Monitorización y visualización EDAR 4.0

Recursos hardware máquina virtual

- **Procesador:** 2 cores del modelo Intel® Xeon® CPU E5-2650 v4 @ 2.20GHz
- **Memoria RAM:** 16GB
- **Memoria física:** 90GB HDD
- **Sistema operativo:** Ubuntu 16.04.5 LTS
- **IP pública:** 193.146.78.62

Usuarios del sistema

Para construir la arquitectura Cloud junto con la aplicación de monitorización y visualización, han sido creados distintos usuarios con distintos fines en el sistema operativo:

- Para administrar cada uno de los servicios disponible en la plataforma HDP, la propia plataforma crea los usuarios necesarios, un usuario por cada servicio, sin contraseña con la cual iniciar sesión.
- Los usuarios creados manualmente para completar esta administración son los siguientes:
    - **Usuario:** administrator, **Password:** eskola2017 (Usuario principal del sistema utilizado para acceder y administrar todo el sistema completo).
    - **Usuario:** bokeh, **Password:** bokeh (Usuario que inicializa y gestiona la aplicación de monitorización y visualización disponible).
    - **Usuario:** rapidminer, **Password:** rapidminer (Usuario responsable de gestionar y administrar la herramienta RapidMiner disponible en el sistema).

Puertos abiertos sistema

Los puertos necesarios por abrir en la arquitectura para permitir un correcto funcionamiento y que los servicios puedan comunicarse de forma correcta son los siguientes:

- 9995: Aplicación de monitorización y visualización (Flask/Gunicorn).
- 9090: Bokeh Server
- 8888: RapidMiner Server
- Puertos para permitir comunicar RapidMiner Studio (versión local) con los servicios de la arquitectura:
    - 8020
    - 8050
    - 8030
    - 1019
    - 10000
    - 1083
    - 88
    - 749
    - 10020
    - 8188
    - 8025

Arquitectura software Cloud

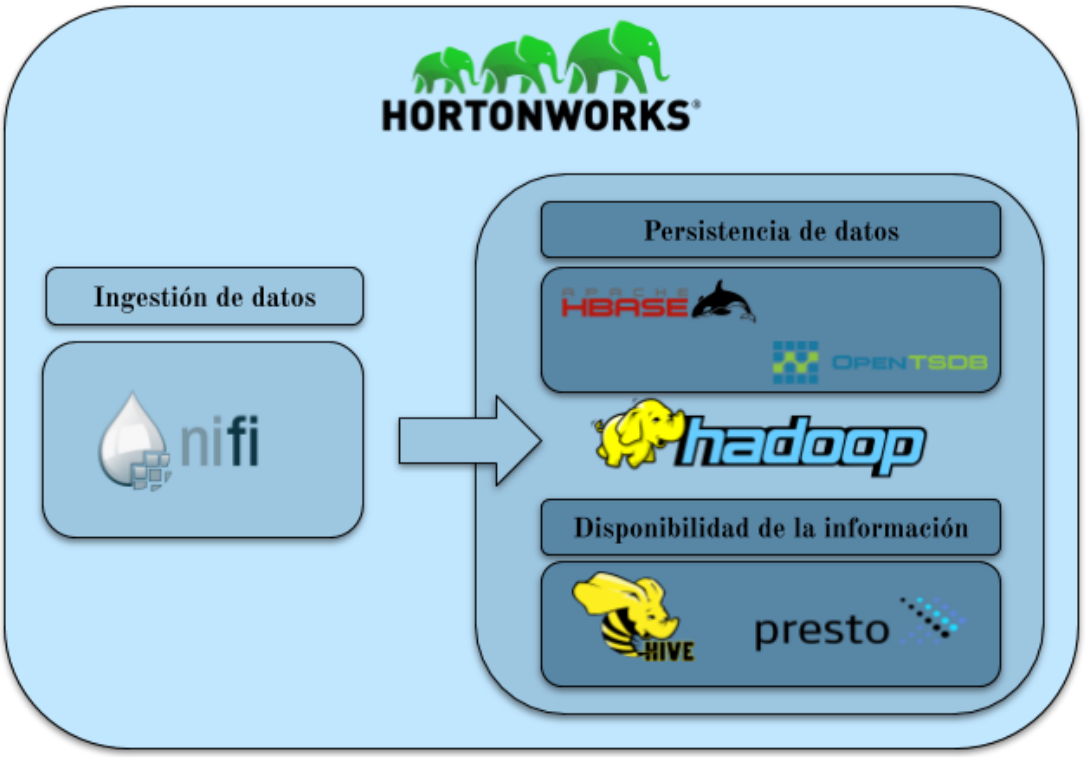

Ilustración 1: Arquitectura general Cloud EDAR 4.0

La arquitectura está construida haciendo uso de la plataforma de datos HDP (Hortonworks Data Platform) disponible por la organización Hortonworks.

Mediante esta distribución basada en el framework Open Source de referencia en el ámbito del Big Data, Apache Hadoop, se facilita la instalación, gestión y administración de todas las herramientas y tecnologías del ecosistema Hadoop con la ayuda de la herramienta Apache Ambari.

La versión de la plataforma HDP instalada en estos momentos es la siguiente: **2.6.5.1050-37**. Aunque la propia organización en su documentación explica los pasos necesarios a llevar a cabo para mantener la plataforma actualizada.

De las herramientas mostradas en la anterior figura, Apache NiFi, OpenTSDB y Presto no están incluidas por defecto en la plataforma de Hortonworks.

Para incluir la herramienta Apache NiFi, la compañía Hortonworks ofrece otras plataformas a distintos casos de uso donde la herramienta Apache NiFi se encuentra incluida. Entre ellas, la solución HDF (Hortonworks DataFlow) destinada a la ingestión y el análisis de datos en tiempo real. Aún y todo, conscientes de los diferentes casos de uso que pueden surgir en una misma solución o una única plataforma, la compañía ofrece la oportunidad de integrar los servicios de la plataforma HDF en un clúster HDP de manera sencilla.

En este caso se ha optado por seguir las instrucciones de Hortonworks e integrar la versión 1.5.0 de Apache NiFi incluida en la plataforma HDF con versión 3.0.0.0-453.

Las herramientas restantes no incluidas en la distribución HDP, OpenTSDB y Presto, han sido descargadas, configuradas e instaladas manualmente. Las rutas de instalación en el sistema son las siguientes:

- **OpenTSDB:** /usr/hdp/2.6.2.0-205/opentsdb/
- **Presto:** /home/administrator/

Junto con las herramientas de la anterior figura, la seguridad de la arquitectura se gestiona mediante herramientas adicionales: mecanismo de autenticación Kerberos y la herramienta de centralización de seguridad del ecosistema Hadoop, Apache Ranger. Además, la mayoría de las comunicaciones de la arquitectura son protegidas, salvaguardando los protocolos mayormente utilizados, mediante el cifrado TLS (Transport Layer Security), el framework SASL (Simple Authentication and Security Layer) y la protección del protocolo DataTransfer Protocol en caso de Hadoop.

Kerberos es un mecanismo de autenticación para entornos informáticos de red abierta. Mediante este mecanismo se previenen y solventan los problemas de autenticación que pueden surgir cuando un usuario necesita acceder a diversos servicios disponibles en distintos equipos en una red no fiable como Internet. Para ello, cada usuario accediendo a un servicio debe especificar su identidad de manera segura, y además, el servidor que contenga ese servicio también debe demostrar su identidad de la misma forma para que el usuario pueda confiar en el. Por lo tanto, Kerberos ofrece una autenticación mutua entre el cliente y el servidor basándose en criptografía de clave simétrica y un tercero de confianza denominado centro de distribución de claves o KDC (Key Distribution Center), formado por los dos servidores distintos, por un lado, un servidor de autenticación o AS (Authentication Server), y por otro lado, un servidor emisor de tickets o TGS (Ticket Granting Server). Este último, como su nombre indica, emite tickets utilizados para demostrar la identidad de los usuarios.

En este caso, para un correcto funcionamiento del protocolo Kerberos en la arquitectura, se ha creado un usuario o principal administrador responsable de realizar todas las acciones de administración. El principal contiene las siguientes credenciales:

- **Principal:** admin/admin@EDAR40.EUS, **Password:** eskola2017

También es conveniente indicar que todos los keytabs utilizados sobre el mecanismo Kerberos para realizar la autenticación se encuentran ubicados en la siguiente ruta:

- /etc/security/keytabs/

Apache Ranger es una herramienta que habilita, monitoriza y gestiona la amplía seguridad de una plataforma Hadoop. Para ello, la herramienta dispone de distintas API RESTs y una interfaz web que permiten especificar de manera centralizada reglas de autorización de grado fino para ejecutar una acción u operación sobre un componente o herramienta de Hadoop. La herramienta permite construir estas reglas siguiendo diferentes métodos de autorización, basado en roles o RBAC (Role Based Access Control) y basado en atributos o ABAC (Attribute Based Access Control).

En la arquitectura Cloud construida se han especificado reglas de autorización basadas en roles, todas ella, centralizadas y manejadas en la interfaz web disponible donde para poder acceder es necesario iniciar sesión con los siguientes valores de identificación:

- **Usuario:** admin, **Password:** eskola2017

Finalmente, para la protección de la comunicación de los datos utilizando el cifrado TLS, se han creado certificados firmados por un CA (Certificate Authority) propio e incluido junto con este documento para poder importarlo en los almacenes de certificados correspondientes y poder confiar así en los certificados.

Detalles de la aplicación de monitorización y visualización

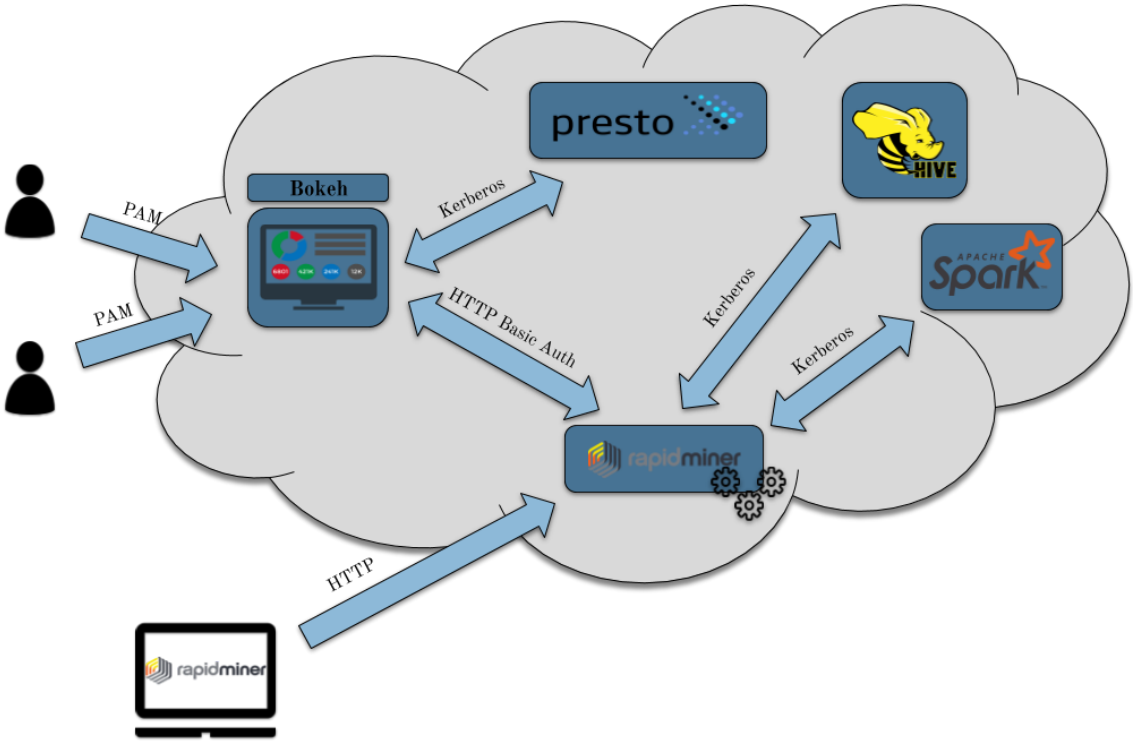

Ilustración 2: Estructura aplicación de monitorización y visualización

Los detalles de desarrollo de aplicación de monitorización y visualización están disponibles en la presentación adjuntada junto con este documento. A pesar de todo, es necesario aclarar algunos conceptos concretamente acerca de la herramienta RapidMiner utilizada.

Para poder comunicar la versión Studio del software RapidMiner y la versión Server del propio software es necesario que coincidan en el número de versión utilizado. En este caso, la versión del software RapidMiner utilizada es **9.2.**

RapidMiner Server hace uso de una interfaz web para realizar la administración completa de la herramienta, para ello, el usuario accediendo necesita identificarse y actualmente debe hacerlo con las siguientes credenciales:

- **Usuario:** rapidminer, **Password:** rapidminer

Para realizar la instalación de tanto el software en la versión Studio como Server se hace uso de una licencia educacional de un año. En ambos casos, se siguen también los pasos establecidos en la documentación del software RapidMiner para realizar la instalación. En caso de RapidMiner Server, la instalación debe realizarse mediante una pequeña aplicación disponible, donde se deben indicar las configuraciones correctas. Al estar utilizando una máquina virtual sin interfaz gráfica, esta pequeña aplicación de

instalación no puede ejecutarse en la propia máquina virtual, y por lo tanto, es necesario ejecutarla en otro equipo y hacer uso de lo que RapidMiner llama la instalación Headless install. Para ello, en la aplicación de instalación ejecutándose es necesario insertar los valores de configuración adecuados correspondientes a la propia máquina virtual.

De estas configuraciones a indicar las siguientes son las más relevantes a cumplir de manera adecuada:

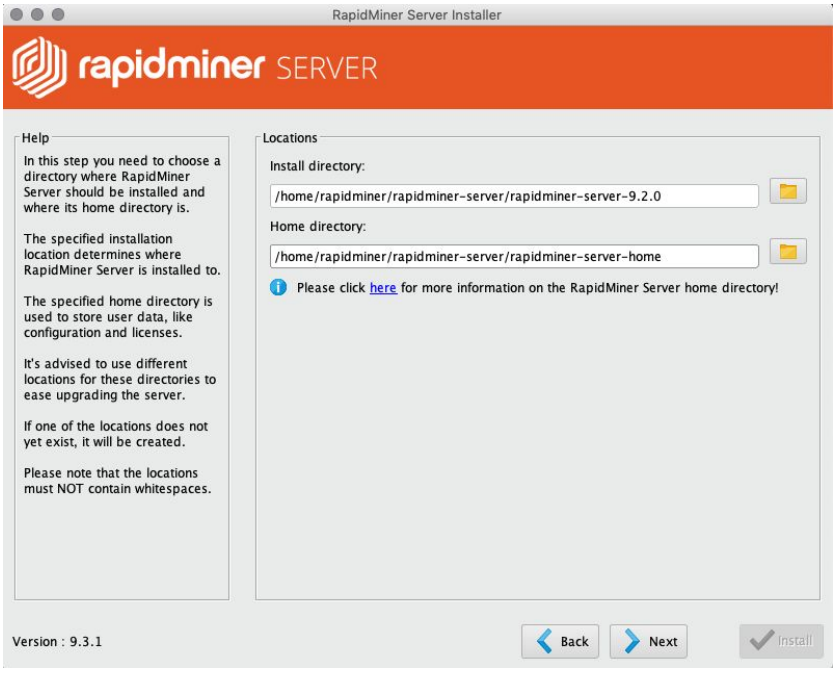

Ilustración 3: Configuraciones de instalación RapidMiner Server (1)

Ilustración 4: Configuraciones de instalación RapidMiner Server (2)

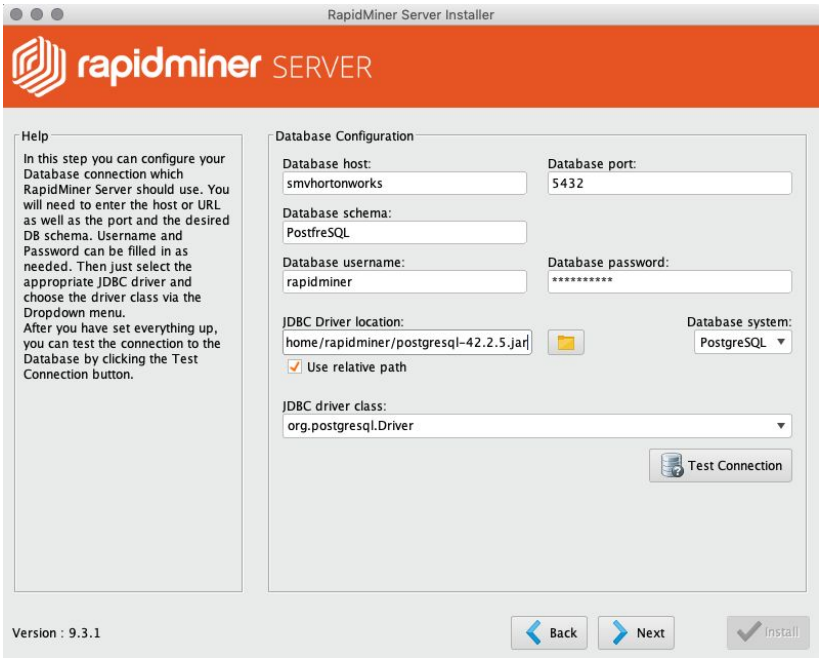

Ilustración 5: Configuraciones de instalación RapidMiner Server (3)

Como se puede observar en las anteriores imágenes, para la gestión de la herramienta RapidMiner Server se hace uso de una base de datos PostgreSQL ya disponible en el sistema. En esta misma base de datos también existe un usuario llamado rapidminer, responsable de realizar toda la gestión completa, con las siguientes credenciales:

- **Usuario**: rapidminer, **Password**: rapidminer

Siguiendo con los conceptos a aclarar respecto a la herramienta RapidMiner Server, los servicios web disponibles para realizar la ejecución del procesado de los datos están protegidos por una autenticación simple sobre el protocolo HTTP, mediante el cual el usuario debe ingresar los valores de identificación correctos, en este caso los siguientes:

- **Usuario:** rapidminer, **Password:** rapidminer

Estos valores de identificación pueden alterarse desde la propia interfaz web de la herramienta o es posible mantener los servicios web públicos para que cualquier usuario pueda tener acceso.

Finalmente, junto con este documento se incluye un documento adicional, en el cual, se especifican los pasos a llevar a cabo para instalar el complemento Radoop tanto en la versión Studio de RapidMiner como en la versión Server y como conectarse a RapidMiner Server desde RapidMiner Studio.