

Earlier, we had mentioned some preliminary results for ResultSet compression using TBinary, TCompact and using a resultSet compressor externally as a plugin. In this report, we will use all the 3 above and also protobufs and snappy.

## Data

We will use the largest table from TPC-H, **lineitem**, for scale factors {2, 3, 4, 5, 6, 15, 30, 50, 75, 100}. For protobufs, we will encode tab-separated values from the table for that given scale factor. For the rest, we will use a modified Hive/HS2 as the server and Simba's leading commercial ODBC driver for Hive as client.

## Results

Scale factor	TBinary (MB)	TCompact (MB)	Protobuf (MB)	Snappy (MB)	Simba (MB)
2G	193	126	172	121	99
3G	289	193	263	183	152
4G	383	259	355	244	207
5G	480	325	446	305	262
6G	578	397	538	366	318
15G	1387	1000	1400	918	812
30G	2774	2079	2800	1832	1685
50G	4620	3588	4500	3061	2884
75G	-	-	-	-	-
100G	-	-	-	-	-