

Schema Evolution related tests

Test conditions

Script used for these tests: [?contrib/dbutils/SchemaEvolTest.py](https://github.com/dbutills/SchemaEvolTest.py)

Test done with 2 tables: Object and ForcedSource, ~DR1 scale.

Object:

- planning 20K partitions
- in DR1: 0.6 million rows, ~440 columns, 1.5 GB data file compressed
- in DR11: 2 million rows, 6 GB data file compressed
- tested with 0.6 million rows, schema similar to baseline schema

ForcedSource:

- planning 20K partitions
- in DR1: 30 million rows, 0.6 GB data file compressed
- in DR11: 1.6 b rows, 35 GB data file compressed
- tested with 1 table (that represents one partition): 30 million rows, schema similar to baseline schema

Discussion

Adding new column for Object table is slow: 22 min per partition, 20K partitions, would require ~600 cores to do the update in parallel in 12 hours.

Tests

Adding new column

Adding new BIGINT column (8 bytes):

```
ALTER TABLE Object ADD COLUMN dummyBigInt BIGINT;  
ALTER TABLE ForcedSource ADD COLUMN dummyBigInt BIGINT;
```

Object: 1347.6 sec (22 min) on a busy server, 2 min 32 sec when server was less busy. ForcedSource: 38.7 sec (< 1 min)

Adding new INT column (4 bytes): Object too 1235.7 sec, ForcedSource 60.1 sec

Filling a column with data

```
UPDATE Object SET dummyBigInt = flux_psf+flux_psf_err;  
UPDATE ForcedSource set dummyBigInt = flux+flux_Sigma;
```

Object took 38 sec, ForcedSource 6 sec

Rebuild indexes

```
ALTER TABLE Object ENABLE KEYS;  
ALTER TABLE ForcedSource ENABLE KEYS;
```

Object too 105 sec, ForcedSorce < 1 sec.