

Capítulo 5

Despliegue, Monitoreo y Métricas

Con los objetivos de disponibilizar el sistema de BI para que los especialistas de la IM puedan hacer uso del mismo, y realizar una transferencia de conocimientos a los encargados de la gestión técnica, para la infraestructura y monitorización de la solución, se describe la etapa de chapter 5 de software, donde se engloban todas las actividades que hacen que un sistema de software esté disponible para su uso, sea posible controlar su correcto funcionamiento y se pueda visualizar métricas de uso.

5.1. Despliegue

Debido a que al momento de la puesta en producción del sistema no se disponen de recursos técnicos para implantar la solución en la infraestructura de la IM, se decide por un despliegue en la nube utilizando tecnologías de contenedores dadas las distintas ventajas que poseen para el caso en cuestión, a saber, compatibilidad y mantenibilidad, simplicidad y configuraciones más rápidas, despliegue rápido, plataformas multi-nube, aislamiento y seguridad.

Si a futuro es requerido realizar una migración a la infraestructura de la IM, solamente es necesario realizar un backup y luego un restore en otro Docker Host interno, debido a que las redes, puertos, conexiones y contenedores son independientes del host en el que se despliegan.

5.1.1. Componentes

Primeramente se describen los distintos componentes que pertenecen al sistema mediante un Diagrama de Despliegue, en fig. 5.1, y se especifican las interacciones entre los mismos:

5.1.1.1. Aplicación web

El aplicativo web de carga, actualización y reportes se instala dentro de un contenedor de Debian Linux en un servidor web NGINX. Dicho aplicativo web es desarrollado utilizando tecnologías estándar como HTML/JS/CSS y la librería React para facilitar el desarrollo y mantenimiento.

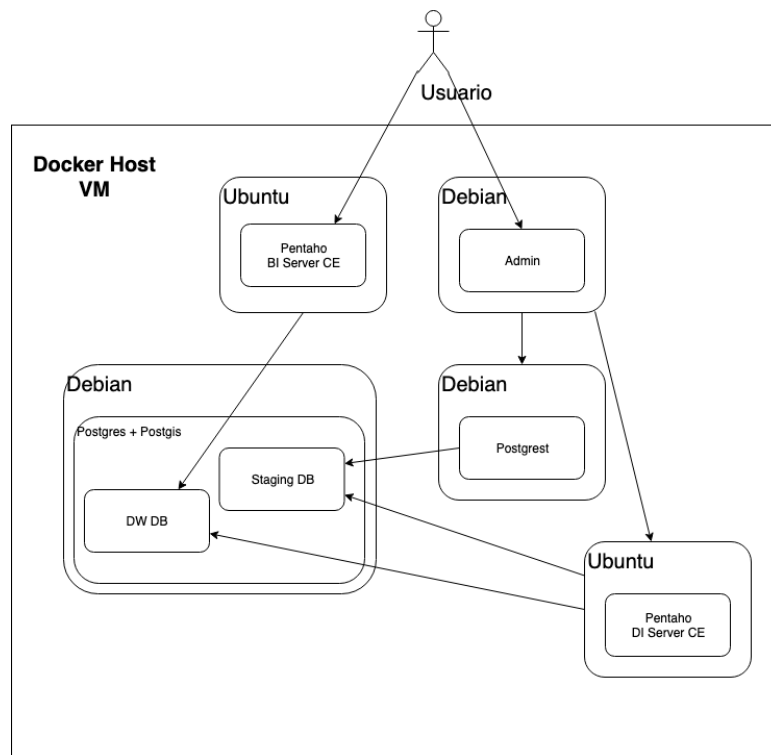


Figura 5.1: Despliegue

5.1.1.2. Base de Datos

Existen distintos contenedores, relacionados al almacenamiento de datos, que forman parte de la solución:

1. PostgreSQL + PostGIS: Este contenedor Debian hace parte como la Base de Datos principal del sistema desplegado. Es utilizado para almacenar las fuentes de datos, para las bases de staging y, como objetivo primordial, para almacenar los datos del DW.
2. PostgREST: Este contenedor Debian hace parte como la capa de servicios utilizada por el aplicativo web para la gestión de las fuentes de datos.
3. pgAdmin: Este contenedor Debian sirve una aplicación web para la gestión de la base de datos PostgreSQL de forma más eficiente.

5.1.1.3. PBI CE

La herramienta core de la suite de Pentaho utilizada es la solución de BI, específicamente, la versión CE. La misma se despliega en un contenedor Ubuntu que contiene el servidor y un contenedor de bases de datos MySQL Server, corriendo en un Oracle Linux, para el almacenamiento de los datos y configuración del contenedor del servidor.

5.1.1.4. PDI CE

Tan indispensable como el componente PBI, es el contenedor PDI, también en su versión CE, cuyo nombre clave es Kettle, y sirve para la planificación y ejecución de los ETL.

La misma se despliega en un contenedor Ubuntu para brindar la capacidad de ejecución de los procesos de forma remota utilizando una interfaz web vía servicios REST denominada Carte.

5.1.1.5. Documentación en línea

Con el objetivo de brindar documentación en línea se despliega un contenedor, basado en Debian, que provee una herramienta web con videos y tutoriales sobre el uso de la plataforma. El mismo es accesible desde la aplicación web de carga, actualización y reportes y puede visualizarse en fig. 5.2.

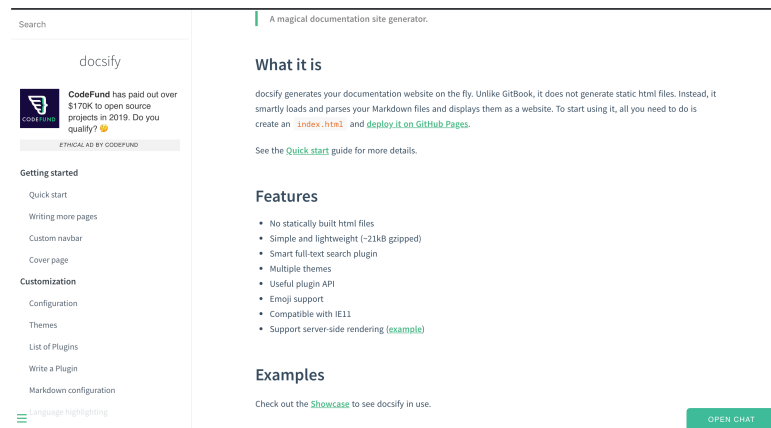


Figura 5.2: Documentación en línea

5.1.1.6. Sistema de archivos

Para lograr desacoplar las distintas responsabilidades que tienen los componentes, se dispone de un contenedor, basado en Debian, que sirve una aplicación web para la gestión de archivos a un SFTP, que se encuentra desplegado en otro contenedor basado en Debian.

Con esta decisión, los archivos que se manejan en la plataforma, a saber, los archivos fuentes de datos y los reportes generados por la aplicación web, se almacenan en un repositorio de archivos que no comparte el espacio lógico de los otros componentes, por ejemplo, el espacio lógico de PDI que es quien utiliza dichos archivos como entrada y salida de la ejecución de los procesos.

5.1.1.7. R

Para la generación de la sección Resultados de Red de Monitoreo del Informe Anual es necesario realizar cálculos estadísticos que los especialistas de la IM realizan manualmente en la aplicación de escritorio R.

Dada la necesidad de ejecución de los métodos estadísticos en los procesos ETL de PDI, es necesario brindarle la capacidad de utilizar dichos métodos en tiempo de ejecución de los procesos. Esto es posible mediante la disponibilización de un contenedor servidor de R con una capa de servicios REST, provista por Plumber, para la integración.

5.1.2. Interacción

El sistema de BI presentado consta de múltiples procesos en los cuales los distintos componentes de la solución interactúan para lograr distintos objetivos. A continuación se presentan algunos procesos y se describe la interacción de los componentes en cada uno.

Como se puede desprender del diagrama presentado el usuario accede a múltiples aplicaciones web:

1. Aplicación web: Carga, actualización y reportes.
2. Documentación en línea sobre la plataforma.
3. Saiku (PBI CE): Realización de análisis dinámicos.

Para la gestión de las fuentes de datos que se utilizarán para cargar el DW la aplicación web se comunica con la base de datos mediante una capa de servicios realizada con PostgREST.

Para la carga y actualización de datos del DW, iniciada por el usuario, en primer paso carga al sistema de gestión de archivos las fuentes de datos a utilizar por los ETL. Luego, mediante la aplicación web, realiza la ejecución de los ETL cargados en PDI utilizando los servicios provistos con el servidor Carte. Los ETL se comunican con el sistema de archivos para tomar las fuentes de datos cargadas por el usuario, dado que no es posible cargarlos a los ETL mediante Carte, y con la base de datos fuentes previamente cargadas por el usuario. Finalmente el ETL impacta los datos en la base de datos del DW.

Para la generación del reporte de BI sobre la sección del Informe Anual, iniciada por el usuario desde la aplicación web, se ejecuta un ETL que se comunica con la base de datos del DW y con el servidor de R para obtener los datos necesarios y realizar los cálculos estadísticos respectivos. Finalmente se genera el archivo PDF con el reporte y se lo envía al usuario por correo.

A su vez, con el objetivo de brindar una mayor visibilidad y claridad en la interacción entre los distintos componentes en el momento de ejecución de ETL, se presenta un Diagrama de Secuencia de Sistema (DSS). En la fig. 5.3 se ilustra el flujo de interacción para la carga del DW:

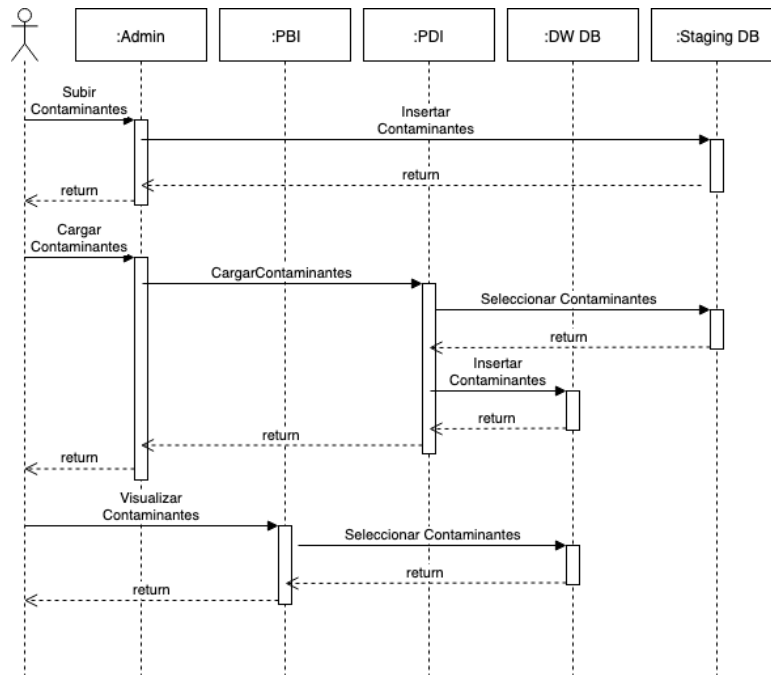


Figura 5.3: Flujo

5.2. Monitoreo

Dentro de las operaciones del sistema requeridas para mantener el sistema de BI en correcto funcionamiento es posible monitorizar los distintos componentes de la solución.

5.2.1. Herramientas

Se presentan distintas herramientas que sirven a monitorizar los distintos componentes según su naturaleza.

5.2.1.1. Portainer

Esta aplicación permite gestionar de forma muy fácil e intuitiva los contenedores Docker a través de una interfaz gráfica web. Mediante la misma un administrador puede tener una visión global más clara de los contenedores que está ejecutando y facilitar su gestión.

A continuación, en fig. 5.4, se presenta la pantalla de inicio de la herramienta donde se muestra el estado de cada componente:

A su vez, y como se puede visualizar en fig. 5.5, es posible visualizar los logs de cada uno de los componentes, desde el servidor PBI, PDI hasta la base de datos PostgreSQL:

5.2.1.2. Grafana + Prometheus

Prometheus es una plataforma de monitoreo de código abierto que es capaz de recopilar métricas de objetivos monitoreados al consultar las métricas.

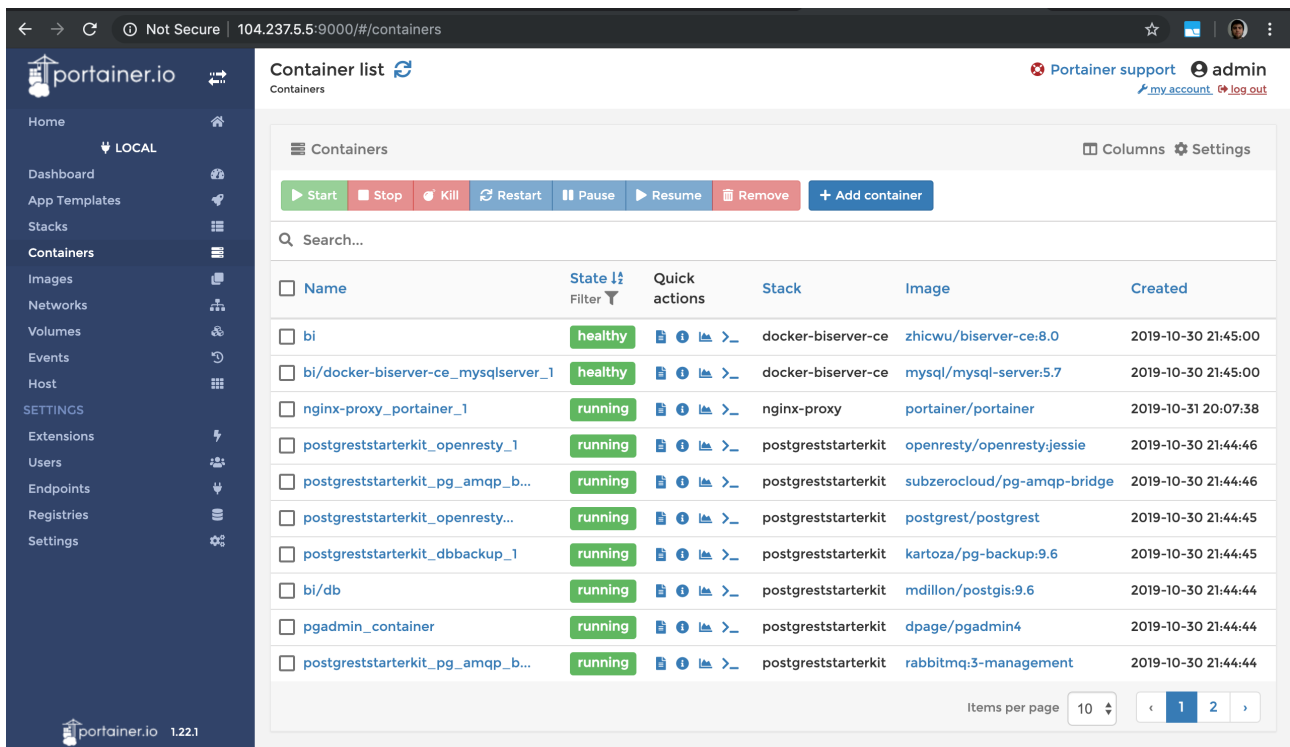


Figura 5.4: Monitoreo - Portainer

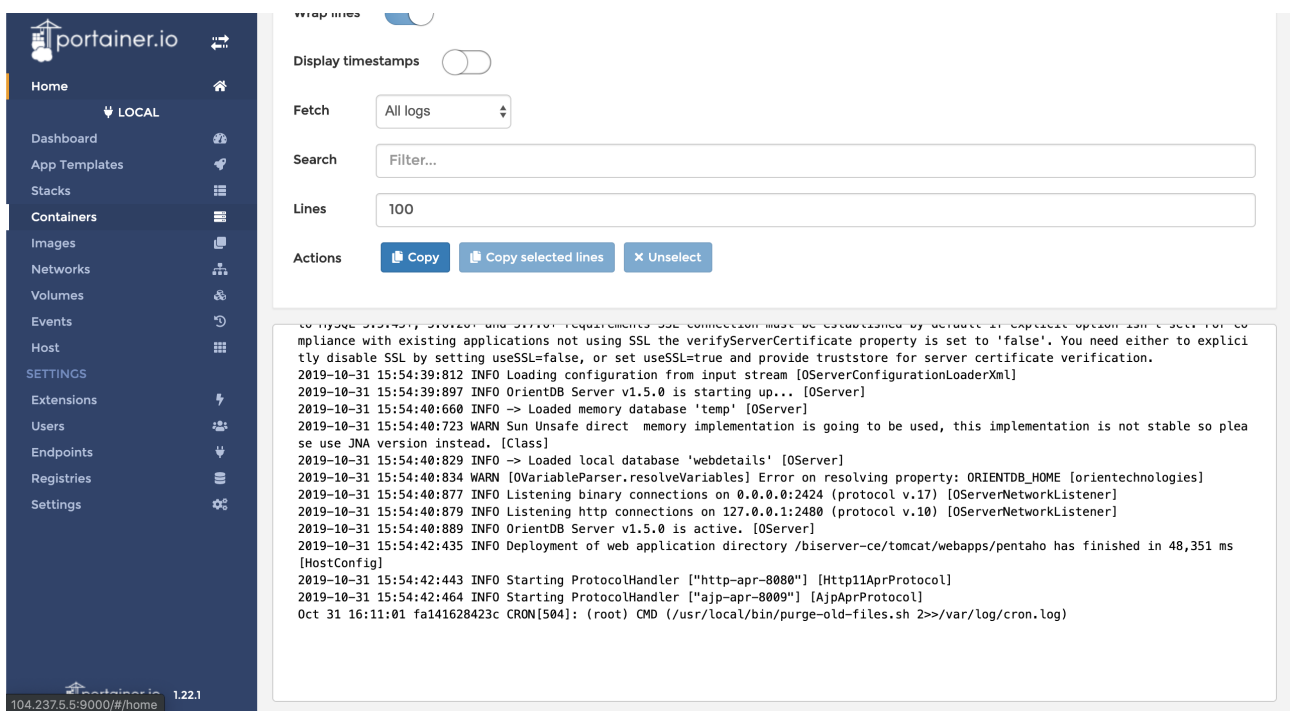


Figura 5.5: Monitoreo - Portainer - Logs

Después de guardar los datos recopilados, Grafana provee una aplicación web donde se puede consultarlos utilizando su lenguaje de consulta y representar todos los resultados en gráficos.

A continuación, en fig. 5.6, se presenta la pantalla de inicio de la herramienta en la cual se muestran indicadores del Docker Host, esto es, la máquina virtual que corre los contenedores:

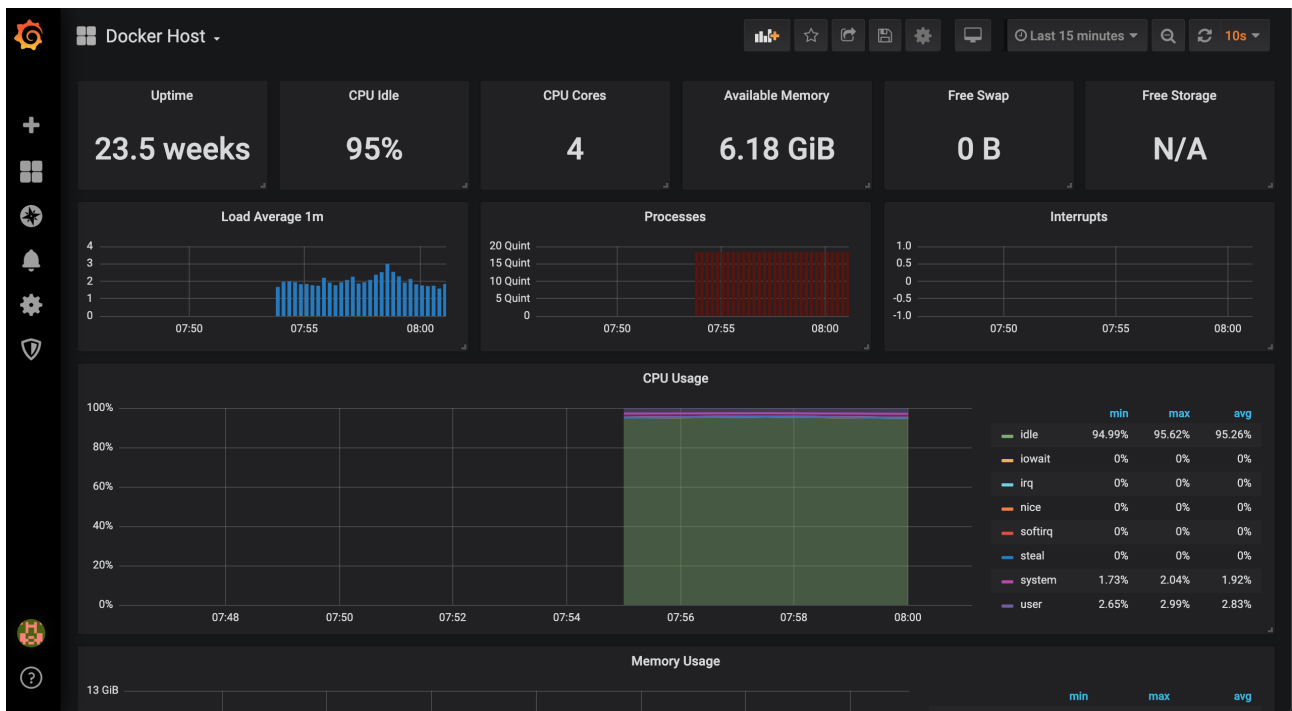


Figura 5.6: Monitoreo - Grafana + Prometheus

Luego, en fig. 5.7, se muestran indicadores de cada uno de los contenedores, aquí se puede ver que el contenedor que utiliza más memoria RAM es el servidor BI:

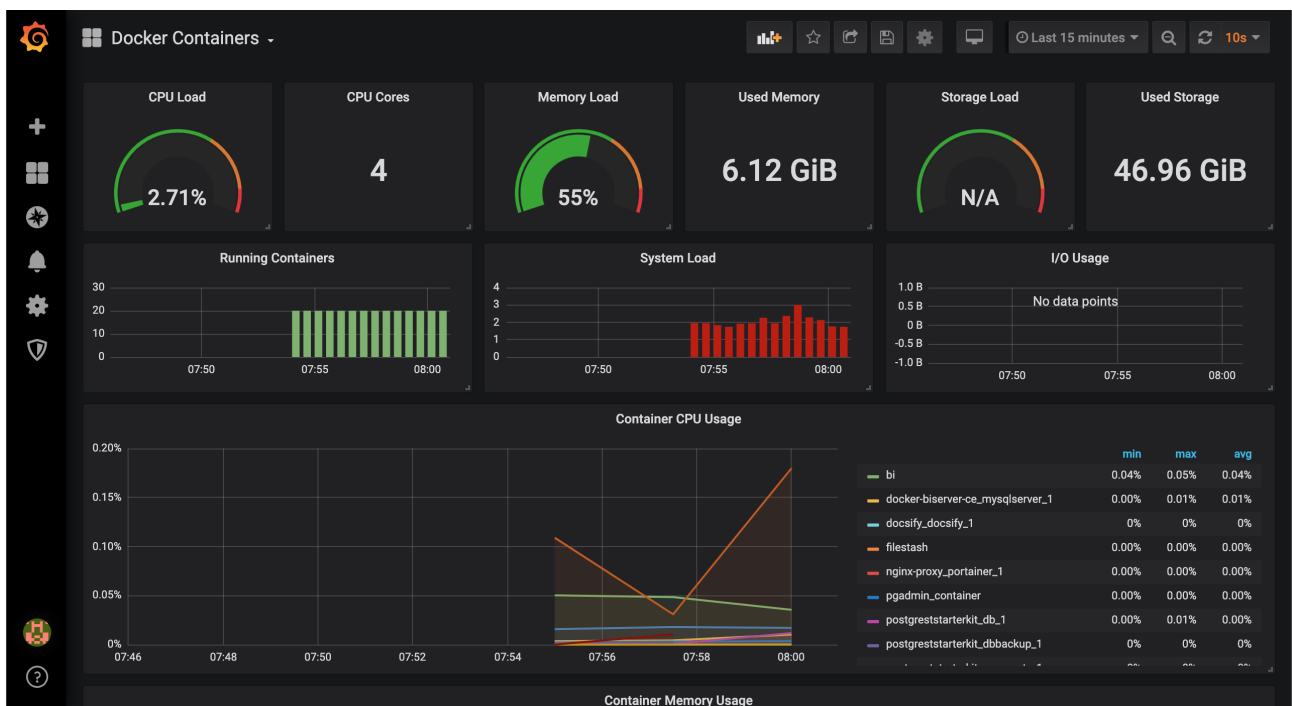


Figura 5.7: Monitoreo - Grafana + Prometheus - Containers

5.3. Métricas

Para poder visualizar y entender cómo los especialistas de la IM están utilizando la herramienta y las funcionalidades brindadas, se utilizan plugins Open Source de Pentaho de analítica web para luego, basado en dicho entendimiento, investigar, decidir y aplicar las oportunidades de mejora correctiva y evolutiva definidos en el sistema de BI.

5.3.1. Herramientas

Se presentan distintas herramientas que sirven a auditar y presentar métricas de uso de la ejecución de análisis realizados en el servidor de BI.

5.3.1.1. Pentaho CE Audit [38]

Total control sobre quién y ingresó a la herramienta y cuándo lo hizo. Además se puede auditar a qué recursos se accedió y cuándo se realizó. En fig. 5.8 se puede visualizar un ejemplo:

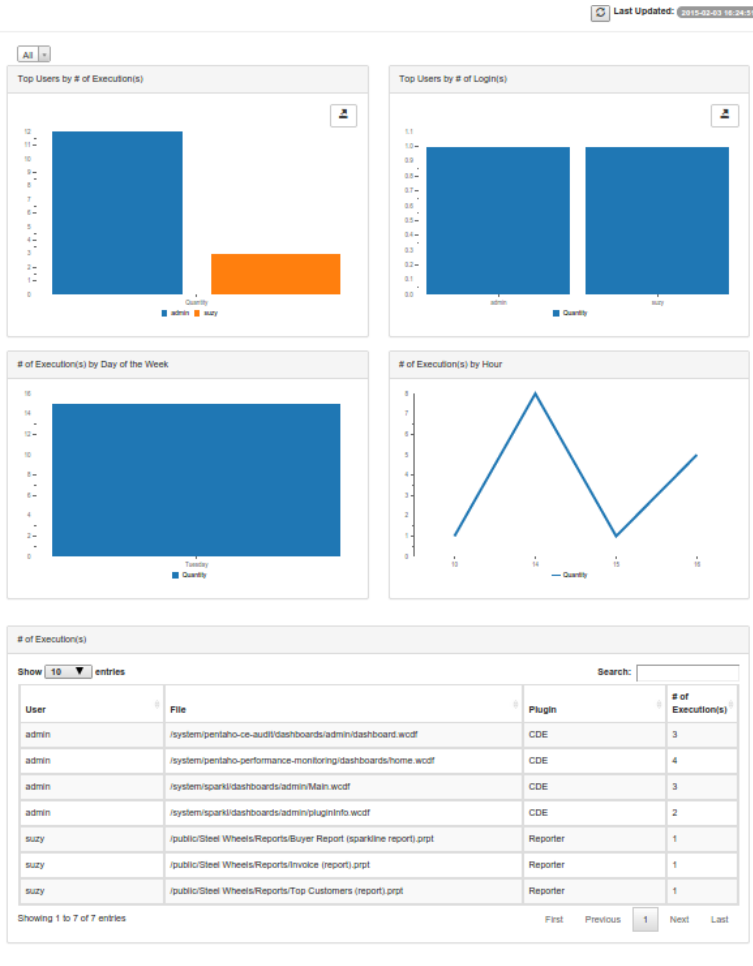


Figura 5.8: Métricas - Pentaho CE Audit

5.3.1.2. Pentaho Performance Monitoring [39]

Total control sobre qué consultas son las más realizadas, cuáles son las más rápidas y cuáles las más lentas. En fig. 5.9 se puede visualizar un ejemplo:



Figura 5.9: Métricas - Pentaho Performance Monitoring

