2ª JORNADA DE INTEGRAÇÃO E INICIAÇÃO CIENTÍFICA
5 A 9 DE JUNHO 2017
FACULDADE CESUSC
Muito além do diploma

UTILIZANDO O BIG DATA PARA ACOMPANHAMENTO DE LICITAÇÕES PÚBLICAS

Ricardo Bresolin[1]; Ibsem Dias[2]; Cristiano Schwening[3]; Gustavo Ribeiro[4].

2ª Jornada de Integração e Iniciação Científica – Faculdade Cesusc – Santa Catarina – SC – Brasil

INTRODUÇÃO

O Grupo de Pesquisa *BIG DATA* do Curso de Análise Desenvolvimento de Sistemas da Faculdade Cesusc uniu-se ao Observatório Social de São José – OSSJ – visando contribuir na concepção de estrutura tecnológica (bases de dados e ferramentas de pesquisa) para gerar informações que o auxiliem na análise das informações disponibilizadas sobre a gestão pública. O OSSJ se caracteriza como um espaço democrático e apartidário que reúne em seu corpo técnico entidades representativas da sociedade civil, com objetivo de contribuir na melhoria da gestão pública, atuando em favor da transparência e da qualidade na aplicação dos recursos públicos. O negócio do OSSJ visa a fiscalização da gestão pública municipal no que tange à aplicação dos recursos públicos.

OBJETIVO

O projeto do grupo de pesquisa proposto tem por objetivo desenvolver pesquisa a dados de gestão pública com técnicas de *Big Data* e *Data Science* por meio da aplicação de métodos e ferramentas livres para extração, tratamento, estruturação e análise de grandes volumes de dados.

DESENVOLVIMENTO PESQUISA

Quanto aos requisitos iniciais: em reunião inicial com os representantes do OSSJ foi definido o escopo de atuação priorizada nas fontes de dados de licitações da prefeitura municipal de São José. O grupo recebeu arquivos texto tabulados originados pela prefeitura e entregues periodicamente para o Ministério Público no formato representado na Figura 1. Estes arquivos contém separadamente dados de licitações, empenhos, fornecedores, produtos, contratos, notas fiscais e pagamentos.

Figura 1 – Documento Esfinge, leiaute arquivo empenho.txt

```
020010000000000003215339014140100000000020000020000000000750.00PELA DESPESA EMPENHADA REFERENTE A 1
                                                              012014070 1Sem Licitação   Sem Contrato
020010000000000003311339033010100000000020000020000000005552.15PELA DESPESA EMPENHADA REFERENTE A CO
                                                              012014070 4PR87/2013       Sem Contrato
020010000000000003343339014140100000000020000020000000001125.00PELA DESPESA EMPENHADA REFERENTE A 1
```

Fonte: autores da pesquisa (2017).

Para a identificação das colunas destes arquivos foi disponibilizado um arquivo denominado *Esfinge* que contém o leiaute da cada arquivo *TXT*. A Figura 2 representa o leiaute de um dos arquivos fornecidos com a identificação das colunas, o tamanho em caracteres, o tipo de dado, além de uma breve descrição do conteúdo.

Figura 2 – Documento Esfinge, leiaute arquivo empenho.txt

Empenho.txt

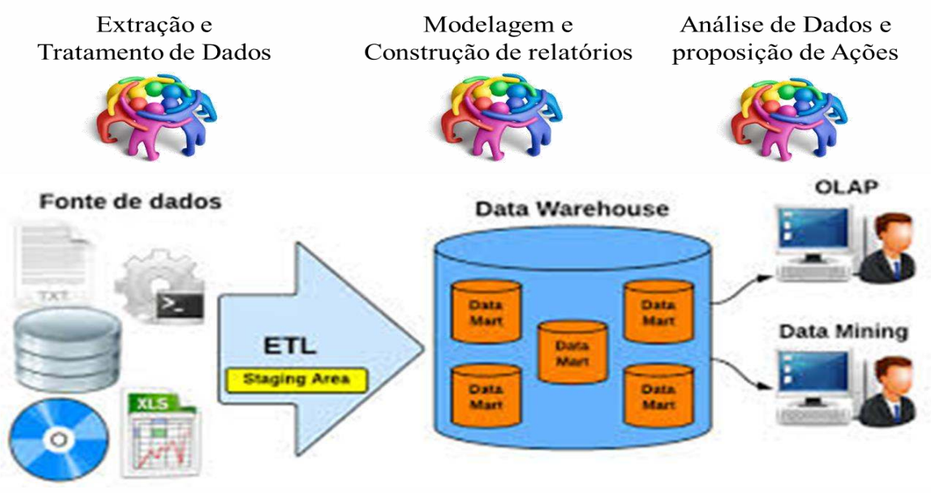
Descrição do campo	Tipo	Tamanho	Observação
#Código da Unidade Orçamentária	Numérico	5	Código da unidade orçamentária a qual o empenho pertence. Obrigatório
#Número do Empenho	Numérico	16	Número seqüencial do empenho dado pela unidade. Obrigatório

Fonte: autores da pesquisa (2017).

Quanto à estruturação do projeto: em reunião com o grupo de alunos interessados, foram apresentadas três grandes áreas de atuação para a construção do projeto conforme Figura 3 e identificadas as áreas de interesse e de conhecimento de cada aluno.

Estes foram divididos em 3 times com atividades de pesquisa e desenvolvimento afins a cada área.

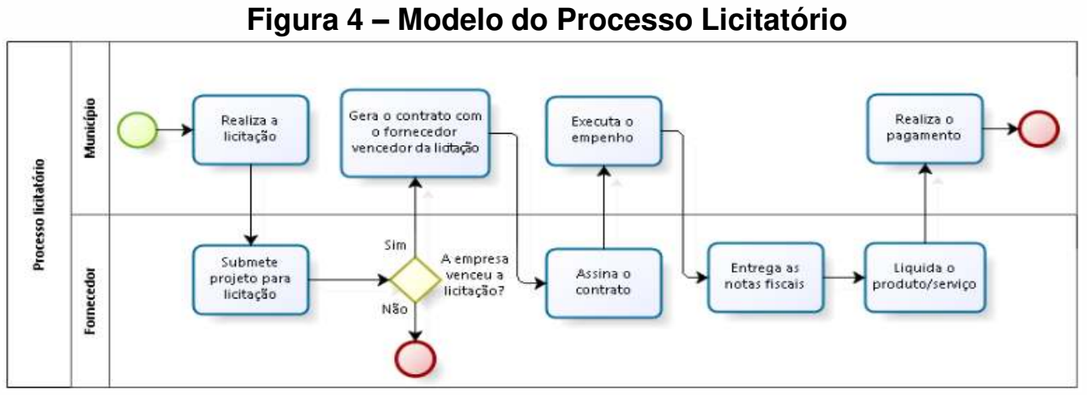
Figura 3 – Áreas de atuação e Arquitetura do Big Data



Fonte: autores da pesquisa (2017).

Quanto à execução das atividades: no time de Tratamento de Dados a primeira etapa a ser desenvolvida aborda a criação de uma tabela com os dados dos leiautes dos arquivos fontes que serão utilizados, o estudo da linguagem de programação Python para o desenvolvimento das Rotinas de ETL (Extração, Tratamento e Carga), a implementação de um algoritmo para extração dos arquivos *TXT*, a carga destes em arquivos com extensão *CSV* e posterior transferência em tabelas de Bancos de Dados. O time de Modelagem e Construção de Relatórios se uniu ao time de Análise de dados para junto com os representantes do OSSJ modelar o *DataWare* que irá armazenar os dados e compreender como funciona o processo licitatório. Das reuniões realizadas foram gerados três artefatos que são base para as análises e codificações do time de extração e tratamento de dados: o Modelo Dimensional, a matriz Fato/Dimensão e o Modelo do Processo Licitatório apresentado na Figura 4.

Figura 4 – Modelo do Processo Licitatório



Fonte: autores da pesquisa (2017).

CONSIDERAÇÕES FINAIS

As reuniões realizadas no grupo de pesquisa com os responsáveis pelo OSSJ proporcionaram a construção do modelo conceitual do processo licitatório, a concepção de um modelo dimensional de um *DataWare* e o desenvolvimento de um algoritmo para extração e conversão de arquivos em formato *TXT* para *CSV*. Para os próximos passos, o grupo vai desenvolver as rotinas para automatizar o processo de Extração, Tratamento e Carga dos dados e realizar pesquisa de ferramentas para a análise das informações geradas.

1 Titulação: Graduando em ADS Instituição atual: Faculdade Cesusc/ ricardo.bresolin@gmail.com
2 Titulação: Mestre em Administração. Instituição atual: Faculdade Cesusc /ibsem.dias@gmail.com
3 Titulação: Mestre em Computação Aplicada. Instituição atual: Faculdade Cesusc /cristiano.schwening@faculdadecesusc.edu.br
4 Titulação: Pós-Graduando em Finanças, Investimentos e Banking Instituição atual: Faculdade PUCRS/ gustavoribeiro@gmail.com

REFERÊNCIAS BIBLIOGRÁFICAS

GONÇALVES, Marcio. **Extração de dados para Data Warehouse**. Rio de Janeiro/RJ: Axcell Books, 2003.
HARRINGTON, Jan L. **Bancos de Dados Relacionais:** teoria e prática. 2002.
SERRA, Laércio. **A essência do Business Intelligence**. São Paulo: Berkeley, 2002.