

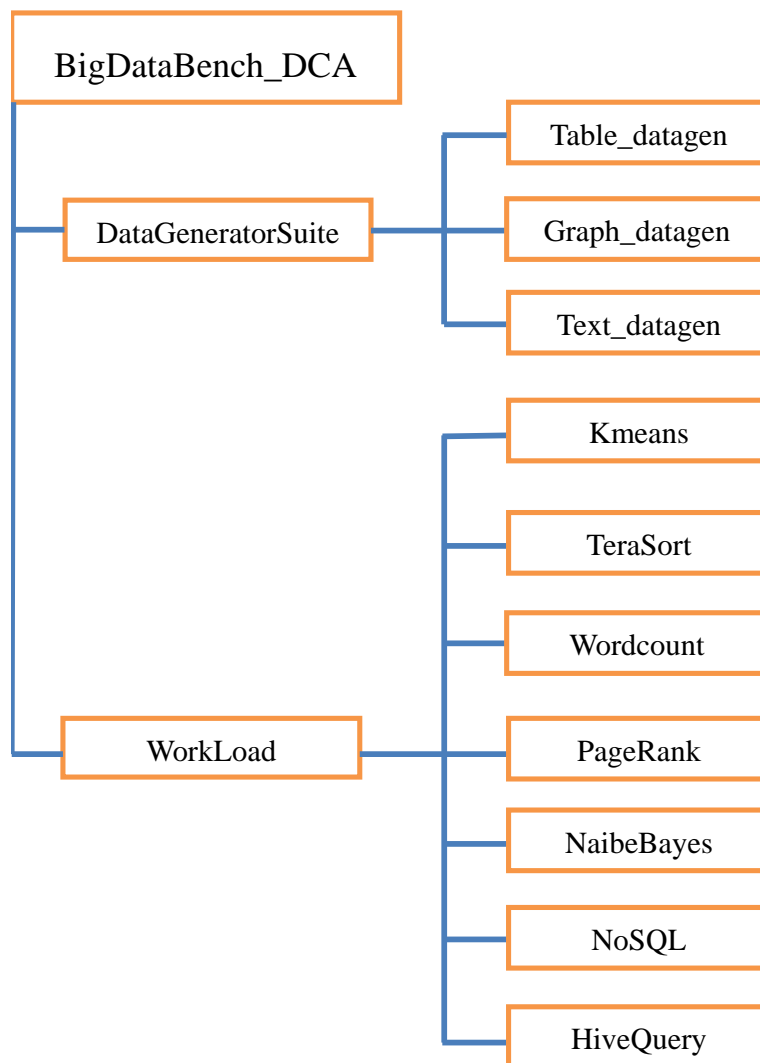
The background features three large, semi-transparent blue circles of varying sizes. Two thin, light blue lines intersect to form a large 'X' shape across the page. The text is centered and rendered in a dark red, serif font.

Big Data Bench DCA

User's Manual

Overview

BigDataBench_DCA is the subset of BigDataBench, the directory structure of BigDataBench_DCA packages is shown as the Figure 1. There are two sub-directory in the packages, one is the DataGeneratorSuite directory which included the packet of data generator suite, another is the Workload directory which included the packet of workloads.



Usage

2.1 Preparing environment

1. Decompression files

```
tar xzf BigDataBench_DCA.tar.gz
```

2. Enter the DIR

```
cd /BigDataBench_DCA
```

3. Source the environment

vim conf.properties and then add JAVA,Hadoop,Hbases,Mahout to path

```
source conf.properties
```

2.2 Generating data

1. `cd /BigDataBench_DCA/WorkLoad`

2. `sh generate_data.sh`

```
[root@hw086 WorkLoad]# ./generate_data.sh
Please select a number to choose the corresponding Workload algorithm
1. Wordcount Generate_data
2. TeraSort Generate_data
3. NaiveBayes Generate_data
4. Pagerank Generate_data
5. Kmeans Generate_data
6. HiveQuery Generate_data
Enter your choice : 1
```

You can choose the data what you want to generate. For example, you can enter “1” to generate the data of wordcount.

```
[root@hw111 WorkLoad]# ./generate_data.sh
Please select a number to choose the corresponding Workload algorithm
1. Wordcount Generate_data
2. TeraSort Generate_data
3. NaiveBayes Generate_data
4. Pagerank Generate_data
5. Kmeans Generate_data
6. HiveQuery Generate_data
Enter your choice : 1
ok. You chose 1 and we'll use Wordcount Workload
Preparing MicroBenchmarks data dir
WORK DIR=/home/ljw/test/BigDataBench_DCA_2/WorkLoad/Wordcount data will be put in /home/ljw/test/BigDataBench_DCA_2/WorkLoad/Wordcount/data-wordcount/in
Fri Feb 13 19:59:14 CST 2015
Fri Feb 13 19:59:15 CST 2015
```

2.3 Running workloads

1. `cd /BigDataBench_DCA/WorkLoad`
2. `sh run_workload.sh`

```
[root@hw086 WorkLoad]# ./run_workload.sh
Please select a number to choose the corresponding Workload algorithm
1. Wordcount Workload
2. TeraSort Workload
3. NaiveBayes Workload
4. Pagerank Workload
5. Kmeans Workload
6. HiveQuery Workload
7. NoSQL Workload
```

You can choose the workload what you want to run. For example, you can enter “1” to run the wordcount workload.

```
[root@hw111 WorkLoad]# ./run_workload.sh
Please select a number to choose the corresponding Workload algorithm
1. Wordcount Workload
2. TeraSort Workload
3. NaiveBayes Workload
4. Pagerank Workload
5. Kmeans Workload
6. HiveQuery Workload
7. NoSQL Workload
Enter your choice : 1
```

The wordcount workload can be run as the following:

```
15/02/13 20:16:24 INFO input.FileInputFormat: Total input paths to process : 20
15/02/13 20:16:24 INFO mapreduce.JobSubmitter: number of splits:20
15/02/13 20:16:24 WARN conf.Configuration: mapred.jar is deprecated. Instead, use mapreduce.job.jar
15/02/13 20:16:24 WARN conf.Configuration: mapred.output.value.class is deprecated. Instead, use mapreduce.job.output.value.class
15/02/13 20:16:24 WARN conf.Configuration: mapreduce.combine.class is deprecated. Instead, use mapreduce.job.combine.class
15/02/13 20:16:24 WARN conf.Configuration: mapreduce.map.class is deprecated. Instead, use mapreduce.job.map.class
15/02/13 20:16:24 WARN conf.Configuration: mapred.job.name is deprecated. Instead, use mapreduce.job.name
15/02/13 20:16:24 WARN conf.Configuration: mapreduce.reduce.class is deprecated. Instead, use mapreduce.job.reduce.class
15/02/13 20:16:24 WARN conf.Configuration: mapred.input.dir is deprecated. Instead, use mapreduce.input.fileinputformat.inputdir
15/02/13 20:16:24 WARN conf.Configuration: mapred.output.dir is deprecated. Instead, use mapreduce.output.fileoutputformat.outputdir
15/02/13 20:16:24 WARN conf.Configuration: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
15/02/13 20:16:24 WARN conf.Configuration: mapred.output.key.class is deprecated. Instead, use mapreduce.job.output.key.class
15/02/13 20:16:24 WARN conf.Configuration: mapred.working.dir is deprecated. Instead, use mapreduce.job.working.dir
15/02/13 20:16:25 INFO mapred.ResourceMgrDelegate: Submitted application application_1423829602840_0002 to ResourceManager at hw111/172.18.11.111:8090
15/02/13 20:16:25 INFO mapreduce.Job: The url to track the job: http://hw111:8088/proxy/application_1423829602840_0002/
15/02/13 20:16:25 INFO mapreduce.Job: Running job: job_1423829602840_0002
15/02/13 20:16:30 INFO mapreduce.Job: Job job_1423829602840_0002 running in uber mode : false
15/02/13 20:16:30 INFO mapreduce.Job: map 0% reduce 0%
15/02/13 20:16:35 INFO mapreduce.Job: map 40% reduce 0%
```