

Session 3

**Getting Started with Hadoop
& HDFS Lab1**

What For Today!!!

- ✓ Setting up VM Hadoop Environment
- ✓ Using the NameNode WebUI
- ✓ Using the Hadoop FileShell
- ✓ Hadoop FileShell Commands
- ✓ Hands On
- ✓ HDFS Federation & HA
- ✓ HDFS 1.0 and 2.0

Sample HDFS Shell Commands

- `bin/hadoop fs -ls`
- `bin/hadoop fs -mkdir`
- `bin/hadoop fs -copyFromLocal`
- `bin/hadoop fs -copyToLocal`
- `bin/hadoop fs -moveToLocal`
- `bin/hadoop fs -rm`
- `bin/hadoop fs -tail`
- `bin/hadoop fs -chmod`
- `bin/hadoop fs -setrep -w 4 -R /dir1/s-dir/`

Sample HDFS Shell Commands

```
hadoop fs -copyFromLocal /home/sam/file.txt  
hdfs://host:port/user/sam/f.txt
```

```
hadoop fs -ls .
```

```
Found 2 items
```

```
drwxr-xr-x – sam group 0 2012-06-17 18:30
```

```
 /user/sam/
```

```
-rw-r--r-- 1 sam group 118 2012-06-13 19:31
```

```
 /user/sam/f.txt
```

FileSystem Check (FSCK)

- % hadoop fsck /
- Status: HEALTHY
- Total size: 9018951544337 B
- Total dirs: 24230
- Total files: 1032578
- Total blocks (validated): 1139580 (avg. block size 7914276 B)
- Minimally replicated blocks: 1139580 (100.0 %)
- Over-replicated blocks: 1 (8.775163E-5 %)
- Under-replicated blocks: 16 (0.001404026 %)
- Mis-replicated blocks: 0 (0.0 %)
- Default replication factor: 2
- Average block replication: 2.0122387
- Corrupt blocks: 0
- Missing replicas: 32 (0.0013954865 %)
- Number of data-nodes: 8
- Number of racks: 1
- FSCK ended at Tue Sep 27 18:57:23 EDT 2011 in 7453 milliseconds

- The filesystem under path '/' is HEALTHY

FileSystem Check (FSCK)

- Which blocks are in a file?

```
% hadoop fsck /user/bob/part-00008 -files -blocks -racks
```

```
/user/bob/part-00008 25582428 bytes, 1 block(s): OK 0. blk_-  
3724870485760122836_1035 len=25582428 repl=3 [/default-  
rack/10.251.43.2:50010, /default-rack/10.251.27.178:50010 ,  
/default-rack/10.251.123.163:50010]
```

Block Scanner

- `dfs.datanode.scan.period.hours` → Default = 504 hours=3 weeks

- <http://datanode:50075/blockScannerReport>

```
Total Blocks : 39422
Verified in last hour : 84
Verified in last day : 2329
Verified in last week : 9012
Verified in last four weeks : 38240
Verified in SCAN_PERIOD : 38240
Not yet verified : 1182
Verified since restart : 45394
Scans since restart : 7492
Scan errors since restart : 0
Transient scan errors : 0
Current scan rate limit KBps : 1024
Progress this period : 108%
Time left in cur period : 2.42%
```

Block Scanner

- <http://datanode:50075/blockScannerReport?listblocks>
blk_6035596358209321442 : status : ok type : none scan time
: 0 not yet verified blk_3065580480714947643 : status : ok
type : remote scan time : 1215755306400 2008-07-11
05:48:26,400 blk_8729669677359108508 : status : ok type :
local scan time : 1215755727345 2008-07-11 05:55:27,345

Balancer

- Balancer runs in background, default speed limit 1 MB/s hdfs-site.xml ---> dfs.balance.bandwidthPerSec

- % start-balancer.sh

```
Time Stamp Iteration# Bytes Already Moved Bytes Left To  
Move Bytes Being Moved
```

```
Mar 14, 2012 5:23:42 PM 0 0 KB 219.21 MB 150.29 MB
```

```
Mar 14, 2012 5:27:14 PM 1 124.24 MB 22.45 MB 150.29 MB
```

```
The cluster is balanced. Exiting...
```

```
Balancing took 32.03452 minutes
```

Thank You