

CAI, Congreso Argentino de AgroInformática

Bancos de Dados Como Tecnologias Habilitadoras para a Agricultura de Precisão: Uma Análise do Contexto Atual

Dhyonatan Santos de Freitas e Sandro da Silva Camargo

Programa de Pós-Graduação em Computação Aplicada, Universidade Federal do Pampa & Embrapa Pecuária Sul, Bagé, Rio Grande do Sul, Brasil
dhyonatanfreitas@alunos.unipampa.edu.br, sandro.camargo@unipampa.edu.br

Resumo Tecnologias de Agricultura de Precisão (AP) tem um papel fundamental no aumento da produtividade na agricultura moderna. Estas tecnologias têm como o aprimoramento do processo de tomada de decisão, o que gera a necessidade de coleta, armazenamento, processamento e análise de quantidades massivas de dados que crescem rapidamente. Assim, a escolha do Sistema de Gerenciamento de Banco de Dados (SGBD) é crucial para o sucesso destas tecnologias. Neste contexto, o presente trabalho apresenta uma revisão na literatura com a finalidade de verificar quais SGBD estão sendo utilizados em técnicas de AP atuais. Os resultados indicam que em aplicações que geram grandes volumes de dados, como redes de sensores e imageamento digital, são preferidos os SGBD NoSQL em detrimento dos SGBD Relacionais convencionais.

Palavras-chave: SGBD, NoSQL, armazenamento, imagens, sensores.

1 Introdução

O aumento da população mundial aliada à pouca margem de expansão das terras aráveis tem conduzido a uma grande pressão por aumento de produtividade na agricultura. Uma das formas de conduzir ao aumento de produtividade é através das técnicas de Agricultura de Precisão (AP). Estas técnicas visam permitir a coleta, armazenamento, processamento e análise de grandes quantidades de dados a fim de permitir o aprimoramento dos processos de tomada de decisão, com o objetivo de conduzir ao aumento da produtividade [6].

Estima-se que nos próximos anos as propriedades rurais estarão informatizadas e repletas de sensores em seus equipamentos agrícolas, os quais serão responsáveis pela coleta de dados atmosféricos, das plantas e do solo. Além de dados de sensores, algumas técnicas de AP utilizam imagens digitais em seu processamento [2] [3], incidindo na geração de grandes volumes de dados [4]. O primeiro passo após a coleta é o armazenamento desses dados, assegurando a possibilidade de comparações futuras e compondo séries históricas. Mesmo que uma triagem descarte dados irrelevantes, é notável que essas tecnologias geram uma quantidade expressiva de dados. Logo, faz-se necessário avaliar as abordagens

no que diz respeito ao armazenamento desses dados em ambiente computacional. Atualmente, os SGBDs estão divididos em dois modelos, os baseados em linguagem de consulta estruturada (SQL) e os que utilizam uma abordagem não estruturada, denominados NoSQL (*Not only SQL*).

Neste contexto, este trabalho tem os seguintes objetivos: i) realizar uma revisão na literatura atual a fim de verificar quais SGBDs estão sendo utilizados no armazenamento de dados gerados por tecnologias de AP; ii) identificar a potencialidade desses SGBDs em tratar dados de sensores e imagens digitais; iii) identificar uma possível tendência de utilização acerca de sistemas de armazenamento que atendam as necessidades da AP atual.

2 Materiais e Métodos

Esse trabalho foi elaborado a partir de uma revisão da literatura nas bases de dados da *IEEE Digital Library*, Portal de Periódicos da Capes e o web site Google Scholar, portal acadêmico da Google.

Com a intenção de buscar trabalhos desenvolvidos recentemente, o período escolhido para a seleção dos artigos foi entre os anos de 2014 e 2018 e as palavras-chave utilizadas foram “dados”, “armazenamento”, “agricultura”, “precisão” e “imagem”, além de suas correspondentes em inglês “*data*”, “*storage*”, “*agriculture*”, “*precision*” e “*image*”. Este trabalho visa analisar as tecnologias de armazenamento de dados na área de agricultura de precisão. Assim, justifica-se o uso das palavras dados, armazenamento, agricultura e precisão. Além destes, dado o extensivo uso de imagens nesta área, também foi incluída a palavra imagem.

Durante a busca nas bases de dados, em um primeiro momento, selecionamos dez trabalhos pelo conteúdo do título e após a leitura do resumo foram descartados seis trabalhos que não atendiam nossos objetivos. Assim, permaneceram quatro trabalhos [1] [3] [2] [5], os quais englobam o interesse deste trabalho.

3 Resultados e Discussão

Os autores De Souza e Dos Santos [5] apresentaram um estudo onde verificaram o amadurecimento e consolidação dos SGBDs não estruturados e testaram a performance e a escalabilidade dos mesmos em relação a um SGBD estruturado. Pelo viés da consolidação no mercado e também na área acadêmica, verificaram que os SGBDs relacionais ainda dominam grande parte das aplicações, porém no tratamento de grandes volumes de dados os SGBDs não relacionais (NoSQL) são preferidos. Para verificar o desempenho de ambos os modelos, utilizaram o *benchmark* YCSB (*Yahoo! Cloud Serving Benchmark*) e compararam os resultados dos SGBDs Redis e Cassandra com o MySQL. Como resultado, verificaram que o Redis foi melhor nos testes de desempenho que o Cassandra e o MySQL. Porém, no quesito escalabilidade o Cassandra se mostrou a melhor opção. Portanto é possível concluir que o Cassandra é uma alternativa eficiente para aplicações que necessitem tratar fluxos de dados que crescem rapidamente, por exemplo dados de sensoriamento remoto voltados para a AP.

Diante das dificuldades em encontrar aplicações que integrem dados gerados por sensores IoT (*Internet of Things*) voltadas para o desenvolvimento de tecnologias de AP, os autores Jayaraman et al. [1] propuseram uma ferramenta que integra diversos sensores IoT em um único produto, o SmartFarmNet. O SmartFarmNet é uma plataforma IoT para aplicativos agrícolas inteligentes que fornece a integração de dispositivos IoT, além disso propicia escalabilidade no processamento e na análise de dados. Segundo os autores, SGBDs relacionais têm sido a escolha histórica para armazenar dados dos sensores IoT para posterior recuperação, análise e visualização. Sabendo da dificuldade que esses sistemas têm no tratamento de grandes fluxos de dados, os desenvolvedores usaram uma forma não estruturada de armazenamento para gerenciar dados de usuário, dados de sensores e metadados. Com isso, foi possível tornar o SmartFarmNet uma ferramenta capaz de dimensionar, armazenar e processar milhares de sensores IoT em conjunto, resultando em bilhões de dados sendo gerados em alta velocidade. Por fim, avaliaram a escalabilidade da plataforma testando seu comportamento diante de um fluxo grande de dados e concluíram que o sistema é capaz de fornecer uma resposta de consulta quase em tempo real e que a velocidade de geração de dados pelos sensores tem pouco ou nenhum impacto no desempenho do sistema comparado aos sistemas tradicionais de IoT baseados em SQL.

Além das redes de sensores, técnicas de AP que envolvem armazenamento de imagens digitais em banco de dados também geram um enorme fluxo de dados e portanto necessitam de um SGBD adequado. Mueller et al. [3] propuseram um sistema que analisa o estado de culturas agrícolas usando imagens captadas por VANT e dados espaciais. O SGBD escolhido por eles foi o SQL Server 2012. Na visão dos autores o SGBD teve um bom desempenho e atendeu as necessidades do sistema, porém ficou evidente a preocupação deles com a escalabilidade do SGBD estruturado. Os autores armazenaram as imagens em um sistema NTFS e apenas as referências aos arquivos no SGBD, com isso evitaram que o banco de dados aumentasse demais, o que poderia prejudicar o desempenho do sistema.

Em contrapartida, Lopez et al. [2] realizaram um estudo sobre a integração dos SGBDs não relacionais na análise de informações espaciais de Sistemas de Informação Geográficas (SIG), incluindo o armazenamento de imagens. Partindo do princípio que SGBDs relacionais tendem a ter dificuldades em tratar grandes quantidades de dados espaciais, os autores propuseram utilizar os SGBDs MongoBD e Cassandra para integrar informações em imagens. As imagens são segmentadas em partes menores e armazenadas no SGBD juntamente com seus dados *Raster*. Para determinar qual SGBD é mais adequado, os autores realizaram uma avaliação de desempenho deste processo utilizando os dados do Instituto de Geografia da Universidad Nacional Autónoma de México (UNAM), a qual possui um volume de dados de aproximadamente 20 Terabytes e normalmente utiliza SGBDs estruturados para armazenar esses dados. Após a integração dos SGBDs não relacionais, concluíram que na integração de dados *Raster* como características das imagens o MongoDB fornece um gerenciamento único na integração de arquivos grandes, proporcionando melhor desempenho em relação ao Cassandra.

Prevendo que essas tecnologias evoluem muito rápido, recomenda-se como trabalhos futuros novas avaliações desses sistemas, utilizando métodos de avaliação de software válidos e já consolidados, como por exemplo o *benchmark* YCSB.

4 Conclusão

Este trabalho teve como objetivo realizar uma revisão na literatura com a finalidade de verificar quais SGBDs estão sendo utilizados em aplicações que constituem a AP atual. Foram analisados quatro trabalhos, onde constatou-se que grandes volumes de dados gerados por aplicações de AP, seja de uma rede de sensores, seja a partir de imagens digitais, têm melhor desempenho em SGBDs não relacionais (NoSQL), como mostram os trabalhos [1] [2] [5]. Além disso, aplicações que utilizaram um SGBD relacional [3] para armazenar grandes volumes de dados, como por exemplo imagens digitais, relataram preocupação com a escalabilidade do banco e como isso pode prejudicar o desempenho da aplicação.

Desta forma, a consolidação dos SGBDs não relacionais (NoSQL) em relação aos relacionais é constatada também em aplicações de AP que necessitem armazenar grandes volumes de dados. Novas aplicações de AP dependem de um armazenamento eficiente e, conforme as referências analisadas, o uso de SGBD NoSQL tem se constituído em uma alternativa muito promissora.

Agradecimentos: Os autores agradecem à Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS) pela concessão da bolsa de mestrado.

Referências

1. Jayaraman, P.P., Yavari, A., Georgakopoulos, D., Morshed, A., Zaslavsky, A.: Internet of things platform for smart farming: Experiences and lessons learnt. *Sensors* 16(11), 1884 (2016)
2. López, M., Couturier, S., López, J.: Integration of nosql databases for analyzing spatial information in geographic information system. In: *Computational Intelligence and Communication Networks (CICN)*, 2016. pp. 351–355. IEEE (2016)
3. Mueller, W., Idziaszek, P., Boniecki, P., Zaborowicz, M., Koszela, K., Kujawa, S., Kozłowski, R., Przybył, K., Niedbała, G.: An it system for the simultaneous management of vector and raster images. In: *Eighth International Conference on Digital Image Processing*. vol. 10033, p. 100332N. IOP (2016)
4. de Oliveira Assis, J., Souza, V.C., Paula, M.M., Cunha, J.B.S.: Performance evaluation of nosql data store for digital media. In: *Information Systems and Technologies (CISTI)*, 2017. pp. 1–6. IEEE (2017)
5. de Souza, V.C.O., dos Santos, M.V.C.: Maturing, consolidation and performance of nosql databases-comparative study. In: *Proceedings of the annual conference on Brazilian Symposium on Information Systems: A Computer Socio-Technical Perspective*. vol. 1, p. 32 (2015)
6. Wolfert, S., Ge, L., Verdouw, C., Bogaardt, M.J.: Big data in smart farming—a review. *Agricultural Systems* 153, 69–80 (2017)