



UNIVERSIDADE FEDERAL DE PERNAMBUCO  
CENTRO DE INFORMÁTICA



---

# Análise comparativa sobre bases de dados para armazenamento e consulta de dados não estruturados no formato JSON.

Proposta de Trabalho de Graduação

**Aluna:** Matheus Dornelas Rodrigues (mdr@cin.ufpe.br)

**Orientador:** Fernando Fonseca (fdfd@cin.ufpe.br)

**Recife, 27 de março de 2016**

## Sumário

Resumo.....	3
Contexto .....	4
Objetivo .....	5
Metodologia .....	6
Cronograma.....	7
Possíveis avaliadores.....	8
Bibliografia .....	9
Assinaturas .....	10

## Resumo

Este trabalho de graduação visa analisar e comparar o funcionamento de bases de dados para o armazenamento e consulta de dados não estruturados, em especial no formato JSON (JavaScript Object Notation). Devido ao fato que esses dados não têm uma estrutura bem definida e por geralmente estarem relacionados a um grande volume, as bases de dados relacionais tradicionais vêm sendo preteridas e a utilização de bases de dados NoSQL (Not only SQL) é mais adequada em cenários desse tipo. Para justificar a opção por NoSQL é necessário uma análise das características desses dois tipos de sistemas de armazenamento, além disso, devido às várias maneiras de como se pode manipular dados em NoSQL é importante mostrar como as diferentes categorias de sistemas de banco de dados desse paradigma funcionam, pois alguns desses sistemas são melhores que outros em determinados cenários.

Uma situação que une dados não estruturados e a movimentação de uma grande massa de dados é a aquisição de informações sobre as atividades dos usuários por partes de sites, aplicativos móveis ou outros dispositivos eletrônicos. Um estudo de caso real será analisado e por meio desse estudo e de experimentos complementares será avaliada qual a melhor alternativa para o armazenamento remoto das atividades dos usuários de um aplicativo móvel.

Palavras Chave: NoSQL, Aplicativos Móveis, Armazenamento Remoto

## Contexto

Quando se fala sobre a criação e gerenciamento de grandes volumes de dados é comum que o termo Big Data (MATHEW e KUMAR, 2015) seja mencionado, mesmo quando não seja esse exatamente o caso. Com o passar dos anos a informação contida nos dados se tornou negócio por si só, ensejando que a informação possa ser tratada como um produto. É por meio da análise desses dados que muitas empresas buscam conhecimento para tomar melhores decisões e agir de maneira estratégica.

Atualmente muitas empresas buscam saber o comportamento dos seus clientes para obter mais conhecimento sobre suas preferências. Essa informação é adquirida a partir do monitoramento das atividades desses clientes no uso do produto em questão. Os dados captados por dispositivos eletrônicos, sites, aplicativos móveis podem ser organizados sem uma estrutura fixa, por isso são chamados de dados não estruturados. Durante muito tempo, sistemas de gerenciamento de base de dados relacionais (SGBD relacionais) eram as únicas alternativas para armazenamento e gerenciamento de dados, mas com crescente volume de informação sendo gerado foi perceptível uma queda no desempenho quando era necessário processar uma grande quantidade de dados, e o particionamento dessas bases de dados para a implementação de um sistema distribuído não é trivial. Outros fatores que não contribuem para utilização de SGBD relacionais é o fato de trabalharem com esquemas de dados bem definidos, em contraposição à realidade de que utilizar dados não estruturados muitas vezes é mais conveniente por aumentar a capacidade de representação de uma informação. Com a proposta de resolver os problemas de desempenho e dar liberdade à estrutura dos dados foi que surgiram as tecnologias NoSQL (HAN, E, *et al.*, 2011) (CATTELL, 2010).

Existem diversos sistemas de bases de dados NoSQL, inclusive alguns motores de busca (*search engines*) podem ser utilizados para armazenamento, como por exemplo o Elasticsearch (BRASETVIK, 2013). Cada categoria de sistema NoSQL (documento, coluna, chave-valor, grafo) lida de maneiras diferentes com questões como: armazenamento, organização, particionamento, capacidade de consultas, entre outras. A escolha de qual base de dados deverá ser utilizada depende do propósito final do sistema ao qual ele estará trabalhando em conjunto.

## Objetivo

O objetivo deste trabalho de graduação é fazer um estudo entre os sistemas SQL e NoSQL para a manipulação de dados não estruturados com um foco no formato JSON. Dentre os critérios de comparação pode-se listar: processo de armazenamento, organização dos dados, particionamento dos dados, disponibilidade, mecanismos que garantem consistência e, capacidade de realizar consultas complexas. É importante lembrar que além de armazenar, organizar a informação e torná-la disponível para a utilização por outros sistemas é essencial.

Será desenvolvido um estudo de caso real com o intuito de mostrar, por meio de conclusões obtidas com o estudo prévio e de experimentos complementares, qual a melhor opção para o armazenamento remoto das atividades dos usuários de aplicativos móveis.

## Metodologia

Este trabalho de graduação será realizado em cinco etapas principais: revisão bibliográfica e estudo do estado da arte, análise dos métodos de armazenamento e comparação das tecnologias, e desenvolvimento de um estudo de caso real.

A fase de revisão bibliográfica e estudo do estado da arte servirá para exploração e fundamentação para o problema proposto. Além disso, será necessário estudar as soluções empregadas atualmente e buscar direcionamentos para a execução dos experimentos do estudo de caso.

Na fase de exploração dos métodos de armazenamento, será analisado com base nas documentações de algumas bases de dados e em experimentos realizados em outras publicações científicas, os funcionamentos de uma base de dados SQL e das 4 categorias de bases de dados NoSQL quando lidam com dados não estruturados. Após feita a análise, serão feitas comparações sobre alguns critérios que são fundamentais no momento da escolha de um banco de dados para cenários que irão lidar com um grande volume de dados não estruturados, dando ênfase ao comportamento dessas tecnologias com o formato JSON. É essencial falar sobre alguns critérios como: particionamento dos dados entre servidores, disponibilidade dos dados, performance no armazenamento, capacidade de realizar consultas complexas, performance nas consultas.

Após as etapas anteriores já se pode chegar a algumas conclusões, mas para o desenvolvimento do estudo de caso outros fatores específicos devem ser analisados. O estudo de caso que será abordado é quando eventos do usuário de aplicativos móveis precisam ser armazenados em servidores remotos através de requisições HTTP. Para realizar os experimentos precisamos obter um modelo e uma base de dados para simular as requisições que os aplicativos farão ao servidor, além de preparar ambientes para serem testadas diversas situações e serem coletados dados para avaliação. A partir daí devem ser avaliadas características como: a existência de uma interface RESTful para comunicação com servidor, capacidade de lidar com grande quantidade de requisições, mecanismo de recuperação de falhas.

Ao final do trabalho espera-se ter evidências suficientes para mostrar o porquê de NoSQL ser tão aplicado quando se trabalha com dados não estruturados, além de chegar a uma conclusão sobre qual é o melhor banco de dados para ser aplicado como solução do estudo de caso.

## Cronograma

Atividade	Março				Abril				Maio				Junho			
Revisão bibliográfica e estudo do estado da arte.	█	█	█	█												
Análise e comparação das base de dados SQL e NoSQL utilizadas no mercado.					█	█	█	█								
Aquisição dos dados para fase de experimentos								█								
Início da escrita da monografia									█	█						
Preparação dos ambientes para execução dos experimentos										█	█	█				
Execução dos experimentos nas ferramentas escolhidas para testes													█	█		
Defesa																█

## Possíveis avaliadores

- Valéria Cesário Times (vct@cin.ufpe.br)
- Robson do Nascimento Fidalgo (rdnf@cin.ufpe.br)



## Bibliografia

BRASETVIK, A. **Elasticsearch as a NoSQL Database**. Disponível em: <https://www.elastic.co/blog/found-elasticsearch-as-nosql>. Acesso em: 27 de março de 2016

CATTELL, R. Scalable SQL and NoSQL Data Stores. **ACM SIGMOD Record**, December 2010. 12-27.

HAN, J. et al. **Survey on NoSQL Database**. Pervasive Computing and Applications (ICPCA). Port Elizabeth, South Africa: IEEE. 2011. p. 363 - 366.

MATHEW, A. B.; KUMAR, S. D. M. Analysis of Data Management and Query Handling in Social Networks using NoSQL Databases. **Advances in Computing, Communications and Informatics (ICACCI), 2015 International Conference on**, August 2015. 800 - 806.

## Assinaturas

---

Fernando da Fonseca de Souza  
**Orientador**

---

Matheus Dornelas Rodrigues  
**Aluno/Autor**