[[1]](#footnote-1)

Comparación de la capacidad de respuesta de la base de datos (BD) entre PostGreSQL y el Stack ELK en un entorno Docker

Lic. Jonathan Fernando Romano - Ing. Agr. Juan Manuel Alonso  
*Junio 2021*

*Resumen*—Se analizó una base de datos de 355.000 registros de Documentos de tránsito vegetal (DTV-E) emitidos por el Servicio Nacional de Sanidad y Calidad Agroalimentaria (SENASA) de Argentina para el traslado de plantas de vivero. Se montó un entorno Docker en Linux a los efectos de comparar la capacidad de respuesta de los motores de base de datos PostGreSQL y Stack ELK en función de consultas de mayor y menor complejidad.

*Index Terms*—Enter key words or phrases in alphabetical order, separated by commas. For a list of suggested keywords, send a blank e-mail to [keywords@ieee.org](mailto:keywords@ieee.org) or visit <http://www.ieee.org/organizations/pubs/ani_prod/keywrd98.txt>

# INTRODUCCION

E

n el marco del curso de posgrado “Captura y Almacenamiento de información” dictado por la Facultad de Informática de la Universidad de La Plata, se realizó el presente trabajo de investigación a los efectos de aplicar los conocimientos adquiridos durante la cursada. En este sentido, se tomó como punto de partida una base de datos del Programa Nacional de Sanidad de Material de Propagación del Servicio Nacional de Sanidad y Calidad Agroalimentaria (SENASA) perteneciente a movimientos de plantas de vivero del año 2016 a 2021 por todo el territorio nacional argentino.

El sistema Docker es un proyecto de código abierto que automatiza el despliegue de aplicaciones dentro de contenedores de software. Permite un montaje ágil de los diferentes entornos y configuraciones, lo que hace esta herramienta ideal para establecer las comparaciones deseadas. Se montó la base de datos en una distribución de GNU Linux utilizando un entorno Docker para comparar los tiempos de respuesta de distintos motores de BD, en este caso PostGreSQL y Stack ELK.

Durante el desarrollo de este trabajo, se describirá la base de datos a analizar, como fue el montaje del entorno Docker y sus complementos, intentando detallar los aciertos y desaciertos que tuvimos durante todo el proceso.

# OBJETIVO

-Evaluar el rendimiento en tiempos de respuesta entre un DBMS (Sistema de gestión de Base de datos) PostgreSQL y el Stack ELK (Elasticsearch, Logstash, Kibana), en el procesamiento de consultas de diferente complejidad y tamaño de BD.

-Montar un entorno con Docker Compose para realizar esta comparativa, y aplicar los conocimientos adquiridos en la materia de Almacenamiento y Captura de la Información de la UNLP.

# MATERIALES Y METODOS

## Base de datos (BD)

La base de datos a analizar corresponde a 355.000 registros distribuidos en 21 campos de datos de Documentos de tránsito Vegetal Electrónicos (DTV-e) emitidos por viveristas de la Argentina en cumplimiento a normativas nacionales y con el fin de amparar el tránsito de estos materiales entre distintos puntos del país.

El archivo original importado esta en formato . CSV (del inglés comma-separated values) y tiene un peso de 109 MB (114.644.979 bytes).

## Configuración y preparación del entorno.

En primer lugar se pensó en montar una máquina virtual con sistema operativo Ubuntu Linux, pero por limitaciones de hardware se tuvo que migrar el montaje a una partición de disco duro e instalación convencional. A partir de allí se pudieron correr los procesos sin inconvenientes y a una velocidad considerablemente mayor.

Durante el proceso aparecieron algunas incompatibilidades que tuvimos que superar, por ejemplo la versión de Java que teníamos instalada en la distribución de Linux era diferente a la que necesita el Logstash. Aunque teníamos una versión más moderna, hubo que hacer un downgrade a la versión JDK 8 para que todo funcione. Este tipo de inconvenientes nos hicieron perder bastante tiempo, en contrapartida buscando la solución logramos un mayor entendimiento de todo el paquete ELK.

Otro punto importante, fue la carga de la base de datos PosGreSQL en el entorno Kibana con Elastic Search y el armado de los archivos de configuración correspondientes. Una vez montada la base de datos en PostGreSQL y en ELK, el manejo de los entornos web en localhost es sencillo.

El paso a paso detallado del montaje del entorno Docker, como de los paquetes de Docker compose, PostGreSQL, PgaAdmin4, Elastic Search, Logstash y Kibana, se pueden consultar en el readme.md disponible en: <https://github.com/warasoft/tp_final#readme>

## Matriz de comparación de tiempos de respuesta.

Se elaboró una matriz de 3 x 3 para evaluar la respuesta de los distintos motores en función de 3 consultas de menor a mayor complejidad y 3 tamaños de la base de datos al 25%, 50% y 100% .

La consulta de menor complejidad, es un filtrado por tipo de movimiento en donde se seleccionan todos los documentos emitidos bajo el tipo "Vivero - Vivero".

La consulta de mediana complejidad, tiene dos condiciones a cumplir, debe mostrar todos los documentos emitidos bajo el tipo "Vivero - Vivero" con origen la provincia de Buenos Aires.

Por último, la consulta de alta complejidad debe cumplir las condiciones anteriores pero solo para el vivero con número de RENFO “BAI00037-A”.

La elección de diferentes tipos de consultas y tamaños de BD busca abarcar la mayor cantidad de situaciones posibles, de forma tal de poder determinar el comportamiento de cada una de las herramientas de forma abarcativa y no con un sesgo que pueda favorecer una u otra.

# RESULTADOS

A continuación en la Tabla 1 se exponen los resultados de los tiempos de respuesta para motores de búsqueda analizados según las condiciones establecidas en la matriz.

TABLa I

resultados de performance de las herramientas elegidas para las diferentes condiciones

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Tamaño BD | 25% | | 50% | | 100% | |
| Consulta | PGS | Esh | PGS | Esh | PGS | Esh |
| Simple |  |  |  |  |  |  |
| Media |  |  |  |  |  |  |
| Alta |  |  |  |  |  |  |
| \*PGS= Posgres, Esh= Elastic Search | | | | | | |

# CONCLUSIONES

A medida que vamos hacia las últimas versiones de este stack de herramientas requieren más hardware.

References

## Cotten. (2021 , May . ). Elastic stack (ELK) on Docker. GitHub. [Online].

Available: [https://github.com/deviantony/docker-elk](http://www.amdahl.com/doc/products/bsg/intra/infra/html)

Autores

**ALONSO, Juan Manuel.** Ingeniero agrónomo, recibido en la Universidad de Morón en el año 2012 y con un posgrado en Alta Dirección de Agronegocios y Alimentos de la Facultad de Ciencias Agrarias de la Universidad de Buenos Aires en el año 2018. Desde el año 2020, se encuentra estudiando la Maestría de inteligencia de datos orientada en Big Data de la Universidad Nacional de La Plata.

Actualmente trabaja como Analista Profesional en la Dirección Nacional de Protección Vegetal del SENASA en la coordinación del Programa Nacional de Sanidad de Material de propagación Vegetal.

**ROMANO, Jonathan Fernando.** Licenciado en Informática, recibido en la Universidad de Palermo, año 2013. Actualmente se encuentra cursando la Carrera de Esp. en Inteligencia de Datos Orientada a Big Data.

Trabaja en la ARMADA ARGENTINA, como Ayudante del Jefe de División Sistemas del Servicio Administrativo Financiero de la Armada.

1. [↑](#footnote-ref-1)