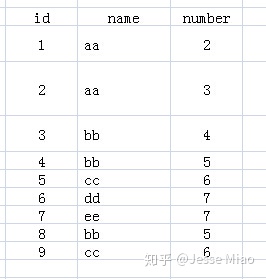
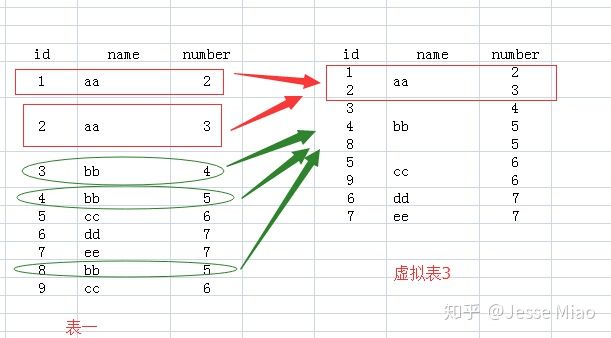
Select ... where ... Group by A

理解的时候是先执行where，筛出符合条件的行，然后对这些行执行按字段A group by，即是按照A分组，然后再看Select是要哪些字段，以及做哪些聚合，如果是做聚合则是按分组来做，一组组来。看一下例子：

先来看下表1，表名为test：



FROM test Group BY name：该句执行后，我们想象生成了**虚拟表3，**如下所图所示，生成过程是这样的：group by name，那么找name那一列，具有相同name值的行，合并成一行，如对于name值为aa的，那么<1 aa 2>与<2 aa 3>两行合并成1行，所有的id值和number值写到一个单元格里面。



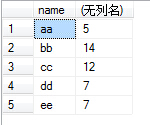
接下来就要针对**虚拟表3**执行Select语句了：

（1）如果执行select \*的话，那么返回的结果应该是虚拟表3，可是id和number中有的单元格里面的内容是多个值的，而关系数据库就是基于关系的，单元格中是不允许有多个值的，所以你看，执行select \* 语句就报错了。

（2）我们再看name列，每个单元格只有一个数据，所以我们select name的话，就没有问题了。为什么name列每个单元格只有一个值呢，因为我们就是用name列来group by的。

（3）那么对于id和number里面的单元格有多个数据的情况怎么办呢？答案就是用**聚合函数，聚合函数就用来输入多个数据，输出一个数据的。**如cout(id)，sum(number)，而每个聚合函数的输入就是每一个多数据的单元格。

（4）例如我们执行select name,sum(number) from test group by name，那么sum就对虚拟表3的number列的每个单元格进行sum操作，例如对name为aa的那一行的number列执行sum操作，即2+3，返回5，最后执行结果如下：



（5）group by 多个字段该怎么理解呢：如group by name,number，我们可以把name和number 看成一个**整体字段**，以他们整体来进行分组的。