

使用分页查询来适应挖掘海量数据之概论 PDF转换可能丢失图片或格式，建议阅读原文

[https://www.100test.com/kao\\_ti2020/253/2021\\_2022\\_\\_E4\\_BD\\_BF\\_E7\\_94\\_A8\\_E5\\_88\\_86\\_E9\\_c102\\_253884.htm](https://www.100test.com/kao_ti2020/253/2021_2022__E4_BD_BF_E7_94_A8_E5_88_86_E9_c102_253884.htm) 数据挖掘各类算法中

，常常需要遍历整个数据库（表）。现实中的数据库可能十分大，往往不可能通过一个简单的Select \*的方式遍历提取数据表内的所有元组。直接用Select \*的方式存在两大问题，一是Select \*过后，可能要等很久数据库才能将所有信息提交完毕，第二是得到的结果可能是很大，远远超过内存的限制。现在各种主流的数据库都支持了分页查询的方式。以Oracle为例，通过rownum关键字可以获取指定的行区间。比如：  
Select \* from XX.TABLE1 where rownum 而实际上，我们可以这样来写：SELECT \* from (0select n.\*,rownum num from XX.TABLE1 n WHERE rownum = 50. 以MySQL为例，提供了limit关键字，更加方便获取中间某区间的行数据。比如：  
Select \* from TABLE1 limit 50,100。MySQL的limit关键字用起来比Oracle要方便一些。不过各个数据库的分页查询的速度我倒没有研究过，网上听一些高手们说，Oracle提供的分页查询效率要高一些。Hibernate这样的数据持久层提供的分页查询，可以屏蔽掉各个不同的数据库之间具体SQL实现差异。Hibernate这样的数据持久层工具一大好处就是可以屏蔽掉不同数据库的之间的某些细节差异。分页查询在不同的数据库上实现的SQL不一样，所以要统一，最好使用Hibernate这样的工具。 Query q = session.createQuery("from Cat as c").  
q.setFirstResult(50). q.setMaxResults(100). List l = q.list(). 网上到处都可以搜寻到关于Hibernate内部实现分页查询的解读的文

章。可以看到，Hibernate内部也是通过rownum，limit这些关键字来实现分页的。100Test 下载频道开通，各类考试题目直接下载。详细请访问 [www.100test.com](http://www.100test.com)