



ITI/CNPq-MAI/DAI

## Criação de um Data Warehouse para avaliação do impacto dos projetos de ação social da Sicredi Pioneira na qualidade educacional em escolas

Autores: Eduardo Eberhardt Pereira, Bianca Libardi, Ana Cristina Fachinelli Bertolini e Daniel Luis Notari



### INTRODUÇÃO / OBJETIVO

A construção de soluções que usam dados para apoio à decisão, como data warehouses, são fundamentais para fundamentar decisões de gestores. Essas soluções tornam-se especialmente poderosas quando associadas ao uso de dados abertos.

Data warehouse é um repositório centralizado de informações provenientes de diversos setores distintos, utilizado para apoiar decisões estratégicas a partir da extração de análises e informações importantes (VIDA et. al, 2021).

Dados abertos, por sua vez, são dados que podem ser usados, estudados e modificados sem restrições, e que podem ser copiados e redistribuídos com ou sem modificações igualmente sem restrição, ou com restrições que garantem que os recebedores seguintes também possam fazer todas essas ações (MURRAY-RUST, 2008).

Para desenvolver essas soluções, metodologias como *Extract-Transform-Load* (ETL) são utilizadas. ETL é o processo de busca, coleta, processamento e carregamento dos dados de forma separada. O processo de ETL garante qualidade e consistência dos dados, permite unir dados de diferentes fontes e entrega os dados em um formato pronto para apresentação em uma aplicação ou dashboard, permitindo que os usuários finais consigam usá-los para tomada de decisão (KIMBALL; CASERTA, 2004).

O objetivo do projeto, portanto, é auxiliar no desenvolvimento de um data warehouse para viabilizar a visualização e interpretação dos dados a partir de métricas baseadas em indicadores de projetos de inovação social do Sicredi. Considerando que o desenvolvimento dessas métricas permanece em construção, este trabalho apresenta um projeto para elaboração deste data warehouse, unindo dados abertos, dados sobre os projetos coletados pelo Sicredi, o processo de ETL, data warehouses e, por fim, dashboards interativos.

Essa pesquisa está vinculada ao projeto de tese, financiado pelo Programa de Bolsas de Iniciação Tecnológica e Industrial - ITI/Cnpq, vinculado ao Programa MAI/DAI UCS, com o título: Desenvolvimento e avaliação de modelos de integração de inovação social em cooperativas de crédito: uma abordagem de Design Science Research baseada em IA para o capitalismo consciente, orientado pela Prof. Dra. Ana Cristina Fachinelli Bertolini.

### RESULTADOS PARCIAIS

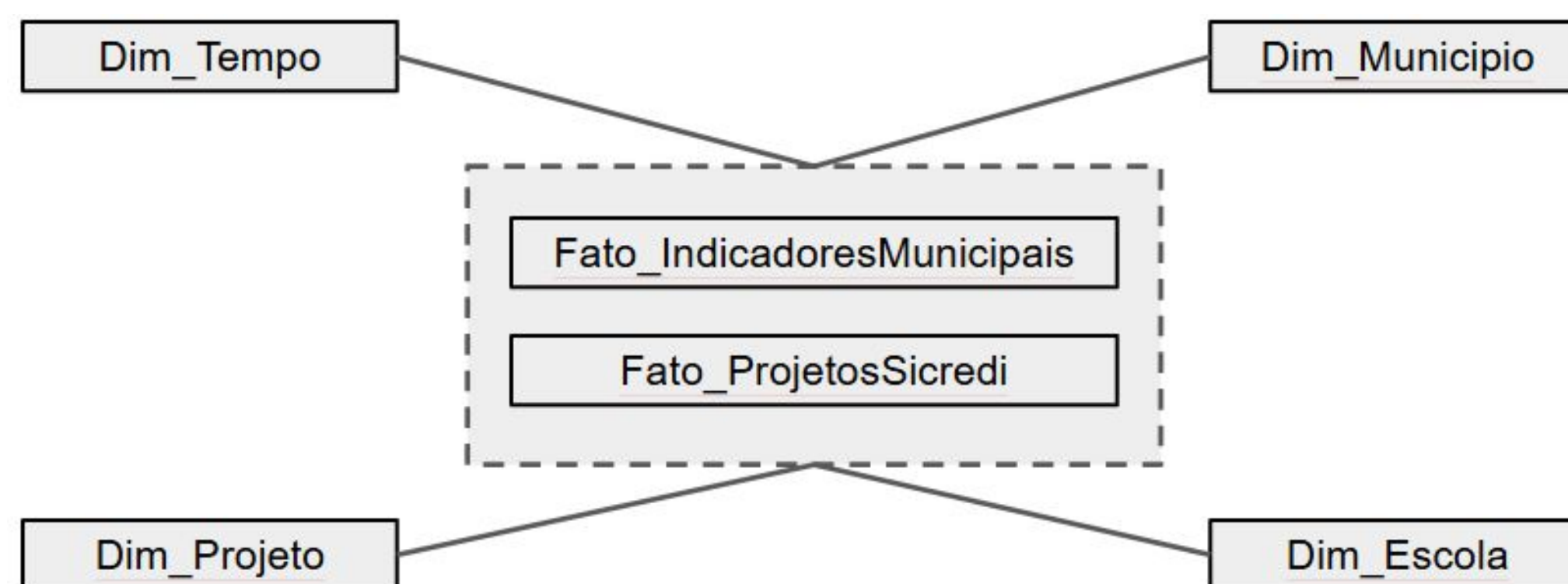
O data warehouse fará uso da modelagem em estrela, que conforme representada na Figura 1, utiliza de redundância dos dados e uma tabela central (chamada de tabela fato) para criar uma organização dos dados otimizada para consultas, facilitando a criação de cubos On-Line Analytical Processing em uma futura etapa de análise dos dados.

As tabelas de dimensão (que se conectam com a tabela fato), incluirão as dimensões de ano (data), municípios, escolas e **projetos**. A tabela fato, que contém informações mais concretas, será dividida em uma tabela para indicadores de municípios e uma para informações sobre os projetos.

Essa divisão é necessária pois as instâncias da tabela de fatos municipais não estão ligadas a projetos específicos, e dados dos projetos não estão diretamente ligados aos dados educacionais das escolas. Portanto, as informações ficam separadas e serão unidas pelo relacionamento entre elas caso uma análise queira correlacioná-las. A Figura 1 é uma demonstração detalhada dessa modelagem.

Essa modelagem é expansível e otimizada para consultas, o que permitirá análises futuras mais refinadas por meio de cubos OLAP e dashboards interativos.

Figura 1 - Modelagem em Estrela



Fonte: O autor

### MATERIAIS E MÉTODOS

Dados abertos foram explorados de forma a auxiliar na definição de indicadores para medir a evolução dos focos de cada projeto do Sicredi. Por exemplo, indicadores de educação, como o IDEB, foram utilizados preliminarmente para verificar a influência dos projetos voltados às escolas na educação.

Após a coleta dos dados abertos referentes a indicadores que poderiam ser atribuídos ao objetivo da pesquisa, o processamento desses dados foi feito utilizando bibliotecas de ciência de dados da linguagem de programação Python, como Pandas, e depois analisados em ferramentas como o R para verificar se havia relação direta da evolução das escolas com o número de alunos atingidos pelos projetos educacionais do Sicredi.

Uma vez que a métrica estiver plenamente definida, o processo irá se repetir, incorporando novos indicadores que representem as esferas dos novos projetos, além de adicionar novas variáveis relativas aos demais projetos.

Após as análises preliminares de adequação, o conjunto de dados passará por um readequação voltada a estruturá-los em um formato seguindo os padrões de Data Warehouses.

### CONSIDERAÇÕES PARCIAIS

Após a conclusão das métricas e indicadores, o *data warehouse* passará da fase de projeto e entrará em período de implementação.

### REFERÊNCIAS BIBLIOGRÁFICAS

- VIDA, Edinilson da S.; ALVES, Nicolli S R.; FERREIRA, Rafael G C.; et al. **Data warehouse**. Porto Alegre: SAGAH, 2021. E-book. p.13. ISBN 9786556901916.
- MURRAY-RUST, P. Open Data in Science. **Nat Prec**, [S.l.], v. 2008, n. 1526, 2008. Disponível em: <https://doi.org/10.1038/npre.2008.1526.1>. Acesso em: 7 mai. 2024.
- KIMBALL, R.; CASERTA, J. **The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming and Delivering Data**. Indianapolis: Wiley Publishing, Inc., 2004.
- Unceta, A., Luna, J., Castro, R., & Wintjes, R. (2018). **Social Innovation Regime: An integrated approach to measure social innovation**. European Planning Studies.
- Economist Intelligence Unit. (2016). Old problems, new solutions: Measuring the capacity for social innovation across the world.