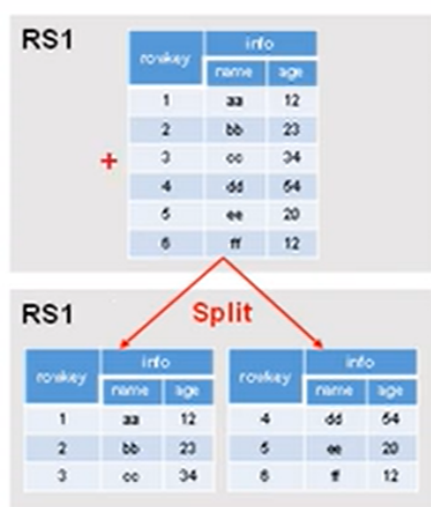


HBase11 Split流程

默认情况下，每个Table起初只有一个Region，随着数据的不断写入，Region会自动进行拆分。刚拆分时，两个子Region都位于当前的Region Server，但处于负载均衡的考虑，HMaster有可能会将某个Region转移给其他的Region Server。

Region Split时机：

1. 当1个region中某一个Store下所有StoreFile的总大小超过hbase.hregion.max.filesize，该Region就会进行拆分（0.94版本之前）
2. 当1个region中某一个Store下所有StoreFile的总大小超过 $\text{Min}(R^2 * \text{"hbase.hregion.memstore.flush.size"}, \text{"hbase.hregion.max.filesize"})$ ，该Region就会进行拆分，其中R为当前Region Server中属于该Table的个数



1. 当1个region中的某个Store下所有StoreFile的总大小超过 "hbase.hregion.max.filesize"，该region就会进行拆分（0.94版之前）

2. 当1个region中的某个Store下所有StoreFile的总大小超过 $\text{Min}(R^2 * \text{"hbase.hregion.memstore.flush.size"}, \text{"hbase.hregion.max.filesize"})$ 就会拆分，其中R为当前RegionServer中属于该table的region个数（0.94版之后）

为什么尽量设计少量的列族

因为flush每个region会生成一个文件，这时候如果一个列族数据很多，一个列族数据很少，会出现有大文件和小文件同时存在的情况。