



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE INFORMÁTICA
GRADUAÇÃO EM ENGENHARIA DA COMPUTAÇÃO



**Um Estudo de Técnicas de Integração de Dados Aplicadas
em Análises de Tempo Real**
PROPOSTA DE TRABALHO DE GRADUAÇÃO

Aluno: Filipe de Melo Silva (fms6@cin.ufpe.br)
Orientador: Robson do Nascimento Fidalgo (rdnf@cin.ufpe.br)
Área: Bancos de dados

14/06/2021

Resumo

A cultura orientada a dados é uma estratégia utilizada em corporações com o intuito de utilizar todas as informações possíveis para auxiliar nas tomadas de decisão. Entretanto, para que essa estratégia funcione, deve ser possível relacionar os dados disponíveis. Por conta disso, se faz necessário o uso de tecnologias de integração, onde estas precisam ser adaptáveis muitas vezes a cenários que requerem análises próximas de tempo real. Neste trabalho serão estudadas as principais soluções de integração de dados oriundos de diferentes fontes de dados, avaliando-as em relação à adequação em cenários que exigem informações próximas de tempo real.

Introdução

O uso eficiente dos dados tem se tornado cada vez mais uma preocupação dentro das corporações. Estima-se que em 2022, 90% das estratégias corporativas irão mencionar a informação como um ativo empresarial crítico [1]. Nesse contexto, a cultura orientada a dados (*data-driven culture*) é uma forte tendência e já é considerada um diferencial competitivo [2]. Seu objetivo é melhorar a tomada de decisão através da captura, limpeza e curadoria de dados significativos de toda a empresa [3].

Entretanto, para que essa estratégia funcione, deve ser possível relacionar os dados disponíveis. O grande impasse dessa questão é a crescente quantidade de sistemas utilizados no dia a dia [4], que fazem com que os dados estejam cada vez mais espalhados em fontes diversas. Por conta disso, diversas soluções de integração de fontes de dados já foram desenvolvidas com o intuito de agregar os dados de diferentes tecnologias de armazenamento, como MySQL, Postgres, HBase, e, além disso, diferentes estruturas de dados em cada uma dessas fontes [5].

Além desses desafios, outros podem surgir. Algumas corporações que possuem serviços como detecção de fraude ou serviços de pontuação em cartão de crédito, necessitam da integração de dados próxima de tempo real [6]. Um outro desafio pode estar presente em cenários com muitas estruturas de dados: o agrupamento dos objetos similares (*schema matching*) nas diferentes bases de dados.

Neste trabalho serão estudadas as principais soluções de integração de dados oriundos de diferentes fontes e, além disso, a adequação dessas soluções para cenários que necessitem de análises próximas de tempo real.

Objetivos

Abordar técnicas de integração de dados oriundos de diferentes fontes de dados, avaliando-as em relação à adequação em cenários que exigem análises próximas de tempo real.

Cronograma

Atividade	Período		
	Junho	Julho	Agosto
Revisão bibliográfica	X		
Investigação dos principais cenários existentes com múltiplas fontes de dados	X		
Categorização dos trabalhos em cada um dos cenários encontrados		X	
Comparação dos trabalhos presentes em uma mesma categoria		X	
Escrita do TG			X
Preparação da apresentação			X

Referências

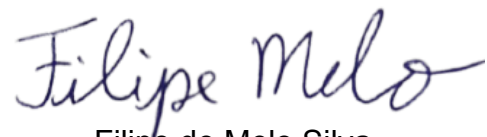
- [1] Garner. 2019. Predicts 2019: Data and Analytics Strategy. Disponível em: <https://www.gartner.com/en/doc/374107-predicts-2019-data-and-analytics-strategy>
- [2] Chatterjee, S., Chaudhuri, R. & Vrontis, D. (2021). Does data-driven culture impact innovation and performance of a firm? An empirical examination. *Ann Oper Res*.
- [3] AWS Cloud Enterprise Strategy Blog. 2019. How to Create a Data-Driven Culture. Disponível em: <https://aws.amazon.com/pt/blogs/enterprise-strategy/how-to-create-a-data-driven-culture>
- [4] The Wall Street Journal. 2019. Employees Are Accessing More and More Business Apps, Study Finds. Disponível em: <https://www.wsj.com/articles/employees-are-accessing-more-and-more-business-apps-study-finds-11549580017>
- [5] Calvanese, Diego & Cogrel, Benjamin & Komla-Ebri, Sarah & Kontchakov, Roman & Lanti, Davide & Rezk, Martín & Rodriguez-Muro, Mariano & Xiao, Guohui. (2016). Ontop: Answering SPARQL queries over relational databases. *Semantic Web*. 8. 10.3233/SW-160217.
- [6] Jon T.S. Quah, M. Sriganesh. “Real-time credit card fraud detection using computational intelligence”. *Expert Systems with Applications*, Volume 35, Issue 4, 2008, Pages 1721-1732, ISSN 0957-4174.

Possíveis Avaliadores

Profª. Bernadette Farias Lóscio

Assinaturas

Recife, 14 de Junho de 2021

A handwritten signature in blue ink that reads "Filipe Melo". The signature is written in a cursive style with a large, stylized 'F' and 'M'.

Filipe de Melo Silva

(Aluno)

Robson do Nascimento Fidalgo

(Orientador)