**Contexto e Aplicação**

Atualmente o Brasil tem passado por um período bem complexo, a famosa crise, seja ela crise financeira, econômica ou política.

Diante tantos “escândalos” sendo anunciados nos telejornais, revistas, internet e redes sociais, nunca estivemos tão por dentro de tantas informações a respeito dos nossos governantes, a respeito de política e leis do nosso País.

Com o crescimento do acesso a informação, acesso a internet e principalmente redes sociais como Facebook, Twitter dentre outras, o grande volume de informações que circulam diariamente se torna inestimável, diante este contexto e curiosidade da população no geral, pensei em extrair informações de usuários, jornalistas e jornais através do Twitter, tendo em vista que o volume de informação é tão grande que fica impossível fazer leitura de todos estes dados, então, o objetivo foi coletar estes dados para um segunda analise mais direcionada a respeito do que estão dizendo pelo Brasil e pelo Mundo.

O tema do meu trabalho foi a respeito da Crise do Brasil, coletei em Real Time dados dos usuários falando em política, corrupção, desvio de dinheiro e palavras que estão ligadas diretamente ao efeito da crise que estamos vivenciando.

**Metodologia**

Para ter acesso a este montante de dados, foi utilizado a API do Twitter para coletar em tempo real tweets / frases / posts contendo determinadas palavras que coloquei explicitamente no código, dentro da API do Twitter.

Esta API nos fornece uma espece de “Listenner” em tempo real onde, capturei todos estes posts e salvei diretamente no Banco de Dados Não Relacional MongoDB, a linguagem de programação utilizada foi Python, isto porque, é uma linguagem de fácil entendimento e com um vasto conteúdo para consultas na internet.

A escolha da Linguagem e da tecnologia de Banco de Dados se deu pelo fato de conseguirem obter o máximo de desempenho, tanto no desenvolvimento da aplicação quanto no momento de salvar dados em Real Time, como foi adicionado no código diversas palavras populares que dizem a respeito de politica, já se imagina que o retorno das informações que são vastas e bem diversificadas contento frases, links (URL´s) dentre outros tipos.

**Resultados**

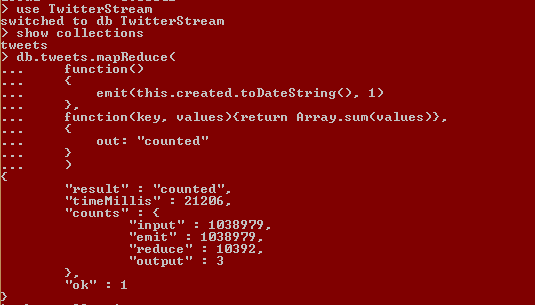
Diante tantas falas que o Banco de Dados não relacional possui um poder de processamento de grandes volumes de dados, certamente me surpreenderia com a velocidade de processamento deste tipo de Banco de Dados, localizar dados, localizar trechos em frases de determinadas “colunas / Atributos”, o tempo de resposta muito bom, tendo em vista a quantidade de dados que possui na Coleção que facilmente criei utilizando o banco MongoDB.

A quantidade de palavras passadas para a localização em mais de um milhão de documentos parece ser indiferente para sua localização, o tempo de resposta parece não alterar como percebemos em Bancos de Dados Relacionais, com os exemplos de consultas que fiz, consegui perceber o quanto é interessante trabalhar com este tipo de tecnologia (NoSQL).

A quantidade total de Twitts capturados em 3 dias de processamento foi de:



Foi utilizado uma função com MapReduce para fazer a contagem diária de posts e inserido em outra collection:



Quantidade:

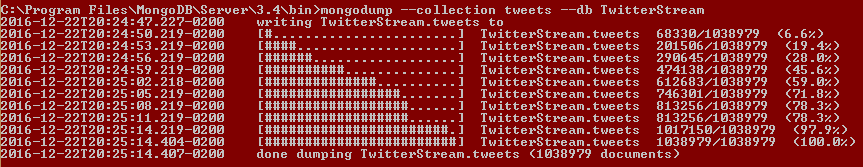


Média de posts por hora nestes três dias de coletagem:

Para calculo desta média utilizei o seguinte comando com a seguinte quantia /hora:



O comando utilizado para a realização do backup da base e collection especifica foi:



Esta base de dados será utilizada para outros testes que farei por conta própria a fim de amadurecer meus conhecimentos nesta tecnologia e para uma melhor aplicação no dia a dia de uma empresa.

**Conclusões**

Este trabalho foi muito agregador, através dele consegui compreender da linguagem de programação Python, coloquei em pratica conceitos visto em sala de aula a respeito do SGBD (Sistema Gerenciador de Banco de Dados) MongoDB bem como o potencial, possibilidades de utilização tanto da linguagem de programação e Banco de Dados.

A forma de trabalho, a facilidade em criar um banco de dados, em armazenar dados dos mais diversos tipos e de forma rápida me chamou bastante atenção, o conceito NoSQL se tornou mais claro com o desenvolver deste trabalho pratico.

Inicialmente os Banco de Dados NoSQL não me despertava praticamente nenhum interesse, depois do trabalho em questão, consegui visualizar, imaginar o tanto que tal tecnologia resolve problemas de gargalos na empresa que trabalho, principalmente em bases de dados relacionais históricas, bases estas que quando precisamos encontrar determinadas informações, o tempo de processamento até a exibição do resultSet se torna inviável para a urgência que determinadas demandas necessitam.

Com esta pratica, principalmente no momento de fazer consultas, a comparação com outras tecnologias de banco de dados, esta sim tem um poder de processamento bem superior aos bancos de dados relacionais.

No geral foi bastante agregador este trabalho uma vez que, após tal execução, abri a cabeça e conseguirei fazer algumas melhorias em meu local de trabalho beneficiando a empresa que trabalho e claro, me agregando mais conhecimento e reconhecimento.