1 Experiments

1.1 E2017120401: Baseline Performance Evaluation

1.1.1 Purpose

To understand the baseline performance by using the queries in ADBIS submission for XMark600.

1.1.2 Settings

- **Hardware** HaoDesk (see 2.1.1)
- **Software** BaseX 6.8.7, Java 1.8.0_151.
- **XML Dataset** XMark600.xml (see Table 2) sized 66.9 GB and a BaseX databases 'xmark600' is created by the server using command:
  `create db xmark600 xmark600.xml`
- **Queries** xm1.org – xm6.org (see).

1.1.3 Experiment Design

A BaseX instance Server first runs in server mode started by the following command on HaoDesk.

```
java -Xmx4g -xms2g -cp BaseX897.jar org.basex.BaseXServer
```

Note the databases in Server are NOT in main memory mode.

Then, a java program APP runs on HaoDesk in charge of sending an input query to Server via local network and saving results returned from Server to memory(short for mem) or disk depending on settings. An input query Query that will be processed in APP is first rewritten into the following XQuery expression:

```
for $node in db:open('xmark600')Query return $node
```

The results are stored either in memory or on disk depending on APP's settings: a) memory means the results are stored in memory and then discarded after the experiments. b) disk means the results are stored in disk and will be preserved after the experiments. One more thing, the maximum memory for APP was set to 12 GB.

1.1.4 Experiment Results

Timing The execution time is measured in APP. The time period between starting sending a query and finishing receiving the results is measured as execution time. Each query is evaluated 5 times.

Process Results We removed the results of the first run and take the average of the rest as the final execution time listed in Table 1.

Table 1: Experiment Results.

query	storage	time(s)	result size
xm1.org	disk	3191.60	60,048,845,586
	memory	N/A	
xm2.org	disk	0.01	0
	memory	0.01	
xm3.org	disk	71.25	922,270,281
	memory	73.34	
xm4.org	disk	113.05	1,583,959,305
	memory	113.84	
xm5.org	disk	83.59	989,346,990
	memory	88.75	
xm6.org	disk	78.42	1,351,708,787
	memory	78.90	

Original Data

All the original results containing execution time of queries and scripts used in the experiments are stored to `experiments/E2017120401` relative to the current folder that stores this report.

Note: The result of xm2.org is always empty (still under investigation).

1.1.5 Observations

- **Storage has small influence on execution time**

  We noticed one thing that the execution time is pretty similar for all the available queries. This is because the bottleneck is on the worker's side but not on the master's side. For example, for xm4.org, it takes 113s to receive about 1500 MB data, i.e. around 13.36 MB/s, which is much slower than the maximum speed of both memory and disk. Thus, the performance are much similar. We also notice that for some queries such as xm3.org and xm5.org, in-memory case is even a bit slower than on-disk case, one possible explanation is that the time was taken by calling System.gc().

- **The execution time is steady**

  Compared with the ADBIS study, the execution time of each run is more steady. My explanation to this result is that for very large scale of data, the fluctuation has a weaker influence on the execution time, which increased from milliseconds to seconds.

Table 2: Statistics of XMark data sets

key	# elements	# attributes	# context	total nodes	file size
xmark1	1,666,315	381,878	1,173,732	3,221,925	113.06 MB
xmark600	1,002,327,042	229,871,111	705,824,967	1,938,023,120	66.99 GB

2 Environments

2.1 Computers

2.1.1 HaoDesk

CPU: Intel Core (TM) i7 3930K@3.2 GHz (turbo to 3.8 GHz), 6 cores 12 threads
Memory: 32 GB DDR3 1333 GHz
Disk: 256GB SSD + 4TB HDD
Windows 7 Professional SP1 64bit

2.1.2 Matsu98

CPU: Two Intel Xeon E5-2620 v3, 6 cores, 2.4GHz, Hyper-Threading off
Memory: 32 GB DDR4 1866 GHz
System: Linux ?.

3 DataSets

3.1 XMark Dataset

The XMark datasets are listed in Table 2.