

# Publicação de Experimentos Científicos em Plataformas de Dados Abertos

## Publication of Scientific Experiments on Open Data Platforms

Caio de Souza e Silva\*

### Resumo

Plataformas abertas são fontes de informação que se mantêm por meio de colaborações de seus usuários. Em um ambiente colaborativo é possível construir um repositório no qual se pode recuperar uma vasta quantidade de dados, os quais ajudam no desenvolvimento de novas pesquisas e experimentos. Esses repositórios podem ser utilizados para compartilhamento e análise de resultados de experimentos científicos baseados em simulações. Este trabalho propõe a facilitação da colaboração de usuários independentes, instituições e empresas que queiram compartilhar e contribuir com informações de diversos bancos de dados, utilizando um ETL (*Extract Transform Load*), também entendido como ferramentas de software que servem para extrair informações da fonte, transformar esses dados da maneira desejada, e carregá-los nos repositórios das plataformas abertas, divulgando a informação para todos aqueles que tem interesse.

Palavra-chave: Dados Abertos; Colaboração; Publicação.

### Abstract

Open Data platforms are information sources that maintain themselves by their user's collaboration. In a collaborative environment, it is possible to build a repository where we can retrieve a big amount of data, which help the development of new experiments and researches. These repositories can be used for sharing and scientific experiments result analysis on simulations. This article proposes the facilitation of collaboration from independent user's, institutes and companies that want to share and contribute with information from various data bases, using a ETL (*Extract Transform Load*), also known as software tools that are useful to extract information from a source, to transform these data to the wanted way, and to load them into the repositories of the opened platforms, publishing the information for all that have interests in it.

Trabalho de conclusão de curso apresentado ao curso de Bacharelado em Sistemas de Informação da Universidade Federal Fluminense como requisito parcial para conclusão do curso.

\* Graduando do Curso de Sistemas de Informação-UFF; caiosilva@id.uff.br

Keywords: Open Data; Collaboration; Publication.

Aprovado em: 14/12/2018 Versão final em: 29/12/2018

## 1 INTRODUÇÃO

Muitas pesquisas são executadas dentro de um instituto, gerando informações para a sociedade e especialmente para a comunidade acadêmica, porém, muitas vezes esses dados estão inacessíveis ou dependem de uma série de etapas administrativas e aprovações para que se possa acessar o repositório onde esses dados se encontram. Esse cenário se torna ainda mais complexo se considerarmos experimentos baseados em simulações computacionais que executam diversos programas científicos encadeados formando um *workflow* científico [Daniel Oliveira, 2012]. Tendo em mente que muitos desses dados não são confidenciais (já que os confidenciais não podem ser publicados) e deveriam ser divulgados abertamente (de forma a facilitar a reprodutibilidade) idealmente deveria existir um arcabouço que permitisse tal compartilhamento. Entretanto, o que vemos hoje é uma grande dificuldade em se compartilhar dados de experimentos baseados em simulação para uma análise colaborativa. Essa dificuldade, seja computacional ou de regras da instituição, acaba impactando negativamente no acesso às informações e impede que outras pessoas façam bom uso das mesmas.

Alguns experimentos geram dados constantemente, e existe uma real necessidade de ser fazer carregamentos periódicos nos respectivos repositórios, fazendo com que estes repositórios sejam fontes dinâmicas de informação, e não páginas estáticas, como páginas que falam sobre fatos históricos, imutáveis (*e.g. Wikipedia*). As plataformas abertas se mostram úteis no momento em que os dados de seus repositórios são consistentes, verídicos, atualizados e passíveis de auditoria por terceiros. Para isso, é necessário se ter toda uma gestão sobre a informação que reside nessas plataformas.

A partir de um alinhamento com o dono da informação, no caso, o autor da pesquisa ou o setor responsável por manter aquele dado, é possível definir um período de atualização das bases nas plataformas abertas. Com este alinhamento, pode-se construir um *job* (sequência de tarefas executadas por um ETL) que fará a extração dos dados da fonte, executará as transformações necessárias e realizará os carregamentos necessários. Porém, devido a possíveis erros contidos no processo, precisa-se ter certeza de que todas as tarefas do ETL foram executadas com sucesso e no momento em que deveriam ser executadas.

Em um primeiro momento, algumas regras de extração devem ser definidas, procurando adequar o momento da extração com o momento em que os dados atualizados estarão disponíveis. A partir daí os dados são extraídos e armazenados em repositórios temporários, para que dali em diante eles sejam refinados e tratados. Uma vez nestes repositórios temporários é possível enriquecer estes dados com informações provenientes de outras fontes, fazendo com que o resultado final seja ainda mais proveitoso para seus usuários.

Após a extração e tratamento desses dados será realizada o carregamento dos dados em questão. O processo consta em criar um output que seja compatível com a plataforma onde tais dados serão inseridos. A saída dos dados após as transformações pode assumir diversas formas e é compatível com muitas outras ferramentas. Um exemplo de um formato de saída, é o arquivo de texto com separadores, como um *CSV*, que tem uma estrutura parecida com a de uma tabela de banco de dados. Ao se construir este arquivo de saída, é possível usá-lo para fazer o carregamento dos dados de maneira estruturada, concatenando aos dados existentes ou simplesmente renovando toda a tabela.

Não somente textos e tabelas podem ser carregados. Pelo processo de *ETL* é possível carregar todo tipo de arquivo, contanto que a fonte desses arquivos os disponibilize e que a plataforma a qual receberá os arquivos também suporte este tipo de artefato. Uma vez no repositório, todos os arquivos estarão disponíveis aos seus usuários, permitindo que eles os usem como fonte de conhecimento para auxiliá-los em novas pesquisas.

A abordagem DColab tem por objetivo extrair, transformar e divulgar informações e dados em plataformas abertas, entendidas simplesmente como plataformas que utilizam *APIs* para receber, enviar, divulgar e compartilhar dados e fornecer serviços abertamente. Essa plataforma irá se basear em um ambiente que pode ser chamado de colaborativo, o qual conta com a colaboração e ajuda de seus usuários para se manter e crescer.

## **2 REFERENCIAL TEÓRICO**

ETL, como explicado na introdução, são ferramentas de software ou uma abordagem, responsável por realizar a extração de fontes externas, transformação para que o dado se adeque a necessidade e carregamento na base de dados alvo [Srividya K Bansal e Sebastian Kagemann, 2013], então, ele pode ser visto como uma peça chave e como o coração de todo o processo aqui proposto. O software responsável por estas etapas utiliza e separa suas

transformações no que pode ser chamado de *job*, que é a sequência de subtarefas responsável por tratar o dado de alguma forma. Dentro desses chamados *jobs*, temos alguns repositórios, que geralmente são usados como locais virtuais temporários para se armazenar dados, sendo que em alguns casos eles são permanentes.

Nesses repositórios podemos armazenar artefatos, vistos como arquivos. Como será mencionado mais a frente, arquivos na extensão CSV (*Comma Separated Values*) podem ser utilizados para armazenar dados estruturados momentaneamente, tendo em vista que eles seguem uma estrutura parecida com uma tabela de um banco de dados. São simplesmente arquivos armazenados como texto, com um separador para simbolizar as colunas de uma tabela.

Os dados estruturados mencionados anteriormente podem ser entendidos como arquivos de textos que seguem uma organização de colunas e linhas, similar a uma tabela, enquanto os dados não estruturados é tudo aquilo que não pode ser visto da forma descrita, como arquivos PDFs, DOCs, imagens, vídeos e sons. Estes ainda podem ser enviados a plataforma aberta como os dados estruturados, porém serão divulgados de maneira diferente, pois sua estrutura não é exatamente igual à de dados estruturados. A colaboração pode ser feita pelas APIs (*Application Program Interface*) citadas anteriormente, que são responsáveis pela integração entre sistemas que funcionam de maneiras diferentes, neste caso, realizando a comunicação entre os artefatos disponibilizados pelo ETL e a plataforma.

Se referindo aos dados governamentais, podemos entender os Portais de Dados Abertos como local onde as, “informações entregues à sociedade devem possibilitar o acesso automatizado sistemas externos em formatos abertos, estruturados e legíveis por máquina, o que transforma a forma do cidadão utilizar o dado ou informação” (Claudio Dutra e Karen Lopes, 2013). Porém, não precisamos ver esses portais somente como utilidade governamental, e sim, como fontes variadas de informação, no caso desta proposta, informações de pesquisas envolvendo simulações computacionais.

### **3 TRABALHOS RELACIONADOS**

A forma com que a abordagem proposta é detalhada (Seção 4) e esclarecida aqui difere um pouco de trabalhos anteriormente feitos, porém, muitos dos termos, tecnologias e serviços já foram objetos de estudo em várias outras pesquisas. Primeiramente, a respeito do ambiente colaborativo e Internet, é possível ver na pesquisa “Colaboração e interação na

WEB 2.0 e biblioteca 2.0 (Ursula Blattmann e Fabiano Couto Corrêa da Silva, 2007) que, “trata-se de um espaço interativo, de trocas, de criação e geração, além de armazenamento de informações, tornando-se uma importante ferramenta de colaboração entre os participantes do mundo digital on-line e repercute na vida de bits e átomos”. Essa citação mostra exatamente o que esta proposta pretende utilizar para divulgar os dados e as informações que estão contidas nos ambientes internos dos institutos, empresas e faculdades.

Ainda nesse estudo, os autores falam sobre a Wiki, serviço definido como “uma coleção livremente expansível de páginas Web interligadas em um sistema de hipertexto para armazenar e modificar informação – um banco de dados, onde cada página é facilmente editada por qualquer usuário com um browser” (Leuf e Cunningham (2001, p. 14)). O sistema Wiki é uma das peças chave dessa proposta, tendo em vista que a base da mesa é a colaboração. Ainda que a divulgação de informação aqui citada não seja feita da mesma forma que um usuário a faria, o objetivo é o mesmo. A grande diferença é que um usuário não teria acesso a um banco interno de um instituto por exemplo, então, alguém que esteja executando um projeto com aquele instituto, conseguiria passar por todos os trâmites necessários para ter o acesso e poderia divulgar os dados na plataforma. Uma vez online, os usuários podem conectar os dados carregados pelo projeto e fazer conexões com outros dados que achem pertinentes, construindo uma rede conectada ainda maior.

Segundo Cunningham (2006), o sistema Wiki deve atender alguns princípios, como aberto, incremental, orgânico, universal, preciso, tolerante, observável e convergente. Além disso, ele adiciona alguns princípios que são essenciais para o funcionamento desse sistema, sendo eles, a confiança, pois deve se confiar na informação dos colaboradores; divertimento, fazendo a colaboração não ser uma obrigação; e compartilhamento, como um espaço de troca.

#### **4 DCOLAB**

O processo de divulgação dos dados em plataformas abertas se inicia no momento em que o autor da informação percebe que sua pesquisa é relevante para outros pesquisadores e interessados no assunto. Ao ver que os produtos gerados por seus experimentos devem ser compartilhados a fim de gerar mais avanços em determinada área, e que esses dados são considerados públicos, o dono de tal informação poderá iniciar o processo de extração de dados de seu repositório para que estes sejam posteriormente transformados e lapidados se

necessário. Após esta etapa, os dados poderão ser carregados em uma plataforma aberta para livre consumo dos usuários.

É possível também, que o processo se inicie a partir de uma demanda pontual de informação, proveniente de qualquer fonte, desde um curioso até uma instituição de pesquisas. Neste caso, identifica-se a existência de um dado que é relevante, porém inacessível e a partir daí, é feita uma solicitação de permissão ao dono de tal pesquisa, e com sua aprovação inicia-se o processo descrito anteriormente.

#### **4.1 EXTRAÇÃO DOS DADOS DE PESQUISAS CIENTÍFICAS**

Pesquisas que lidam com dados e informações digitais muitas vezes geram dados estruturados e não estruturados. De qualquer forma os artefatos gerados como produtos podem ser armazenados, transferidos e inseridos em outros repositórios dependendo da necessidade. Para que estes sejam transferidos de um local para outro, primeiramente eles devem ser extraídos de sua fonte.

A maneira com que a informação será extraída depende de sua forma, tendo em vista que colunas e linhas não podem ser tratados do mesmo jeito que imagens, vídeos e sons. A partir dessa ideia de tratamento diferenciado, a ferramenta responsável por extrair os dados deve ser usada a fim de se construir um fluxo de transformação adequado ao tipo de dado. Logo, o fluxo do *ETL* não será genérico, e não poderá ser reutilizado para diferentes demandas e tipos de dados.

Para extração de dados estruturados pela ferramenta de *ETL*, algumas funcionalidades são necessárias para que o processo ocorra sem problemas. Primeiramente, é preciso realizar a conexão com o repositório em questão, podendo ser uma tabela em um banco de dados, um arquivo de texto estruturado ou uma planilha eletrônica. No caso de um banco de dados, o *host*, porta, *scheme*, nome do banco, usuário e senha com permissão de leitura, nome da tabela e query de extração, são informações necessárias para que se consiga o dado necessário na forma correta. Após o mapeamento dos campos e preenchimentos dos dados de acesso, será possível recolher todas as linhas com os dados pertinentes. No caso de uma fonte estática, como um arquivo de texto estruturado ou uma planilha eletrônica, a etapa de mapeamento de campos continua sendo essencial, porém a informação de conexão será diferente, sendo necessário apenas o nome do arquivo e o diretório no qual ele se localiza.

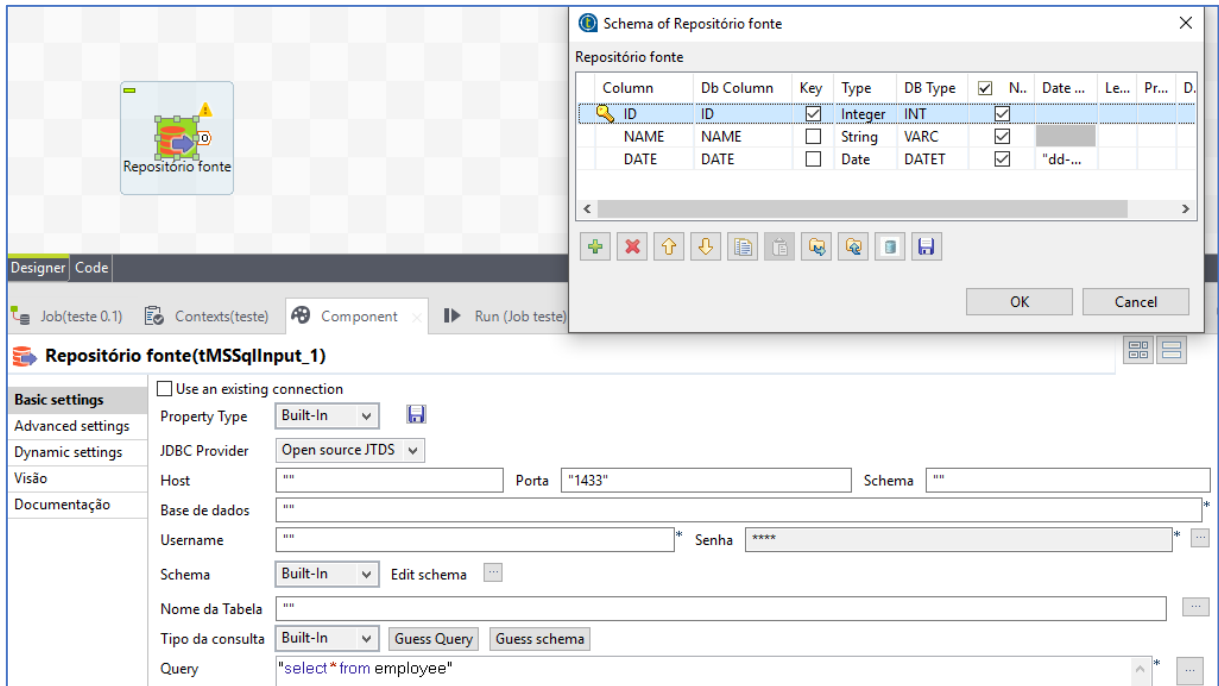


Imagem 1. – Passo responsável pela extração dos dados de uma tabela em um banco de dados

Após a conexão com a fonte dos dados e a execução da funcionalidade que é responsável por fazer a recuperação, os dados estão armazenados temporariamente dentro do *ETL* e deverão seguir o fluxo, seja de transformação ou de inserção em outro local. A partir deste ponto, o dado já pode ser considerado extraído e poderá seguir para a etapa de transformação de maneira estruturada.

Artefatos considerados como não estruturados não seguem exatamente a mesma linha, pois não são tratados como fluxo de dados, ou seja, como linhas de uma tabela. Logo, nesses casos é feita apenas uma busca no repositório fonte, verificando a existência desses arquivos nos diretórios de pastas, e os trazendo para o repositório temporário do *ETL*, onde eles irão permanecer até seguir para seu local de destino. Os dados não estruturados também não serão transformados como os estruturados, porém ainda podem ser duplicados, excluídos e compactados.

#### **4.2 DATA CLEANSING, DATA ENRICHMENT E TRANSFORMAÇÕES GENÉRICAS**

*Data Cleansing*, também conhecido como *Data Cleaning* ou *Scrubbing*, lida com a detecção e remoção de erros e inconsistências dos dados a fim de se melhorar a qualidade do dado (E. Rahm and H. Hai Do, 2000). Essa é a primeira parte da etapa de transformação dos dados e será responsável pela qualidade do dado divulgado no futuro. Durante a pesquisa do

autor e geração dos produtos, pode ser que partes do dado não sejam relevantes para outros interessados na pesquisa, e por ser algo que ocorre ou não, esta é uma etapa considerada opcional, dependendo da complexidade e estrutura do dado.

Na sequência do fluxo do processo, o dado recebido pela etapa de extração irá ser tratado em relação à limpeza do dado, conforme o conceito de *Data Cleansing* descrito. Após essa etapa, mas ainda na fase de transformação, será avaliado se os dados precisam ser enriquecidos. *Data Enrichment*, ou *Data Enhancement*, é o processo de aumentar o valor do conjunto de informações combinando dados de múltiplas fontes e seu objetivo é disponibilizar mais conhecimento de um conjunto de dados (David Loshin, 2001).

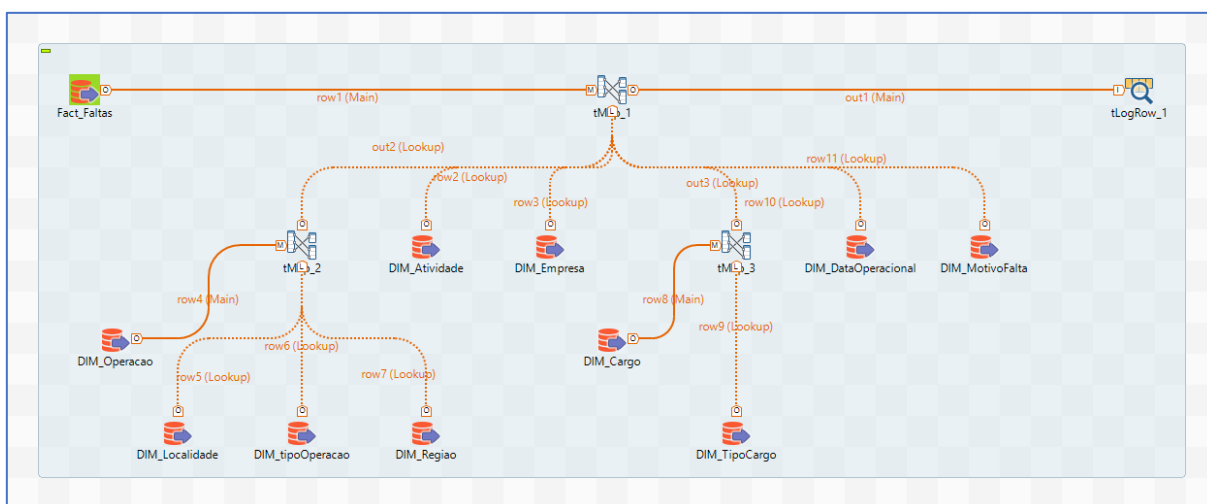


Imagem 2. Exemplo de *Data Enrichment*

Além do *Data Cleansing* e *Data Enrichment*, a etapa de transformação em dados estruturados lida com qualquer tipo de cálculo genérico que tenha por objetivo, modificar, adicionar ou excluir dados. Nesta fase, é possível criar condições e tratar linhas separadamente, utilizando funções já existentes na linguagem de programação da ferramenta *ETL*.

Após as transformações desejadas, o fluxo de dados estruturados segue para a próxima etapa, o carregamento. Caso seja útil, antes da fase de carregamento, é possível disponibilizar um produto em um repositório. Este produto pode conter todos os dados, parte dele ou simplesmente um log de execução entre outros. A ideia desta divulgação é visualizar os dados que estão sendo carregados e se estes estão corretos.



### 4.3 CARREGAMENTO DOS DADOS DE PESQUISAS CIENTÍFICAS

A etapa de carregamento de dados pode ser considerada como a última fase desta proposta abordada, divulgando os dados nas plataformas abertas de maneira que a informação esteja disponível para seu público alvo.

A forma com que os dados serão divulgados irá depender de alguns fatores importantes, como a arquitetura da plataforma que irá receber, guardar e mostrar os dados. Dependendo dessa estrutura, ambos os tipos de informação poderão seguir fluxos diferentes. O fluxo de processo para esta etapa consiste primeiramente em identificar a maneira com que a plataforma recebe os dados a serem enviados. Após identificar como o receptor irá coletar os dados, será necessário adaptar o *ETL* para que este consiga fazer a interface com a plataforma escolhida. O armazenamento será de preocupação da plataforma, tendo em vista que os dados já foram formatados para serem recebidos como predefinido pelo receptor. Será também responsabilidade da plataforma mostrar os dados, então, a maneira como esses dados serão divulgados abertamente independe do processo proposto.

Uma vez feito o carregamento, a plataforma ficará responsável pela próxima etapa do processo, o qual consiste em dar visibilidade dos dados para os usuários daquele serviço ou website.

A periodicidade e intervalos de tempo de carregamento deverão ser alinhados com o autor da informação e com a entidade responsável por disponibilizar os dados para extração. Além de também, verificar se a plataforma receptora dos dados tem algum tipo de janela de tempo predefinida para carregamentos, tendo em vista que inúmeros carregamentos possam ser feitos ao mesmo tempo, levando a uma baixa performance do serviço que recebe a informação.

A periodicidade pode não ser o único impeditivo para o carregamento, levando em consideração que a receptora dos dados pode definir permissões de carregamento, permitindo que apenas hosts cadastrados possam fazer carregamentos, alterações e remoções de informações em suas dependências.

Após todo o desenvolvimento dos *jobs* que farão as etapas descritas, uma etapa de documentação se inicia, procurando registrar a maneira com que os dados estão sendo extraídos, transformados e carregados na plataforma, para que seja possível e fácil realizar-se manutenções e alterações no processo caso necessário.

Por fim, caso seja de interesse do desenvolvedor e dos envolvidos, será possível agendar os *jobs ETL*, para que estes sejam executados automaticamente, sem a necessidade de intervenção humana, utilizando os conceitos de *RPA (Robotic Process Automation)*, termo que representa a execução de tarefas de serviço uma vez feito por pessoas, são feitos por automatizações, scripts e softwares (Mary Lacity, Leslie Willcocks e Andrew Craig, 2015). Os agendamentos das transformações podem ser feitos em orquestradores existentes no mercado, ou simplesmente em ferramentas disponibilizadas pelos sistemas operacionais, conhecidas como *Schedulers*. Com os devidos acessos, a ferramenta de agendamento poderá executar o arquivo construído pela ferramenta de *ETL* nos dias e horas definidas, gerando *logs* de execução.

## 5 ESTUDO DE CASO

Buscando fazer a divulgação de uma base de dados provenientes de uma pesquisa científica do instituto de computação da UFF, o processo descrito nessa proposta será mostrado a seguir. O software *ETL Talend* será utilizado para executar o processo.

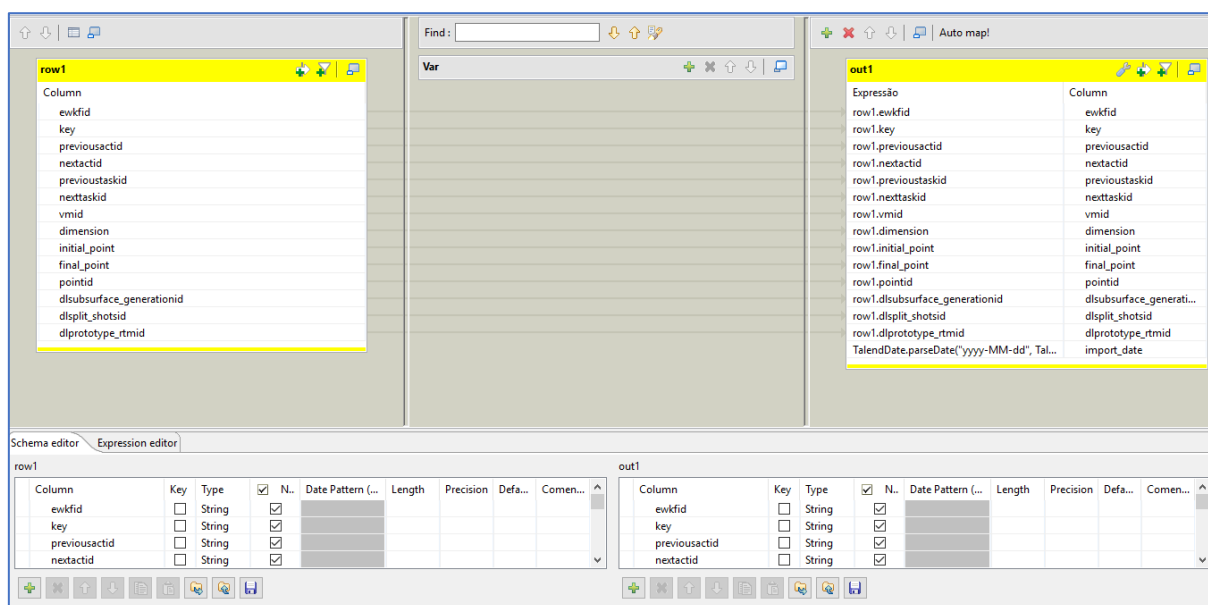
Primeiramente, cria-se o *job* que irá conter os passos que executarão as tarefas do processo. Agora, o passo de leitura de uma fonte de dados é criado, e para isso, é necessário definir a tecnologia ser utilizada, ter todos os dados de conexão, permissão de acesso e entendimento de que dados devem ser lidos.

The screenshot displays the Talend Job Designer interface for a job named "Job import\_from\_seismic". The main workspace shows a data flow from a "tPostgresInput\_1" component to a "tLogRow\_1" component. A status bar indicates "1000 rows in 1,26s" and "792,39 rows/s". Below the workspace, the "Execution" tab is active, showing a log of the job's execution. The log includes the start time "Starting job import\_from\_seismic at 17:08 27/11/2018." and a table of data extracted from the source database.

ewkfid	key	previousactid	nextactid	previousaskid	nexttaskid	vaid	dimension	initial_point	final_point	pointid	dlsurface_genera
345	2	371	372	7412	8407	1	8	1	1000	8	11
345	3	371	372	7418	8407	1	8	1	1000	14	23
345	4	371	372	7420	8407	1	8	1	1000	16	27
345	5	371	372	7414	8407	1	8	1	1000	10	15
345	6	371	372	7409	8407	1	8	1	1000	5	5
345	7	371	372	7415	8407	1	8	1	1000	11	17
345	8	371	372	7410	8407	1	8	1	1000	6	7
345	9	371	372	7411	8407	1	8	1	1000	7	9

Imagem 3. Extração da tabela do banco de dados fonte

Uma vez criado, os dados a serem lidos já estão dentro do software e irão seguir para os passos de tratamento, os quais podem variar dependendo da necessidade e do artefato. Neste caso, a divulgação é referente a dados em uma tabela de um banco de dados interno da faculdade. Logo, esses dados estruturados serão lidos com um passo de *Postgresql Input*, e o fluxo será seguido para cada linha da tabela. Neste caso, a etapa de transformação apenas adiciona um campo a tabela original, inserindo a data de importação dos dados.



Por fim, com os dados prontos, um passo de escrita de arquivo *CSV* é criado e será responsável por guardar todos os dados lidos e tratados, utilizando um separador específico para manter a estrutura de tabela. A partir deste momento, pode-se decidir se o *ETL* ficará responsável pela conexão com a plataforma e carregamento do artefato ou se isso será feito manualmente.

Após a criação de uma página na plataforma aberta em questão, conseguimos fazer o carregamento manual na página. Neste ponto, o processo se finaliza. A informação já está disponível para os usuários da plataforma.

## 6 CONCLUSÃO

A abordagem proposta procurou entender a atual necessidade e os desafios relacionados ao processo de divulgação de informação em outros ambientes, processo esse

que pode ser dividido em várias etapas, cada uma com suas respectivas dificuldades. Além do entendimento, procurou também mapear e executar o processo fim-a-fim de divulgação.

Todo o processo de extração de dados é vastamente utilizado atualmente em muitos processos manuais e automáticos, tanto em empresas, como em institutos ou em trabalhos individuais, ou seja, não é uma parte custosa, pois já existem tecnologias e estratégias das mais diversas formas para executar esta etapa. A etapa de transformação, quando os artefatos são simples, não traz grandes dificuldades pois também é vastamente utilizado. Porém, ao se deparar com um volume muito grande de dados e artefatos, surgirá a necessidade de se preocupar com os recursos envolvidos no processo, sendo necessário estudar quais estratégias são melhores para se utilizar, a fim de chegar-se ao produto desejado. Já a divulgação deste produto pode ser complexa dependendo de muitos fatores, como em quais formatos disponibilizar os artefatos, que tecnologia utilizar para fazer a interface com a plataforma, horários em que a interface será utilizada e verificação da informação, ou seja, garantir a integridade dos dados.

Por fim, conclui-se que o foco não recai apenas em como executar o processo, quais tecnologias e recursos utilizar, e sim nas pessoas envolvidas, autores, donos da informação, donos dos recursos, gerentes da informação e público alvo. A partir do momento em que o processo de divulgação é alinhado e entendido com todos os envolvidos, o fluxo de dados dificilmente sofrerá gargalos e impedimentos, tendo em vista a abundância de estratégias, tecnologias e recursos existentes. A aliança entre pessoas e tecnologias permitirá que a informação não fique mais escondida, mas sim, divulgada através de seus colaboradores, ajudando todos aqueles que tem a informação em questão como objeto de estudo.

## REFERÊNCIAS

LOSHIN, David. **Enterprise Knowledge Management**. Estados Unidos da América: Morgan Kaufmann, 2001

OLIVEIRA, Daniel. **Uma Abordagem De Apoio À Execução Paralela De Workflows Científicos Em Nuvens De Computadores**. Rio de Janeiro, 2012

BANSAL, Srividya K e KAGEMANN Sebastian. **Semantic Extract-Transform-Load framework for Big Data Integration**. Estados Unidos da América: IEEE, 2013

RAHM, E and DO, H. Hai, **Data Engineering**. Estados Unidos da América: IEEE Computer Society, 2000

BATIMARCHI, Susana. **A diferença entre dados estruturados e não estruturados**. São Paulo: Guia , 2015

CUNNINGHAM, Ward .**The Wiki Way Quick Collaboration on the Web**. Boston, Estados Unidos da América: Addison-Wesley Longman Publishing Co, 2006

PFEIFER Gabriel, **API: o que é e como funciona**. Porto Alegre, 2018

BLATTMANN, Úrsula e SILVA, Fabiano Couto Corrêa da. **Colaboração e Interação na Web 2.0 e Biblioteca 2.0**. Florianópolis: Revista ACB, 2007

LEUF, Bo e CUNNINGHAM, Ward. **The Wiki Way**. Boston, Massachusetts: Adison-Wesley Professional, 2001

DUTRA, Claudio e LOPES, Karen. **Dados Abertos: Uma Forma Inovadora de Transparência**. Brasília DF, 2013

LACITY, Mary, WILLCOCKS, Leslie e CRAIG, Andrew. **Robotic Process Automation at Telefónica O2**. Londres: The London School of Economics and Political Science, 2015