# Solr入门参考：

了解solr启动过程和原理—重点！

了解solrCloud启动过程和原理—重点！

Java –jar start.jar 运行jetty启动solr

<http://www.cnblogs.com/luxiaoxun/p/4442770.html> 配置启动运行

<http://blog.sina.com.cn/s/blog_8d955f8c0102uyaw.html> solr 的启动过程研究

<http://www.cnblogs.com/likehua/p/4353608.html> [Solr初始化源码分析-Solr初始化与启动](http://www.cnblogs.com/likehua/p/4353608.html)

<http://blog.csdn.net/yangbutao/article/details/11176469> [solrcloud集群启动管理过程基于源码的分析](http://blog.csdn.net/yangbutao/article/details/11176469)

<http://itindex.net/detail/50759-%E5%85%A8%E6%96%87%E6%A3%80%E7%B4%A2-%E5%BC%95%E6%93%8E-solr> 全文检索引擎Solr系列—–全文检索基本原理

# 用分布式搜索扩展索引大小

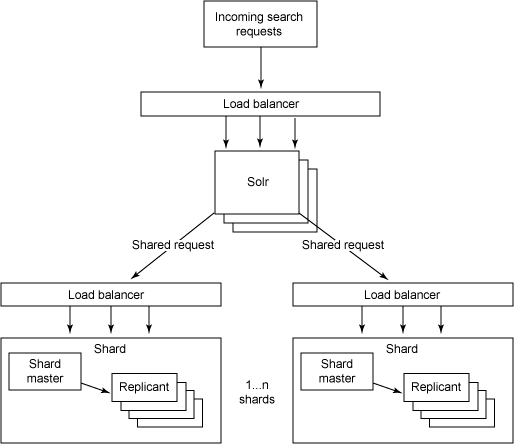
直到 1.3 版本，Solr 才能通过复制轻松进行扩展，以满足更大容量的查询需求。但是，如果没有应用程序帮助完成大部分工作，要提供超出单个机器的承载额度的索引还是很困难的。例如，通常可以在 Solr 中设置多个服务器，其中每一个服务器都有自己的索引，然后再让应用程序来管理搜索 —但这需要大量的自定义代码。在 1.3 版本中，Solr 添加了分布式搜索功能。应用程序将文档分布到几个计算机上，Solr（和其他程序）通常称之为 片（shard）。每一个片都包含自己的独立索引，而且 Solr 能够跨片协调索引查询。不幸的是，应用程序仍然需要将要创建索引的文档发送到每一个片，但这可能会添加到将来的 Solr 版本中。同时，可以使用一个简单的散列函数根据文档的唯一 ID 确定将文档发送到什么片。与此同时，我将关注搜索的等式方面。

Solr 机器容量

一个机器可以容纳的索引大小取决于机器的配置（RAM、CPU、磁盘等）、查询和索引的量，以及文档的大小和搜索模式。但是通常一个机器能够容纳的文档数量约在几百万到 1 亿之间。

要开始使用分布式搜索，用户需要花些时间考虑架构。如果仅需要几个片，而且不考虑复制的话，那么可以在每个机器上放置一个片，并且每一个片都能够创建索引和提供搜索。但如果索引和查询量很大的话，就必须复制每一个片。设置这种系统的常用的方法就是将每一个片及其复制放到一个载入平衡器的后面。图 2 展示了这个架构：

图 2. 分布式和复制 Solr 架构



分布式搜索的问题

Solr 的分布式搜索有几个缺点。第一，master 节点不容错，因此它一旦停止，系统就不能为新的文档创建索引或执行复制了。这并不会阻止搜索，而且对于可以手动或通过脚本和外部监控工具来管理片的较小分布式设置而言，这并不是个问题。换句话说，您不会使用当前的架构来构建下一个 Google，但一定能够提供一个大索引。第二，并不是所有的 1.3 版本中的SearchComponent都是分布式感知的。搜索、分类、调试和突出显示组件是分布是感知的；而其他的不常用的组件，正在努力实现这个功能。在分布式搜索的 wiki 链接中，可以看到一些小警告和所有细节信息（参见 [参考资料](http://www.ibm.com/developerworks/cn/java/j-solr-update/#resources)）。

注意，图 2 中输入的请求可以进入任何一个复制的片中，因为它们是功能齐全的 Solr 实例。然后，检索节点会将请求发送到其他片。这些请求仅仅是普通的 Solr 请求。要将请求提交到 Solr 服务器并分发请求，需要将 shards参数添加到请求，比如：

http://localhost:8983/solr/select?

shards=localhost:8983/solr,localhost:7574/solr&q=ipod+solr

在这个例子中，我假定在本地主机上运行了两个 Solr 服务器（它不是真正的分布式的；它适合于这里的论述，但不能用于您的设置），主服务器在端口 8983 上，从服务器在端口 7574 上。输入的请求进入端口 8983 上的实例，然后它将请求发送到片式服务器上。应用程序很可能将 shards参数值设置成 solrconfig.xml 文件中的 SolrRequestHandler的默认配置的一部分，这样就不需要在每次查询时都传入所有片式服务器的名称了。

# 常用命令

Suggest:

<http://localhost:1888/solr/new_core/suggest?spellcheck.build=true&q=c>