



ZOOMBOX: ECOSISTEMA DE INTELIGÊNCIA DE NEGÓCIOS

ZOOMBOX: BUSINESS INTELLIGENCE ECOSYSTEM

Gabriel Covello Furlanetto ¹

Luciano Sasso Vieira ²

Evandro Filie Alampi ³

Fernando Gibotti ⁴

RESUMO: Neste artigo é apresentado o ZoOmBOX, um ecossistema de captura e processamento de dados massivos referentes ao comportamento de consumo nos mais variados tipos de segmentos, como farmácias, supermercados, hipermercados, redes varejistas e atacadistas. Seus módulos permitem a interação entre o consumidor e o varejista, gerando dados que quando enriquecidos por meio de fontes secundárias e análises geoespaciais possibilitam uma análise mais precisa do grau de captação do potencial de consumo, das áreas de influência das lojas e dos raios de atuação. Tal tarefa torna-se desafiadora devido ao volume de dados captados e à variedade de bases de fontes secundárias. Porém, as barreiras são superadas a partir de uma plataforma veloz e intuitiva que atua sobre uma infraestrutura robusta para o processamento de dados.

Palavras-chave: Inteligência de negócios. Big Data. Varejo.

ABSTRACT: *In this article we present ZoOmBOX, an ecosystem for capturing and processing massive data referring to consumption behavior in the most varied types of segments, such as pharmacies, supermarkets, hypermarkets, retail chains and wholesalers. Its modules allow the interaction between the consumer and the retailer, generating data*

¹ Mestre em Ciência da Computação, UNESP,

² Bacharel em Sistemas de Informação, Grupo FAIMI, 2004.

³ Mestre em Geografia, UNESP,

⁴ Doutor em Computação Aplicada, INPE, 2006.

that when enriched by means of secondary sources and geospatial analyzes allow a more precise analysis of the degree of capture of the consumption potential, the areas of influence of the stores and the radius of actuation . Such a task becomes challenging because of the volume of data collected and the variety of databases from secondary sources. However, the barriers are overcome by a fast and intuitive platform that acts on a robust infrastructure for data processing.

Keywords: *Business intelligence. Big data. Retail.*

1 Introdução

A forte competição por preços entre as empresas do setor varejista e a diminuição das margens de lucro provocou, nas últimas décadas, a busca constante pela diferenciação no setor. Além das formas de atendimento, da contenção de gastos e da inovação nas plataformas de comercialização de produtos, algumas frentes vêm se destacando permitindo um aprimoramento baseado na experiência e no comportamento de consumo do cliente.

Mediante a necessidade de o varejista compreender quem são os seus clientes, o que consomem e com qual frequência, a captura de dados de vendas acaba por se tornar uma ferramenta de forte relevância. No entanto, para o armazenamento e controle de tais dados não bastam apenas uma caderneta ou uma planilha, uma vez que o volume se tornou grandioso demais. Figura, então, o conceito de Big Data descrito pelo modelo dos 3Vs (Volume da massa de dados, Variedade dos dados englobados, Velocidade da análise a ser feita) (McAfee et al, 2012) utilizado para agrupar conjuntos de dados massivos estruturados, ou não, que não podem ser processados por computadores comuns (CHEN; MAO; LIU, 2014).

Deste modo, surge também o valor dos dados massivos que, quando processados em tempo real, possibilitam análise de negócios (Business Intelligence) e tomada de decisões, estimulando o consumo e aumentando a lucratividade no varejo (GOLE; TIDKE, 2015). Podem ser observados inúmeros exemplos de sucesso, como a Macy's que consegue estabelecer preços em tempo real para os milhões de itens em suas lojas elaborando

promoções a partir do comportamento da demanda e do estoque; e o Walmart, um dos pioneiros na análise de grandes massas de dados, que tem melhorado a finalização de compras on-line gerando faturamento de milhões de dólares (MAÇADA; BRINKHUES; JÚNIOR, 2015).

Até então, as ferramentas estavam muito voltadas para a automação e para as melhorias nos processos internos do varejista, desconsiderando o principal agente desta interação: o consumidor. Nesse contexto, o ZoOmBOX foi concebido para ser o ecossistema que engloba o varejista e o cliente a fim de gerar benefícios entre ambas as partes: o cliente recebe benefícios e incentivos (participação em sorteios, acúmulo de pontos ou parte do seu consumo em descontos e vale compras) e, em contrapartida, fornece seus dados cadastrais ao varejista. Com isso, o varejista tem uma informação valiosa em mãos; os dados cadastrais desse cliente aliados às ferramentas de processamento de dados, modelos matemáticos, informações secundárias e análises geográficas geram conteúdos que apoiam novas estratégias para o negócio.

A participação do cliente, associada às informações mais precisas, com o apoio de ferramentas apropriadas, permite ao varejista captar o comportamento do cliente e relacionar-se com ele simultaneamente de forma personalizada, gerando conhecimento profundo do seu hábito de consumo e resultando em melhores desempenhos e vantagens competitivas em relação aos concorrentes.

2 Contextualização temporal e desafios de desenvolvimento

Como observado, os dados possuem extrema importância, sendo considerados de grande valia para obtenção de recursos financeiros, assim como foi o petróleo em décadas atrás. Neste meio, sem dar-se conta, as pessoas são as principais fontes geradoras de dados, os quais são fornecidos em troca da utilização de diversos softwares e aplicativos.

As mídias sociais, como Twitter, Facebook, Instagram, entre outras, aparecem como exemplo explícito. Para utilizá-las não é necessário realizar compensações financeiras, porém a imensa massa de informações gerada durante a navegação é armazenada pelas plataformas, permitindo que seja obtido conhecimento sobre atividades, gostos e consumo pessoal dos usuários (BATRINCA; TRELEAVEN, 2015).

A partir do processamento e análise das informações coletadas diariamente, os dados obtidos são transformados em lucro financeiro, uma vez que são utilizados em ofertas de produtos, serviços e até mesmo na detecção de tendências e comportamentos. No entanto, devido ao volume e heterogeneidade desses dados, existe um desafio para tratá-los em tempo hábil sendo necessárias técnicas e ferramentas de Big Data para seu armazenamento, acesso e processamento e de algoritmos de Machine Learning para sua análise, com a extração de comportamentos.

Além disso, com a exigência dos consumidores e o aumento da competitividade em todos os mercados, torna-se necessária a inovação como premissa básica nos negócios. Deste modo, pode-se obter informações de compras e vendas por meio de sistemas internos das empresas e cruzá-las com as dos próprios consumidores obtidas pelas mídias sociais, por exemplo sua opinião sobre produtos e até mesmo sobre o fornecedor do mesmo, o próprio varejista. Sendo assim, a velocidade para capturar, tratar e analisar a informação é parte fundamental do sucesso que o Big Data pode proporcionar às empresas.

O ecossistema ZoOmBOX possui como premissa atingir êxito em todos esses desafios apresentados. Para que isso ocorra parte de um software para realizar promoção, relacionamento e fidelidade de clientes responsável por coletar grande volume de dados nos varejistas dos mais diversos ramos comerciais, como alimentício, farmacêutico, industrial, em tempo hábil. Dados que passarão por etapas de modelagem, limpeza e enriquecimento de informações com a agregação de informações secundárias e geográficas.

Tais etapas são necessárias para estudos futuros com maior nível de detalhamento, voltados aos serviços de inteligência de negócios, análises de comportamento de consumo, potencial de mercado e satisfação de clientes. Serviços que atualmente são oferecidos ao setor varejista somente a partir de dados secundários e coletas de campo, não levando necessariamente em consideração o real comportamento do consumidor. O impacto dessas análises transformam o negócio, alterando formatos de lojas, níveis de atendimento, público-alvo, mix de produtos, resultando no aumento do faturamento e na diminuição de perdas.

2.1 Características do ecossistema ZoOmBOX

O ecossistema ZoOmBOX tem como principal objetivo capturar e processar dados off-line de grandes redes de farmácias, supermercados, hipermercados varejistas e atacadistas, o varejo físico, identificando o cliente que realizou cada compra, ou seja, apontando cada cupom fiscal com um identificador do cliente, o CPF. Isso permite capturar o carrinho de compra de cada cliente (produtos comprados contidos no cupom fiscal), possibilitando rastrear todos seus gastos na loja e seu hábito de consumo de forma bastante profunda.

Como incentivo para identificação, o motor de regras do ZoOmBOX gera benefícios que podem ser cupons promocionais, pontos em programa de fidelidade próprio, entre outros.

Todos os dados capturados são tratados e enriquecidos na plataforma com mais de 150 fontes de dados de órgãos públicos a fim de aumentar ainda mais a profundidade do conhecimento sobre o cliente. Por exemplo, pode-se analisar o grau de captação do potencial (share de mercado) de uma determinada loja utilizando dados secundários de diferentes órgãos oficiais, estimando o potencial de consumo desse cliente e o quanto ele realmente consumiu. Por último são acrescentados dados geográficos, ou seja, a localização residencial do cliente no mapa, permitindo melhor precisão nos estudos de áreas de influência e raios de atuação da loja.

Esses são os pilares do ecossistema, dados primários (carrinho de compra), dados secundários e contexto geográfico. Desta forma, resultam aos gestores e tomadores de decisão análises descritivas, preditivas e prescritivas do seu negócio, além de meios de relacionamento e ativação direta do consumidor, pois para ele poderá ser ofertado a experiência/produto adequado na hora certa e da forma certa.

3 Organização estrutural do ZoOmBOX

As Figuras 1 e 2 apresentam a arquitetura ZoOmBOX de forma conceitual, abstraindo detalhes tecnológicos. Ela está dividida em duas grandes plataformas: uma implantada no ambiente do varejista, utilizando sua infraestrutura local de rede, servidores e internet; e outra na nuvem, que recebe e centraliza os dados de todos os varejistas. A primeira denominada ZoOmBOX Core e a segunda ZoOmBOX Discovery, respectivamente.

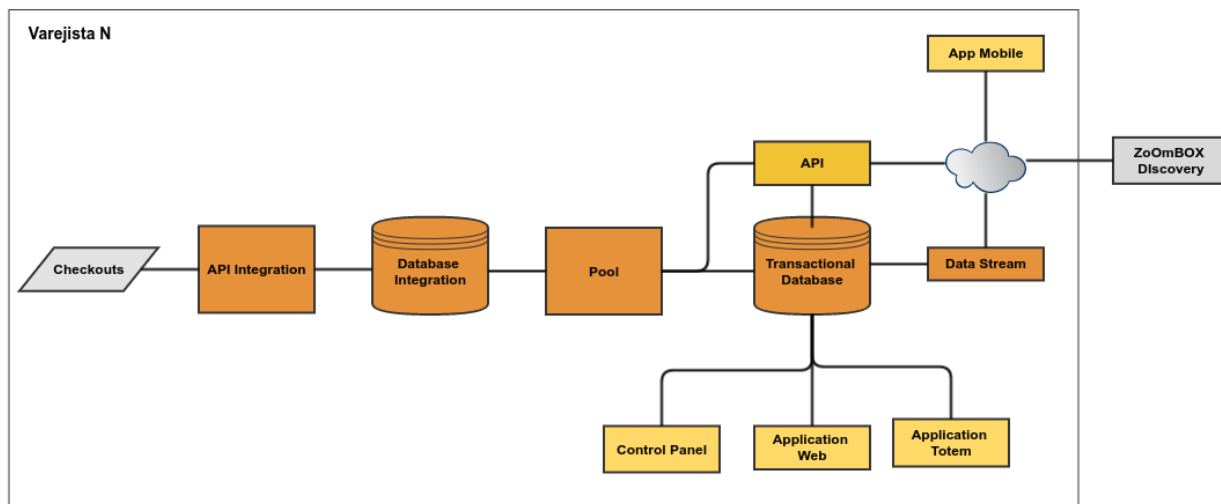
3.1 ZoOmBOX Core

A Figura 1 mostra o modelo conceitual da plataforma; seu principal objetivo é capturar os dados de vendas do varejista e realizar toda a gestão do clube de fidelidade, ações promocionais e ainda toda a gestão da base cadastral de clientes. Executa na infraestrutura local da loja, por questões de desempenho, evitando latência de rede no processo de integração com o checkout, e também por questões de segurança, pois os varejistas ainda têm receio de manter seus sistemas totalmente em nuvem, apesar deste cenário tender a mudar rapidamente.

3.1.1 API Integration

É o módulo responsável por capturar as vendas de cada cliente. Os dados capturados são a forma de pagamento, os produtos comprados ou o carrinho de compra e a identificação do cliente pelo CPF. Após a captura de cada venda é realizada uma checagem de integridade da mesma; caso a venda esteja íntegra, ela é armazenada no repositório Database Integration.

Figura 1 – ZoOmBox Core



Fonte: Elaborado pelas autores.

Todas as vendas íntegras recebidas são armazenadas no repositório em uma espécie de fila, para que o próximo módulo as processe. Outra importante função desse módulo é isolar as especificidades do sistema de checkouts, ou seja, ele funciona como um driver. Para cada sistema de checkout, temos um módulo específico que captura os dados e os normaliza no formato ZoOmBOX. Desta forma, qualquer alteração ou mudança no sistema de checkouts não se propagam para o restante do sistema.

3.1.2 Pool

É o módulo que recupera as vendas do repositório e realiza todo o processamento de geração de benefícios para cada cliente. O Pool obtém os metadados referentes aos parâmetros utilizados para a geração dos benefícios, por exemplo, para a geração dos pontos do programa de fidelidade, o parâmetro seria, a cada R\$ 50,00, um ponto. Para a geração de cupons promocionais, o parâmetro seria, a cada R\$ 20,00, um cupom promocional para concorrer a sorteio. Caso algum produto configurado apareça no carrinho de compras, mais pontos ou cupons podem ser gerados. Existe um completo mecanismo de combinações de parâmetros para geração de benefícios. Baseado no valor da compra,

produtos no carrinho, dia da semana, forma de pagamento são alguns exemplos. As configurações de todos os parâmetros são realizadas pelo módulos Control Panel e armazenados no Transactional Database. Todos os dados gerados referentes aos benefícios e à venda processada são armazenados no repositório Transactional Database.

3.1.3 Control Panel

A aplicação web é responsável pela configuração de todos os parâmetros utilizados pelo Pool, bem como toda a gestão cadastral da base de clientes. Possui também todos os relatórios operacionais necessários para gestão dos benefícios, como também as configurações de front-end para os módulos Application Web e Application Totem. Esse módulo é apenas utilizado por gestores da plataforma e não pelo cliente consumidor final.

3.1.4 Application Web e Totem

Esses dois módulos são responsáveis pela interação com o cliente, ou seja, por meio dessas aplicações web os clientes podem realizar e atualizar o seu cadastro, consultar e utilizar os seus benefícios, consultar ofertas exclusivas e personalizadas para o cliente, responder a pesquisas e ganhar prêmios. Application Web permite o acesso a partir de qualquer navegador independentemente do lugar onde o cliente esteja. Enquanto isso, o Application Totem executa em um terminal de autoatendimento localizado fisicamente na loja do varejista.

3.1.5 API e App Mobile

Este módulo provê todo back-end para o aplicativo mobile, além de disponibilizar inúmeros endpoints para integração com terceiros, tanto para importação quanto exportação de dados. O App Mobile permite ao usuário toda a gestão de seu cadastro, visualização e utilização dos benefícios. Por fim, recebe mensagens e notificações contendo promoções personalizadas, lembrete de compras e sugestões para o seu carrinho de compra.

3.1.6 Data Stream

Este módulo envia os dados da base transacional local para a nuvem, onde são processados e analisados, gerando indicadores, dashboards e análise em tempo real.

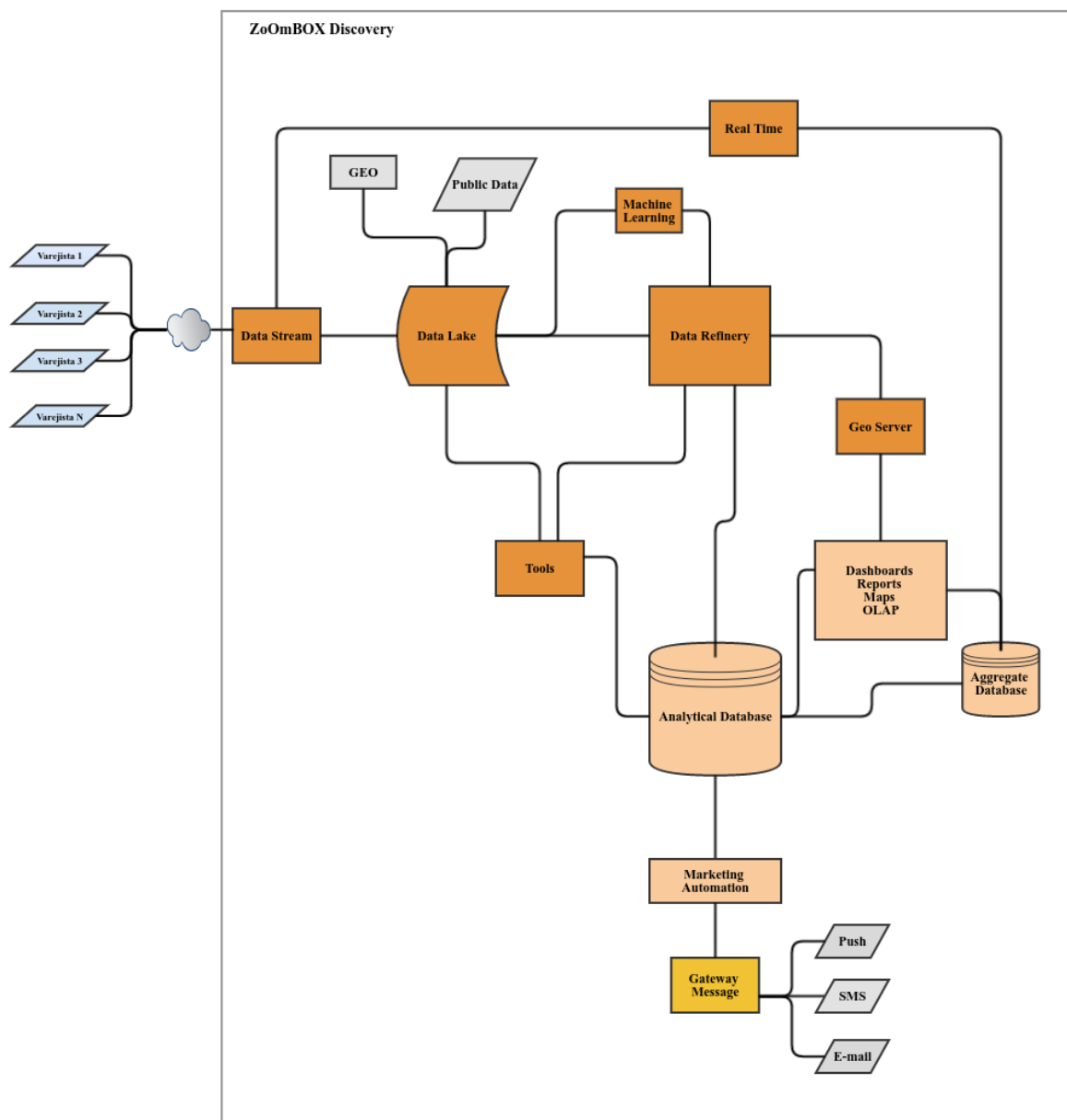
3.2 ZoOmBOX Discovery

A Figura 2 mostra o modelo conceitual do ZoOmBOX Discovery, plataforma que executa em nuvem, agrupando os dados de todos os varejistas e disponibilizando análises descritivas e preditivas sobre os mesmos.

Os dados são recebidos e armazenados em um Data Lake, em seguida processados, organizados e enriquecidos. Os mesmos ficam dispostos em estrutura colunar para maior eficiência na execução de consultas analíticas. Os dados tratados e enriquecidos são armazenados em uma estrutura mais granular permitindo consultas em qualquer nível e hierarquia, e também em uma estrutura agregada para apresentar indicadores em tempo real.

Esses dados estruturados são disponibilizados para usuários gestores no varejista por meio de dashboards. Além da visualização dos dados, existem os módulos Marketing Automation e Gateway Message responsáveis pela segmentação e ativação de clientes.

Figura 2 – ZoOmBox Discovery



Fonte: elaborado pelos autores.

3.2.1 Data Stream

É responsável pela captura dos dados enviados pelo varejista (ZoOmBOX Core) e armazenamento no Data Lake, além do envio para o módulo Real Time para agregação e análise em tempo real.

3.2.2 Data Lake

É responsável pelo armazenamento dos dados brutos dos varejistas, com as diversas bases de dados secundários. Acrescenta-se a todos esses dados, o contexto geográfico.

Todos os clientes da base ZoOmBOX são geolocalizados por meio das coordenadas de latitude e longitude, podendo assim serem relacionados a layers com informações geográficas, gerando vários contextos de análises. O objetivo do Data Lake é armazenar os dados brutos para, se necessário, serem reprocessados e agregados de acordo com novas necessidades de análises do negócio. Quando os dados processados são apenas armazenados, corre-se o risco de perder detalhes que não poderão ser mais recuperados.

Essa estrutura também possui outra vantagem, que é a possibilidade de armazenamento de dados não estruturados, além dos já tradicionais dados estruturados.

3.2.3 Data Refinery

Este módulo contempla todo o pipeline de processamento dos dados, usando a estratégia de processamento/armazenamento, ou seja, a cada etapa os dados intermediários são armazenados, para que outro processo continue, desacoplando e isolando todas as etapas.

Existe, também, o processo que normaliza os dados primários do varejista, estruturando-os e gerando o catálogo de metadados, . A normalização pode ser, por exemplo, o mapeamento dos produtos para descrição e categorização comum a todos os varejistas, considerando que cada um deles possui um modelo próprio de categoria.

Por meio do endpoint Tools e seu console (por exemplo, SQL/R/Python) é possível explorar detalhes dos dados e subir catálogos de mapeamentos de acordo com necessidades específicas.

Outro processo agrega aos dados primários normalizados e os dados secundários. Por fim, além dos dados secundários, são agregados também vários layers de mapas, construídos por meio de ferramentas de GIS (Geographic Information System) e inseridos

no Data Lake. Esses layers são disponibilizados por meio do módulo Geo Server, e os dados normalizados e enriquecidos são enviados ao repositório Analytical Database.

Esse módulo ainda realiza a execução dos modelos estatísticos de Machine Learning, retroalimentando os dados processados, como pode-se citar a detecção de padrões e segmentação de clientes segundo seu comportamento de consumo, ou seja, produtos comprados.

3.2.4 Machine Learning

Arelado ao grande volume de dados recebidos e armazenados pelo ecossistema ZoOmBOX podem ser realizadas análises preditivas. Tais análises possuem a finalidade de detectar padrões de consumo e revelar quais os produtos têm maior chance de serem procurados pelo cliente em uma futura compra. Nesse meio, figura o conceito de Machine Learning.

Tal conceito utiliza de algoritmos supervisionados ou não para a tomada de decisões, aprendizado de máquina, como seu próprio nome diz, e previsões de comportamentos futuros. Dentre os algoritmos mais utilizados estão árvores de decisão, regras de associação e classificação e redes neurais (AHA; KIBLER; ALBERT, 1991).

Assim, esse módulo é responsável pela análise dos carrinhos de compras dos consumidores, classificando quais tipos de compras o cliente realiza, pelo número de itens que aparecem no carrinho. Além disso, também é feita a predição dos produtos que possivelmente estarão presentes nas próximas compras do cliente, para que, assim, possam ser enviadas ofertas ao mesmo.

3.2.5 Tools

Este módulo é dedicado à equipe de ciência de dados, que realiza a exploração dos dados, mapeamentos para normalização e a construção de testes dos modelos estatísticos e de Machine Learning. Disponibiliza um endpoint para conexão e execução de comandos SQL ou execução de scripts, permitindo a exploração dos dados, download e upload dos mesmos.

3.2.6 Analytical Database

É um banco de dados analítico que armazena os dados em uma estrutura colunar e de forma compactada, permitindo alto desempenho em consultas em grandes volumes de dados, mesmo em estruturas desnormalizadas; todas as tabelas nesta base estão desnormalizadas a fim de minimizar o máximo possível latência na execução de joins. Possui um motor de processamento massivamente paralelo, realizando shard dos dados. Realiza todo backend de processamento analítico para todos os módulos subsequentes. A fim de melhorar ainda mais o tempo de resposta do módulo de Dashboards, alguns dados são agregados em um schema à parte denominado Aggregate Database.

3.2.7 Dashboards/Report/Maps/OLAP

É a interface com os gestores do varejista, permitindo análises descritivas e preditivas por meio de dashboards interativos e indicadores do negócio. Possui uma ferramenta analítica que permite drill down dos dados, a fim de explorar o nível mais granular possível. Algumas análises disponibilizadas são: evolução do faturamento nas dimensões de tempo, tíquete médio, gasto médio, curva de Pareto - que são os 20% dos clientes responsáveis por 80% do faturamento da loja (LIPOVETSKY, 2009) -, análise do perfil cadastral desses clientes e análise no contexto geoespacial.

Por meio da interface de mapas é possível visualizar e selecionar os clientes de acordo com a área de influência da loja, podendo analisar o faturamento real realizado frente ao faturamento em potencial, gerando o indicador se a loja já está faturando seu teto ou ainda existe margem para crescimento.

3.2.8 Marketing Automation e Gateway Message

Todas as análises realizadas pelo módulo anterior são importantes, pois mostram como está o desempenho do negócio e também subsidia tomada de decisões; além disso, essas análises podem ser retroalimentadas para o sistema. Porém, o objetivo final de todo

esse processo é o conhecimento profundo do cliente para fidelizá-lo à loja, aumentando sua recorrência de compra, seu tíquete médio, e, conseqüentemente, o faturamento da loja. Tudo isso é possível com a ativação do cliente no momento certo, oferecendo a experiência ou produto certo, da forma certa e pelo canal certo. O cliente deve se sentir único e individualizado, recebendo uma comunicação de acordo com o seu perfil e as suas preferências.

Frente a essa necessidade e ao grande volume de dados e clientes, todo esse processo de ativação deve ser automatizado, que é o objetivo desse módulo. Segmenta a massa de clientes de acordo com filtros pré-configurados, enviando a mensagem necessária por um ou todos os canais disponíveis, SMS, E-mail ou Push. Realiza os disparos de forma agendada, periódica ou ainda por meio de gatilhos de eventos específicos, por exemplo, quando o cliente realiza uma compra e não compra um produto que sempre costuma comprar, ele recebe uma mensagem de lembrete. Uma outra possibilidade é a geração de tabloide digital de ofertas individual para cada cliente, com ofertas relevantes de acordo com o seu hábito de consumo; suponha, por exemplo, uma loja com 100 mil clientes, cada um receberá uma mensagem individualizada de acordo com o seu perfil de consumo.

4 Resultados e Desempenho

A seguir estão os resultados obtidos até o momento com o ecossistema ZoOmBOX divididos em duas etapas: a primeira relacionada ao desempenho em nível tecnológico e de capacidade de processamento e a segunda em nível de negócio.

Números relacionados aos dados já processados até o momento:

- 2.4 Mil lojas de varejo e atacado;
- 7 Milhões de clientes identificados e analisados;
- 2.6 Trilhões de produtos processados e analisados;
- 200 Milhões de compras processadas e analisadas;
- 1.6 Milhões de tabloides digitais gerados e entregues;
- Capacidade de processamento diário de mais de 15 Milhões de vendas;

- Consultas analíticas complexas em menos de 30 segundos, com várias agregações - tabelas temporárias, tabelas com mais de 500 milhões de registros.

5 Conclusões e Trabalhos Futuros

Após a grande quantidade de dados processados e analisados, o ecossistema ZoOmBOX mostrou-se uma ferramenta eficiente para a captura e processamento eficaz do comportamento do cliente e tem sido utilizado com satisfação, tornando-se útil para o varejo brasileiro.

Deste modo, como trabalhos futuros, tem-se como objetivo melhorar os mecanismos de análise e Machine Learning com vista à geração de inteligência acionável, permitindo automatizar e personalizar a comunicação entre o varejo e o cliente.

Referências

- AHA, David W.; KIBLER, Dennis; ALBERT, Marc K. Instance-based learning algorithms. *Machine learning*, v. 6, n. 1, p. 37-66, 1991.
- BATRINCA, Bogdan; TRELEAVEN, Philip C. Social media analytics: a survey of techniques, tools and platforms. *AI & SOCIETY*, v. 30, n. 1, p. 89-116, 2015.
- CHEN, Min; MAO, Shiwen; LIU, Yunhao. Big data: A survey. *Mobile Networks and Applications*, v. 19, n. 2, p. 171-209, 2014.
- GOLE, Sheela; TIDKE, Bharat. A survey of big data in social media using data mining techniques. In: *Advanced Computing and Communication Systems, 2015 International Conference on*. IEEE, 2015. p. 1-6.
- LIPOVETSKY, Stan. Pareto 80/20 law: derivation via random partitioning. *International Journal of Mathematical Education in Science and Technology*, v. 40, n. 2, p. 271-277, 2009.
- MAÇADA, Antônio Carlos Gastaud; BRINKHUES, Rafael Alfonso; JÚNIOR, José Carlos Freitas. Big data e as capacidades de gestão da informação. *ComCiência*. Disponível: <<http://www.comciencia.br/comciencia>>.
- MCAFEE, Andrew et al. Big data: the management revolution. *Harvard business review*, v. 90, n. 10, p. 60-68, 2012.