LanguageManual Indexing

Indexing Is Removed since 3.0

There are alternate options which might work similarly to indexing:

- Materialized views with automatic rewriting can result in very similar results. Hive 2.3.0 adds support for materialzed views.
- Using columnar file formats (Parquet, ORC) – they can do selective scanning; they may even skip entire files/blocks.

Indexing has been **removed** in version 3.0 (HIVE-18448).

Overview of Hive Indexes

The goal of Hive indexing is to improve the speed of query lookup on certain columns of a table. Without an index, queries with predicates like 'WHERE tab1.col1 = 10' load the entire table or partition and process all the rows. But if an index exists for col1, then only a portion of the file needs to be loaded and processed.

The improvement in query speed that an index can provide comes at the cost of additional processing to create the index and disk space to store the index.

Versions

Hive indexing was added in version 0.7.0, and bitmap indexing was added in version 0.8.0.

Indexing Resources

Documentation and examples of how to use Hive indexes can be found here:

- Indexes – design document (lists indexing JIRAs with current status, starting with HIVE-417)
- Create/Drop/Alter Index – HiveQL Language Manual DDL
- Show Indexes – HiveQL Language Manual DDL
- Bitmap indexes – added in Hive version 0.8.0 (HIVE-1803)
- Indexed Hive – overview and examples by Prafulla Tekawade and Nikhil Deshpande, October 2010
- Tutorial: SQL-like join and index with MapReduce using Hadoop and Hive – blog by Ashish Garg, April 2012

Configuration Parameters for Hive Indexes

The Configuration Properties document describes parameters that configure Hive indexes.

Simple Examples

This section gives some indexing examples adapted from the Hive test suite.

Case sensitivity

In Hive 0.12.0 and earlier releases, the index name is case-sensitive for CREATE INDEX and DROP INDEX statements. However, ALTER INDEX requires an index name that was created with lowercase letters (see HIVE-2752). This bug is fixed in Hive 0.13.0 by making index names case-insensitive for all HiveQL statements. For releases prior to 0.13.0, the best practice is to use lowercase letters for all index names.

Create/build, show, and drop index:

```
CREATE INDEX table01_index ON TABLE table01 (column2) AS 'COMPACT';
SHOW INDEX ON table01;
DROP INDEX table01_index ON table01;
```

Create then build, show formatted (with column names), and drop index:

```
CREATE INDEX table02_index ON TABLE table02 (column3) AS 'COMPACT' WITH DEFERRED REBUILD;
ALTER INDEX table02_index ON table2 REBUILD;
SHOW FORMATTED INDEX ON table02;
DROP INDEX table02_index ON table02;
```

Create bitmap index, build, show, and drop:

```
CREATE INDEX table03_index ON TABLE table03 (column4) AS 'BITMAP' WITH DEFERRED REBUILD;
ALTER INDEX table03_index ON table03 REBUILD;
SHOW FORMATTED INDEX ON table03;
DROP INDEX table03_index ON table03;
```

Create index in a new table:

```
CREATE INDEX table04_index ON TABLE table04 (column5) AS 'COMPACT' WITH DEFERRED REBUILD IN TABLE
table04_index_table;
```

Create index stored as RCFile:

```
CREATE INDEX table05_index ON TABLE table05 (column6) AS 'COMPACT' STORED AS RCFILE;
```

Create index stored as text file:

```
CREATE INDEX table06_index ON TABLE table06 (column7) AS 'COMPACT' ROW FORMAT DELIMITED FIELDS TERMINATED BY
'\t' STORED AS TEXTFILE;
```

Create index with index properties:

```
CREATE INDEX table07_index ON TABLE table07 (column8) AS 'COMPACT' IDXPROPERTIES ("prop1"="value1", "prop2"="
value2");
```

Create index with table properties:

```
CREATE INDEX table08_index ON TABLE table08 (column9) AS 'COMPACT' TBLPROPERTIES ("prop3"="value3", "prop4"="
value4");
```

Drop index if exists:

```
DROP INDEX IF EXISTS table09_index ON table09;
```

Rebuild index on a partition:

```
ALTER INDEX table10_index ON table10 PARTITION (columnX='valueQ', columnY='valueR') REBUILD;
```