

HBase 扫描器

主讲人： Louis



Scanner扫描器



HBase在扫描数据的时候，使用scanner表扫描器。

HTable通过一个Scan实例，调用getScanner(scan)来获取扫描器。可以配置扫描起止位，以及其他的过滤条件。

通过迭代器返回查询结果，使用起来虽然不是很方便，不过并不复杂。但是这里有一点可能被忽略的地方，就是返回的scanner迭代器，每次调用next的获取下一条记录的时候，默认配置下会访问一次RegionServer。这在网络不是很好的情况下，对性能的影响是很大的。

建议配置扫描器缓存。

扫描器缓存

`hbase.client.scanner.caching`配置项可以设置HBase scanner一次从服务端抓取的数据条数，默认情况下一次一条。通过将其设置成一个合理的值，可以减少scan过程中`next()`的时间开销，代价是scanner需要通过客户端的内存来维持这些被cache的行记录。

有三个地方可以进行配置：

- 1) 在HBase的`conf`配置文件中配置
- 2) 通过调用`HTable.setScannerCaching(int scannerCaching)`进行配置
- 3) 通过调用`Scan.setCaching(int caching)`进行配置。三者的优先级越来越高。