Apache Solr 指南

## 介绍

Apache Solr是一种开放源码的、基于 Lucene Java 的搜索服务器，易于加入到 Web 应用程序中。Solr 提供了层面搜索、命中醒目显示并且支持多种输出格式（包括 XML/XSLT 和 JSON 格式）。它易于安装和配置，而且附带了一个基于 HTTP 的管理界面。您可以坚持使用 Solr 的表现优异的基本搜索功能，也可以对它进行扩展从而满足企业的需要。Solr 还拥有一个活跃的开发者群体，如有需要，您可以随时向他们寻求帮助。

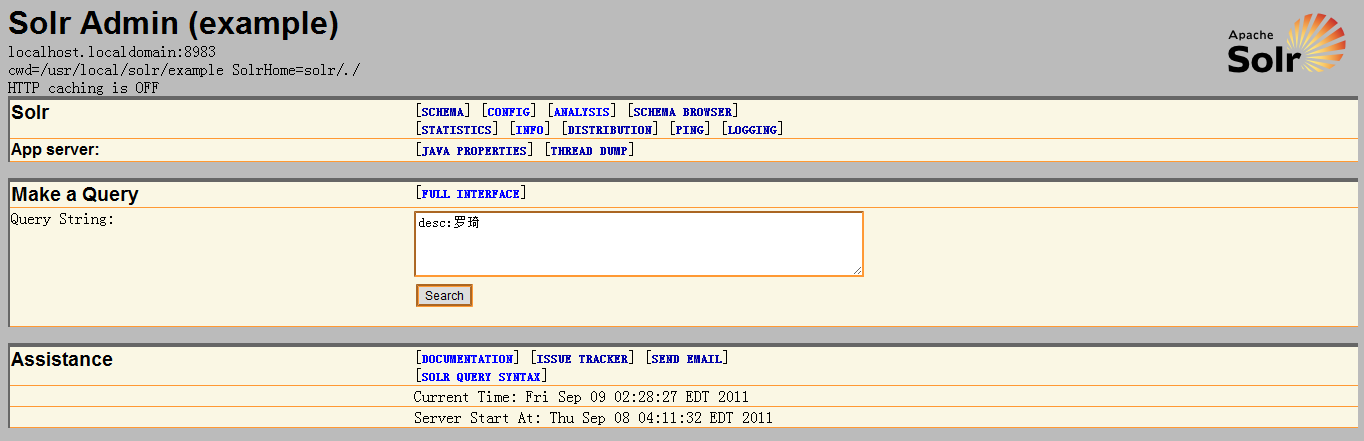
## 安装

这里我们以 基于 centOS 5.x 来演示整个安装过程

安装Java 6  
yum install java-1.6.0-openjdk.i386

下载 Apache Solr  
<http://www.apache.org/dyn/closer.cgi/lucene/solr/>  
在这里选择适当的版本进行安装

wget <http://apache.etoak.com/lucene/solr/3.3.0/apache-solr-3.3.0.zip>  
unzip apache-solr-3.3.0.zip  
mv apache-solr-3.3.0 /usr/local/solr

启动solr  
一种方式是通过servlet 容器进行启动， 当然这里我们介绍一种最简单 快速的方法 只需要运行下面命令即可  
cd /usr/local/solr/example  
java –jar start.jar  
会默认打开8983 端口  
  
这时候打开浏览器 输入  <http://localhost:8983/solr/admin/> 便会看到如下的 控制台界面  


## 配置

Solr 的配置是位于 /usr/local/solr/conf 中得 schema.xml.   
配置大概可以分为三部分：  
[类型] [字段] [其他声明]

类型主要是 描述字段的类型 每一种类型都可以配置很多Analyzer 具体用法 我也不太清  
  
字段主要是 描述有哪些字段需要进行索引, 每个字段是什么样的类型  
  
其他声明( 暂时没研究 )  
  
当然还有俩个比较重要的配置 一个是   
<uniqueKey>id</uniqueKey>   
主要表明建立索引的时候 以哪个字段为主键  
<defaultSearchField>desc</defaultSearchField>  
这个表明搜索的时候 默认以哪个字段进行全文搜索

## 索引

建立索引  
由于solr 提供了HTTP 的访问接口 所以在建立索引的时候 我们可以通过HTTP POST 的方式进行索引的建立，当然solr 也有一个 post.jar 可以方便的让我们来进行索引的建立。 举例说明 一个rocky.xml 文档  
  
<add>

<doc>

<field name="id">1001</field>

<field name="name">Rocky Article</field>

<field name="desc">Rocky is a very good Guy, and we do a wonderful work.</field>

<field name="date\_add">2006-01-17T00:00:00.000Z</field>

</doc>

<doc>

<field name="id">1002</field>

<field name="name">Zach Article</field>

<field name="desc">Zach is a vest Guy, test.</field>

<field name="date\_add">2006-01-17T00:00:00.000Z</field>

</doc>

</add>

然后将这个xml 文档post 到 solr 进行索引的建立  
java -jar post.jar rocky.xml

删除索引：  
java -Ddata=args -jar post.jar "<delete><query>id:1001</query></delete>"

## 搜索

进入<http://localhost:8983/solr/admin/form.jsp> 进行高级搜索  
因为solr 搜索核心采用的是 java Lucene 所以搜索的语法 和 Lucene 一样 这里我只举例一些简单的搜索方式 复杂的搜索方式可以访问  
<http://lucene.apache.org/java/2_9_1/queryparsersyntax.html>  
  
以rocky.xml 为案例 如果搜索内容为:

1. Rocky

这说明对默认字段进行搜索 相当于 desc:Rocky

1. name:rocky desc:rocky

对俩个字段进行搜索, 相当于 name:rocky OR desc:rocky

1. name:rocky AND desc:rocky

表明俩个字段都必须含有 rocky 才行

## Linux 下solr 的自动启动脚本

Vim /etc/init.d/solr

#!/bin/sh

SOLR\_DIR="/usr/local/solr/example"

JAVA\_OPTIONS="-Xmx256m -DSTOP.PORT=8079 -DSTOP.KEY=stopkey -jar start.jar"

LOG\_FILE="/var/log/solr.log"

JAVA="/usr/bin/java"

case $1 in

start)

echo "Starting Solr"

cd $SOLR\_DIR

$JAVA $JAVA\_OPTIONS 2> $LOG\_FILE &

;;

stop)

echo "Stopping Solr"

cd $SOLR\_DIR

$JAVA $JAVA\_OPTIONS --stop

;;

restart)

$0 stop

sleep 1

$0 start

;;

\*)

echo "Usage: $0 {start|stop|restart}" >&2

exit 1

;;

Esac

Chmod +x solr

然后把 在 /etc/rc.local 里加入  
/etc/init.d/solr start

## 参考文献

<http://www.ibm.com/developerworks/cn/java/j-solr1/index.html>  
<http://blog.chenlb.com/2009/05/apache-solr-quick-start-and-demo.html>  
<http://drupalsh.cn/node/2358>  
<http://www.blogjava.net/RongHao/archive/2007/11/06/158621.html>  
<http://lucene.apache.org/solr/#intro>