尚硅谷大数据技术之Ranger

(作者：尚硅谷大数据研发部)

版本：V1.0

第1章 Ranger概述

1.1 什么是Ranger

Apache Ranger是一个用来在Hadoop平台上进行监控，启用服务，以及全方位**数据安全访问管理**的安全框架。

Ranger的愿景是在Apache Hadoop生态系统中提供全面的安全管理。随着企业业务的拓展，企业可能在多用户环境中运行多个工作任务，这就要求Hadoop内的数据安全性需要扩展为同时支持多种不同的需求进行数据访问，同时还需要提供一个可以对安全策略进行集中管理，配置和监控用户访问的框架。Ranger由此产生！

Ranger的官网：https://ranger.apache.org/

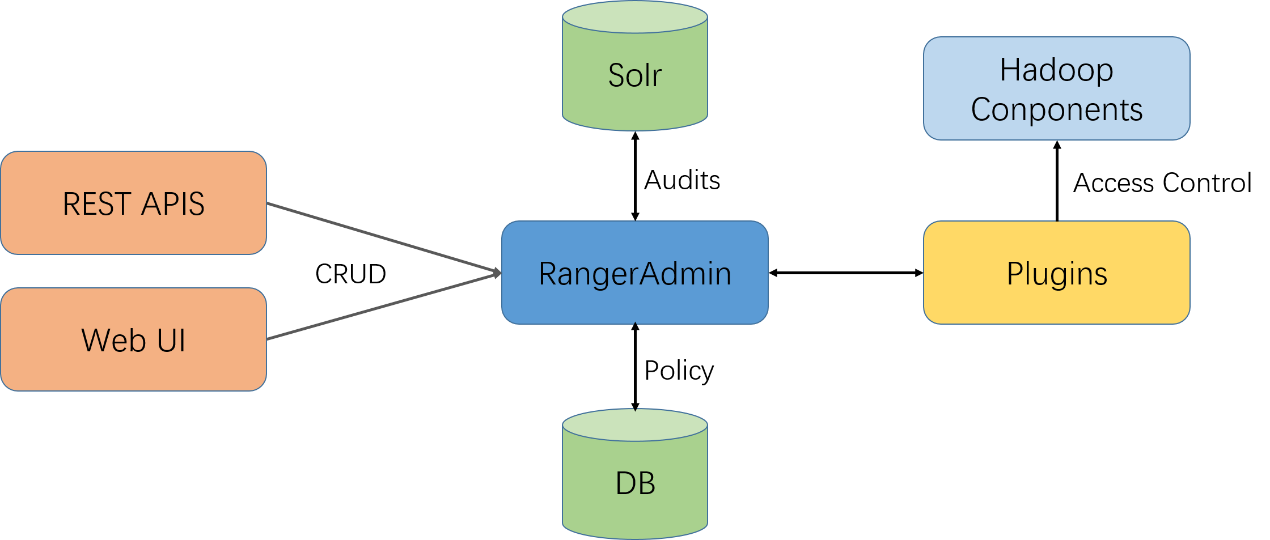
1.2 Ranger的目标

* 允许用户使用UI或REST API对所有和安全相关的任务进行集中化的管理
* 允许用户使用一个管理工具对操作Hadoop体系中的组件和工具的行为进行细粒度的授权
* 支持Hadoop体系中各个组件的授权认证标准
* 增强了对不同业务场景需求的授权方法支持，例如基于角色的授权或基于属性的授权
* 支持对Hadoop组件所有涉及安全的审计行为的集中化管理

1.3 Ranger支持的框架

* Apache Hadoop
* Apache Hive
* Apache HBase
* Apache Storm
* Apache Knox
* Apache Solr
* Apache Kafka
* YARN
* NIFI

1.4 Ranger的架构



1.5 Ranger的工作原理

Ranager的核心是web应用程序，也成为RangerAdmin模块，此模块由管理策略，审计日志和报告等三部分组成。

管理员角色的用户可以通过RangerAdmin提供的web界面或REST APIS来定制安全策略。这些策略会由Ranger提供的轻量级的针对不同Hadoop体系中组件的插件来执行。插件会在Hadoop的不同组件的核心进程启动后，启动对应的插件进程来进行安全管理！

第2章 Ranger的安装

2.1 环境准备

Ranger2.0要求对应的Hadoop为3.x以上，Hive为3.x以上版本，JDK为1.8以上版本！

2.2 安装RangerAdmin

2.2.1 数据库环境准备

在Mysql数据库中创建Ranger存储数据的数据库:

mysql> create database ranger;

创建用户:

mysql> grant all privileges on ranger.\* to ranger@'%' identified by 'ranger';

2.2.2 安装RangerAdmin

解压软件：

[atguigu@hadoop103 target]tar -zxvf ranger-2.0.0-admin.tar.gz -C /opt/module/ranger

进行配置：

[atguigu@hadoop103 ranger-2.0.0-admin]$ vim install.properties

修改以下配置内容：

#mysql驱动

SQL\_CONNECTOR\_JAR=/opt/software/mysql-connector-java-5.1.27-bin.jar

#mysql的主机名和root用户的用户名密码

db\_root\_user=root

db\_root\_password=000000

db\_host=hadoop103

#ranger需要的数据库名和用户信息，和2.2.1创建的信息要一一对应

db\_name=ranger

db\_user=ranger

db\_password=ranger

#其他ranger admin需要的用户密码

rangerAdmin\_password=atguigu123

rangerTagsync\_password=atguigu123

rangerUsersync\_password=atguigu123

keyadmin\_password=atguigu123

#ranger存储审计日志的路径，默认为solr，这里为了方便暂不设置

audit\_store=

#策略管理器的url,rangeradmin安装在哪台机器，主机名就为对应的主机名

policymgr\_external\_url=http://hadoop103:6080

#启动ranger admin进程的linux用户信息

unix\_user=atguigu

unix\_user\_pwd=atguigu

unix\_group=atguigu

#hadoop的配置文件目录

hadoop\_conf=/opt/module/hadoop-3.1.3/etc/hadoop

之后切换到root用户，执行安装

[root@hadoop103 ranger-2.0.0-admin]# ./setup.sh

出现以下信息，说明安装完成

2020-04-30 13:58:18,051 [I] Ranger all admins default password change request processed successfully..

Installation of Ranger PolicyManager Web Application is completed.

创建ranger的配置文件软连接到web应用下：

[root@hadoop103 ranger-2.0.0-admin]# ./set\_globals.sh

usermod：无改变

[2020/04/30 13:58:47]: [I] Soft linking /etc/ranger/admin/conf to ews/webapp/WEB-INF/classes/conf

2.2.3 启动RangerAdmin

配置RangerAdmin web应用的配置信息：

[root@hadoop103 ranger-2.0.0-admin]# cd /etc/ranger/admin/conf/

[root@hadoop103 conf]# vim ranger-admin-site.xml

<property>

<name>ranger.jpa.jdbc.password</name>

<value>ranger</value>

<description />

</property>

<property>

<name>ranger.service.host</name>

<value>hadoop103</value>

</property>

启动：

[root@hadoop103 conf]# ranger-admin start

Starting Apache Ranger Admin Service

Apache Ranger Admin Service with pid 7058 has started.

ranger-admin在安装时已经配设置为开机自启动，因此之后无需再手动启动！

查看启动后的进程：

[root@hadoop103 ranger-2.0.0-usersync]# jps

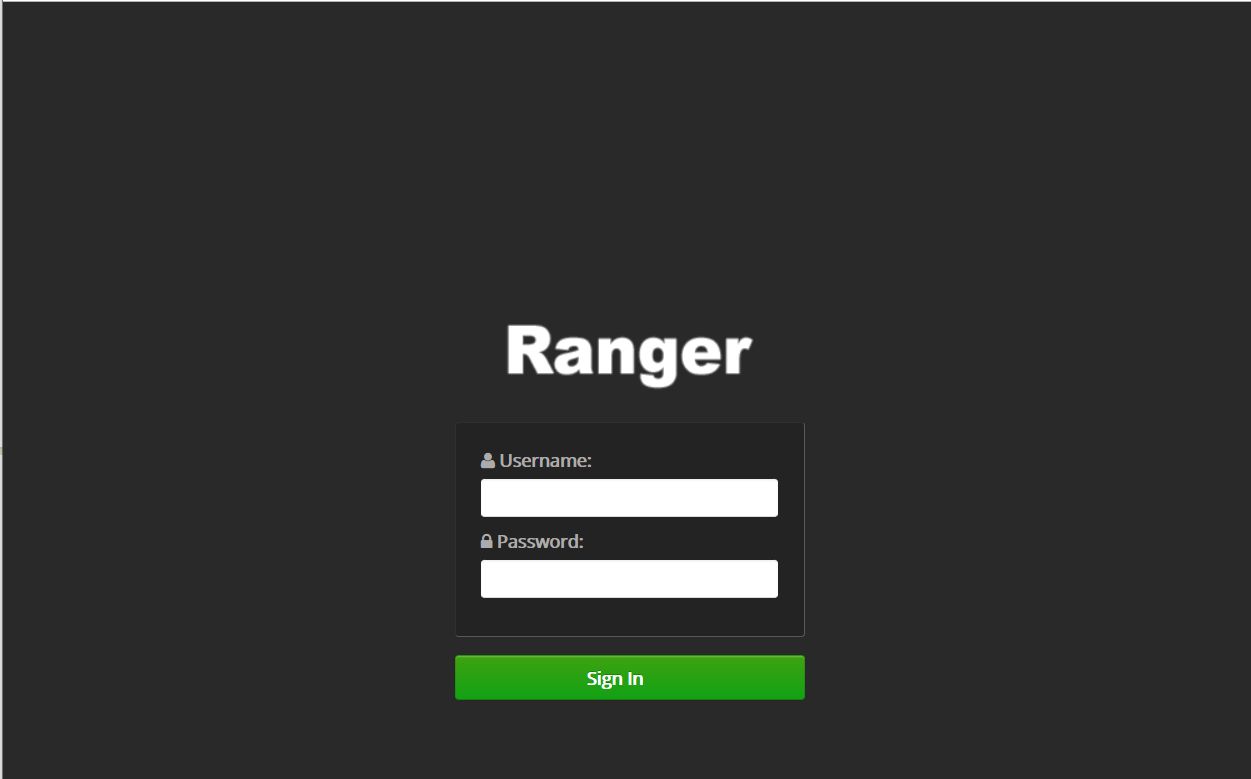
7058 EmbeddedServer

8132 Jps

停止：

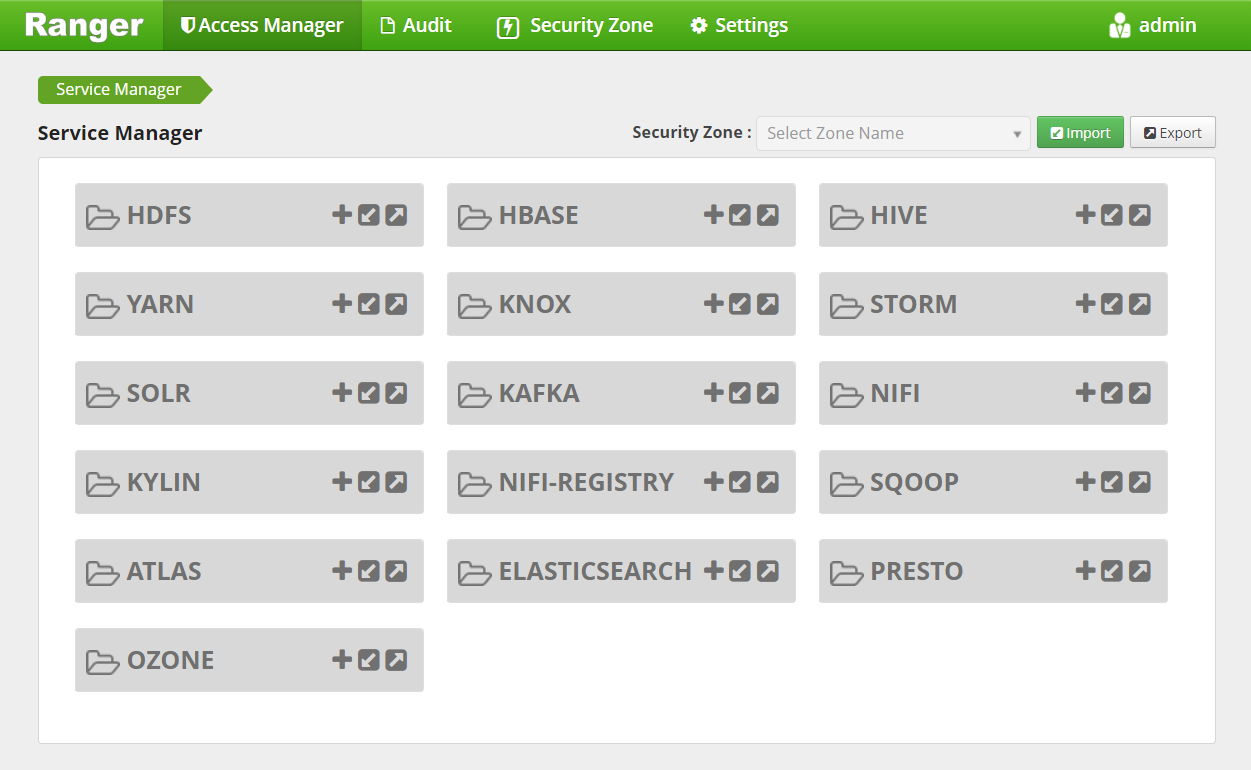
[root@hadoop103 conf]# ranger-admin stop

之后可以通过访问<http://hadoop103:6080>，如出现以下界面，说明ranger-admin启动完成！



2.2.4 登录管理员用户

默认可以使用用户名:admin,密码为之前配置的atguigu123进行登录！登录后界面如下：



第3章 安装 RangerUsersync

3.1 RangerUsersync简介

RangerUsersync作为Ranger提供的一个管理模块，可以将Linux机器上的用户和组信息同步到RangerAdmin的数据库中进行管理！

3.2 RangerUsersync安装

解压软件：

[root@hadoop103 conf]# tar -zxvf /opt/software/apache-ranger-2.0.0/target/ranger-2.0.0-usersync.tar.gz -C /opt/module/ranger/

配置软件：

[root@hadoop103 ranger-2.0.0-usersync]# vim install.properties

修改以下配置信息：

#rangeradmin的url

POLICY\_MGR\_URL =http://hadoop103:6080

#同步间隔时间，单位(分钟)

SYNC\_INTERVAL = 1

#运行此进程的linux用户

unix\_user=atguigu

unix\_group=atguigu

#rangerUserSync的用户密码，参考rangeradmin中install.properties的配置

rangerUsersync\_password=atguigu123

#hadoop的配置文件目录

hadoop\_conf=/opt/module/hadoop-3.1.3/etc/hadoop

使用root用户进行安装

[root@hadoop103 ranger-2.0.0-usersync]# ./setup.sh

出现以下信息，说明安装完成

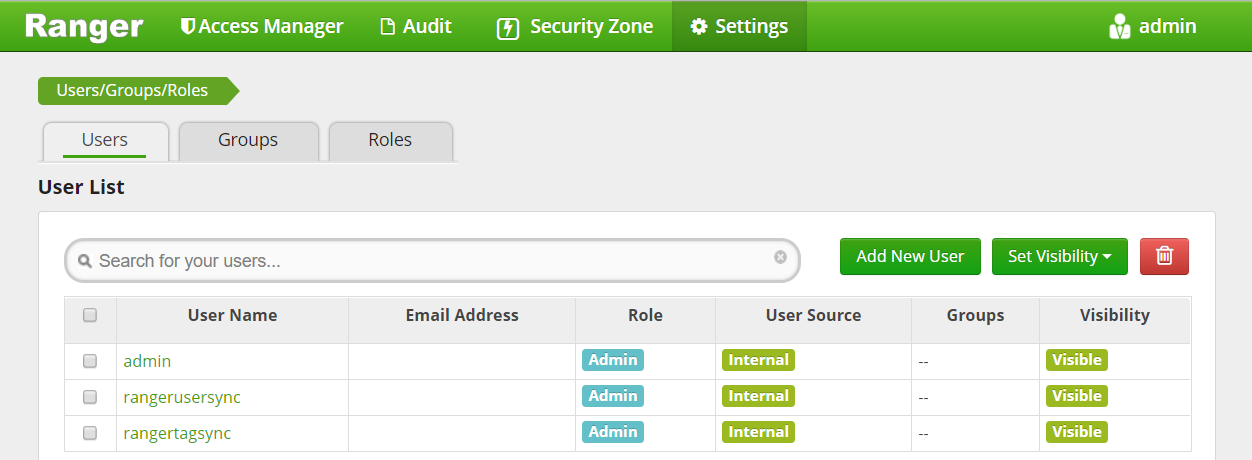
ranger.usersync.policymgr.password has been successfully created.

Provider jceks://file/etc/ranger/usersync/conf/rangerusersync.jceks was updated.

[I] Successfully updated password of rangerusersync user

3.3 RangerUsersync启动

启动之前，在ranger admin的web-UI界面，查看用户信息如下：



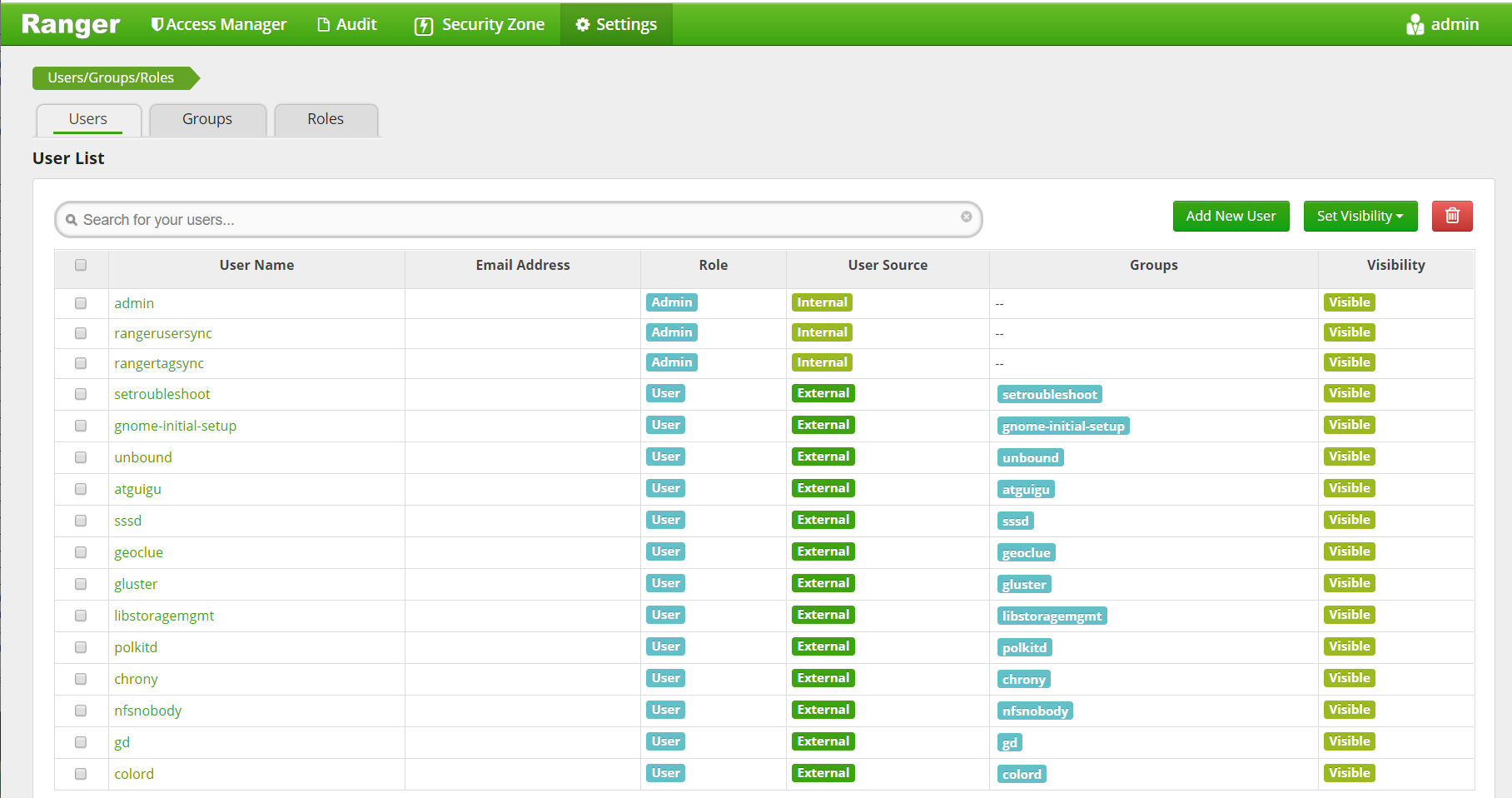
使用root用户启动

[root@hadoop103 ranger-2.0.0-usersync]# ranger-usersync start

Starting Apache Ranger Usersync Service

Apache Ranger Usersync Service with pid 7510 has started.

启动后，再次查看用户信息：



说明ranger-usersync工作正常！

ranger-usersync服务也是开机自启动的，因此之后不需要手动启动！

第4章 安装Ranger Hive-plugin

4.1 Ranger Hive-plugin简介

Ranger Hive-plugin是Ranger对hive进行权限管理的插件。Ranger Hive-plugin只能对使用jdbc方式访问hive的请求进行权限管理，hive-cli并不受限制！

4.2 Ranger Hive-plugin安装

解压软件：

[root@hadoop103 ranger-2.0.0-usersync]# tar -zxvf /opt/software/apache-ranger-2.0.0/target/ranger-2.0.0-hive-plugin.tar.gz -C /opt/module/ranger/

配置软件：

[root@hadoop103 ranger-2.0.0-hive-plugin]# vim install.properties

#策略管理器的url地址

POLICY\_MGR\_URL=http://hadoop103:6080

#组件名称可以自定义

REPOSITORY\_NAME=hivedev

#hive的安装目录

COMPONENT\_INSTALL\_DIR\_NAME=/opt/module/hive

#hive组件的启动用户

CUSTOM\_USER=atguigu

#hive组件启动用户所属组

CUSTOM\_GROUP=atguigu

将hive的配置文件作为软连接安装到Ranger Hive-plugin目录下：

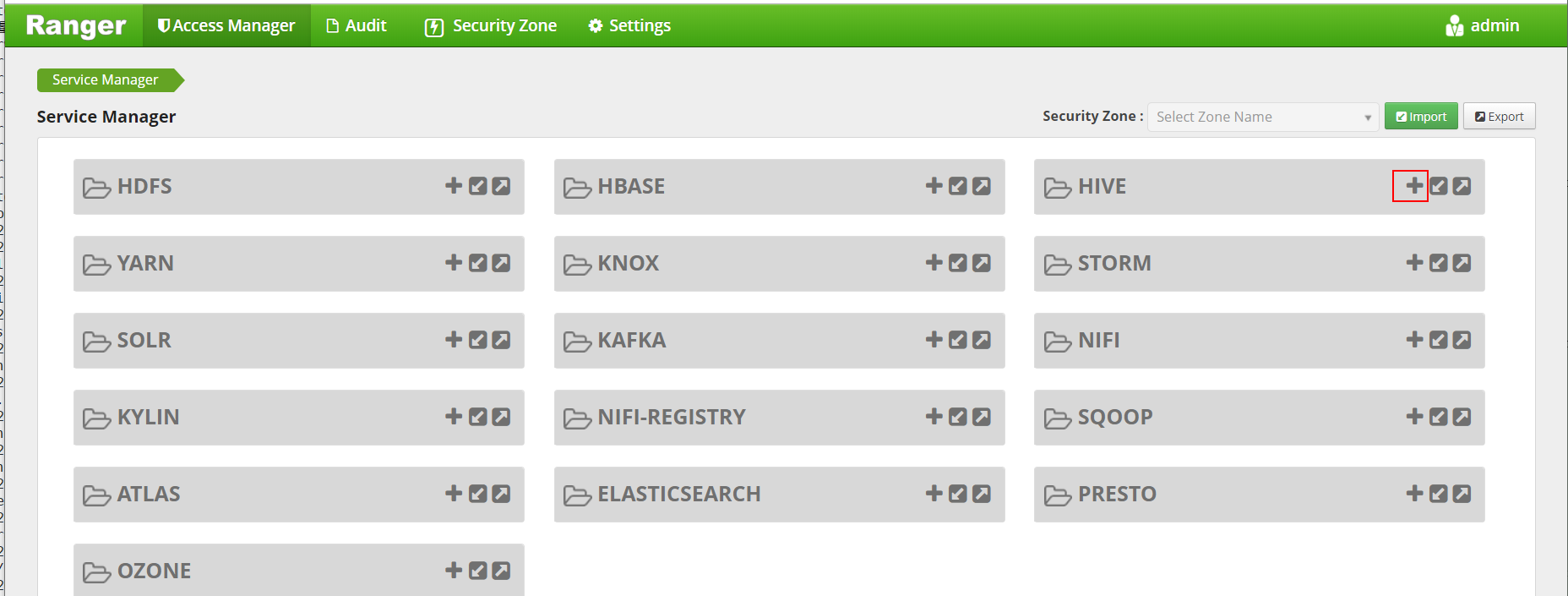
[root@hadoop103 ranger-2.0.0-hive-plugin]# ln -s /opt/module/hive/conf/ conf

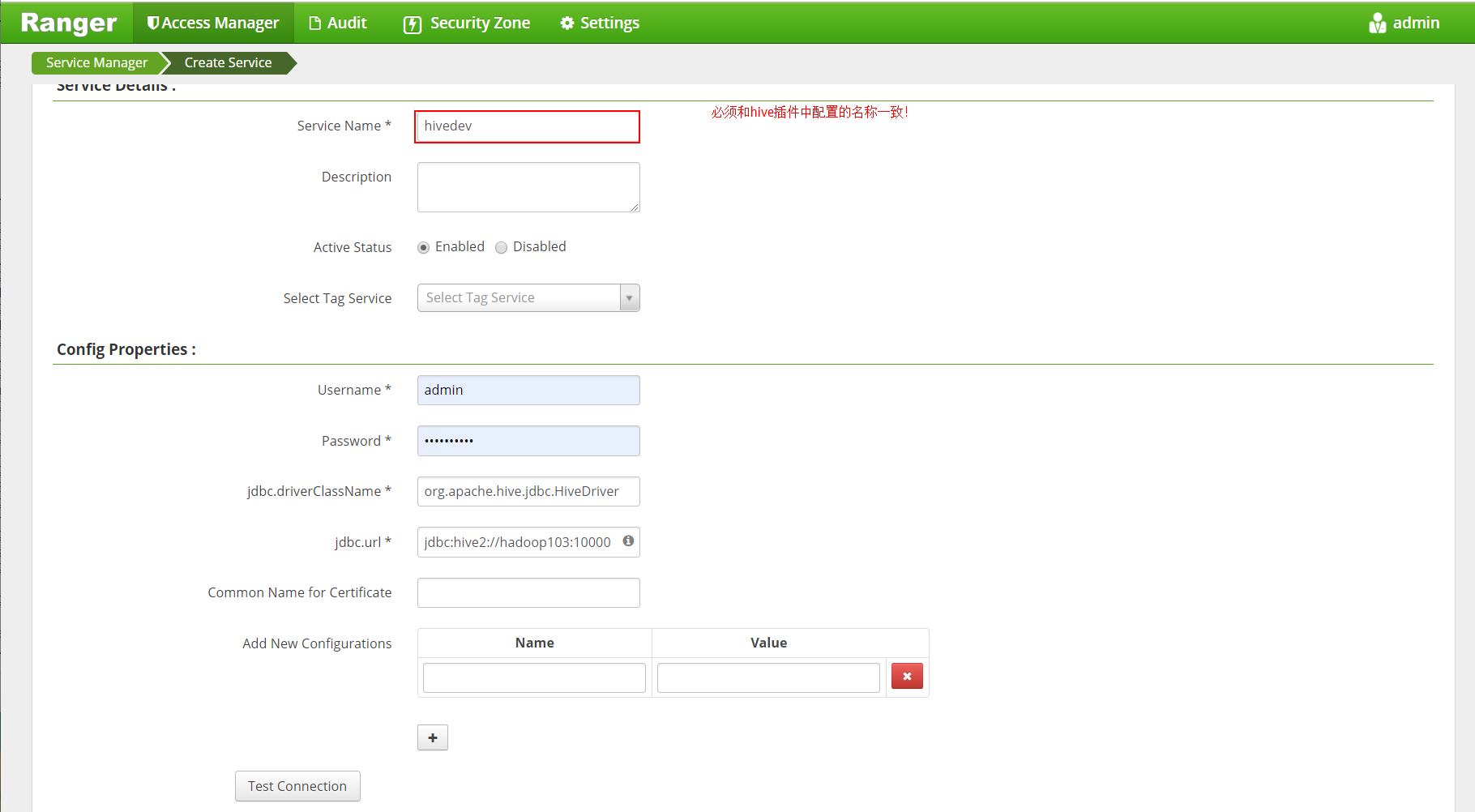
使用root用户启用Ranger Hive-plugin

[root@hadoop103 ranger-2.0.0-hive-plugin]# ./enable-hive-plugin.sh

之后需要重启hive才能生效！

4.3 在ranger admin上配置hive插件





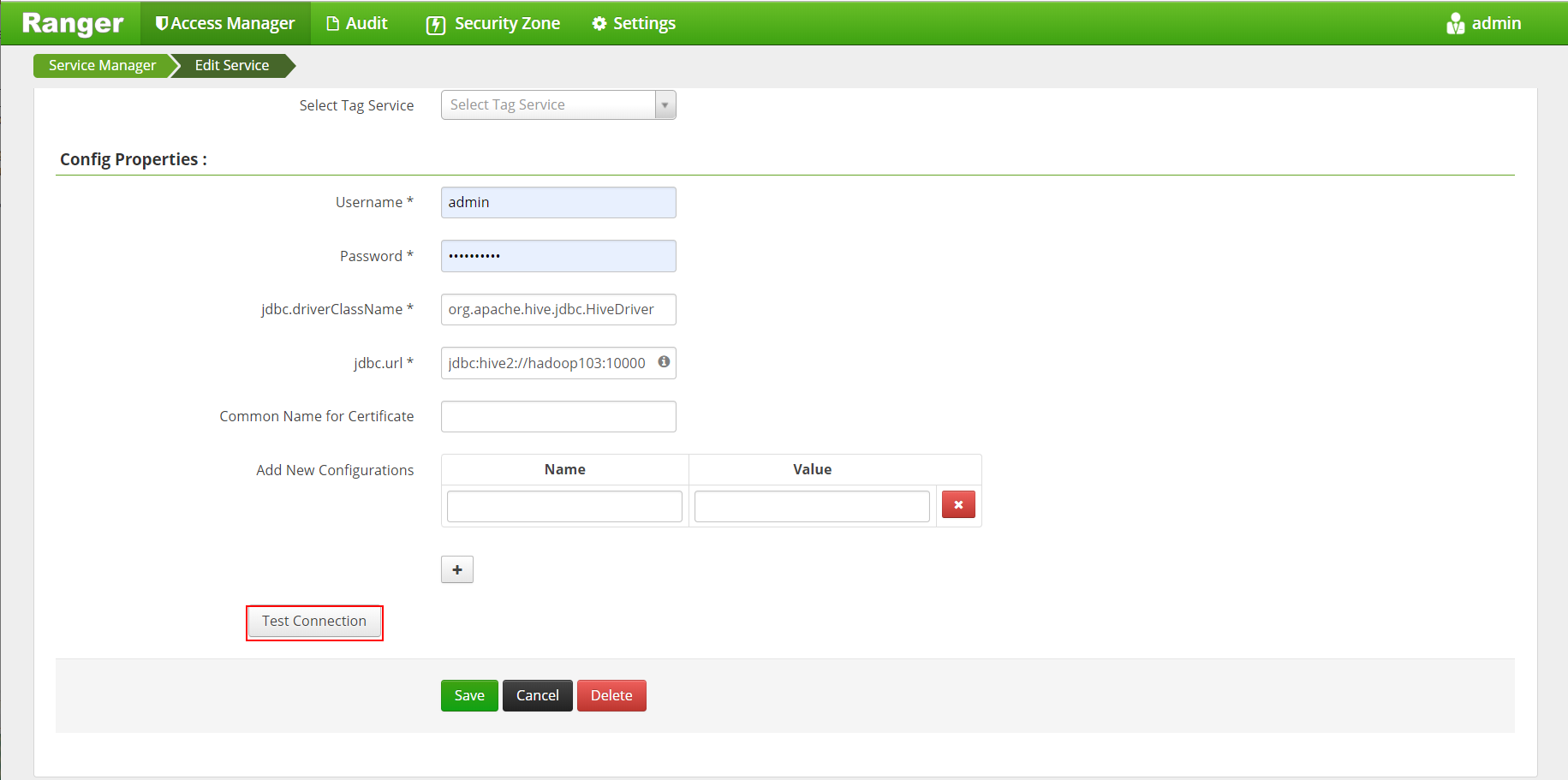
4.4 测试连接hiveserver2

启动hiveserver2：

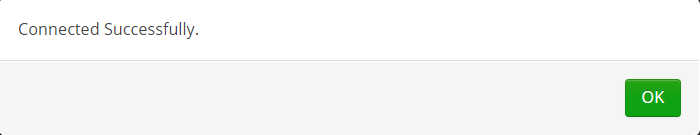
[atguigu@hadoop103 hive]$ hive --service metastore &

[atguigu@hadoop103 hive]$ hiveserver2

测试插件是否可以连接hiveserver2



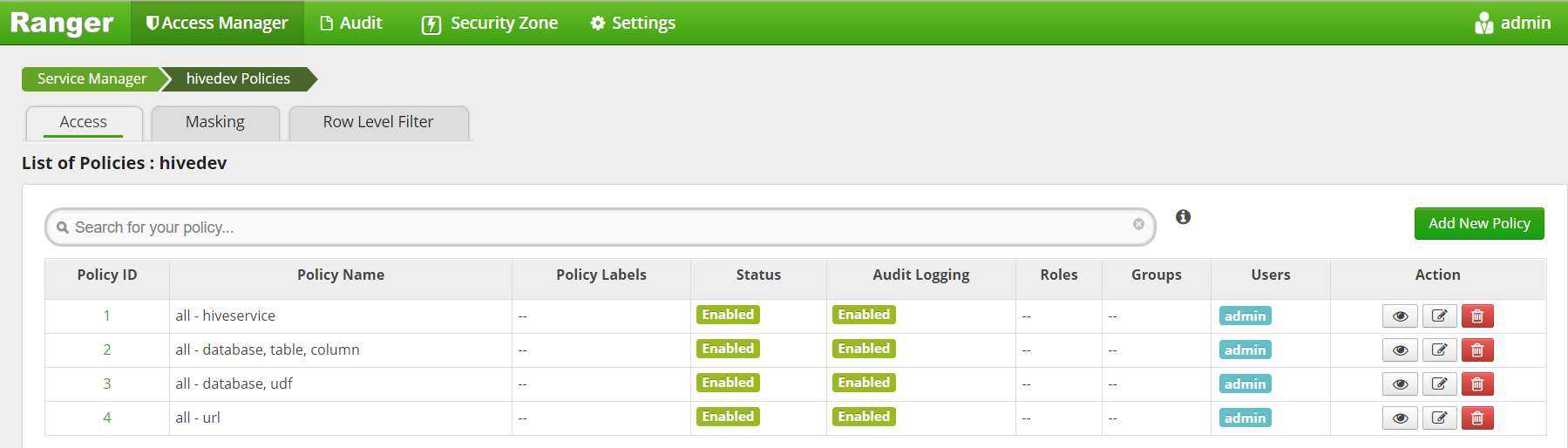
出现以下提示说明连接成功！



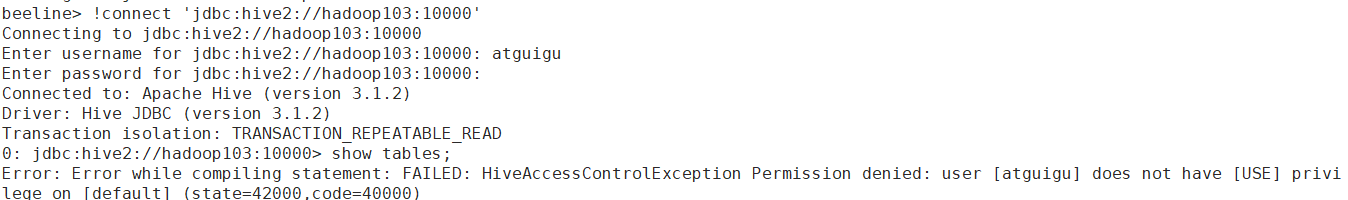
第5章 使用Ranger对Hive进行权限管理

5.1 权限控制初体验

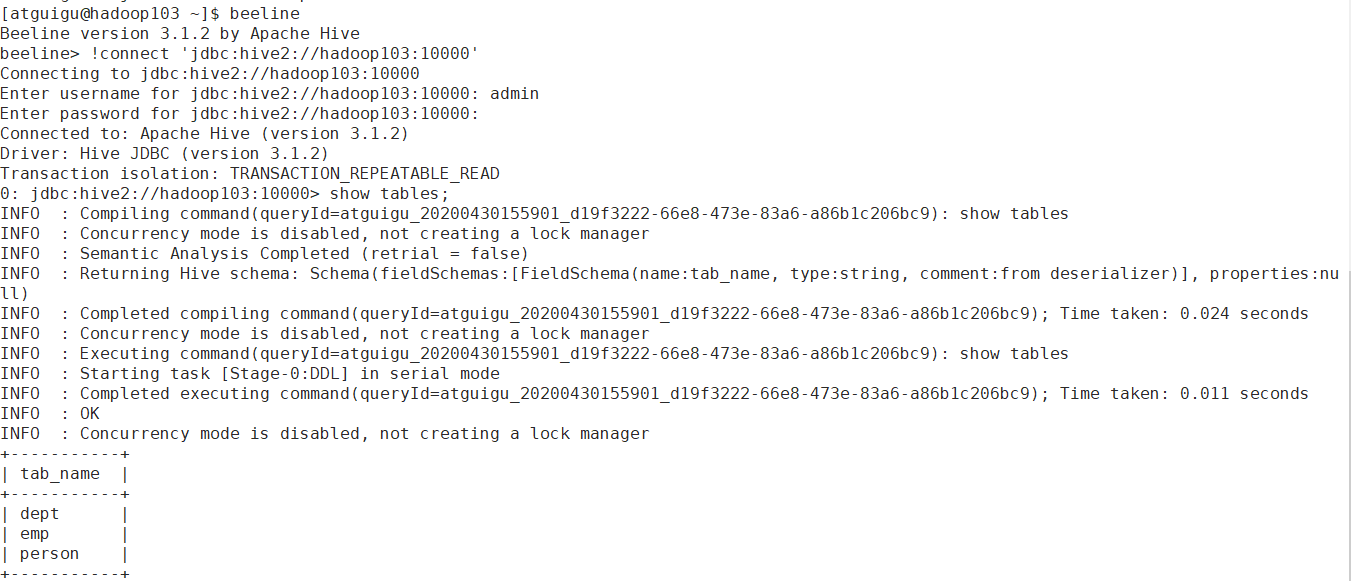
查看默认的访问策略，此时只有admin用户拥有对所有库、表和函数的访问权限



验证：使用atguigu用户尝试进行登录，登录成功后，执行查询语句：



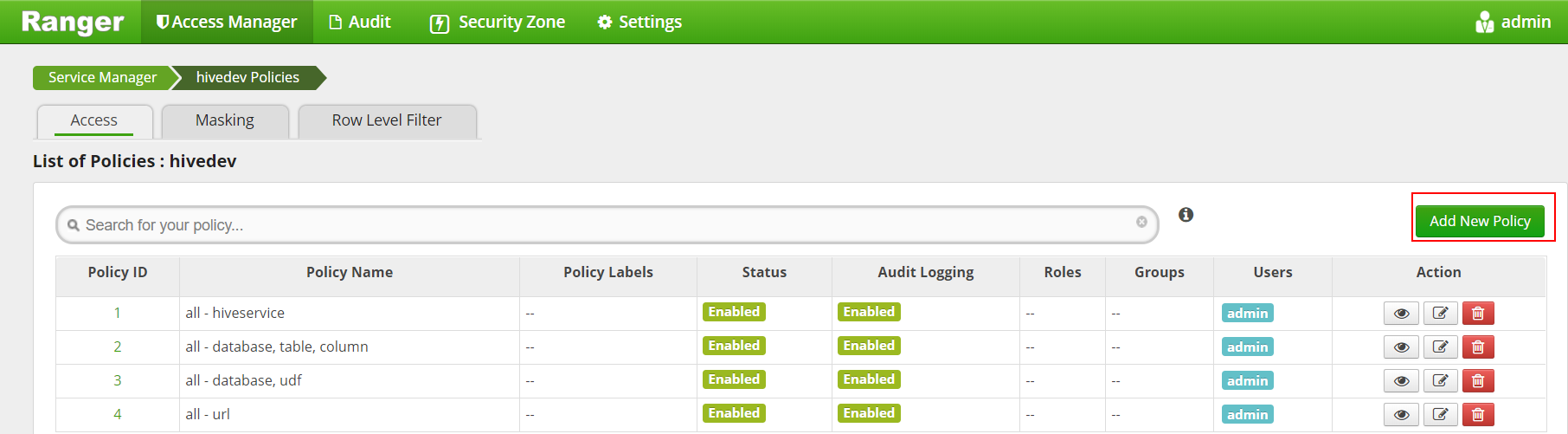
之后使用admin用户进行登录，可以完成Hive的所有操作



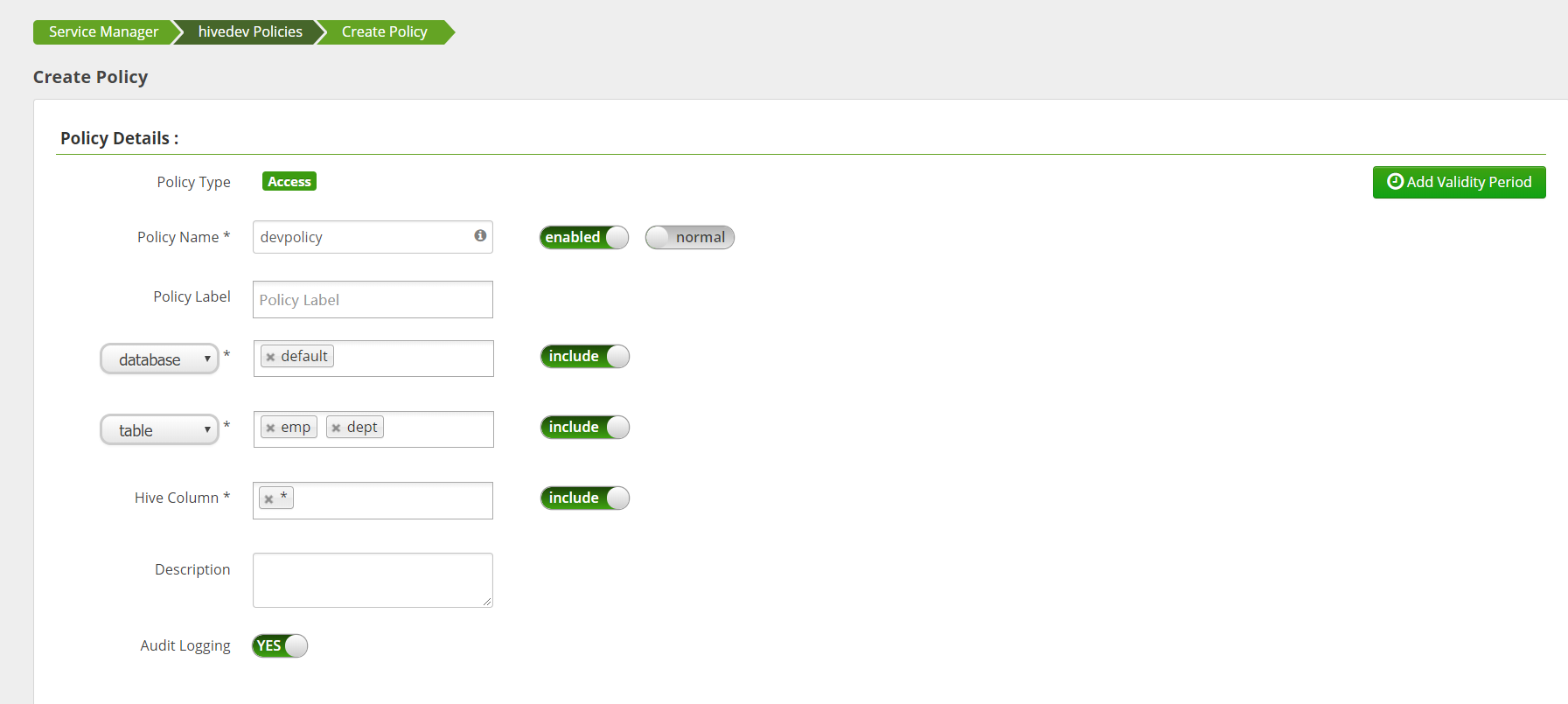
5.2 为用户配置权限

例如为atguigu用户配置default库emp和dept表的所有列的读权限，为jack用户配置default库emp和dept表的所有列的读写权限。

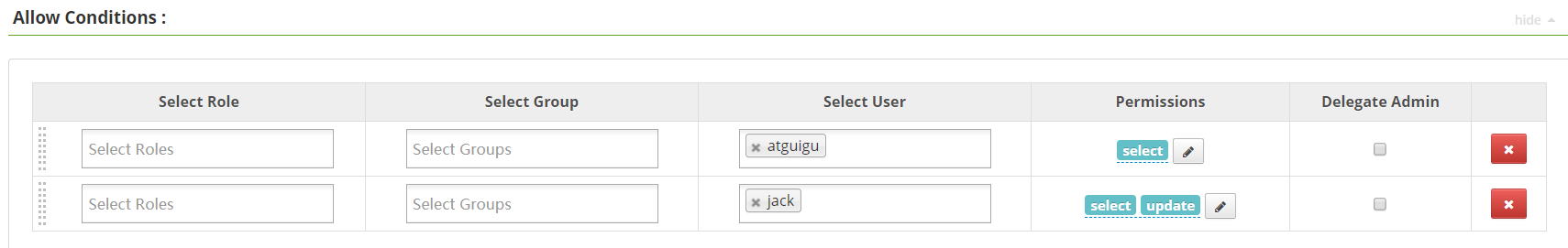
点击Add New Policy按钮：



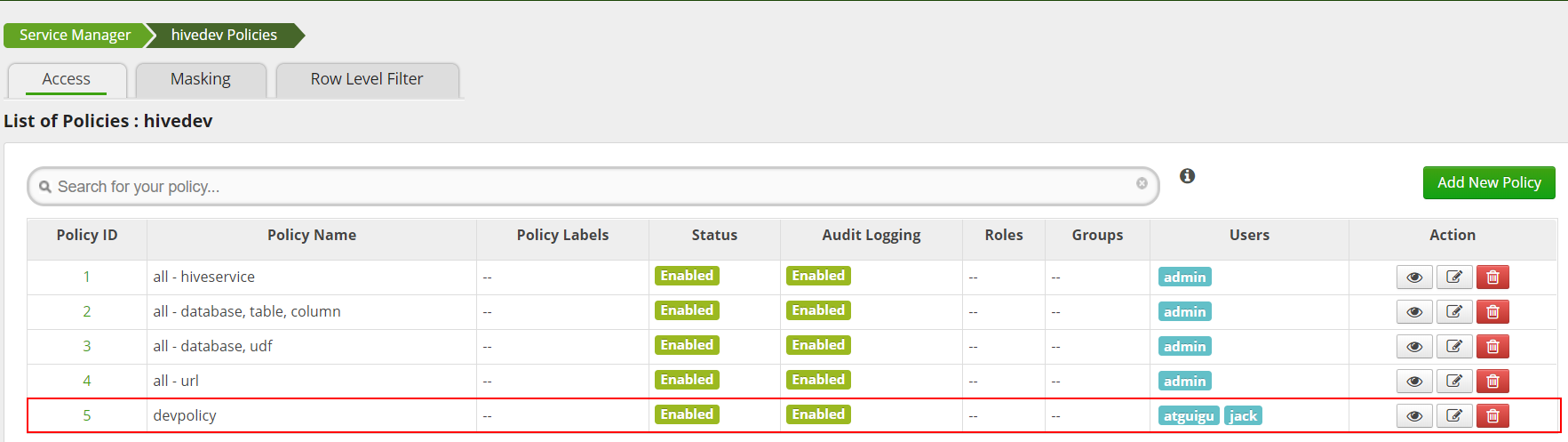
填写策略名称，以及此策略设计的库、表、列等信息：



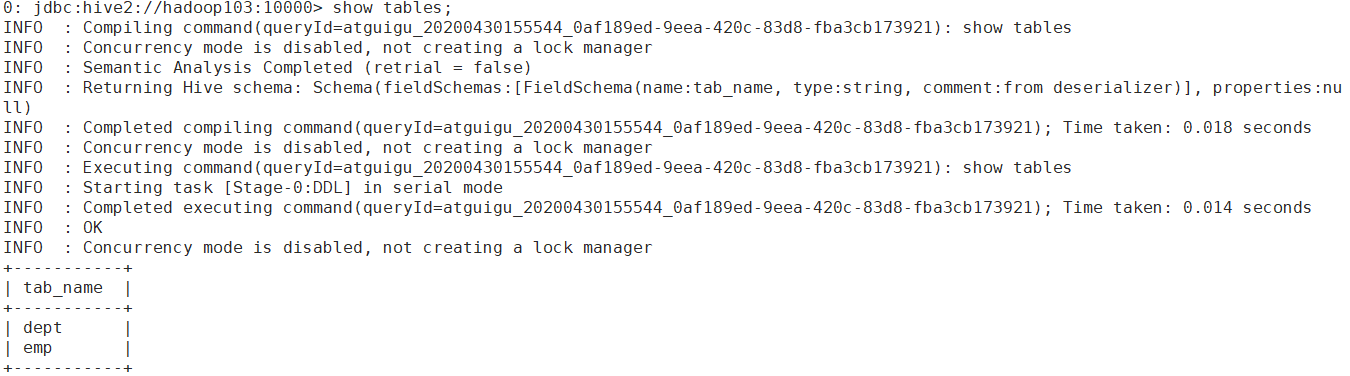
填写设计此策略的允许的用户权限：



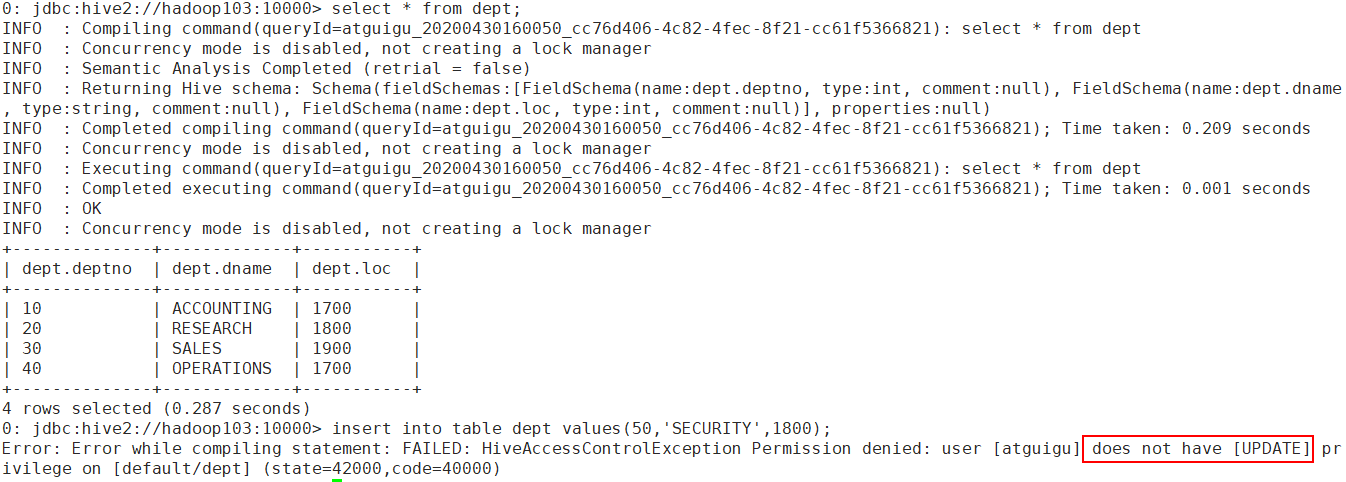
之后点击Add添加按钮，发现在面板上已经添加完成：



测试：beeline无需重新连接hiveserver2，再次执行查询，发现atguigu用户已经可以进行查询，但是只能查询自己有权限查询的表信息



对以下两个表，有读权限，没有写权限：



再次测试jack用户，尝试向dept表写入数据后查询：

0: jdbc:hive2://hadoop103:10000> insert into table dept values(50,'SECURITY',1800);

0: jdbc:hive2://hadoop103:10000> select \* from dept;



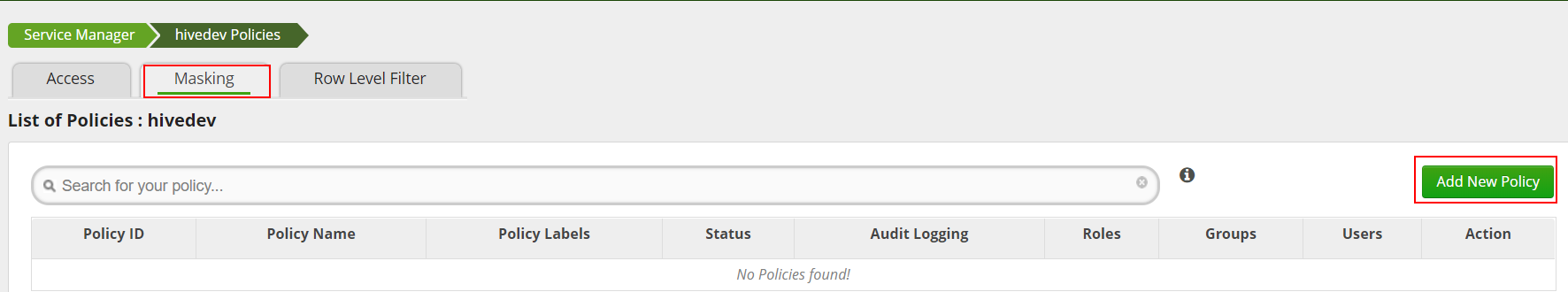
5.3 脱敏操作

通过脱敏操作可以限制用户对某一列的访问，将敏感数据不暴露给用户！

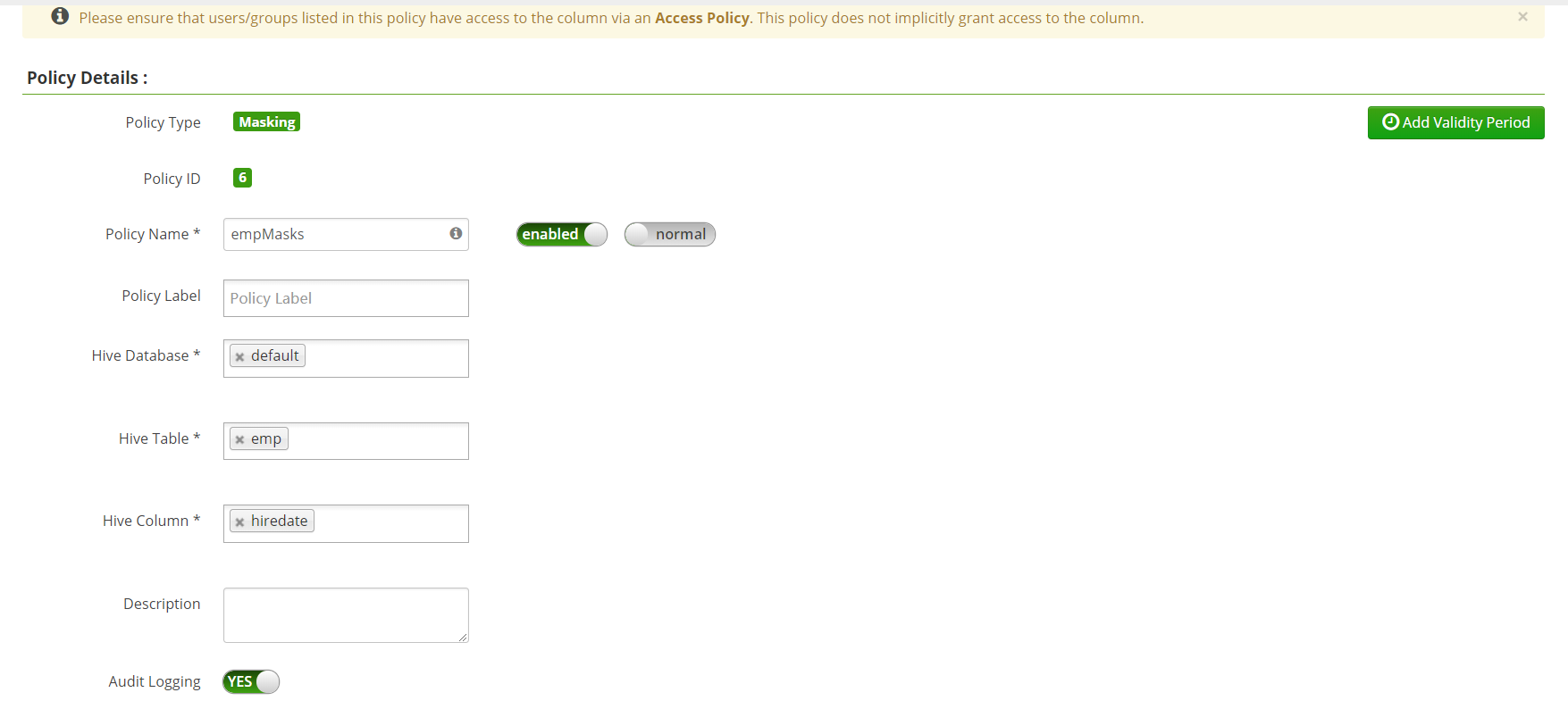
案例：指定atguigu用户在查询emp表时，对hiredate的年月部分脱敏！

首先需要保证用户对指定的列有访问权限，可以参考5.2进行配置！

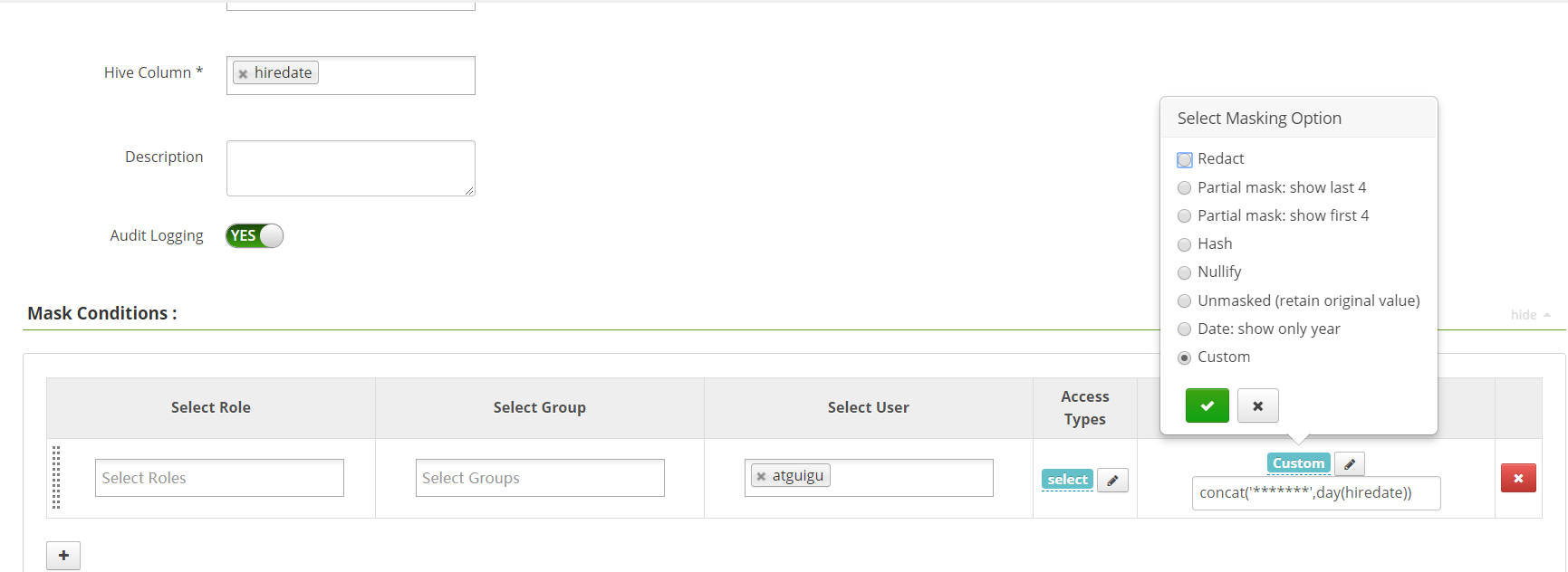
点击Masing标签，再点击Add New Policy



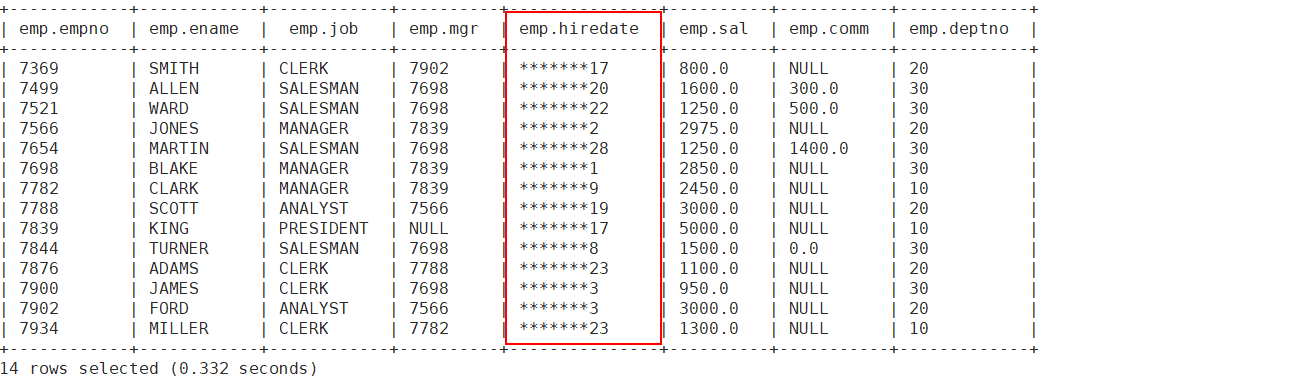
指定表和列



指定用户和脱敏操作：



之后点击save按钮！那么只有atguigu用户在查询时，会触发此策略！



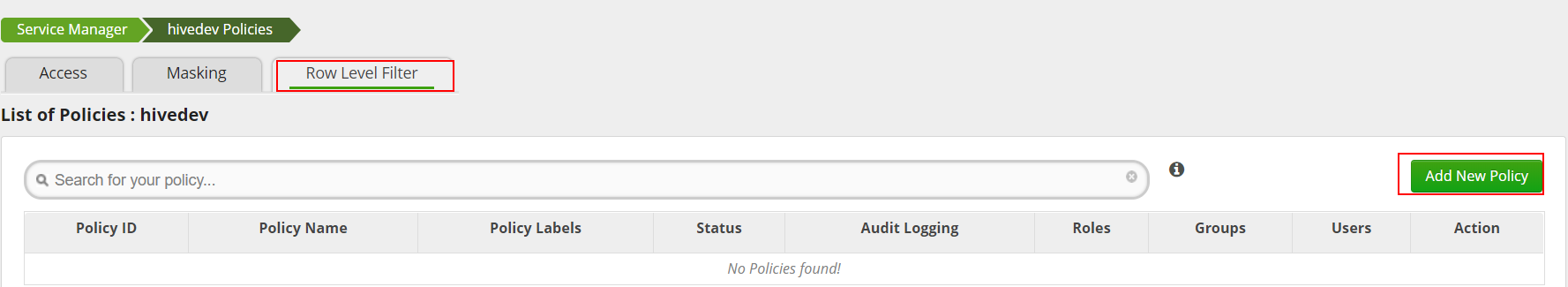
5.4 行级别过滤

通过行级别过滤可以将表中的数据进行条件过滤后再暴露给用户！

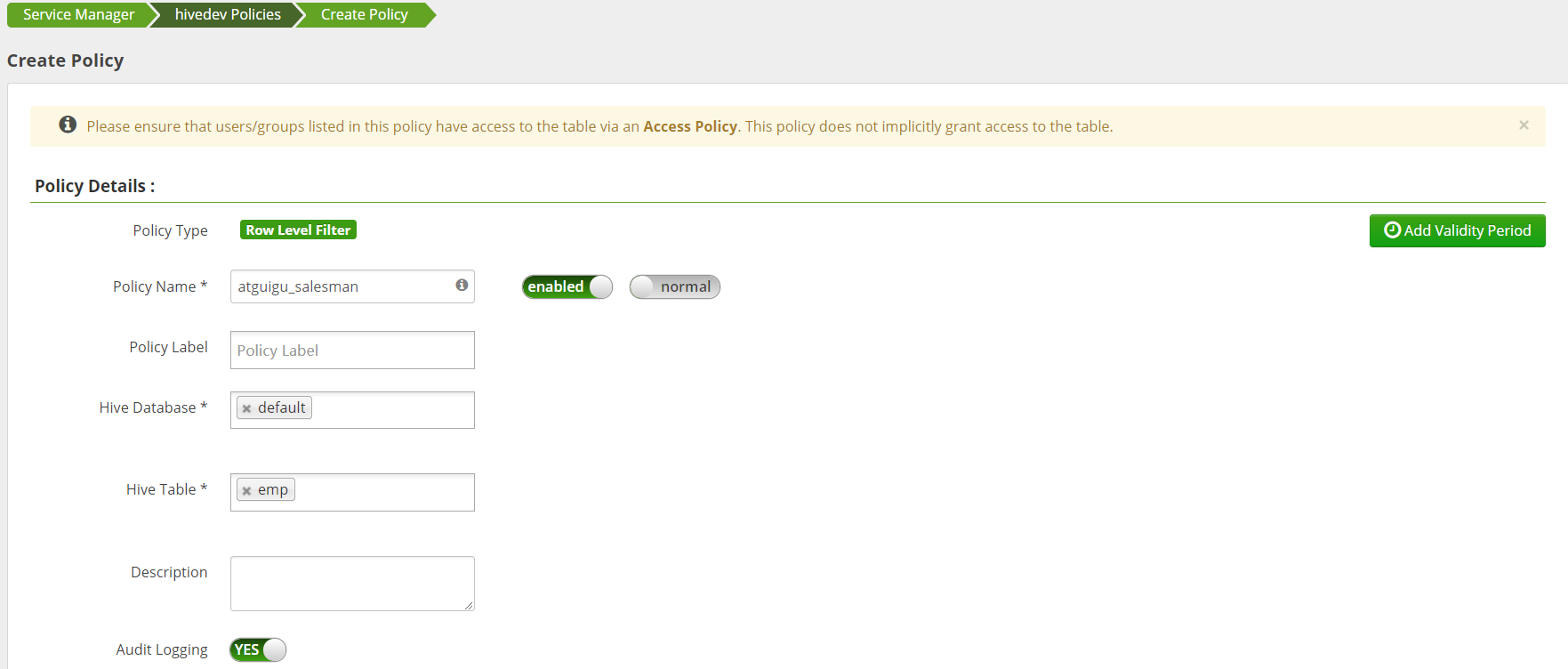
例如：atguigu用户只允许查询emp表中job类型为SALESMAN的用户信息。

同理，行级别过滤也要求用户对指定表有access权限！参考5.2的配置！

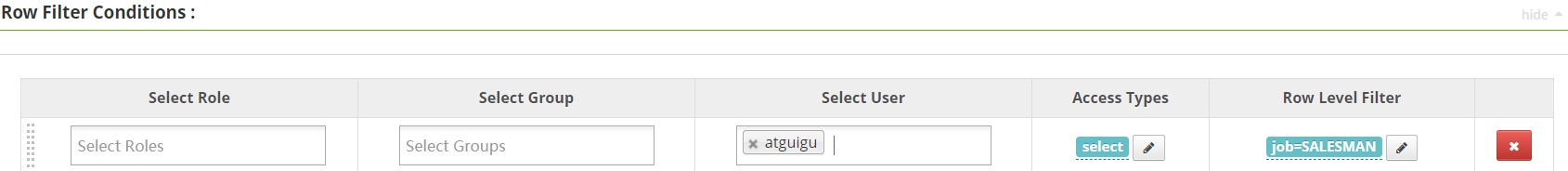
选择Row Level Filter标签，点击Add New Policy:



选择对应的库和表：

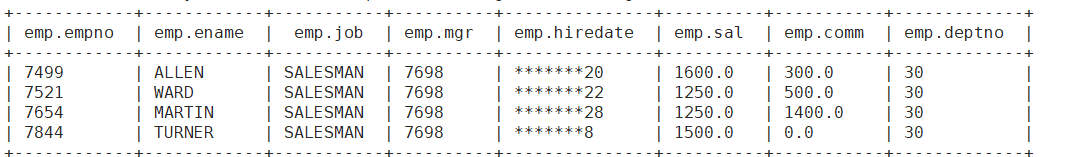


添加过滤规则和用户：



之后点击add按钮！

验证：



更多配置，可以参考官网介绍：https://cwiki.apache.org/confluence/display/RANGER/Row-level+filtering+and+column-masking+using+Apache+Ranger+policies+in+Apache+Hive