Faculdade de Informática e Administração Paulista – “FIAP”

**CURSO SUPERIOR DE TECNOLOGIA EM BANCO DE DADOS**

**Turma 2º TBDA**

**André Vieira Santos Silva**

**João Pedro dos Santos**

**Kevin Fernandes Redling**

**Maria Aparecida Mendes Loureiro**

**Ralf Silva Nascimento**

**AM Arquiteturas Disruptivas e Big Data – 2º TBDA**

**São Paulo**

**2016**

**Turma 2º TBDA**

**André Vieira Santos Silva**

**João Pedro dos Santos**

**Kevin Fernandes Redling**

**Maria Aparecida Mendes Loureiro**

**Ralf Silva Nascimento**

**AM Arquiteturas Disruptivas e Big Data – 2º TBDA**

Trabalho apresentado à Faculdade de Informática e Administração Paulista – FIAP, Curso Superior de Tecnologia em Banco de Dados, para AM do 3º Semestre – Prof. Ricardo Rezende.

**São Paulo**

**2016**

**Sumário**

[**1 Integrantes** **4**](#_Toc451456451)

[**2 Descrição do Projeto** **5**](#_Toc451456452)

[**3 Identificar um problema no projeto onde uma solução *NoSQL* seria o ideal para resolver o problema. Apresentar o problema e quais os motivos que indicam que *NoSQL* é a melhor alternativa para resolução do problema**. **6**](#_Toc451456453)

[**4 Definir qual o modelo de banco de dados *NoSQL* (Chave/Valor, Baseado em Coluna, Grafos) que será utilizado e qual o motivo que torna este modelo o ideal para o caso**. **7**](#_Toc451456454)

[**5 Definir qual o SGBD que será utilizado (dentro do modelo escolhido) e qual o motivo pela escolha deste SGBD**. **8**](#_Toc451456455)

[**6 Definir a arquitetura que será utilizada, descrevendo-a**. **9**](#_Toc451456456)

[**7 Referências** **10**](#_Toc451456457)

# Integrantes

|  |  |  |
| --- | --- | --- |
| **SELECT** | | |
| **RM** | **Nome do Integrante** | **E-mail** |
| 73920 | André Vieira Santos Silva | [andrevieirass@gmail.com](mailto:andrevieirass@gmail.com) |
| 75312 | João Pedro dos Santos | [jpsdba@gmail.com](mailto:jpsdba@gmail.com) |
| 73671 | Kevin Fernandes Redling | redlingkevin@yahoo.com.br |
| 74305 | Maria Aparecida Mendes Loureiro | [cida\_loureiro@bol.com.br](mailto:cida_loureiro@bol.com.br) |
| 74296 | Ralf Silva Nascimento | [ralf-silva@msn.com](mailto:ralf-silva@msn.com) |

# Descrição do Projeto

Criar dimensões para analisar os perfis dos consumidores de café frente à sazonalidade, identificar os diferentes tipos de harmonização do café com outros produtos, compreender a mudança lenta do consumidor, tendências e analisar o faturamento da empresa, construindo um ambiente de *BI* que municie os gestores com informações relevantes para a tomada de decisões.

# Identificar um problema no projeto onde uma solução *NoSQL* seria o ideal para resolver o problema. Apresentar o problema e quais os motivos que indicam que *NoSQL* é a melhor alternativa para resolução do problema.

Foi encontrado no projeto a necessidade de tratar as análises de sentimentos da concorrência visando uma inteligência competitiva. A solução mais adequada para esta finalidade é o *NoSQL*, pois o banco relacional é menos performático.

Como argumentação será utilizado o *NoSQL* para solucionar diversos problemas relacionados à escalabilidade, performance, disponibilidade e proporcionar uma alternativa de alto armazenamento com velocidade e grande disponibilidade, a fim de eliminar certas regras e estruturas que norteiam o modelo relacional. O ganho de performance, flexibiliza os sistemas de banco de dados para as diversas características que são específicas da empresa.

# Definir qual o modelo de banco de dados *NoSQL* (Chave/Valor, Baseado em Coluna, Grafos) que será utilizado e qual o motivo que torna este modelo o ideal para o caso.

O modelo de banco de dados escolhido foi o de Grafos.

Com a escolha do banco de dados de grafos, o projeto terá flexibilidade e performance para proporcionar análises de sentimentos dos clientes da concorrência com dados factíveis e consistentes, cenário em que um ambiente relacional seria menos eficiente se comparado ao *NoSQL*.

Com esse modelo de banco será possível a criação da inteligência competitiva, onde haverá uma análise sob os comentários referentes às empresas concorrentes em redes sociais, visando atingir a estratégia de vendas com dados por análises de sentimentos, com alta performance.

# Definir qual o SGBD que será utilizado (dentro do modelo escolhido) e qual o motivo pela escolha deste SGBD.

O SGBD escolhido foi o Neo4J.

Com seu esquema flexível teremos o modelo de grafo de propriedade nativo, que realiza a captura de dados, uma vez que ocorre naturalmente, com ambos os nós e relacionamentos tendo propriedades.

Poderemos adicionar e remover propriedades "*on the fly*" de acordo com a evolução do modelo de dados, com restrições de esquema opcionais para flexibilidade.

Enquanto muitos bancos de dados *NoSQL* rejeitam o ACID (Atomicidade, Consistência, Isolamento e Durabilidade), a confiabilidade dos dados é uma consideração de projeto chave para o Neo4J.

O Neo4J permitirá dimensionar volume, leitura e escrita através de cada dimensão chave, tudo ao mesmo tempo proporcionando consultas extremamente rápidas, tempos de resposta consistentes e integridade de dados sólida, essencial para o desenvolvimento.

Integrações com outras tecnologias de banco de dados e ferramentas de análise municiarão o projeto com maior flexibilidade.

O Neo4J clustering fornecerá recursos de dimensionamento para leitura, permitindo espalhar seu gráfico na memória, ao mesmo tempo assegurando que cada instância é capaz de chegar a qualquer nó ou relacionamento usando a sua própria cópia local.

# Definir a arquitetura que será utilizada, descrevendo-a.

Com base na calculadora disponível do site[[1]](#footnote-1) do Neo4J foi feita a prospecção da estrutura a ser utilizada.

Abaixo segue as informações:

|  |  |
| --- | --- |
| **Calculation Input** | |
| Numbers of nodes | 3000000 |
| Number of relationships | 4500000 |
| Property storage per node (bytes) | 15 B |
| Property storage per relantionship (bytes) | 31 B |
| Concurrent requests per second | 1 |
| Number of servers | 1 |
| Number of CPU cores per server | 4 |
| Amount of RAM per server | 4096 MB |

|  |  |
| --- | --- |
| **Calculation Results** | |
| Recommended number of servers | 1 |
| Recommended number of cores per server | 4 |
| Recommended amount of RAM | 12700 MB |
| Recommended Java heap size | 10800 MB |
| Estimated size on disk | 898 MB |
| Current infrastructure fit factor | 57 (100+ recommneded) |

Cluster com 5 máquinas conectadas ao mesmo servidor (ambiente acadêmico).

# Referências

http://Neo4J.com/product/

http://Neo4J.com/developer/guide-sizing-and-hardware-calculator/#\_hardware\_sizing\_calculator

1. http://Neo4J.com/developer/guide-sizing-and-hardware-calculator/#\_hardware\_sizing\_calculator [↑](#footnote-ref-1)