



UNIVERSIDADE
FEDERAL
DE PERNAMBUCO



UNIVERSIDADE FEDERAL DE PERNAMBUCO

CENTRO DE INFORMÁTICA – Cin

Proposta de Trabalho de Graduação

Desenvolvimento de um *Data Warehouse* para um cenário de *Big Data* utilizando a ferramenta *Hadoop*

Aluno: Bruno de Assis Pereira (bap2@cin.ufpe.br)

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

Recife, setembro de 2016

Resumo

Dados são a representação mais simples das informações e do conhecimento. Realizar um gerenciamento sobre essa informação envolve um processo de organizar, adquirir, armazenar, recuperar, e gerenciar esses dados. Estes são coletados por meio de diferentes processos e usados para auxiliar tomadas de decisão, de modo que aqueles que estão executando e consumindo os resultados do processo possam compreender toda a informação e atender aos seus diversos requisitos.

Um elemento que evoluiu para dar suporte ao processo de tomada de decisões foi o *Data Warehouse* (DW) [2]. Por meio da capacidade de coletar, armazenar, e gerenciar dados, aplicando métodos tradicionais e estatísticos de medições para criar relatórios e plataformas de análise, o *DW* se tornou um elemento chave no processo de tomada de decisões. Esta última pode se tornar ainda mais segura caso se possa analisar um número maior de dados.

O *Big Data* [1] tem se popularizado principalmente pela seguinte razão: as plataformas tecnológicas que surgiram, junto com esse complexo e amplo conjunto de dados, provêm a capacidade de processar vários formatos e estruturas de dados sem ter que se preocupar com as restrições associadas aos sistemas de bancos de dados tradicionais. Como exemplo de ferramentas tem-se o *Hadoop* [3] que é uma arquitetura proposta como solução para o processamento de *Big Data* em uma plataforma mais barata com rápida escalabilidade e processamento paralelo.

Construir um *Data Warehouse* que é alimentado por um *Big Data* possibilitará o desenvolvimento de análises sobre esses dados, e esta atividade trará conhecimento valioso que antes estava escondido ou não estaria disponível tão facilmente.

Neste trabalho serão explorados projetos relacionados que realizaram análises sobre *Big Data*. Ao fim dos estudos, a construção de um DW sobre um cenário de *Big Data* utilizando o *Hadoop* será realizada a fim de apresentar por meio da criação de relatórios contendo aplicação de métodos estatísticos, os conceitos que foram extraídos e o potencial que esse tipo de desenvolvimento tem.

Sumário

Objetivo _____	4
Cronograma _____	5
Proposta de Estutura _____	7
Possíveis Avaliadores _____	7
Referências _____	8
Assinaturas _____	9

Objetivo

Este trabalho tem como objetivo analisar e compreender mais sobre o processo de análises sobre *Big Data*, utilizando técnicas de DW.

Objetivos Específicos

- ❖ Criação de um *Data Warehouse* para um cenário de *Big Data*;
- ❖ Utilização da ferramenta *Hadoop*;
- ❖ Apresentar por meio da criação de relatórios contendo aplicação de métodos estatísticos, os conceitos que foram extraídos; e
- ❖ Destacar o potencial de uso que esse tipo de desenvolvimento tem.

Cronograma

Atividade \ Mês	Setembro			Outubro			Novembro			Dezembro		
Revisão Bibliográfica	■	■	■									
Exploração sobre trabalho de análise sobre <i>Big Data</i>		■	■	■								
Desenvolvimento do <i>Data Warehouse</i>				■	■	■	■					
Documentação e formatação do TG							■	■	■	■		
Defesa											■	

Proposta de Estutura

O trabalho deverá ser organizado como segue:

- Capítulo 1 - Introdução: descrição do contexto do trabalho; motivação; objetivos: geral e específicos; metodologia utilizada; estrutura do trabalho;
- Capítulo 2 - Fundamentação conceitual: Data Warehousing sobre big data; Hadoop;
- Capítulo 3 - Trabalhos relacionados: descrição e análise. Para isto devem ser descritos e justificados os critérios adotados para análise;
- Capítulo 4 - Especificação e Implementação da ferramenta;
- Capítulo 5 - Conclusões: contribuições, limitações; sugestões de trabalhos futuros; e
- Referências.

Possíveis Avaliadores

São possíveis avaliadores desse trabalho: Valéria Cesário Times (CIn/UFPE); Ana Carolina Salgado (CIn/UFPE); Robson do Nascimento Fidalgo (CIn/UFPE).

Referências

- [1] – MELOROSE, J.; PERROY, R.; CAREAS, S. **Data Warehouse in the Age of Big Data**. [s.l: s.n.]. v. 1
- [2] – SUN, L. et al. Present situation and prospect of data warehouse architecture under the background of big data. **Proceedings - 2013 International Conference on Information Science and Cloud Computing Companion, ISCC-C 2013**, p. 529–535, 2014.
- [3] – JURNEY, R. **Agile Data Science: Building Data Analytics Applications with Hadoop**. [s.l: s.n.].

Assinaturas

Recife, 16 de setembro de 2016.

Orientador

Fernando da Fonseca de Souza

Aluno

Bruno de Assis Pereira