# Comparação de Performance entre PostgreSQL e Cassandra NoSQL utilizando um Dataset de Vagas de Emprego do LinkedIn

Amanda de Medeiros Zepechouka, Juan C. F. R. R. de Morais, Leandro de Lima Minchuel, Leonardo J. Zanotti, Victoria K. Vieira, Wellington H. Kania

# Introdução

- Linkedin job posts (Kaggle)

**postings.csv** (515.2 MB)

Detail Compact Column 10 of 28 columns

**About this file** Add Suggestion

Main table. Supplemental files in folders (companies, job skill/industry tags, etc)

| job_id | company_name | title | description | max_salary | pay_pe |
|---|---|---|---|---|---|
| 922k 3.91b | [null] 1%<br>Liberty Healthcare ... 1%<br>Other (121022) 98% | 72521 unique values | 107828 unique values | 1 120m | [null]<br>YEARLY<br>Other (15 |
| 921716 | Corcoran Sawyer Smith | Marketing Coordinator | Job descriptionA leading real estate firm in New Jersey is seeking an administrative Marketing Coord... | 20.0 | HOURLY |
| 1829192 | | Mental Health Therapist/Counselor | At Aspen Therapy and Wellness , we are committed to serving clients with best practices to help them... | 50.0 | HOURLY |

https://www.kaggle.com/datasets/arshkon/linkedin-job-postings

# Tecnologias e comparação

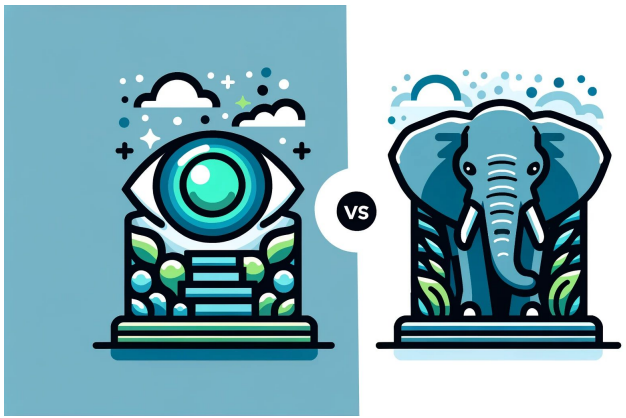
01 | Velocidade

02 | Simplicidade (de instalação, configuração e escrita de código)

03 | Modelagem de dados

04 | Error report detalhado e facilidade de correção

05 | Abrangência/Especificação




https://medium.com/open-source-journal/apache-cassandra-vs-postgresql-when-and-why-choosing-the-right-database-for-your-project-986b888225aa

# Estudo de caso

1. **Consulta Envolvendo Máximo (Agrupamento):** Qual é o maior salário presente no banco de dados?

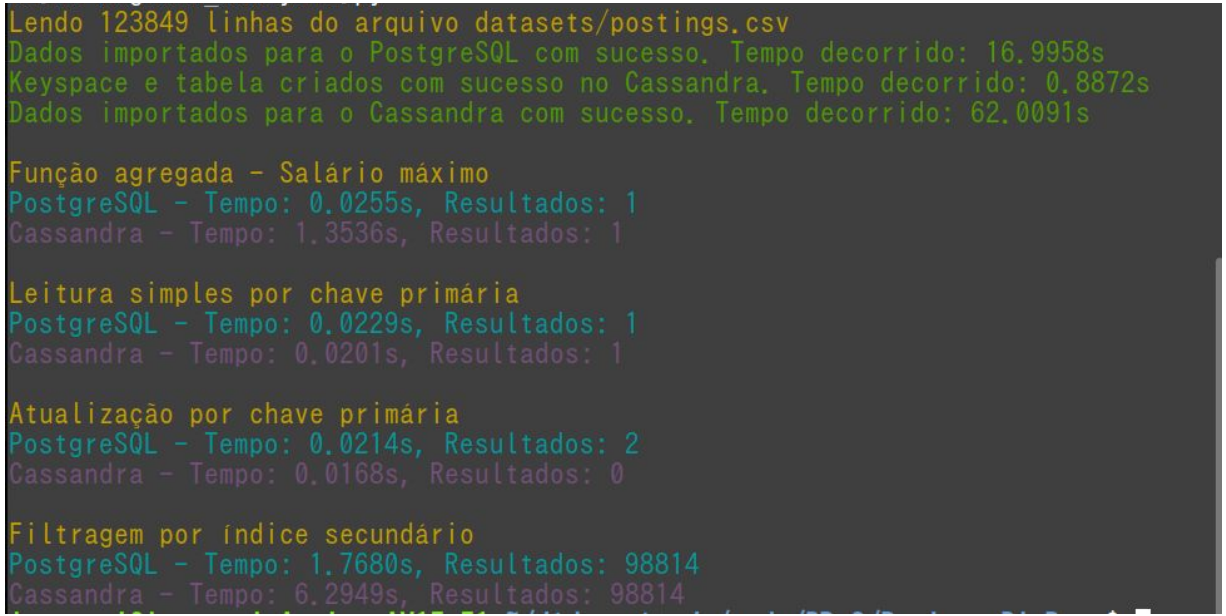
2. **Leitura simples por chave primária**: Qual a vaga com job_id igual a ”2147609816”?

3. **Atualização de dados por chave primária:** Atualize o tíıtulo da vaga com job_id igual a ”2974397965”

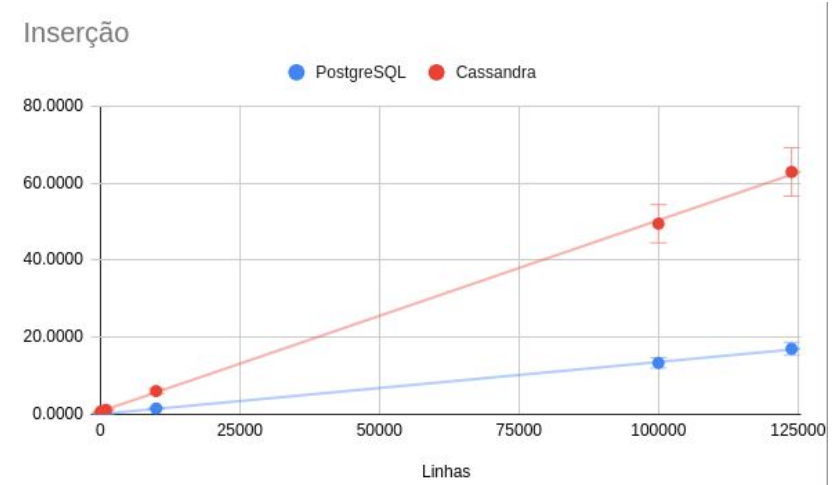
4. **Filtragem por índice secundário:** Selecione todas as vagas que são de tempo integral

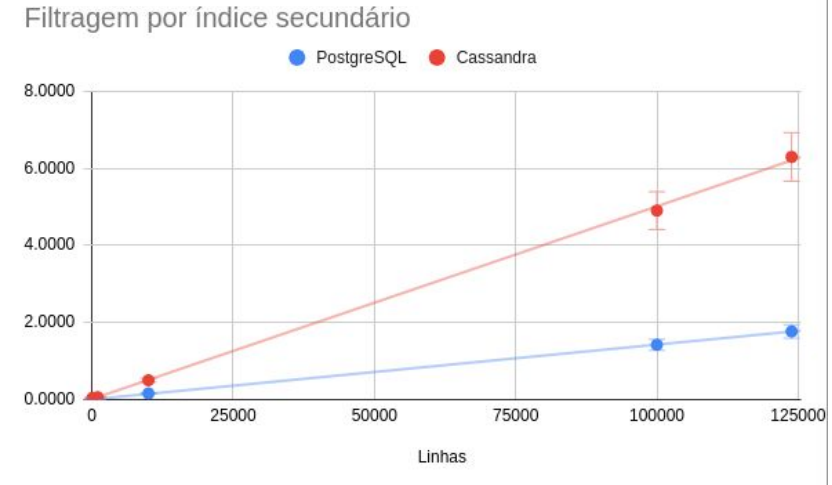
# Resultados

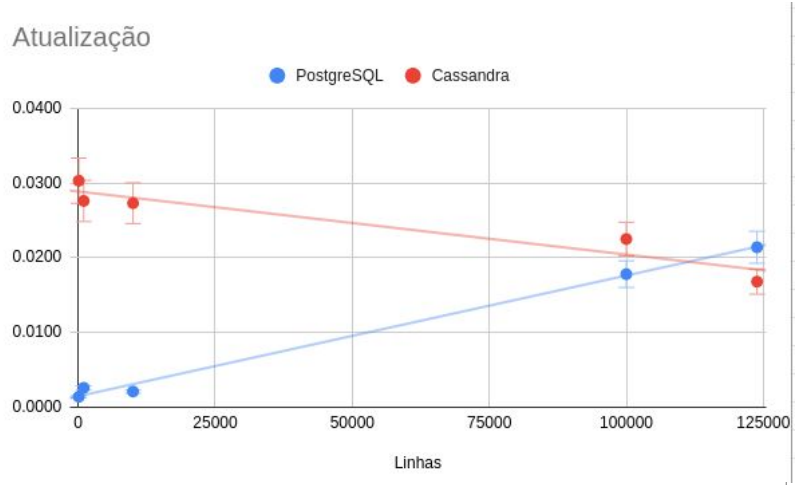
- PostgreSQL com performance superior geral
- Cassandra se destaca nas consultas por chave primária

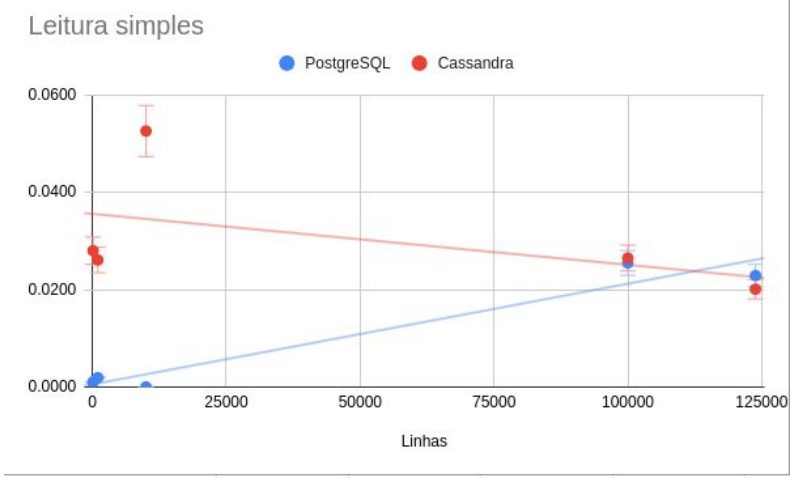


# Regressão linear dos resultados por tempo



# Simplicidade, abrangência e erros

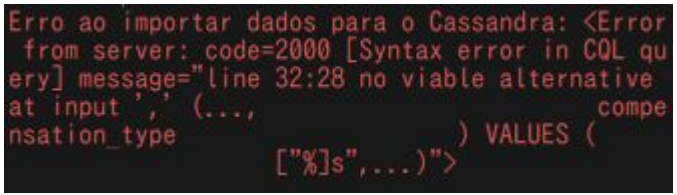
**2**

**Linhas de código para inserir no PostgreSQL**

**178**

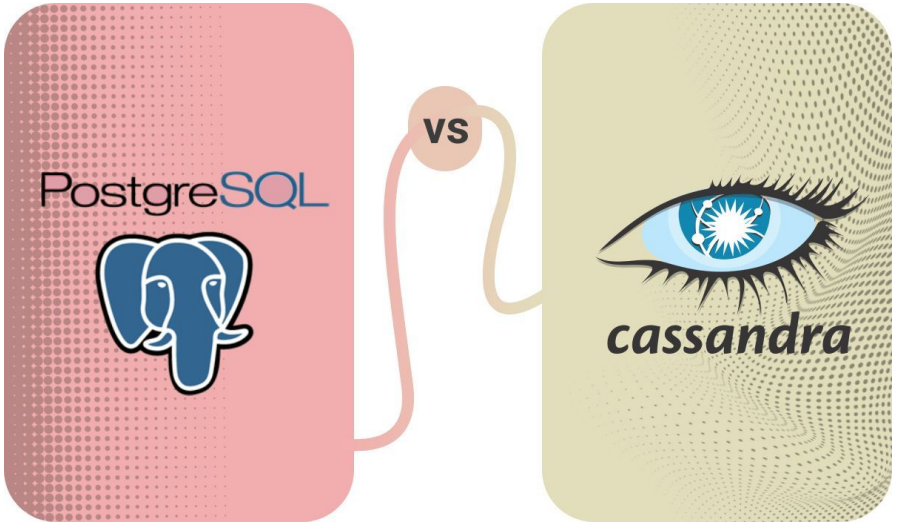
**Linhas de código para converter dados, remodelar e inserir no Cassandra**

- Match exatos
- Group By, JOIN, LIKE %word%
- ElasticSearch, Apache Spark
- Pré-agregação dos dados, contadores e sumários
- Atualizar uma coluna é inviável



# Conclusão

- Resultados foram os esperados
- Dataset maior (milhões de linhas)
- Usar outra modelagem de dados
- Usar outra ferramenta para fazer a comparação
- Padronizar o ambiente




https://devathon.com/blog/sql-vs-nosql-mysql-vs-postgresql-vs-mongodb-vs-cassandra/

# Thank you.

**“Pensaram que iam deixar o pai offline, mas não...” - Neymar Jr.**