



Programa de Pós-Graduação em Computação Aplicada - PPGCAP

Aplicação de Técnicas de BI em uma Instituição de Ensino Superior do Rio Grande do Sul

Alunos: Fernando Fagonde e Luciano Brum.

Disciplina: Sistemas Avançados de Banco de dados.

Professor: Dr. Sandro da Silva Camargo.

Roteiro

- Conceitos Básicos de *Business Intelligence*;
- Objetivo;
- Metodologia;
- Resultados, Discussões e Dificuldades;
- Conclusões;

Conceitos Básicos



Figura 1: Arquitetura de uma solução de BI. Fonte: <http://rafaelpiton.com.br/data-warehouse-o-que-e/>

Conceitos Básicos

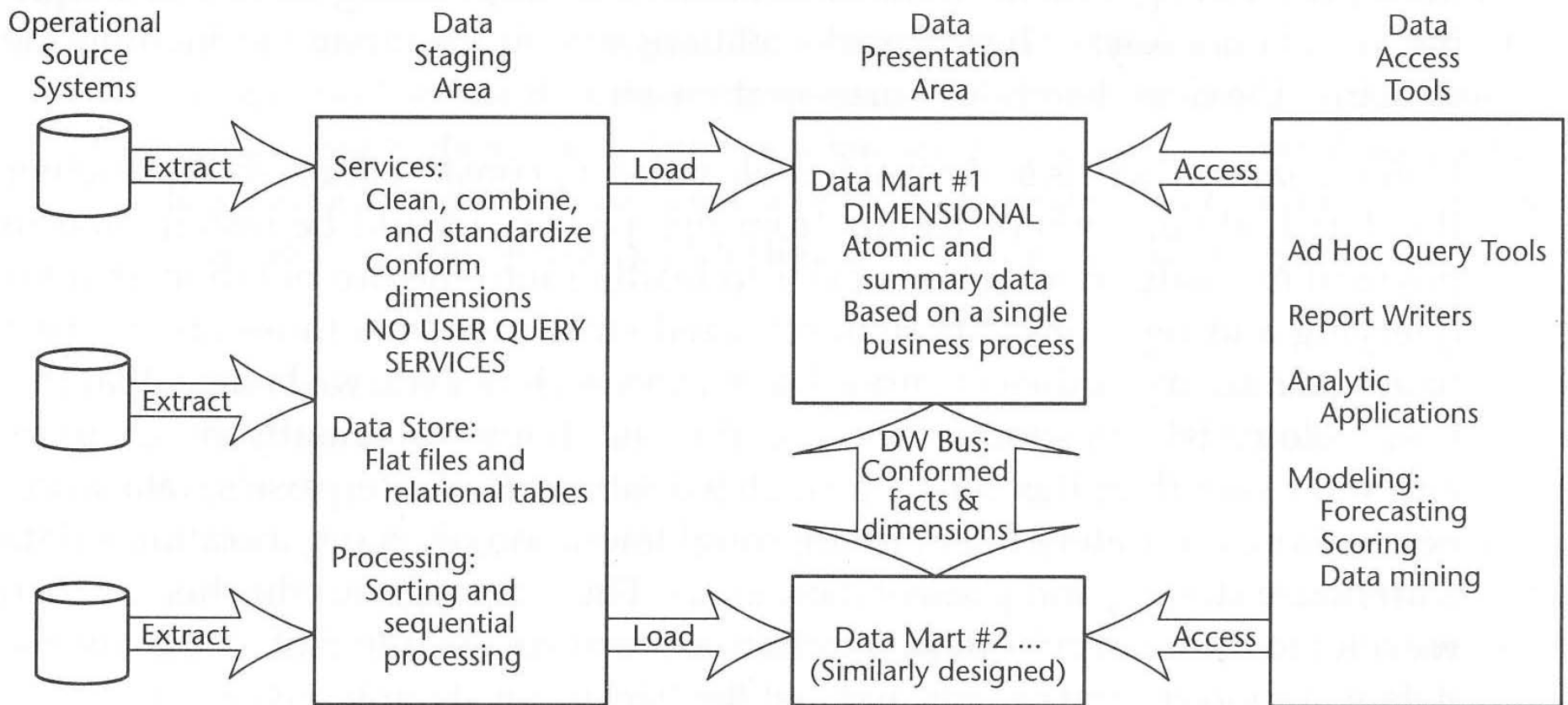


Figura 2: Elementos Básicos de um DW. Fonte: Kimball e Ross, 1996.

Conceitos Básicos



Figura 3: Indicadores-chave de performance (do inglês: *key-performance indicators* - KPI). Fonte: Elaborada pelos autores, 2017..

Conceitos Básicos

What they are:

- Quantifiable/measurable and actionable
- Measure factors that are critical to the success of the organization
- Tied to business goals and targets
- Limited to 5-8 key metrics
- Applied consistently throughout the company

What they are not:

- Metrics that are vague or unclear
- “Nice-to-know’s” or metrics that are not actionable
- Reports (e.g., top search engines, top keywords)
- Exhaustive set of metrics
- Refutable

Conceitos Básicos



Figura 4: Indicadores-chave de performance. Fonte: Elaborada pelos autores, 2017.

Conceitos Básicos

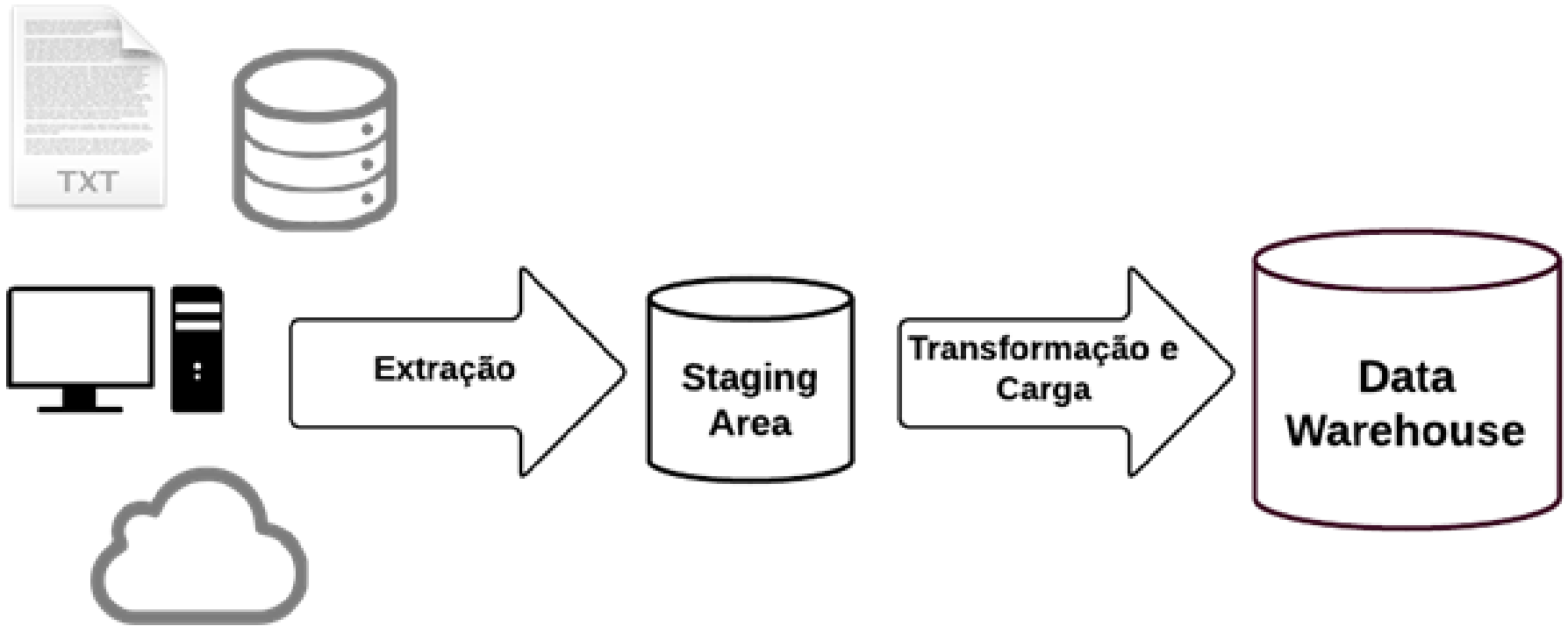


Figura 5: Processo de ETL (Extração-Transformação-Carga). Fonte: <https://corporate.canaltech.com.br/noticia/business-intelligence/entendendo-o-processo-de-etl-22850/> (2017).

Conceitos Básicos

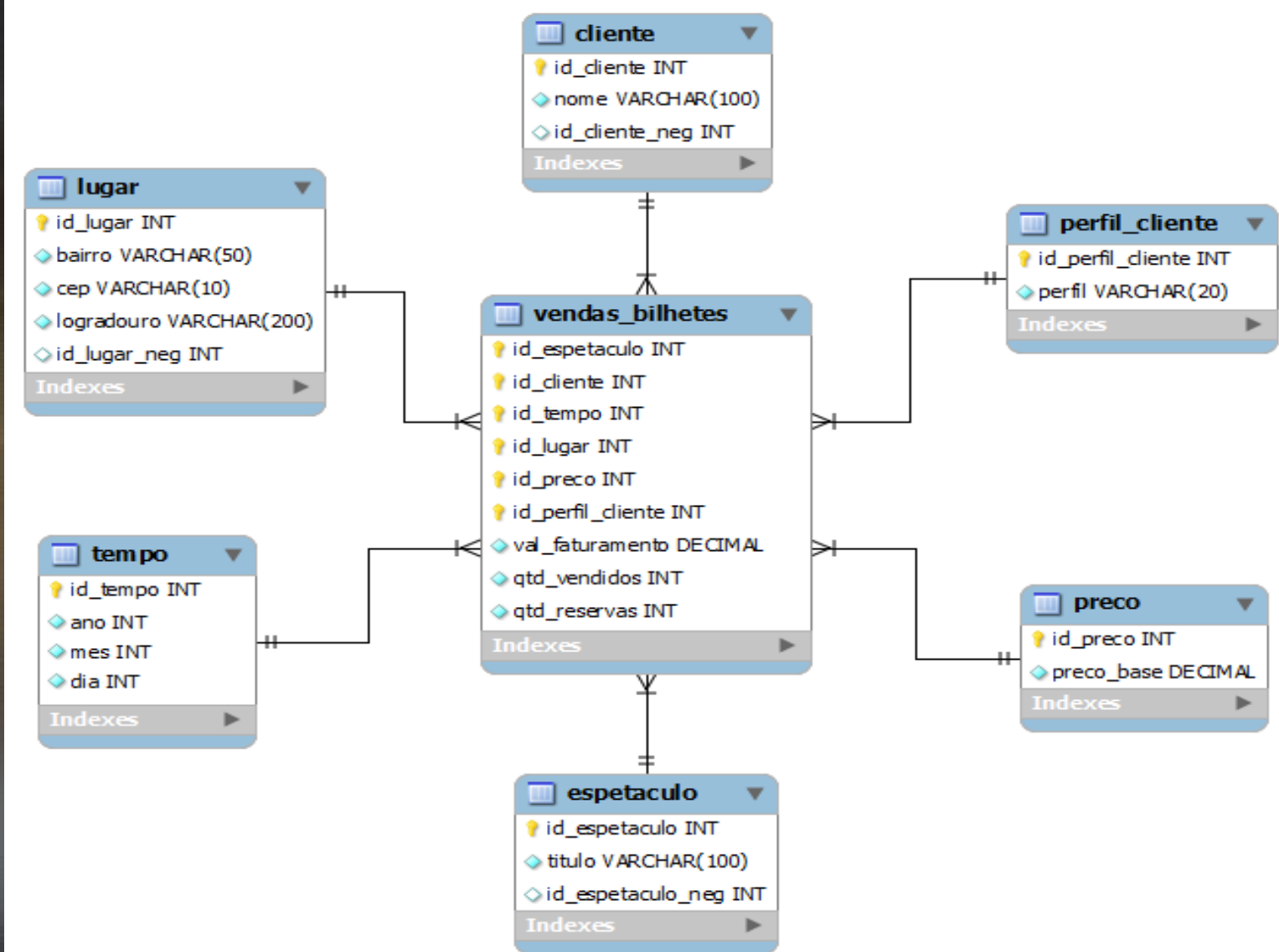


Figura 6: Exemplos de tabelas fato e dimensões (modelagem dimensional do DW). Fonte: http://bdtickets.blogspot.com.br/2012_07_01_archive.html/ (2017).

Objetivos

- Definição dos *KPI's* com os gestores da IES para a solução de *BI*;
- Definição da granularidade dos dados;
- Definição da latência dos dados;
- Definição das ferramentas a serem utilizadas;
- Projeto do *Data Warehouse*;
- Processo de mapeamento *ETL* para o *Data Warehouse*;

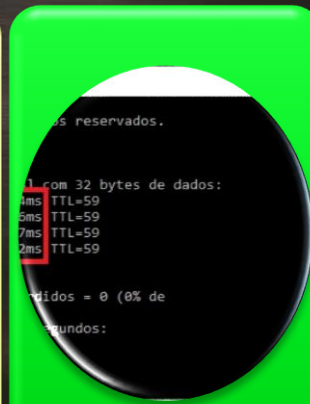
Metodologia



Definição
dos *KPIs*.



*Definição da
Granularidade.*



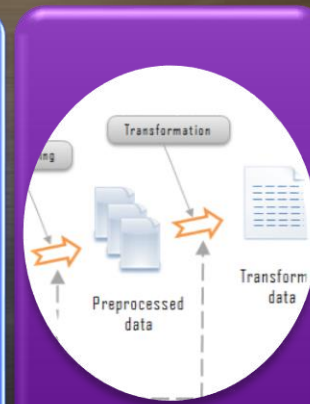
*Definição
da latência.*



*Definição de
ferramentas.*



*Projeto do
DW.*



*Mapeamento
ETL.*



Metodologia



Definição dos *KPIs*



- **Evolução de matrículas.**
- Evolução dos créditos brutos.
- Evolução de bolsas por curso.
- Evolução de créditos educativos.
- Evolução do número de créditos por curso.
- Evolução da inadimplência.
- Evolução dos Trancamentos e Cancelamentos.

Metodologia



Definição da Granularidade



- Definição do maior nível de detalhamento que podemos chegar.
- No caso específico deste trabalho: Quantidade de alunos por curso, campus, turno e período.

Metodologia



Definição da Latência



- Foi utilizada a latência de 1 dia (D-1) para que as tabelas do DW estivessem com dados disponíveis para os tomadores de decisão.

Metodologia



Definição das
Ferramentas



➤ PostgreSQL v. 9.4 (DW).

➤ Pentaho Data Integration v. 7.0 (ETL).

Metodologia



Projeto do
DW



- Modelagem Dimensional.
- Esquema Conceitual.
- Esquema Lógico (ER) das tabelas fato e dimensões (*star schema*).

Metodologia

- Esquema Conceitual:

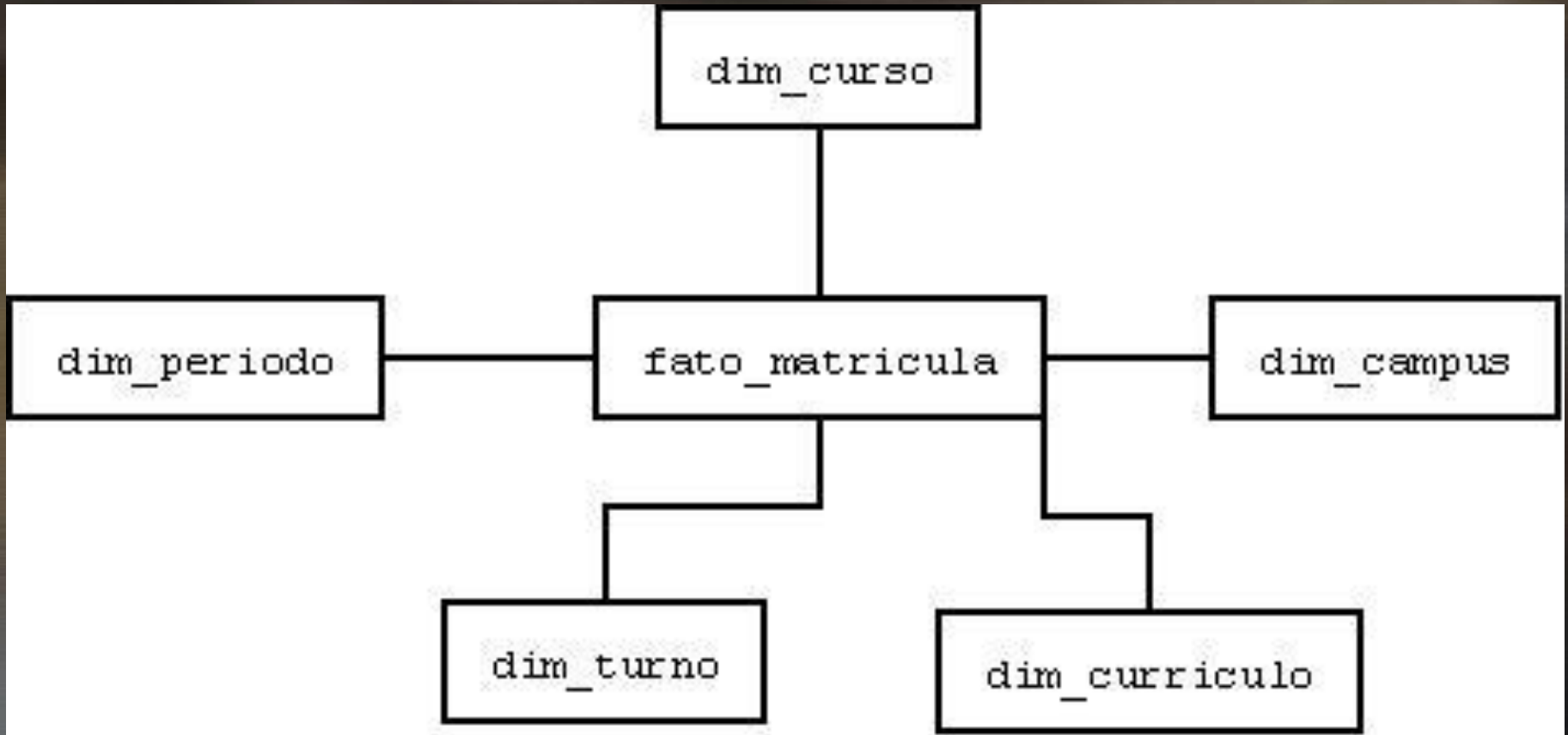


Figura 7: Modelo conceitual do DW. Fonte: Elaborada pelos autores, 2017.

Metodologia



Processo de ETL



- Carga dos dados da base de dados para as dimensões (Extração - Carga).
- *Slowly changing dimensions.*
- Carga dos dados da base de dados para a tabela fato (Extração – Carga)
- Transformações: *PK* para *SK*, consulta *substr.*

Metodologia

- Mapa ETL das dimensões

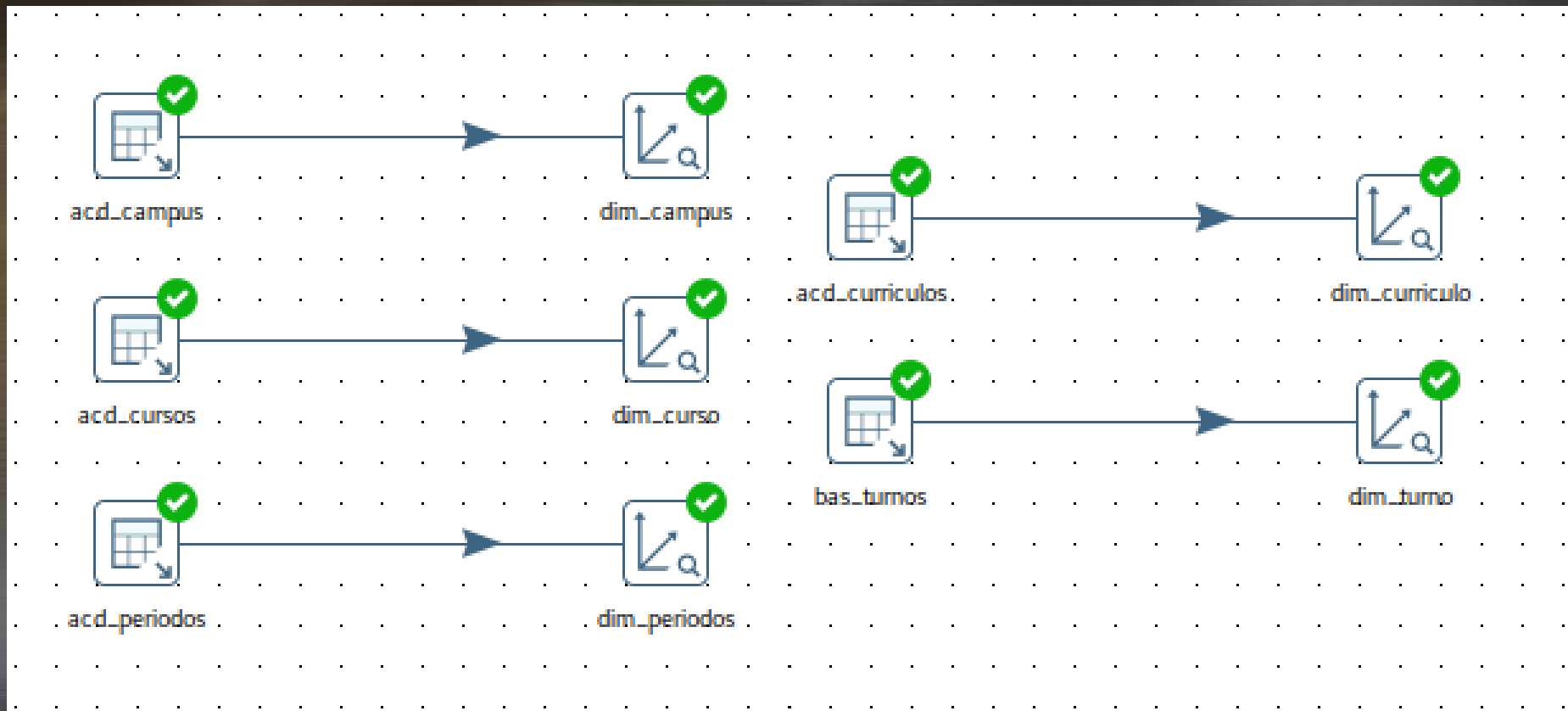


Figura 9: Mapa ETL das dimensões. Fonte: Elaborada pelos autores, 2017.

Metodologia

- Mapa ETL das dimensões:
- Busca os dados das tabelas de origem;
- Verifique se há alterações nos dados existentes nas dimensões;
- Insere os valores diferentes nas dimensões;
- *Surrogate Key;*

Metodologia

- Estatísticas (filtro de problemas):

Nome do step	Copia nr	Lidos	escritos	Entrada	Saída	Atualizados	Rejected	Erros	Ativo	Tempo	Velocidade (r/s)
bas_turnos	0	0	6	6	0	0	0	0	Finished	0.0s	167
acd_periodos	0	0	220	220	0	0	0	0	Finished	0.0s	4.583
acd_campus	0	0	9	9	0	0	0	0	Finished	0.1s	123
acd_curriculos	0	0	636	636	0	0	0	0	Finished	0.1s	7.663
acd_cursos	0	0	47	47	0	0	0	0	Finished	0.1s	635
dim_campus	0	9	9	9	0	0	0	0	Finished	0.2s	58
dim_periodo	0	220	220	220	0	0	0	0	Finished	0.3s	756
dim_curriculos	0	636	636	636	0	0	0	0	Finished	0.5s	1.322
dim_turno	0	6	6	6	0	0	0	0	Finished	0.1s	42
dim_cursos	0	47	47	47	0	0	0	0	Finished	0.2s	258

Figura 10: Estatísticas do mapeamento das dimensões. Fonte: Elaborada pelos autores, 2017.

Metodologia

- Estatísticas (filtro de problemas):
 - Quantidade de registros da origem;
 - Quantidade de registros gravados no destino (quando há);

Metodologia

- Mapa ETL da tabela fato:

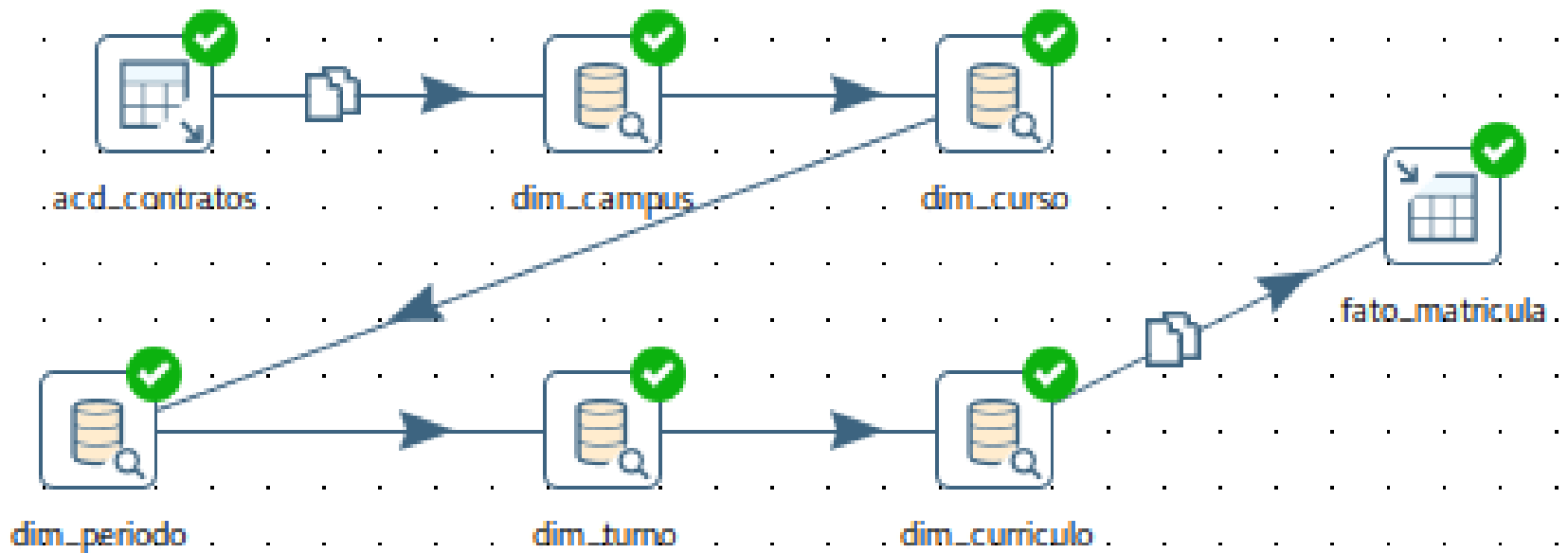


Figura 11: Mapeamento da tabela fato_maturricula. Fonte: Elaborada pelos autores, 2017.

Metodologia

- Objeto *input table*: dados de entrada, consulta ou integração com diversas fontes de dados;
- *Dimension lookups*: verifica e substitui as chaves no fluxo dos dados pelas *surrogates*;
- Objeto *Table Output*: Registra os resultados na tabela fato;

Metodologia

- Estatísticas (filtro de problemas):

Nome do step	Copia nr	Lidos	escritos	Entrada	Saída	Atualizados	Rejected	Erros	Ativo	Tempo	Velocidade (r/s)
acd_contratos	0	0	15788	15788	0	0	0	0	Finished	21.7s	728
dim_campus	0	15788	15788	15788	0	0	0	0	Finished	25.8s	613
dim_curso	0	15788	15788	12125	0	0	0	0	Finished	26.4s	598
dim_periodo	0	15788	15788	15788	0	0	0	0	Finished	27.0s	584
dim_turno	0	15788	15788	15788	0	0	0	0	Finished	27.1s	583
dim_curriculo	0	15788	15788	0	0	0	0	0	Finished	27.1s	583
fato_matricula	0	15788	15788	0	15788	0	0	0	Finished	28.2s	560

Figura 12: Estatísticas do mapeamento da tabela fato_matricula. Fonte: Elaborada pelos autores, 2017.

Resultados e Discussões

- 15788 registros no sistema de origem (*acd_contratos*) e 15788 registros inseridos na tabela fato.
- Esse fato ocorre em virtude dos dados estarem trabalhados na tabela de origem.
- Resta para o mapeamento *ETL* as *lookups* com as dimensões e inserção dos dados na tabela fato.

Resultados e Discussões

- Dos indicadores definidos com a gestão, foi trabalhado apenas o Evolução de alunos matriculados por período.
- Uma consulta SQL com um *inner join* da fato com as dimensões, possibilita a visualização dos totais com as descrições dos indicadores analisados.
- O resultado desta consulta pode ser inserido em alguma ferramenta de *dashboard* que permitiria a visualização dos dados por parte da área cliente responsável pela tomada de decisão.

Dificuldades

- Durante o processo de LOOKUP com as tabelas dimensões, foi detectado um problema com o dimensionamento da janela de propriedades do passo, que impede o usuário de executar o procedimento.
- Um segundo problema detectado é que ao executar o ETL, como ferramenta de depuração dos resultados utilizou-se um elemento *text file output*, que, em sua definição padrão utilizava para colunas do tipo *string* o tamanho de 2147483647 bytes (*Out Of Memory Error*).

Conclusões

- Este trabalho auxiliou na solução ao criar um ambiente que permite o registro de dados históricos, identificação de gargalos, riscos, oportunidades, comparações, entre outras possibilidades.
- Como trabalhos futuros, sugere-se a utilização de ferramentas que permitam a análise e visualização das informações consolidadas no DW e ainda, aplicação de técnicas de mineração de dados para obtenção de conhecimento que auxilie nos processos decisórios e gerenciais da IES.

Referências Bibliográficas

- BARBIERI, C. BI2-Business Intelligence. 1ª ed. RJ: Elsevier, 2011.
- HAN J.; MICHELINE, K. Data Mining Concepts and Techniques. Morgan Kauffman, 2001.
- INMON, W. H. *Building the Data Warehouse*. JohnWiley & Sons, 1996.
- KIMBALL, R.; ROSS, M.. The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling. 1996.
- SINGH, H. S. Data Warehouse: Conceitos, Tecnologias, Implementação e Gerenciamento. 1ª ed. São Paulo: Makron Books, 2001.
- WIJAYA, R.; PUDJOATMODJO, B. An overview and implementation of extraction-transformation-loading (ETL) process in data warehouse (Case study: Department of agriculture). *3rd International Conference on Information and Communication Technology (IColCT)*. p. 70-74, 2015.