Arenadata™ Platform Security

Версия - v1.0.0

Руководство по работе с Apache Ranger

2021 Arenadata™

Оглавление

1 Функции безопасности Hadoop **3**
1.1 Аутентификация . . . 3
1.2 Авторизация (контроль доступа) . . . 3
1.3 Отчетность (аудит и мониторинг безопасности) . . . 3
1.4 Защита данных . . . 4

2 Настройка авторизации в Hadoop **5**
2.1 Ranger. Введение . . . 5
2.2 Установка Ranger . . . 8
2.3 HDFS Policy . . . 14

В документе приведены настройки функций безопасности Hadoop, авторизация в кластере, установка Apache Ranger и создание HDFS Policy.

Документ может быть полезен администраторам, программистам, разработчикам и сотрудникам подразделений информационных технологий, осуществляющих внедрение кластера.

Important: Перед установкой убедитесь, что версия бандла Arenadata Hadoop не ниже чем 2.1.3.0

Important: Контактная информация службы поддержки – e-mail: info@arenadata.io

Глава 1

Функции безопасности Hadoop

Для организаций, которые хранят конфиденциальные данные в экосистеме **Hadoop**, например, запатентованные или персональные данные, которые подлежат соблюдению нормативных требований (**HIPPA**, **PCI**, **DSS**, **FISAM** и т.д.), безопасность очень важна. Многие организации также должны придерживаться строгой внутренней политики безопасности.

Arenadata Platofrm Security (далее – **ADPS**) обеспечивает комплексный подход к безопасности в следующих ключевых областях: безопасность периметра, аутентификация, авторизация, отчетность, защита данных.

1.1 Аутентификация

ADH предоставляет единую точку аутентификации для сервисов и пользователей, которая интегрируется с существующими системами идентификации предприятия и доступа. **ADH** поддерживает следующие службы аутентификации:

- LDAP;
- Локальная система Unix;

1.2 Авторизация (контроль доступа)

ADH предоставляет функции, позволяющие системным администраторам контролировать доступ к данным **Hadoop** с использованием авторизации на основе ролей. **ADH** поддерживает следующие модели авторизации:

- Высококачественное управление доступом для данных, хранящихся в HDFS;
- Контроль доступа на уровне ресурсов для YARN;
- Контроль доступа на крупнозернистом уровне для операций MapReduce;
- Контроль доступа на уровне семейства таблиц и столбцов для данных HBase;
- Контроль доступа на уровне таблицы для наборов данных Apache Hive.

1.3 Отчетность (аудит и мониторинг безопасности)

ADH позволяет отслеживать активность **Hadoop**, используя Native Auditing, журналы аудита безопасности периметра на шлюзе **Knox**, и из центрального расположения консоли администрирования **ADH**, включая:

- Запросы доступа;
- Операции обработки данных;
- Изменение данных.

1.4 Защита данных

ADH предоставляет механизмы для шифрования данных при их передаче и требует использования партнерских решений для шифрования данных в состоянии покоя, их поиска и маскирования. **ADH** поддерживает следующие методы шифрования:

- SSL для компонентов ADH (данный режим подходит не для всех окружений и сервисов стека, в частности возможны проблемы при взаимодействии внутренних сервисов по данному протоколу);
- Шифрование RPC;
- Протокол передачи данных.

Глава 2

Настройка авторизации в Hadoop

- *Ranger. Введение*
- *Предварительные требования к установке*
- *Установка Ranger*
- *HDFS Policy*

2.1 Ranger. Введение

Apache Ranger можно установить при помощи пользовательского интерфейса **ADCM**. В отличие от ручного процесса установки, требующего выполнения ряда шагов, установка **Ranger** с использованием интерфейса **ADCM** проще и легче. Опция службы **Ranger** доступна через мастер **Add Service** после инсталляции кластера **ADH** с помощью установщика.

Служба Ranger включает в себя следующие компоненты:

- Ranger Admin
- Ranger UserSync
- Ranger Key Management Service

После установки и запуска этих компонентов можно включить плагины **Ranger**, перейдя к каждому отдельному сервису (**HDFS**, **HBase**, **Hiveserver2**, **Storm**, **Knox**, **YARN** и **Kafka**) и изменив конфигурацию в расширенном режиме *ranger-<service>-plugin-properties*.

Important: При включении плагина Ranger необходимо перезапустить компонент

Important: Включение Apache Kafka требует включения Kerberos

2.1.1 Предварительные требования к установке

Перед установкой **Ranger** необходимо убедиться, что кластер отвечает следующим требованиям:

- Установлен и настроен кластер Arenadata Hadoop версии не ниже 2.1.3
- Рекомендуется хранить аудиты как в HDFS, так и в Solr. Конфигурация по умолчанию для Ranger Audits в Solr использует общий экземпляр Solr;

- Чтобы обеспечить принудительную авторизацию на уровне групп LDAP/AD в Hadoop, необходимо настроить сопоставление групп Hadoop для LDAP/AD для LDAP (*Настройка сопоставления групп Hadoop для LDAP/AD*);
- Должен быть запущен и доступен экземпляр базы данных MySQL, который будет использоваться Ranger. Установщик Ranger создаст двух новых пользователей (имена по умолчанию: *rangeradmin* и *rangerlogger*) и две новые базы данных (имена по умолчанию: *ranger* и *ranger_ audit*).

Настройка сопоставления групп Hadoop для LDAP/AD

Для обеспечения принудительной авторизации на уровне групп **LDAP/AD** в **Hadoop**, необходимо настроить сопоставление групп **Hadoop** для **LDAP/AD**.

Important: Доступ к LDAP и сведения о подключении: настройки LDAP могут различаться в зависимости от используемой реализации LDAP

Существует три способа настройки сопоставления групп **Hadoop**.

- **Настройка сопоставления групп Hadoop для LDAP/AD с использованием SSSD (рекомендуется)**

Для сопоставления групп рекомендуется использовать SSSD или один из следующих сервисов подключения ОС **Linux** к **LDAP**:

- Centrify
- NSLCD
- Winbind
- SAMBA

Большинство перечисленных сервисов позволяет не только искать пользователя и перечислять группы, но также выполнять другие действия на хосте. При этом ни одно из этих действий не требуется для сопоставления групп **LDAP** в **Hadoop**. Поэтому, оценивая эти сервисы, необходимо понимать разницу между модулем **NSS** (который выполняет разрешение пользователь/группа) и модулем **PAM** (который выполняет аутентификацию пользователя). Для возможности поиска (или "валидации") пользователя в **LDAP** и перечисления групп требуется **NSS**. А **PAM** может представлять угрозу безопасности.

- **Настройка сопоставления групп Hadoop в файле core-site.xml**

Настройка **Hadoop** для использования сопоставления групп на основе **LDAP** в файле *core-site.xml* осуществляется в следующем порядке:

1. Добавить свойства, показанные в приведенном ниже примере, в файл *core-site.xml*. Необходимо указать значение для привязанного пользователя, его пароль и другие свойства, специфичные для экземпляра LDAP, и убедиться, что фильтры классов объектов, пользователей и групп соответствуют значениям, указанным в экземпляре LDAP.

```
<property>
<name>hadoop.security.group.mapping</name>
<value>org.apache.hadoop.security.LdapGroupsMapping</value>
</property>

<property>
<name>hadoop.security.group.mapping.ldap.bind.user</name>
<value>cn=Manager,dc=hadoop,dc=apache,dc=org</value>
</property>

<!-
```

```
<property>
<name>hadoop.security.group.mapping.ldap.bind.password.file</name>
<value>/etc/hadoop/conf/ldap-conn-pass.txt</value>
</property>
->

<property>
<name>hadoop.security.group.mapping.ldap.bind.password</name>
<value>hadoop</value>
</property>

<property>
<name>hadoop.security.group.mapping.ldap.url</name>
<value>ldap://localhost:389/dc=hadoop,dc=apache,dc=org</value>
</property>

<property>
<name>hadoop.security.group.mapping.ldap.url</name>
<value>ldap://localhost:389/dc=hadoop,dc=apache,dc=org</value>
</property>

<property>
<name>hadoop.security.group.mapping.ldap.base</name>
<value></value>
</property>

<property>
<name>hadoop.security.group.mapping.ldap.search.filter.user</name>
<value>(&amp;(|(objectclass=person)(objectclass=applicationProcess))(cn={0}))</value>
</property>

<property>
<name>hadoop.security.group.mapping.ldap.search.filter.group</name>
<value>(objectclass=groupOfNames)</value>
</property>

<property>
<name>hadoop.security.group.mapping.ldap.search.attr.member</name>
<value>member</value>
</property>

<property>
<name>hadoop.security.group.mapping.ldap.search.attr.group.name</name>
<value>cn</value>
</property>
```

2. В зависимости от конфигурации можно обновлять сопоставления пользователей и групп с помощью следующих команд HDFS и YARN:

```
hdfs dfsadmin -refreshUserToGroupsMappings
yarn rmadmin -refreshUserToGroupsMappings
```

3. Проверить сопоставление групп LDAP, выполнив команду *hdfs groups*. Команда отображает группы из LDAP для текущего пользователя. При настроенном сопоставлении групп LDAP разрешения HDFS могут использовать группы, определенные в LDAP для контроля доступа.

- **Ручное создание пользователей и групп в среде Linux**

Также можно вручную создавать пользователей и группы в среде Linux.

2.2 Установка Ranger

Установка **Ranger** с помощью **ADCM** заключается в три этапа:

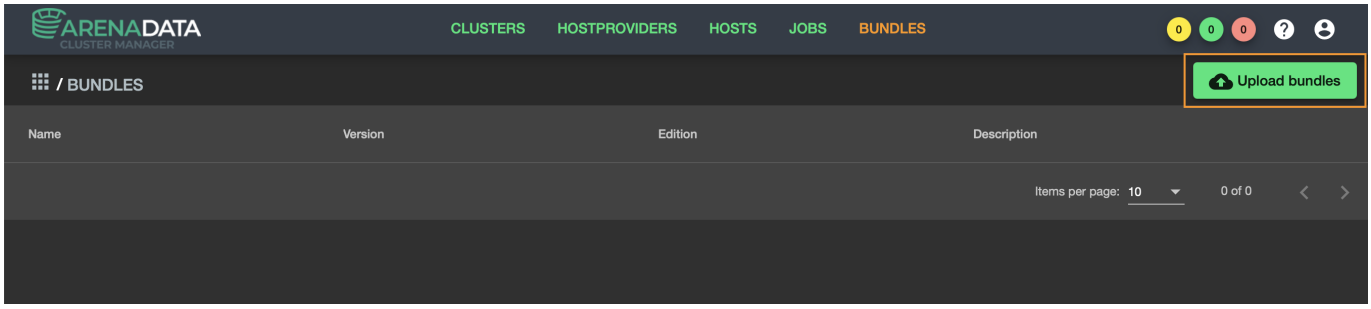
- *Загрузка бандла Arenadata Platform Security*
- *Создание кластера*
- *Конфигурирование сервисов*
- **‘Запуск установки‘_**

Смежные темы:

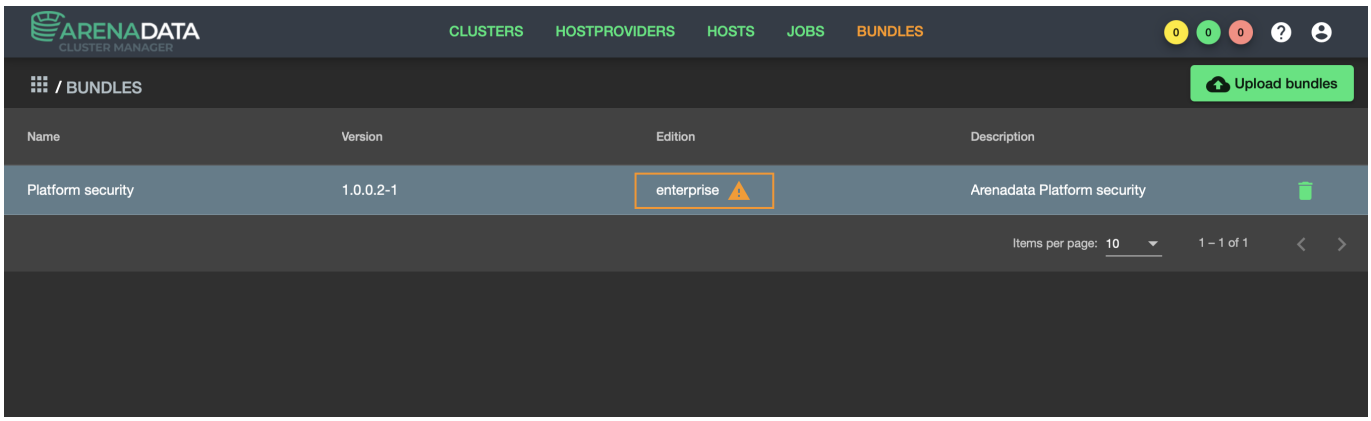
- **‘Расширенные настройки пользователей‘_**
- **‘Настройка пользователей без использования учетных данных DBA‘_**
- **‘Обновление паролей администратора Ranger‘_**
- *Включение плагинов Ranger*

2.2.1 Загрузка бандла Arenadata Platform Security

Для начала работы с компонентами Apache Ranger, необходимо загрузить бындл в текущий инстанс ADCM, с помощью действия “Upload Bundle”

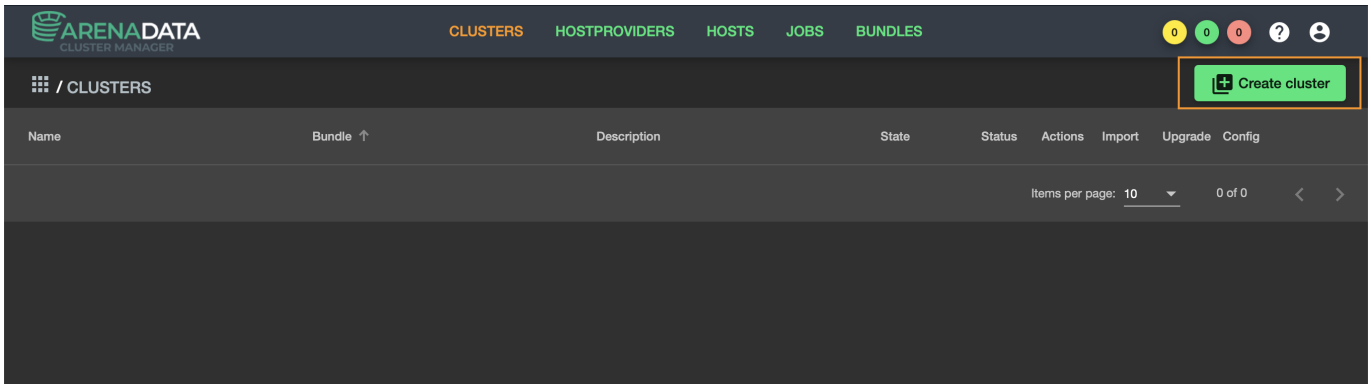


Далее необходимо принять соглашение об использовании (EULA)

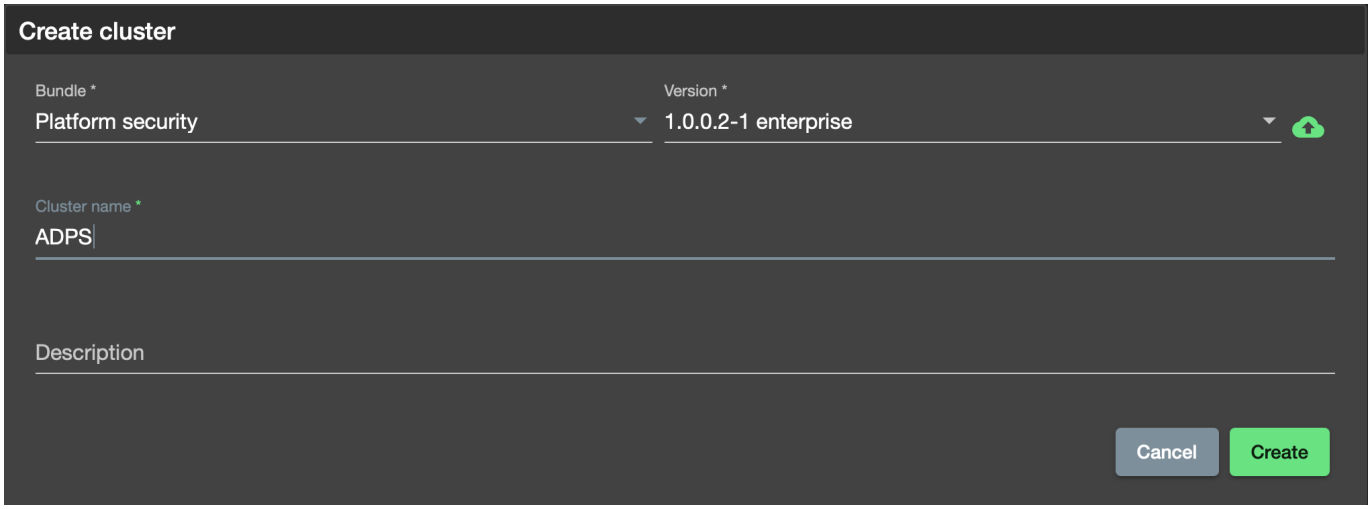


2.2.2 Создание кластера

Следующим шагом является создание кластера Arenadata Platform Security с помощью действия “Create cluster” в разделе “Clusters”



После чего, выберите имя кластера и завершите конфигурацию с помощью кнопки “Create”.



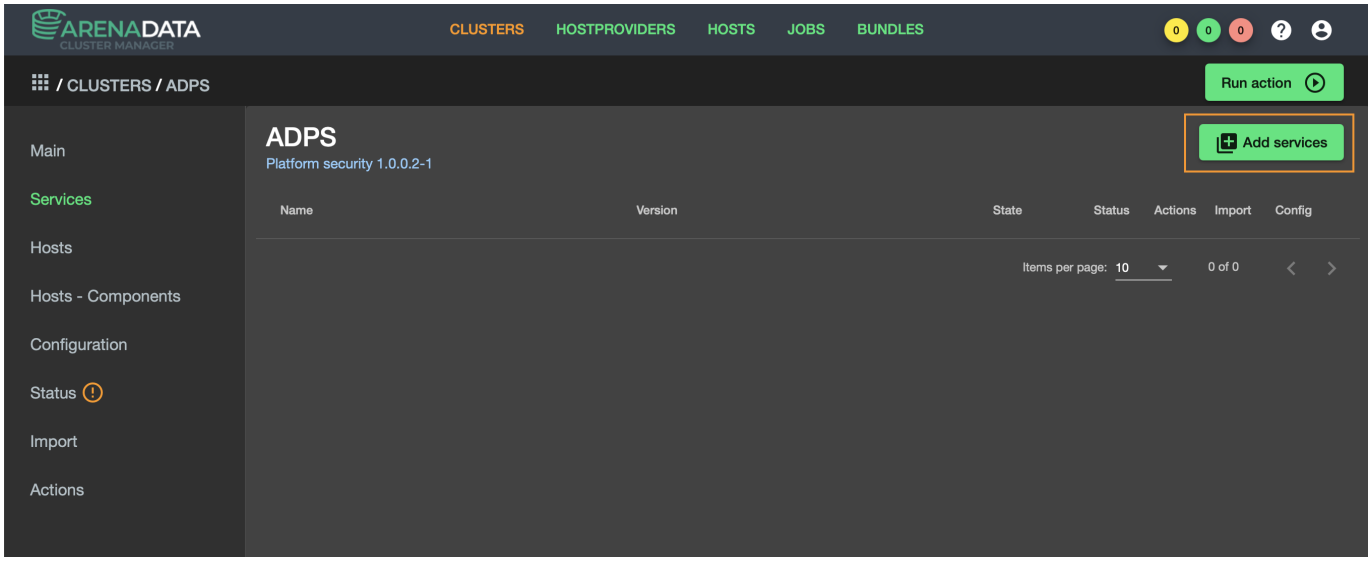
Следующим шагом является ыбор требуемых сервисов и распределение топологии компонентов. Для этого перейдите в конфигурацию кластера и выберите раздел “Services” и нажмите “Add services”

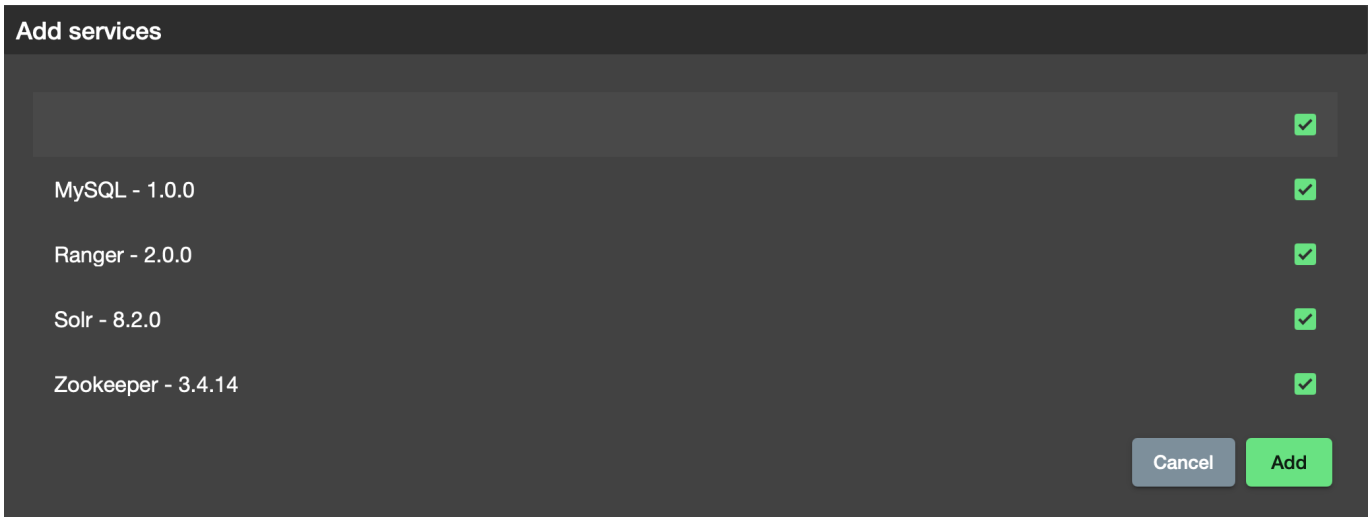
Выберите необходимые компоненты в интерфейсе ADCM

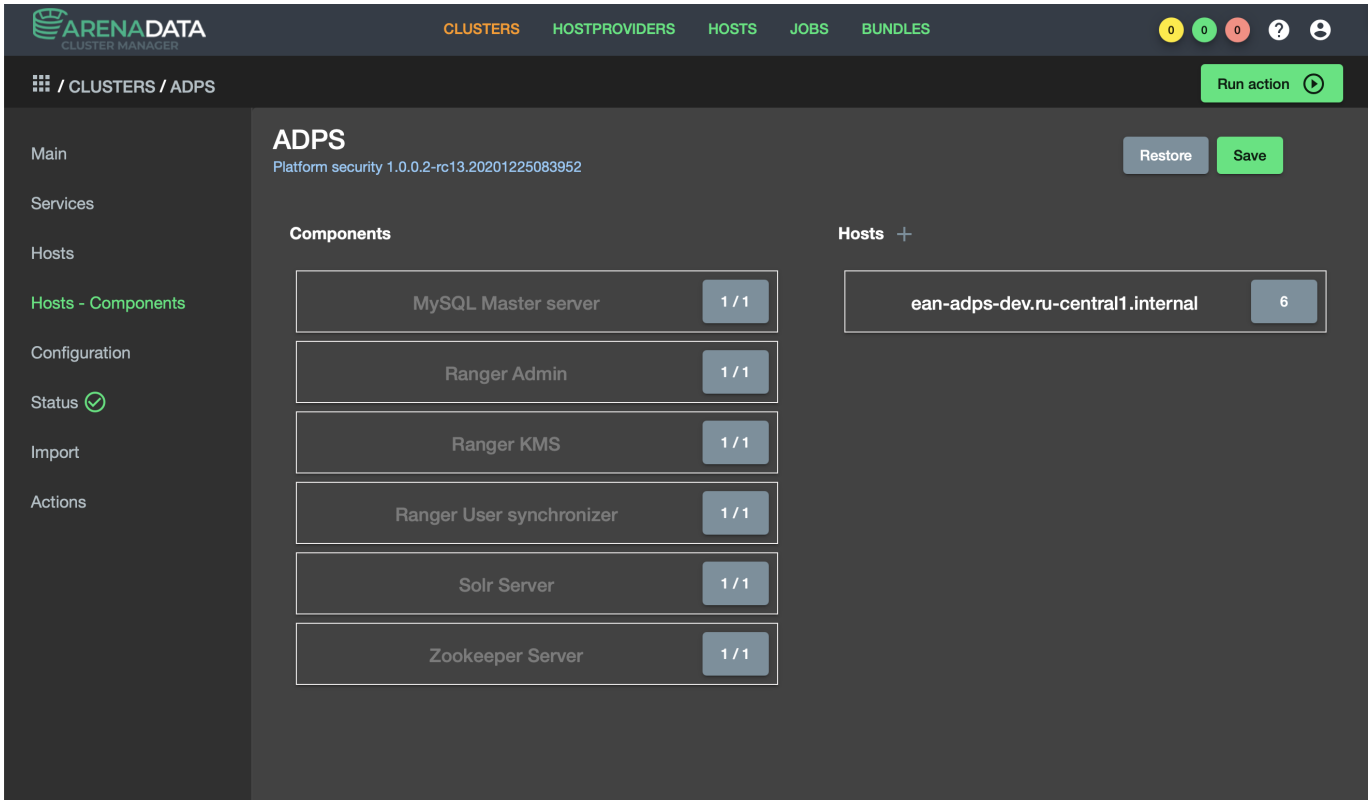
Далее перейдите в раздел “Host-Components” и распределите компоненты по хостам

Important: Все необходимые хосты должны быть созданы перед установкой компонентов

Important: Решение позволяет произвести установку всех компонентов в рамках одного хоста, но для промышленных инсталляций рекомендуется разнести компоненты между различными хостами для обеспечения большей отказоустойчивости


ARENADATA
CLUSTER MANAGER
CLUSTERS
HOSTPROVIDERS
HOSTS
JOBS
BUNDLES
0
0
0
/ CLUSTERS / ADPS
Run action
Main
Services
Hosts
Hosts - Components
Configuration
Status
Import
Actions
ADPS
Platform security 1.0.0.2-1
Add services
Name
Version
State
Status
Actions
Import
Config
Items per page: 10
0 of 0


Add services
MySQL - 1.0.0
Ranger - 2.0.0
Solr - 8.2.0
Zookeeper - 3.4.14
Cancel
Add



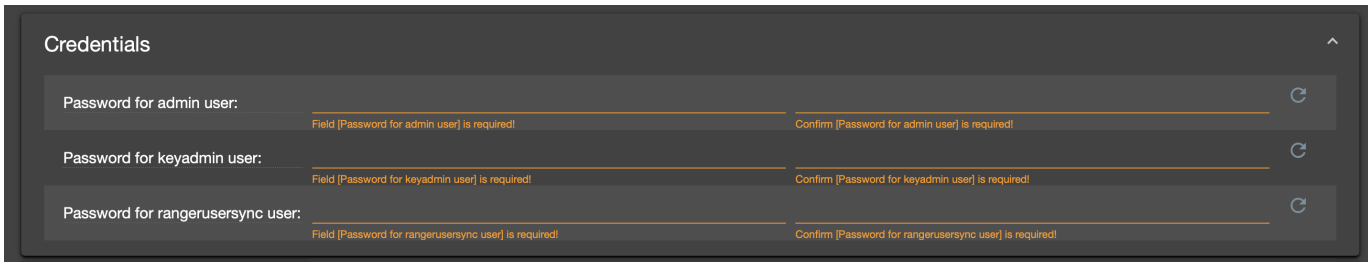
2.2.3 Конфигурирование сервисов

Следующим шагом в процессе установки **Ranger** является задание настроек на странице сервиса Ranger “Configuration”:

- *Credentials*
- *dbks-site.xml*
- *ranger-admin-site.xml*
- *ranger-ugsync-site.xml*

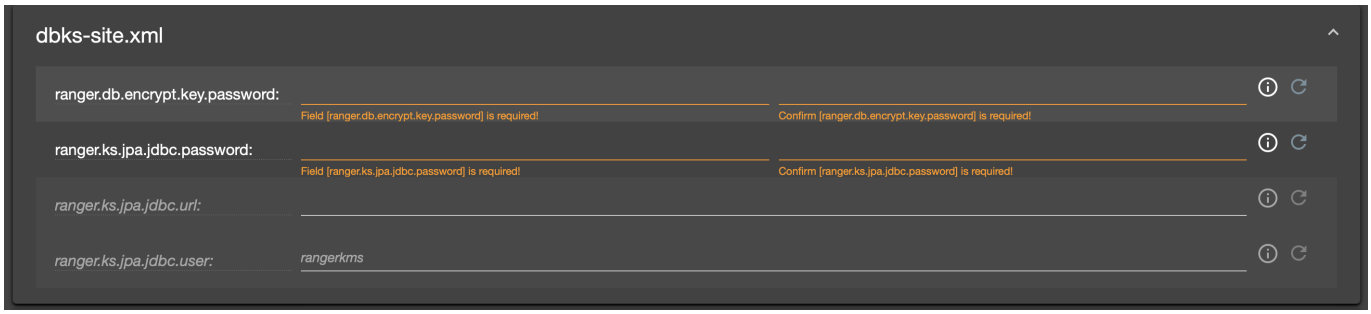
Credentials

В данном разделе необходимо указать учетные данные создаваемых технологических пользователей для доступа к интерфейсу и компонентам сервиса Ranger



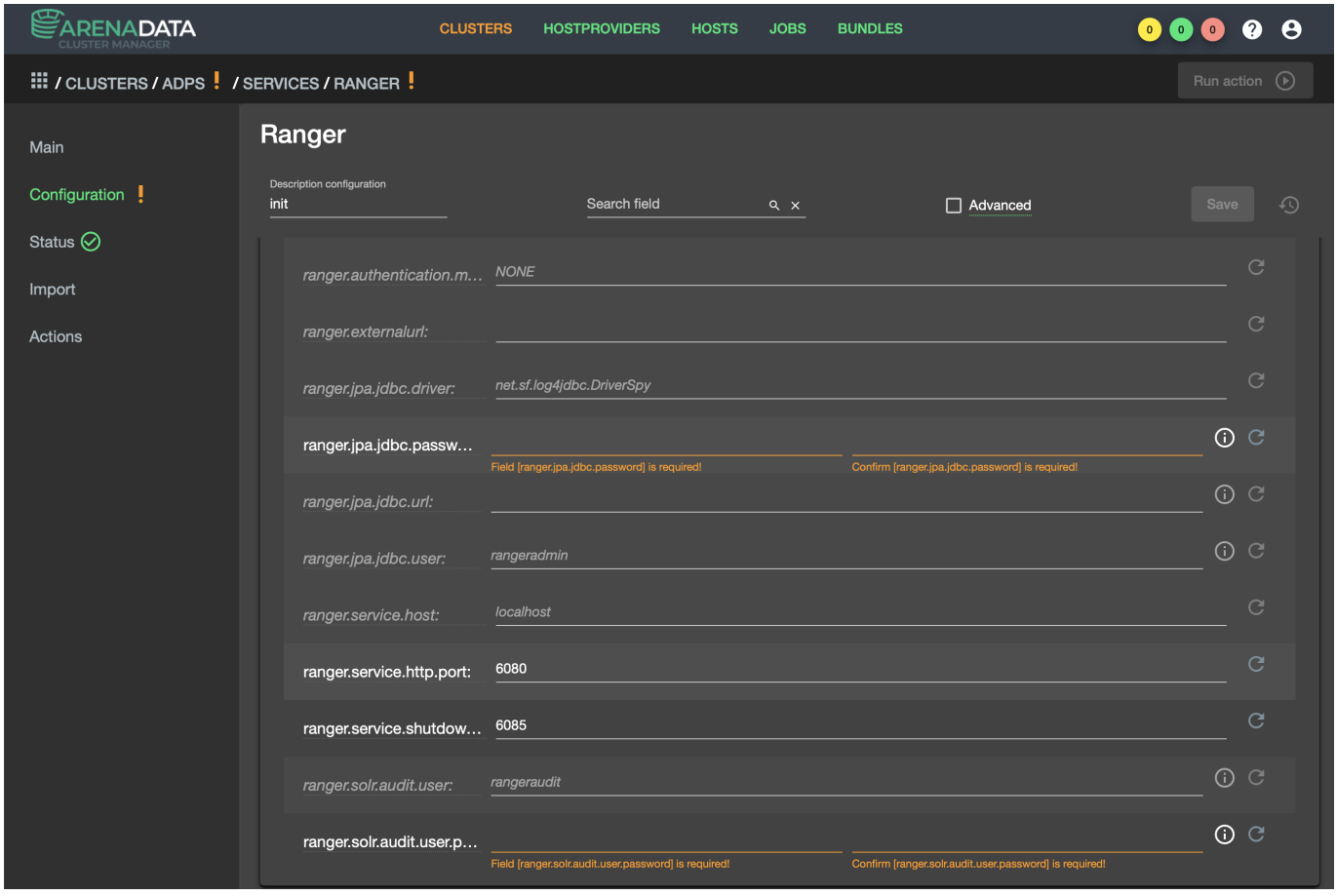
dbks-site.xml

В данном разделе необходимо указать пароль доступа для ключей шифрования и пароль подключения к БД



ranger-admin-site.xml

В данном разделе необходимо указать пароль доступа для подключения к БД и инстансу Solr для обеспечения аудита действий пользователей



ranger-ugsync-site.xml

В разделе описывается настройка **Ranger User Sync** для **UNIX** и **LDAP/AD**.

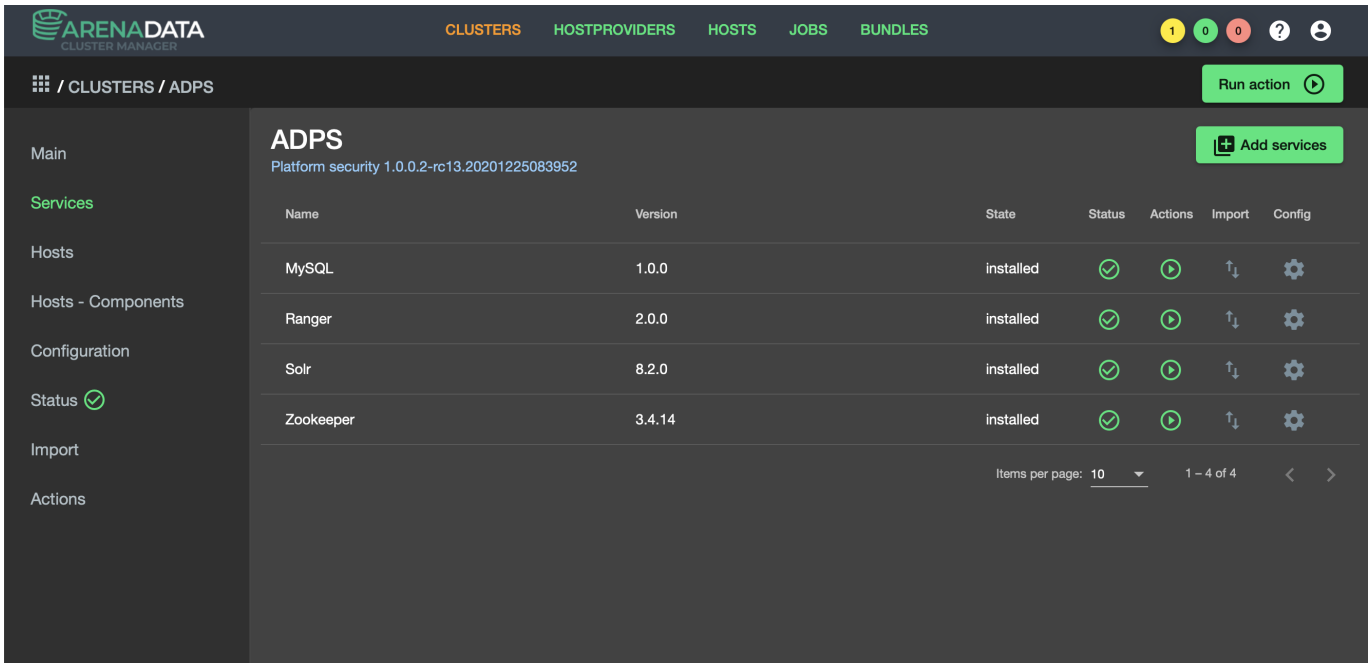
- **‘Настройка синхронизации пользователей Ranger для UNIX‘_**
- **‘Настройка синхронизации пользователя Ranger для LDAP/AD‘_**
- **‘Автоматическое назначение роли ADMIN/KEYADMIN для внешних пользователей‘_**

2.2.4 Завершение установки

Important: Перед запуском установки убедитесь что в интерфейсе более нет предупреждений в части конфигурации сервисов.

Для завершение процесса установки **Ranger** перйдите в раздел Main кластера и выберите действие “Install” в разделе “Run action”

После заершения установки, все компоненты должны иметь “зеленый” статус



2.2.5 Включение плагинов Ranger

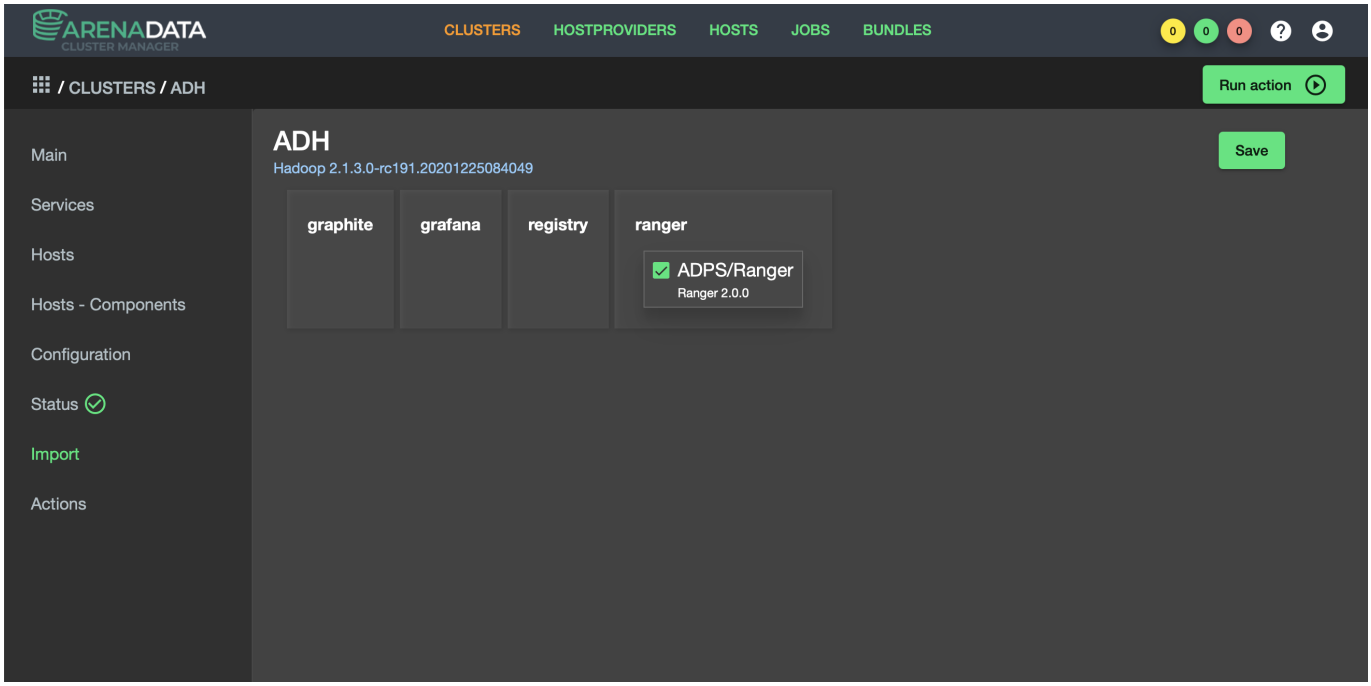
Плагины **Ranger** могут быть включены для нескольких сервисов **ADH**. По соображениям производительности рекомендуется хранить аудиты в **Solr**, а не в базе данных.

Доступны следующие плагины **Ranger**: *HDFS*, **Hive**, **HBase**, **YARN**, **Sqoop**.

Important: Перед включением плагинов необходимо сделать import в том кластере, который будет интегрирован с ADPS

HDFS

Для включения плагина **Ranger HDFS** необходимо выполнить следующие действия:



На странице сервиса HDFS кластера ADH в разделе “Run action”, выбрать Enable Ranger Plugin. После чего запустится процесс активации плагина и создание дефолтной политике на стороне Ranger.

2.3 HDFS Policy

2.3.1 Ranger для авторизации в Hadoop

После проверки подлинности пользователя необходимо определить его права доступа. Права доступа пользователя к ресурсам определяет авторизация. Например, пользователю может быть разрешено создание политики и просмотр отчетов, но не разрешено редактирование пользователей и групп. **Ranger** можно использовать для настройки и управления доступом к сервисам **Hadoop**.

Ranger позволяет создавать сервисы для определенных ресурсов **Hadoop** (**HDFS**, **HBase**, **Hive** и др.) и добавлять права доступа к этим сервисам. Можно также создавать сервисы на основе тегов и добавлять политики доступа к ним. Использование политик на основе тегов позволяет управлять доступом к ресурсам нескольких компонентов **Hadoop** без создания отдельных сервисов и политик в каждом компоненте. Можно также использовать **Ranger TagSync** для синхронизации хранилища тегов **Ranger** с внешним сервисом метаданных, таким как **Apache Atlas**.

2.3.2 Создание HDFS Policy

Благодаря конфигурации **Apache Ranger** позволяет проверять для запроса пользователя как политики **Ranger**, так и разрешения **HDFS**. Когда **NameNode** получает пользовательский запрос, плагин **Ranger** проверяет политики, установленные через **Ranger Service Manager**, и если их нет, проверяет разрешения, установленные в **HDFS**.

Рекомендуется создавать разрешения в **Ranger Service Manager** и иметь ограниченные разрешения на уровне **HDFS**.

Добавление новой политики к существующему сервису **HDFS** осуществляется по следующему алгоритму:

1. На странице “Service Manager” выбрать существующий сервис в разделе HDFS (Рис.2.1.).

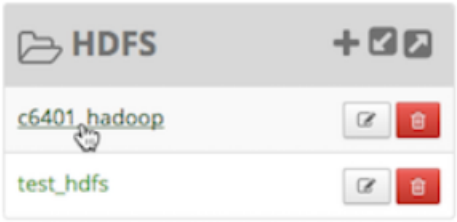


Рис.2.1.: Выбор сервиса HDFS

При этом открывается страница “List of Policies”, на которой необходимо нажать кнопку “Add New Policy” (Рис.2.2.).

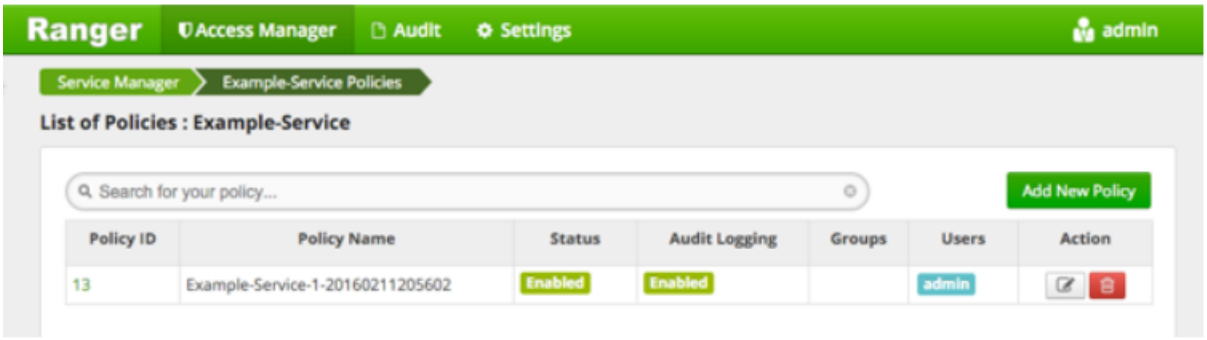


Рис.2.2.: List of Policies

2. Открывается страница “Create Policy” (Рис.2.3.).

На странице необходимо заполнить поля. Раздел “Policy Details”:

- *Policy Name* – ввести уникальное имя для данной политики (имя не может быть продублировано нигде в системе);
- *Resource Path* – определить путь к ресурсу для папки/файла политики. Во избежание необходимости указывать полный путь или включать политику для всех вложенных папок или файлов, можно заполнить это поле с помощью подстановочных знаков (например, /home*) либо указать, что политика должна быть рекурсивной;
  - Подстановочные знаки могут быть включены в путь ресурса, имя базы данных, таблицы или столбца: “*” – указывает ноль или более символов; “?” – указывает один символ;
- *Description* – (опционально) указать цель политики;
- *Audit Logging* – указать, выполняется ли аудит данной политики (снять флажок, чтобы отключить аудит).

Раздел “Allow Conditions”:

- *Select Group* – указать группу, к которой применяется данная политика. Чтобы назначить группу в качестве администратора для выбранного ресурса, выбрать *Admin permissions* (администраторы могут создавать дочерние политики на основе существующих). Группа *public* содержит всех пользователей, поэтому предоставление доступа к ней предоставляет доступ ко всем пользователям;
- *Select User* – указать конкретного пользователя, к которому применяется данная политика (за пределами уже указанной группы), или назначить определенного пользователя администратором данной политики (администраторы могут создавать дочерние политики на основе существующих);
- *Permissions* – добавить или изменить права: *Read* (чтение), *Write* (запись), *Create* (создание), *Admin* (Администратор), *Select/Deselect All* (выбрать/отменить все);

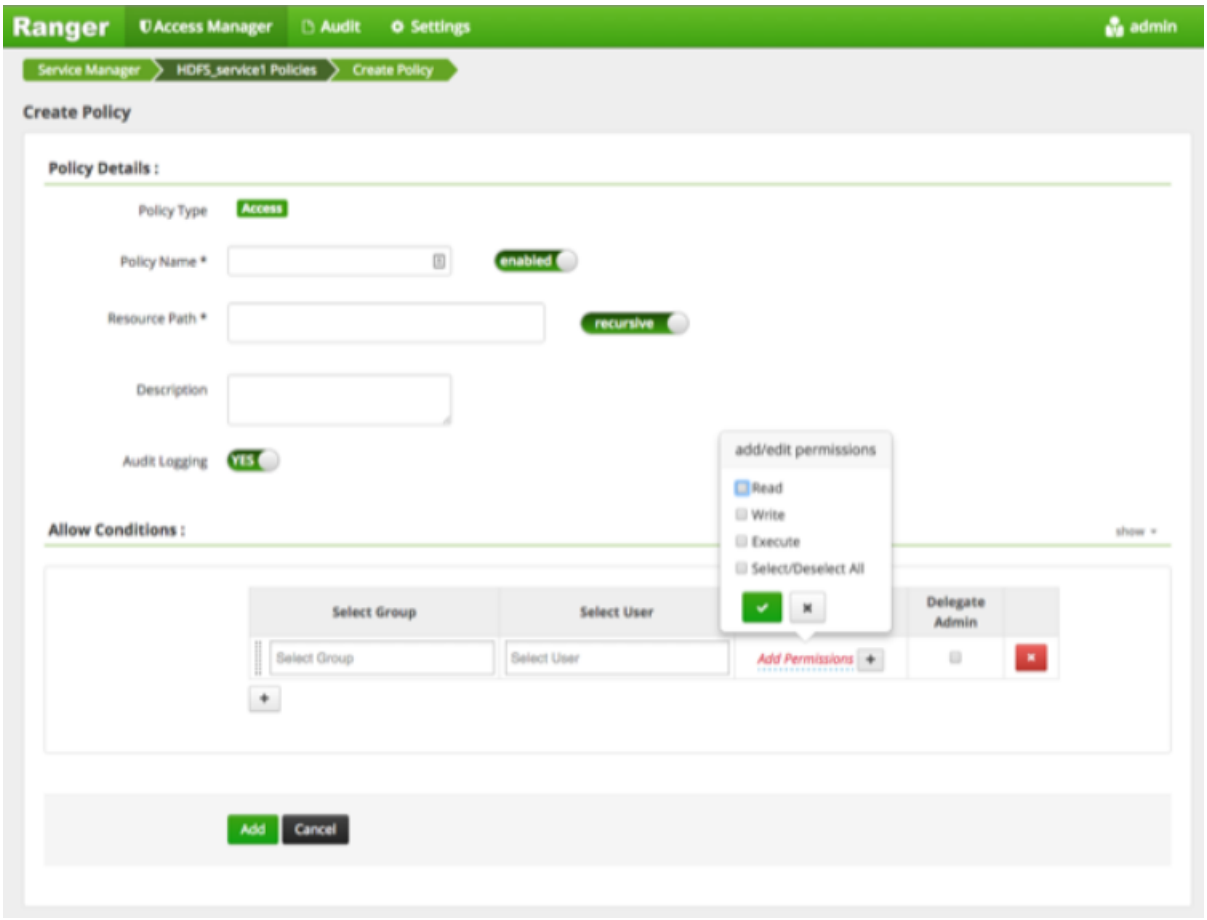


Рис.2.3.: Create Policy

- *Delegate Admin* – когда политика назначается пользователю или группе пользователей, данные пользователи становятся делегированными администраторами. Делегированный администратор может обновлять, удалять политики. Он также может создавать дочерние политики на основе исходной (базовой);

3. Для добавления дополнительных условий можно использовать символ плюс “+”. Условия оцениваются в порядке, указанном в списке – сначала применяется условие в верхней части списка, затем второе, третье и так далее;
4. Нажать кнопку *Add* для сохранения новой политики.