

# DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN

## “Evaluación comparativa de rendimiento de PostgreSQL y Cassandra en operaciones CRUD”

### **Autores:**

Jaramillo Sandoval, Angélica Alexandra  
Puchaicela Soliz, Wendy Lisseth

### **Director:**

Ing. Coronel Guerrero, Christian Alfredo, Mgtr.



# AGENDA

01

## INTRODUCCIÓN

- Problemática
  - Objetivos
  - Marco Teórico
- 

02

## IMPLEMENTACIÓN

- Entorno de Trabajo
- 

03

## PRUEBAS Y RESULTADOS

- Análisis de Fase de Carga
  - Análisis de Fase Transaccional
  - Análisis de Rendimiento
- 

04

## FINALIZACIÓN

- Conclusiones
- Recomendaciones



01

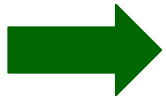
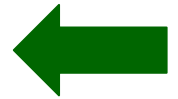
# INTRODUCCIÓN



**ESPE**  
UNIVERSIDAD DE LAS FUERZAS ARMADAS  
INNOVACIÓN PARA LA EXCELENCIA

# PROBLEMÁTICA

Las organizaciones requieren de unas bases de datos que cumplan con los requerimientos funcionales del cliente, que cuenten con un alto rendimiento, que permita la gestión de procesos de manera eficaz.



Las bases no relacionales pueden ser una **alternativa de solución para las organizaciones con grandes volúmenes de datos**. En base a ello, se implementó un estudio comparativo entre PostgreSQL (SGBD) y Apache Cassandra (NoSQL).

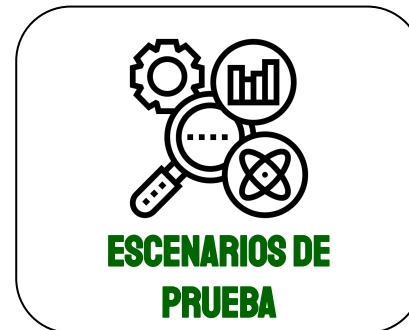
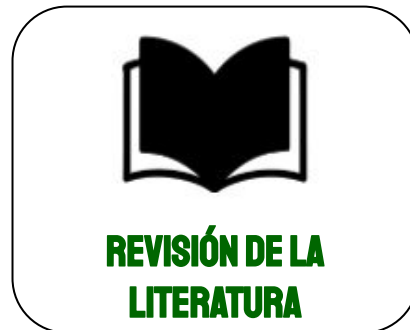


# OBJETIVOS

## Objetivo General

- Evaluar una comparativa de rendimiento de PostgreSQL y Cassandra en operaciones CRUD

## Objetivo Específicos



# MARCO TEÓRICO

**BASES DE DATOS  
RELACIONALES**



PostgreSQL

**POSTGRESQL  
VS  
CASSANDRA**

**BASES DE DATOS NO  
RELACIONALES**



*cassandra*

**YCSB**



**ESPE**  
UNIVERSIDAD DE LAS FUERZAS ARMADAS  
INNOVACIÓN PARA LA EXCELENCIA

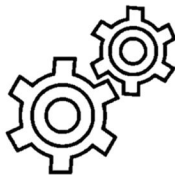
# POSTGRESQL VS CASSANDRA

**Tabla 1.** Comparativa de funciones de PostgreSQL y Cassandra

Funciones	PostgreSQL	Cassandra
Lenguaje de consulta	SQL	CQL
Modelo de datos	Relacional Base de datos	No relacional Base de datos Columnas
Ordenación rápida de resultados	Si	Si
Consistencia inmediata	Si	Configurable a nivel de consulta
Escrituras de alto rendimiento	No	Si
Fragmentación transparente	No	Si
Uniones relacionales	Si	No
Índices secundarios	Si	Si
Registros Estructurados	Si	Si
Colecciones de escritura discreta	Si	Si
Ningún Punto único de falla	No	Si



# YAHOO CLOUD SERVING BENCHMARK (YCSB)



Ofrece  
resultados  
concisos en  
sus  
comparaciones

Permite realizar  
pruebas

Brinda sus  
resultados en  
formatos de  
texto

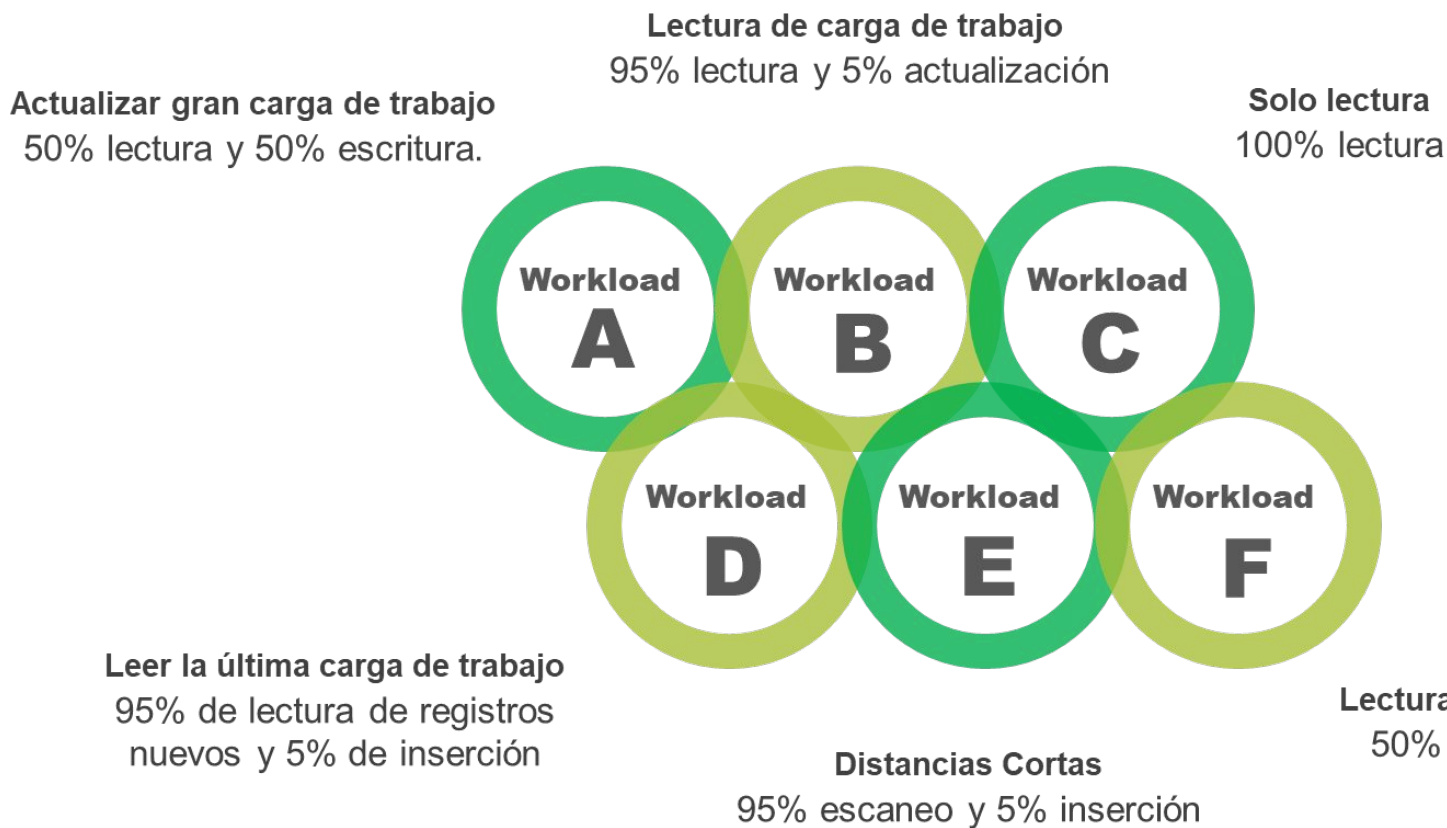
Resume el  
rendimiento de  
latencia y  
tiempo de  
ejecución



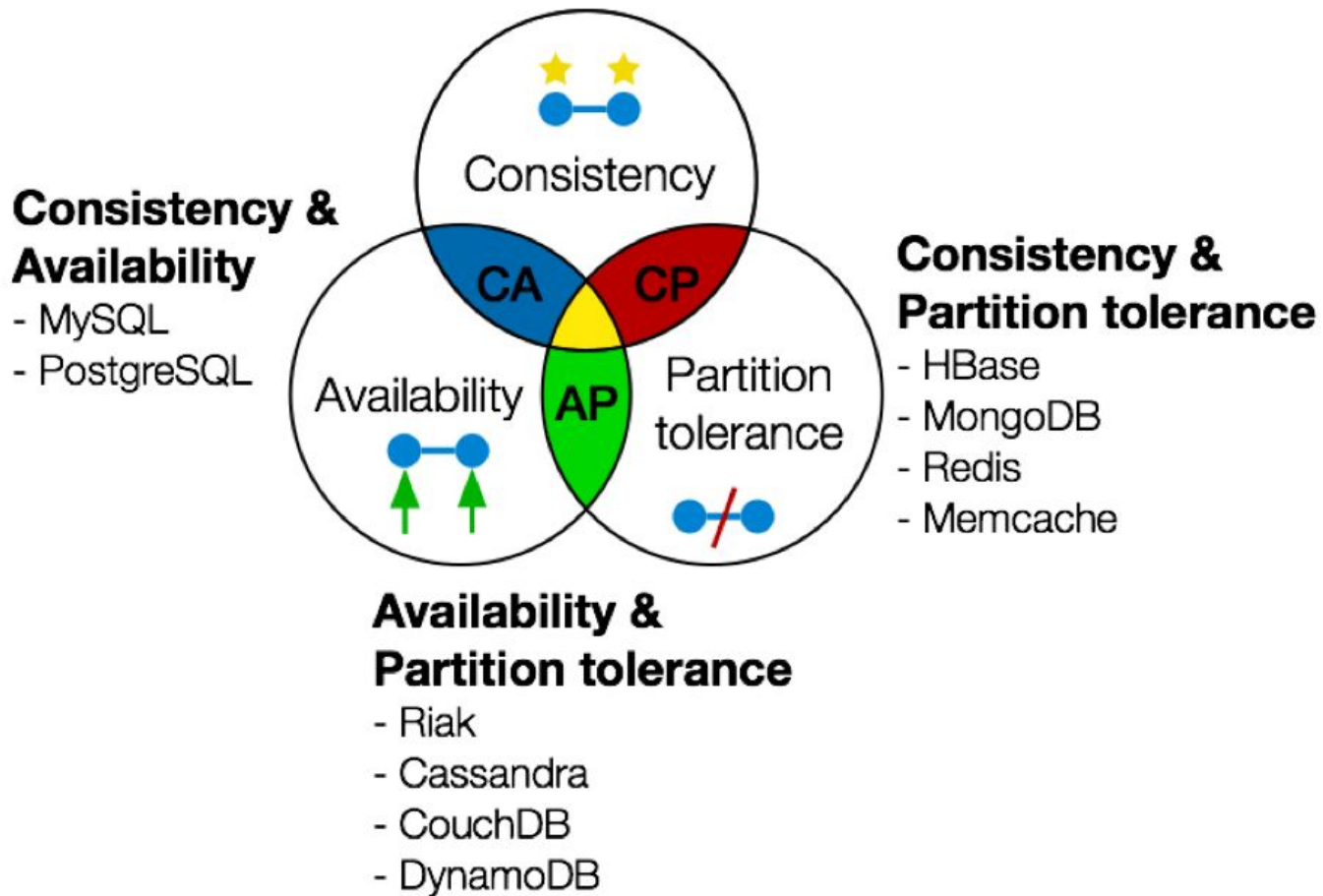
**ESPE**  
UNIVERSIDAD DE LAS FUERZAS ARMADAS  
INNOVACIÓN PARA LA EXCELENCIA



# YAHOO CLOUD SERVING BENCHMARK (YCSB)



# TEOREMA CAP



# 02 IMPLEMENTACIÓN



# METODOLOGÍA

Desarrollo del modelo  
YCSB

Entorno de trabajo

Escenario de trabajo  
POSTGRESQL - CASSANDRA



# DESARROLLO DEL MODELO YCSB

## FASE DE CARGA

- Recordcount
- Threads
- S
- Comando de carga

## FASE TRANSACCIONAL

- Operationcount
- Threads
- S
- Comando de carga

```
\bin\ycsb load [GESTOR A UTILIZAR] -P .\workloads\workloada -P  
.\[DIRECCION DE ARCHIVO DE PROPIEDADES] >  
[ARCHIVO DONDE GUARDA LOS DATOS].txt
```



# ENTORNO DE TRABAJO

## PROCESADOR CORE I7

---

### *Especificaciones del procesador core i7*

---

**Procesador** Intel Core I7-6700 CPU (2.80 GHz 2.81 GHz)

**RAM** 8,00 GB DDR4 (7,88 GB usable)

**SO** Windows pro de 64 bits, procesador basado en x64

---

## RIG DE PROCESAMIENTO

---

Especificaciones del dispositivo	Características
Procesador	AMD Ryzen Threadripper 2920X 12-Core Processor (3.50 GHz)
RAM	16,00 GB DDR4
SO	Windows pro de 64 bits, procesador basado en x64

---



# ESCENARIO DE TRABAJO

## POSTGRESQL

01

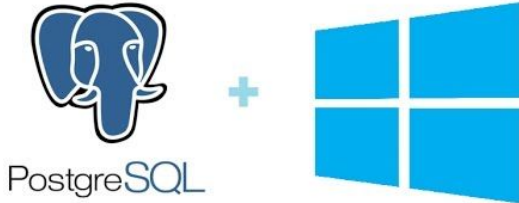
Instalación de PostgreSQL

02

Configuración en YCSB

03

Creación de Base de Datos



```
db.driver=org.postgresql.Driver
db.url=jdbc:postgresql://127.0.0.1:5432/testa
db.user=postgres
db.passwd=[REDACTED]

operationcount = 1000000
recordcount = 1000000
threads = 16
```

Usetable	
PK	YCSB_KEY
	FIELD0
	FIELD1
	FIELD2
	FIELD3
	FIELD4
	FIELD5
	FIELD6
	FIELD7
	FIELD8
	FIELD9



**ESPE**  
UNIVERSIDAD DE LAS FUERZAS ARMADAS  
INNOVACIÓN PARA LA EXCELENCIA

# ESCENARIO DE TRABAJO

## CASSANDRA

01

Instalación de  
Cassandra



3.11



02

Configuración en  
YCSB

```
hosts=127.0.0.1
port=9042
debug=true
keyspace=ycsb
cassandra.username=Cassandra
cassandra.password=XXXXXXXXXX

operationcount=1000000
recordcount=1000000
threads=16
```

03

Creación de Base  
de Datos

Usertable	
PK	YCSB_KEY
	FIELD0
	FIELD1
	FIELD2
	FIELD3
	FIELD4
	FIELD5
	FIELD6
	FIELD7
	FIELD8
	FIELD9



**ESPE**  
UNIVERSIDAD DE LAS FUERZAS ARMADAS  
INNOVACIÓN PARA LA EXCELENCIA



03

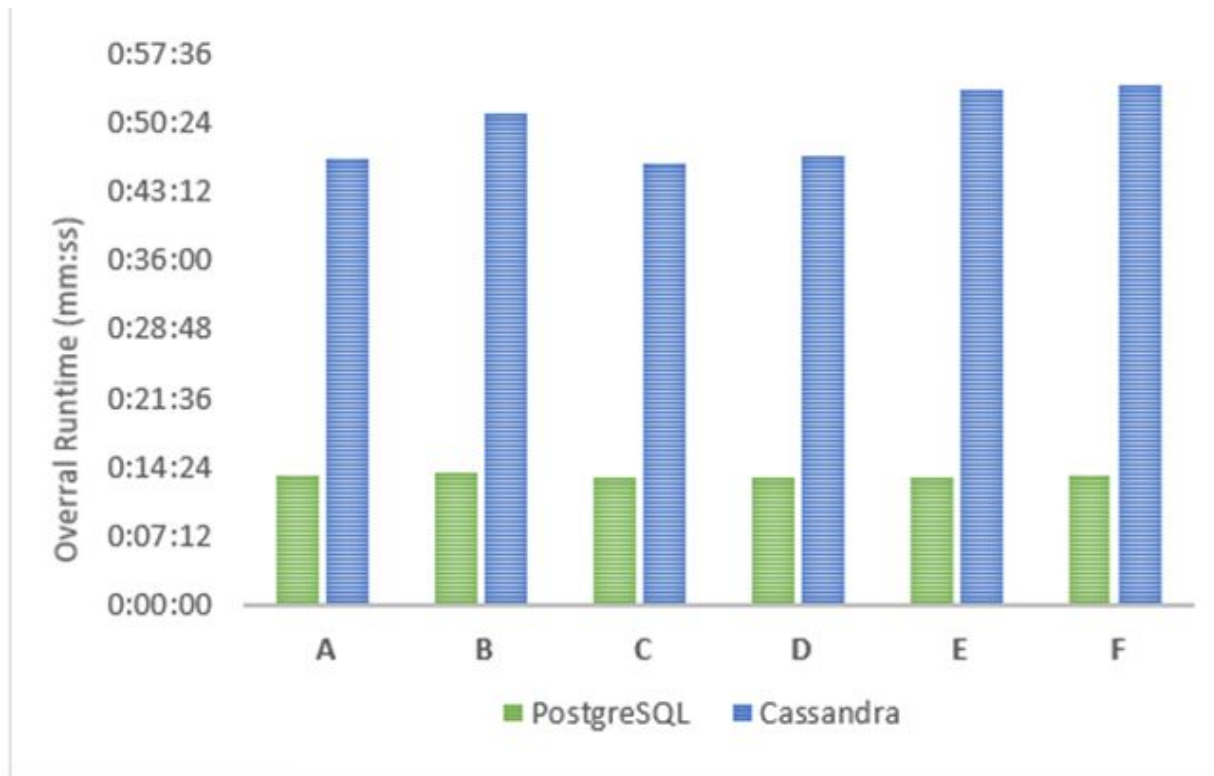
# PLAN DE PRUEBAS Y RESULTADOS



**ESPE**  
UNIVERSIDAD DE LAS FUERZAS ARMADAS  
INNOVACIÓN PARA LA EXCELENCIA

# FASE DE CARGA - CORE I7

## OVERRRAL RUNTIME (MM:SS)

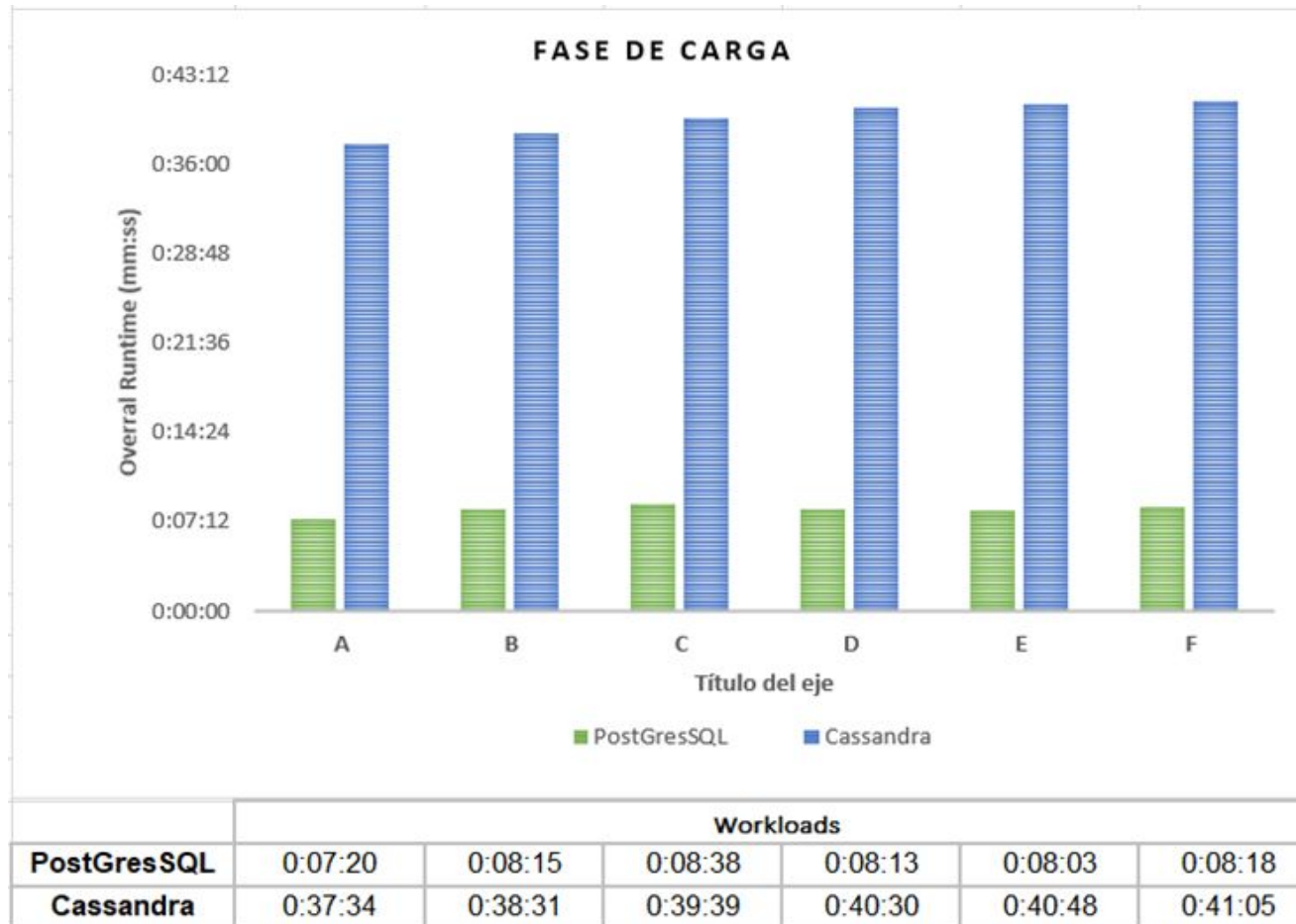


	Workloads					
PostgreSQL	0:13:26	0:13:44	0:13:16	0:13:10	0:13:13	0:13:29
Cassandra	0:46:40	0:51:18	0:46:04	0:46:54	0:53:36	0:54:16

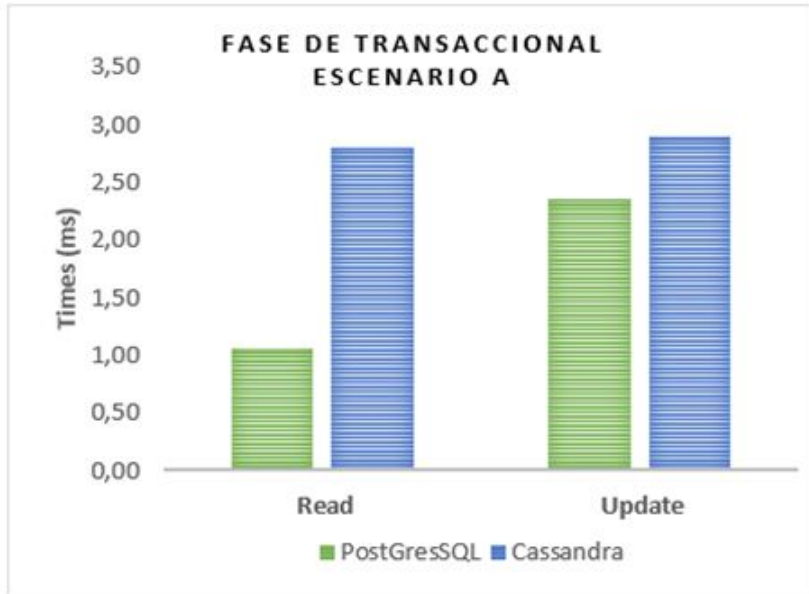


# FASE DE CARGA - RIG DE PROCESAMIENTO

## OVERRRAL RUNTIME (MM:SS)



# FASE TRANSACCIONAL - CORE I7

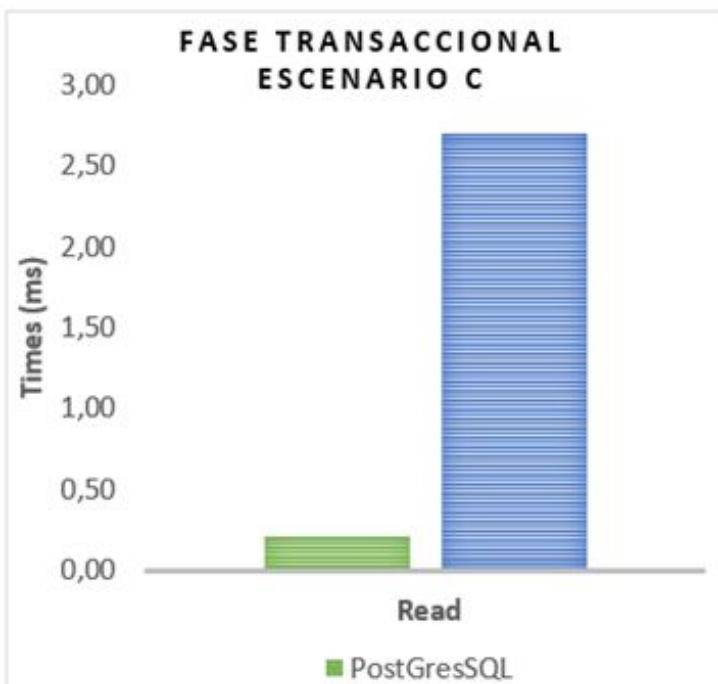


<b>PostGresSQL</b>	1,06	2,35
<b>Cassandra</b>	2,80	2,90

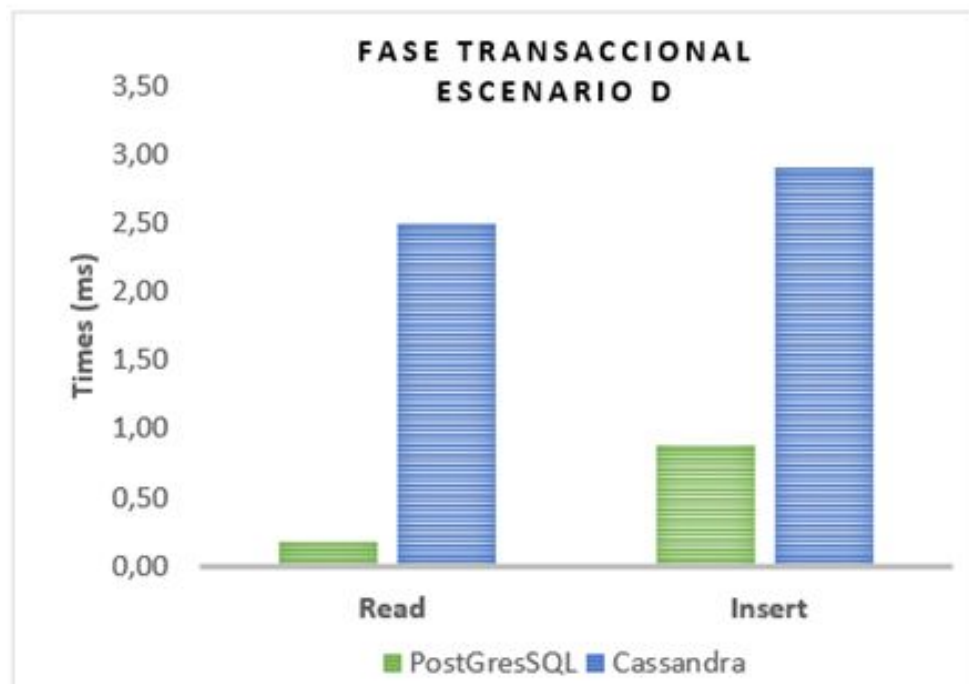


<b>PostGresSQL</b>	0,22	0,66
<b>Cassandra</b>	2,90	2,90



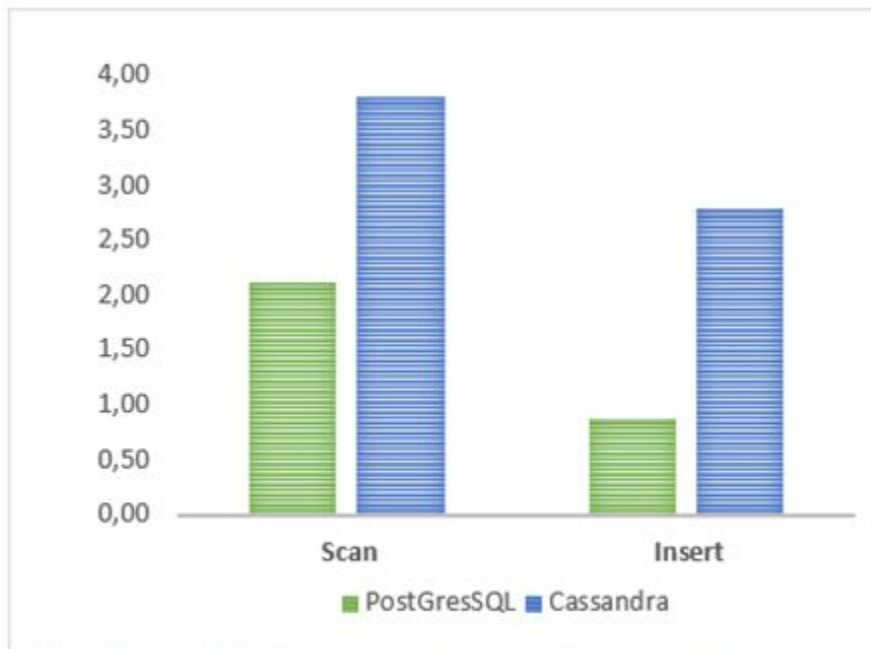


<b>PostGRES SQL</b>	0,22
<b>Cassandra</b>	2,70

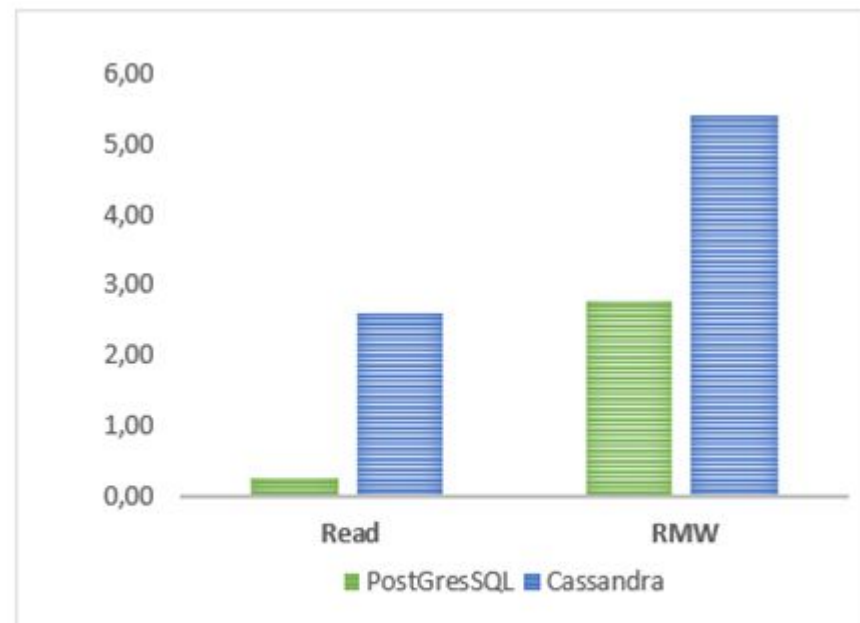


<b>PostGRES SQL</b>	0,18	0,89
<b>Cassandra</b>	2,50	2,90





<b>PostGresSQL</b>	2,12	0,88
<b>Cassandra</b>	3,80	2,80

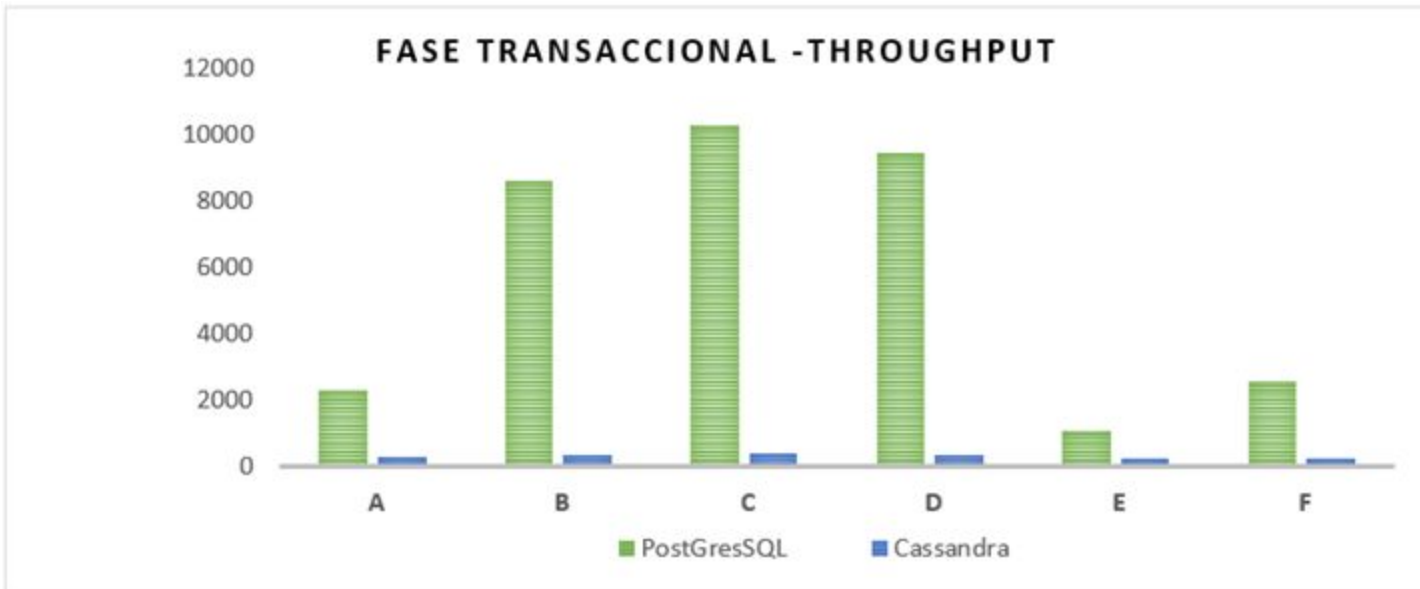


<b>PostGresSQL</b>	0,28	2,77
<b>Cassandra</b>	2,60	5,40



# FASE TRANSACCIONAL - CORE I7

## RENDIMIENTO - (THROUGHPUT)

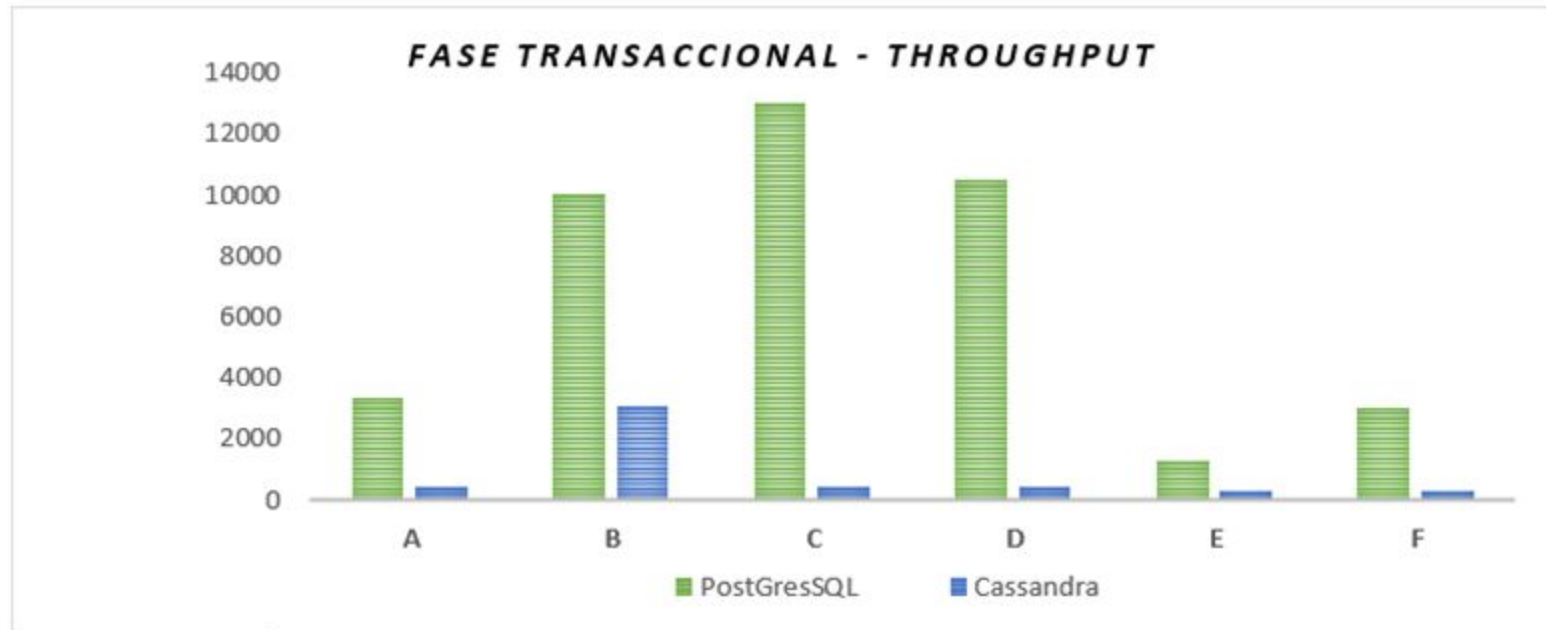


	Workloads					
PostGresSQL	2308	8598	10292	9465	1110	2549
Cassandra	334	352	413	381	263	247



# FASE TRANSACCIONAL - RIG DE PROCESAMIENTO

## RENDIMIENTO - (THROUGHPUT)



	Workloads					
PostGresSQL	3328	10038	13027	10496	1282	3001
Cassandra	442	3073	439	437	334	314





# ANÁLISIS DE RESULTADOS

## Procesador core I7

< - 16 Subprocesos - >

Carga de trabajo	Descripción	Cassandra	PostgreSQL	Desempeño Relativo
A	50 % Lectura / 50 % Actualización	344.087	2308.583	6.70x
B	95 % Lectura / 5% Actualización	352.367	8598.156	24.40x
C	100 % Lectura	413.690	10292.936	24.88x
D	95 % Lectura / 5% Inserción	381.227	9465.125	24.82x
E	95 % Escaneo / 5% Inserción	263.291	1110.528	4.21x
F	50 % Lectura / 50% RMW	247.150	2549.271	10.31x

## RIG de Procesamiento

< - 32 Subprocesos - >

A	50 % Lectura / 50 % Actualización	442.181	3328.518	7.52x
B	95 % Lectura / 5% Actualización	3.073	10038.548	3.26x
C	100 % Lectura	439.075	13027.279	29.66x
D	95 % Lectura / 5% Inserción	437.187	10496.924	24.01x
E	95 % Escaneo / 5% Inserción	334.718	1282.848	3.83x
F	50 % Lectura / 50% RMW	314.676	3001.209	9.53x



# ANÁLISIS DE RENDIMIENTO

## USO DE MEMORIA CASSANDRA

### Memoria



En uso (comprimido)	Disponibile	Velocidad:	2400 MHz
<b>9,1 GB (0 MB)</b>	<b>54,7 GB</b>	Ranuras usadas:	4 de 8
Confirmada	En caché	Factor de forma:	DIMM
<b>12,3/73,4 GB</b>	<b>18,9 GB</b>	Reservada para hardware:	131 MB
Bloque paginado	Bloque no paginado		
484 MB	230 MB		

## USO DE MEMORIA POSTGRESQL

### Memoria



En uso (comprimido)	Disponibile	Velocidad:	2400 MHz
<b>5,9 GB (0 MB)</b>	<b>57,7 GB</b>	Ranuras usadas:	4 de 8
Confirmada	En caché	Factor de forma:	DIMM
<b>7,2/73,4 GB</b>	<b>11,3 GB</b>	Reservada para hardware:	131 MB
Bloque paginado	Bloque no paginado		
369 MB	197 MB		



# ANÁLISIS DE RENDIMIENTO

## CONSUMO DE DISCO POSTGRESQL

Disco 0 (C:)

Tiempo de actividad



60 segundos

Velocidad de transferencia de disco



60 segundos

Tiempo de actividad	Tiempo promedio de respuesta	Capacidad:	224 GB
61%	52,7 ms	Con formato:	224 GB
Velocidad de lectura	Velocidad de escritura	Disco del sistema:	Si
603 KB/s	21,7 MB/s	Archivo de paginación:	Si

## CONSUMO DE DISCO CASSANDRA

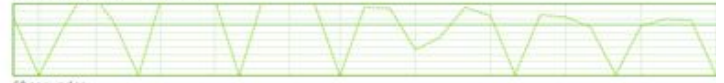
Disco 0 (C:)

Tiempo de actividad



60 segundos

Velocidad de transferencia de disco



60 segundos

Tiempo de actividad	Tiempo promedio de respuesta	Capacidad:	224 GB
6%	28,1 ms	Con formato:	224 GB
Velocidad de lectura	Velocidad de escritura	Disco del sistema:	Si
0 KB/s	6,4 MB/s	Archivo de paginación:	Si



**ESPE**  
UNIVERSIDAD DE LAS FUERZAS ARMADAS  
INNOVACIÓN PARA LA EXCELENCIA

# CONCLUSIONES

- En base a la evaluación comparativa y pruebas de desarrollo se obtuvo que PostgreSQL obtuvo mayor rendimiento en el manejo de las operaciones CRUD frente a la herramienta YCSB, destacándose en mostrar mejores tiempos de latencia y alto consumo en servicios del sistema como en disco duro, CPU y memoria.
- Mediante la revisión de la literatura e investigaciones similares sobre la comparativa entre bases de datos relaciones y no relaciones, se determinó que las bases de datos no relacionales son capaces de procesar grandes volúmenes de datos con relación a los métodos CRUD.



# CONCLUSIONES

- Con los seis escenarios de pruebas realizadas con el Framework YCSB se facilitó el manejo de las operaciones CRUD, generando datos precisos con el manejo de un millón de datos con 16 subprocesos.
- Respecto a las mediciones de rendimiento al tiempo general ambas bases de datos se destacaron en mayor rapidez, estableciendo que PostgreSQL trabajó con una latencia de 13.10 (mm/ss) para el escenario D. Por otra parte, Cassandra trabajó con una latencia de 46.04 (mm/ss) para el escenario C.



# RECOMENDACIONES

- Cassandra al ser una base de datos distribuida es capaz de procesar grandes cantidades de datos, se recomienda trabajar con un clúster mínimo de 5 nodos para obtener mejores tiempos al momento de realizar operaciones o transacciones.
- Se recomienda realizar el mismo trabajo de investigación con distintas bases de datos relacionales y no relacionales para comparar los resultados futuros ante los resultados obtenidos en este proyecto.
- Para obtener una mayor velocidad en cuanto a las operaciones de ambas bases de datos, se recomienda agregar índices, que permitirá un rápido acceso a los registros de las tablas.

