: 1 upacity: 1 upacity: 3 upacity: 3 upacity: 3 upacity: 3 upacity: 4 upacity: 4 u	Apps Complete 7 1 (over capacit 8 (0000 capaci	NEW,I ed Containers Rur 24 97 Max Cap	NEW_SAVII	Memory Total 56 08 0	MITTED, AC	CCEPTE	ED,RUNNIN VCores Total VCores	IG Applic
App Submitted A App Submitted A App Submitted A Application Queue: Legend: Cape	ops Pending Apps 3 s acity Used Absolute Used C Absolute Used C Absolute Applic Schedulable Applic	e State: F pacity: 4 pacity: 2 pacity: 1 pacity: 1 pacity: 1 pacity: 2 stations: 3	Apps Completi 7 8 (over capacit 848.2% 48.2% 48.2% 48.2% 48.2% 49.00.0% 49.2%	NEW, I 24 Max Cap: 48, vCores:24>	NEW_SAVII	Merrory Total 56 G B 0	MITTED, AC	CCEPTE	ED,RUNNIN VCores Total VCores	IG Applic
Luster Metrics App Statemited A App Stat	opps Pending Apps 3 s Queu Used C: Absolute Used C: Absolute Weid C: Absolute Weid C: Schedulable Appli Schedulable Appli Schedulable Appli	e State: F pacity: 4 apacity: 4 apacity: 1 purces: c ations: 0 tainers: 2 ations: 4	Apps Complete 7 1 (over capacit 842,2% 100,0% 100,0% 462,2% 400,0% 100,0% 20,0	NEW, I Containers Rur 24 Max Cap. 48, vCores:24>	NEW_SAVII	Merrory Total 56 0B 0	MITTED, AC	CCEPTE	Cores Total VCores	IG Applic
Lutter Metrics Apps Submitted A Apps Submitted A Application Queue Legend: Capa	opps Pending Apps 3 s acity Used C Absolute Used C Absolute Used C Absolute Used Res Schedulable Applic Num Con Max Applications P	e State: F upacity: 4 upacity: 4 upacity: 1 upacity: 1 upacity: 1 upacity: 3 upacity: 3	Apps Complety 7 1 (over capacit 848.2% 48.2% 48.2% 100.0% smemory.276- 3 0 24 10000	NEW, I 24 24 24 24 24 24 24 24 24 24 24 24 24	NEW_SAVII	Memory Total 55 GB 0	MITTED, AC	CEPTE	Cores Total VCores	IG Applic
App Statement of App St	opps Pending Apps 3 s acity Used Absolute Used C Absolute Used C Absolute Used C Absolute Appli Num Con Max Applications P Schedulable Appli Schedulable Appli Schedulable Appli Schedulable Appli Schedulable Appli	e State: F pacity: 4 apacity: 4 apacity: 2 apacity: 3 ations: 3 cations: 3 cations: 1 ations: 6 cations: 6 cat	Apps Complete 7 8 1 (over capacit RUNNING 48.2% 48.2% 100.0% commony.276-3 0 24 10000 5	NEW,I ed Containers Rur 24 97 Max Cap: 48, vCores:24>	NEW_SAVII	Memory Total 55 CB 0	MITTED, AC	CCEPTE	ED,RUNNIN VCores Total VCores s	IG Applic
Luster Metrics Apps Submitted A App Submitted	Ogueu Gueu Gueu Used C: Absolute Used C: Absolute Used C: Absolute Ward Schedulable Appli Num Com Max Applications Py Configured C: Configured C:	e State: F apacity: 4 apacity: 2 apacity: 2 ations: 2 ations: 2 ations: 1 ations: 1 ations: 2 ations: 2 ations: 2 ations: 2 ations: 3 ations: 4 ations: 4 at	Apps. Completiv 7 8 (over capaciti 848,2% 48,2% 48,2% 48,2% 48,2% 48,2% 400,0% 5 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	NEW, I	NEW_SAVII	Mernery Total 56 0B 0	MITTED, AC	CCEPTE	ED,RUNNIN	IG Applic
Lutter Metrics App Submitted A	pps Pending Apps 3 s acity : Used Absolute Used C: Absolute Used C: Absolute Max Applications P: Schedulable Applications P: Schedulable Applications P: Schedulable Applications P: Configured C: Configured C: Configured Max C2:	e State: F pacity: 2 pacity: 2 pacity: 2 pacity: 1 pacity: 1 purces: 2 cations: 0 cations: 0 cations: 1 cations: 1 cations: 1 cations: 1 cations: 2 cations: 1 cations: 1 cation	Apps Complete 7 1 (over capacit 2 UNNING 48 2% 48 2% 100.0% 5 0 2 24 10000 5 5 5 5 100.0%	NEW, I ad Containers Rur 24 Max Cap	NEW_SAVII	Merrory Total 56 OB 0	MITTED, AC	CCEPTE	Cores Total VCores	IG Applic
Lutter Metrics Apps Submitted A 10 0 0 0 Legend: Capa Castor Castor Num Num Non Max Max Schedul Configured M	opps Pending Apps 3 s acity: Used Absolute Used C: Absolute Used C: Absolute Used C: Absolute Used C: Schedulable Applic Num Con Max Applications P: Schedulable Applications P: Schedulable Applications P: Configured User Configured User Limit Mur Vaer Limit Factor International Max Paplications P: Configured User Limit	e State: F pacity: 2 pacity: 2 pacity: 2 pacity: 3 pacity: 1 pacity: 1 pacity: 1 pacity: 1 ations: 2 taliners: 2 ations: 1 r User: 6 pacity: 1 pacity: 1 pacity: 1 ercent: 1 ercent: 1	Apps Complete 7 (over capacit RUNNING 48.2% 100.0% 100.0% 100.0% 10000 6 5 100.0% 1	NEW, I ad Containers Rur 24 Max Cap	NEW_SAVII	Memory Total 55 GB 0	MITTED, AC	CCEPTE	CCres Total VCores	IG Applic
App Submitted A App Submitted A App Submitted A Application Queue: Legend: Cape Cape Cape Cape Cape Cape Cape Cape	pps Pending Apps] g g Queu Queu Queu Queu Queu Queu Queu Queu Queu Queu Queu Queu Queu Queu Schedualbe Appli Absolute Max Appli Absolute Max Appli Max Applications P Schedualbe Appli Schedualbe Appli Schedualbe Appli Configured Max C Configured Max C C Configured Max C C C C C C C C C C C C C C	Running 7 7 Used state: f 4 vsata states f states f state	Apps Complete 7 1 (over Capace 8 UNNING 48.2% 48.2% 49.2% 100.0% 100.0% 100.0% 100.0% 100.0% 100.0% 100.0% 100.0%	NEW, I 24 Max Cap: 48, vCores:24>	NEW_SAVII	Memory Total 55 CB 0	MITTED, AC	CCEPTE	Cores Total VCores S	IG Applic
App Submitted A App Submitted A App Submitted A Application Queue: Legend: Cape Cape Cape Cape Cape Cape Cape Cape	pps Pending Apps] g g Queu Queu Queu Queu Queu Queu Queu Queu Queu Queu Queu Queu Queu Queu Queu Queu Queu Queu Schedualbe Appli Astronolite Appli Astronolite Appli Astronolite Appli Astronolite Appli Astronolite Appli Astronolite Appli Astronolite Appli Astronolite Appli Configured Astronolite Configured Max Configured Configured Astronolite Configured Max Configured Astronolite Astronolite Astronolite Astronolite Configured Max Configured Astronolite Configured Max Configured Max Config	Running 7 7 Used State: F 4 states 5 tate: F 4 states 5 tates 7 states 5 tates	Apps Complete 7 1 (Over Capacity 8 (UNNING 48.2% 48.2% 49.2% 100.0% 100.0% 100.0% 100.0% 100.0% 100.0% 100.0% 100.0% 100.0% 100.0%	NEW, I 24 Max Cap: 48, vCores:24>	NEW_SAVII	Memory Total 56 CB 0	MITTED, AC	CCEPTE	Cores Total VCores	IG Applic
App Submitted A App Submitted A App Submitted A Application Queue Legend: Caps Application Queue Capped A Max Schedul Configured M Co	ops Pending Apps 3 s acity Used Absolute Used C Absolute Used C Absolute Used C Absolute Used R Absolute Max C Schedulable Applic Num Con Max Applications P Schedulable Applic Num Con Max Applications P Configured Cser Limit Active Accessible Node	Running 7 7 Uted state:	Apps Complete 7 1 (Over Capacity 14 (Over Capaci	NEW, I 24 Max Cap: 48, vCores:24>	NEW_SAVII	Memory Total 55 GB 0	MITTED, AC	CCEPTE Cores Used 1 SI	Cores Total VCores	IG Applic
Luster Metrics Apps Submitted A App Submitted	pps Pending Apps 3 3 acity 3 Control Control C	2 Unning 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7	Apps Complete 7 1 (over capacit 1(over capacit 82% 48.2% 48.2% 48.2% 48.2% 49.2% 49.2% 40.00.0% 100.0% 5 6 6 6 6 6 6 100.0% 100.0% 100.0% 100.0% 100 0 0 Name	NEW, I	NEW_SAVII	Memory Total 56 0B 0	MITTED, AC	CCEPTE Cores Used 's Si	Cores Total VCores Cores Total VCores of used resources in Projektime	IG Applic
ter Metrics ps Submitted A ps Submitted A ps Submitted A ps Capa plication Queue egend: Capa pret metrics Num Non Num Non Max Max Schedulu Configured M Co station_143322336	pps Panding Apps 3 5 acity Dead C Absolute Used C Absolute Used C Absolute Ward C Absolute Max Schedulable Applications P Schedulable Applications P Max Applications P Max Applications P Max Applications P Schedulable Applications P Schedu	7 7 7 7 2 Used	Apps Complete 7 2 (over capaciti (ver capaciti 48,2% 48,2% 48,2% 48,2% 48,2% 48,2% 48,2% 49,2% 40,00% 100.0% 5 5 5 100.0% 1.00 0% 1.00	NEW, I Containers Rur 24 Max Cap Max Cap 48, vCores:24> moy: 27648 (<span 1<="" td=""><td>NEW_SAVII</td><td>Memory Total 56 0B 0</td><td>MITTED, AC Memory Peserved VI B 24 00%-(span>), vCores: RMB 0 State State Wed, 10 Jun Wed, 10 Jun Wed, 10 Jun</td><td>CCEPTE Cores Used i 51</td><td>Cores Total VCores Cores Total VCores Cores Cor</td><td>IG Applic</td>	NEW_SAVII	Memory Total 56 0B 0	MITTED, AC Memory Peserved VI B 24 00%-(span>), vCores: RMB 0 State State Wed, 10 Jun Wed, 10 Jun Wed, 10 Jun	CCEPTE Cores Used i 51	Cores Total VCores Cores Total VCores Cores Cor	IG Applic
Num	pps Pending Apps] 3 3 acity Dead Absolute Used C Absolute Used C Absolute Used C Absolute Max C Absolute Max C Absolute Max C Schedulable Applications P Configured User Limit Max Applications P Configured User Limit Active Accessible Node	Running 7 7 Used state: F state: F state: State: State state	Apps Complete 7 1 (over capacit 8 2% 48 2% 48 2% 100.0% 5 24 20.0% 24 20.0% 20.0% 20.0% 100.0% 2	NEW, I ad Containers Rur 24 Max Capi 48, vCores:24> 48, vCores:24> mory: 27648 (<span to<br="">MAPREDUC MAPREDUC	NEW_SAVII aning Memory Used 27 GB acity tide="of used resources i Application Type E	Memory Total 56 GB 0	MITTED, AC	CCEPTE Cores Used (5 5 24 (<span title<br="">24 (<span title<br="">2015 10 21:37 2015 10 21:34	Cores Total VCores	IG Applic

19. When you're happy with your first DataNode, and you're ready to add another, go through steps 1-8 and 10-13 to create each subsequent node. If you've performed any tests, loaded any data, you will have to clear down the /hdfs/tmp directory to ensure all nodes are in-sync. When you start DFS from the NameNode (node1) it will recreate the file structure required on each node.

The Finished Cluster



Next Steps/Still to do:

Get a router so that it's on its own network. This will enable the use of the web interface URL's that use the node names. Currently my home router is trying to DNS the IP addresses so these links don't work.

Write some scripts, or find tool, for running commands across all nodes. Especially for commands like startup/shutdown and clearing down the /hdfs partition. I'm thinking Chef or Puppet at this stage as I think they would be useful tools to learn about.