



Modelamento de SGBD NoSQL visando a otimização das configurações em função do desempenho, disponibilidade e consumo de energia

Aluno: Carlos Gomes Araújo
cga@cin.ufpe.br

Orientador: Prof. Dr. Eduardo Antonio Tavares Guimarães
eagt@cin.ufpe.br

Coorientador: Prof. Dr. Meuse Nogueira de Oliveira Junior (IFPE)
meusejunior@recife.ifpe.edu.br

Centro de Informática (CIn)
Pós-Graduação em Ciência da Computação
Universidade Federal de Pernambuco (UFPE)



UNIVERSIDADE
FEDERAL
DE PERNAMBUCO



CIn.ufpe.br



Agenda

1. Introdução
2. SGBD NoSQL
3. Trabalho Realizado (Mestrado)
4. Proposta Atual
5. Atividade Atual



Introdução

O consumo de energia contribui significativamente nos custos em *data centers* (NRDC 2015)

- Em 2013, *data centers* consumiram nos EUA o equivalente a \$ 9 bilhões USD em energia
- Para 2020, estima-se este consumo em \$ 13,7 bilhões USD



Introdução

Armazenamento de dados

- Importante subsistema em *data centers*
- Evoluiu e aumentou a capacidade devido ao advento de novos paradigmas:
 - *Big Data*



2. SGBD NoSQL



SGBD NoSQL

- O gerenciamento de grande volume de informações encontra barreiras nas tecnologias tradicionais com características de atomicidade, consistência, isolamento e durabilidade (ACID)
- NoSQL prioriza o desempenho, pois a consistência fica em segundo plano
- SGBDs que aderem a características ACID fazem o oposto



SGBD NoSQL

O dimensionamento de sistemas que utilizam as transações ACID se mostra um problema, os conflitos decorrem de diferentes aspectos da alta disponibilidade de sistemas distribuídos que não são de fácil resolução



SGBD NoSQL

Características:

- Livre de esquema
- Fácil replicação
- API simples
- Consistência eventual (não ACID)



3. Trabalho Realizado (Mestrado)



Trabalho Realizado (Mestrado)

Consideramos três SGBDs NoSQL de modelos distintos:

- Modelo: Coluna
 - SGBD: Cassandra
- Modelo: Orientado a documento
 - SGBD: MongoDB
- Modelo: Chave-valor
 - SGBD: Redis



SGBD NoSQL

NoSQL:

1. MongoDB
2. Cassandra
3. Redis

Rank			DBMS
Jun 2016	May 2016	Jun 2015	
1.	1.	1.	Oracle
2.	2.	2.	MySQL +
3.	3.	3.	Microsoft SQL Server
4.	4.	↑ 5.	MongoDB +
5.	5.	↓ 4.	PostgreSQL
6.	6.	6.	DB2
7.	7.	↑ 8.	Cassandra +
8.	8.	↓ 7.	Microsoft Access
9.	↑ 10.	9.	SQLite
10.	↓ 9.	10.	Redis +

Fonte: DB-Engines





Metodologia

Base em *Design of Experiments*

Experimento fatorial (1^k)

- Dois fatores ($k = 2$)
 - SGBD, Comando

- Três níveis ($l = 3$)
 - (i) SGBD: Cassandra, MongoDB, Redis
 - (ii) Comando: escrita, leitura, atualização

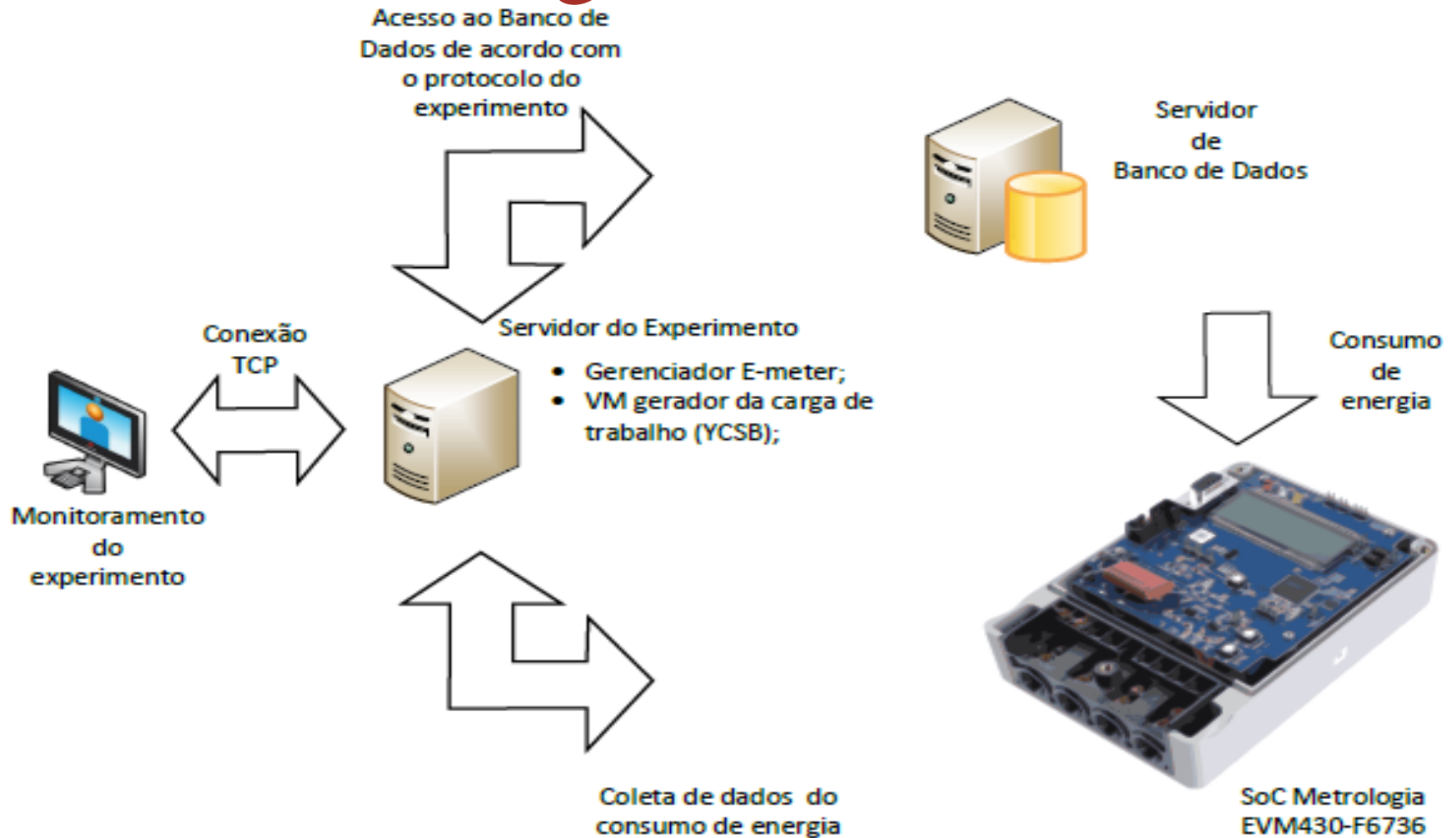


Metodologia

- Consideramos cargas de 1.000, 10.000 e 100.000 operações
- Métricas
 - Consumo de energia (J) e tempo de execução (ms)



Metodologia





Metodologia

Yahoo! Cloud Serving Benchmark (YCSB) é um *framework* para comparação de SGBDs, basicamente:

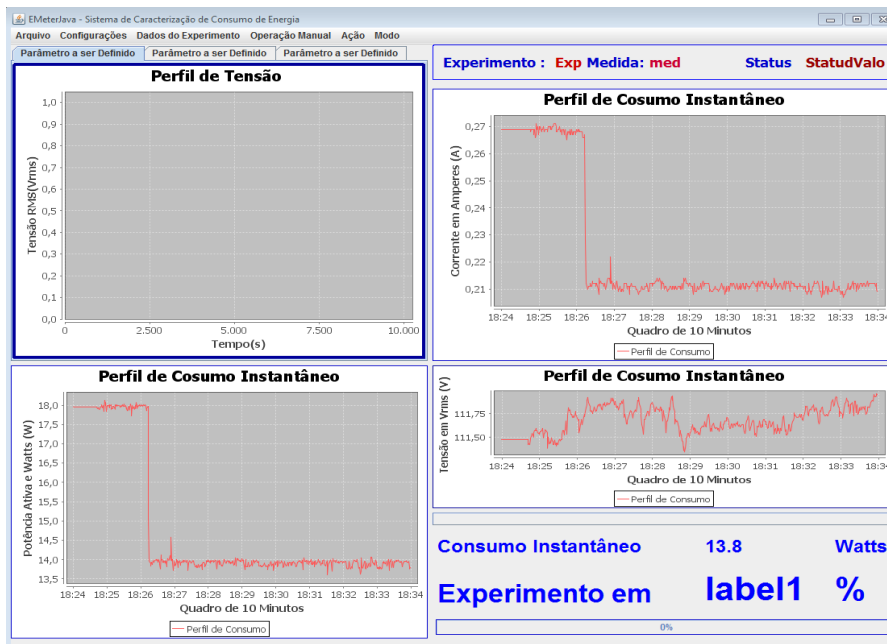
- Define um conjunto de dados a ser trabalhado no banco de dados
- Executa um dado procedimento parametrizado
- Afere o desempenho



Metodologia

Emeter

- Software Emeter (JAVA)
- Hardware EVM430-F6736



Metrology SoC
EVM430-F6736



Resultados Experimentais

Redis	Cassandra	MongoDB
Bom desempenho em 1.000 operações	Escrita e atualização	Resultados estáveis
Menor escalabilidade para 10.000 e 100.000 operações	Bom comportamento no incremento da carga de trabalho	Escrita tem menor desempenho que leitura e atualização
	Menor desempenho em leitura	



Resultados Experimentais

Ranking

Tempo de Execução	Consumo de Energia
1 - Cassandra	1 - MongoDB
2- MongoDB	2 - Cassandra
3 - Redis	3 Redis



Comentários

- Não há um melhor SGBD em todos os tratamentos estudados
- O SGBD mais rápido não é necessariamente o que consome menos energia



Conclusão

- O consumo de energia pode variar significativamente entre os SGBDs avaliados para comandos e cargas de trabalho diferentes
- As diferenças de consumo de energia encontradas, justificam a continuidade de estudos em ambiente de cluster e outros espaços não avaliados neste trabalho.



4. Proposta Atual



Proposta Atual

Desenvolver um mecanismo autônomo de otimização de configurações de SGBD NoSQL, baseado em modelos que respondam com o melhor conjunto de configurações, dado uma carga de trabalho e restrições de desempenho, disponibilidade e consumo de energia. Este modelo é retroalimentado com os dados de utilização anteriores, de modo a otimizar o seu funcionamento em função dos níveis de desempenho, disponibilidade e consumo de energia esperados.



5. Atividade Atual



Atividade Atual

Avaliação do desempenho SGBD Cassandra NoSQL em *Cluster*:

- Arquitetura Física:
 - Hardware Dell R710 8 núcleos, 32GB RAM;
 - VMs (2, 4, 6)
 - Processadores (2, 4, 8 núcleos);
 - Memória RAM GB (1, 2, 4);



Atividade Atual

Avaliação do desempenho SGBD Cassandra NoSQL em *Cluster*:

- Cassandra:
 - Fator de Replicação 02;
- YCSB:
 - 180 threads (clientes);
 - Carga 1000000 operações;
 - Mix 50/50 leitura e escrita



Obrigado !

Carlos Gomes Araújo, Centro de Informática - UFPE
cga@cin.ufpe.br