How to Use BigDataBench 4.0

Jianfeng Zhan, Chen Zheng, and Wanling Gao

http://prof.ict.ac.cn

ICT, Chinese Academy of Sciences

ASPLOS 2018, Williamsburg, VA, USA

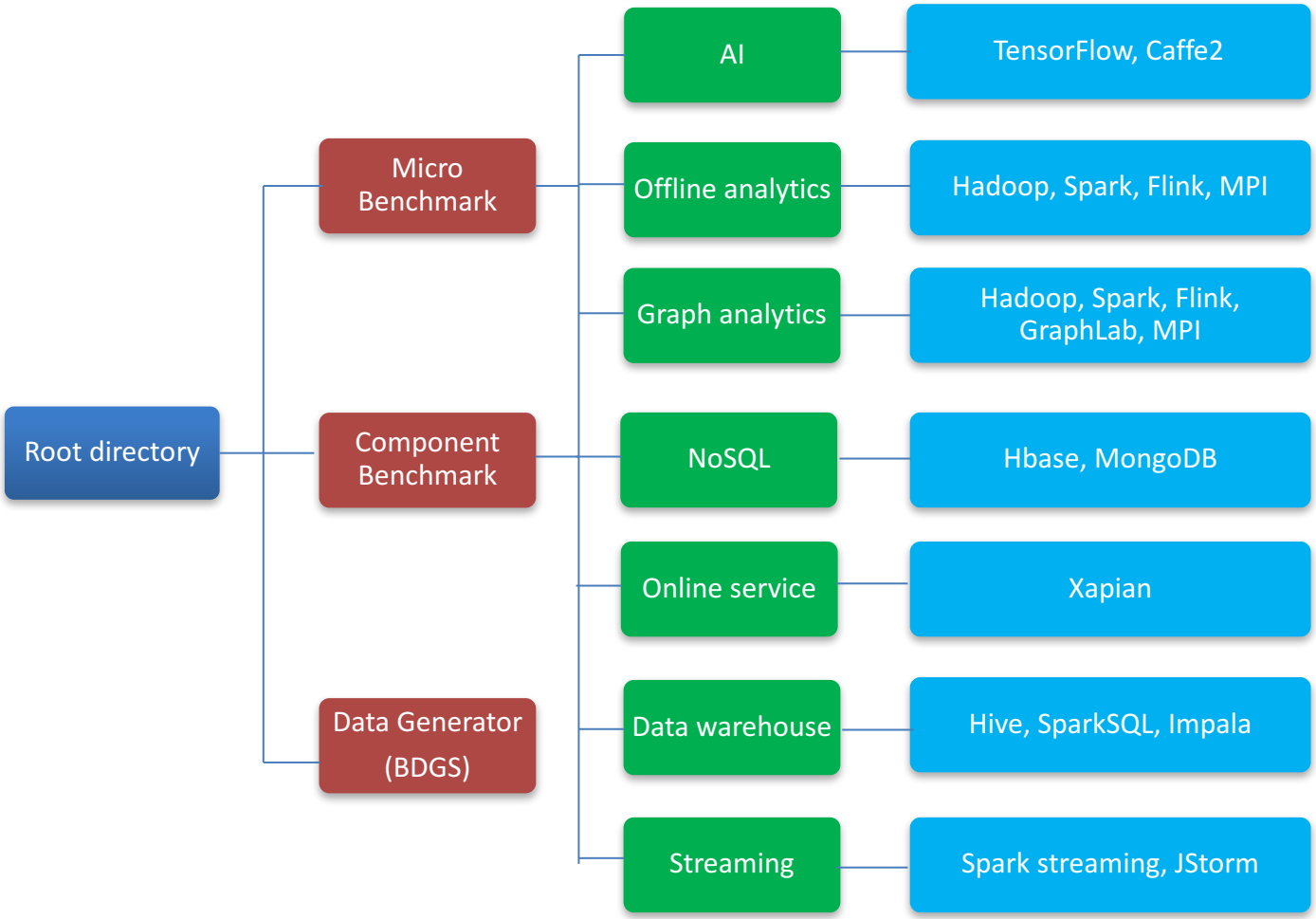
中科院计算所

INSTITUTE OF COMPUTING TECHNOLOGY

General Steps to Use BigDataBench

- Current release
  - Version 4.0 on http://prof.ict.ac.cn
- General steps to run the benchmarks
  - Prepare the package of BigDataBench
  - Prepare the environments of the selected software stack
  - Generate data sets as you need
    - *You can find a genDate* or a prepare* shell script in each directory of the benchmarks*
  - Run the scripts or commands (User Manual!)

Directory Structure



BDGS - Text

- Text_datagen
  - Wikipedia generator - 3 trained models
    - lda_wiki1w, wiki_1w5, wiki_noSW_90_Sampling
  - Amazon movie review generator – 2 models
    - amazonMR1, AMR1_noSW_95_Sampling
  - Use “gen_text_data.sh”

```
Text_datagen/gen_text_data.sh $MODEL_NAME $FILE_NUM $LINES_PER_FILE $WORDS_PER_LINE $Out_Dir
```

Wiki example:	e.g. lda_wiki1w	e.g. 10	e.g. 100	e.g. 10000
Amazon example:	e.g. amazonMR1	e.g. 10	e.g. 100	e.g. 10000

BDGS - Graph

- Graph_datagen
  - Kronecker Model
    - Weighted graph
    - Un-weighted graph

```
Amazon un-weighted graph:
    To Run: Graph_datagen/gen_weighted_graph.sh

Amazon weighted graph:
    To Run: Graph_datagen/gen_kronecker_graph -o:amazon_gen_16.txt -m:"0.9532 0.5502; 0.4439 0.2511" -i:16

Facebook graph:
    To Run: Graph_datagen/gen_kronecker_graph -o:facebook_g_16.txt -m:"0.9999 0.5887; 0.6254 0.3676" -i:16

Google graph :
    To Run: Graph_datagen/gen_kronecker_graph -o:google_g_16.txt -m:"0.8305 0.5573; 0.4638 0.3021" -i:16
```

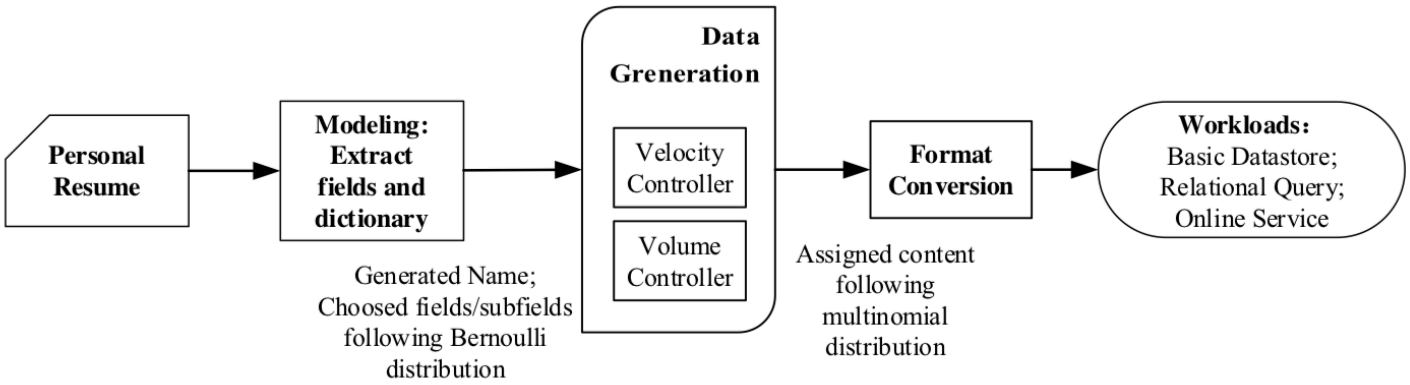
e.g. kronecker model parameter

Vertex: 2^16

BDGS - Table

- Table_datagen
  - E-commerce data generation
    - PDGF: uses XML configuration files for data description and distribution `Table_datagen/e-com/generate_table.sh`
  - Personal Resume generation

```
Table_datagen/personal_generator/gen_resume.sh "NUM_RESUME" "NUMBER_OF_FILES" "OUT_DATA_DIR"
```



Micro Benchmark

- Offline analytics & Graph analytics
- Streaming

Micro Benchmark	Involved Dwarf	Application Domain	Workload Type	Data Set	Software Stack
Sort	Sort	SE, SN, EC, MP, BI	Offline analytics	Wikipedia entries	Hadoop, Spark, Flink, MPI
Grep	Set		Offline analytics	Wikipedia entries	Hadoop, Spark, Flink, MPI
			Streaming	Random Generate	Spark streaming
WordCount	Basic statistics		Offline analytics	Wikipedia entries	Hadoop, Spark, Flink, MPI
MD5	Logic		Offline analytics	Wikipedia entries	Hadoop, Spark, MPI
Connected Component	Graph	SN	Graph analytics	Facebook social network	Hadoop, Spark, Flink, GraphLab, MPI
RandSample	Sampling	SE, MP, BI	Offline analytics	Wikipedia entries	Hadoop, Spark, MPI
FFT	Transform	MP	Offline analytics	Two-dimensional matrix	Hadoop, Spark, MPI
Matrix Multiply	Matrix	SE, SN, EC, MP, BI	Offline analytics	Two-dimensional matrix	Hadoop, Spark, MPI

Offline Analytics - RandSample

- Target: run RandSample microbenchmark
- General steps:

```
[root@hw125 tpds2017]# $HADOOP_HOME/bin/hadoop jar dwarf-hadoop/sample-operations/RandSample/out/artifacts
t1 0.2
18/03/22 16:30:16 INFO client.RMProxy: Connecting to ResourceManager at hw125/172.18.11.125:8032
18/03/22 16:30:16 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Im
medy this.
18/03/22 16:30:17 INFO input.FileInputFormat: Total input paths to process : 200
18/03/22 16:30:17 INFO mapreduce.JobSubmitter: number of splits:800
18/03/22 16:30:17 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1521170223299_0040
18/03/22 16:30:17 INFO impl.YarnClientImpl: Submitted application application_1521170223299_0040
18/03/22 16:30:17 INFO mapreduce.Job: The url to track the job: http://hw125:8088/proxy/application_152117
18/03/22 16:30:17 INFO mapreduce.Job: Running job: job_1521170223299_0040
18/03/22 16:30:21 INFO mapreduce.Job: Job job_1521170223299_0040 running in uber mode : false
18/03/22 16:30:21 INFO mapreduce.Job:  map 0% reduce 0%
```

  - hadoop jar RandSample.jar RandSample <*input*> <*output*> <*sample_ratio*>

Offline Analytics – FFT example

- Target: run “FFT” micro benchmark using hadoop
- General steps:
  - Prepare Hadoop environment
  - Prepare matrix data
    - cd /BigDataBench_V4.0_Hadoop/MicroBenchmark/OfflineAnalytics/FFT
    - sh genData_FFT.sh <log2_co>: (auto-generated by run_FFT.sh)

```
root@s002:/home/BigDataBench_V4.0_Hadoop/MicroBenchmark/OfflineAnalytics/FFT# ./run_FFT.sh
Java HotSpot(TM) Server VM warning: ignoring option MaxPermSize=512m; support was removed in 8.0
Java HotSpot(TM) Server VM warning: ignoring option MaxPermSize=512m; support was removed in 8.0
Java HotSpot(TM) Server VM warning: ignoring option MaxPermSize=512m; support was removed in 8.0
Java HotSpot(TM) Server VM warning: ignoring option MaxPermSize=512m; support was removed in 8.0
Java HotSpot(TM) Server VM warning: You have loaded library /home/w/hadoop-2.7.1/lib/native/libhadoop.so.1.0.0 which might have disabled stack guard. The VM will try to fix
 the stack guard now.
It's highly recommended that you fix the library with 'execstack -c <libfile>', or link it with '-z noexecstack'.
18/03/22 15:17:13 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/03/22 15:17:14 INFO client.RMProxy: Connecting to ResourceManager at /192.168.1.202:8032
18/03/22 15:17:14 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolR
unner to remedy this.
18/03/22 15:17:15 INFO input.FileInputFormat: Total input paths to process : 1
18/03/22 15:17:15 INFO mapreduce.JobSubmitter: number of splits:1
18/03/22 15:17:15 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1521686472463_0004
18/03/22 15:17:15 INFO impl.YarnClientImpl: Submitted application application_1521686472463_0004
18/03/22 15:17:15 INFO mapreduce.Job: The url to track the job: http://s002:8088/proxy/application_1521686472463_0004/
18/03/22 15:17:15 INFO mapreduce.Job: Running job: job_1521686472463_0004
18/03/22 15:17:20 INFO mapreduce.Job: Job job_1521686472463_0004 running in uber mode : false
18/03/22 15:17:20 INFO mapreduce.Job:  map 0% reduce 0%
18/03/22 15:17:24 INFO mapreduce.Job:  map 100% reduce 0%
18/03/22 15:17:30 INFO mapreduce.Job:  map 100% reduce 100%
18/03/22 15:17:30 INFO mapreduce.Job: Job job_1521686472463_0004 completed successfully
18/03/22 15:17:30 INFO mapreduce.Job: Counters: 49
```

Streaming – Grep example

- Target: run grep benchmark using Spark streaming
- General steps:
  - Prepare Spark streaming environment
  - cd /BigDataBench_V4.0_Streaming/MicroBenchmark/Streaming/Grep
  - ./run-sparkstreaming-grep.sh

```
#########################################################################
#Usage: Grep <numStreams> <host> <port> <batchMillis>
# *   <numStream> is the number rawNetworkStreams, which should be same as number
# *               of work nodes in the cluster
# *   <host> is "localhost".
# *   <port> is the port on which RawTextSender is running in the worker nodes.
# *   <batchMillise> is the Spark Streaming batch duration in milliseconds.
#########################################################################
```

Micro Benchmark

- AI

Convolution	Transform	SN, EC, MP, BI	AI	Cifar, ImageNet	TensorFlow, pyTorch	Caffe,
Fully Connected	Matrix	SN, EC, MP, BI	AI	Cifar, ImageNet	TensorFlow, pyTorch	Caffe,
Relu	Logic	SN, EC, MP, BI	AI	Cifar, ImageNet	TensorFlow, pyTorch	Caffe,
Sigmoid	Matrix	SN, EC, MP, BI	AI	Cifar, ImageNet	TensorFlow, pyTorch	Caffe,
Tanh	Matrix	SN, EC, MP, BI	AI	Cifar, ImageNet	TensorFlow, pyTorch	Caffe,
MaxPooling	Sampling	SN, EC, MP, BI	AI	Cifar, ImageNet	TensorFlow, pyTorch	Caffe,
AvgPooling	Sampling	SN, EC, MP, BI	AI	Cifar, ImageNet	TensorFlow, pyTorch	Caffe,
CosineNorm [36]	Basic Statistics	SN, EC, MP, BI	AI	Cifar, ImageNet	TensorFlow, pyTorch	Caffe,
BatchNorm [37]	Basic Statistics	SN, EC, MP, BI	AI	Cifar, ImageNet	TensorFlow, pyTorch	Caffe,
Dropout [38]	Sampling	SN, EC, MP, BI	AI	Cifar, ImageNet	TensorFlow, pyTorch	Caffe,

AI – Conv2d example

- Target: run conv2d micro benchmark using TensorFlow
- General steps:
  - Prepare TensorFlow environment
  - Prepare image data
  - Config image directory in conv2d.py
  - python conv2d.py

Micro Benchmark

- NoSQL

Read	Set	SE, SN, EC	NoSQL	ProfSearch resumes	HBase, MongoDB
Write	Set	SE, SN, EC	NoSQL	ProfSearch resumes	HBase, MongoDB
Scan	Set	SE, SN, EC	NoSQL	ProfSearch resumes	HBase, MongoDB

NoSQL – Write example

- Target: run “write” operations using HBase
- General steps:
  - Prepare HBase according to the office guide
    - sh /hbase-0.94.5/bin/hbase shell
    - create 'usertable','f1','f2','f3'
  - Prepare YCSB as the workload generator
    - YCSB is in the directory of BasicDatastoreOperaOons/ycsb-0.1.4
  - Run YCSB commands like this:
    - • sh bin/ycsb load hbase -P workloads/workloadc -p threads=<thread-numbers> -p columnfamily=<family> -p recordcount=<recordcount-value> -p hosts=<hosOp> -s>load.dat

Component Benchmark

- AI

Alexnet	Matrix, Transform, Sampling, Logic, Basic statistics	SN, MP, BI	AI	Cifar, ImageNet	TensorFlow, pyTorch	Caffe,
Googlenet		SN, MP, BI	AI	Cifar, ImageNet	TensorFlow, pyTorch	Caffe,
Resnet		SN, MP, BI	AI	Cifar, ImageNet	TensorFlow, pyTorch	Caffe,
Inception Resnet V2		SN, MP, BI	AI	Cifar, ImageNet	TensorFlow, pyTorch	Caffe,
VGG16		SN, MP, BI	AI	Cifar, ImageNet	TensorFlow, pyTorch	Caffe,
DCGAN		SN, MP, BI	AI	LSUN	TensorFlow, pyTorch	Caffe,
WGAN		SN, MP, BI	AI	LSUN	TensorFlow, pyTorch	Caffe,
GAN	Matrix, Sampling, Logic, Basic statistics	SN, MP, BI	AI	LSUN	TensorFlow, pyTorch	Caffe,
Seq2Seq		SE, EC, BI	AI	TED Talks	TensorFlow, pyTorch	Caffe,
Word2vec	Matrix, Basic statistics, Logic	SE, SN, EC	AI	Wikipedia entries, Sogou data	TensorFlow, pyTorch	Caffe,

AI – Alexnet Example

- Target: run “Alexnet” micro benchmark using Tensorflow
- General steps:
  - Prepare Tensorflow environment
  - Run Alexnet:
    - cd /BigDataBench_V4.0_Tensorflow/ComponentBenchmark/AI/Alexnet
    - python alexnet_cifar10.py
    - Choosing CPU or GPU environment

```
root@s002:/home/BigDataBench_V4.0_Tensorflow/ComponentBenchmark/AI/Alexnet# python alexnet_cifar10.py
hdf5 is not supported on this machine (please install/reinstall h5py for optimal experience)
Downloading CIFAR 10, Please wait...
100.0% 170500096 / 170498071
('Succesfully downloaded', 'cifar-10-python.tar.gz', 170498071, 'bytes.')
File Extracted in Current Directory
WARNING:tensorflow:From /usr/local/lib/python2.7/dist-packages/tflearn/initializations.py:119: __init__ (from tensorflow.python.ops.init_ops) is deprecated and will be remo
ved in a future version.
Instructions for updating:
Use tf.initializers.variance_scaling instead with distribution=uniform to get equivalent behavior.
WARNING:tensorflow:From /usr/local/lib/python2.7/dist-packages/tflearn/objectives.py:66: calling reduce_sum (from tensorflow.python.ops.math_ops) with keep_dims is deprecat
ed and will be removed in a future version.
Instructions for updating:
keep_dims is deprecated, use keepdims instead
---------------------------------
Run id: alexnet_cifar10
Log directory: /tmp/tflearn_logs/
---------------------------------
Training samples: 50000
Validation samples: 10000
--
Training Step: 200  | total loss: 2.29368 | time: 284.832s
| Momentum | epoch: 001 | loss: 2.29368 - acc: 0.1339 | val_loss: 2.29204 - val_acc: 0.1292 -- iter: 12800/50000
--
Training Step: 400  | total loss: 2.26979 | time: 562.061s
| Momentum | epoch: 001 | loss: 2.26979 - acc: 0.1446 | val_loss: 2.26635 - val_acc: 0.1243 -- iter: 25600/50000
--
Training Step: 415  | total loss: 2.26876 | time: 582.255s
| Momentum | epoch: 001 | loss: 2.26876 - acc: 0.1563 -- iter: 26560/50000
```

Component Benchmark

- Offline analytics & Graph analytics
- Streaming

PageRank	Matrix, Sort, Basic statistics, Graph	SE	Graph analytics	Google web graph	Hadoop, Spark, Flink, GraphLab, MPI
Index	Logic, Sort, Basic statistics, Set	SE	Offline analytics	Wikipedia entries	Hadoop, Spark
Rolling top words	Sort, Basic statistics	SN	Streaming	Random generate	Spark streaming, JStorm
Kmeans	Matrix, Sort, Basic statistics	SE, SN, EC, MP, BI	Offline analytics	Facebook social network	Hadoop, Spark, Flink, MPI
			Streaming	Random generate	Spark streaming
Collaborative Filtering	Graph, Matrix	EC	Offline analytics	Amazon movie review	Hadoop, Spark
		EC	Streaming	MovieLens dataset	JStorm
Naive Bayes	Basic statistics, Sort	SE, SN, EC	Offline analytics	Amazon movie review	Hadoop, Spark, Flink, MPI
SIFT	Matrix, Sampling, Transform, Sort	MP	Offline analytics	ImageNet	Hadoop, Spark, MPI
LDA	Matrix, Graph, Sampling	SE	Offline analytics	Wikipedia entries	Hadoop, Spark, MPI

Offline Analytics – SIFT example

- Target: run “SIFT” component benchmark using hadoop
- General steps:
  - Prepare Hadoop environment
  - Prepare SIFT data
    - cd /BigDataBench_V4.0_Hadoop/ComponentBenchmark/OfflineAnalytics/SIFT
    - Put the image data under SIFT directory

```
[root@hw125 hpca2018]# hadoop jar /home/gwl/hipi/tools/sift/build/libs/sift.jar /img-data/image9G.hib /image-out/siftout
18/03/22 15:52:49 INFO client.RMProxy: Connecting to ResourceManager at hw125/172.18.11.125:8032
18/03/22 15:52:50 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to re
medy this.
18/03/22 15:52:50 INFO input.FileInputFormat: Total input paths to process : 1
Spawned 69map tasks
18/03/22 15:52:50 INFO mapreduce.JobSubmitter: number of splits:69
18/03/22 15:52:51 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1521170223299_0039
18/03/22 15:52:51 INFO impl.YarnClientImpl: Submitted application application_1521170223299_0039
18/03/22 15:52:51 INFO mapreduce.Job: The url to track the job: http://hw125:8088/proxy/application_1521170223299_0039/
18/03/22 15:52:51 INFO mapreduce.Job: Running job: job_1521170223299_0039
18/03/22 15:52:56 INFO mapreduce.Job: Job job_1521170223299_0039 running in uber mode : false
18/03/22 15:52:56 INFO mapreduce.Job:  map 0% reduce 0%
```

hadoop jar sift.jar <out.hib> <outsif>

<out.hib>:genData_SIFT.sh generate data

<outsif>:the result to save path

Streaming – Kmeans example

- Target: run kmeans benchmark using Spark streaming
- General steps:
  - Prepare Spark streaming environment
  - cd /BigDataBench_V4.0_Streaming/ ComponentBenchmark/Streaming/Kmeans
  - ./run-sparkstreaming-kmeans.sh

```
#########################################################################################
#StreamingKMeans <trainingDir> <testDir> <batchDuration> <numClusters> <numDimensions>
#<trainingDir>: The directory used to save the training file
#<testDir>: The directory used to save the test file
#<batchDuration>: Batch's time interval
#<numClusters>: The number of clusters
#<numDimensions>: The dimension of vector (which is number of columns), it needs to be same as " columnNum" in
#                 KMeansTestDataGenerator.sh and KMeansTrainDataGenerator.sh
#########################################################################################
```

Graph Analytics – PageRank

- Target: run “PageRank” component benchmark using hadoop
- General steps:

```
root@s002:/home/BigDataBench_V4.0_Hadoop/ComponentBenchmark/GraphAnalytics/PageRank# ./run_PageRank.sh 1
[#_of_nodes] : number of nodes in the graph is 2
[#_of_reducers] : number of reducers to use in hadoop is 16
[makesym or nosym] : makesym-duplicate reverse edges, nosym-use original edge file makesym
[#_Iterations_of_GenGragh] : Iterations_of_GenGragh is 1
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.
-----===[PEGASUS: A Peta-Scale Graph Mining System]===-----

[PEGASUS] Computing PageRank. Max iteration = 1024, threshold = 0.1, cur_iteration=1

Creating initial pagerank vectors....
Java HotSpot(TM) Server VM warning: You have loaded library /home/w/hadoop-2.7.1/lib/native/libhadoop.so.1.0.0 which might have disabled stack guard. The VM will try to fix
 the stack guard now.
It's highly recommended that you fix the library with 'execstack -c <libfile>', or link it with '-z noexecstack'.
18/03/22 15:21:41 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/03/22 15:21:42 INFO client.RMProxy: Connecting to ResourceManager at /192.168.1.202:8032
18/03/22 15:21:42 INFO client.RMProxy: Connecting to ResourceManager at /192.168.1.202:8032
18/03/22 15:21:42 INFO mapred.FileInputFormat: Total input paths to process : 2
18/03/22 15:21:42 INFO mapreduce.JobSubmitter: number of splits:3
18/03/22 15:21:42 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1521686472463_0006
18/03/22 15:21:43 INFO impl.YarnClientImpl: Submitted application application_1521686472463_0006
18/03/22 15:21:43 INFO mapreduce.Job: The url to track the job: http://s002:8088/proxy/application_1521686472463_0006/
18/03/22 15:21:43 INFO mapreduce.Job: Running job: job_1521686472463_0006
18/03/22 15:21:48 INFO mapreduce.Job: Job job_1521686472463_0006 running in uber mode : false
18/03/22 15:21:48 INFO mapreduce.Job:  map 0% reduce 0%
18/03/22 15:21:53 INFO mapreduce.Job:  map 33% reduce 0%
18/03/22 15:21:57 INFO mapreduce.Job:  map 67% reduce 0%
18/03/22 15:22:01 INFO mapreduce.Job:  map 100% reduce 0%
18/03/22 15:22:05 INFO mapreduce.Job:  map 100% reduce 13%
18/03/22 15:22:09 INFO mapreduce.Job:  map 100% reduce 25%
18/03/22 15:22:13 INFO mapreduce.Job:  map 100% reduce 38%
18/03/22 15:22:17 INFO mapreduce.Job:  map 100% reduce 50%
18/03/22 15:22:18 INFO mapreduce.Job:  map 100% reduce 100%
18/03/22 15:22:19 INFO mapreduce.Job: Job job_1521686472463_0006 failed with state KILLED due to: Kill job job_1521686472463_0006 received from root (auth:SIMPLE) at 192.16
8.1.202
Job received Kill while in RUNNING state.
```

Online Service – Xapian (cont')

- Target: run searching using Xapian
- General steps:
  - 3) Online searching
    - Run xapian/run_networked.sh

```
[root@hw114 xapian]# ./run_networked.sh
TBENCH_SERVER = TBENCH_CLIENT_THREADS = Client left, removing
All clients exited. Server finishing
```

Online Service – Xapian

- Target: run searching using Xapian
- General steps:
  - 1) Install Xapian according to user manual
    - ./build.sh to install harness (gcc version > 4.8)
    - xapian/build.sh to install xapian

```
[root@hw114 xapian]# ./build.sh
g++ -o xapian_integrated main.o server.o client.o ../harness/client.o ../harness/tbench_server_integrated.o
`./xapian-core-1.2.13/install/bin/xapian-config --libs` -lpthread -lrt
g++ -o xapian_networked_server main.o server.o ../harness/tbench_server_networked.o `./xapian-core-1.2.13/in
stall/bin/xapian-config --libs` -lpthread -lrt
g++ -o xapian_networked_client client.o ../harness/client.o ../harness/tbench_client_networked.o `./xapian-c
ore-1.2.13/install/bin/xapian-config --libs` -lpthread -lrt
```

Online Service – Xapian (cont')

- Target: run searching using Xapian
- General steps:
  - 2) Configuration
    - vim xapian/run_networked.sh

```
#The server number configuration
NSERVERS=12
#Queries per second
QPS=5000
#Warmup configuration: the query numbers before testing
WARMUPREQS=10
#The total queries
REQUESTS=1000000
```

Component Benchmark

- Data warehouse

OrderBy	Set, Sort	EC	Data warehouse	E-commerce transaction	Hive, Spark-SQL, Impala
Aggregation	Set, Basic statistics	EC	Data warehouse	E-commerce transaction	Hive, Spark-SQL, Impala
Project	Set	EC	Data warehouse	E-commerce transaction	Hive, Spark-SQL, Impala
Filter	Set	EC	Data warehouse	E-commerce transaction	Hive, Spark-SQL, Impala
Select	Set	EC	Data warehouse	E-commerce transaction	Hive, Spark-SQL, Impala
Union	Set	EC	Data warehouse	E-commerce transaction	Hive, Spark-SQL, Impala

Data Warehouse – Select example

- Target: run “Select” benchmark using hadoop hive
- General steps:
  - Prepare Hadoop and hive environment
  - Run the data generation script

```
root@s002:/home/BigDataBench_V4.0_Hadoop/ComponentBenchmark/Datawarehouse/Select# ./run_Select.sh
Java HotSpot(TM) Server VM warning: ignoring option MaxPermSize=512m; support was removed in 8.0
Java HotSpot(TM) Server VM warning: ignoring option MaxPermSize=512m; support was removed in 8.0
Java HotSpot(TM) Server VM warning: ignoring option MaxPermSize=512m; support was removed in 8.0
Java HotSpot(TM) Server VM warning: ignoring option MaxPermSize=512m; support was removed in 8.0
Java HotSpot(TM) Server VM warning: ignoring option MaxPermSize=512m; support was removed in 8.0
Java HotSpot(TM) Server VM warning: ignoring option MaxPermSize=512m; support was removed in 8.0
Query ID = root_20180322153226_5bbb36f8-d227-4cd5-b0a9-ecdc6e2fff51
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1521686472463_0007, Tracking URL = http://s002:8088/proxy/application_1521686472463_0007/
Kill Command = /home/w/hadoop-2.7.1/bin/hadoop job  -kill job_1521686472463_0007
Hadoop job information for Stage-1: number of mappers: 3; number of reducers: 0
2018-03-22 15:32:35,610 Stage-1 map = 0%,  reduce = 0%
2018-03-22 15:32:49,169 Stage-1 map = 17%,  reduce = 0%, Cumulative CPU 13.98 sec
2018-03-22 15:32:54,319 Stage-1 map = 33%,  reduce = 0%, Cumulative CPU 19.12 sec
2018-03-22 15:33:06,754 Stage-1 map = 50%,  reduce = 0%, Cumulative CPU 33.22 sec
2018-03-22 15:33:10,863 Stage-1 map = 67%,  reduce = 0%, Cumulative CPU 36.9 sec
2018-03-22 15:33:23,253 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 50.27 sec
```

Conclusion

- Website： http://prof.ict.ac.cn

- Please refer to user manual for more details !


QUESTIONS
And
Answers