**R语言与HANA建模**

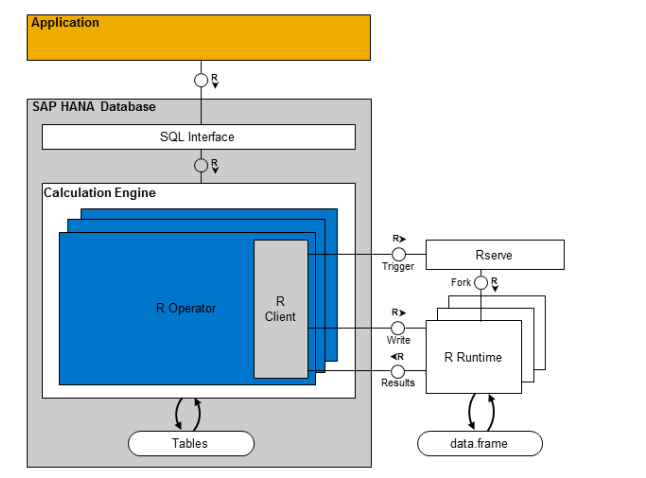
**（一）HANA与Rserve的连接及R建模示例**

**by Hujue**

1. 概要

R语言主要用于统计分析、数据挖掘，有着丰富的统计计算包资源和挖掘算法。在HANA SPS5开始，HANA结合了R语言，在存储过程中可以使用R代码，调用R中的统计函数。

Rserve与HANA之间的交互如下：



如图中可知，在hana计算引擎中整合了R Client，在存储过程PROC（或是计算视图CV）准备执行前，如果计算引擎识别出了R操作符，便由R client发出一个请求到Rserve，并发送相关的参数，触发执行，在执行完毕之后，将结果数据发送到计算引擎。

Rserve是R的一个开发包，用于实现远程连接R进行调用函数。

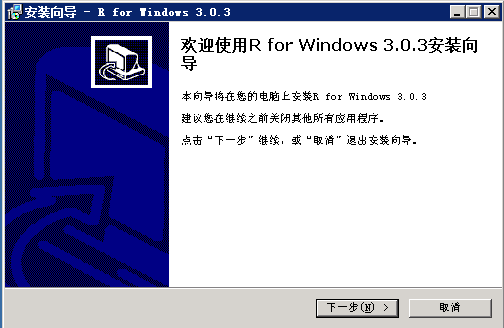
1. 安装R、Rserve与配置

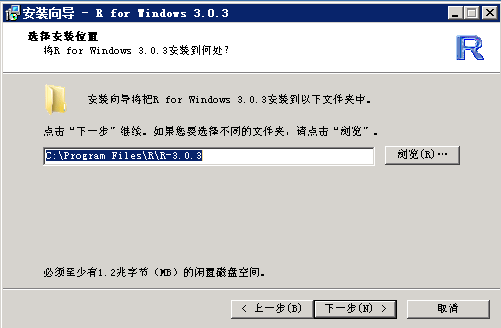
R及Rserve可以安装在windows平台，也可以安装在linux平台上，因为linux平台安装较为复杂，以下是分别以windows平台和linux平台安装为例。

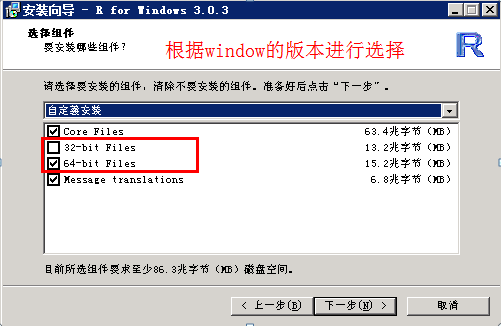
* 1. Windows平台安装R及Rserve
     1. 下载

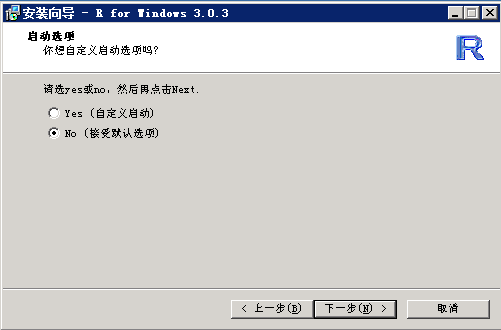
下载链接：<http://cran.r-project.org/bin/windows/base/R-3.1.0-win.exe>

* + 1. 安装

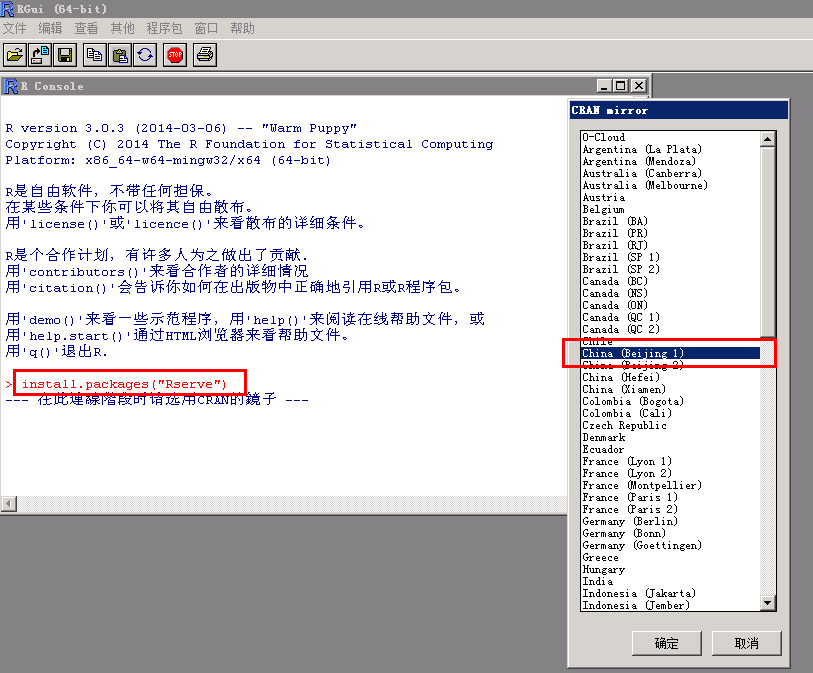








安装R语言完成后，打开R Console，输入命令 install.packages("Rserve")，然后选择下载镜像，如图示



出现如下图所示，说明Rserve安装成功



* + 1. 启动Rserve

在CMD中，定位到%R\_HOME%\library\Rserve\libs\x64，

输入命令

|  |
| --- |
| R CMD Rserve --RS-port 30120 --no-save --RS-encoding utf8 --RS-enable-remote |



**--RS-port 30120** 是定义Rserve对外listen的端口

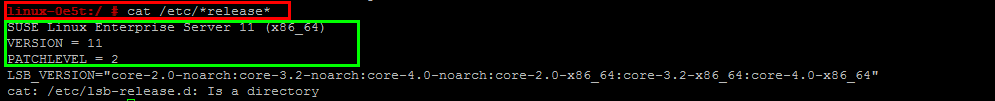
**--RS-encoding utf8** 定义传输数据的字符编码

**--RS-enable-remote**  是开放远程连接，可以让远程主机进行连接（非常重要）

* 1. Linux平台安装R及Rserve
     1. 确认安装环境

首先确认Linux平台是哪种，以及版本，输入以下命令

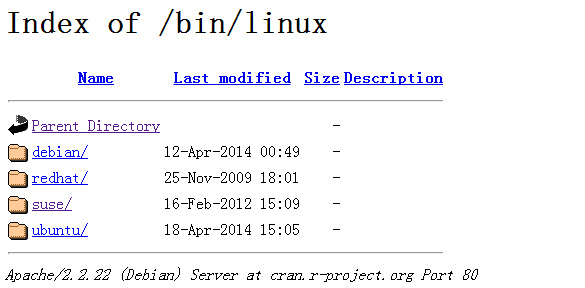
|  |
| --- |
| cat /etc/\*release\* |



如上图所知，此次安装环境为SUSE11企业版SP2，即SLE\_11\_SP2（SUSE同时还有开源版本，为openSUSE）

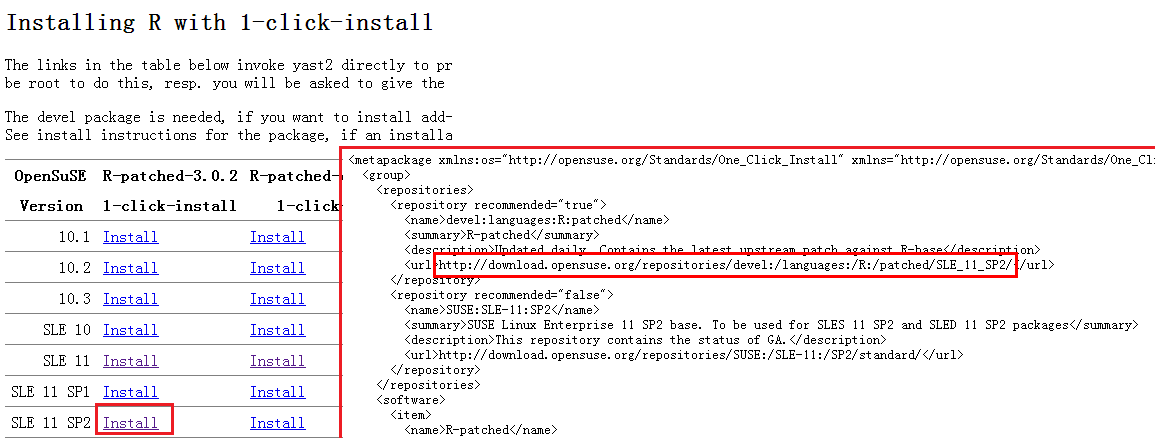
* + 1. 查找R软件源

访问网址http://cran.r-project.org/bin/linux/



查找到与自己安装环境相符的软件源

点击网页中的地址可以查询到软件源链接



以上可以得知，软件源链接为

|  |
| --- |
| http://download.opensuse.org/repositories/devel\:/languages\:/R\:/patched/SLE\_11\_SP2/ |

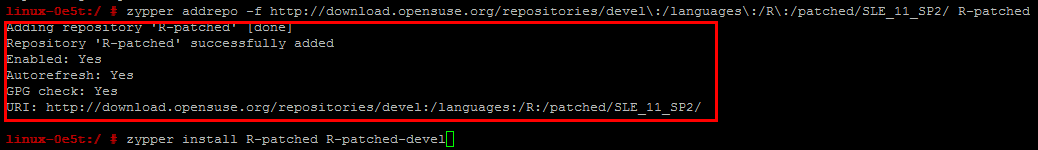
* + 1. 设置安装环境

以root用户登录服务器，在命令行中输入以下命令

|  |
| --- |
| zypper addrepo –f \  http://download.opensuse.org/repositories/devel\:/languages\:/R\:/patched/SLE\_11\_SP2/ \  R-patched |

（软件源链接中“：”前面的“\”为转义符，如果不添加转义符，会报错）

zypper是SUSE环境下安装软件的命令，以上操作是添加安装软件源的定义



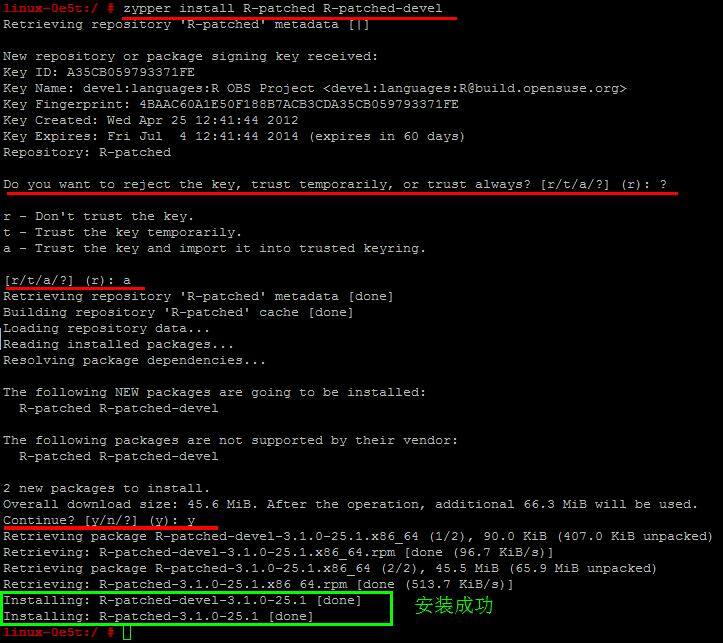
如图，添加软件安装源定义成功

* + 1. 安装R

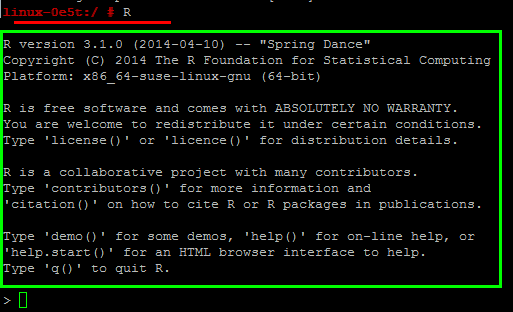
|  |
| --- |
| zypper install R-patched R-patched-devel |

在安装期间，会有R-patched的key让你确认，选择信任即可；

同时会自动下载关联的rpm包



在命令行中输入R，出现如下提示，则表明R安装成功



* + 1. 安装Rserve

在R命令行中输入以下命令

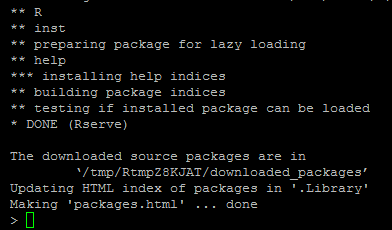
|  |
| --- |
| install.packages("Rserve") |

然后会出现下载镜像让你选择



按自己所在位置选择即可，这里选择了“20：Beijing 2”，然后安装自动开始

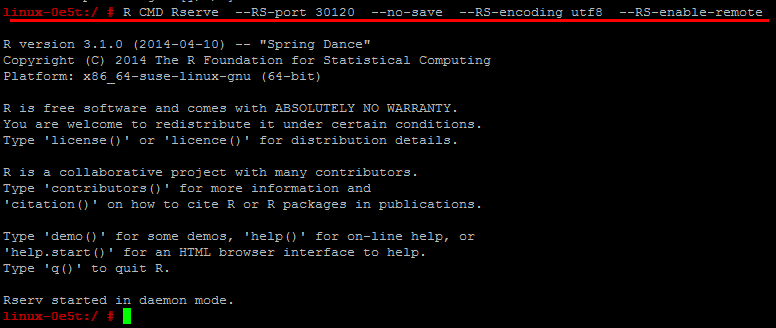
出现如下信息，说明Rserve安装成功



* + 1. 启动Rserve

退出R平台，在系统命令行输入以下命令

|  |
| --- |
| R CMD Rserve --RS-port 30120 --no-save --RS-encoding utf8 --RS-enable-remote |

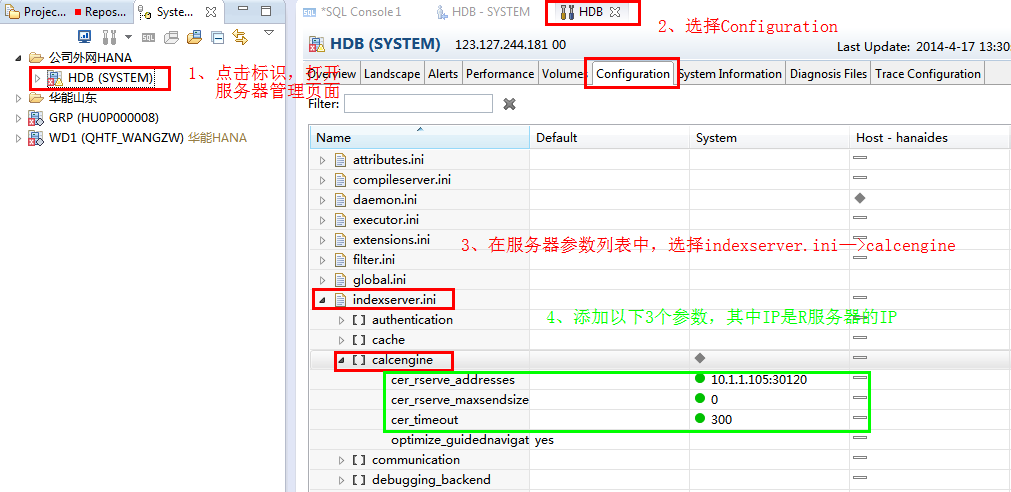


查看Rserve



如图所知，Rserve已经正常运行，监听端口为30120

* 1. 配置HANA



1. 示例及测试
   1. 示例及测试
      1. 代码准备

打开HANA studio，在sql console中输入以下代码

|  |
| --- |
| **CREATE** **ROW** **TABLE** "PRIME" ( "NUMBER" INT CS\_INT );  **insert** **into** "PRIME" **values**(2);  **insert** **into** "PRIME" **values**(3);  **insert** **into** "PRIME" **values**(5);  **insert** **into** "PRIME" **values**(7);  **CREATE** **ROW** **TABLE** "PRIME\_SQR" ( "NUMBER" INT CS\_INT );  **CREATE** **PROCEDURE** MY\_F(**IN** input1 PRIME,**OUT** result PRIME\_SQR)  **LANGUAGE** RLANG **AS**  **BEGIN**        result<-**as**.data.frame(input1$NUMBER^2);  **names**(result)<-c("NUMBER");  **END**; |

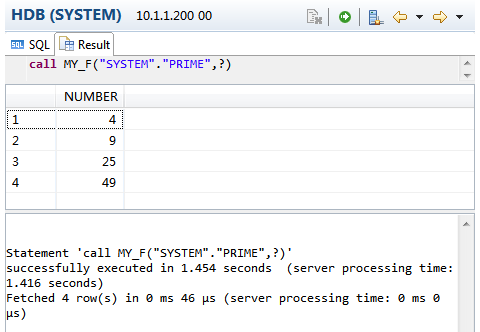
MY\_F是采用RLANG编写的SQLscript存储过程，在存储过程内，调用了R语言的函数，如**as**.data.frame ，“<-”是调用R语言函数的标识，当存储过程执行时识别到“<-”，就会远程连接到HANA configuration中定义的Rserve，然后通过Rserve调用R语言函数。

* + 1. 代码测试

MY\_F执行示例，在sqlconsole中输入以下代码

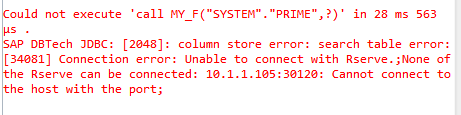
|  |
| --- |
| **call** MY\_F("SYSTEM"."PRIME",?); |

执行后结果为



* + 1. 常见错误说明
       1. 错误1：没有任何Rserve可连接

执行存储过程后，错误提示如下



原因：

启动Rserve时，未加上--RS-enable-remote参数，或是IP、端口不正确

* + 1. RLANG语法说明

暂缺

1. 特别说明
   1. R安装平台的问题

在《SAP HANA R Integration Guide》中提到R、Rserve必须安装在独立服务器上，不得安装在HANA服务器上，而且R最好安装在SLES Linux上，其他R宿主环境不被支持。

实际安装中证实，R安装在windows上也能连接，并且按原理来说，R安装在hana服务器上也是可行的；

Rserve推荐在Linux下使用，因为在Windows平台上还很不完善，Windows上运行Rserve还是有很大的局限性的。

最大的问题是，所有的连接都会在一个工作空间内运行。所以当客户端是多线程程序时，可能会造成线程间互相干扰。我们需要在客户端做一些额外的工作，使它能够以线程安全的方式运行。Java端可以通过使用synchronize使得同一时间只有一个线程使用R服务。

Rserve的另一个缺点是，它的错误处理过于简单化，如果R端出现错误，它抛出的异常对于找到问题的根源几乎没有任何用处。

其次，开发中需要注意的是，Rserve同时只允许一个客户端连接。因此，如果第二个线程试图连接时，它就会一直处在等待状态。另外，如果要更新R包，也需要关闭并重新启动Rserve。所以，如果发现客户端连接时出现异常情况，查看一下任务管理器，看是不是有多个Rserve的线程在运行，或者你的服务有没有重启，往往可以帮助解决问题。

1. 参考资料

1、[R语言包安装并实现与HANA的整合](http://scn.sap.com/community/chinese/hana/blog/2014/02/14/r%E8%AF%AD%E8%A8%80%E5%8C%85%E5%AE%89%E8%A3%85%E5%B9%B6%E5%AE%9E%E7%8E%B0%E4%B8%8Ehana%E7%9A%84%E6%95%B4%E5%90%88)

2、《SAP HANA R Integration Guide》，<http://help.sap.com/hana_platform>

3、Rserve与java的跨平台通信，<http://blog.fens.me/r-rserve-java/>

4、深入理解SAP HANA与R整合的原理（一），<http://scn.sap.com/community/chinese/hana/blog/2014/04/16/%E6%B7%B1%E5%85%A5%E7%90%86%E8%A7%A3sap-hana%E4%B8%8Er%E6%95%B4%E5%90%88%E7%9A%84%E5%8E%9F%E7%90%86%E4%B8%80>

5、在linux下安装R语言程序及bioconductor程序包，<http://blog.sina.com.cn/s/blog_700aa8830100s5np.html>

6、R语言服务器程序Rserve详解http://blog.fens.me/r-rserve-server/