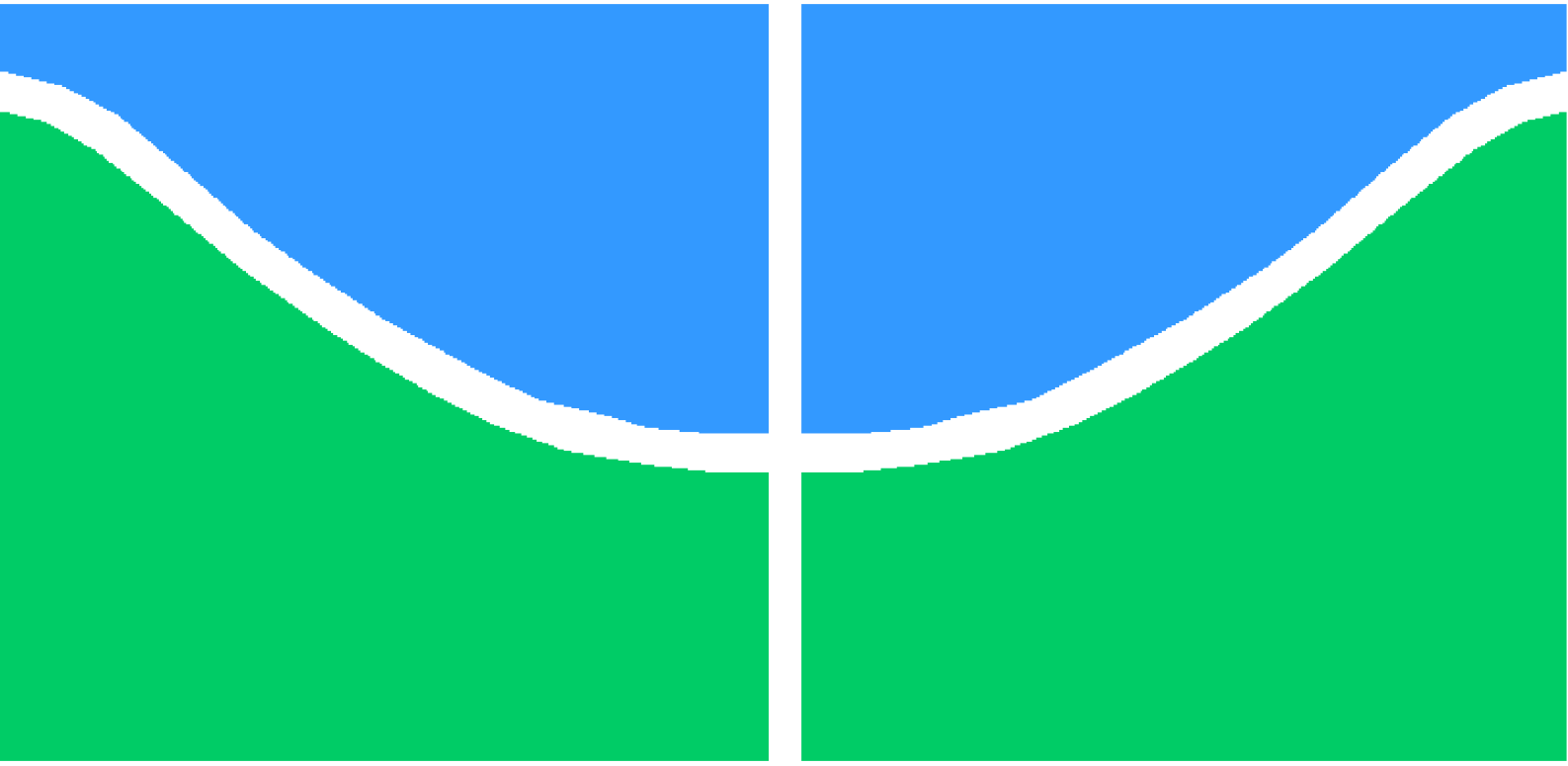

Universidade de Brasília – UnB
Faculdade UnB Gama – FGA
Engenharia de Software

Plataforma Web de Análise de Mercado de Jogos para Desenvolvedoras

Autor: João Paulo Busche da Cruz
Orientador: Prof. Matheus de Sousa Faria

Brasília, DF
2018

João Paulo Busche da Cruz

Plataforma Web de Análise de Mercado de Jogos para Desenvolvedoras

Monografia submetida ao curso de graduação em Engenharia de Software da Universidade de Brasília, como requisito parcial para obtenção do Título de Bacharel em Engenharia de Software.

Universidade de Brasília – UnB

Faculdade UnB Gama – FGA

Orientador: Prof. Matheus de Sousa Faria

Coorientador: Prof. Dr. Edson Alves da Costa Júnior

Brasília, DF

2018

João Paulo Busche da Cruz
Plataforma Web de Análise de Mercado de Jogos para Desenvolvedoras/ João Paulo Busche da Cruz. – Brasília, DF, 2018-
45 p. : il. (algumas color.) ; 30 cm.

Orientador: Prof. Matheus de Sousa Faria

Trabalho de Conclusão de Curso – Universidade de Brasília – UnB
Faculdade UnB Gama – FGA , 2018.

1. Mercado de Jogos. 2. Análise de Dados. I. Prof. Matheus de Sousa Faria. II. Universidade de Brasília. III. Faculdade UnB Gama. IV. Plataforma Web de Análise de Mercado de Jogos para Desenvolvedoras

CDU 02:141:005.6

João Paulo Busche da Cruz

Plataforma Web de Análise de Mercado de Jogos para Desenvolvedoras

Monografia submetida ao curso de graduação em Engenharia de Software da Universidade de Brasília, como requisito parcial para obtenção do Título de Bacharel em Engenharia de Software.

Prof. Matheus de Sousa Faria
Orientador

Prof. Dr. Edson Alves da Costa Júnior
Convidado 1

Prof. Dra. Carla Silva Rocha Aguiar
Convidado 2

Brasília, DF
2018

Agradecimentos

A Deus por me dar conhecimento e oportunidades para realizar meus sonhos, e por me dar força para superar as dificuldades.

Aos meus pais por sempre me apoiarem, idenpendente do quão insano fosse meu objetivo, e por sempre estarem ao meu lado nos momentos de dificuldades.

Ao meu orientador Matheus Faria, que sempre entendeu minha dificuldades e sempre se mostrou disponível para qualquer dificuldade que eu enfrentei no decorrer do trabalho.

Ao meu coorientador Edson Alves por suas correções e incentivos.

E a todos que contribuiram diretamente ou indiretamente para a minha formação, o meu muito obrigado.

Resumo

O mercado de jogos nos dias atuais é um dos mercados mais lucrativos e também um dos que mais cresce nos decorrer dos anos. Tendo um aumento considerável no número de desenvolvedoras *indies*, houve um aumento no número de desenvolvedoras que buscam por métricas de jogos, porém a extração de informações destas métricas é um trabalho que requer muitos recursos humanos/tempo. Por isso uma plataforma web que disponibilize estas métricas e *insights* poderia beneficiar o desenvolvimento de jogos. Para o desenvolvimento desta plataforma foram feitas as seguintes atividades: definição dos requisitos e desenvolvimento dos protótipos do software de extração e da plataforma web. O protótipo do software de extração é capaz de extrair os dados de suas fontes, manipulá-los e inseri-los no Elasticsearch.

Palavras-chaves: *game analytics*, *business intelligence*, métricas de jogos.

Abstract

The gaming market is one of most profitable and one with most growing over the years. Having a considerable increase in the number of indie developers, there has been an increase in the number of developers looking for game metrics, however extracting information from these metrics is a work that requires a lot of human/time resources. For that a web platform that provides metrics and insights, could benefit the development of games. or the development of this platform the following activities were done: definition of requirements and development of prototypes of extraction software and web platform. The extraction software prototype is capable of extracting data from its sources, manipulating it, and inserting it into ElasticSearch.

Key-words: game analytics, business intelligence, game metrics.

Lista de ilustrações

Lista de tabelas

Lista de Algorithms

Lista de abreviaturas e siglas

BI	*Business Intelligence*
GUR	*Game User Research*
API	*Application Programming Interface*
XML	*Extensible Markup Language*
HTTP	*Hypertext Transfer Protocol*
URI	*Uniform Resource Identifier*
XP	*Extreme Programming*
MVC	*Model-View-Controller*
REST	*Representational State Transfer*
TPS	Sistema Toyota de Produção
TCC	Trabalho de Conclusão de Curso

Sumário

1 Introdução

A indústria de jogos é uma das áreas de mercado mais lucrativa atualmente. No mercado brasileiro isto não seria diferente: sendo apenas o 13º maior mercado, no ano de 2017 movimentou 1,3 bilhões de dólares de acordo com o Newzoo (NEWZOO, 2017a), empresa que estuda o mercado de jogos. Um valor pequeno se comparado ao mercado chinês, o maior mercado de jogos do mundo, o qual movimentou 27,5 bilhões de dólares no ano de 2017 (NEWZOO, 2017b).

Sendo um mercado bastante lucrativo, houve um grande aumento no número de empresas desenvolvedoras de jogos. No mercado brasileiro, houve um aumento de 600 % entre 2008 e 2016 no número de desenvolvedoras (SILVEIRA, 2016). Este aumento não se limita apenas ao mercado brasileiro: de acordo com o site Steamspy (GALYONKIN, 2018), plataforma web que exibi dados de jogos, apenas no ano de 2017 foram lançados 7.672 jogos na plataforma Steam, plataforma onde desenvolvedoras disponibilizam seus jogos, uma média de 21 jogos por dia. Porém com este grande crescimento de concorrência fica cada vez mais difícil desenvolver um jogo que seja aceito pela comunidades de jogadores, assim, os jogadores tem muitas opções na hora comprar, o que faz que com algumas *features* se tornem diferencias, como multyplayer, localização, entre outros.

Levando em conta a dificuldade de se criar um jogo nos tempos atuais, muitas desenvolvedoras *indies* buscam por métricas que auxiliem na hora do desenvolvimento. Estas métricas, apesar de estarem disponíveis no mercado, necessitam um grande numero de recursos humanos/tempo para extrair informações que agreguem valor ao desenvolvimento. Desenvolvedoras *indies*, geralmente por falta deste tipo de recursos, muitas vezes pagam para outras empresas fazerem essas análises ou desenvolvem jogos sem as análises.

Por isso o objetivo deste trabalho e a criação de uma plataforma web onde será disponibilizada, de uma forma mais simples, diferentes métricas e agregações de dados, que permite o usuário moldar a necessidade dele. Incluindo *insights* sobre as métricas.

1.1 Objetivos

O objetivo deste trabalho è desenvolver um banco de dados inicial, com os dados de diversas fontes, para que futuramente possa ser criado as métricas de negócio.

Os objetivos espécificos são:

- elaborar uma arquitetura responsável por fazer a extração, manipulação dos dados dos jogos;

- elaborar uma rotina para a extração constantes dos dados;
- desenvolver um software responsável por extrair, manipular e inserir os dados dos jogos num banco de dados.

1.2 Estrutura do Documento

Este documento será dividido em 5 capítulos, sendo o primeiro a introdução. O Capítulo 2 é responsável pelo referencial teórico do projeto e análise de concorrentes. O Capítulo 3 é responsável pela metodologia, mostrando como será feito o projeto. O Capítulo 4 é responsável pelos resultados obtidos no projeto. O Capítulo 5 é responsável pela conclusão do projeto, onde também será mostrados possíveis trabalhos futuros.

2 Fundamentação Teórica

Neste capitulo será abordada a fundamentação teórica para o entendimento do propósito da implementação do projeto. Nele são explicados os conceitos de *application programming interface*, *business intelligence* e *game analytics* e também serão abordados softwares parecidos com o que será desenvolvido.

2.1 *Application Programming Interface*

A sigla API (*Application Programming Interface*) é um conjunto de rotinas e padrões de programação para acesso a um aplicativo de software ou plataforma baseado na Web. A utilização destas rotinas servem principalmente para que softwares externos tenham acesso a funcionalidade de um produto, sem que esse tenha que envolver-se em detalhes da implementação (RETHANS, 2016).

Para a utilização de APIs Webs existem várias arquiteturas que podem ser utilizadas a mais comum e a arquitetura REST (*Representational State Transfer*), que define um conjunto de restrições e propriedades baseadas no HTTP (*Hypertext Transfer Protocol*).

2.1.1 *Representational State Transfer*

A sigla REST, originalmente, se referia a um conjunto de princípios de arquitetura, nos tempos atuais é utilizada no sentido mais amplo, descrevendo qualquer interface web que use XML (*Extensible Markup Language*) e HTTP, sem as abstrações adicionais dos protocolos baseados em padrões de trocas de mensagem. De acordo com Fielding (FIELDING, 2000) é possível projetar um sistema de serviços web com a arquitetura REST, completando que a própria *World Wide Web*, utilizou REST como base para o desenvolvimento do protocolo HTTP.

A arquitetura REST possui quatro princípios principais, sendo eles:

- **Protocolo cliente/servidor sem estado**: cada mensagem HTTP contém toda a informação necessária para compreender o pedido.
- **Operações bem definidas**: que se aplicam a todos os recursos de informações, as mais importantes são ***POST***, ***GET***, ***PUT*** e ***DELETE***.
- **Sintaxe universal**: para identificar os recursos, na arquitetura REST este recurso é direcionado pela sua URI(*Uniform Resource Identifier*).

- **Hipermídia**: para a informação da aplicação, como para as transições de estado da aplicação, normalmente são representados pelo HTTP e XML

2.2 *Business Intelligence*

No decorrer dos anos houve uma grande mudança em relação a criação, coleta e do uso de dados. Enquanto houve uma grande evolução na maneira que esses dados eram gerenciados, sempre houve um desejo de extrair valores de negócios nas grandes pilhas de dados, que são capturadas hoje em diversas fontes, sejam elas estruturas de dados ou arquivos.

Os resultados da análise destas pilhas de dados e da criação de conhecimentos sobre esta análise, criou-se uma vantagem no mercado de negócios inimagináveis. Com esse conhecimento foi possível criar políticas de escolhas de ações para negócios em diferentes cenários. A exposição e exploração destes conhecimento é apenas uma das vantagens do uso de BI (*Business Intelligence*).

Uma das maiores vantagens do uso do BI está na otimização do processo de tomadas de decisão, pois para cada processo de negócios é associado a sua performance, e num mundo perfeito cada escolha deve ser a mais otimizada, ou seja, a que têm a melhor performance (LOSHIN, 2012).

A utilização de BI no âmbito de trabalho apresenta uma evolução na performance nas seguintes dimensões de negócio:

- No valor financeiro associado ao crescimento da lucratividade, sejam elas derivadas de custos ou do aumento de receitas.
- No valor de produtividade associado a diminuição da carga de trabalho, diminuição do tempo necessário para a execução de processos ponta-a-ponta e no aumento da porcentagem de produtos de alta qualidade.
- No valor de confiança, como maior satisfação do cliente, funcionário ou fornecedor, assim como aumento na confiança de previsões, consistência operacional e relatórios gerenciais, reduções no tempo gasto com "paralisia de análise" e melhor resultados de decisões.
- No valor de risco associado com a uma melhor visibilidade da exposição do crédito, confiança no investimentos em capitais e conformidade auditável com a jurisdição e normas e regulamentos da indústria.

The Data Warehousing Institute (ECKERSON, 2002), uma instituição especializada na educação e treinamento em armazenamento de dados, define que BI é os processos, as tec-

nologias e as ferramentas necessárias para transformar dados em informação, informação em conhecimento e conhecimento em planos que dirigem rentáveis planos de negócios, ou seja BI engloba armazenamento de dados, ferramentas de análiticas e gestao de informação/conhecimento.

Business intelligence geralmente considerar as informações dos dados como ativos, por isso, pode-se valer a pena examinar o uso de informações no contexto de como o valor é criado dentro de uma organização. Para isso existem três tipos diferentes de perspectivas, sendo elas:

- **Perspectiva funcional**: Neste tipo de perspectiva, os processos focam nas tarefas relacionadas a algum tipo particular de negócio, como vendas, *marketing*, entre outros. Processos funcionais confiam nos dados que operam dentros dos padrões de atividades comerciais.
- **Perspectiva interfuncional**: Como a maioria das empresas funcionam como um aglomerado de processos funcionais, e isto reflete em informações mais complexas. Para esta perspectiva a atividade foi um sucesso quando todas as tarefas foram completadas. Pela sua própria natureza, os processos envolvidos compartilham informações em diferentes funções, e o sucesso é medido tanto em termos de conclusão bem-sucedida, bem como as características do desempenho geral
- **Perspectiva empresarial**: Num ponto de vista organizacional e observando as características de desempenho dos processos interfuncionais, pode-se informar arquitetos empresariais e analistas de negócios maneiras na qual a organização pode mudar e melhorar o jeito que as coisas são feitas. Neste ponto de vista, o dado não é mais usado apenas para executar negócios, dados são utilizados para melhorar os negócios

2.3 *Game Analytics*

O desenvolvimento de jogos hoje pode se mostrar como um grande desafio, e grande parte deste desafio se dá pelo fato do grande número de jogos publicados. Para auxiliar as desenvolvedoras a criarem jogos eficientemente foram criados várias ferramentas e técnicas, um destes métodos é o *analytics*.

Analytics é o processo de descobrir e comunicar padrões em dados, solucionando problemas de negócios ou suportar decisões de gerenciamento de empresas. Está metodologia possui seus fundamentos em mineração de dados, na matemática, estatística, programação e operações de busca, como também na visualização dos dados, de forma a comunicar padrões relevantes. Vale mencionar que o *analytics* não é apenas perguntar e relatar dados de BI, e sim análises atual daqueles dados (DAVEPORT; HARRIS, 2007).

Game analytics é uma aplicação do *analytics* para o contexto de desenvolvimento de jogos (ANDERS; EL-NASR; CANOSSA, 2013). Um dos maiores benefícios em utilizar o *game analytics* é o suporte na hora de fazer decisões em todos os níveis e áreas organizacionais. Este método é direcionado tanto como a análise de um jogo com um produto, tanto como a análise de um jogo como projeto.

A aplicação padrão do *game analytics* é na hora de informar o GUR (*Game User Research*). GUR é a aplicação de várias técnicas e metodologias para avaliar a maneira na qual os jogadores jogam, e o nível de interação entre o jogador e o jogo. Vale mencionar que *game analytics* não é só GUR, já que o GUR é focado nos dados obtidos a partir dos usuários, já o *game analytics* considera todos os tipo de dados obtidos no desenvolvimento do jogo.

2.3.1 Telemetria

Telemetria são os dados obtidos à distância, geralmente digitais, porém qualquer dado transmitido à distância e telemetria. No contexto de jogos, telemetria seria algum jogo transmitindo dados sobre a interação do jogador com o jogo.

Telemetria de jogos é o termo utilizado para qualquer dado obtidos a distância que pertence durante o desenvolvimento ou evolução de um jogo, e isto inclui o monitoramento e análise de: servidores, dispositivos celulares e comportamento dos usuários. A fonte que produz mais dados por telemetrias, são os de usuário, por exemplo, interação com jogos, comportamento de compra e interações com outros jogadores ou aplicativos (BOHANNON, 2010).

2.3.2 *Game Metrics*

Em sua forma pura, os dados obtidos a partir da telemetria, não são de muito auxílio, por isso estes dados devem ser transformado em várias métricas interpretativas, como: o número de jogadores ativos por dia, bugs arrumados por semana, entre outros. Essas métricas são chamadas de game metrics. *Game metrics* possuem os mesmo potencial que outras fontes de BI. *Game metrics* geralmente são definidas como um medição quantitativa de um ou mais atributos, de um ou mais objetos que operem no contexto de um jogo.

Métricas podem ser variáveis ou agregações mais complexas, como a soma de várias variáveis, em outras palavras as métricas podem ser simples variáveis que geram uma análise básica, ou a combinação de várias variáveis para gerar uma análise mais complexa e completa. Métricas que não estão relacionadas diretamente ao jogo, são chamadas de métricas de negócios. Durante a utilização do *game analytics* é essencial a distinção entre as métricas de negócio e as *game metrics*.

As game metrics foram categorizadas em três tipos por Mellon (MELLON, 2009), ou seja, as *game metrics* podem ser definidas como:

- **Métricas de usuário**: São métricas relacionadas aos usuários que jogam aquele jogo, pela perspectiva de jogadores, ou de clientes. A perspectiva de cliente é utilizada quando as métricas são relacionadas a receita. A perspectiva de jogador é utilizada para investigar como é a interação das pessoas com o sistema do jogo e seus componentes.

- **Métricas de performance**: São métricas relacionadas a performance da tecnologia e arquitetura utilizada no jogo, muito relevantes para jogos onlines. Essas métricas geralmente são utilizadas no monitoramento dos impactos causado por alguma atualização no jogo.

- **Métricas de processo**: São métricas relacionadas ao processo de desenvolvimento de jogos. Similares as métricas de performance, são utilizadas para gerenciar e monitorar métodos que foram adotados, ou que foram adotados na hora do desenvolvimento do jogo.

2.4 Softwares Correlatos

Nesta parte de documento é levantado os softwares correlacionados, que possuem características ou objetivos parecidos com a plataforma a ser desenvolvido. Os principais são o Steam DB, uma ferramenta que disponibiliza informações sobre o banco de dados da Steam; a Steam Spy, uma plataforma web que disponibiliza informações sobre os jogos e suas vendas por região e o DFC Intelligence, uma ferramenta de pesquisa sobre o mercado de jogos.

2.4.1 Steam DB

Steam DB é uma ferramenta *open source third-party* com objetivo de dar um melhor conhecimento sobre os jogos e suas atualizações disponíveis no banco de dados da Steam (STEAMDB, 2010). Está ferramenta disponibiliza *rankings* e gráficos de jogos para que o usuário tenha uma melhor visualização destes. Atualmente o Steam DB apresenta métricas individuais de cada jogo, não oferencendo nenhuma maneira de agregação. Na tabela 1 podemos ver as vantagens e desvantagens do Steam DB.

Vantagens	**Desvantagens**
Open source	Não possui política de contribuição
Rankings e gráficos	Informações individuais de um jogo
Gratuito	Precisa de login na Steam
Informações individuais de um usuário	

Tabela 1 – Vantagens e Desvantagens Steam DB

2.4.2 Steam Spy

Steam Spy é uma plataforma Web que a partir de informações sobre os usuários da Steam, disponibiliza informações como o número de donos de algum jogo ou vendas por região de determinado jogo (GALYONKIN, 2018). Um dos objetivos especificados pelo Steam Spy é o auxílio a desenvolvedores *indies*, porém para se conseguir todos os dados disponíveis pela plataforma é preciso pagar por eles. Atualmente a Steam Spy para que o usuário possua total acesso aos dados dela, o usuário precisa pagar. Na tabela 2 podemos ver as vantagens e desvantagens do Steam Spy.

Vantagens	**Desvantagens**
Disponibiliza número de donos e vendas por região	Não é totalmente gratuito
Gráficos genéricos e confusos	Não apresenta métricas sobre os jogos
Disponibiliza API para seus dados	Não é *open source*

Tabela 2 – Vantagens e Desvantagens Steam Spy

2.4.3 DFC Intelligence

DFC Intelligence é uma ferramenta paga que auxiliam desenvolvedoras a tomar conhecimentos sobre estatísticas de seus jogos no mercado. Desenvolvedoras que utilizem essa ferramenta receberão informações sobre o número de vendas de seus jogos, picos de vendas, regiões que mais venderam, entre outras. Também será disponibilizado gráficos e métricas que informam como está o mercado de jogos (COLE, 1994). Atualmente a ferramenta DFC Intelligence é focada apenas nos jogos da desenvolvedora, que contratou seus serviços. Na tabela 3 podemos ver as vantagens e desvantagens do Steam DB.

Vantagens	**Desvantagens**
Disponibiliza informações sobre o mercado de jogos	Não é gratuita
Auxiliam desenvolvedores na hora da criação de um jogo	Não mostrar o mercado de jogos como um todo
	Não é *open source*
	Foca apenas nos jogos de uma desenvolvedora

Tabela 3 – Vantagens e Desvantagens DFC Intelligence

3 Metodologia

A metodologia escolhida para o desenvolvimento da plataforma será a metodologia ágil, a metodologia principal que vai será utilizada é o Kanban, porém, não será utilizada sua forma pura, e sim com algumas modificações, que atendem as necessidades do projeto. A escolha dessa metodologia se dá pelas seguintes características: o projeto será desenvolvido de maneira incremental, ou seja, ele poderá ser modificado no decorrer da implementação; a equipe consiste em apenas uma pessoa, o que descarta a possibilidade de utilizar o Scrum ou XP (*Extreme Programming*); o escopo do projeto será divididos em tarefas, e a utilização do Kanban facilita no descobrimento de gargalos.

A metodologia Kanban surgiu no Japão com o TPS (Sistema Toyota de Produção) (OHNO, 1997) para controlar a fabricação de automóveis e foi inserida no meio de desenvolvimento de software no ano de 2007. Kanban é um termo japonês para sinal visual e uma das grandes características dessa metodologia é evidenciar os problemas existentes no processo.

A metodologia ágil surgiu no ano de 2001, com a reunião de especialistas em processos de desenvolvimento de software para discutir maneiras de melhorar o desempenho em projetos, com isso foi criado o Manifesto Ágil (BECK et al., 2001). Uma das características das metodologias ágeis são sua capacidade de adaptar a novos fatores durante o desenvolvimento do projeto, ao invés de tentar prever o que pode acontecer e o que não pode

3.1 Análise de Ferramentas

Nesta seção são feito estudos sobre as ferramentas que serão utilizadas no desenvolvimento do projeto.

3.1.1 Banco de Dados

A ferramenta de banco de dados é responsável pelo armazenamento dos dados extraídos das fontes de dados.

Elasticsearch

Elasticsearch é uma ferramenta *open source*, desenvolvida pela Elastic[1], de análise e busca REST capaz de resolver um grande número de casos. É a parte principal de Elastic Stack, servindo como um centro de armazenamento de dados (ELASTIC, 2010a).

[1] <https://www.elastic.co/>

Elasticsearch suporta qualquer tipo de dado, além de agregar grande quantidades de dados para se ter uma visão melhor. Entre suas características as que mais se destacam são sua rapidez de busca, capacidade de detecção de falhas, múltiplos tipos de dados e suporte a múltiplas linguagens de programação.

3.1.2 Frequência na Extração dos Dados

Esta ferramente será responsável pela criação de uma rotina na hora de extrair os dados e atualizar o banco de dados.

Celery

Celery é uma ferramenta *open source* focada em operações em tempo real numa fila de tarefas assincronas baseadas na passagem de mensagens distribuidas, tambem oferece suporte a operações de agendamento (CELERY, 2007).

Celery possui funções que auxiliam a criação de rotinas, funcionando principalmente com Python[2]. Como Celery trabalha com a utilização de tarefas, podemos agendar seus funcionamentos utilizando *cronjobs*[3] para suas frequências.

3.1.3 Visualização dos Dados

Esta ferramenta será responsável por demonstrar os dados do banco de uma maneira mais intuitiva.

Kibana

Kibana é um ferramenta *open source*, desenvolvida pela Elastic, que permite a visualização dos dados guardados no Elasticsearch, possuindo diversas maneiras diferentes de disponibilização visual dos dados. É a parte de visualização do Elastic Stack, servindo como um centro de monitoramento (ELASTIC, 2010b).

3.1.4 Criação de Containers

Esta ferramente é responsável por dividir as partes da arquitetura em diferentes containers.

Docker

Docker é uma ferramenta *open source* que proveem containers de *softwares* para o auxilios de desenvolvedores. Outras funcionalidades que são disponibilizadas são a integra-

[2] <https://www.python.org/>
[3] <https://cron-job.org/en/>

ção com o desenvolvimento, suporte a multiplataformas e acesso a uma extensa biblioteca de servidores (DOCKER, 2014).

3.2 Arquitetura do Projeto

A arquitetura do projeto é dividida em quatro partes: as fontes de dados, o extractor, o banco de indexação e a API RESTful. A ligação entre as partes e sua posição na arquitetura ficam evidenciado na figura 1.

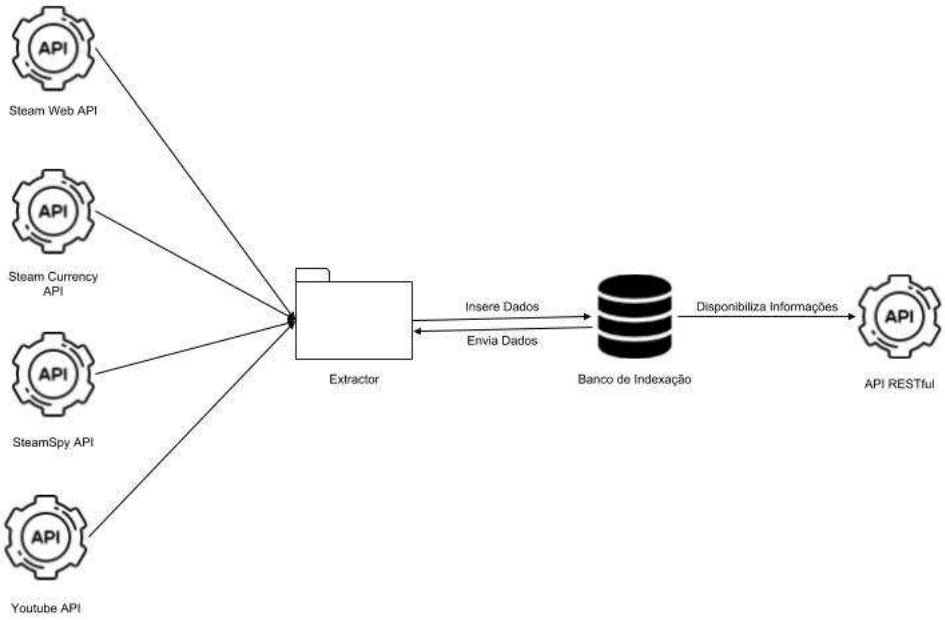

Figura 1 – Arquitura Geral do Projeto

Fonte de Dados são locais onde se encontrar os dados que serão extraidos, manipulados e inseridos no banco de indexação, estes dados poderão vir de quaisquer fonte, sejam elas um banco de dados, de *crawlers*, de APIs ou de arquivos locais.

Banco de Indexação é um banco de dados responsável por guardar os dados manipulados pelo Extractor. Como será utilizado o banco de jogos da Steam, o banco de indexação precisa suportar um grande número de dados e rapidez na atualização dos dados e na suas buscas.

API RESTful é a parte responsável por disponibilizar os dados do banco de indexaçõa, sem que o usuário precisa conectar diretamente a ele. Como será desenvolvida um API, outros e quaisquers projetos poderão usufruir dela.

3.2.1 Arquitetura Extractor

A arquitetura do Extractor e composta de três classes principais e de dois arquivos. Esta arquitetura está representada na figura 2.

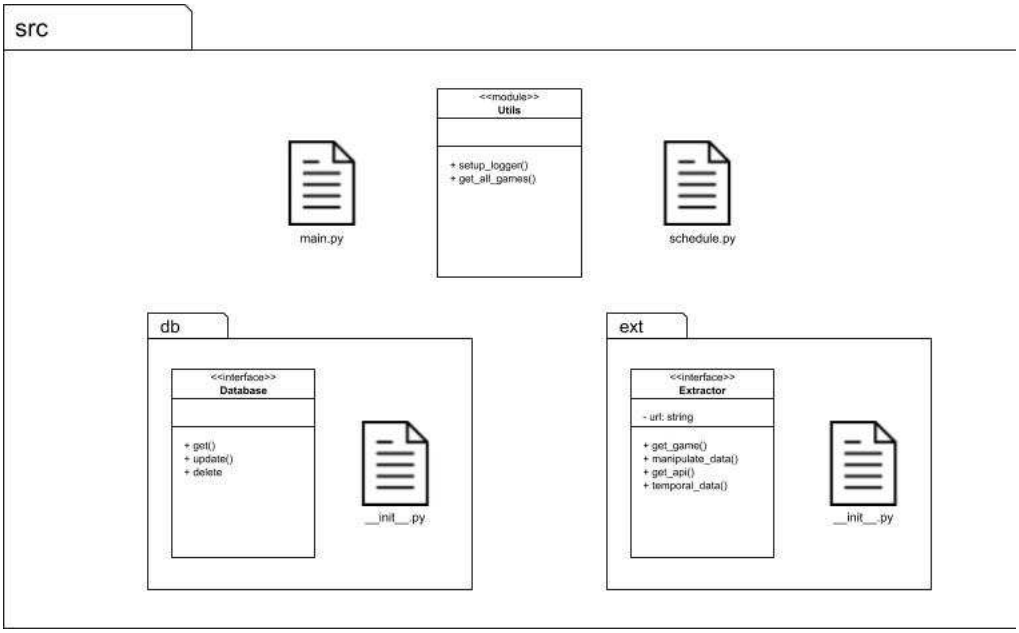

Figura 2 – Arquitura do Extractor

A interface ***Extractor*** deverá ser herdada por todos os plugins. Esta interface possui funções responsáveis por fazer requisições na API, extrair dados desta API e manipular estes dados, sejam eles estáticos ou temporais.

A interface ***Database*** deverá ser herdada por todos os bancos de dados que possam a ser utilizado. Esta interface possui funções responsáveis por inserir/atualizar, delete e mostrar os dados guardados pelo banco.

O módulo ***Utils*** é responsável por implementar funções que não se encaixam em nenhuma classe do software. A princípio este módulo possui funções responsaveis por configurar os logs e extrair todos os jogos que possuem.

Os arquivos ***Main*** e ***Schedule*** são resposáveis por, respectivamente, fazer a primeira inserção no banco de dados e por manter um rotina de atualizações nos dados dos jogos.

3.3 Extractor

O Extractor é responsável pela extração, manipulação das informações das fontes de dados e também por inserir e atualizar o banco de dados. Como o Extractor lidará

com dados que se modificam com o tempo, será preciso criar rotinas que lidarão com esses dados temporais. No escopo inicial serão utilizados quatro APIs.

A API ***Steam WEB API*** é fornecida pela Steamworks[4], sendo assim é uma API oficial da Steam (STEAMWORKS, 2018). Ela possui métodos públicos e métodos privados. Os métodos públicos são abertos para qualquer usuário utiliza-los, já os métodos privados é necessário uma chave de desenvolvedor cedido pela própria Steamworks. Para acessar quaisquer dados da API é necessário um id, seja de usuário ou de jogo.

A API ***Steam Store API*** é fornecida pela Steam. Ela disponibiliza informações sobre os jogos guardados no banco de dados da Steam. Para acessar estes dados é necessário o id do jogo na Steam, porém também é possível passar filtros para pesquisas mais específicas.

A API ***Steam Spy API*** é fornecida pelo Steam Spy. Ela também disponibiliza informações sobre os jogos, porém ela também dispõe de outras informações como o número de donos ou número de avaliações positivas e negaticas de um determinado jogo. Para acessar estes dados é necessário o id do jogo na Steam, porém também é possível passar filtros para pesquisas mais específicas

A API ***Youtube API*** é fornecida pelo Google Developers[5], sendo assim a API oficial do Youtube. Ela dispõe de informações sobre cada vídeo como seus números de likes, dislikes e visualizações. Para extrair os dados de cada vídeo são necessário conseguir a lista dos 20 vídeos mais relevantes de um jogo, ápos isso com o id dos vídeos extrair suas informações.

A relação entre os dados que serão extraídos e de qual API será utilizado esta representado na tabela 4.

[4] <https://partner.steamgames.com>
[5] <https://developers.google.com/>

	Steam WEB API	***Steam Spy API***	***Steam Store API***	***Youtube API***
Nome				
Descrição				
Imagem *Header*				
Imagem *Background*				
Website				
Data Lançamento				
Steam Id				
Metacritic *Score*				
Avaliação Positiva				
Avaliação Negatica				
Média de Horas				
Donos				
Jogadores Online				
Visualizações				
Likes				
Dislikes				
Userscore				
Gêneros				
Categorias				
Linguagens				
Screenshots				
Desenvolvedoras				
Publicadoras				
Plataformas				
Preço				
Legendas				
Dados Estáticos				
Dados Temporais				

Tabela 4 – Dados x APIs

3.3.1 Dados Temporais

Dados temporais são aqueles dados que serão atualizados de acordo com alguma frequência. Como serão guardados seus valores no decorrer do tempo, será possível comporar o mesmo dado de um jogo num determinado espaço de tempo, podendo assim extrair mais informações e ocasionalmente mais métricas. A frequência que estes dados serão atualizado estão representado na tabela 5.

Tipo do Dado	**Frequência**
Média de Horas	1 vez por semana
Donos	1 vez por semana
Jogadores Online	1 vez por dia
Visualizações	1 vez por dia
Likes	1 vez por dia
Dislikes	1 vez por dia
Userscore	1 vez por semana
Preço	1 vez por semana

Tabela 5 – Dados x Frequência

Outro função que será chamanda numa determinada frequência será a inserção de novos jogos no banco de dados, pois novos jogos são lançandos a cada dia. Para não causar uma sobrecarga no Celery, a inserção de novos jogos ocorrerão uma vez por semana.

Referências

ANDERS, D.; EL-NASR, M. S.; CANOSSA, A. Game analytics: Maximizing the value of player data. Londres, 2013. Citado na página 26.

BECK, K. et al. *Manifesto Ágil*. 2001. Disponível em: <http://agilemanifesto.org/iso/ptbr/manifesto.html>. Citado na página 31.

BOHANNON, J. Game-miners grapple with massive data. 2010. Citado na página 26.

CELERY. *Celery*. 2007. Disponível em: <http://www.celeryproject.org/>. Citado na página 32.

COLE, D. *DFC Intelligence*. 1994. Disponível em: <https://www.dfcint.com>. Citado na página 28.

DAVEPORT, T. H.; HARRIS, J. G. Competing on analytics: The new science of winning. Boston, 2007. Citado na página 25.

DOCKER. *Docker*. 2014. Disponível em: <https://www.docker.com/>. Citado na página 33.

ECKERSON, W. The rise of analytic applications: Build or buy. 2002. Citado na página 24.

ELASTIC. *Elasticsearch*. 2010. Disponível em: <https://www.elastic.co/products/elasticsearch>. Citado na página 31.

ELASTIC. *Kibana*. 2010. Disponível em: <https://www.elastic.co/products/kibana>. Citado na página 32.

FIELDING, R. T. Architectural styles and the design of network-based software architectures. 2000. Citado na página 23.

GALYONKIN, S. *Steam Spy*. 2018. Disponível em: <http://steamspy.com>. Citado 2 vezes nas páginas 21 e 28.

LOSHIN, D. Business intelligence: The savvy manager's guide. 2012. Citado na página 24.

MELLON, L. Apply metrics driven development to mmo costs and risks. República Tcheca, 2009. Citado na página 27.

NEWZOO. *The Brasilian Gamer 2017*. 2017. Disponível em: <https://newzoo.com/insights/infographics/the-brazilian-gamer-2017>. Citado na página 21.

NEWZOO. *The Chinese Gamer 2017*. 2017. Disponível em: <https://newzoo.com/insights/infographics/chinese-gamer-2017>. Citado na página 21.

OHNO, T. O sistema toyota de produção. Porto Alegre, 1997. Citado na página 31.

RETHANS, J. *APIs: Leverage For Digital Transformation*. 2016. Disponível em: <https://www.forbes.com/sites/forbestechcouncil/2017/05/08/apis-leverage-for-digital-transformation/#70b07fa17140>. Citado na página 23.

SILVEIRA, D. *Crescimento Desenvolvedoras*. 2016. Disponível em: <https://g1.globo.com/economia/negocios/noticia/numero-de-desenvolvedores-de-games-cresce-600-em-8-anos-diz-associacao.ghtml>. Citado na página 21.

STEAMDB. *SteamDB*. 2010. Disponível em: <https://steamdb.info>. Citado na página 27.

STEAMWORKS. *Steam Web API*. 2018. Disponível em: <https://partner.steamgames.com/doc/webapi_overview>. Citado na página 35.

Anexos

ANEXO A – Código dos *plugins* e do Elasticsearch

```
1 from ext.extractor import Extractor, GameNotFound
2
3 class SteamAPI(Extractor):
4
5     url = 'https://store.steampowered.com/api/appdetails/?
    appids={}'
6
7     def get_game(self, identifier):
8         response = self.get_api(identifier)
9         data = response[str(identifier)]
10        if data['success'] and data is not None:
11            result = self.manipulate_data(data['data'],
    identifier)
12            return result
13        else:
14            raise GameNotFound("Game not found!!!")
15
16    def manipulate_data(self, data, identifier):
17        result = {}
18        result['genres'] = []
19        result['publishers'] = []
20        result['developers'] = []
21        result['platforms'] = []
22        result['screenshots'] = []
23        result['price'] = []
24        result['name'] = data['name']
25        result['steam_id'] = data['steam_appid']
26        if data['is_free'] or not 'price_overview' in data:
27            result['price'] = self.temporal_data(identifier,
    0.0, 'price')
28            result['is_free'] = True
29        else:
```

```
30              result['price'] = self.temporal_data(identifier,
    data['price_overview']['initial'] / 100, 'price')
31              result['is_free'] = False
32          result['description'] = data['about_the_game']
33          result['header_image'] = data['header_image']
34          result['background_image'] = data['background']
35          if data['website'] is not None:
36              result['website'] = data['website']
37          else:
38              result['website'] = ""
39          if 'genres' in data:
40              for gnr in data['genres']:
41                  result['genres'].append(gnr['description'])
42          for pbl in data['publishers']:
43              result['publishers'].append(pbl)
44          if 'developers' in data:
45              for dvp in data['developers']:
46                  result['developers'].append(dvp)
47          for plt in data['platforms']:
48              if plt: result['platforms'].append(str(plt).
    capitalize())
49          if data['release_date']['date'] != "":
50              result['release_date'] = data['release_date']['
    date']
51          else:
52              result['release_date'] = '12 Set, 2003'
53          if 'metacritic' in data:
54              result['metacritic_score'] = data['metacritic']['
    score']
55          else:
56              result['metacritic_score'] = 0
57          for screen in data['screenshots']:
58              result['screenshots'].append(screen['path_full']
```

Algorithm A.1 – Código do *plugin* ***Steam Store API***

```
59 import requests
60 import json
61 from elasticsearch import Elasticsearch
62 from db.database import Database, DataNotFound
```

```
63
64 class Elastic(Database):
65
66     def __init__(self, host, index):
67         self.hostname = host
68         self.index_name = index
69         try:
70             self.elastic = self.connect_elasticsearch()
71         except ElasticNotConnected as enc:
72             raise ElasticNotConnected('Elasticsearch couldn\'t
    be connected!!!')
73
74     def get(self, identifier):
75         res = self.elastic.search(index=self.index_name, body
    ={"query": {"match": {"steam_id": int(identifier)}}})
76         if res['hits']['total'] == 1:
77             data = res['hits']['hits'][0]['_source']
78             return data
79         else:
80             raise DataNotFound('Data has not found in the
    Database!!!')
81
82     def update(self, identifier, data):
83         if self.has_data(identifier):
84             body = {}
85             body['doc'] = data
86             self.elastic.update(index=self.index_name,
    doc_type='game', id=identifier, body=body)
87         else:
88             self.elastic.index(index=self.index_name, doc_type
    ='game', id=identifier, body=data)
89
90     def delete(self, identifier):
91         url = 'http://{}/{}/game/{}'
92         response = requests.delete(url.format(self.hostname,
    self.index_name, identifier))
93         if response.status_code != requests.codes.ok:
94             raise DataNotFound('Data has not found in the
    Database!!!')
```

```
95
96      def has_data(self, identifier):
97          url = 'http://{}/{}/game/{}'
98          response = requests.get(url.format(self.hostname, self
        .index_name, identifier))
99          if response.status_code == requests.codes.ok:
100             return True
101         else:
102             return False
103
104     def connect_elasticsearch(self):
105         elastic = Elasticsearch([self.hostname])
106         if elastic.ping():
107             return elastic
108         else:
109             raise ElasticNotConnected('Elasticsearch couldn\'t
        be connected!!!')
110
111     def delete_index(self):
112         self.elastic.indices.delete(index=self.index_name,
        ignore=[400, 404])
113
114     def get_all(self):
115         res = self.elastic.search(index=self.index_name, body
        ={"query": {"match_all":{}}})
116         size = res['hits']['total']
117         res = self.elastic.search(index=self.index_name, body
        ={"size": size, "query": {"match_all":{}}})
118         games = []
119         for game in res['hits']['hits']:
120             pair = (int(game['_id']), game['_source']['name'])
121             games.append(pair)
122         return games
123
124     def create_index(self, index_body):
125         if self.elastic.indices.exists(self.index_name): self.
        delete_index()
126         self.elastic.indices.create(index=self.index_name,
        body=index_body, ignore=400)
```

```
127
128
129 class ElasticNotConnected(Exception):
130     pass
```

Algorithm A.2 – Código do ***Elasticsearch***