

Armazém de dados para o Censo da Educação Superior: uma experiência no Centro de Computação da UFMG

Patrícia Nascimento Silva¹, Josemar Pereira dos Santos¹, Vitor Fonseca de Melo¹

¹Centro de Computação – Universidade Federal de Minas Gerais (UFMG)
Minas Gerais – MG – Brasil

{patricians, josemar, vitorfm}@ufmg.br

***Abstract.** This article describes the process of planning and constructing a warehouse for the Higher Education Census from the existing databases at the Federal University of Minas Gerais (UFMG). In view of the volume and complexity of the information requested by CENSO, the UFMG's Computing Center implemented a data warehouse to assist in the extraction of information and the monitoring of data reported annually. When using the solution in CENSO 2016 it is possible to see an optimization in the tasks to be performed and greater possibility of validation of the data to be sent, since the warehouse allows a better visualization and cross-referencing of information.*

***Resumo.** Este artigo descreve o processo de planejamento e construção de um armazém para o Censo da Educação Superior a partir das bases de dados existentes na Universidade Federal de Minas Gerais (UFMG). Diante do volume e da complexidade das informações solicitadas pelo CENSO o Centro de Computação da UFMG implementou um armazém de dados para auxiliar na extração de informações e na conferência dos dados reportados anualmente. Ao utilizar a solução no CENSO 2016 percebe-se uma otimização nas tarefas a serem realizadas e maior possibilidade de validação dos dados a serem enviados, já que o armazém permite uma melhor visualização e cruzamento de informações.*

1. Introdução

O CENSO da Educação Superior divulga informações sobre o ensino superior sendo realizado anualmente pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). A coleta de dados para o CENSO é realizada através de um sistema web onde são preenchidos os dados solicitados às Instituições de Ensino Superior (IES), desta forma cada IES precisa reunir todas as informações solicitadas e enviá-las dentro do prazo estipulado.

Diante do volume e da complexidade das informações solicitadas pelo CENSO o Centro de Computação da Universidade Federal de Minas Gerais (CECOM - UFMG) implementou um armazém de dados para auxiliar na extração de informações para o Censo, otimizando a conferência dos dados reportados anualmente pela UFMG. Para o armazém de dados monta uma base de dados alimentada periodicamente com as informações necessárias para o Censo e demais informações relevantes para os gestores da Universidade. Este armazém foi desenvolvido utilizando soluções de software livre como: Mondrian, Kettle e Pentaho BI Server.

O desenvolvimento desta solução foi iniciado em 2015 e está em utilização para o CENSO de 2016 que deverá ser enviado até abril de 2017. O objetivo deste artigo é apresentar o processo de planejamento e construção do armazém para o Censo que auxilia os diversos atores envolvidos na coleta dos dados.

2. Métodos

Este artigo é um estudo de caso da implementação realizada no CECOM - UFMG para a coleta de dados para o CENSO 2016. Uma pesquisa bibliográfica foi realizada a fim de identificar os conceitos sobre *data warehouse* e ferramentas disponíveis. Além disso, foram revisados os arquivos e documentos relativos ao CENSO, assim como seu processo de forma geral. Identificou-se que o processo para envio dos dados para o CENSO envolve três etapas: A primeira etapa corresponde a compatibilização entre os conceitos do CENSO e os conceitos da UFMG, a segunda etapa está relacionada a coleta dos dados e a terceira etapa que corresponde criação e envio do arquivo. Neste contexto, o processo de elaboração do CENSO foi revisto dentro da Universidade cabendo ao CECOM - UFMG o desenvolvimento de uma solução tecnológica para auxiliar todos os órgãos envolvidos no processo.

3. Resultados

Neste capítulo serão apresentadas as etapas do processo para envio dos dados para o Censo.

3.1. Compatibilização de conceitos do CENSO e da UFMG

O CENSO disponibiliza um glossário com todos os conceitos solicitados anualmente. Os conceitos e classificações atribuídas aos alunos nos sistemas acadêmicos da UFMG podem apresentar diferenças, desta forma foi necessário comparar todos os conceitos do CENSO com os conceitos da UFMG e fazer adaptações nos conceitos que apresentavam alguma divergência. Estas alterações incluíam a reclassificação ou a criação de novos atributos para adaptação ao CENSO. Como resultado desta etapa foi desenvolvida uma documentação única com as informações instrumentais relevantes para coleta e verificação dos dados.

3.2. Coleta de dados

Uma das maiores dificuldades encontradas no CENSO está relacionada à coleta dos dados. Por ter informações abrangentes, o CENSO utiliza dados de diversas bases de dados e, além disso, estes dados precisam ser convertidos e tratados para a formatação solicitada. Com isso, a solução tecnológica proposta pelo CECOM – UFMG foi o desenvolvimento de um armazém de dados que reunisse as várias bases de dados e uma interface para que o usuário pudesse visualizar / conferir os dados e gerar o arquivo a ser enviado para o CENSO através do portal do INEP. Essa solução também seria importante para definição de papéis no processo interno de coleta de dados do CENSO, já que o Centro de Computação assumia várias tarefas de coleta e formatação de dados, que agora podem ser facilmente gerenciadas por outros órgãos da Universidade.

Para construção do armazém de dados do CENSO foram realizadas várias pesquisas na literatura envolvendo a construção de armazém de dados. Conforme

INMON (1997) as etapas para construção de um *Data Warehouse* (DW) são: (1) Análise de Dados: Elaborar o modelo de dados para o DW; (2) Dimensionamento: Estimar o volume de dados do DW; (3) Avaliação Técnica: Definição das configurações técnicas; (4) Preparação do Ambiente Técnico: Identificar, tecnicamente, como a configuração definida será acomodada; (5) Análise das áreas de Interesse: Seleção da área de negócio para ser povoada; (6) Projeto do DW: Projeto físico do banco de dados; (7) Análise do Sistema Fonte: Identificar, nos sistemas existentes, a fonte de dados para o DW, e efetuar um mapeamento entre eles; (8) Especificações: Definir, em especificações de programas, as rotinas para carga dos dados; (9) Programação: Codificação das especificações para as rotinas de carga e (10) Povoamento: Execução das rotinas de carga do DW.

Com base nestas etapas foi elaborado o armazém de dados do Censo com o seguinte projeto físico do banco de dados, conforme apresentado na Figura 1:

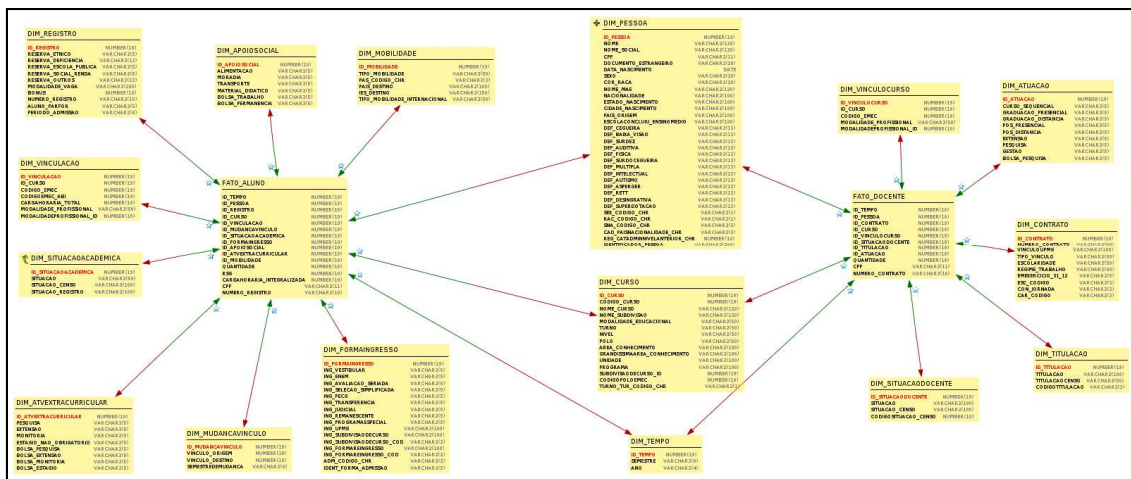


Figura 1: Modelo Físico Armazém do CENSO

Fonte: Próprio autor

A partir do modelo físico é possível obter uma visualização dos eventos que o compõem e seus relacionamentos. A Figura 1 apresenta algumas tabelas dimensão e fato do modelo. Para a extração, transformação e carga dos dados foi usada a ferramenta Kettle, para criar os cubos foi utilizada a ferramenta Mondrian e para a visualização foi utilizada a ferramenta Pentaho. Hoje a atualização do armazém é feita sob demanda quando há alterações nos sistemas em produção ou correções de erros.

3.3. Criação e envio do arquivo para o CENSO

O armazém de dados possibilitou reunir os dados de diversas fontes de informação. Para facilitar a gestão dessas informações foi desenvolvida uma interface para gestão do arquivo do Censo. Nesta interface é possível fazer a importação manual de arquivos que ainda não estão disponibilizados em bancos de dados, gerar o arquivo para importação no site do INEP e visualizar/conferir as informações do armazém de dados do Censo.

Através da definição dos processos foi possível identificar melhor os papéis dos envolvidos. Através da ferramenta Pentaho foi criado um ambiente de visualização dos dados do armazém para maior independência dos usuários que podem conferir e validar a qualidade dos dados.

3.4 Processos e Artefatos

A figura 2 apresenta o novo processo para geração do arquivo CENSO considerando as soluções desenvolvidas:

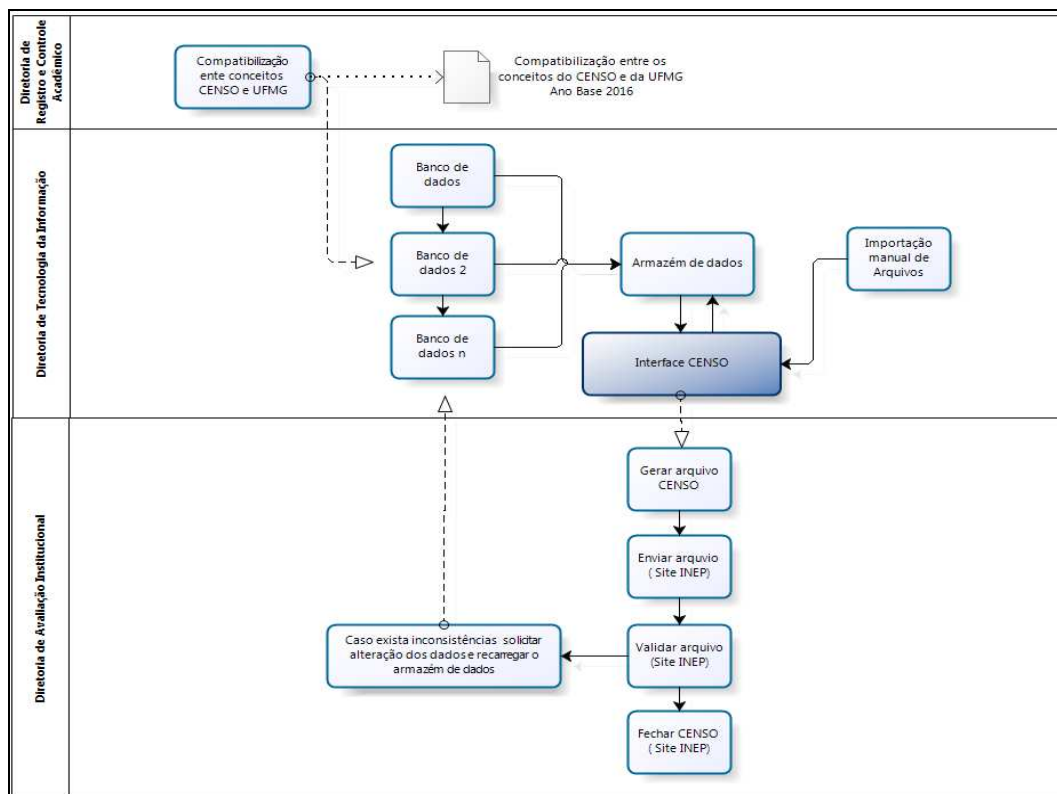


Figura 2: Processo CENSO

Fonte: Próprio autor

A Figura 3 apresenta a tela da aplicação desenvolvida e seus menus. O sistema faz algumas validações e correções básicas do arquivo. A figura 4 apresenta a funcionalidade relatórios que permite ao usuário diversas formas de visualização dos dados do armazém. Essa funcionalidade permite que o usuário valide as informações garantindo maior qualidade ao processo.

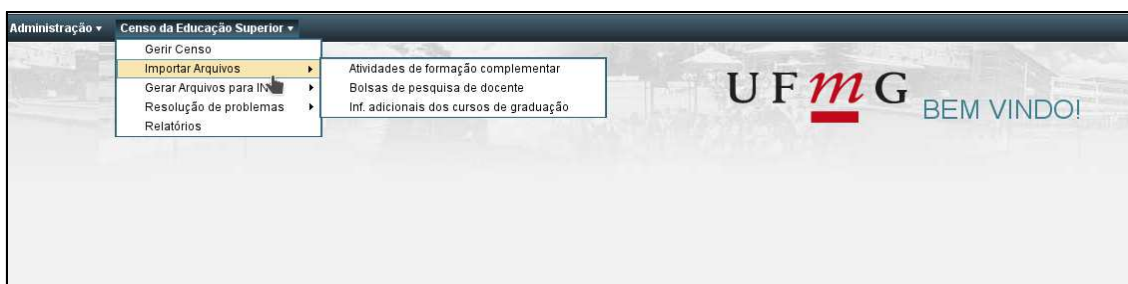


Figura 3: Interface CENSO

Fonte: Próprio autor

ARMAZÉM DE DADOS

SELECIONAR RELATÓRIO: ▾

Censo: Relatórios de vagas de curso

Filtros OCULTAR FILTROS

Semestre: 1966/2, 1966/1, TODOS

Código Emec: TODOS, 12941 - ADMINISTRAÇÃO-I.2001PD001/PRESENCIAL/DIURNO, 12941 - ADMINISTRAÇÃO-I.2001PN002/PRESENCIAL/NOTURNO, 12942 - ARQUITETURA E URBANISMO-09001PD001/PRESENCIAL/DIURNO

Forma de visualização: Tabela Gráfico barra Gráfico pizza PESQUISAR

Figura 4: Relatórios interface CENSO

Fonte: Próprio autor

4. Conclusão

O quantitativo de informações a serem coletadas no CENSO é um desafio para as universidades já que as informações precisam ser tratadas e organizadas no formato solicitado pelo INEP. A diversidade de sistemas coexistentes e as diferentes formas de gestão vivenciada em cada órgão da Universidade aumentam a complexidade para a centralização das informações. O desenvolvimento de um armazém de dados foi a solução encontrada para centralizar o tratamento da informação e automatizar o processo de coleta de dados. Além disso, foi desenvolvida uma interface para gestão de todas as etapas que envolvem o processo de envio do arquivo para o INEP.

O armazém de dados Censo foi implementado pelo CECOM – UFMG em 2016 e mesmo estando em fase de teste, ele já foi utilizado na geração dos dados para o CENSO 2016. Com sua utilização foi possível coletar os dados de forma mais rápida, o que possibilitou maior conferência e validação dos dados a serem enviados. A partir da revisão do processo os papéis dos atores envolvidos foi melhor definido, contribuindo para a comunicação e correção das não conformidades encontradas. As divergências encontradas estão sendo tratadas para melhoria e evolução do armazém de dados para as próximas edições do CENSO.

Referências

INMON, W. H. Como construir o Data Warehouse. Rio de Janeiro: Campus, 1997.