**Requisitos da disciplina: Introdução ao Big Data**

**Descrição.** Considerando o data warehouse construído (conforme especificado nos requisitos anteriores), **porém sem levar em conta o armazenamento de vetores de características**, o grupo deve discutir como ele poderia ser implantado em um NoSQL. O relatório (item I) deve descrever essa transformação baseando-se no seguinte guia:

* Qual é o NoSQL escolhido e por que da sua escolha?
* Qual é o modelo de dados do NoSQL escolhido?
* Por que seria interessante migrar do PostgreSQL ao NoSQL escolhido? detalhar a motivação.
* Como implantar as tabelas de dimensão e de fato no NoSQL escolhido?
* Qual seria a sintaxe das consultas analíticas no NoSQL escolhido? apresente exemplos.

**Observação:** Não há a necessidade de implementar efetivamente o NoSQL e as consultas podem ser apresentadas em linhas gerais (mas utilizando a sintaxe do NoSQL escolhido).

**Arquivos solicitados para entrega.** O relatório (item I) deve conter o consenso do grupo em relação ao plano de migração para o NoSQL escolhido, esclarecendo, no mínimo, as questões apontadas anteriormente (utilize gráficos, desenhos, tabelas, …, se necessário).

Optamos pelo MongoDB, pois contém um grupo de documentos que é a coleção, existe dentro de um banco de dados, pode possuir um conjunto de diferentes campos, os documentos possuem uma estrutura similar, não precisa de um esquema rígido, escala bem horizontalmente, é aprimorado para dados no volume de petabytes e segue o modelo BASE.

Escolhemos o modelo orientado a documentos que é uma extensão do modelo chave-valor, onde o conceito básico do dado é o documento e cada documento pertence a uma coleção que por sua vez pertence a um database. Outras características é que é possível gerenciar informações de dados semi-estruturados.

Seria interessante migrar para o MongoDB pois o foco é em flexibilidade e desempenho, é projetado para dados não-estruturados, é possível escalar horizontalmente por meio de sharding que particiona os dados por intervalos e atribui dados em varias instâncias e oferece grande disponibilidade pela reaplicação de dados que é uma solução que permite servidos de banco de dados tenham os mesmos dados, assim deixando mais fácil o balanceamento de carga e a redundância.

É possível ter as tabelas de fato e dimensão da maneira abaixo através das várias coleções e gerar referencias entre os documentos delas, como abaixo:



Abaixo algumas sintaxes de consultas no MongoDB:

* db.localizacao.find( {‘pais’:’Brasil’} )
* db.flor.fin().sort( {‘nome’,’quantidade’} )
* db.jardim.find().limit(5)
* db.localizacao.distinct(‘cidade’)
* db.flor.find({quantidade: { $gte: 10000}})