

Atividade 10 - Pesquisa

Caroline Dantas e Lucas Mazim de Sousa

04/04/2021

O mundo está na era do *big data*. Por isso, empresas e desenvolvedores de sistema devem se manter atualizados do estado-da-arte da pesquisa nesta área para se manterem competitivos e tomarem a melhor decisão quando tiverem que escolher o conjunto de ferramentas que irão utilizar. Tendo isto em vista, para desenvolver um sistema de e-mail marketing competitivo com as ferramentas de mercado, as decisões de implementação do grupo 2 foram tomadas observando as melhores práticas desta área.

Existem diversas definições para o termo *big data* com origens acadêmicas e industriais. Em (GANTZ; REINSEL, 2011), *big data* é definido como "uma nova geração de tecnologias e arquiteturas, construídas para extrair valor de maneira econômica de um grande volume de dados variados, permitindo captura em alta velocidade, descobrimento e/ou análise." Para (CHEN; MAO; LIU, 2014), o significado e a necessidade do termo pode ser entendido a partir da teoria dos 4 V's: volume, velocidade, variedade e valor. Isto é, exploração e descoberta de grandes valores comerciais escondidos.

Para atingir o objetivo de construir uma ferramenta de valor para nossos clientes, devemos nos atentar as principais dificuldades encontradas na implementação de sucesso de um sistema que usufrui de tecnologias *big data*. Por isso optamos por utilizar computação em nuvem.

O principal objetivo da computação em nuvem é levar recursos de computação e armazenamento sob gerenciamento concentrado para aplicações de *big data*. A computação em nuvem resolve o problema de armazenamento de *big data* com tecnologias de armazenamento distribuído, enquanto a eficiência de aquisição e análise é ampliada com computação paralela (CHEN; MAO; LIU, 2014). A computação em nuvem é um estilo de fornecer recursos compartilhados ilimitados, como (Hardware/ Software) para o cliente, assim que solicitado através da internet, estes recursos podem ser escalados de acordo com a demanda (BOKHARI; MAKKI; TAMANDANI, 2018). De acordo com o National Institute of Standard Technology (NIST, Departamento de Comércio dos EUA), existem cinco características essenciais da computação em nuvem: serviços sob demanda, acesso amplo via internet, agrupamento de recursos, elasticidade e marcação de serviços (BOKHARI; MAKKI; TAMANDANI, 2018).

Além disso, computação em nuvem possui diferentes modelos de arquitetura para serviços de negócio. Dentre eles, destacamos o modelo *Infrastructure-as-a-Service* (IaaS). Este modelo fornece infraestrutura como um serviço para o cliente, fazendo com o que os custos de serviços de TI e armazenamento sejam calculados de acordo com o uso. O que vai de encontro com os nossos desejos para o sistema de e-mail marketing.

Por fim, analisamos as vantagens e desvantagens de utilizar um banco de dados não relacional no sistema. Segundo (GYORODI et al., 2015), bancos de dados relacionais são amplamente usados na maioria dos aplicativos e têm bom desempenho quando lidam com uma quantidade limitada de dados. Porém, para lidar com volumes massivos de dados, como arquivos multimídia ou dados de mídias social, o uso destes bancos tradicionais é ineficiente, assim sendo, não podem atender as demandas de armazenamento e escalabilidade trazidos pelo *big data*. Por conta disso, bancos de dados NoSQL estão se tornando a tecnologia principal para *big data* (CHEN; MAO; LIU, 2014). Dentre os bancos de dados disponíveis, escolhemos utilizar o MongoDB. MongoDB é um banco de dados de código aberto e orientado a documentos. Documentos são armazenados como objetos Binary JSON (BSON). Um banco de dados no MongoDB é um conjunto de coleções. Uma base de dados no MongoDB possui coleções, coleções possuem documentos e documentos contém múltiplos campos. Uma coleção não tem *schema* predefinido como uma tabela, por isso os documentos dentro de uma tabela não precisam ter o mesmo formato, deixando o processo de criação de bases de dados bem mais dinâmico e flexível. Diversos estudos comparativos foram feitos entre bancos de dados relacionais e não-relacionais demonstrando as vantagens de se usar bases não relacionais para lidar com *big data* (ABOUTORABI et al., 2015; PADHY et al., 2021; WEI-PING; MING-XIN; HUAN, 2011; KANADE; GOPAL; KANADE, 2014).

Este trabalho elucida as escolhas feitas pelo grupo 2 no que diz respeito as tecnologias de infraestrutura que serão utilizadas na criação do sistema de computação, buscando tornar a ferramenta competitiva e utilizando as melhores boas práticas e tecnologias a disposição.

References

- ABOUTORABI, S. et al. Performance evaluation of sql and mongodb databases for big e-commerce data. In: . [S.l.: s.n.], 2015. page.11
- BOKHARI, M.; MAKKI, Q.; TAMANDANI, Y. K. A survey on cloud computing. In: _____. [S.l.: s.n.], 2018. p. 149–164. ISBN 978-981-10-6619-1. page.11
- CHEN, M.; MAO, S.; LIU, Y. Big data: A survey. *Mobile Networks and Applications*, v. 19, 04 2014. page.11
- GANTZ, J.; REINSEL, D. Extracting value from chaos. *IDC IView*, p. 1–12, 01 2011. page.11
- GYORODI, C. et al. A comparative study: Mongodb vs. mysql. In: . [S.l.: s.n.], 2015. page.11
- KANADE, A.; GOPAL, A.; KANADE, S. A study of normalization and embedding in mongodb. In: . [S.l.: s.n.], 2014. p. 416–421. ISBN 978-1-4799-2572-8. page.11
- PADHY, R. et al. Rdbms to nosql: Reviewing some next-generation non-relational database's. 04 2021. page.11
- WEI-PING, Z.; MING-XIN, L.; HUAN, C. Using mongodb to implement textbook management system instead of mysql. 05 2011. page.11