

Cronos: Uma aplicação para particionamento de base de dados

Cleyciane F. de Lima ¹, Juliana A. da Cunha ¹, Raul Vitor L. da Costa ¹

¹Curso de Bacharelado em Sistemas de Informação

Universidade Federal do Acre (UFAC) – Rio Branco, AC – Brazil

{fariascleycianedelima, juliana.1419981, raulawliet}@gmail.com

Abstract. *This article presents an application developed with Java, designed to help researchers / students reduce the time spent splitting their datasets for further analysis through an interface that loads, metrics, partitions and saves smaller files based on on user-defined criteria. Thus, the application allows partitioning of any database that has at least one attribute of the type “date” in order, reducing the preprocessing time that is usually large in this type of operation.*

Resumo. *Este artigo apresenta uma aplicação desenvolvida em Java, destinada a auxiliar pesquisadores/estudantes a reduzir o tempo gasto com a divisão de seus datasets. Por meio de uma interface, a ferramenta carrega os dados, define métricas e particiona a base em arquivos menores com base nos critérios temporais definidos pelo usuário. Assim, a aplicação permite o particionamento de qualquer base de dados que possua pelo menos um atributo do tipo data e que esteja ordenado, reduzindo o tempo de pré-processamento que é geralmente grande nesse tipo de operação.*

1. Introdução

A análise temporal de padrões sobre pull requests pode, potencialmente, estabelecer um novo método para avaliação de *datasets*. Isso permite compreender melhor tendências existentes nos projetos no decorrer do tempo, o que permite adoção de medidas que visem fortalecer padrões positivos ou enfraquecer os negativos (SOARES, 2017, p.96). O particionamento de bases de dados pode, em diversos casos, apresentar dados mais refinados para auxiliar na extração de conhecimentos (RIOS, 2018).

Nesse contexto o presente trabalho se propôs a desenvolver um software denominado Cronos, que tem como objetivo particionar uma base de dados em sub-bases correspondentes. O Cronos, ao ser carregado com uma base de dados, recebe como parâmetro o atributo separador, tamanho da janela temporal, o tamanho do salto¹ que será dado para o próximo intervalo (caracterizando a janela móvel) e salva em um diretório indicado pelo usuário as N novas bases de dados geradas conforme os parâmetros estabelecidos anteriormente.

¹ Tamanho do salto: Definir uma instância de tempo de dias, meses ou ano em que as janelas serão separadas

2. Tecnologias Utilizadas

Nesta seção serão apresentadas as tecnologias que foram utilizadas na composição da aplicação ou que fizeram parte de sua implementação.

2.1 Eclipse

O Eclipse² consiste em uma plataforma de desenvolvimento de software livre extensível, baseada em Java. É uma comunidade de código aberto, cujos projetos são concentrados na criação de uma plataforma de desenvolvimento aberta, composta de estruturas, ferramentas e tempos de execução extensíveis para desenvolvimento, implementação e gerenciamento de software por todo o ciclo de vida.

2.2 Git

O Git³ é um sistema de controle de versão distribuído e de código aberto, projetado para lidar com projetos pequenos e grandes, com velocidade e eficiência. É um sistema usado principalmente no desenvolvimento de software, mas que pode ser usado para registrar o histórico de edições de qualquer tipo de arquivo. Cada diretório de trabalho do Git é um repositório com histórico completo e habilidade total de acompanhamento das revisões, não dependente de acesso a uma rede ou a um servidor central.

2.3 Java Runtime Environment (JRE)

O software Java Runtime Environment (JRE)⁴, contém a máquina virtual Java (JVM – Java Virtual Machine), as bibliotecas de classes de tempo de execução e o ativador de aplicativos Java necessários para executar programas escritos na linguagem de programação Java. Não é um ambiente de desenvolvimento, como compiladores ou depuradores, para ferramentas de desenvolvimento, ele atua como interpretador, onde carrega o “bytecodes” da aplicação compilada em Java, interpreta esses códigos para a arquitetura do computador que está sendo usado e depois executa a sua aplicação.

2.4 Java Virtual Machine (JVM)

O Java Virtual Machine (JVM)⁵, consiste em um ambiente padrão compatível com Java que executa qualquer aplicativo Java puro. Ele suporta o formato binário padrão Java e as APIs Java padrão. A JVM é detalhada por uma especificação que descreve formalmente o que é necessário em uma implementação da JVM.

² <https://www.ibm.com/developerworks/br/library/os-eclipse-platform/index.html>

³ <https://git-scm.com/>

⁴ <https://www.oracle.com/technetwork/java/javase/jre-8-readme-2095710.html>

⁵ <https://docs.oracle.com/en/database/oracle/oracle-database/18/jjdev/Oracle-JVM-overview.html#GUID-391B265A-D656-4589-A4F5-C4F801340886>

3. Metodologia

A Figura 1 detalha o processo metodológico utilizado no desenvolvimento do Software Cronos.

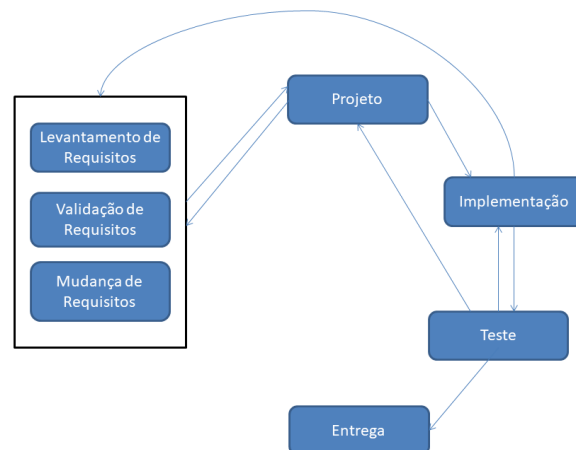


Figura 1. Fluxo Metodológico

O Processo metodológico utilizado para o desenvolvimento representado pela Figura 1, foi baseado nas metodologias já firmadas, que são: Scrum (Revisão de Sprints) e XP (Versões pequenas, Propriedade Coletiva, Padrões de Codificação).

Desta forma, o processo adotado iniciou-se com a definição dos requisitos. Com os processos de levantamento, validação e evolução dos requisitos, foi gerado o documento de requisitos, onde foram detalhadas todas as necessidades do cliente e especificações necessárias ao software. No fim da primeira etapa obteve-se uma ferramenta de particionamento com janela fixa, onde é possível definir o tamanho da janela, mas não o salto, que é 1 por padrão, conforme na Figura 2.

Na segunda etapa, foram realizados uma bateria de testes, que culminaram na melhoria posterior do sistema. Ao final da etapa de teste, foi realizada de fato a evolução do software, incluindo o requisito de janela móvel, onde pode-se definir o tamanho do intervalo e do salto, que na imagem é 4, mas pode ser de N, conforme na Figura 3.

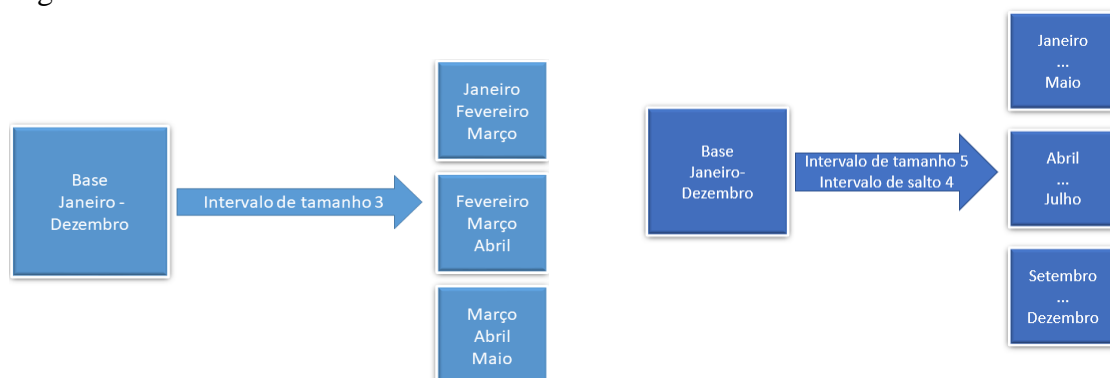


Figura 2. Janela Fixa

Figura 3. Janela Móvel

4. Cronos

O Cronos⁵ é uma ferramenta que tem como objetivo proporcionar o particionamento de bases de dados no formato Comma-separated-values (CSV) e que possuam pelo menos um atributo temporal do tipo data. Para o correto funcionamento da ferramenta, as datas que serão usadas como parâmetro devem estar padronizadas e ordenadas de maneira crescente. Na tela principal, é possível observar as métricas necessárias para o particionamento ao lado de uma explicação, conforme consta na Figura 4, o usuário pode realizar o particionamento de suas bases, definir o diretório final de salvamento e um padrão de nome para ser seguido entre os vários arquivos resultantes do processamento. Caso o atributo ou formato desejado não tenha sido corretamente especificado, o sistema informa quais campos estão incorretos e solicita que o usuário popule-os novamente da maneira correta, conforme é possível observar na Figura 5. Por fim, o sistema notifica e fornece um *Feedback* visual informando a conclusão do particionamento, conforme na Figura 6.



Figure 4. Tela Principal

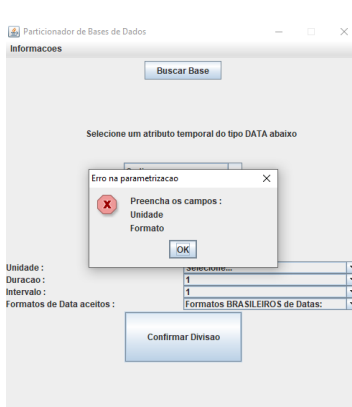


Figura 5. Informe de Erro

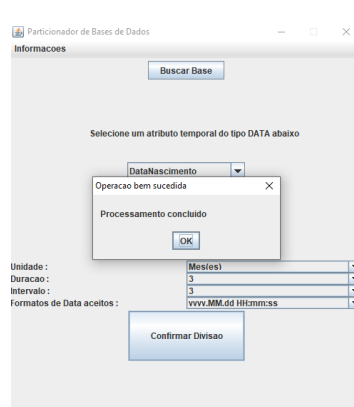


Figura 6. Sucesso

5. Considerações Finais

Espera-se que a ferramenta Cronos, auxilie pesquisadores e estudantes no gerenciamento do processo de particionamento de *datasets* que contem com atributos temporais do tipo data, possibilitando a criação de janelas temporais móveis, com intervalos definidos dinamicamente pelo usuário. O sistema inclusive está disponível para download no site [raulvitorl.github.io](https://github.com/raulvitorl/CRONOS) para toda comunidade acadêmica que porventura desejem realizar pesquisas com enfoque em particionamento de dados.

Referências

SOARES, Daricélio Moreira. (2017) “On The Nature Of Pull Requests: A Study About This Collaboration Paradigm Over Open-Source Projects Using Association Rules,” Rio de Janeiro: Universidade Federal Fluminense.

RIOS, Álvaro Lopes. (2018) “Análise temporal de pull requests usando regras de associação. ” Monografia (Graduação em Sistemas de Informação) – Universidade Federal do Acre, Acre.

⁵<https://github.com/raulvitorl/CRONOS>