# Hbase 表设计

HBase与RDBMS的区别在于：HBase的Cell（每条数据记录中的数据项）是具有版本描述的（versioned），行是有序的，列（qualifier）在所属列簇（Column families）存在的情况下，由客户端自由添加。以下的几个因素是Hbase Schema设计需要考虑的问题：

1. Hbase中没有joins的概念

大表的结构可以使得不需要joins，而解决这一问题。

2、Row keys 设计

主键，在Region里按字母顺序来排序(byte数组存储)。

写入要分散，如订单表： order\_id做reverse逆排序后做rowkey，以便分布式存储，避免数据只保存在个别节点上。

多条件查询时，设为组合row key，左匹配

注：读取数据只能按row key（及其range）或scan全表扫描，确保查询高效

3、列族CF设计

尽量少，建议CF数量在1-2个。

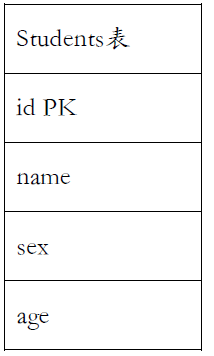
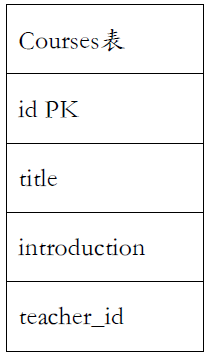
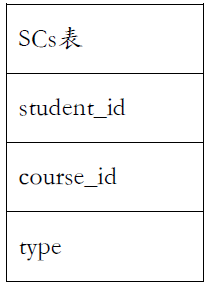
设计Hbase schema的时候，要尽量只有一个column family。

flush和compaction触发的基本单位都是Region级别。当一个CF有大量的数据的时候会触发整个region里面的其他CF的memstore（其实这些memstore可能仅有少量的数据，还不需要flush的）也发生flush动作；

另外compaction触发的条件是当store file的个数（不是总的store file的大小）达到一定数量的时候会发生，而flush产生的大量store file通常会导致compaction，flush/compaction会发生很多IO相关的负载，这对Hbase的整体性能有很大影响，所以选择合适的column family个数很重要。

案例1、学生表和课程表，多对多

关系数据库中设计：

HBase中：

Student 表

|  |  |  |
| --- | --- | --- |
| Row Key | CF | CF |
|  | info | course |
| Student\_id（reverse逆排序） | Info:name  Info:age  Info:sex | Course:c1  Course:c2  … |

Course表

|  |  |  |
| --- | --- | --- |
| Row Key | CF | CF |
|  | info | Student |
| Course\_id（reverse逆排序） | Info:title  Info:introduction  Info:teacher\_id | Student:t1=student\_id  Student:t2  … |

案例2、person 和 身份证card表

关系数据库：

|  |
| --- |
| Person表 |
| Pserson\_id |
| Name |
| sex |
| Card |

|  |
| --- |
| Card表 |
| ID |
| City\_id |

HBase中：

Person表

|  |  |
| --- | --- |
| Row Key | CF |
|  | info |
| Person\_id | Info:name  Info:sex  Info:age  Info:card |

……

案例2、订单order 和 订单明细表order\_item，一对多

关系数据库中：

Order表

|  |
| --- |
| Order\_Id |
| Order\_mount |
| User\_id |
| City\_id |
| …. |

Order\_item 表

|  |
| --- |
| Item\_Id |
| Order\_Id |
| Product\_id |
| Unit\_price |
| Order\_num |
| ….. |

HBase 中：

方案一：

Order表

|  |  |
| --- | --- |
| Row key | CF |
|  | info |
| Order\_Id | Info: Order\_mount  Info:user\_id  Info:city\_id  ….. |

Order\_item 表

|  |  |
| --- | --- |
| Row key | CF |
|  | info |
| Order\_id-item\_id | Info: Order\_mount  Info:user\_id  Info:city\_id  Info:order\_id  ….. |

方案二：

Order表

|  |  |  |
| --- | --- | --- |
| Row key | CF | CF |
|  | info | item |
| Order\_Id | Info: Order\_mount  Info:user\_id  Info:city\_id  Info:create\_time  ….. | Item:i1<item\_id>  Item:i2  ….. |

Order\_item 表

|  |  |
| --- | --- |
| Row key | CF |
|  | info |
| item\_id | Info: Order\_mount  Info:user\_id  Info:city\_id  Info:order\_id  ….. |

多条件查询时，需要增加一个index表

Row key为查询条件组合

|  |  |
| --- | --- |
| Row key | CF |
|  | info |
| [create\_time]\_[order\_id]\_[city\_id] | Info:order\_id |