Hive实战

徐冬
frankxus@gmail.com

Agenda

- 简介
- 部署/配置
- Hive QL 编程
- 查询优化
- Hive QL vs. SQL

简介

什么是Hive？

建立在 Hadoop 上的数据仓库基础构架

什么是Hive？

- Hadoop
  - Open Source MapReduce framework
- Hive
  - 支持SQL语义的大规模数据分析工具
  - 离线/数据仓库应用
  - File Processor

为什么选择Hive？

- 为超大数据集设计的计算/扩展能力
  - based on Hadoop
- 支持SQL like查询语言
- 统一的元数据管理

为什么选择Hive？

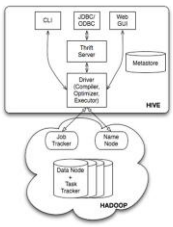
- 简单

```
select word, count(*)
from (
select
explode(split(sentence. ' ')) word
from article
) t
group by word
```

What Hive can/cannot do？

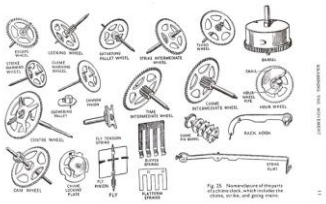
- Cans
  - 大规模数据处理（依赖Hadoop）
  - 支持大部分SQL语义（select/join/group by …）
- Can’ts
  - 在线应用/事务 （OLTP）
  - update
  - …

Architect



- Client端应用程序
- 元数据
- 编程接口

配置/部署



- 客户端
- 元数据服务器

客户端部署

- 依赖
  - Hadoop Client
- 安装

配置

- 环境变量
  - HIVE_HOME
  - HIVE_CONF_DIR
  - HIVE_AUX_JARS_PATH
  - HADOOP_HOME/HADOOP_CONF_DIR
- 配置文件
  - hive-default.xml/hive-site.xml
  - hive-log4j.properties

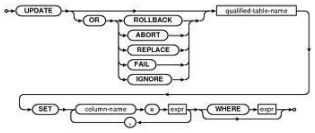
HDFS路径配置

- hive.metastore.warehouse.dir
- hive.exec.scratchdir

元数据服务器

- 配置
  - javax.jdo.option.ConnectionURL
  - javax.jdo.option.ConnectionUserName
  - javax.jdo.option.ConnectionPassword
- 选择一个元数据服务器
  - Derby
  - MySQL
  - javax.jdo.option.ConnectionDriverName
- JDO配置
  - datanucleus.fixedDatastore
  - datanucleus.autoCreateTables
  - datanucleus.autoCreateSchema

Hive QL 编程



SQL like, but simpler

数据模型

- Database
- Table
- Partition
- File

数据类型

- Primitive
  - int / bigint / smallint / tinyint
  - boolean
  - double / float
  - string
- Array
- Map
- Struct
- 没有精度/长度设定
- Java style

文件类型

- TextFile
- SequenceFile
- RCFile
- 自定义类型
  - 自定义InputFormat/OutputFormat
  - i.e. XMLInputFormat

DDL

CREATE [EXTERNAL] TABLE [IF NOT EXISTS] table_name
(col_name data_type, ...)
[PARTITIONED BY (col_name data_type, ...)]
[[ROW FORMAT row_format] [STORED AS file_format] | [WITH SERDEPROPERTIES (...)]]
[LOCATION hdfs_path]

DDL

- CTAS
  - CREATE [EXTERNAL] TABLE [IF NOT EXISTS] table_name
    (col_name data_type, ...)
    …
    AS SELECT …

DML/加载数据

- Load data
  - LOAD DATA [LOCAL] INPATH 'filepath' [OVERWRITE] INTO TABLE tablename [PARTITION (partcol1=val1, partcol2=val2 ...)]

DML/加载数据

- Insert
  - INSERT OVERWRITE TABLE tablename [PARTITION (partcol1=val1, partcol2=val2 ...)] select_statement FROM from_statement
- Multiple insert
  - FROM from_statement
    INSERT OVERWRITE TABLE tablename1 [PARTITION...)] select_statement1
    [INSERT OVERWRITE TABLE tablename2 [PARTITION ...] select_statement2] ...
- Dynamic partitioning
  - INSERT OVERWRITE TABLE tablename PARTITION (partcol1[=val1], partcol2[=val2] ...) select_statement FROM from_statement
- (HDFS)不支持UPDATE !

DML/DDL加载数据

- Add partition (常用)
  - ALTER TABLE table_name ADD PARTITION (partcol1=val1, partcol2=val2 ...) [LOCATION 'filepath']

Query

- select
  - SELECT [ALL | DISTINCT] select_expr, select_expr, ...
    FROM table_reference
    [WHERE where_condition]
    [GROUP BY col_list]
    [CLUSTER BY col_list | [DISTRIBUTE BY col_list] [SORT BY col_list] | [ORDER BY col_list]]
    [LIMIT number]

Query

- Join
  - join_table: table_reference JOIN table_factor [join_condition] | table_reference [{LEFT|RIGHT|FULL} OUTER | LEFT SEMI] JOIN table_reference join_condition
  - table_reference: table_factor | join_table
  - table_factor: tbl_name [alias] | table_subquery alias | (table_references)
  - join_condition: ON equality_expression (AND equality_expression)* equality_expression: expression = expression
- 等值Join
- 合并Join的原则

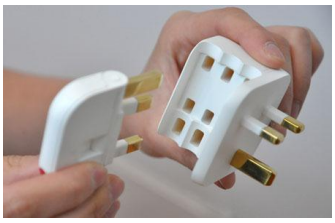
Query

- Subqueries
  - SELECT ... FROM (subquery) name ...
  - 不支持exist in子查询
  - select_statement UNION ALL select_statement UNION ALL select_statement ...

View

- CREATE VIEW [IF NOT EXISTS] view_name [(column_name...)] [TBLPROPERTIES (property_name = property_value, ...)] AS SELECT ...

自定义扩展



- UDF
- Transform
- SerDe

UDF

- Java语言编写的用户自定义函数
- 类别
  - UDF – 1:1
  - UDAF – N:1
  - UDTF – 1:N
- 使用
  - CREATE TEMPORARY FUNCTION function_name AS class_name

UDF

- Implement UDF
  - extends UDF / GenericUDF
  - implement evaluate() function
- Implement UDAF
  - extends UDAF / GenericUDAF
  - implement
    - iterate
    - merge
    - terminatePartial
    - terminate

UDF

- Implement UDTF
  - extends GenericUDTF
  - Implement process()
- UDTF的限制
  - 不支持UDTF/列混合的select
  - 不支持嵌套
  - 不支持相同子查询中的GROUP BY / CLUSTER BY / DISTRIBUTE BY / SORT BY
- 使用Lateral view

Transform

- Hive QL中直接支持MapReduce

```
FROM (
  FROM src
  MAP expression (',' expression)*
  USING 'my_map_script'
  ( AS colName (',' colName)* )?
  ( clusterBy? | distributeBy? sortBy? ) src_alias
)
REDUCE
  expression (',' expression)*
  USING 'my_reduce_script'
  ( AS colName (',' colName)* )?
```

UDxF vs. Transform

	UDF/UDAF	M/R scripts
language	Java	any language
data format	in-memory objects	serialized streams
1/1 input/output	supported via UDF	supported
n/1 input/output	supported via UDAF	supported
1/n input/output	supported via UDTF	supported
Speed	faster	Slower

- 优先选择UDxF
- 其他灵活性
  - n/n?

SerDe

- SerDe处理的是？
  - 文件格式
  - 字段序列化/反序列化格式
- Default: delimited
- RegexSerDe

查询优化

- 优化器
- 更好的执行计划
- 执行优化

优化器

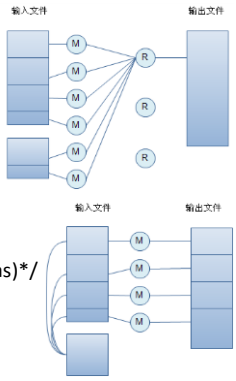
- Rule based optimizer (Semantic query optimize)
  - Partition Pruning (ppr)
  - Predicate Push down (ppd)
  - Column Pruning (cp)
  - Union Transformer
- Physical optimizer
  - Map join Transformer
  - Skew join

更好的执行计划

- 数据倾斜
  - 什么是数据倾斜？
  - 倾斜的原因？
    - group by/distinct
    - join

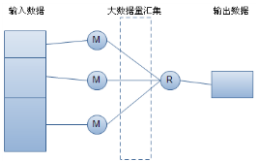
更好的执行计划

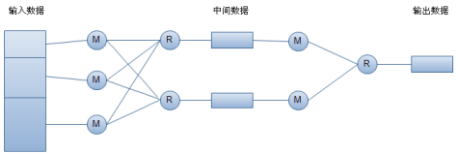
- Join数据倾斜
  - Map Join
  - 限制
    - 内存
    - 语义
  - 代价
  - 用法
    - select /*+ MAPJOIN(tb_alias)*/
  - Bucketed map join
  - Sort merge bucketed map join



更好的执行计划

- GroupBy数据倾斜
  - skewindata优化
  - 用法
    - set hive.groupby.skewindata=true



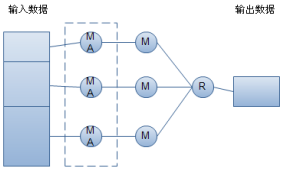


执行优化

- 内存优化
  - 驱动表
    - 使用大表做驱动表，避免内存溢出
    - 默认Join中最右边的表是驱动表
    - MapJoin无视Join顺序，使用大表做驱动表
    - STREAMTABLE hint

执行优化

- I/O优化
  - Map aggregation
  - 相关参数
    - hive.map.aggr
    - hive.groupby.mapaggr.checkinterval (100000)
    - hive.map.aggr.hash.percentmemory(0.5)
    - hive.map.aggr.hash.min.reduction(0.5)
  - 注意控制内存



执行优化

- I/O优化
  - 合并小文件
    - 减少后续任务的map数
    - 代价：额外的MR过程
    - 参数：
      - hive.merge.mapfiles
      - hive.merge.mapredfiles
      - hive.merge.size.per.task
      - hive.merge.size.smallfiles.avgsize

执行优化

- MR任务合并
  - multi-insert

```
FROM from_statement
INSERT OVERWRITE TABLE tablename1 [PARTITION (partcol1=val1,
    partcol2=val2 ...)] select_statement1
[INSERT OVERWRITE TABLE tablename2 [PARTITION ...]
    select_statement2] ...
```

select statement中的过滤条件不能做分区裁剪

执行优化

- MR任务合并
  - union

```
select col_list from tbl where …
union all
select col_list from tbl where …
union all
select col_list from tbl where …
```

执行优化

- MR任务合并
  - Joins
    - 相同Join key的Join可被优化为1次MR过程
    - 注意写法
  - Map Joins
    - 多次查表操作可以用一个Map Join做完
    - 不需要是相同Join key

Hive QL vs. SQL

- 关系/约束
  - 主键约束
  - 外键约束
- 数据类型

Hive QL vs. SQL

- 不支持的语义
  - update/delete
  - 嵌套子查询
    - exist in子查询
    - 只支持等值Join
  - having
  - 有限的order by
- 不支持的特性
  - Index (hash index of Hive 0.7.0?)
  - 窗口函数