TPC-DS Queries executed:

Query Engine	Query number
Hawq	3, 7, 19, 27, 34, 42, 43, 46, 52, 53, 55, 59, 63, 65, 68, 79, 89, 98, 99
Spark	3, 7, 19, 27, 34, 42, 43, 46, 52, 53, 55, 59, 63, 65, 68, 79, 89, ssmax

NOTE: No manipulation of queries were performed.

Experimental Setup:

Operating System: Ubuntu14.04

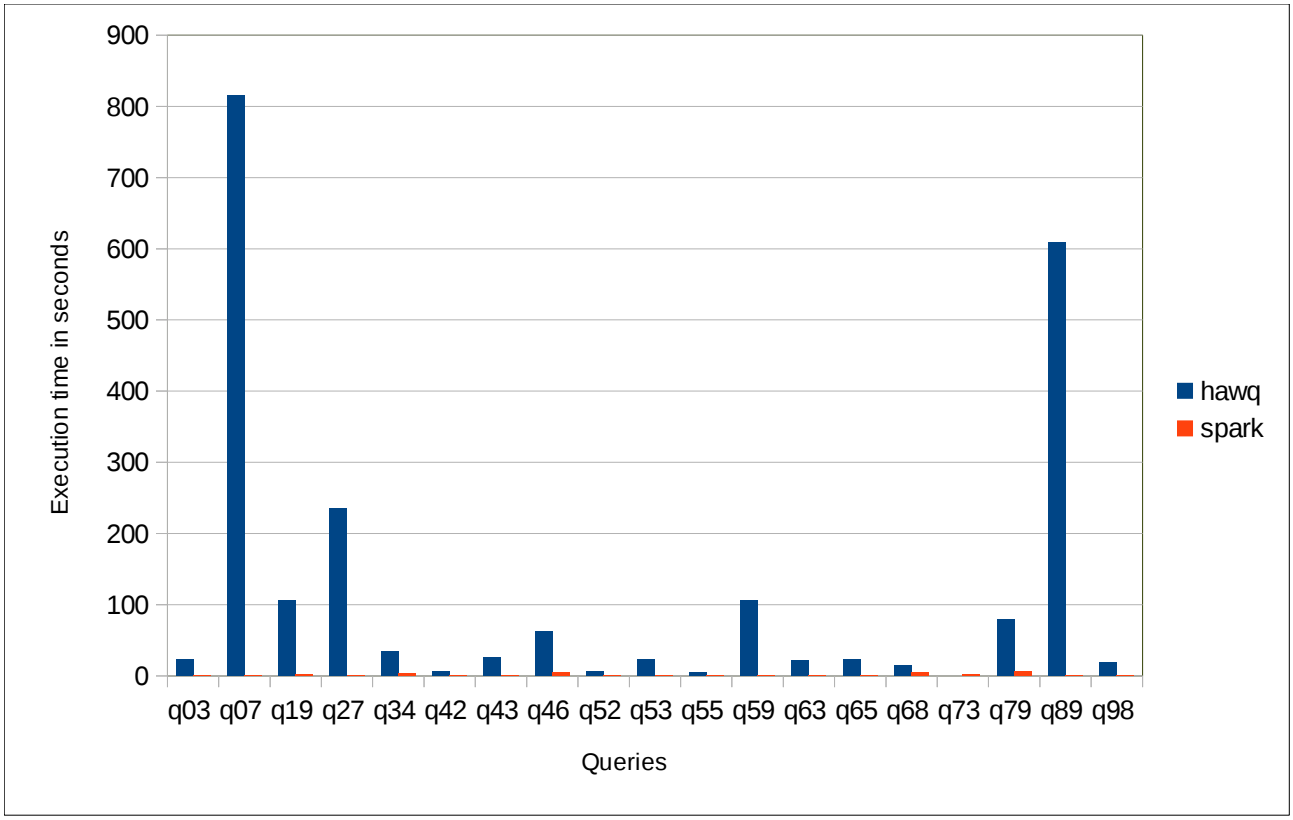
	Processor details	**No of cores**	**Memory per node**	**Disks per node**	**HDD Space after RAID**	**Space Allocated for Benchmark Execution**
Master-1	Intel Xeon E5-4640v2@2.20GHz	CPU(s): 80 Thread(s)/core: 2 Core(s)/socket: 10 Socket(s): 4	524GB	4*1TB	2.725TB	1TB
Slaves-5	Intel Xeon E5-2640v2@2.60GHz	CPU(s): 32 Thread(s)/core: 2 Core(s)/socket: 8 Sockets: 2	128GB	4*1TB	2.725TB	1TB

Note: Only the slaves were involved in data load & execution of the queries. Master was excluded to act as workers in Spark and executing-segment in Hawq. Same 5 nodes hosted Hadoop and Hawq/Spark.

Experiment Results:

Versions: Spark 2.1
Hawq 2.0.0
Hadoop 2.7.3

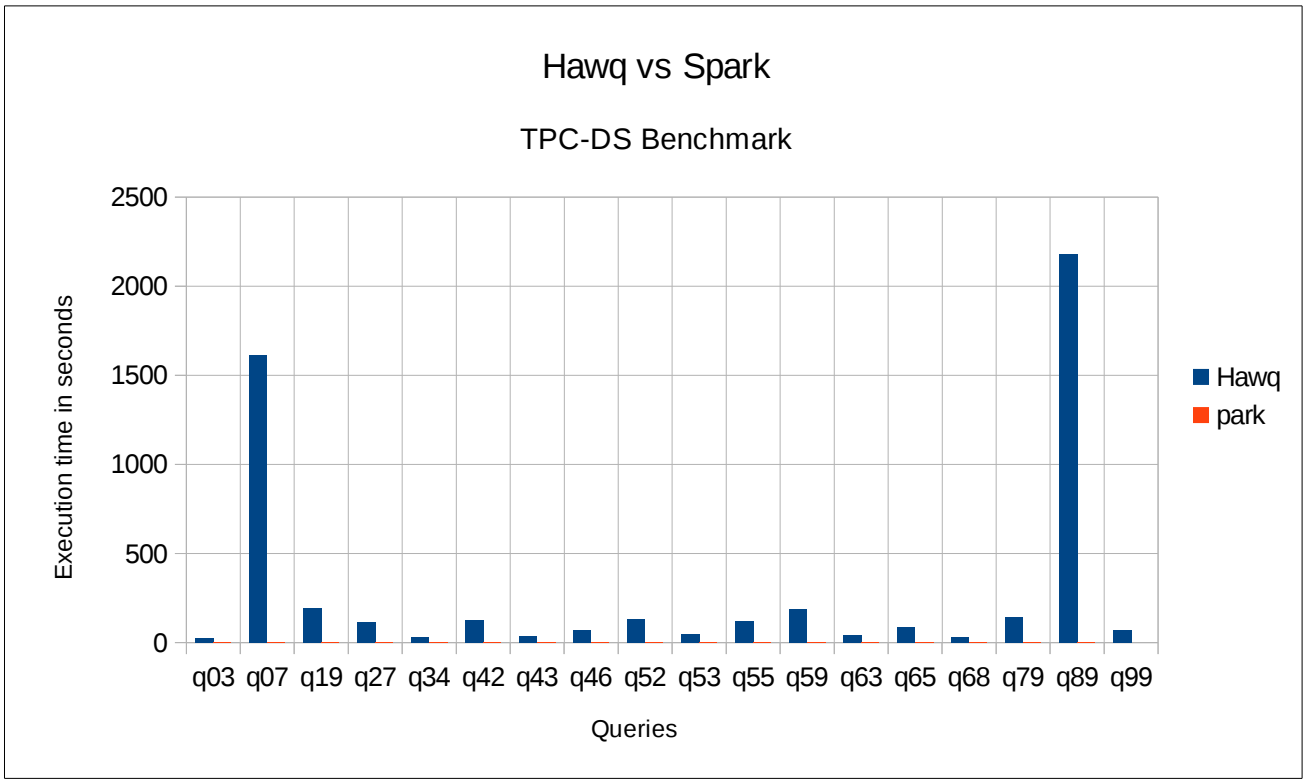
Test 1
DataLoad : 500GB



Query execution time of Hawq & Spark in seconds for 500GB dataload

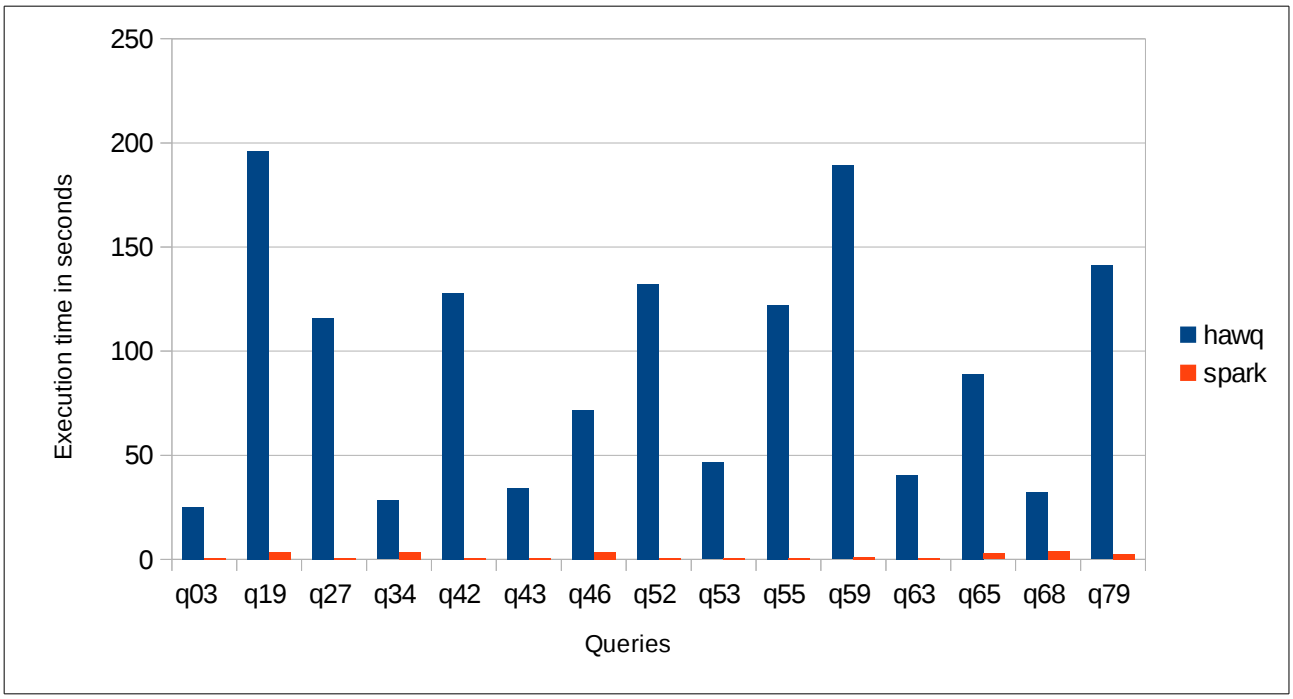
Queries	hawq (in seconds)	spark (in seconds)
q03	22.23	0.66
q07	814.82	0.91
q19	106.11	2.63
q27	235.24	0.81
q34	34.34	3.30
q42	6.63	0.41
q43	25.26	0.43
q46	61.61	4.34
q52	5.59	0.35
q53	22.23	0.47
q55	4.48	0.41
q59	106.11	1.07
q63	21.21	0.58
q65	22.23	1.09
q68	14.15	4.32
q73	0	1.23
q79	79.79	6.64
q89	609.61	0.58
q98	19.19	0.63
Total time	2210.83	30.83

Test 2
Data Load : 1TB



Hawq vs Spark (1TB Dataload)

(Excluded query 07, 89 & 99)



Hawq vs Spark (1TB Dataload)

Query execution time of Hawq & Spark in seconds, for 1TB dataload (Total time excludes query 99)

Query	Hawq	Spark
q03	25.25	0.47
q07	1613.16	0.77
q19	196.2	3.72
q27	116.12	0.80
q34	28.28	3.60
q42	128.13	0.59
q43	34.34	0.46
q46	71.71	3.63
q52	132.13	0.45
q53	46.46	0.62
q55	122.12	0.47
q59	189.19	1.04
q63	40.41	0.67
q65	88.89	2.85
q68	32.33	4.09
q79	141.14	2.73
q89	2179.22	0.59
q99	71.71	
Total exe time	23.21min	.4365min

Reference:

1. https://github.com/pivotalguru/TPC-DS
2. https://www.pivotalguru.com/?p=1045
3. https://github.com/databricks/spark-sql-perf
4. http://www.spark.tc/apache-spark-2-0-2-spark-sql/
5. http://www.tpc.org/tpcds/