Dataflair Web Services Pvt. Ltd.
Ph: +91-8451097879 / 7718877477
http://www.data-flair.com

Sar Log Analytics

Objective:

Analyze sar logs, generated by server to identify system resource utilization. The objective of the project is to get the insights of system resource utilization of all the nodes of the cluster. This application will be useful to identify issues of frameworks which work on a cluster (Hadoop, Greenplum, etc.). Using this app we can identify issues like if there is uneven load distribution when submitting a Map-Reduce Job or Greeplum query.

Steps:

Install sysstat

sudo apt-get install sysstat

This will be used for scripts which we will schedule to generate system logs

Develop scripts to generate system logs

Prepare a script to generate CPU utilization

```
echo "`hostname` `date +%d-%m-%y,%H:%M` ` sar 1 59 |tail -1 `" >> /home/hadoop/sar/logs/`date +%m-%y`-cpu-sar.txt &
```

Prepare a script to generate Memory utilization

```
echo "`hostname` `date +%d%m%y,%H:%M` ` sar -r 1 59 |tail -1 `" >> /home/hadoop/sar/logs/`date +%m-%y`-memory-sar.txt &
```

Prepare a script to generate Disk utilization

```
echo "`hostname` `date +%d%m%y,%H:%M` `df -h |head -3|tail -1 `" >> /home/hadoop/sar/logs/`date +%m-%y`-disk-sar.txt &
```

Schedule the scripts in crontab

```
crontab –e
* * * * * /PATH/TO/SCRIPT/CPU-SCRIPT
* * * * * /PATH/TO/SCRIPT/MEMORY-SCRIPT
* * * * * /PATH/TO/SCRIPT/DISK-SCRIPT
```

The scheduled scripts will generate logs in the respective files.

The format of logs:

CPU Logs:

```
hostname date-time                    CPU    %user %nice %system %iowait %steal %idle
```

```
hdtr001 240613,20:44 Average:      all     4.05      0.00     10.17      0.02      0.00     85.76
```

Memory Logs:

```
hostname date-time kbmemfree kbmemused  %memused kbbuffers  kbcached  kbcommit   %commit
kbactive   kbinact
hdtr001 240613,20:50 Average:      473633    319179     40.26     77812     63504    936325     71.31
208009     63161
```

Disk Logs:

```
hostname date-time              Size  Used Avail Use% Mounted on
hdtr001 240613,20:50              19G  2.9G   16G  16% /
```

Load the data in HDFS and process using Map-Reduce to identify average Cpu, Memory, Disk utilization.

Process CPU Logs

```
hadoop dfs –mkdir sarlogs/cpu-logs
hadoop dfs -put sar/logs/08-14-cpu-sar.txt sarlogs/cpu-logs
hadoop jar sarProcessorJob.jar org.tr.hd.log.cpu. SarCpuAggregator sarlogs/cpu-logs sarlogs/processed-cpu-logs
```

Process Memory Logs

```
hadoop dfs –mkdir sarlogs/memory-logs
hadoop dfs -put sar/logs/08-14-memory-sar.txt sarlogs/memory-logs
hadoop jar sarProcessorJob.jar org.tr.hd.log.mem.SarMemoryAggregator sarlogs/memory-logs sarlogs/processed-memory-logs
```

Process Disk Logs

```
hadoop dfs –mkdir sarlogs/disk-logs
hadoop dfs -put sar/logs/08-14-disk-sar.txt sarlogs/disk-logs
hadoop jar sarProcessorJob.jar org.tr.hd.log.disk.SarDiskAggregator sarlogs/disk-logs sarlogs/processed-disk-logs
```