大数据Hadoop高薪直通车课程

Hive 高级进阶

讲师：轩宇（北风网版权所有)

课程大纲

1 HiveServer2

2 Hive 数据压缩

3 Hive 数据存储

4 Hive 企业优化

5 Hive 实战案例

课程大纲

1 **HiveServer2**

2 **Hive 数据压缩**

3 **Hive 数据存储**

4 **Hive 企业优化**

5 **Hive 实战案例**

HiveServer2

- hiveserver2

1) hiveserver2

2) hiveserver2 --hiveconf hive.server2.thrift.port=14000

- beeline (beeline --help查看参数信息)

1) beeline -u jdbc:hive2://hadoop001:10000/default -n hadoop

2) beeline -u jdbc:hive2://hadoop001:**14000**/default -n hadoop

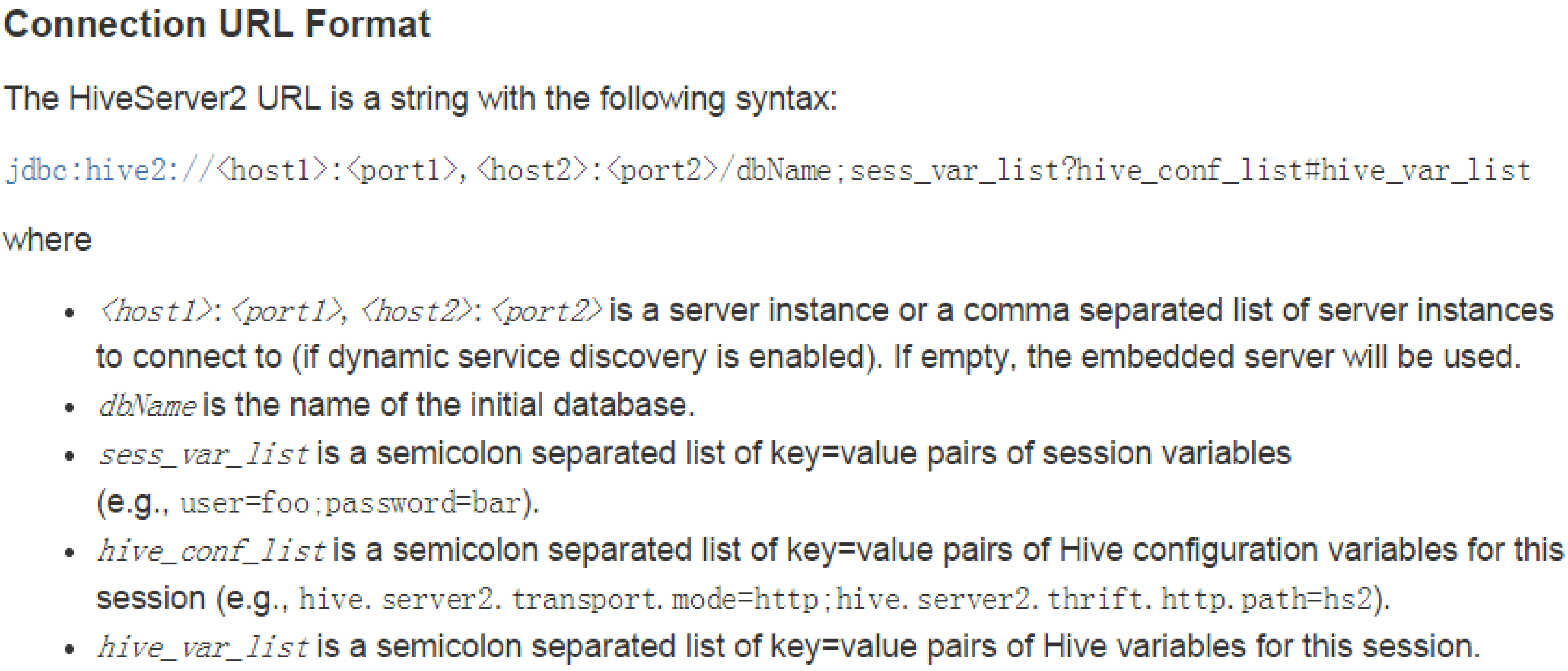
HiveServer2

- 使用Java代码连接Hive

使用JDBC连接之前一定要先启动hiveserver2

https://cwiki.apache.org/confluence/display/Hive/HiveServer2+Clients

Using JDBC



Connection URL Format

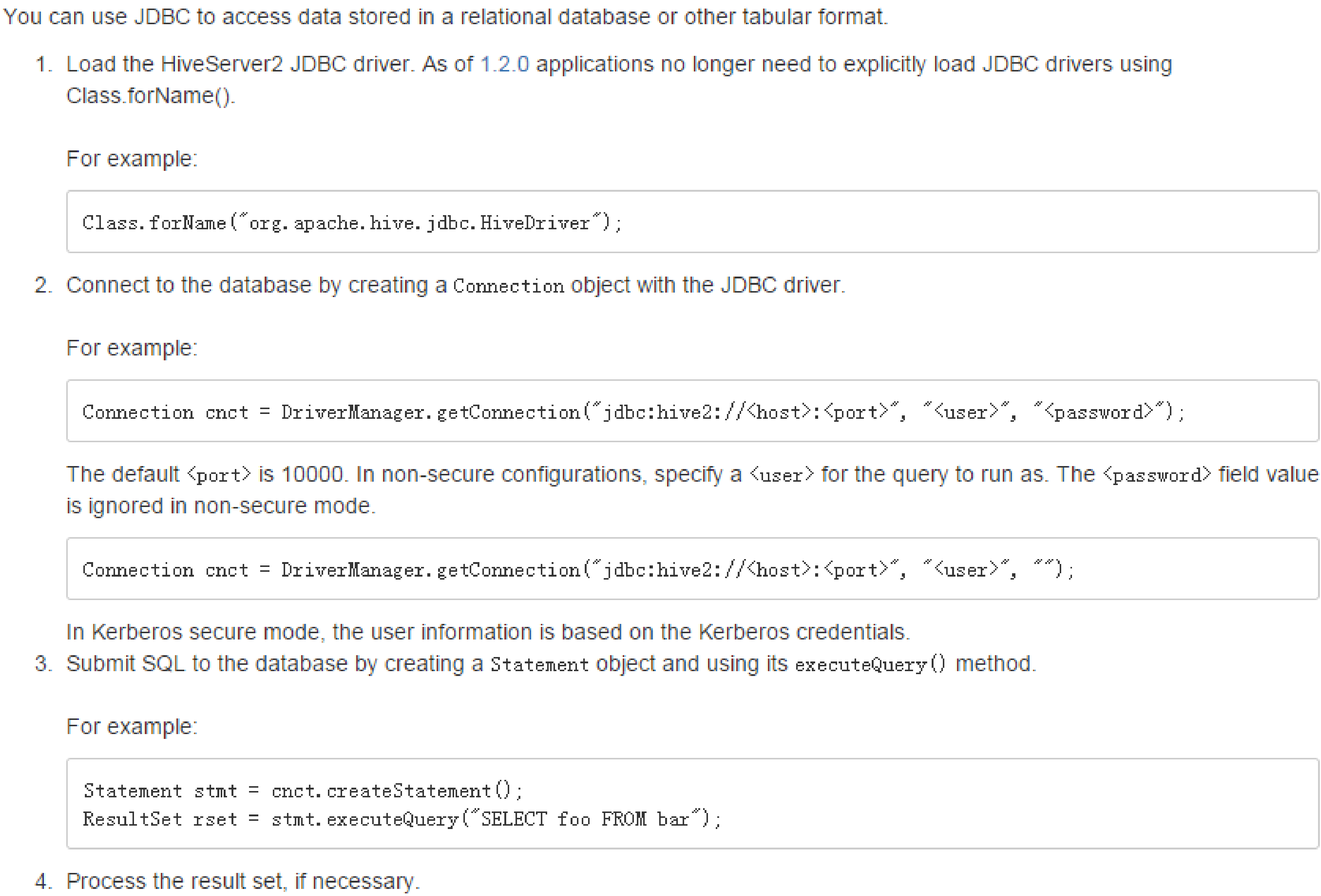
The HiveServer2 URL is a string with the following syntax:

```
jdbc:hive2://〈host1〉:〈port1〉,〈host2〉:〈port2〉/dbName;sess_var_list?hive_conf_list#hive_var_list
```

where

- *〈host1〉:〈port1〉,〈host2〉:〈port2〉* is a server instance or a comma separated list of server instances to connect to (if dynamic service discovery is enabled). If empty, the embedded server will be used.
- *dbName* is the name of the initial database.
- *sess_var_list* is a semicolon separated list of key=value pairs of session variables (e.g., `user=foo;password=bar`).
- *hive_conf_list* is a semicolon separated list of key=value pairs of Hive configuration variables for this session (e.g., `hive.server2.transport.mode=http;hive.server2.thrift.http.path=hs2`).
- *hive_var_list* is a semicolon separated list of key=value pairs of Hive variables for this session.

Using JDBC



You can use JDBC to access data stored in a relational database or other tabular format.

1. Load the HiveServer2 JDBC driver. As of 1.2.0 applications no longer need to explicitly load JDBC drivers using Class.forName().

   For example:

   ```
   Class.forName("org.apache.hive.jdbc.HiveDriver");
   ```

2. Connect to the database by creating a `Connection` object with the JDBC driver.

   For example:

   ```
   Connection cnct = DriverManager.getConnection("jdbc:hive2://<host>:<port>", "<user>", "<password>");
   ```

   The default `<port>` is 10000. In non-secure configurations, specify a `<user>` for the query to run as. The `<password>` field value is ignored in non-secure mode.

   ```
   Connection cnct = DriverManager.getConnection("jdbc:hive2://<host>:<port>", "<user>", "");
   ```

   In Kerberos secure mode, the user information is based on the Kerberos credentials.
3. Submit SQL to the database by creating a `Statement` object and using its `executeQuery()` method.

   For example:

   ```
   Statement stmt = cnct.createStatement();
   ResultSet rset = stmt.executeQuery("SELECT foo FROM bar");
   ```

4. Process the result set, if necessary.

课程大纲

1 HiveServer2

2 Hive 数据压缩

3 Hive 数据存储

4 Hive 企业优化

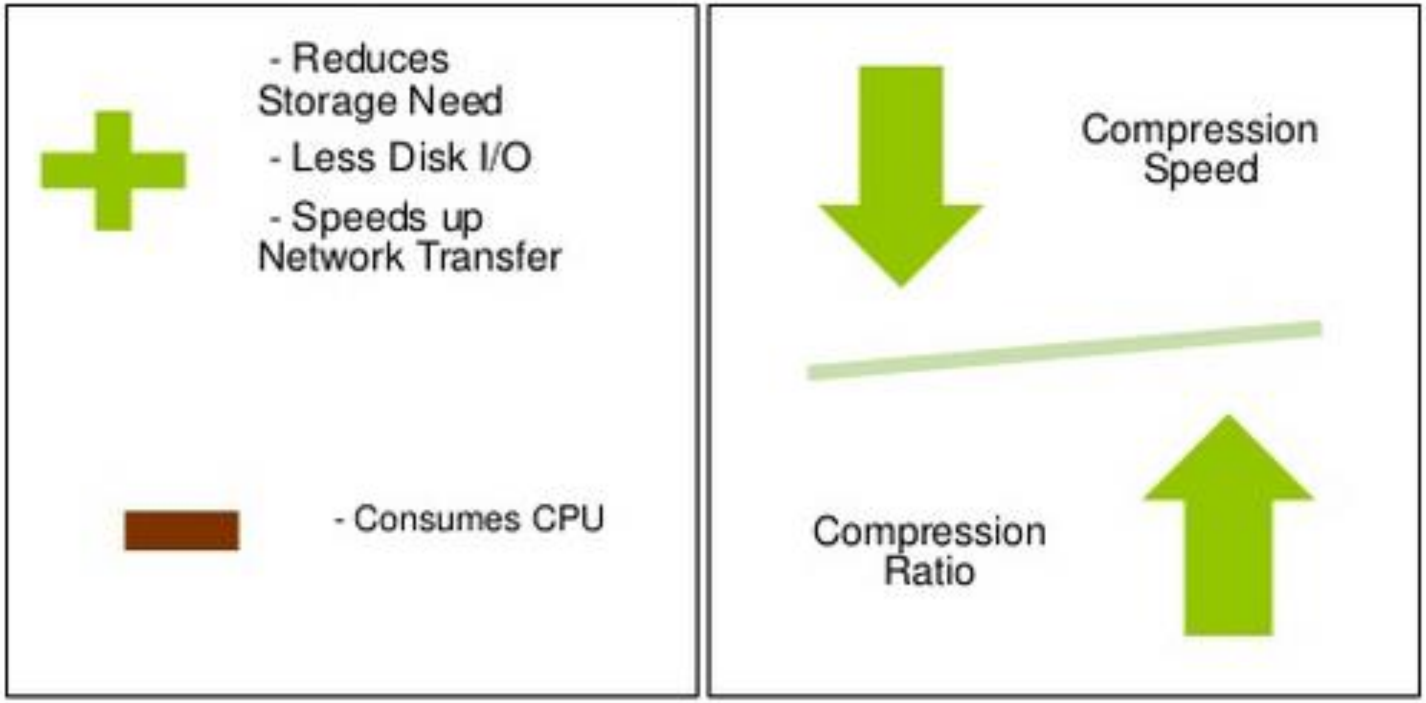
5 Hive 实战案例

常见数据压缩技术

压缩格式: bzip2, gzip, lzo, **snappy**等

压缩比：bzip2>gzip>lzo   bzip2最节省存储空间

解压速度：lzo>gzip>bzip2   lzo解压速度是最快的



Data Compression in Hadoop

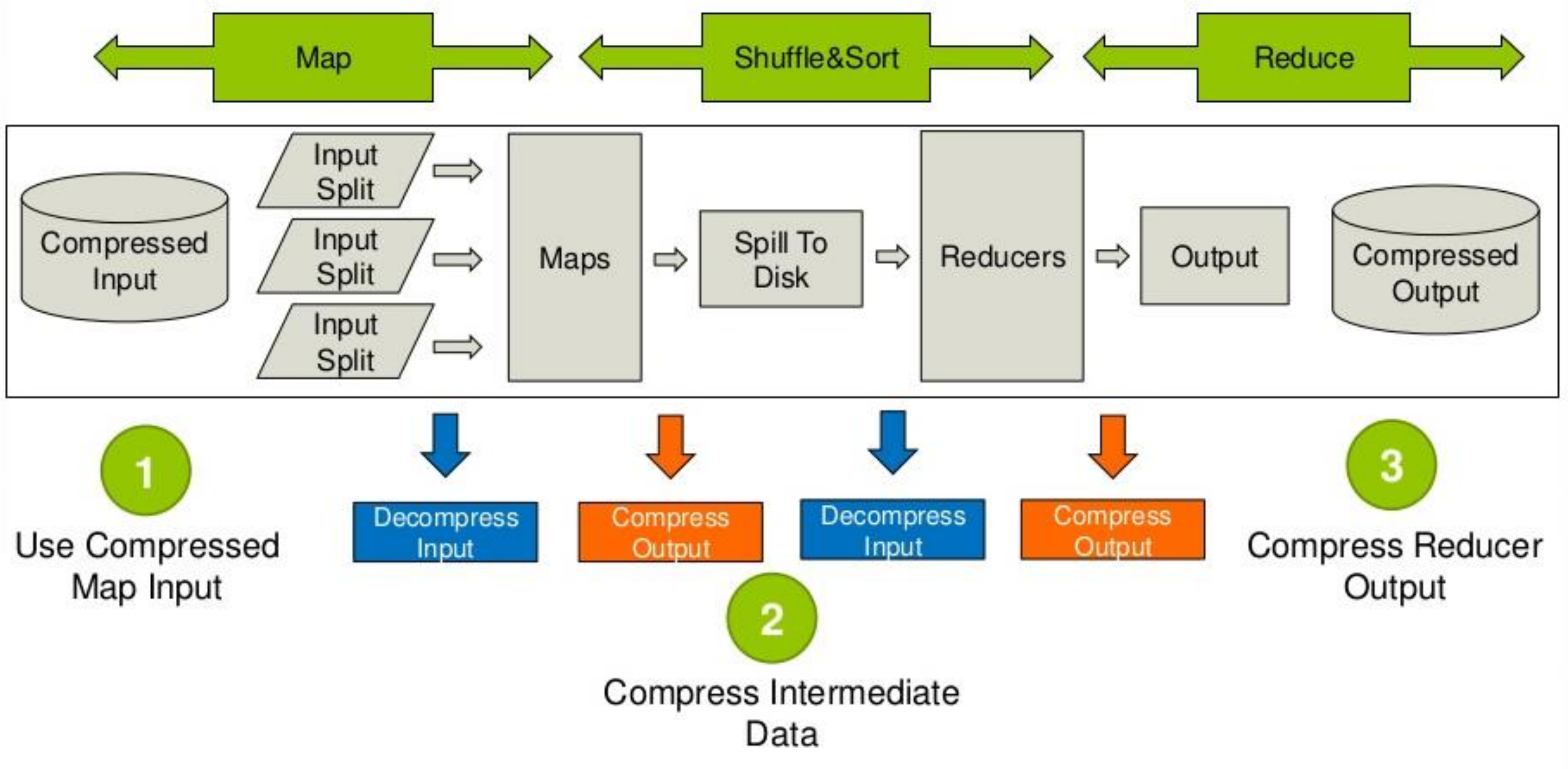
✓Hadoop jobs are usually IO bound;

✓Compression reduces the size of data transferred across network;

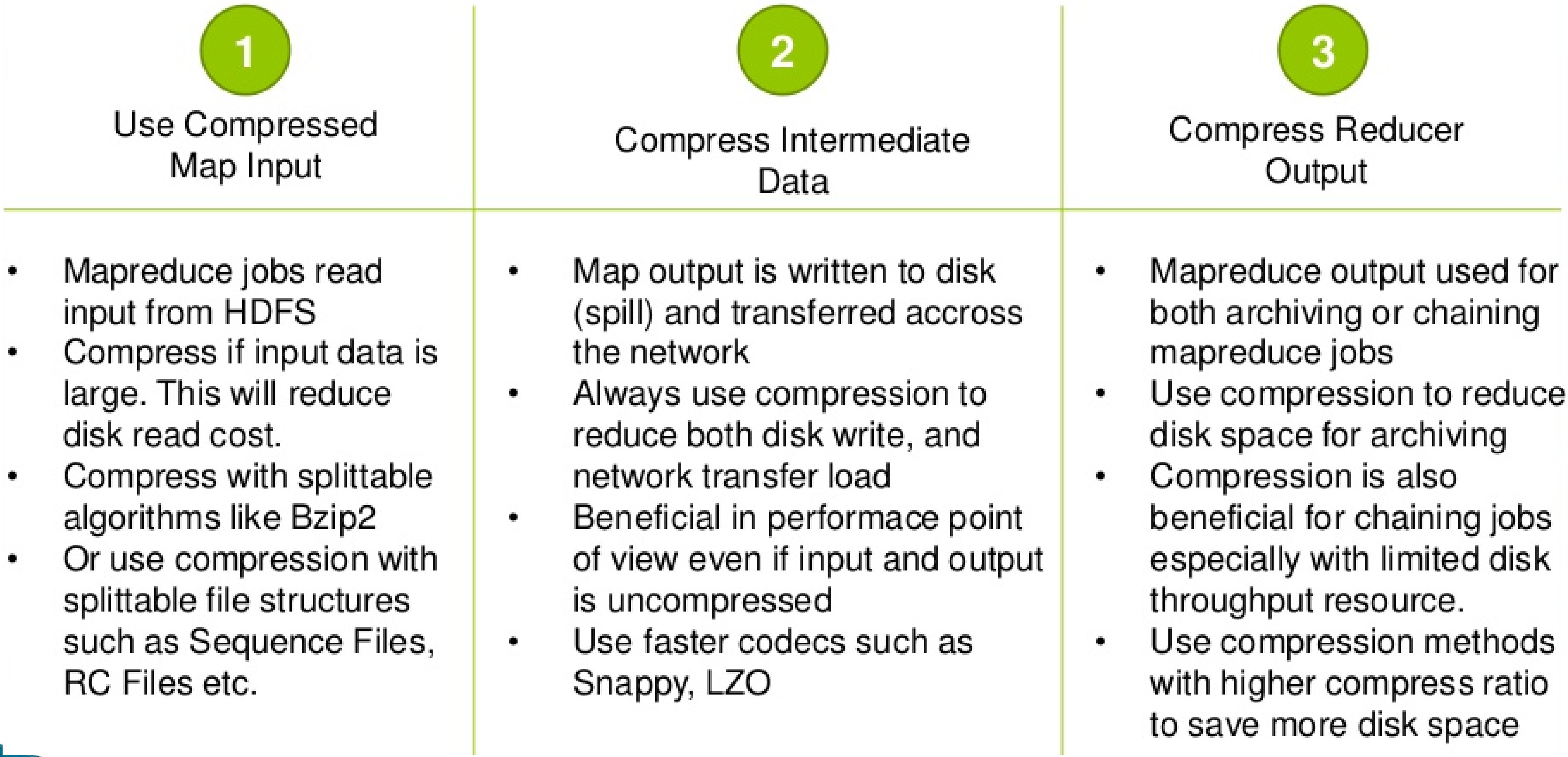
✓Overall job performance may be increased by simple enabing compression;

✓Splittability must be taken into account;

Compression in MR Pipeline

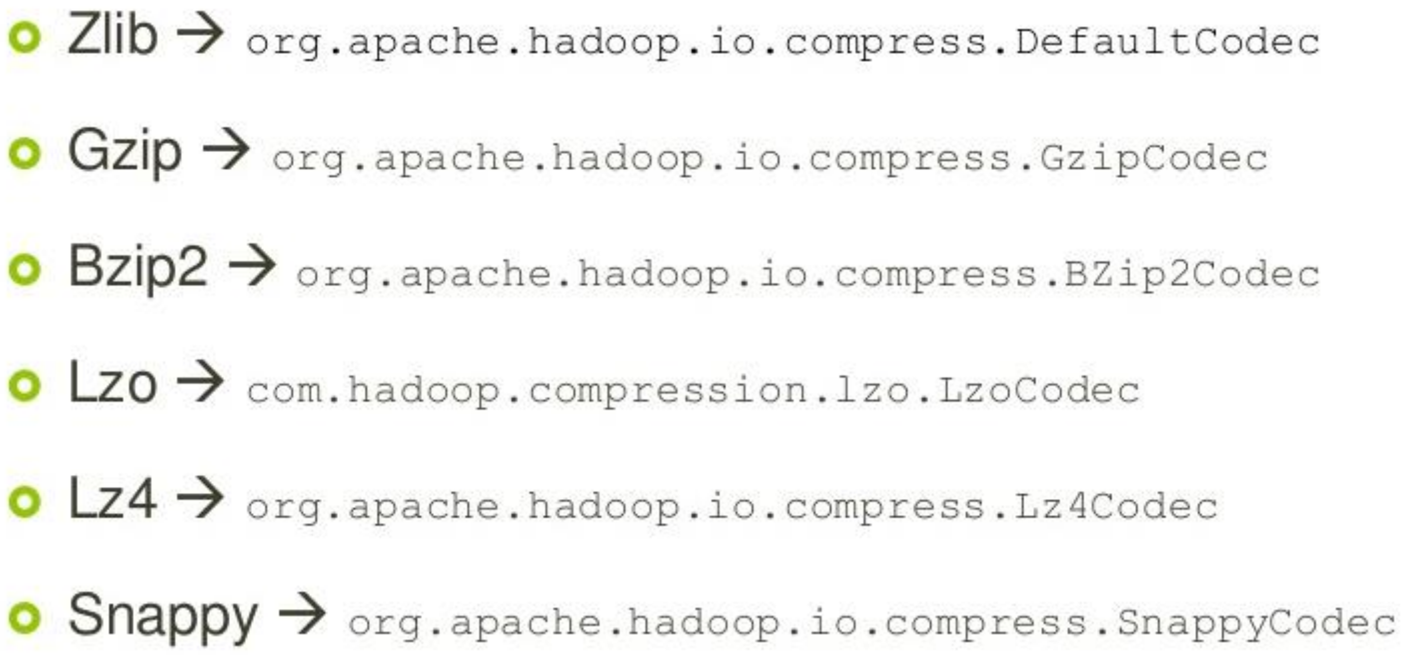


When to Compress?



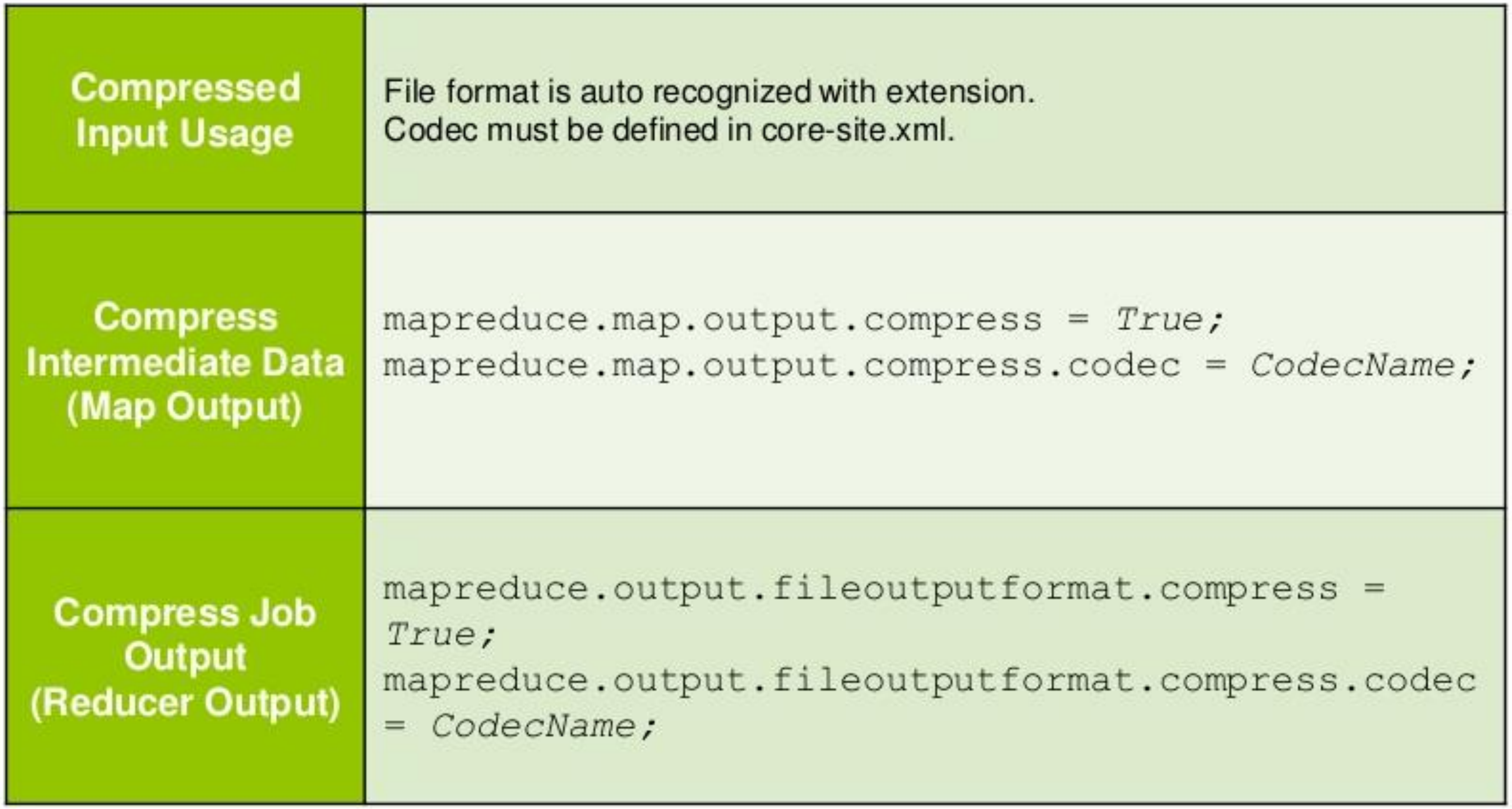
1 Use Compressed Map Input	2 Compress Intermediate Data	3 Compress Reducer Output
• Mapreduce jobs read input from HDFS • Compress if input data is large. This will reduce disk read cost. • Compress with splittable algorithms like Bzip2 • Or use compression with splittable file structures such as Sequence Files, RC Files etc.	• Map output is written to disk (spill) and transferred accross the network • Always use compression to reduce both disk write, and network transfer load • Beneficial in performace point of view even if input and output is uncompressed • Use faster codecs such as Snappy, LZO	• Mapreduce output used for both archiving or chaining mapreduce jobs • Use compression to reduce disk space for archiving • Compression is also beneficial for chaining jobs especially with limited disk throughput resource. • Use compression methods with higher compress ratio to save more disk space

Supported Codecs in Hadoop



- Zlib → `org.apache.hadoop.io.compress.DefaultCodec`
- Gzip → `org.apache.hadoop.io.compress.GzipCodec`
- Bzip2 → `org.apache.hadoop.io.compress.BZip2Codec`
- Lzo → `com.hadoop.compression.lzo.LzoCodec`
- Lz4 → `org.apache.hadoop.io.compress.Lz4Codec`
- Snappy → `org.apache.hadoop.io.compress.SnappyCodec`

Compression in MapReduce

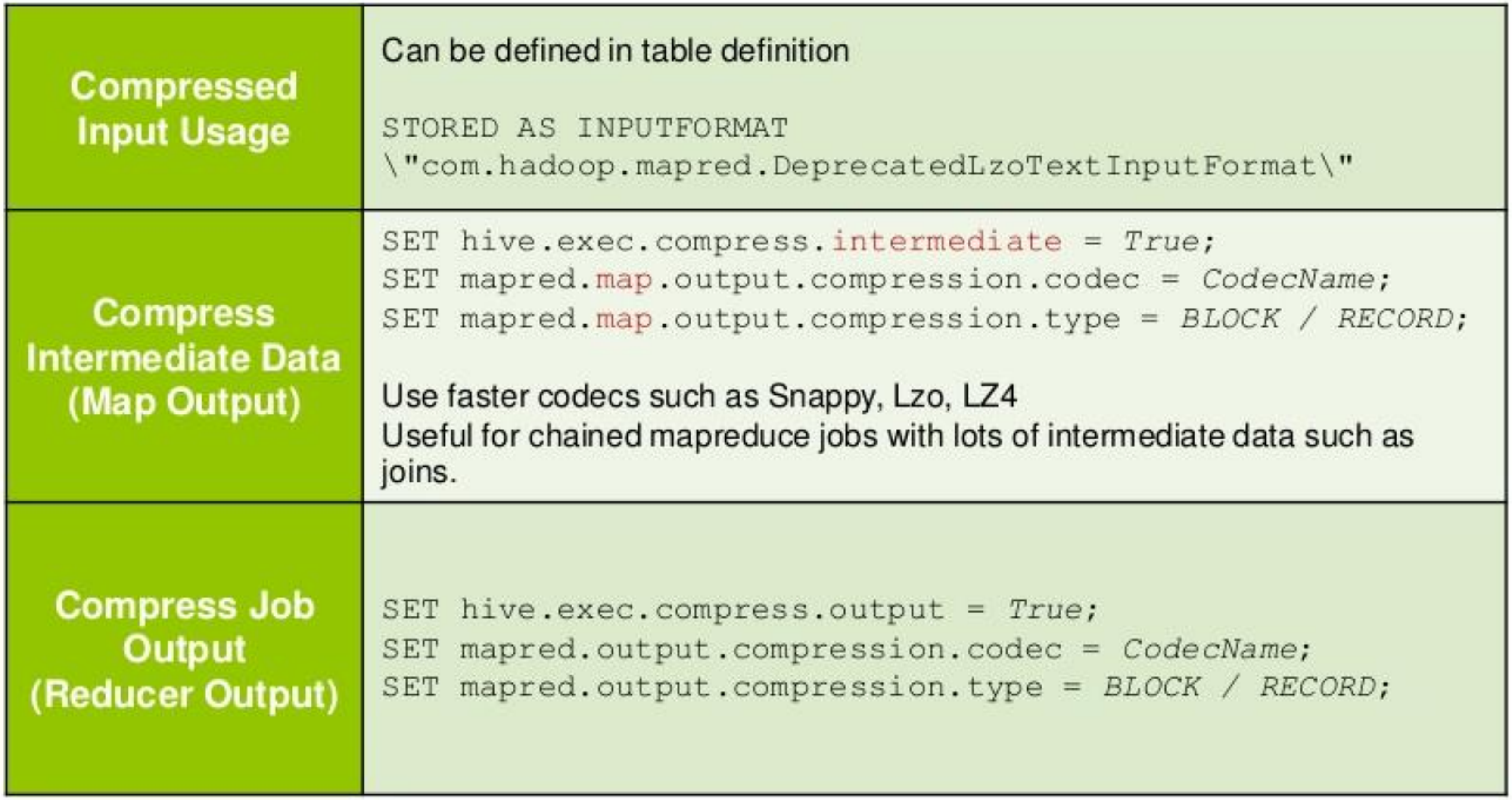


Compressed Input Usage	File format is auto recognized with extension. Codec must be defined in core-site.xml.

Compress Intermediate Data (Map Output)	mapreduce.map.output.compress = *True;*
	mapreduce.map.output.compress.codec = *CodecName;*

Compress Job Output (Reducer Output)	mapreduce.output.fileoutputformat.compress = *True;*
	mapreduce.output.fileoutputformat.compress.codec = *CodecName;*

Compression in Hive



Compressed Input Usage	Can be defined in table definition
	`STORED AS INPUTFORMAT`
	`\"com.hadoop.mapred.DeprecatedLzoTextInputFormat\"`

Compress Intermediate Data (Map Output)	`SET hive.exec.compress.intermediate = True;`
	`SET mapred.map.output.compression.codec = CodecName;`
	`SET mapred.map.output.compression.type = BLOCK / RECORD;`
	Use faster codecs such as Snappy, Lzo, LZ4
	Useful for chained mapreduce jobs with lots of intermediate data such as joins.

Compress Job Output (Reducer Output)	`SET hive.exec.compress.output = True;`
	`SET mapred.output.compression.codec = CodecName;`
	`SET mapred.output.compression.type = BLOCK / RECORD;`

课程大纲

1 HiveServer2

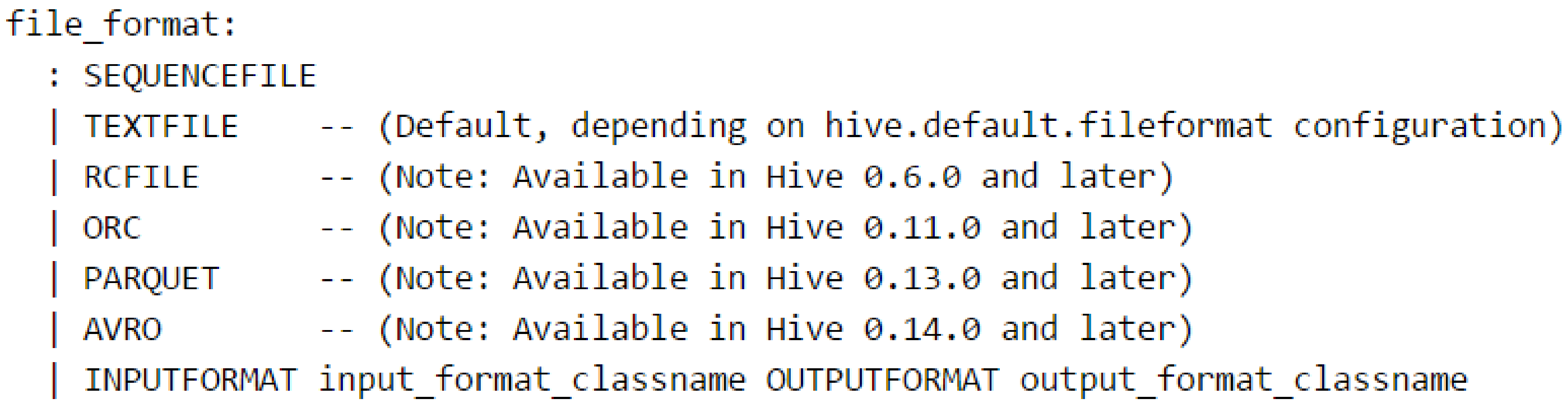
2 Hive 数据压缩

3 Hive 数据存储

4 Hive 企业优化

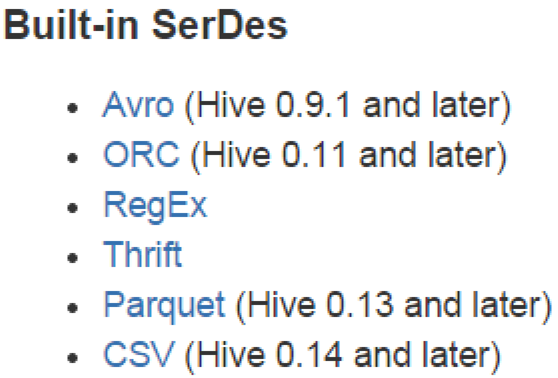
5 Hive 实战案例

数据文件存储格式



```
file_format:
  : SEQUENCEFILE
  | TEXTFILE    -- (Default, depending on hive.default.fileformat configuration)
  | RCFILE      -- (Note: Available in Hive 0.6.0 and later)
  | ORC         -- (Note: Available in Hive 0.11.0 and later)
  | PARQUET     -- (Note: Available in Hive 0.13.0 and later)
  | AVRO        -- (Note: Available in Hive 0.14.0 and later)
  | INPUTFORMAT input_format_classname OUTPUTFORMAT output_format_classname
```

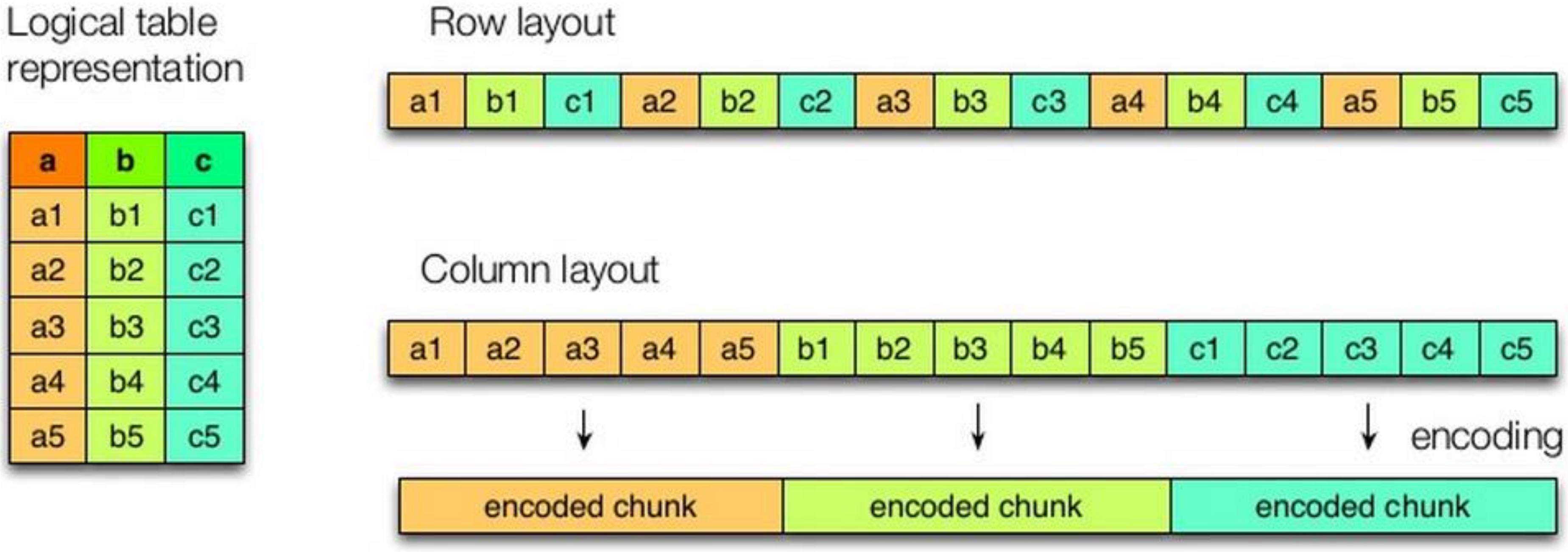
https://cwiki.apache.org/confluence/display/Hive/SerDe



Built-in SerDes

- Avro (Hive 0.9.1 and later)
- ORC (Hive 0.11 and later)
- RegEx
- Thrift
- Parquet (Hive 0.13 and later)
- CSV (Hive 0.14 and later)

Columnar vs Row-based



Logical table representation

a	b	c
a1	b1	c1
a2	b2	c2
a3	b3	c3
a4	b4	c4
a5	b5	c5

Row layout

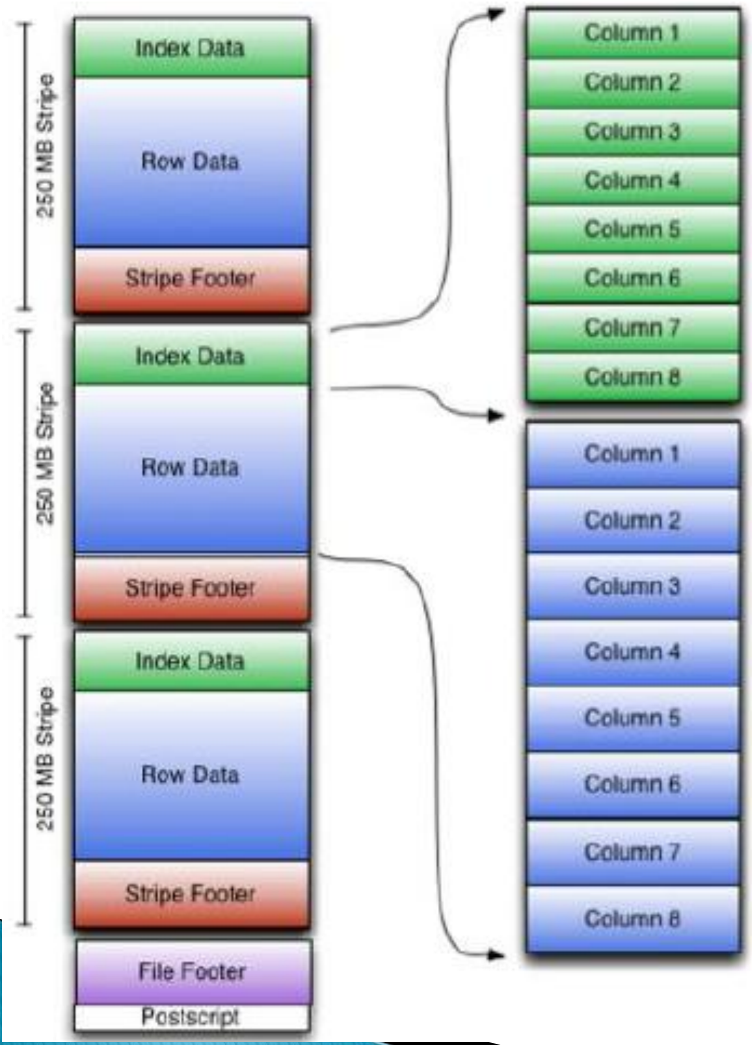
a1	b1	c1	a2	b2	c2	a3	b3	c3	a4	b4	c4	a5	b5	c5

Column layout

a1	a2	a3	a4	a5	b1	b2	b3	b4	b5	c1	c2	c3	c4	c5

ORCFile

ORCFile（Optimized Row Columnar File）：hive/shark/spark支持



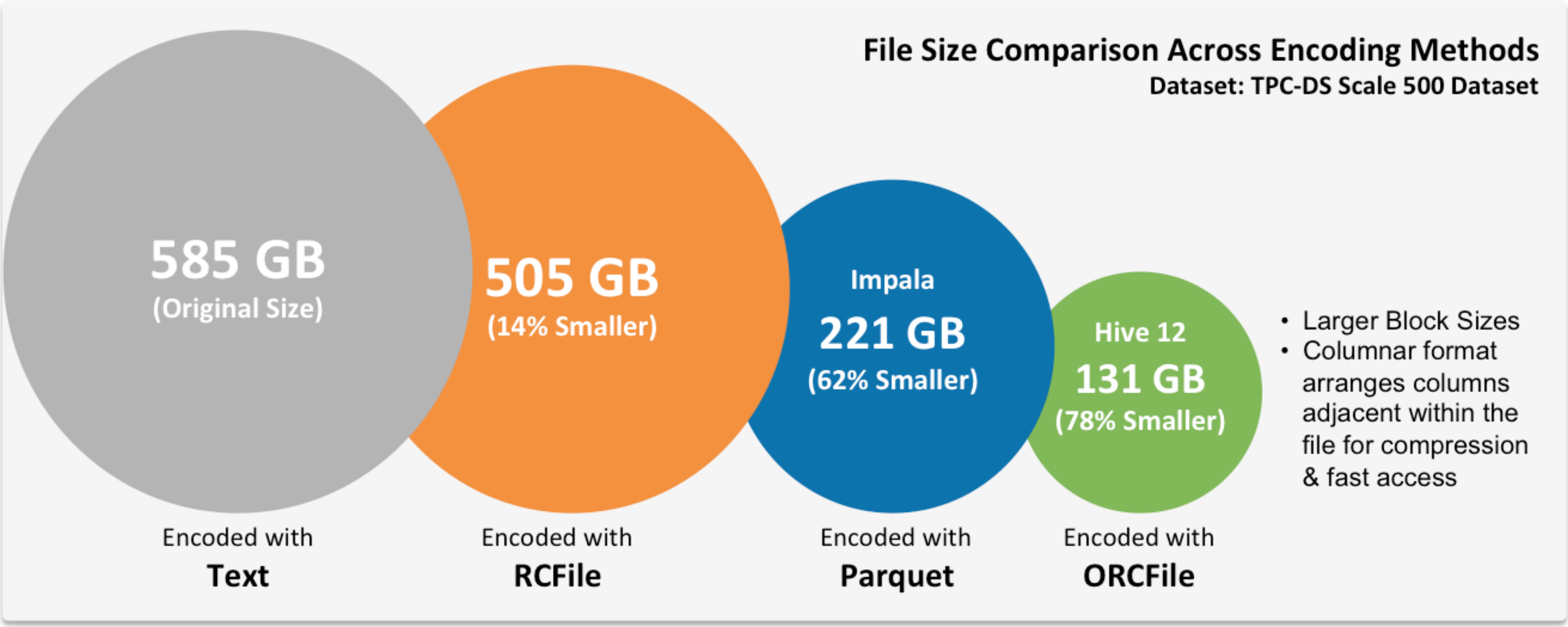
使用ORCFile格式存储列数较多的表

Improve speed of query:
Only read required columns

http://zh.hortonworks.com/blog/orcfile-in-hdp-2-better-compression-better-performance/

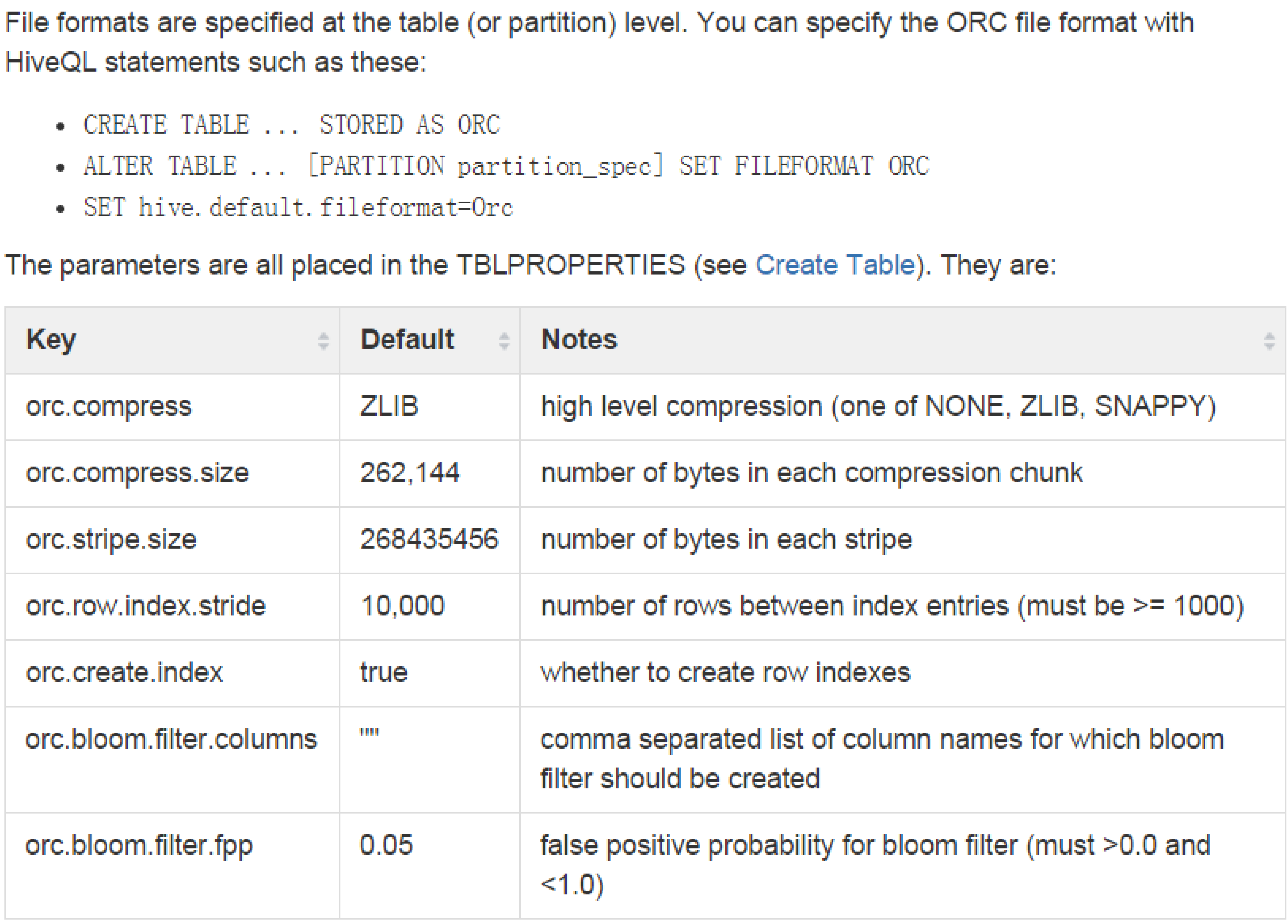
https://cwiki.apache.org/confluence/display/Hive/LanguageManual+ORC

High Compression Compare



- Larger Block Sizes
- Columnar format arranges columns adjacent within the file for compression & fast access

ORC Files



File formats are specified at the table (or partition) level. You can specify the ORC file format with HiveQL statements such as these:

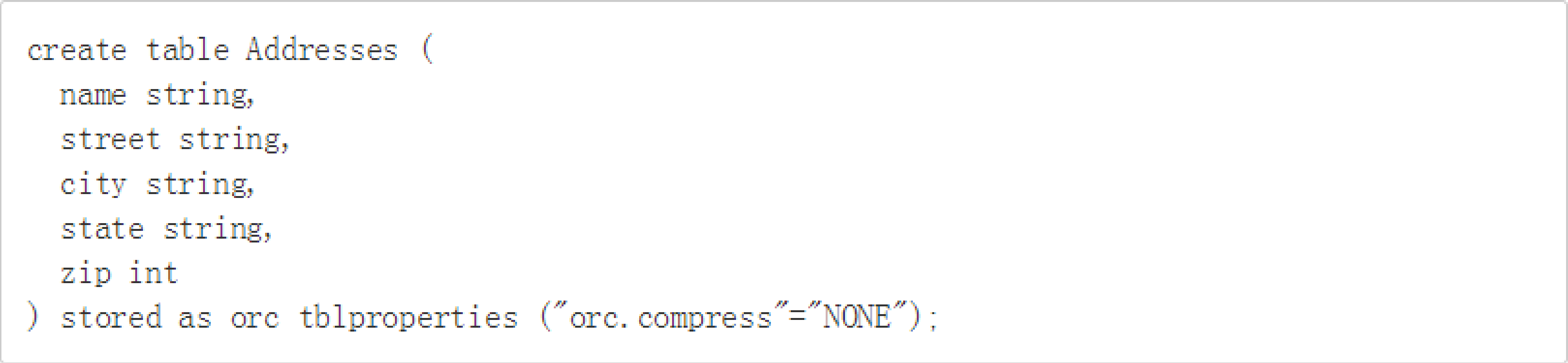
- `CREATE TABLE ... STORED AS ORC`
- `ALTER TABLE ... [PARTITION partition_spec] SET FILEFORMAT ORC`
- `SET hive.default.fileformat=Orc`

The parameters are all placed in the TBLPROPERTIES (see Create Table). They are:

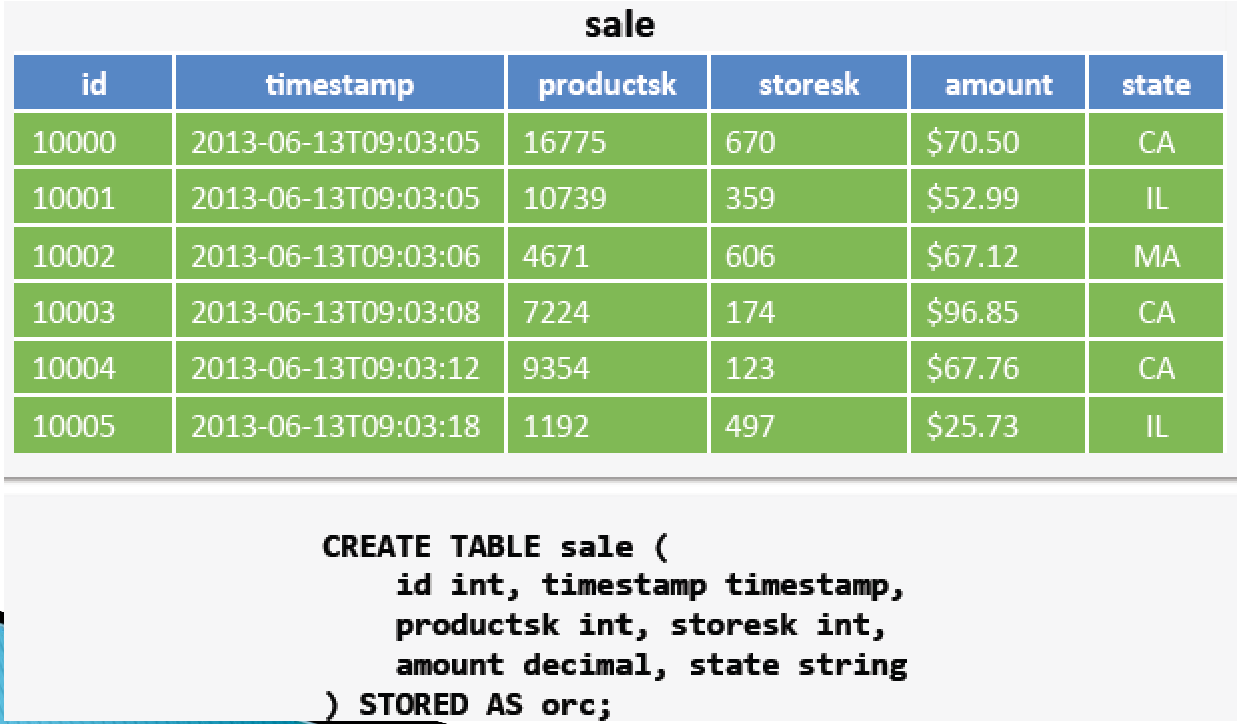
Key	Default	Notes
orc.compress	ZLIB	high level compression (one of NONE, ZLIB, SNAPPY)
orc.compress.size	262,144	number of bytes in each compression chunk
orc.stripe.size	268435456	number of bytes in each stripe
orc.row.index.stride	10,000	number of rows between index entries (must be >= 1000)
orc.create.index	true	whether to create row indexes
orc.bloom.filter.columns	""	comma separated list of column names for which bloom filter should be created
orc.bloom.filter.fpp	0.05	false positive probability for bloom filter (must >0.0 and <1.0)

ORC Files

For example, creating an ORC stored table without compression:



```
create table Addresses (
  name string,
  street string,
  city string,
  state string,
  zip int
) stored as orc tblproperties ("orc.compress"="NONE");
```



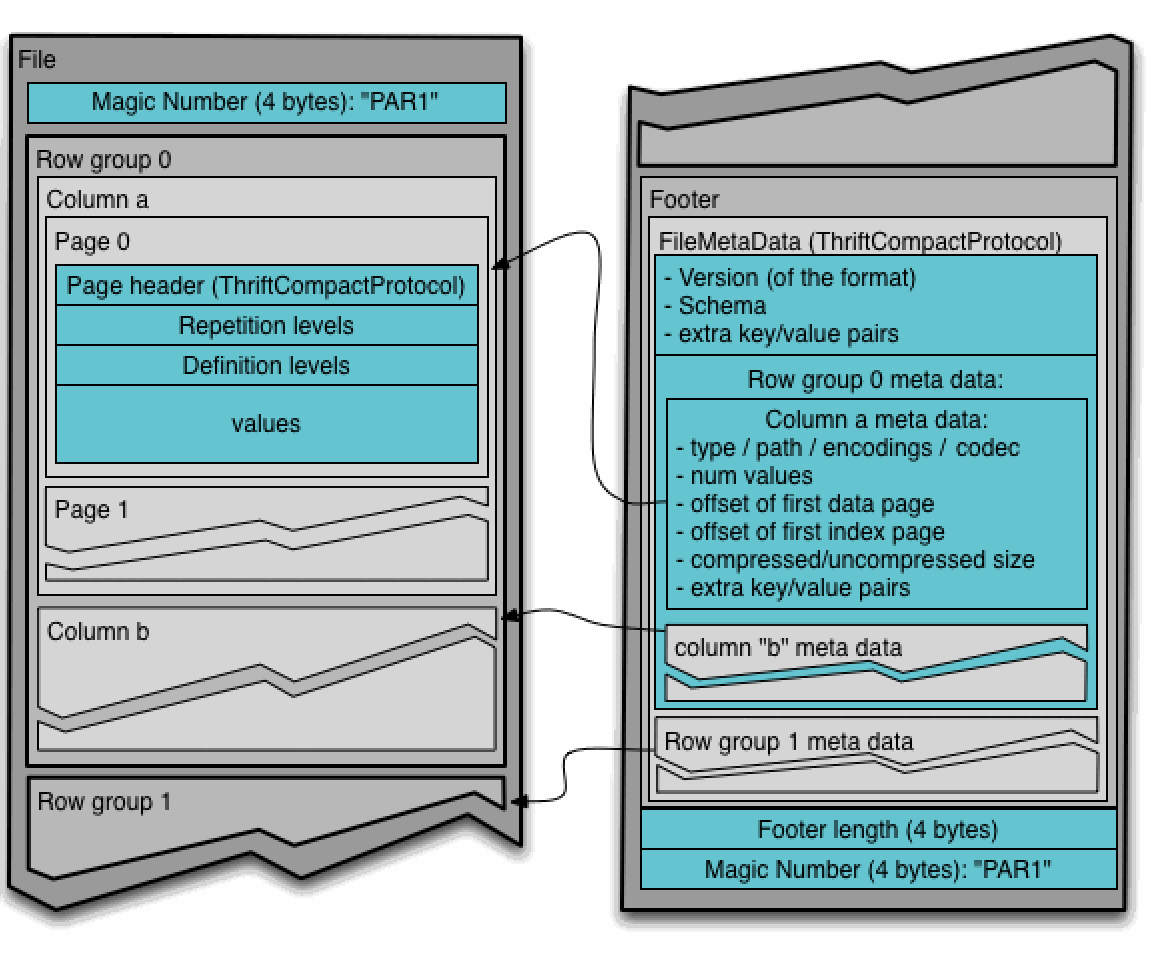
sale

id	timestamp	productsk	storesk	amount	state
10000	2013-06-13T09:03:05	16775	670	$70.50	CA
10001	2013-06-13T09:03:05	10739	359	$52.99	IL
10002	2013-06-13T09:03:06	4671	606	$67.12	MA
10003	2013-06-13T09:03:08	7224	174	$96.85	CA
10004	2013-06-13T09:03:12	9354	123	$67.76	CA
10005	2013-06-13T09:03:18	1192	497	$25.73	IL

```
CREATE TABLE sale (
    id int, timestamp timestamp,
    productsk int, storesk int,
    amount decimal, state string
) STORED AS orc;
```

Parquet

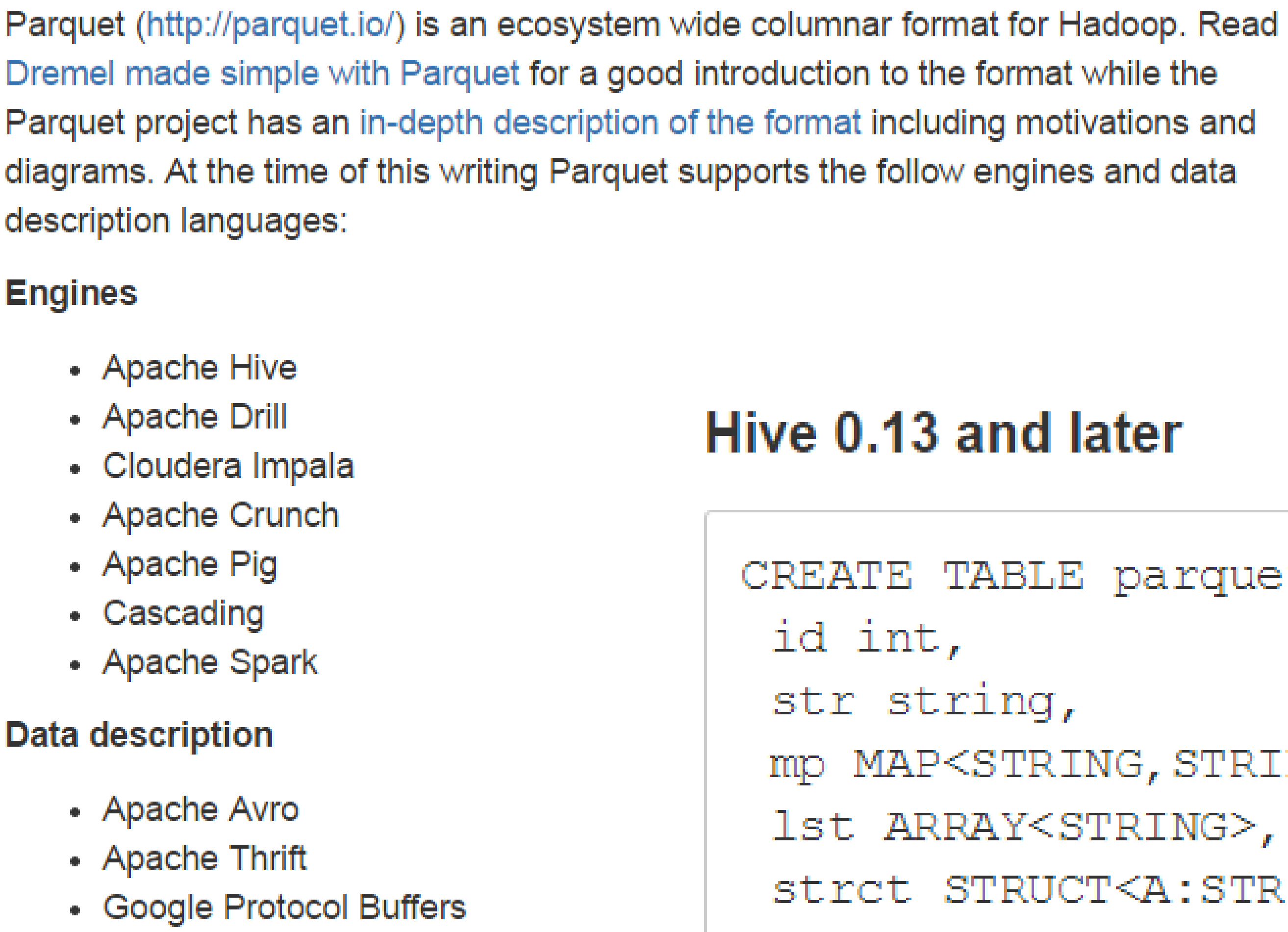
Parquet(twitter+cloudera开源，被Hive、Spark、Drill、Impala、Pig等支持)


File
Magic Number (4 bytes): "PAR1"
Row group 0
Column a
Page 0
Page header (ThriftCompactProtocol)
Repetition levels
Definition levels
values
Page 1
Column b
Row group 1
Footer
FileMetaData (ThriftCompactProtocol)
- Version (of the format)
- Schema
- extra key/value pairs
Row group 0 meta data:
Column a meta data:
- type / path / encodings / codec
- num values
- offset of first data page
- offset of first index page
- compressed/uncompressed size
- extra key/value pairs
column "b" meta data
Row group 1 meta data
Footer length (4 bytes)
Magic Number (4 bytes): "PAR1"

Parquet比较复杂，其灵感主要来自于dremel，parquet存储结构的主要亮点是**支持嵌套数据结构**以及**高效且种类丰富算法**(以应对不同值分布特征的压缩)。

https://cwiki.apache.org/confluence/display/Hive/Parquet

Parquet



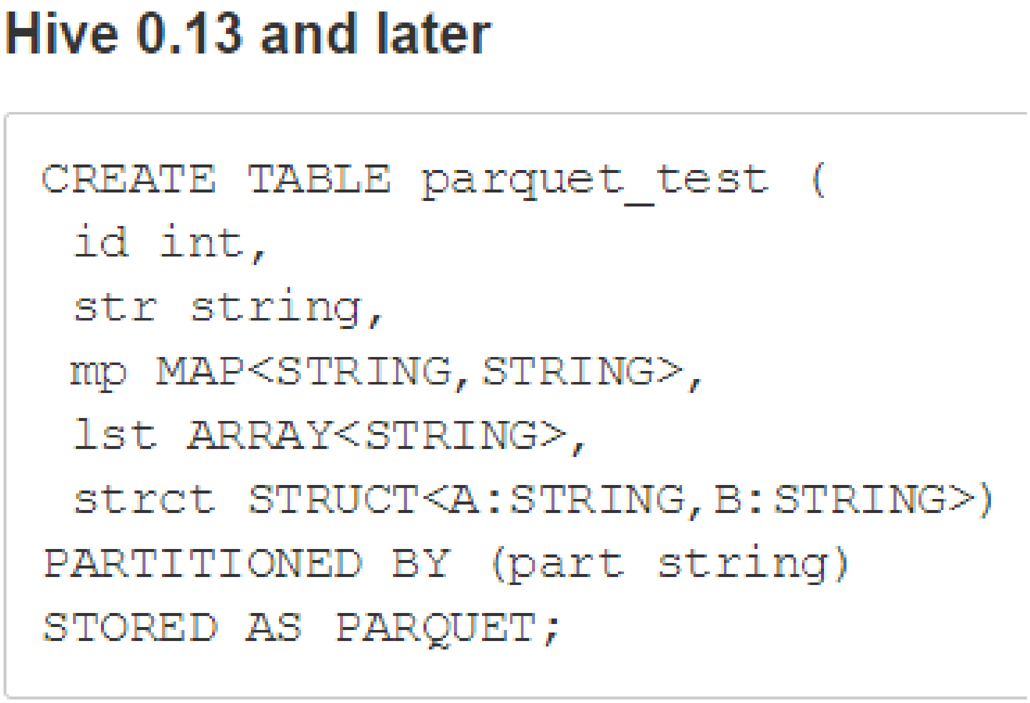
Parquet (http://parquet.io/) is an ecosystem wide columnar format for Hadoop. Read Dremel made simple with Parquet for a good introduction to the format while the Parquet project has an in-depth description of the format including motivations and diagrams. At the time of this writing Parquet supports the follow engines and data description languages:

Engines

- Apache Hive
- Apache Drill
- Cloudera Impala
- Apache Crunch
- Apache Pig
- Cascading
- Apache Spark

Data description

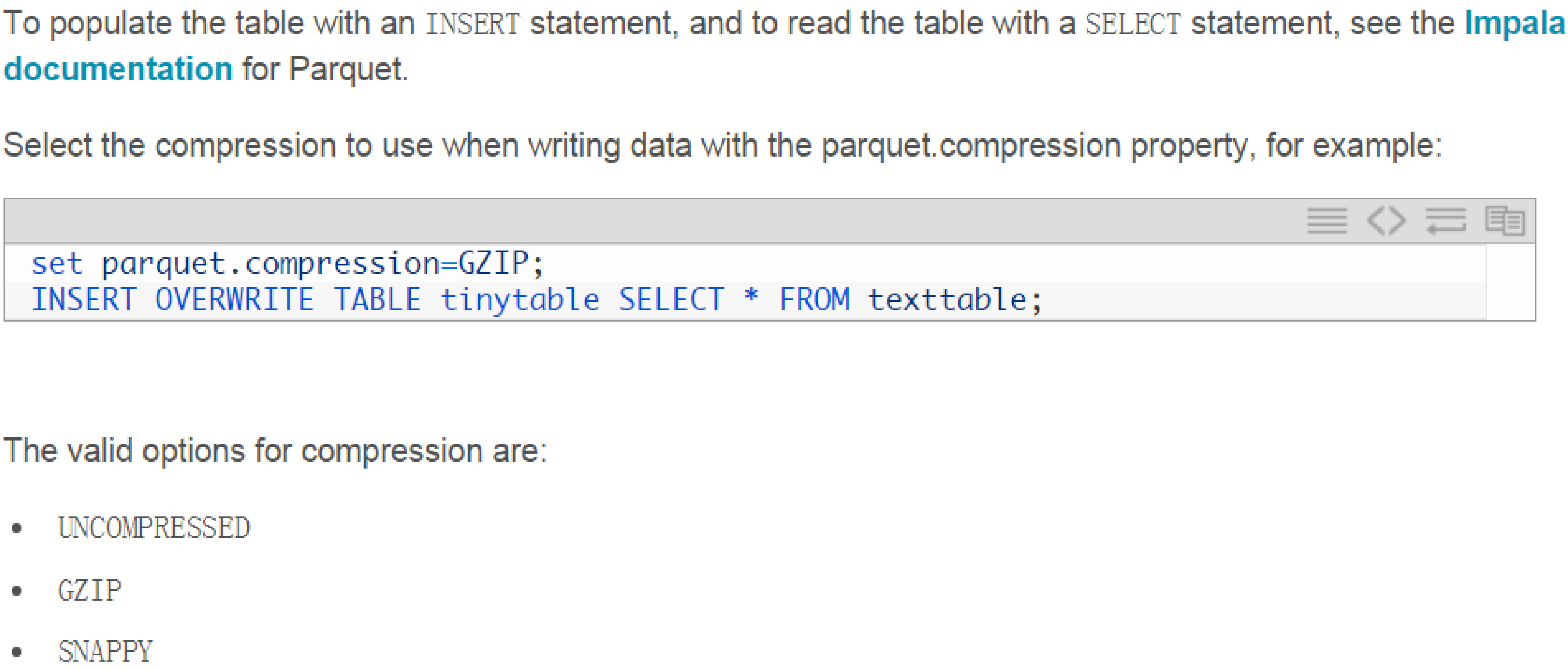
- Apache Avro
- Apache Thrift
- Google Protocol Buffers



Hive 0.13 and later

```
CREATE TABLE parquet_test (
 id int,
 str string,
 mp MAP<STRING,STRING>,
 lst ARRAY<STRING>,
 strct STRUCT<A:STRING,B:STRING>)
PARTITIONED BY (part string)
STORED AS PARQUET;
```

Parquet



To populate the table with an `INSERT` statement, and to read the table with a `SELECT` statement, see the **Impala documentation** for Parquet.

Select the compression to use when writing data with the parquet.compression property, for example:

```
set parquet.compression=GZIP;
INSERT OVERWRITE TABLE tinytable SELECT * FROM texttable;
```

The valid options for compression are:

- `UNCOMPRESSED`
- `GZIP`
- `SNAPPY`

课程大纲

1 HiveServer2

2 Hive 数据压缩

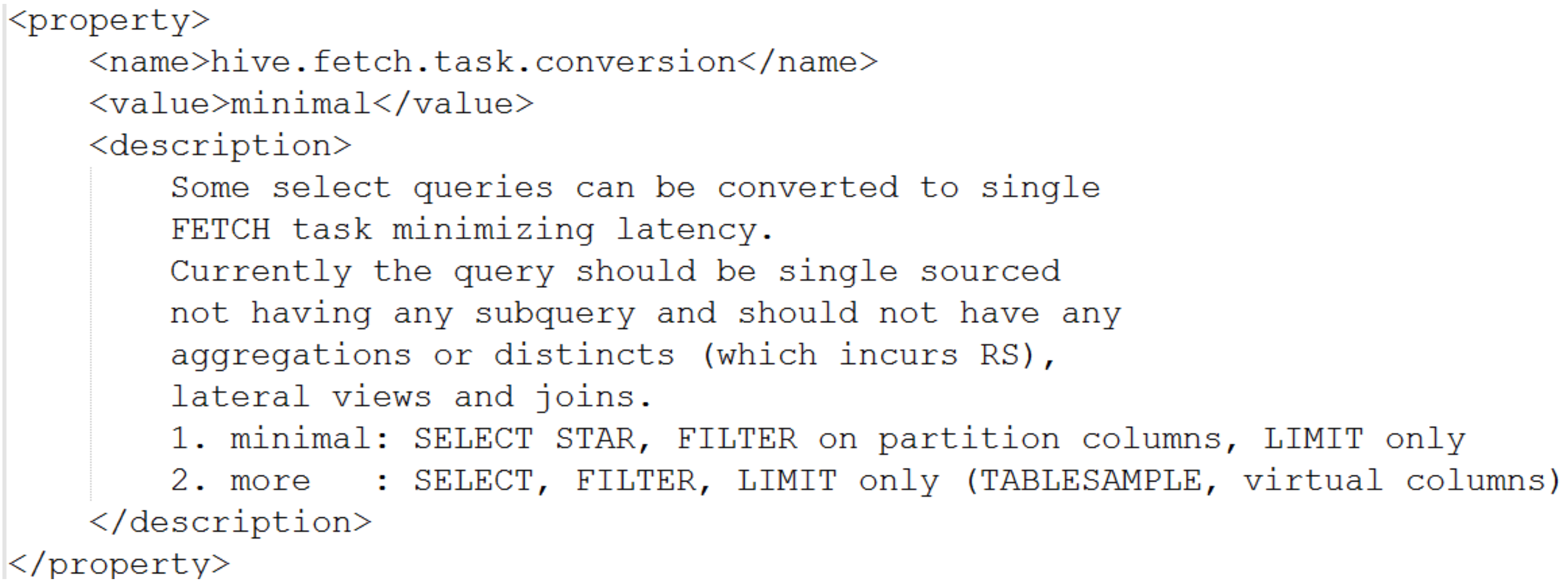
3 Hive 数据存储

4 Hive 企业优化

5 Hive 实战案例

FetchTask

为什么有的sql执行mapreduce，而有的确不？



```
<property>
    <name>hive.fetch.task.conversion</name>
    <value>minimal</value>
    <description>
        Some select queries can be converted to single
        FETCH task minimizing latency.
        Currently the query should be single sourced
        not having any subquery and should not have any
        aggregations or distincts (which incurs RS),
        lateral views and joins.
        1. minimal: SELECT STAR, FILTER on partition columns, LIMIT only
        2. more   : SELECT, FILTER, LIMIT only (TABLESAMPLE, virtual columns)
    </description>
</property>
```

hive.fetch.task.conversion

Explain Execution Plan

Hive provides an EXPLAIN command that shows the execution plan for a query. The syntax for this statement is as follows:

```
EXPLAIN [EXTENDED|DEPENDENCY] query
```

```
explain select deptno, avg(sal)avg_sal from emp group by deptno ;
```

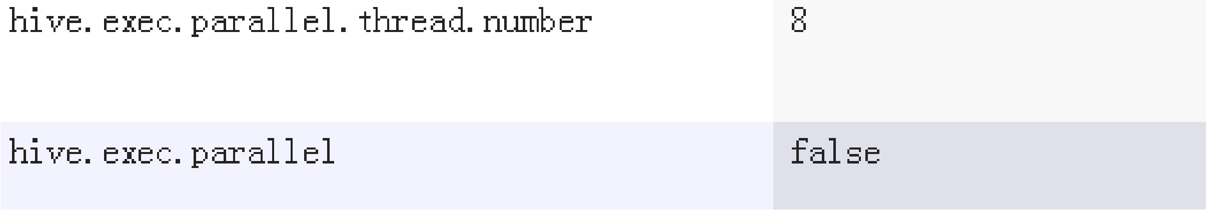
Hive 高级优化

- 大表【拆分】
  - 子表
- 外部表、分区表
  - 结合使用
  - 多级分区
- 数据
  - 存储格式（textfile、orcfile、parquet）
  - 数据压缩（snappy）

- SQL
  - 优化SQL语句
  - join，filter
- MapReduce
  - Reduce Number
  - JVM重用
  - 推测执行

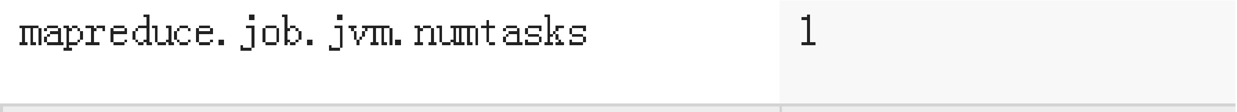
Hive 高级优化

并行执行



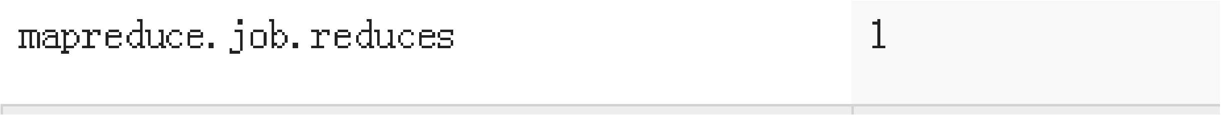
hive.exec.parallel.thread.number	8
hive.exec.parallel	false

JVM重用



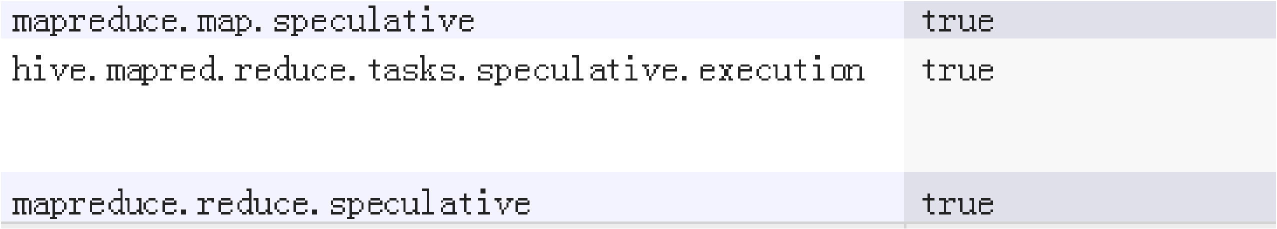
mapreduce.job.jvm.numtasks	1

Reduce数目



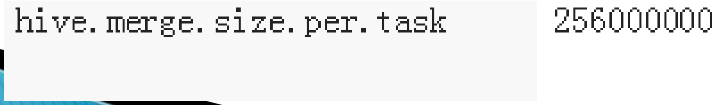
mapreduce.job.reduces	1

推测执行



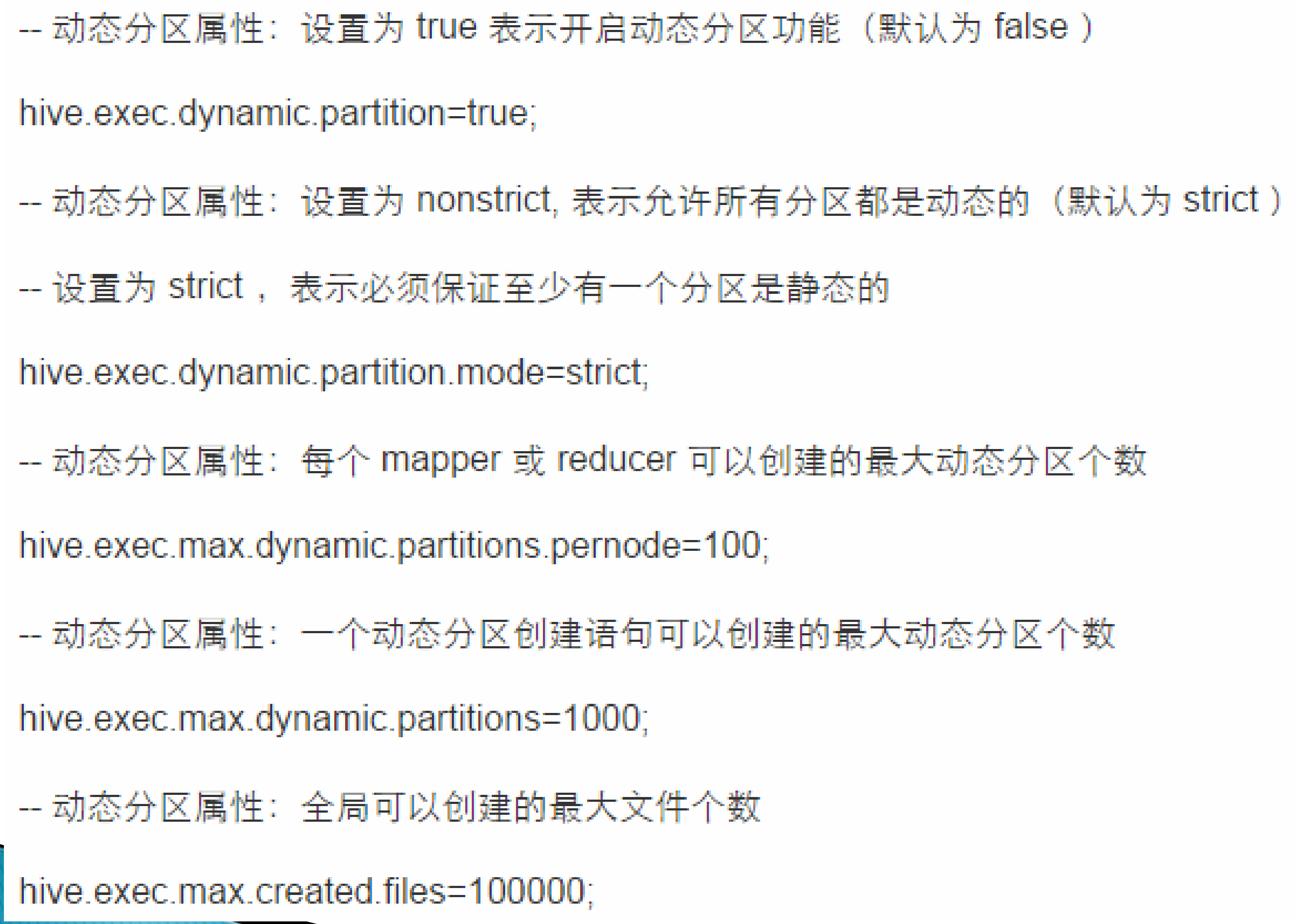
mapreduce.map.speculative	true
hive.mapred.reduce.tasks.speculative.execution	true
mapreduce.reduce.speculative	true

Map数目



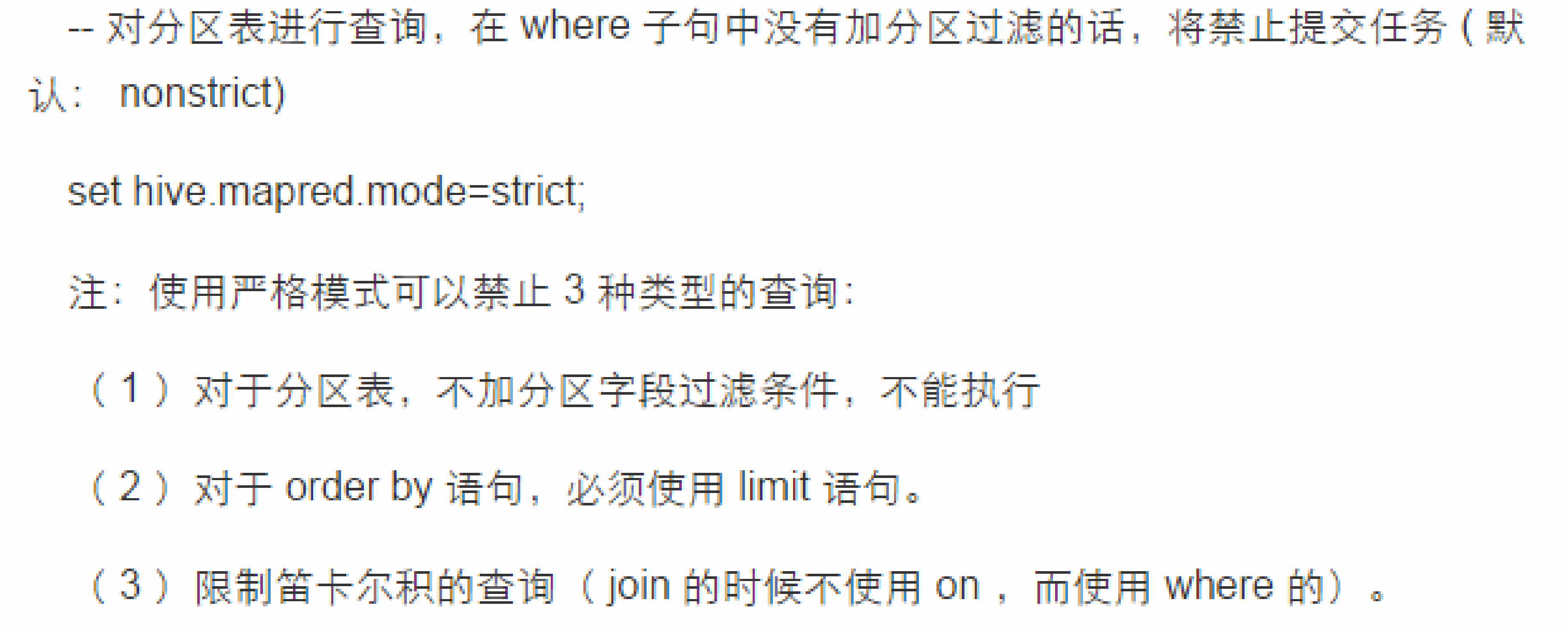
hive.merge.size.per.task	256000000

动态分区调整



```
-- 动态分区属性：设置为 true 表示开启动态分区功能（默认为 false ）

hive.exec.dynamic.partition=true;

-- 动态分区属性：设置为 nonstrict, 表示允许所有分区都是动态的（默认为 strict ）

-- 设置为 strict ，表示必须保证至少有一个分区是静态的

hive.exec.dynamic.partition.mode=strict;

-- 动态分区属性：每个 mapper 或 reducer 可以创建的最大动态分区个数

hive.exec.max.dynamic.partitions.pernode=100;

-- 动态分区属性：一个动态分区创建语句可以创建的最大动态分区个数

hive.exec.max.dynamic.partitions=1000;

-- 动态分区属性：全局可以创建的最大文件个数

hive.exec.max.created.files=100000;
```

Strict Mode



--对分区表进行查询，在where子句中没有加分区过滤的话，将禁止提交任务(默认：nonstrict)

set hive.mapred.mode=strict;

注：使用严格模式可以禁止3种类型的查询：

（1）对于分区表，不加分区字段过滤条件，不能执行

（2）对于order by语句，必须使用limit语句。

（3）限制笛卡尔积的查询（join的时候不使用on，而使用where的）。

数据倾斜

- Common/Shuffle/Reduce Join
- Map Join
- SMB Join

https://cwiki.apache.org/confluence/display/Hive/LanguageManual+Joins

- small table join big table
- big table join big table
- group by xx
- count(distinct xx)

Shuffle/Reduce/Common Join

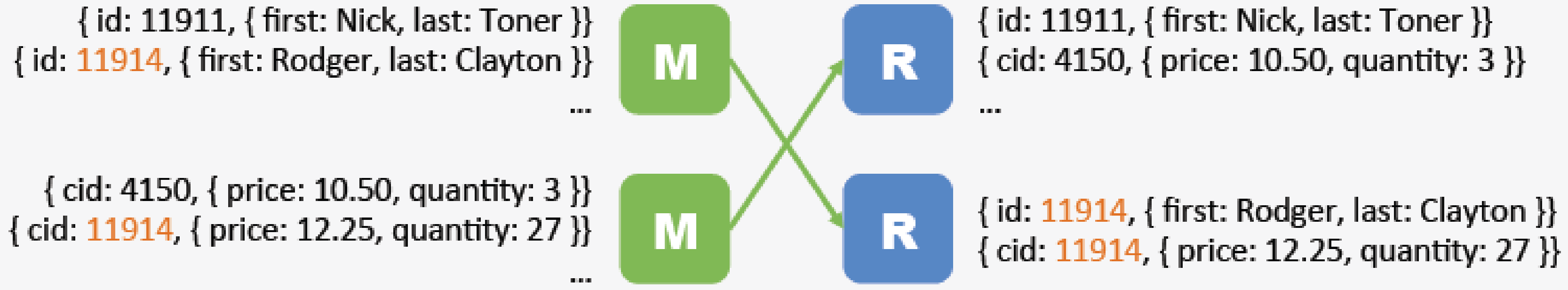
customer

first	last	id
Nick	Toner	11911
Jessie	Simonds	11912
Kasi	Lamers	11913
Rodger	Clayton	11914
Verona	Hollen	11915

order

cid	price	quantity
4150	10.50	3
11914	12.25	27
3491	5.99	5
2934	39.99	22
11914	40.50	10

```
SELECT * FROM customer join order ON customer.id = order.cid;
```



{ id: 11911, { first: Nick, last: Toner }}
{ id: 11914, { first: Rodger, last: Clayton }}
...

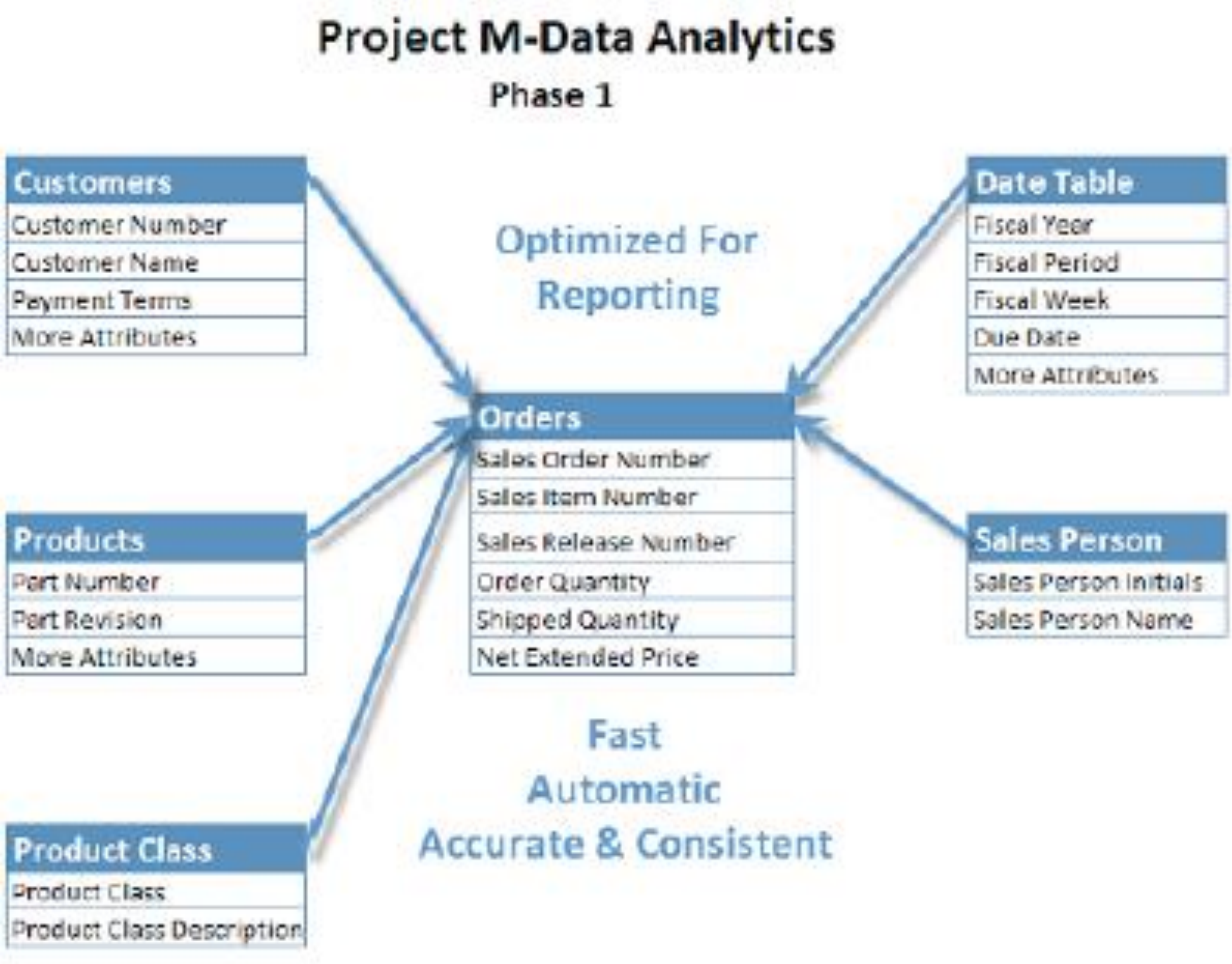
{ cid: 4150, { price: 10.50, quantity: 3 }}
{ cid: 11914, { price: 12.25, quantity: 27 }}
...

{ id: 11911, { first: Nick, last: Toner }}
{ cid: 4150, { price: 10.50, quantity: 3 }}
...

{ id: 11914, { first: Rodger, last: Clayton }}
{ cid: 11914, { price: 12.25, quantity: 27 }}

Identical keys shuffled to the same reducer. Join done reduce-side.
Expensive from a network utilization standpoint.

Map Join



Project M-Data Analytics

Phase 1

Customers
- Customer Number
- Customer Name
- Payment Terms
- More Attributes

Date Table
- Fiscal Year
- Fiscal Period
- Fiscal Week
- Due Date
- More Attributes

Optimized For Reporting

Orders
- Sales Order Number
- Sales Item Number
- Sales Release Number
- Order Quantity
- Shipped Quantity
- Net Extended Price

Products
- Part Number
- Part Revision
- More Attributes

Sales Person
- Sales Person Initials
- Sales Person Name

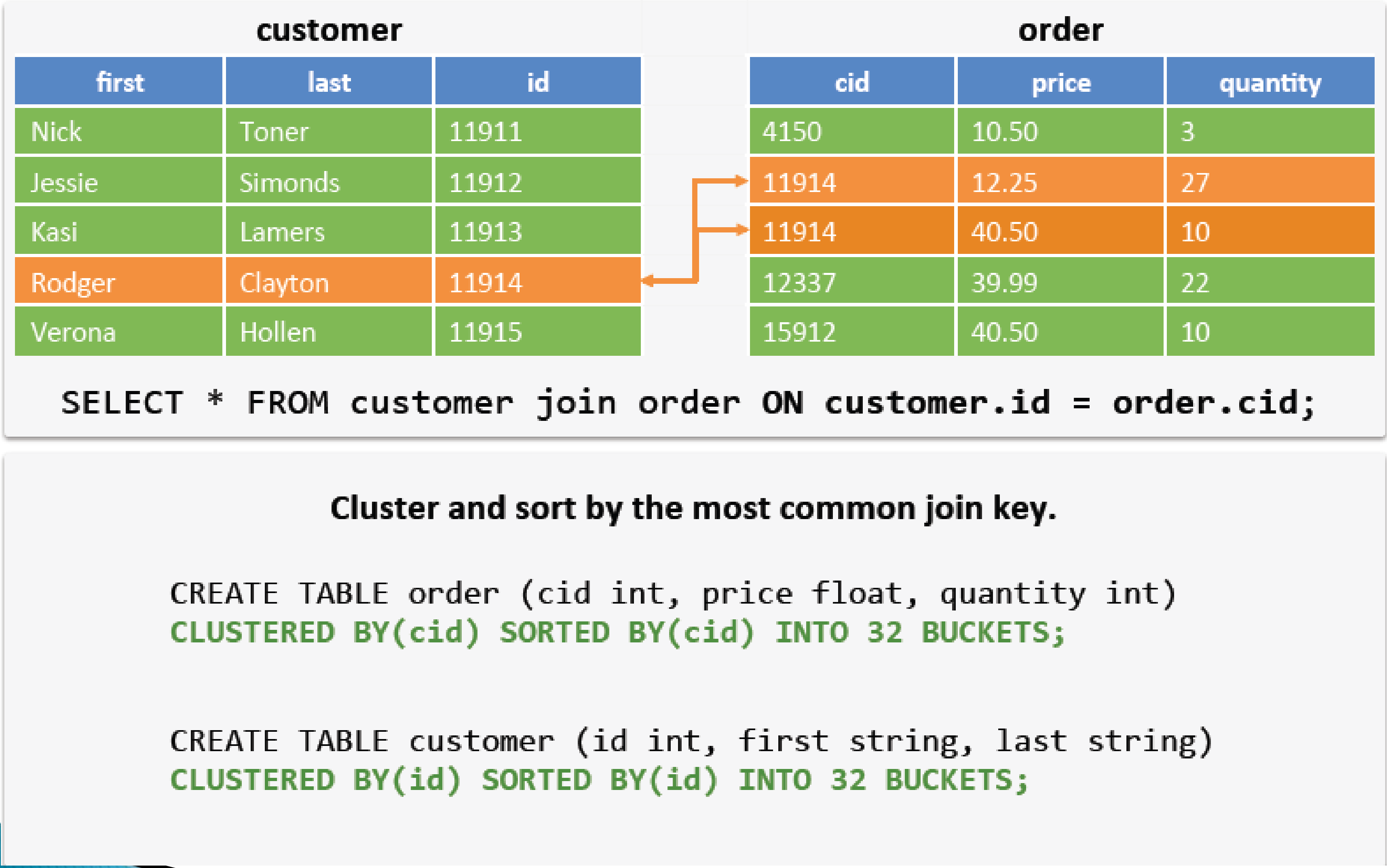
Product Class
- Product Class
- Product Class Description

Fast

Automatic

Accurate & Consistent

SMB（Sort-Merge-Bucket） Join



customer

first	last	id
Nick	Toner	11911
Jessie	Simonds	11912
Kasi	Lamers	11913
Rodger	Clayton	11914
Verona	Hollen	11915

order

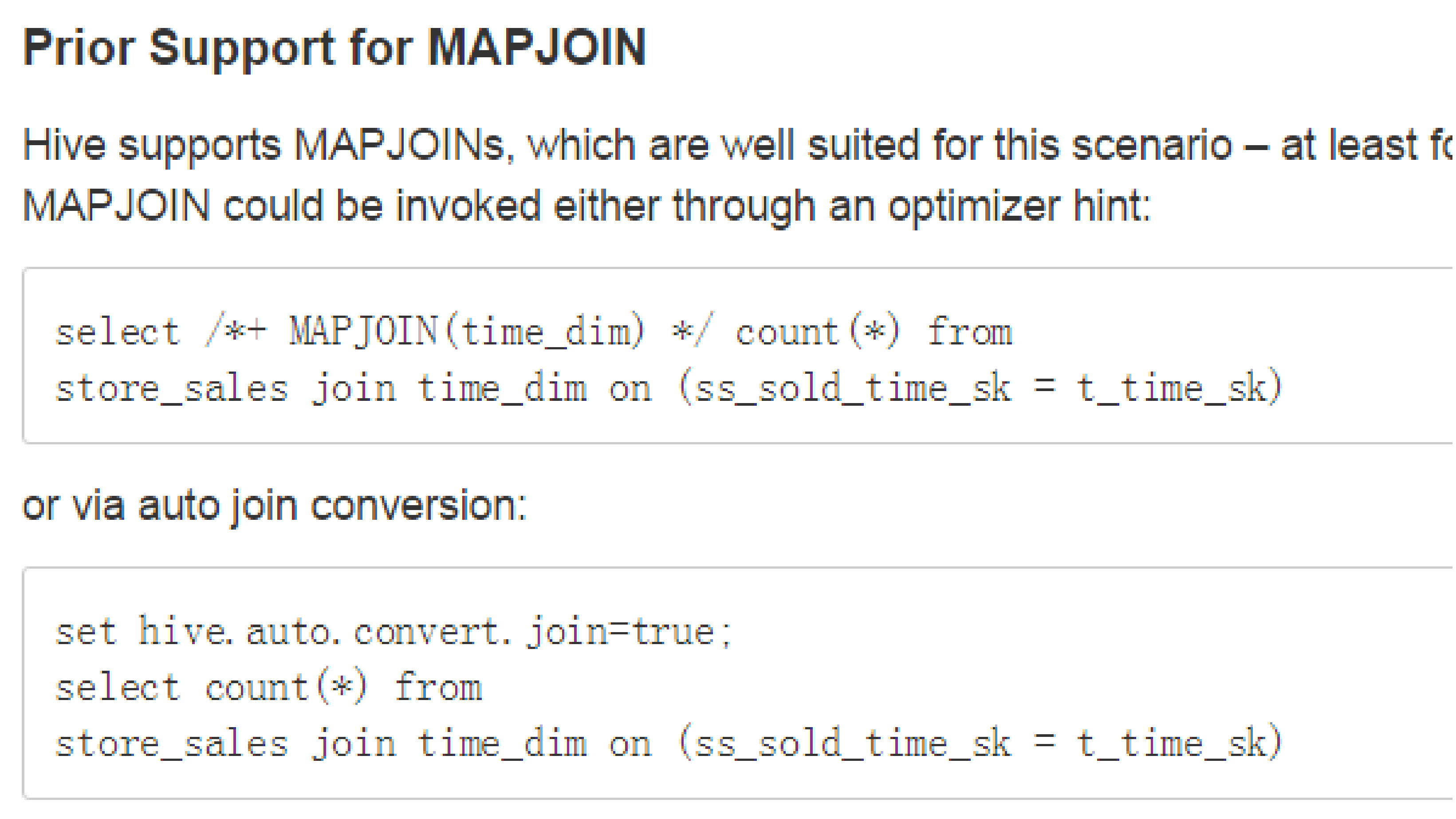
cid	price	quantity
4150	10.50	3
11914	12.25	27
11914	40.50	10
12337	39.99	22
15912	40.50	10

```
SELECT * FROM customer join order ON customer.id = order.cid;
```

Cluster and sort by the most common join key.

```
CREATE TABLE order (cid int, price float, quantity int)
CLUSTERED BY(cid) SORTED BY(cid) INTO 32 BUCKETS;


CREATE TABLE customer (id int, first string, last string)
CLUSTERED BY(id) SORTED BY(id) INTO 32 BUCKETS;
```

http://my.oschina.net/leejun2005/blog/178631



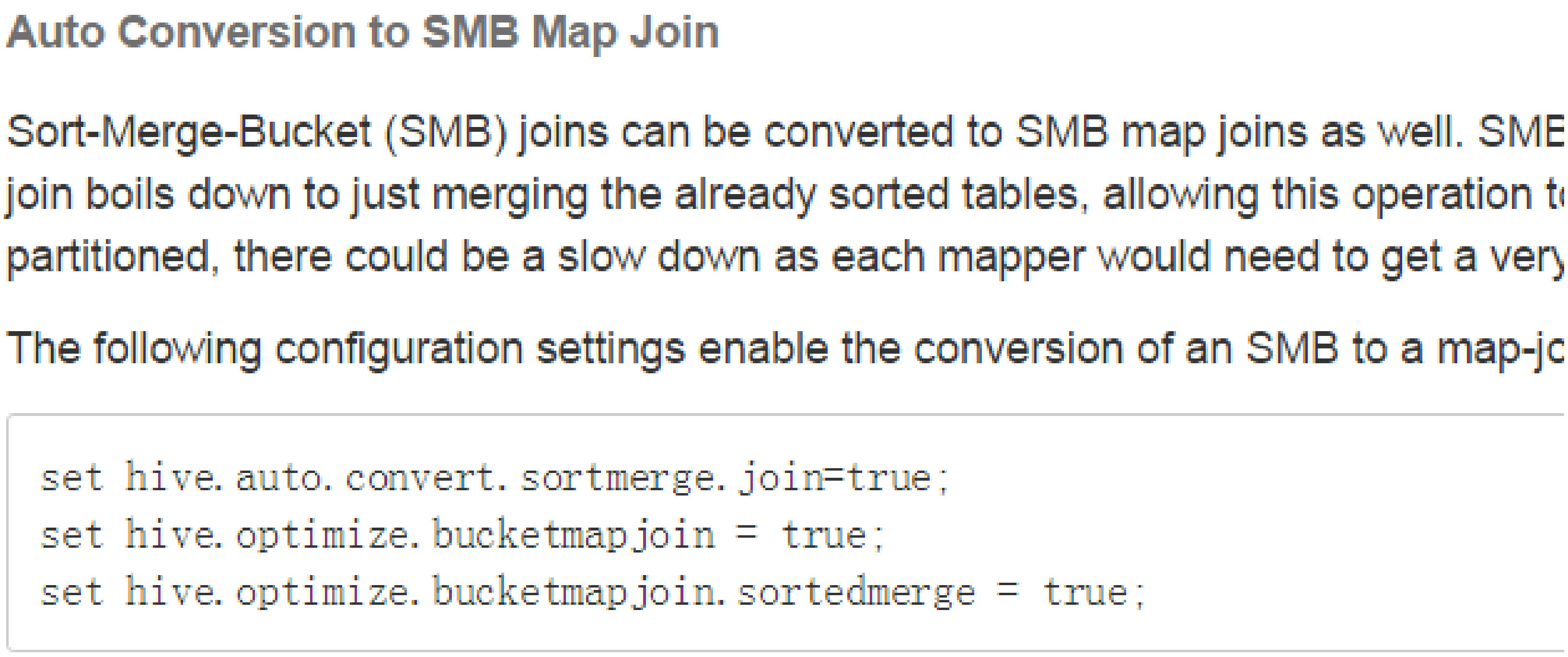
Prior Support for MAPJOIN

Hive supports MAPJOINs, which are well suited for this scenario – at least fo
MAPJOIN could be invoked either through an optimizer hint:

```
select /*+ MAPJOIN(time_dim) */ count(*) from
store_sales join time_dim on (ss_sold_time_sk = t_time_sk)
```

or via auto join conversion:

```
set hive.auto.convert.join=true;
select count(*) from
store_sales join time_dim on (ss_sold_time_sk = t_time_sk)
```



Auto Conversion to SMB Map Join

Sort-Merge-Bucket (SMB) joins can be converted to SMB map joins as well. SMB
join boils down to just merging the already sorted tables, allowing this operation t
partitioned, there could be a slow down as each mapper would need to get a very

The following configuration settings enable the conversion of an SMB to a map-jo

```
set hive.auto.convert.sortmerge.join=true;
set hive.optimize.bucketmapjoin = true;
set hive.optimize.bucketmapjoin.sortedmerge = true;
```

课程大纲

1 HiveServer2

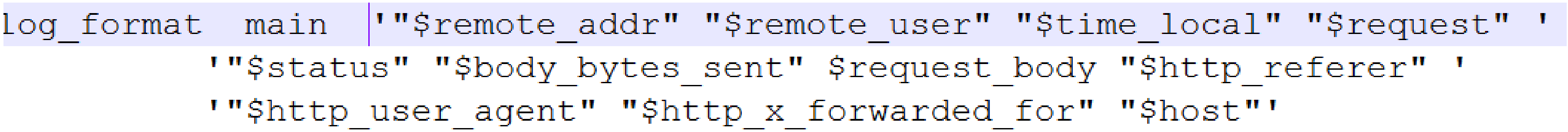
2 Hive 数据压缩

3 Hive 数据存储

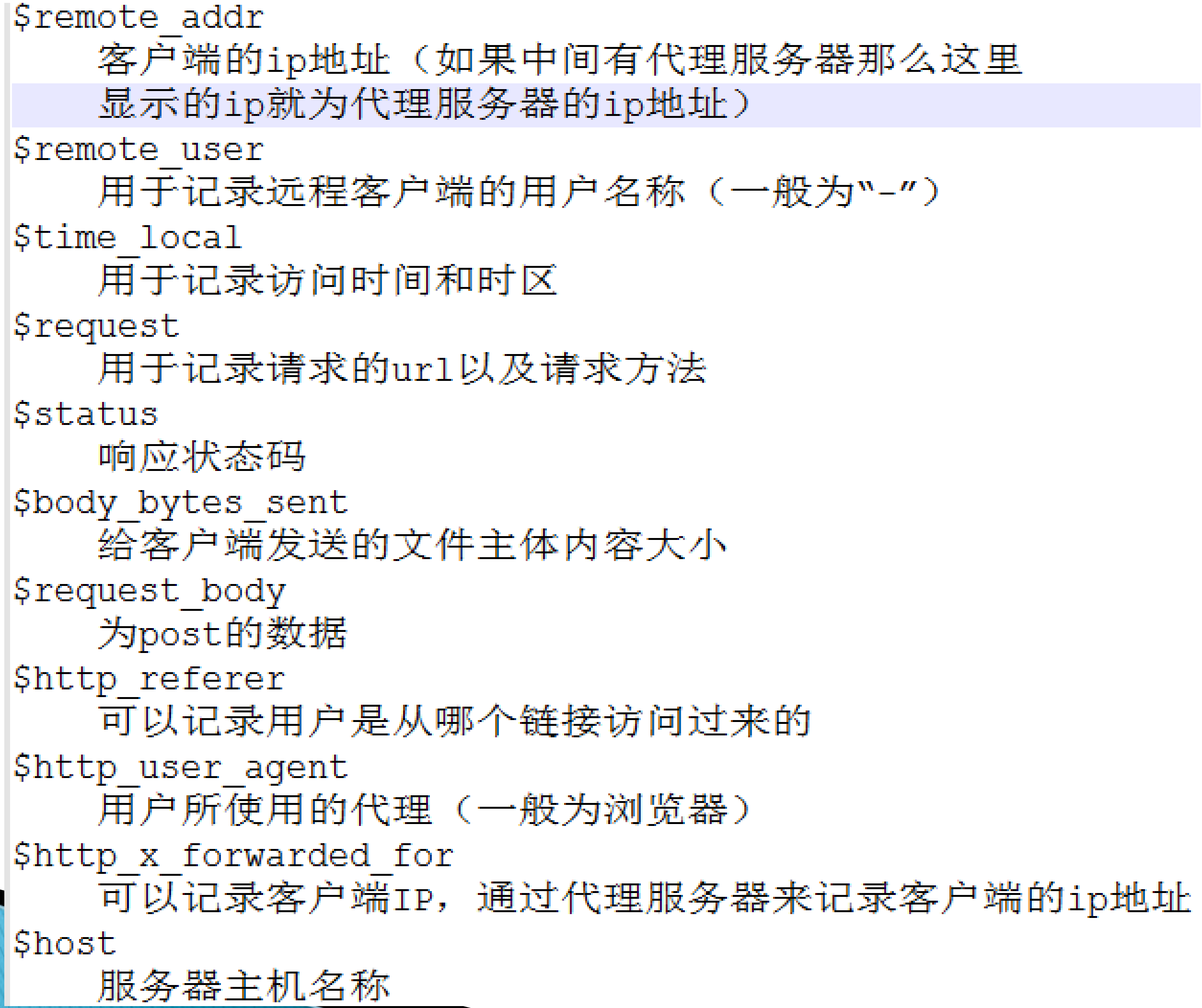
4 Hive 企业优化

5 Hive 实战案例

【北风网】日志格式



```
log_format  main  '"$remote_addr" "$remote_user" "$time_local" "$request" '
          '"$status" "$body_bytes_sent" $request_body "$http_referer" '
          '"$http_user_agent" "$http_x_forwarded_for" "$host"'
```



```
$remote_addr
    客户端的ip地址（如果中间有代理服务器那么这里
    显示的ip就为代理服务器的ip地址）
$remote_user
    用于记录远程客户端的用户名称（一般为“-”）
$time_local
    用于记录访问时间和时区
$request
    用于记录请求的url以及请求方法
$status
    响应状态码
$body_bytes_sent
    给客户端发送的文件主体内容大小
$request_body
    为post的数据
$http_referer
    可以记录用户是从哪个链接访问过来的
$http_user_agent
    用户所使用的代理（一般为浏览器）
$http_x_forwarded_for
    可以记录客户端IP，通过代理服务器来记录客户端的ip地址
$host
    服务器主机名称
```

创建表

```
create table IF NOT EXISTS default.bf_log_src (
remote_addr string,
remote_user string,
time_local string,
request string,
status string,
body_bytes_sent string,
request_body string,
http_referer string,
http_user_agent string,
http_x_forwarded_for string,
host string
)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ' '
stored as textfile ;

load data local inpath '/opt/datas/moodle.ibeifeng.access.log' into table bf_log_src ;
```

可行?

项目案例

- 依据业务数据表
  - 方式一、原始表bf_log_src，加载数据（预先处理）
  - 方式二、创建正则表RegexSerDe

  https://cwiki.apache.org/confluence/display/Hive/GettingStarted

  http://wpjam.qiniudn.com/tool/regexpal/

- 数据ETL
  - 拆分表（子表）、数据存储格式
  - 数据预处理ETL（udf、python）

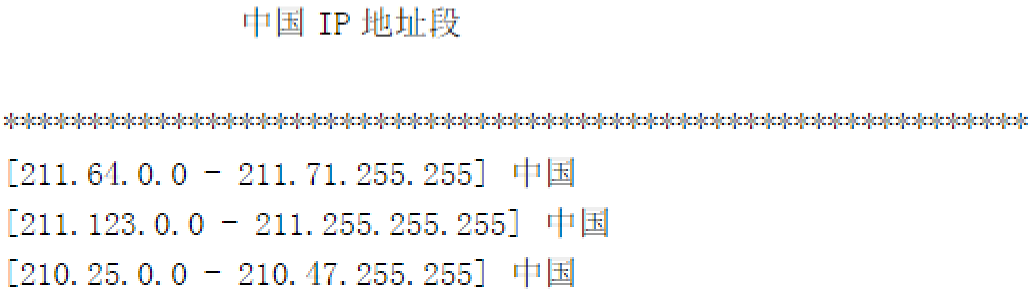
- 数据分析HQL

业务需求之IP地址

- 【$remote_addr】

　　客户端的ip地址（如果中间有代理服务器那么这里显示的ip就为代理服务器的ip地址）。

- 业务
  - 依据ip地址确定区域，定向营销，【IP地址 -> 地域】



中国 IP 地址段

```
*****************************************************************
[211.64.0.0 - 211.71.255.255] 中国
[211.123.0.0 - 211.255.255.255] 中国
[210.25.0.0 - 210.47.255.255] 中国
```

  - 用户统计，访问某一网站数
    - ● 准确性（同一外网，不同内网）

业务需求之访问时间

- 【$time_local】

  用于记录访问时间和时区。

- 业务
  - 分析用户访问网站的时间段
  - 针对销售来说，合理安排值班，销售课程

业务需求之请求地址

- 【$request】

  用于记录请求的url以及请求方法。

- 业务

  - 用户最关注的网站 -> 课程
  - 定向投放此套课程，做好的相关课程

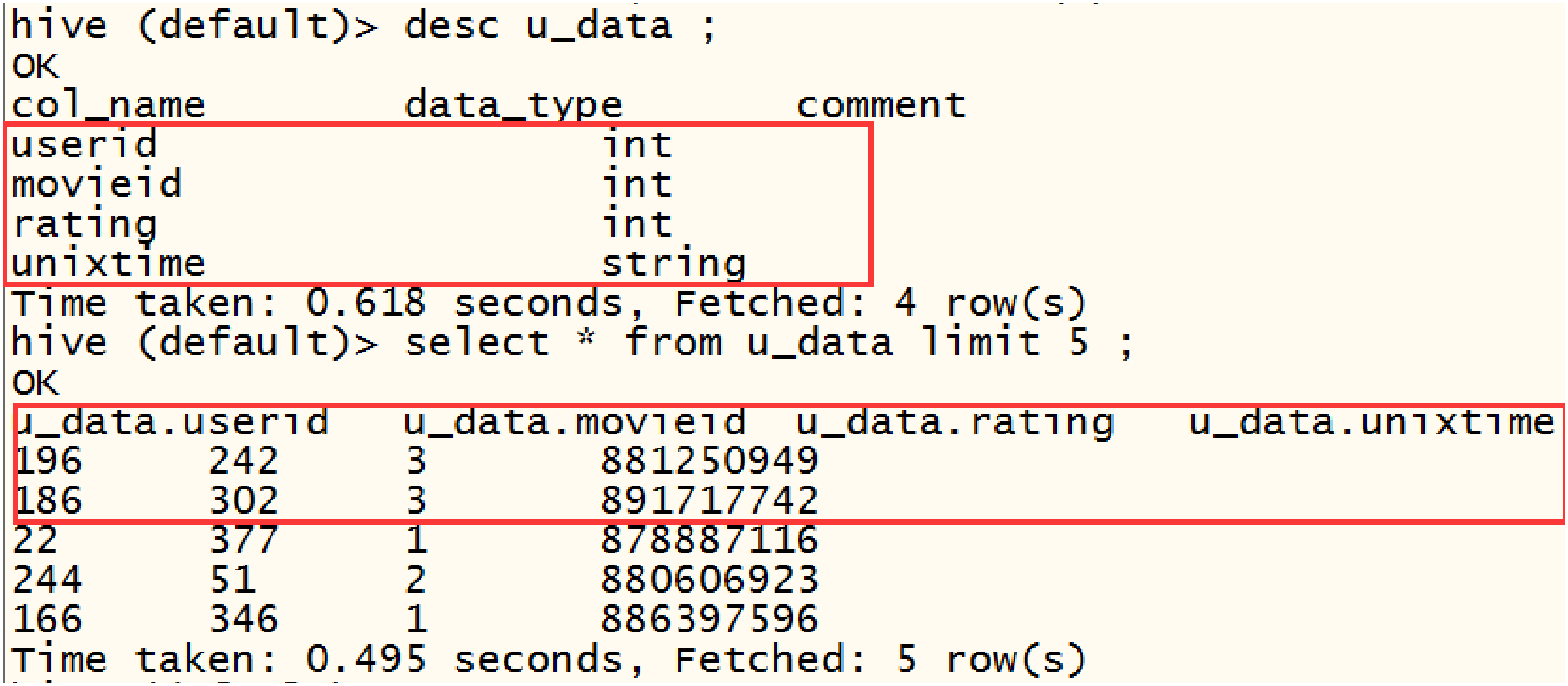
业务需求之转入链接

- 【$http_referer】

  可以记录用户是从哪个链接访问过来的。

- 业务

  - 关注用户如何，访问我们的课程，定向某个区域，进行广告投放



```
hive (default)> desc u_data ;
OK
col_name                data_type               comment
userid                  int
movieid                 int
rating                  int
unixtime                string
Time taken: 0.618 seconds, Fetched: 4 row(s)
hive (default)> select * from u_data limit 5 ;
OK
u_data.userid   u_data.movieid  u_data.rating   u_data.unixtime
196     242     3       881250949
186     302     3       891717742
22      377     1       878887116
244     51      2       880606923
166     346     1       886397596
Time taken: 0.495 seconds, Fetched: 5 row(s)
```