

BSBM V3 Results (February 2011) Interpretation

Raw results available at <http://www4.wiwiw.fu-berlin.de/bizer/BerlinSPARQLBenchmark/results/V6/index.html#comparison>

Task	Weight	Best	4store		BigData		BigOWLIM		Jena TDB		Virtuoso	
			result	score	result	score	result	score	result	score	result	score
100M dataset												
Loading (minutes, lower is better)	15	17	27	9	63	4	17	15	75	3	109	2
Explore use case, single user	10	7,352	5,589	8	2,428	3	3,534	5	2,274	3	7,352	10
Explore and update use case	50	5,311	5,311	50		-10	2,809	26	680	6		-10
Multi-Client, Explore use case, AVG	30	26,490		-6	4,192	5	12,477	14	3,114	4	26,490	30
200M dataset												
Loading (minutes, lower is better)	15	39	72	8	204	3	39	15	165	4	240	2
Explore use case, single user	10	4,669	4,593	10	1,795	4	1,795	4	1,443	3	4,669	10
Multi-Client, Explore use case, AVG	30	16,031		-6	2,965	6	3,792	7	1,840	3	16,031	30
Total	160			73		14		86		27		75

Penalty for lack of functionality -20%

Weights summary

- 19% 30 Loading performance
- 13% 20 Single-user query performance
- 31% 50 Query and update performance
- 38% 60 Multi-user query performance

RAW RESULTS

Loading time (converted to minutes) Source: section 4

SUT	100M	200M
4store	27	72
BigData	63	204
BigOwlim	17	39
TDB	75	165
Virtuoso	109	240

Explore Use case, single user Source: section 6.1.1

	100m	200m
4store	5,589	4,593
BigData	2,428	1,795
BigOwlim	3,534	1,795
TDB	2,274	1,443
Virtuoso	7,352	4,669

Explore & Update Source: section 6.1.2

	100m
4store	5,311
BigOwlim	2,809
TDB	680

Multiple Clients (Explore query mix only) Source: section 6.2

Dataset Size 100M	Number of				AVERAGE
	1	4	8	64	
4store	5,589	*	*	*	
BigData	2,428	4,153	4,286	4,136	4,192
BigOwlim	3,534	9,349	12,798	15,285	12,477
TDB	2,274	4,065	3,035	2,242	3,114
Virtuoso	7,352	25,194	36,269	18,008	26,490

Dataset Size 200M	Number of				AVERAGE
	1	4	8	64	
4store	4,593	*	*	*	
BigData	1,795	3,040	3,167	2,689	2,965
BigOwlim	1,795	3,713	4,041	3,622	3,792
TDB	1,443	2,206	1,474	**	1,840
Virtuoso	4,669	13,265	18,264	16,564	16,031

Notes

1. For tasks where higher results are better, the repository with the best result gets the maximum score, which is also the relative weight of the task. The other repositories get scores, which represent a fraction of the maximal score that is proportional to the ratio between their results and the best one
2. For loading, where lower results are better, the scores are determined analogously, but using inverse proportion
3. There is a negative penalty score for tasks which were not completed by the respective repository. It is calculated as a percentage of the weight of this task
4. For the multi-client tests, the average results across the runs with different number of concurrent clients is used as a result value; the results for single client are not included in the average as those appear as separate task