

Hive9 索引

创建索引

- 先简历一张表

```
CREATE TABLE employees(  
    name          STRING,  
    salary        FLOAT,  
    subordinates  ARRAY<STRING>,  
    deductions    MAP<STRING, FLOAT>,  
    adress        STRUCT<street:STRING, city:STRING, state:STRING, zip:INT>  
)  
PARTITION BY (country STRING, state STRING);
```

- 对分区建立索引

```
CREATE INDEX employees_index  
ON TABLE employees(country)  
AS 'org.apache.hadoop.hive.ql.index.compact.CompactIndexHandler'  
WITH DEFERRED REBUILD  
IDXPROPERTIES ('creator' = 'me', 'ceated_at' = 'some_time')  
IN TABLE employees_index_table  
PARTITIONED BY (country, name)  
COMMENT 'Employees indexed by country and name.';
```

AS: 语句制定了索引处理器，也就是实现了索引接口的JAVA类，Hive本身包含了一些典型的索引实现。

IN TABLE: 要求索引处理器在一张新表中保留索引数据。

Bitmap索引

bitmap索引普遍应用于排重后较少的列。

```
CREATE INDEX employees_index  
ON TABLE employees(country)  
AS 'BITMAP'  
WITH DEFERRED REBUILD  
IDXPROPERTIED('creator' = 'me', 'ceated_at' = 'some_time')  
IN TABLE employees_index_table  
PARTITIONED BY (country, name)  
COMMENT 'Employees indexes by country and name.';
```

重建索引

如果用户指定了 DEFERRED REBUILD，那么新索引将呈现空白状态，在任何时候，都可以进行第一次索引创建或者使用 ALTER INDEX对索引进行重建。

```
ALTER INDEX employees_index  
ON TABLE employees  
PARTITION (country = 'US')  
REBUILD;
```

如果忽略掉PARTITION，那么将会对所有分区进行重建索引。

显示索引

```
SHOW FORMATTED INDEX ON employees;
```

关键字FORMATTED是可选的，增加这个关键字可以使输出中包含有列名称。用户还可以替换INDEX 为INDEXES。

删除索引

```
DROP INDEX IF EXISTS employees_index ON TABLE employees;
```