



ESPE

UNIVERSIDAD DE LAS FUERZAS ARMADAS

INNOVACIÓN PARA LA EXCELENCIA

PROYECTO DE TITULACIÓN

Presentado por:

-Peña Kevin

-Simbaña Estefanía

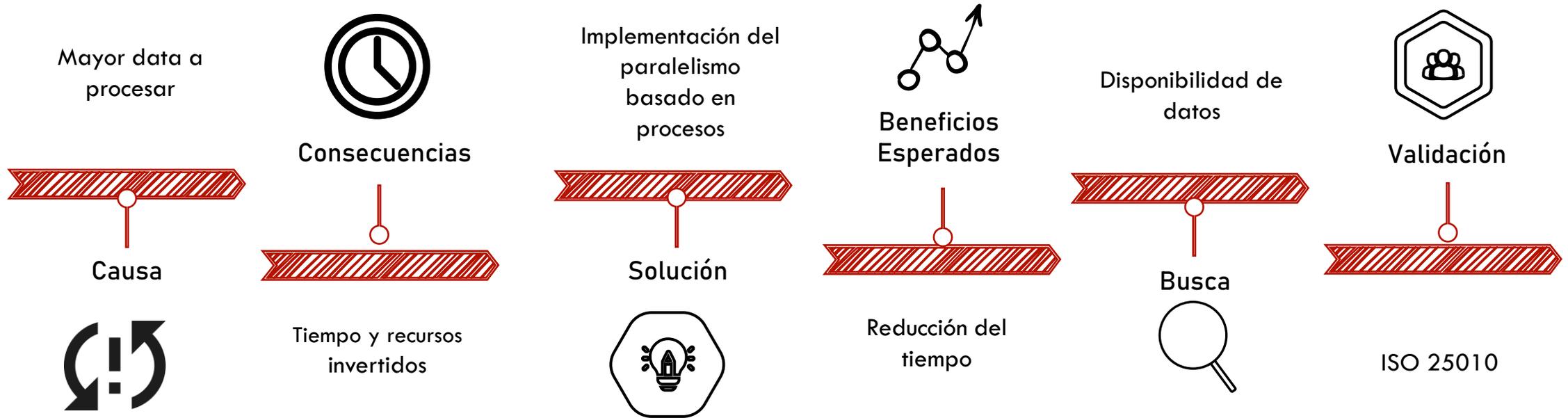
AGENDA

01. Planteamiento del problema
02. Objetivos
03. Estado del arte
04. Desarrollo
05. Análisis y resultados
06. Conclusiones y recomendaciones



01. Planteamiento del problema

Motivación y Contexto



“

Optimización del proceso ETL en Big Data de la empresa “Casa de Incentivos Casintour” mediante técnicas de paralelismo basado en procesos durante el año 2023

”

02. Objetivos

OBJETIVO GENERAL

Proponer una estrategia de implementación que aproveche el paralelismo basado en procesos para mejorar la eficiencia y el rendimiento del ETL en Big Data de la empresa “Casa de Incentivos Casintour” durante el año 2023.

02. Objetivos

Objetivos específicos

1. Investigar y comprender los conceptos y principios fundamentales del proceso ETL en entornos de Big Data, así como las técnicas y herramientas del paralelismo para la optimización de dicho proceso.
2. Analizar en profundidad la infraestructura actual de procesamiento de datos en la empresa 'Casa de Incentivos Casintour', tanto la arquitectura de hardware como software utilizado para el proceso ETL en Big Data.

02. Objetivos

3.

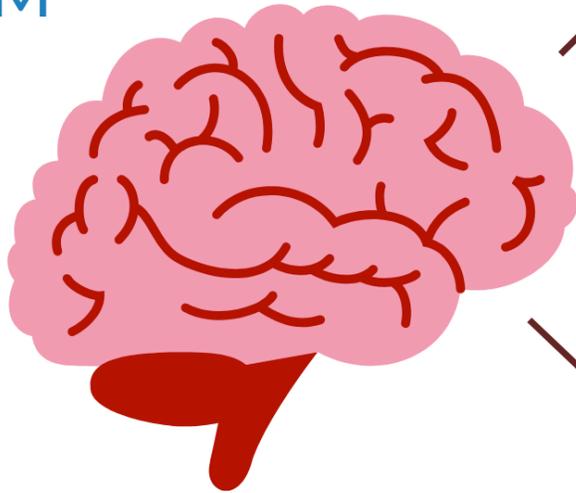
Desarrollar un prototipo de la implementación propuesta, con tecnologías y herramientas adecuadas para habilitar el procesamiento paralelo de datos en ETL.

4.

Recopilar datos relevantes sobre el rendimiento, la escalabilidad y la eficiencia del proceso mejorado y compararlo con el proceso ETL tradicional para validar los resultados obtenidos y analizar el impacto de la optimización en términos de tiempo de ejecución, consumo de recursos y calidad de los datos procesados.

03. Estado del arte

Trabajos relacionados



1

Mejores prácticas de ETL (extraer, transformar, cargar)

(Seenivasan, 2023)

- Destaca la importancia del procesamiento paralelo para distribuir la carga de trabajo, reduciendo el tiempo de procesamiento y aumentando la eficiencia.

2

Mejora del rendimiento de los trabajos ETL

(Seenivasan, 2023)

- Implementación de diversas estrategias de paralelismo según los recursos disponibles.

3

Marco ETL de próxima generación para abordar los desafíos que plantean los Big Data

(Ali, 2018)

- Efectividad en flujos de trabajo ETL mediante paralelismo considerando el rendimiento.

04. Desarrollo

LUGAR

- Una de las empresas más grandes del Ecuador en Servicios Integrales de Turismo
- Consolidadora para emisión

M&M® | Group



7 años
Promedio de
antigüedad



Boletos aéreos

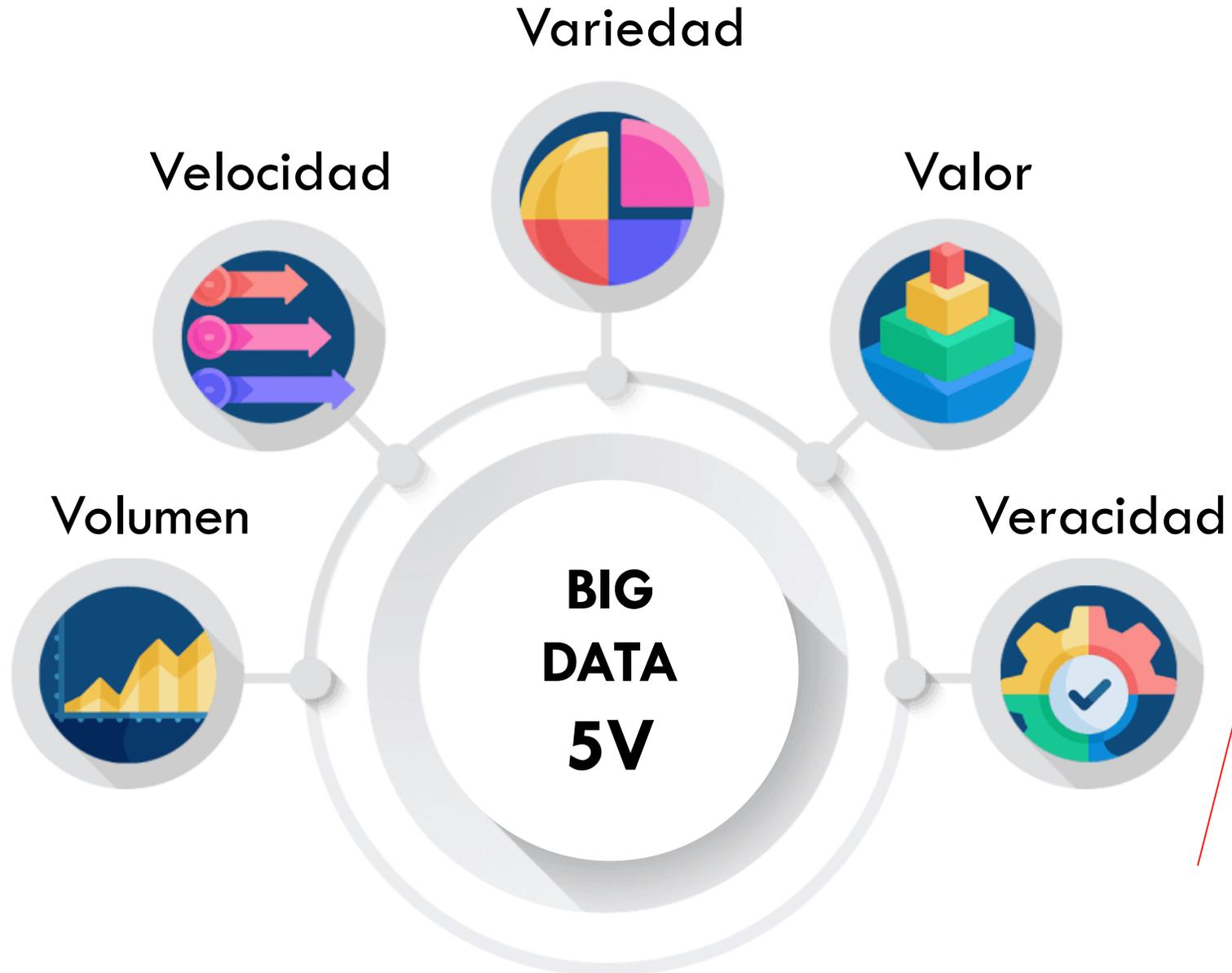
248 Operadoras



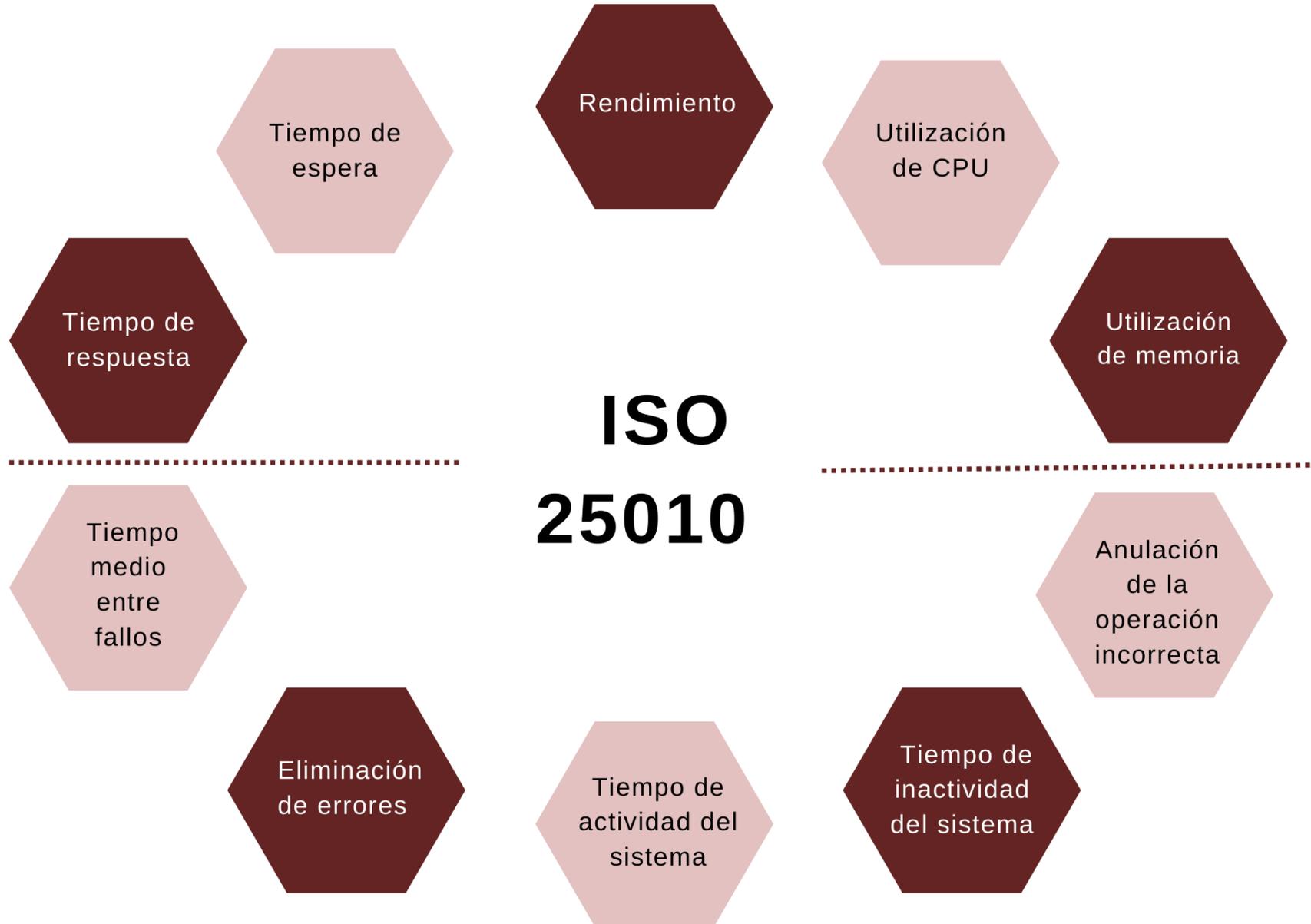
Turísticas

04. Desarrollo

Bourany, T, 2018



04. Desarrollo



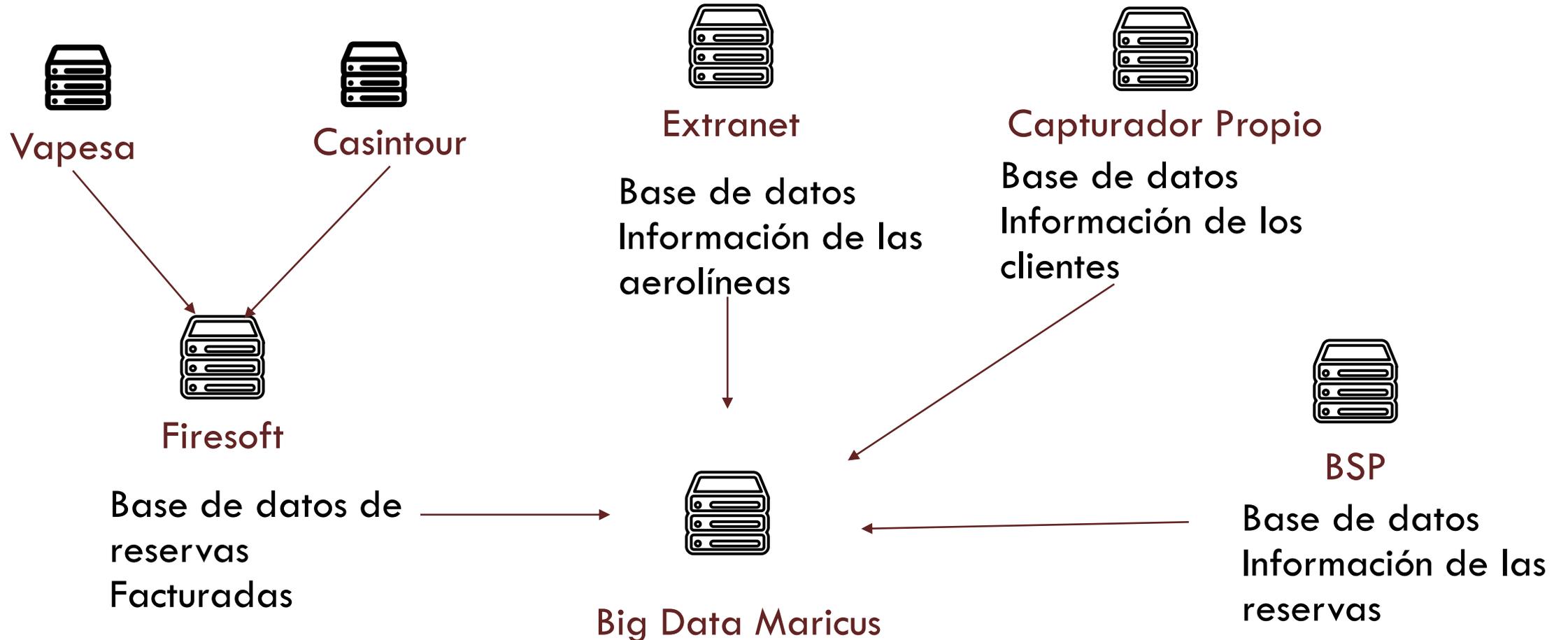
04. Desarrollo

Infraestructura



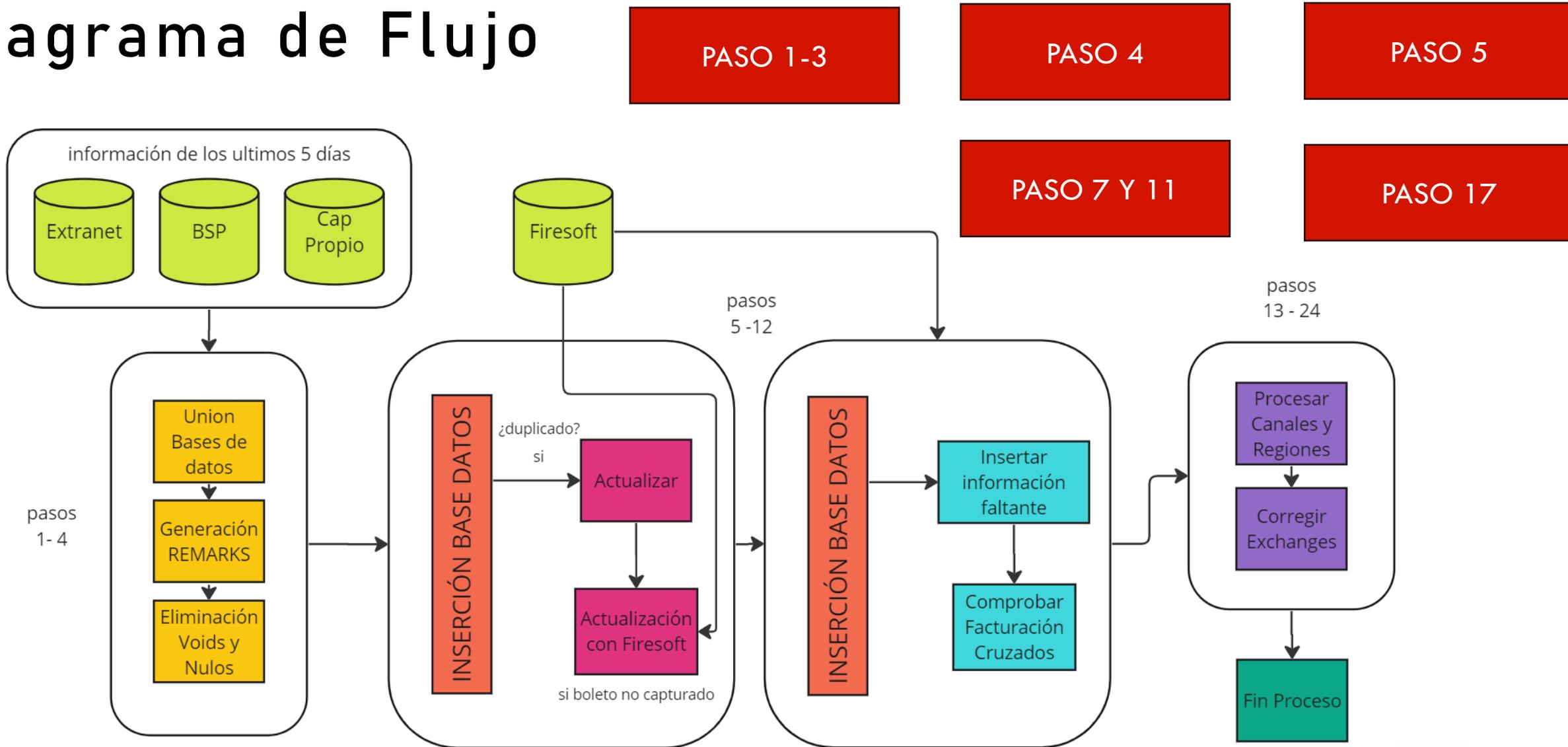
CentOS

8 vCPU - 4 core
64 GB



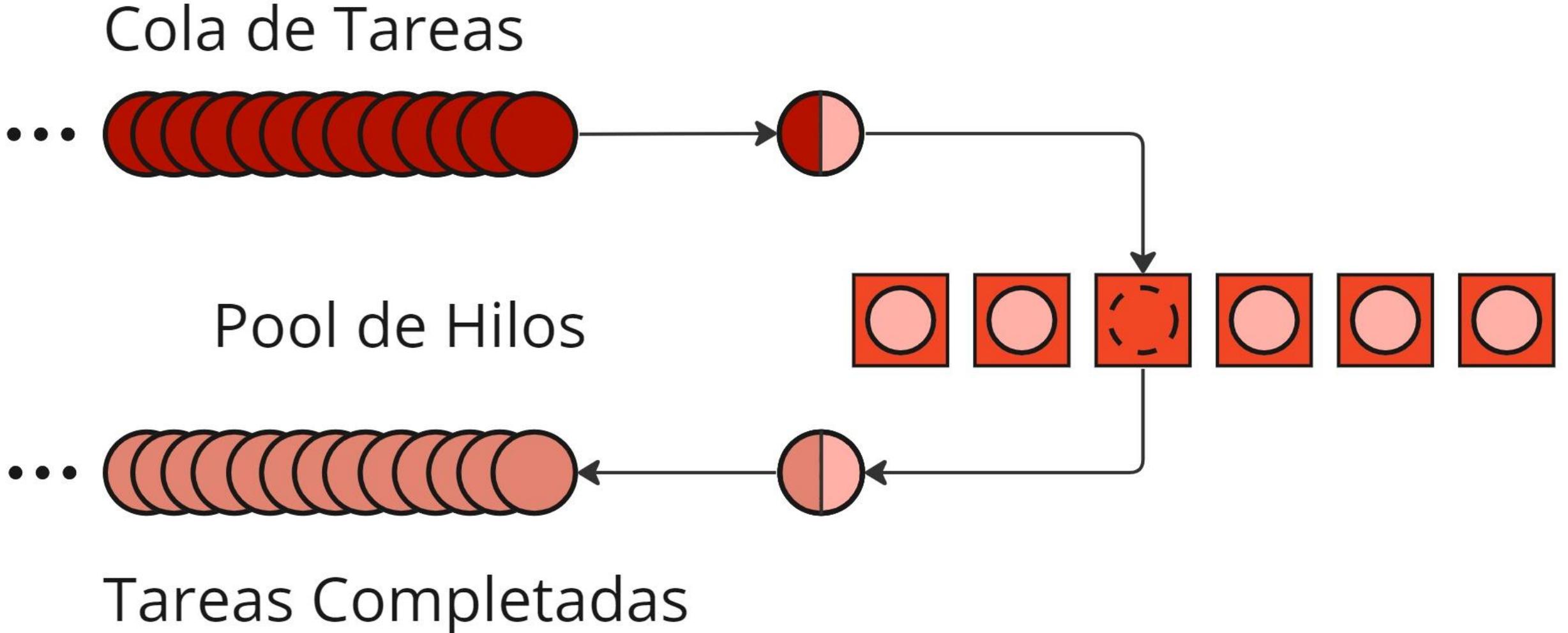
04. Desarrollo

Diagrama de Flujo



04. Desarrollo

Paralelismo basado en procesos



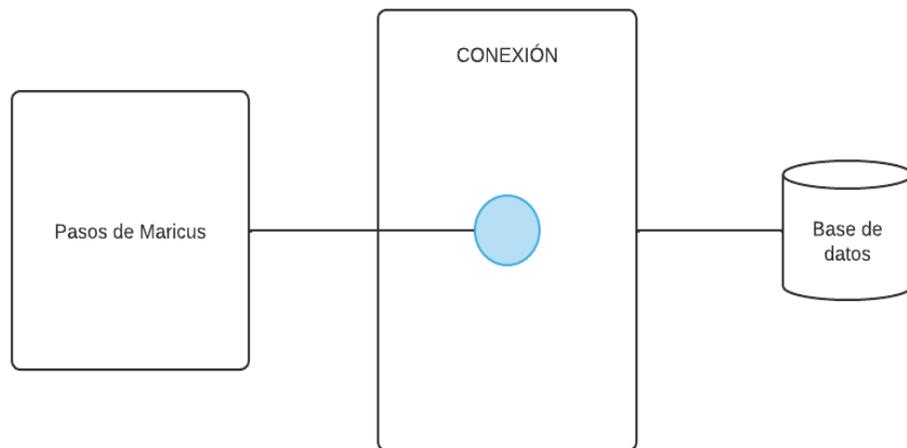
04. Desarrollo

Maricus Original

Motor de almacenamiento

InnoDB

Conexión

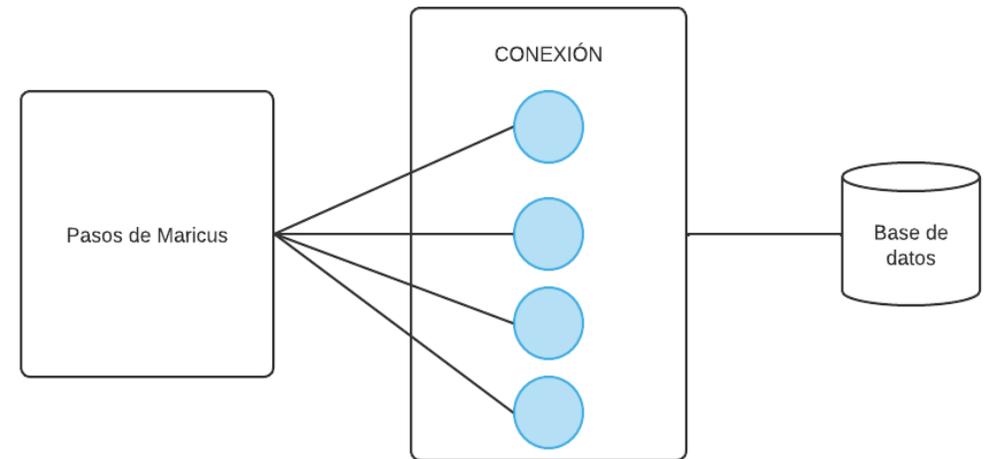


Maricus con cambios

Motor de almacenamiento

MyISAM

Conexión



05. Análisis y resultados

RESULTADOS

Puntuación global



Original

3.73



Con cambios

7.30

Mejora en
Tiempo

37.16%

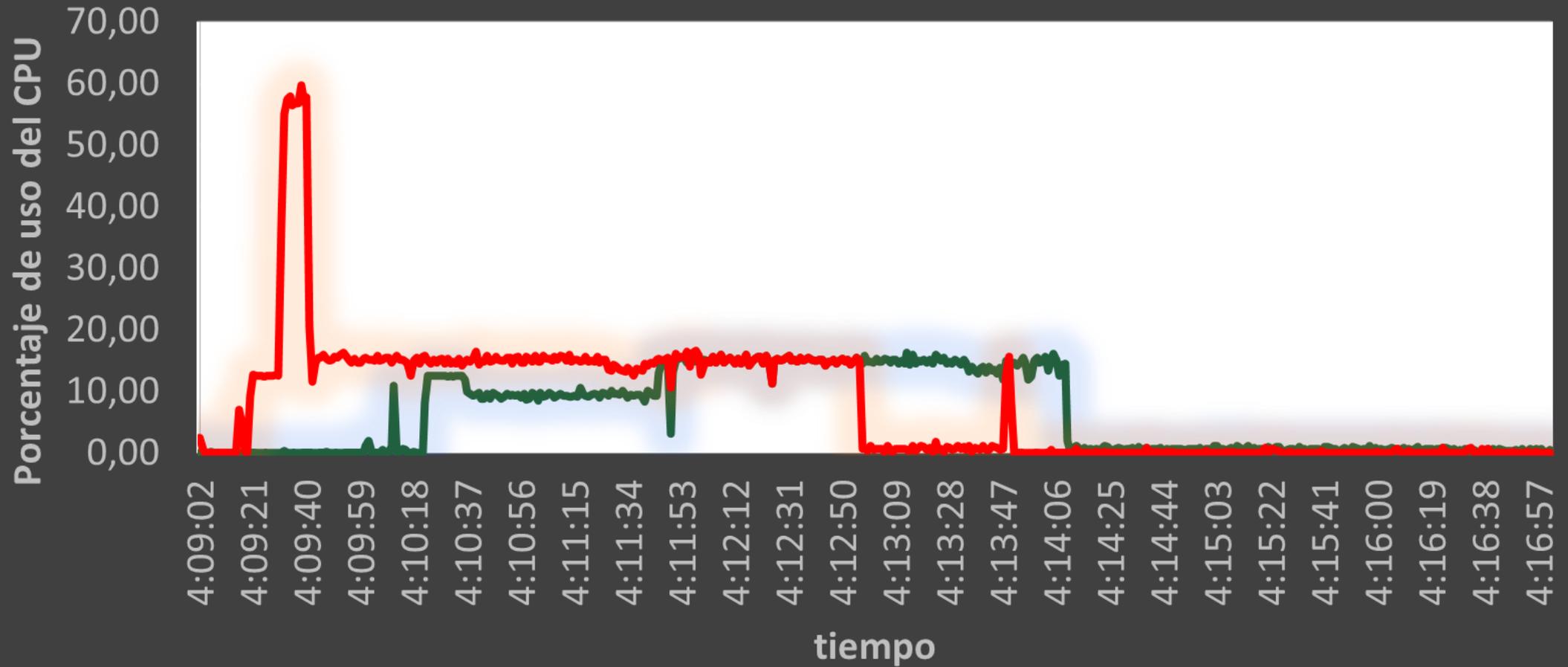
Mejora ISO

95.9%

Parámetros	Maricus original	Maricus con Cambios
Tiempo de espera	7 min 24 seg	4 min 39 seg
Utilización de CPU	6.41% - 0.073	8.43 % - 0.084
Utilización de memoria	1.53 GB - 0.0250	1.55 GB - 0.0254

Utilización de CPU

— CPU ANTES — CPU DESPUES



Maricus Original

```
orte-maricuz]$ python ProcesadorMaricuz.py
- Logs Maricuz - INFO      - Enabled Modo DEV
- Logs Maricuz - INFO      - INICIO DEL PROCESO DEL REPORTE
- Logs Maricuz - INFO      - Fecha de proceso BETWEEN "2023-04-01" AND "2023-04-22"
- Logs Maricuz - INFO      - Conectando a Bases de Datos.
- Logs Maricuz - INFO      - 1. Consultar remarks corporativos
- Logs Maricuz - INFO      - - Numero de registro: 81
- Logs Maricuz - INFO      - 2. Consultar al BSP.
```

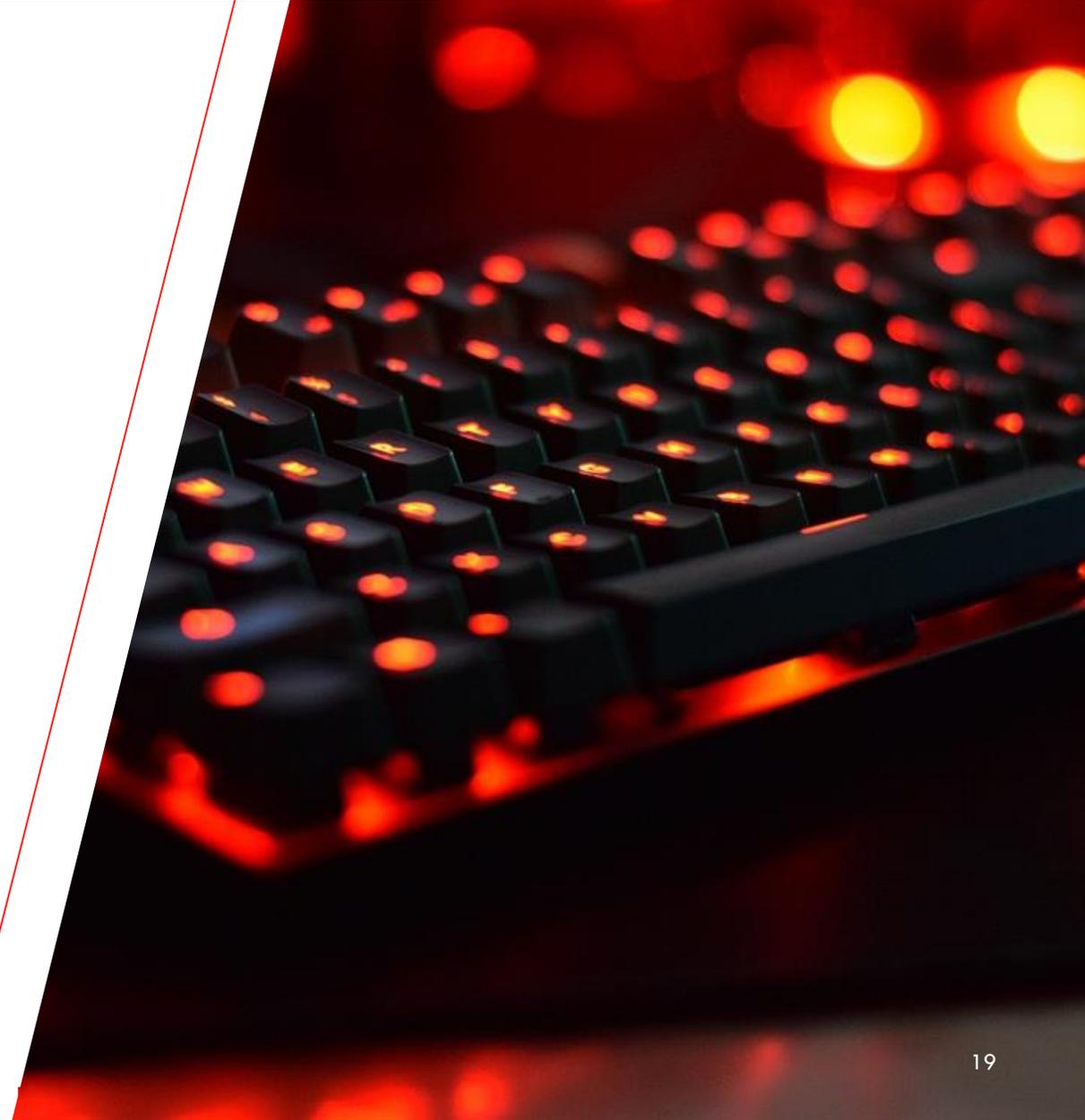
Maricus con cambios

```
e-maricuz]$ python ProcesadorMaricuz.py
```

06. Conclusiones y recomendaciones

CONCLUSIÓN

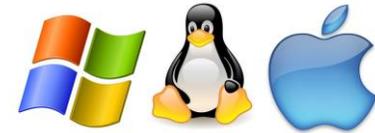
- Paralelismo basado en procesos
- Mejora del rendimiento en tiempo del 37.16%
- ISO 25010: Eficiencia del desempeño y fiabilidad



06. Conclusiones y recomendaciones

Recomendaciones

- Número de hilos.
- Considerar el sistema operativo.
- Utilización de MyISAM
- Las sugerencias y buenas prácticas proporcionadas por la documentación de las librerías.



TRABAJOS FUTUROS

MUCHAS GRACIAS

