

Only average time have been recorded here

All times are in ms

Patch	Description	Is Write Impacted?	Is Read Impacted?
BASE:	master with SpecificInternalRow changes in https://github.com/apache/spark/pull/29366	N/A	N/A
X:	1) AvroSerializer: Use GenericData instead of ReflectData when writing avro data 2) AvroDeserializer: do not allocate extra primitive array for primitive arrays 3) Reuse objects when reading/writing avro	Yes	Yes
Y:	Native avro datum reader and writers	Yes	Yes

AvroReadBenchmark

Benchmark	Patch=>	BASE	X	Y
SQL Single TINYINT Column Scan:		3671	3276	2724
SQL Single SMALLINT Column Scan:		3806	3244	2931
SQL Single INT Column Scan:		3759	3064	2955
SQL Single BIGINT Column Scan:		4302	3715	3183
SQL Single FLOAT Column Scan:		3576	3080	2787
SQL Single DOUBLE Column Scan:		3905	3295	3086
Int and String Scan:		6252	5459	5326
Partitioned Table:	Data column	4369	4175	3444
	Partition column	3708	3451	2798
	Both columns	4382	4083	3481
Repeated String:		4571	4587	3819
String with Nulls Scan		7030	8079	6402
String with Nulls Scan (50.0%):		4759	5422	3820
String with Nulls Scan (95.0%):		3065	3900	2289
Single Column Scan from 100 columns:		6114	5944	4543
Single Column Scan from 200 columns:		12508	12000	8927
Single Column Scan from 300 columns:		18983	17755	13206
Nested Struct Scan:		50379	46283	44397
Array Scan:		5898	5800	5348
Array of Struct Scan:		19639	18886	18469
w/o filters		12826	12247	14286
pushdown disabled		12752	13093	15895
w/ filters		5662	5257	3462

AvroWriteBenchmark

<u>Benchmark</u>	<u>Patch=></u>	<u>BASE</u>	<u>X</u>	<u>Y</u>
Output Single Int Column		3750	2296	2096
Output Single Double Column		4099	2629	2270
Output Int and String Column		8903	6770	6104
Output Partitions		8426	6573	5951
Output Buckets		9262	7179	6611
Nested Struct		142209	81020	29134
Array		31617	18050	8192
Array of Struct		130682	78307	27832

AvroNestedSchemaPruningBenchmark

Benchmark	Patch=>	BASE	X	Y
Selection:	Top-level column	4122	3762	3545
	Nested column	4575	4131	4187
	Nested column in array	5389	4951	5015
Limiting:	Top-level column	4541	4124	3842
	Nested column	5180	4559	4713
	Nested column in array	6308	5767	6040
Repartitioning:	Top-level column	4650	4038	3689
	Nested column	5080	4471	4574
	Nested column in array	6241	5545	5813
Repartitioning by exprs:	Top-level column	4472	4040	3831
	Nested column	5183	4604	4698
	Nested column in array	6721	5996	6336
Sample:	Top-level column	4102	3683	3407
	Nested column	4681	4146	4153
	Nested column in array	5514	5019	5151
Sorting:	Top-level column	4981	4274	4101
	Nested column	5509	4896	4944
	Nested column in array	6898	6422	6723