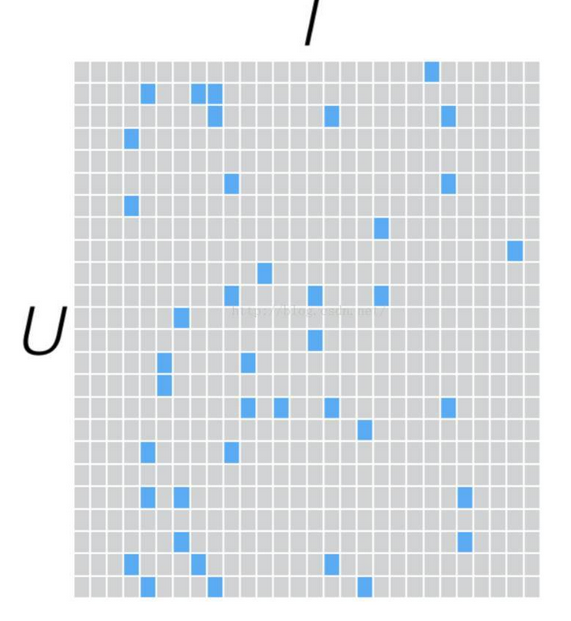
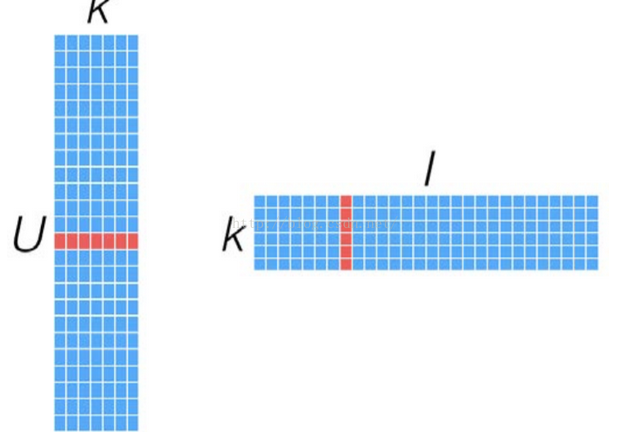
RecSys 说明：

计划任务，每天运行一次，将推荐表存在mysql

1. 开启hadoop：
   1. start-all.sh 打开所有节点的hadoop
   2. 如有必要，先 stop-all.sh, 再start-all.sh
2. 运行程序：
   1. cd /home/hadoop/resys 进入目录
   2. spark-submit resys.py 执行程序，，注意，，不是运行python resys.py
3. 程序描述：
   1. 从数据库中将rating表导出到本地/home/hadoop/resys/all\_rating.csv 中
   2. 启动并初始化spark
   3. 读取csv文件到hdfs中，生成rdd对象，并对数据进行必要的格式化（以逗号分割，分别为user\_id, item\_id, rating）
   4. 使用协同过滤算法对数据进行分析和建模（spark中的mllib的als算法：使用交替最小二乘法的系统过滤）。
   5. 算法原理：
      1. user\_id, item\_id, rating 的数据用二位表表示是一个及其稀疏的矩阵：
      2. 通过矩阵分解（MF），可分解为两个纬度较小的因子矩阵，因为这种矩阵算出的是矩阵内在的行为的结构特征，并不能被直接解释，所以叫隐性特征（LMF），但它包含了某些方面的偏好因子。如下图表示：
      3. 通过U-K矩阵或者I-K矩阵可以很容易的计算出用户相似度和物品相似度，对于推荐的求解也相对容易。
      4. 在我们的推荐系统中，我们取得维度k=50，迭代次数为10次，学习收敛速率为0.01，相对来说能够使结果更加精确，但是k越大，迭代次数越多计算的时间成本和空间成本就会越大，在本次对于模型的训练，花了大概40分钟（感觉时间也太久了，，，，）