

Armazenamento de Dados Abertos com NoSQL

Jorge Luiz Andrade

Universidade de Brasília

24 de maio de 2017

Roteiro

- 1 Introdução
- 2 Estudo de Caso
- 3 Resultados

Cassandra

- Distribuído e Descentralizado
- Elasticamente Escalável
- Altamente Disponível e Tolerante a Falhas
- Variavelmente Consistente

Bolsa Família

- Programa de transferência de renda para famílias em pobreza ou extrema pobreza
- 13,9 milhões de família com um custo anual de R\$27,4 bilhões

Estudo de Caso

- Portal da Transparência
- 24 arquivos no total
- 16Gib e quase 14 milhões de registros por arquivo

Campos: **UF, Código SIAFI Município, Nome Município, Código Função, Código Subfunção, Código Programa, Código Ação, NIS Favorecido, Nome Favorecido, Fonte-Finalidade, Valor Parcela e Mês Competência**

Modelo de Dados

- Fator de replicação 1(sem tolerância a falhas)
- UF, Código SIAFI Município, Nome Município, NIS Favorecido, Nome Favorecido, Fonte-Finalidade, Valor Parcela, Mês Competência
- Chave primária: **UF**, NIS Favorecido, Valor Parcela e Mês Competência

Aplicação

Foi desenvolvida uma aplicação em Java, com uso do driver da Datastax. A aplicação realiza ETL dos arquivos .csv originais, e consultas no banco por meio de CQL.

Testes

Inserção e Busca realizadas por meio de comandos CQL na aplicação, com uso do driver da Datastax

- 2, 4, 6 máquinas
- 6 meses, 1 ano, 2 anos



Figure: Ambiente

Resultados Obtidos

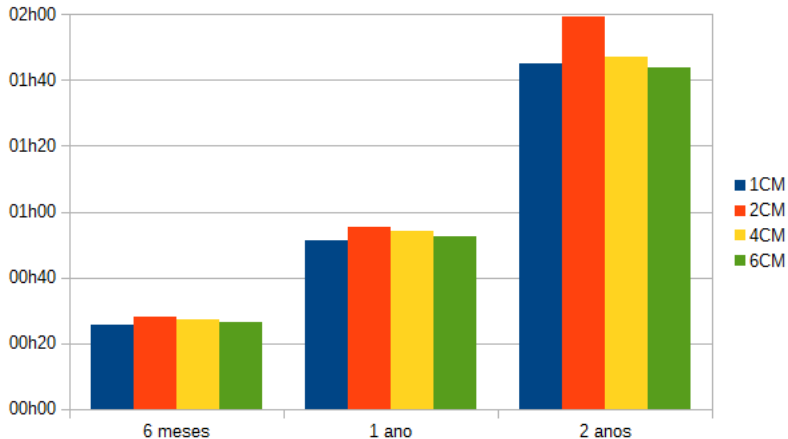


Figure: Inserção de dados