

CARLOS AUGUSTO MATUGUMA

**ANÁLISE COMPARATIVA DE FERRAMENTAS DE DATA DISCOVERY – UM
ESTUDO DE CASO**

São Paulo

2014

CARLOS AUGUSTO MATUGUMA

**ANÁLISE COMPARATIVA DE FERRAMENTAS DE DATA DISCOVERY – UM
ESTUDO DE CASO**

Monografia apresentada ao PECE – Programa de Educação Continuada da Universidade de São Paulo, como parte dos requisitos para obtenção do título de MBA em Tecnologia da Informação.

Orientadora: Profa. Dra. Solange Nice Alves de Souza

São Paulo

2014

MBA/TI
2014
Matuguma

551



Escola Politécnica - EPEL



31500023551

FICHA CATALOGRÁFICA

2014AHV X

Matuguma, Carlos Augusto

ANÁLISE COMPARATIVA DE FERRAMENTAS DE DATA
DISCOVERY – UM ESTUDO DE CASO/ C. A. Matuguma. --
São Paulo, 2014.54 p.

Monografia (MBA em Tecnologia da Informação) –
Escola Politécnica da Universidade de São Paulo. Programa
de Educação Continuada em Engenharia.

PCG

1.Data Discovery 2.Business Intelligence. Universidade de
São Paulo. Escola Politécnica. Programa de Educação
Continuada em Engenharia II.t.

[267017301]

CARLOS AUGUSTO MATUGUMA

**ANÁLISE COMPARATIVA DE FERRAMENTAS DE DATA DISCOVERY – UM
ESTUDO DE CASO**

Monografia apresentada ao PECE – Programa de Educação Continuada da Universidade de São Paulo, como parte dos requisitos para obtenção do título de MBA em Tecnologia da Informação.

Orientadora: Profa. Dra. Solange Nice Alves de Souza

Aprovação em:

São Paulo

2014

RESUMO

As soluções de inteligência de negócios, BI ou *business intelligence*, estão em constante evolução e cada vez mais precisam ser capazes de rapidamente gerar conhecimento a partir dos dados. O cenário de BI dos últimos anos revelou soluções inovadoras e criou uma nova categoria de produtos, as ferramentas de Data Discovery. Essas soluções dão autonomia aos usuários de negócio, deixando-os livres para explorar os dados, montando e publicando análises visuais. Fez parte deste trabalho uma revisão na literatura buscando artigos com foco em BI, análises visuais e seus benefícios para as organizações. Após esse estudo, foi proposta uma análise comparativa entre três ferramentas de Data Discovery fornecidas por: TABLEAU, QLIK e TIBCO. A comparação foi realizada utilizando um projeto piloto, que funcionou como um caso de uso, em uma empresa da área de cosméticos. A seleção da área em que o projeto foi conduzido ocorreu de forma criteriosa pois ela deveria refletir as necessidades da empresa dado que a adoção da ferramenta escolhida ocorreria de forma institucional. Em outras palavras, a ferramenta escolhida será utilizada por todas as demais áreas da organização. O trabalho incluiu a sintetização de critérios para a análise qualitativa das ferramentas, bem como o detalhamento dos benefícios trazidos pela utilização das mesmas. Assim, esse trabalho apresenta os estudos realizados para a escolha da área do projeto piloto para comparação das 3 ferramentas de Data Discovery, os critérios utilizados para a avaliação dessas ferramentas, os resultados obtidos na aplicação desses critérios. Discute-se ainda a importância de tais ferramentas dentro do contexto de BI.

ABSTRACT

Business intelligence suites are constantly evolving therefore there is the need to be able to quickly generate knowledge from data. The BI scenario of recent years revealed innovative solutions and created a new category of solutions, the Data Discovery tools. These solutions provide autonomy to business users, leaving them free to explore the data, assemble and publish visual analysis. Is part of this paperwork a review in the literature by searching for articles containing the following themes: BI, visual analysis and its benefits for organizations, Big Data and Data Discovery tools. After this study, a comparative analysis among three suppliers: TABLEAU, QLIK and TIBCO was done. The comparison was made using a pilot project, working as a use case, within the scenario of a cosmetics company. The selection of the area in which the project was conducted was carefully defined because it should reflect the company's needs since the adoption of the chosen tool would occur in an institutional manner. The work included the synthesis of criteria for qualitative analysis tools, as well as the details of the benefits brought by its utilization. Thus, this paper presents the studies conducted for the choice of the pilot project area to compare the three tools of Data Discovery, the criteria used to evaluate these tools, the results obtained in applying these criteria. It also discusses the importance of such tools within the BI environment.

LISTA DE FIGURAS

Figura 1: Representação da arquitetura de um sistema de BI genérico.....	5
Figura 2: Representação do momento da informação e as análises necessárias para cada um.....	9
Figura 3: Representação dos tipos de análises dos sistemas BI.	10
Figura 4: Representação gráfica de um painel construído na ferramenta QlikView. .	16
Figura 5: Representação de dados sobre diabetes nos Estados Unidos, construída na ferramenta Tableau.	17
Figura 6: Infográfico sobre os estados americanos, construído na ferramenta Spotfire.....	18
Figura 7: Exemplo de nuvem de palavras.	19
Figura 8: Representação gráfica dos tipos de análises dos sistemas de BI tradicionais.	21
Figura 9: representação gráfica dos tipos de análises dos sistemas de BI tradicionais sobreposto pelas ferramentas de Data Discovery.....	22
Figura 10: Representação dos quadrantes mágicos do Gartner para ferramentas de DataDiscovery.	25

LISTA DE TABELAS

Tabela 1 - Especificação dos produtos de Data Discovery selecionados para avaliação	31
Tabela 2 - Avaliação das ferramentas	34
Tabela 3 - Pontuação das ferramentas de Data Discovery testadas, para os critérios (por grupo e subgrupo) definidos. Destacam-se os maiores pontuadores de cada subgrupo.	37

LISTA DE GRÁFICOS

Gráfico 1 - Gráfico que mostra a pontuação dada a cada uma das ferramentas analisadas de acordo com os critérios (grupo e subgrupo) de avaliação que foram determinados.....38

SUMÁRIO

1. INTRODUÇÃO	1
1.1 Contexto	1
1.2 Objetivo.....	2
1.3 Metodologia	2
1.4 Estrutura do trabalho.....	2
2. FUNDAMENTOS TEÓRICOS	4
2.1 Business intelligence (BI) ou inteligência de negócios.....	4
2.1.1 Conceito de business intelligence	4
2.1.2 Componentes dos sistemas de BI.....	5
2.1.3 Maturidade do sistema de BI.....	7
2.1.4 Business analytics ou análise de negócios	8
2.1.5 BI e o valor para os negócios.....	12
2.2 BIG DATA.....	13
2.3 DATA DISCOVERY	14
2.3.1 BI Tradicional x Data Discovery	19
2.3.2 Data Discovery e estratégia Big Data.....	22
2.3.3 Mercado e principais fornecedores de ferramentas de Data Discovery	23
2.4 TRABALHOS CORRELATOS	25
3. EXPERIMENTO	29

3.1 Critérios	31
3.2 Análise comparativa e resultados	36
4. CONCLUSÕES E CONSIDERAÇÕES FINAIS	40
REFERENCIAS.....	42

1. INTRODUÇÃO

1.1 CONTEXTO

O presente trabalho tem como contexto o mercado competitivo dos dias atuais e como as novas soluções analíticas, denominadas de *Data Discovery* (Descoberta de Dados) ou, conforme alguns autores, *Agile BI (Business Intelligence – Inteligência de Negócio Ágil)*, ajudam a conquistar vantagens estratégicas fundamentais para a permanência das empresas nos negócios.

Os investimentos nesse tipo de solução estão acontecendo de forma expressiva nos últimos anos e é visível como a procura por profissionais nesse ramo vêm aumentando. Contudo, por ser um assunto ainda em rápida ascensão, não existe uma fórmula correta e nem um método eficaz para a implementação desse tipo de solução. Temos ainda que os benefícios não aparecem ser tão claros para o negócio de modo a justificar o investimento, apesar de existir algumas experiências de sucesso na literatura, como por exemplo nas aplicações citadas em (TIEN, 2013).

As soluções analíticas atuais possuem um alto custo de desenvolvimento e também um ciclo de desenvolvimento longo para a implementação de um projeto analítico (AIGNER, 2013). Assim, o surgimento de ferramentas de rápida implantação e menor complexidade para a utilização e customizações, é visto como um meio de contornar essa dificuldade. Tais ferramentas são oferecidas nos pacotes de soluções para inteligência de negócios (*Business Intelligence, BI*) fornecidos por empresas relativamente novas em comparação aos grandes fornecedores. O foco dessas ferramentas é permitir ao usuário final a capacidade de explorar as informações de forma *self-service*, o que significa que ele próprio formula as consultas que deseja, e ainda pode gerar e publicar análises gráficas customizadas.

As vantagens da utilização de tais ferramentas diante das tradicionalmente usadas são várias, podendo-se destacar:

- Redução de custos na compra de soluções específicas para análise de dados, uma vez que tais funcionalidades já são oferecidas nas soluções de BI;
- Não necessidade de profissionais de TI para elaboração de diferentes consultas analíticas para a obtenção e/ou visualizações diversas da informação de valor para o negócio. Isso pois, dado a facilidade de utilização, os próprios usuários de negócio podem formular as consultas e definir a forma de visualização que deseja da informação obtida como resultado da consulta.

1.2 OBJETIVO

O principal desafio que se objetiva neste trabalho foi realizar uma análise comparativa entre ferramentas de Data Discovery, buscando relacioná-las com os recursos computacionais adequados as necessidades e especificidades da organização.

Importante destacar que o estudo conduzido neste trabalho considerou apenas o aspecto técnico das ferramentas analisadas. Alguns pontos relacionados ao negócio podem ter sido abordados e considerados, porém, a análise econômica e financeira não foi refletida por este trabalho.

Destacam-se como objetivos específicos, sintetizar critérios para a análise qualitativa das ferramentas e destacar benefícios trazidos pela utilização das mesmas.

1.3 METODOLOGIA

Inicialmente foi feita uma pesquisa bibliográfica em que foram buscadas publicações relacionadas ao tema, tanto ferramentas de Data Discovery quanto

arquitetura Big Data, e também os impactos da inteligência de negócios (*business intelligence*) para a atividade comercial.

Na sequência, realizou-se a análise dos artigos, verificando similaridades e divergências com a finalidade de verificar como seria a aplicação das ferramentas de Data Discovery na arquitetura de BI e Big Data, e como isso pode influenciar os negócios e tomadas de decisões.

Os artigos estudados foram utilizados para a definição de critérios para uma análise comparativa entre três ferramentas de fornecedores de soluções de Data Discovery: QLIK, TABLEAU e TIBCO.

1.4 ESTRUTURA DO TRABALHO

O presente trabalho está organizado em 4 capítulos. No capítulo 1 tem-se a introdução do trabalho, onde se discute a motivação, além de apresentar o objetivo, metodologia seguida para a elaboração deste trabalho.

O capítulo 2 apresenta a fundamentação teórica, abordando os principais conceitos envolvidos com o tema do trabalho aqui apresentado. Este capítulo traz ainda um resumo de alguns artigos que apoiaram a elaboração deste trabalho.

O capítulo 3 detalha o experimento realizado para a comparação das ferramentas 3 ferramentas de Data Discovery escolhidas, discute os critérios empregados para essa comparação, além de apresentar e discutir os resultados obtidos.

O capítulo 4 traz comentários finais, conclusões e propostas de trabalhos futuros.

2. FUNDAMENTOS TEÓRICOS

2.1 BUSINESS INTELLIGENCE (BI) OU INTELIGÊNCIA DE NEGÓCIOS

2.1.1 Conceito de BUSINESS INTELLIGENCE

O conceito de Inteligência pode ser definido como “uso com destreza da razão ou habilidade de aplicar o conhecimento para manipular o meio ou pensar de forma abstrata”. Segundo (HANS, 2013), a aplicação da inteligência aos negócios requer o uso de informações e se relaciona diretamente à melhora sustentável do desempenho dos negócios, ou seja, os dados são transformados em informação e conhecimento, que são utilizados para dar mais eficiência e precisão aos processos de tomada de decisão, guiando a organização a um futuro mais próspero.

Da mesma forma, CALOF e SKINNER citados por (HANS, 2013) definem BI ou Inteligência de negócios de uma forma um pouco mais aprofundada:

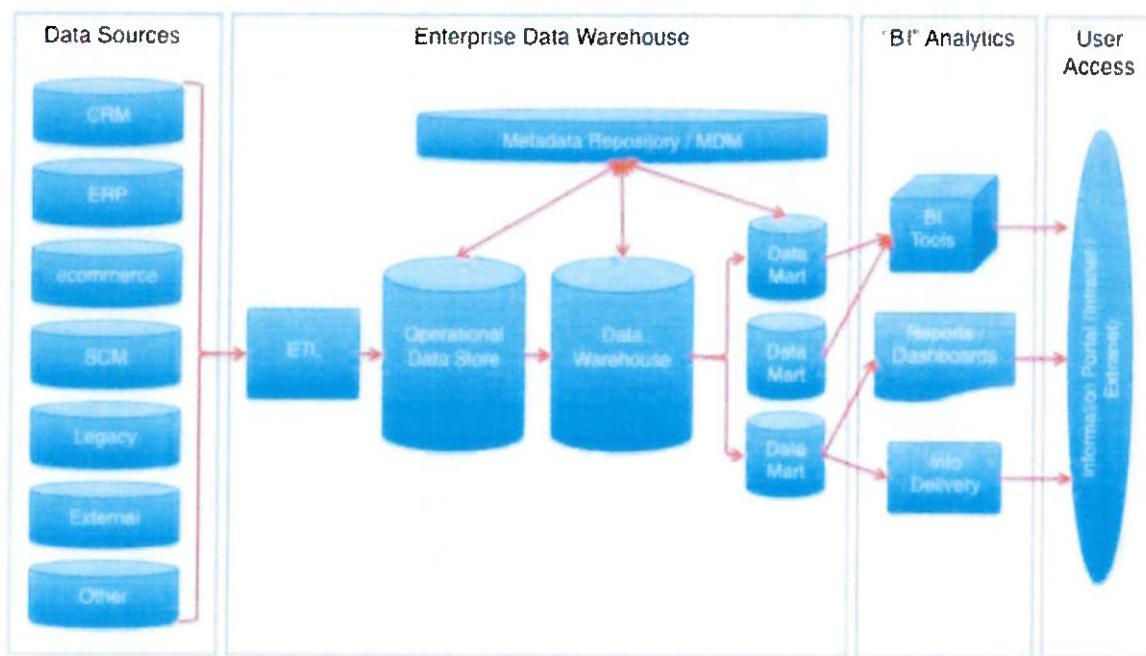
“(...) a arte e ciência de preparar as empresas para o futuro pelo caminho de um processo de gerenciamento de conhecimento sistemático. Ou seja, gerar conhecimento das informações disponíveis utilizando um processo sistemático que envolva planejamento, coleta, análise, comunicação e gerenciamento, cujo resultado é a ação de uma tomada de decisão.”

Assim sendo, várias definições e conceitos podem ser relacionadas a BI dependendo do campo de aplicação. Independente da maneira que as pessoas vejam é importante que o BI seja tratado como um valioso núcleo de competência, posto que em todas as visões têm referência ao grande valor trazido às organizações pelo BI.

Neste trabalho, BI é interpretado como um ambiente composto por ferramentas e procedimentos, conforme a Figura 1, que manipulam dados corporativos, gerando

informação de valor para a organização, que apoiam tomadas de decisões estratégicas.

Figura 1: Representação da arquitetura de um sistema de BI genérico.



2.1.2 Componentes dos sistemas de BI

Um ambiente de BI é composto por um conjunto de elementos, cada um dos quais fornece entrada para o elemento seguinte (MOHAMADINA, 2012). Esses elementos, ilustrados na Figura 1, são apresentados de forma detalhada a seguir:

Fontes de dados: as fontes de dados representam qualquer meio que a organização utiliza para obter dados, e variam desde bases de dados transacionais até dados coletados da internet ou pesquisas de mercado de empresas terceirizadas. As informações das fontes de dados podem se apresentar em diferentes formatos, como tabelas, planilhas ou até dados não estruturados como arquivos texto, imagens e informações multimídia.

ETL (Extração, transformação e carga): o termo genérico ETL representa a etapa que reúne os processos que culminarão na carga dos dados a partir dos sistemas operacionais para o repositório de Data Warehouse (DW) ou Data Marts da organização. Ferramentas específicas são usadas no processo ETL, que é dividido em três etapas, sendo a primeira a leitura das fontes de dados para a extração dos mesmos; a segunda a transformação dos dados adquiridos para a formatação correta de entrada para o DW, através de funções de conversões e cruzamentos com tabelas de dimensões; e a terceira a carga dos dados no DW. Faz parte do processo ETL o tratamento da qualidade dos dados fontes.

Repositório de dados (DW): o DW é um dos principais componentes do sistema de BI, que suporta uma grande quantidade de dados. O DW contém dados históricos, normalizados e organizados em tabelas fato e dimensões, de forma a facilitar a visualização de dados agrupados, que facilitam a visualização de informações de negócio não facilmente identificadas nos sistemas fontes.

Processamento Analítico On-Line (OLAP): a camada OLAP representa uma categoria de softwares e ferramentas que possibilita aos analistas a capacidade de buscar informações do DW de forma rápida, consistente e interativa. As ferramentas OLAP permitem aos usuários analisar diferentes dimensões de informações multidimensionais, como por exemplo a navegação dos dados na dimensão do tempo. Em geral, esse tipo de ferramenta e solução é desenhada para trabalhar com um volume de dados alto para apoiar o processo de tomadas de decisões estratégicas, e por essa razão a implementação da camada OLAP é feita de modo a ser utilizado por diversos usuários simultaneamente (cliente-servidor) e deve apresentar um bom desempenho, independente da complexidade da requisição. A solução é desenhada de modo a responder perguntas do tipo “e se?”, auxiliando a descoberta de soluções, caso existam nos próprios dados. O OLAP corresponde a camada “BI Analytics” na Figura 1

Mineração de dados (*Data Mining*): A mineração de dados é um processo automatizado de descoberta de padrões. Esse processo ajuda a obter relações uteis à tomada de decisão e dificilmente perceptíveis pesquisando diretamente nas fontes

de dados. Também correspondem à camada “BI Analytics” demonstrada na Figura 1.

Relatórios: os relatórios (representados na Figura 1) representam as ferramentas e soluções que permitem a criação de relatórios utilizando outros relatórios como fonte e baseados em *querys* (consultas) ad-hocs.

Camada de apresentação (Figura 1): representa o estágio final que apresenta interfaces com recursos gráficos opções multimídia para os usuários finais.

A Figura 1 esquematiza uma arquitetura de BI, nesta o componente BI TOOLS abriga as ferramentas OLAP e Data Mining.

2.1.3 Maturidade do sistema de BI

CMMI, Modelo Integrado de Capacitação para Maturidade (*Capability Maturity Model Integration*) consiste na evolução do CMM, Modelo de Capacitação para Maturidade (*Capability Maturity Model*), que foi criado no Instituto de Engenharia de Software da Universidade de Carnegie Mellon. O framework proposto por (CHUAH, 2010) avalia a maturidade dos sistemas de BI das organizações, classificando tais sistemas em estágios similares ao do CMM.

A relação do CMM com o BI separa a maturidade dos sistemas de BI em cinco, são eles:

Nível 1, inicial: nesse estágio não existem regras ou normas para o gerenciamento das informações, que residem em diversas fontes de diferentes formatos, o que gera muita redundância e replicação de dados. A qualidade dos dados depende muito da habilidade dos programadores e analistas de dados, e a geração de conhecimento é feita de forma individualizada, focando nos problemas e entregas do dia a dia, ao invés de um alinhamento com estratégia da organização.

Nível 2, repetitivo: nesse estágio existe uma política de como e quando os dados estruturados devem ser criados, armazenados, bem como regras e requisitos de atualização. Define-se ainda o responsável pelo gerenciamento. Devem existir documentações a serem seguidas a fim de garantir a qualidade das informações. As informações são coletadas e armazenadas e podem ser consultadas no nível de cada departamento.

Nível 3, definido: Os dados possuem tratamento de um bem valioso para a organização. São implementadas as melhores práticas para desenvolvimento, manutenção e utilização do DW em todos os setores. A política de gerenciamento de dados se torna uma componente chave no ciclo de vida do desenvolvimento de aplicações. Dados e métricas criadas por diversos setores são alinhados, compartilhados e integrados em um mesmo local.

Nível 4, gerenciado: nesse estágio é introduzido um ambiente gerencial de METADADOS (dados sobre os próprios dados), permitindo que as informações corporativas sejam catalogadas e gerenciadas de forma estruturada. Essa gestão permite que a organização possa fazer previsões de desempenho com uma boa acurácia. Ainda, mais recursos são disponibilizados para a atividade de manutenção dos processos da qualidade das informações, e os conceitos e o gerenciamento do conhecimento gerado são aplicados e aceitos pela organização.

Nível 5, otimizando: esse estágio configura um constante processo de melhoria no acesso as informações e desempenho do DW, na qualidade das informações geradas e da gestão de conhecimento adquirido a cada aprendizado.

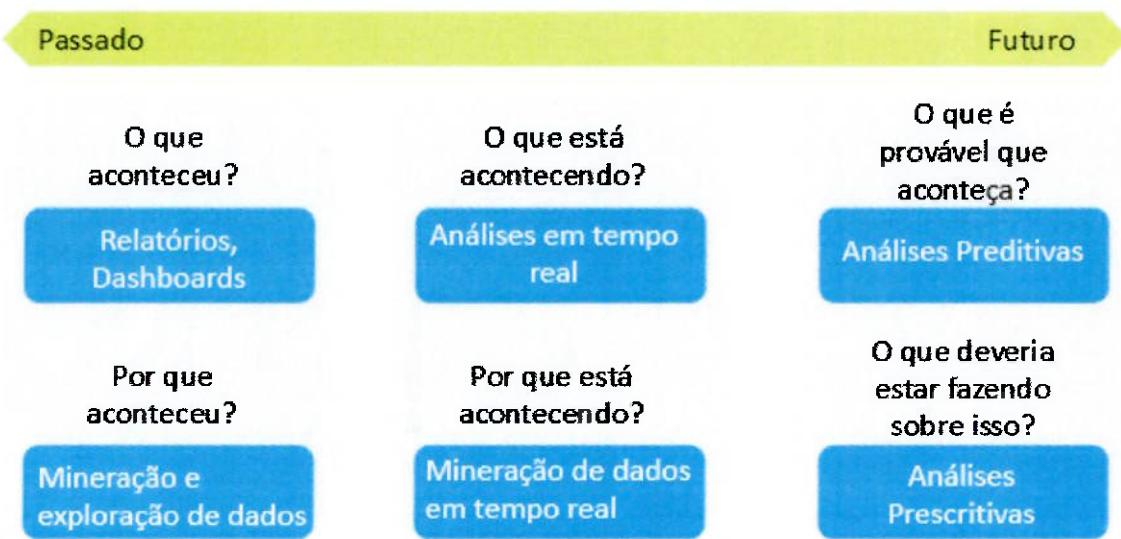
2.1.4 BUSINESS ANALYTICS ou análise de negócios

As ferramentas de BI oferecem capacidades para promover análises, fazer descobertas sobre os dados estruturados da organização e dessa forma apoiar as tomadas de decisões através de indicadores baseados em modelos estatísticos. É importante ter conhecimento sobre os fatos que aconteceram, estão acontecendo e que virão a acontecer, e para cada um desses momentos é necessário um tipo de

estudo ou ferramenta analítica para extrair os aprendizados necessários para a organização reagir corretamente.

A Figura 2 a seguir tem a representação das análises necessárias para cada um desses momentos da informação.

Figura 2: Representação do momento da informação e as análises necessárias para cada um.



(Fonte: Gartner Advanced Analytics Summit 2014, São Paulo)

De acordo com (SHARDA, 2013), as análises de negócio podem ser divididas em três fases distintas (da menor para a de maior complexidade de implementação): descriptiva, preditiva e prescritiva. Na literatura estudada, é citado também outro tipo de análise: a análise diagnóstica; e junto com as demais compõe os tipos de análises existentes atualmente no BI. Existem diversas ferramentas e soluções completas de BI no mercado e cada uma delas tem maior ou menor adequação a um tipo de análise.

A análise descritiva refere-se ao conhecimento do que ocorreu ou está acontecendo na organização.

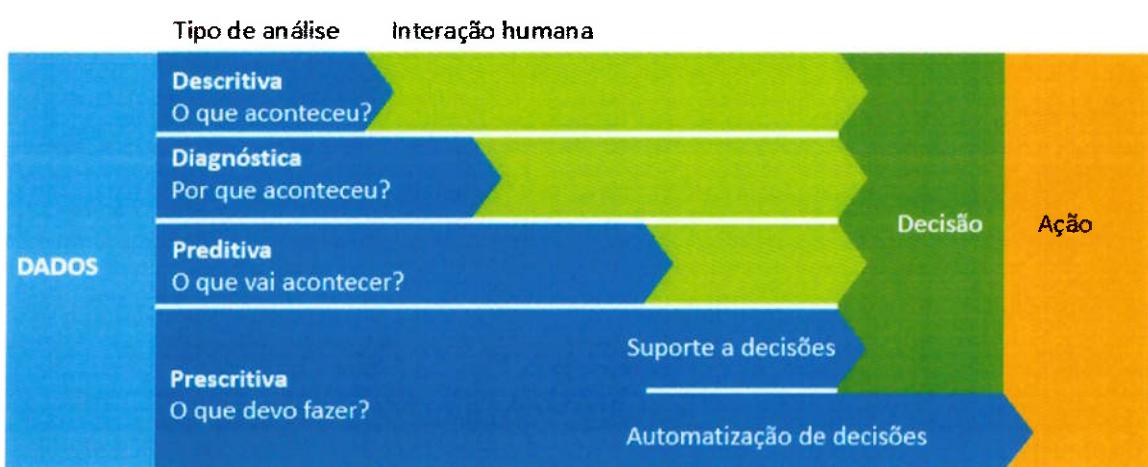
A análise diagnóstica representa as análises que são feitas buscando entender os porquês e os fatos acontecidos para os resultados encontrados na análise descritiva, gerando aprendizados que podem evitar que os mesmos erros sejam cometidos.

As análises preditivas têm como foco o estudo de previsão de informações com confiança, e dessa forma ajudar a planejar melhor as ações e decisões de negócio a serem tomadas.

Um passo adiante, as análises prescritivas têm o objetivo de realizar ou sugerir ações a serem tomadas baseadas em modelos estatísticos avançados.

Na Figura 3, abaixo, estão representados os tipos de análises existentes no BI e como cada tipo de análise depende da interação humana para chegar no seu objetivo. Pode-se observar que conforme a complexidade da análise cresce, a necessidade da interação humana até que seja tomada uma ação para o negócio diminui.

Figura 3: Representação dos tipos de análises dos sistemas BI.



(Fonte: Gartner Advanced Analytics Summit 2014, São Paulo)

Cada um dos tipos de análise precisa de um ou mais ferramentas ou métodos estatísticos para ser executado com precisão, dessa forma é importante pontuar qual tipo de ferramenta se adequa a cada tipo.

Nas análises descritivas temos que é necessária a consolidação dos dados e a disponibilização de todas as informações relevantes para que a análise seja viabilizada através de relatórios ou análises pontuais (AD-HOCS). As soluções para esse tipo de análise, em geral, disponibilizam ferramentas de relatórios e agendamentos que permitem a geração de relatórios operacionais estáticos bem definidos

A análise preditiva tem como meta determinar o que vai ocorrer no futuro com pequena margem de erro. Essa análise tem como base técnicas estatísticas e outras que podem ser qualificadas como técnicas de mineração de dados. Os casos de uso vão desde identificar quais clientes estão propensos a mudar para o concorrente (CHURN, 2010) até identificar qual a promoção ele estaria mais inclinado a comprar. Modelos de árvore de decisão e redes neurais são alguns exemplos de técnicas aplicadas. Para isso são necessários softwares e soluções com boas capacidades estatísticas para os analistas desenvolverem os algoritmos de cada análise.

Na análise prescritiva examina-se as tendências e previsões atuais como base para a tomada de decisões. O histórico das análises estatísticas são recursivamente utilizados para a melhoria e otimização do desempenho do sistema. O resultado dessas análises são respostas ou recomendações específicas, como a resposta sim/não para alguma pergunta ou então uma quantidade específica. Por exemplo o preço de algum artigo ou então limite de crédito. Essas respostas podem ser apresentadas para alguém tomar a decisão ou então automatizada de acordo com uma regra específica definida.

De uma forma geral, o tipo de análise de negócios que é possível fazer pode ser uma maneira de medir a qualidade do BI da organização, posto que para a implementação de uma análise prescritiva com boa acurácia é necessário que o BI

esteja bem estruturado. Contudo, podem existir casos de analistas extremamente talentosos em que conseguem fazer análises sem o BI estruturado.

2.1.5 BI e o valor para os negócios

De acordo com (MOHAMADINA, 2012), a área de tecnologia da informação se tornou indispensável para qualquer organização, e o setor de BI possibilita o gerenciamento das informações geradas de forma eficiente, disponibilizando-as aos gerentes para monitorar, controlar e melhorar os processos de negócio, além de vislumbrar novas possibilidades de negócios.

Indo além, argui-se na literatura (INDRAJANI e LISANTI *apud* HANS, 2013), que o BI tem potencial para promover descobertas e geração de conhecimento o que permitiria ao negócio a tomada de decisões de forma mais assertiva, gerando maior valor econômico, tornando então a organização mais competitiva. Para tanto se faz necessário o uso de ferramentas analíticas aos dados internos e externos da organização para transformá-los em informação e conhecimento sobre as capacidades da empresa, seu posicionamento no mercado, ajudando no estabelecimento de estratégias para a organização.

Dessa forma, o BI fornece o desempenho do passado, auxilia o processo atual e como melhorar as operações no futuro, permitindo as organizações inovarem, manter a excelência dos negócios e planejar o futuro com mais tranquilidade para os interessados. O acompanhamento do desempenho da organização através do BI, ainda de acordo com INDRAJANI e LISANTI, citados por (HANS, 2013) é o caminho mais correto para descobertas e geração de conhecimento de fatores que levem ao sucesso.

De forma mais profunda, SAHAY e RANJAN, também citados por (HANS, 2013), afirmam que o BI deve fornecer informações consistentes para clientes, fornecedores, colaboradores, logísticos e infraestrutura. E quando feito em tempo real ou quase em tempo real é possível identificar, antecipar problemas de negócio e até tomar ações impedindo problemas mais severos. As organizações com esses

conhecimentos possuem vantagens estratégicas sobre seus competidores, em outras palavras, o uso efetivo das informações nos mais variados setores está diferenciando as organizações de seus competidores.

2.2 BIG DATA

No século 21 a economia mundial passou de industrializada à uma economia de serviços. Para apoiar esse novo modelo econômico, observa-se um aumento na oferta de serviços de computação, que incluem infraestrutura flexível e processamento eficiente para suportar todas as demandas. A computação em nuvem tem permitido a viabilização de tais serviços, O número de aplicações e usuários, impulsionados pelo uso de dispositivos móveis e de redes sociais, crescem vertiginosamente a cada dia e a geração de informações e dados também.

Na área de tecnologia da informação o termo BIG DATA surge, atraindo as atenções de governos, indústrias e acadêmicos. O conceito de BIG DATA é definido por 3 Vs, grande volume, velocidade e variedade de informações que precisam de um novo meio de ser processado e analisado. O objetivo da captação, manutenção e processamento de Big Data é encontrar “insights” para novos produtos e/ou serviços, ou para a geração de informação de valor que apoiem decisões estratégicas.

O investimento em BIG DATA teve um aumento enorme nos últimos anos, principalmente na parte de pesquisas e desenvolvimento. Após os investimentos, a valorização da indústria, o surgimento de softwares, tecnologias e empresas especializadas ainda reside uma pergunta: Como criar valor de forma eficiente e efetiva utilizando BIG DATA?

Cada vez mais novos sistemas orientados a serviços surgem e se conectam a outros serviços e plataformas, em estruturas bastante complexas. Geralmente, os dados gerados por esses sistemas são bastante heterogêneos, de vários tipos (imagens, áudio, texto e outros formatos) e altamente dinâmicos.

Um novo “V”, além de velocidade, variedade e volume foi incorporado ao conceito de BIG DATA. Na definição do GARTNER, a veracidade ou variabilidade dos dados é a característica de confiabilidade dos dados gerados, sendo necessário tratar e corrigir as inconsistências para a condução de análises confiáveis dos dados. O estabelecimento da confiança nos dados é um desafio cada vez maior à medida que o número de fontes de dados aumenta e é necessário processá-lo em janelas de tempo específicas e muito pequenas.

2.3 DATA DISCOVERY

Uma necessidade constantemente encontrada nas organizações é a capacidade de gerenciar mudanças de forma eficiente. Este cenário se torna complexo a medida que é necessário rapidez no processo de tomada de decisões com base na rapidez da agregação do conhecimento sobre clientes e mercado. Atualmente, agilidade está associada a habilidade de adotar, reagir e ser bem sucedido no meio de uma variedade de necessidades de negócios que não estão sendo atendidas.

As soluções de BI de fornecedores tradicionais atuais possuem alta escalabilidade e robustez, além de centralizar as informações em um repositório único para que toda a organização tenha a mesma visão das informações, acordando com a arquitetura representada na Figura 1. Em contrapartida, todo desenvolvimento de uma nova informação ou análise diferente precisa passar por um novo ciclo de implementação, que em regra é demorado e é feito por especialistas de TI após o entendimento dos requisitos do negócio.

Muitas vezes o retorno dos investimentos feitos nesse novo desenvolvimento não vale a pena devido a demora que em geral leva o processo, pois cada vez mais as oportunidades devem ser aproveitadas rapidamente para otimizar o capital investido.

Pode-se, portanto, dizer que as ferramentas atuais de BI são inflexíveis ao usuário final de negócios, posto que é necessário uma nova demanda e novo

desenvolvimento para adequar as informações às mudanças exigidas pelo ambiente externo.

Nesse contexto de mudanças frequentes e novas necessidades do negócio é que as ferramentas de Data Discovery foram crescendo e ganhando espaço nos últimos anos, devido a autonomia que oferecem aos usuários finais de negócio para fazer análises customizadas sem a necessidade de envolver especialistas de TI no desenvolvimento.

O termo Data Discovery descreve a utilização de técnicas analíticas visuais dentro da camada de apresentação da plataforma que tem o objetivo de sintetizar grandes volumes de informação em elementos visuais para poderem ser assimilados com maior facilidade e extrair o conhecimento necessário.

Os principais representantes da categoria de Data Discovery, ou também chamados de plataformas ágeis de BI, de acordo com (EVELSON, 2014), podem ser classificados de acordo com algumas características, como segue:

- Self Service BI: apresentam funcionalidades que permitem ao usuário de negócios desenvolver e publicar ele mesmo sua análise com autonomia. Por exemplo: o próprio usuário final de negócios consegue se conectar às fontes de informação oficiais da empresa ou pessoais, como uma tabela de vendas de produtos oficial e uma planilha de metas desenvolvida por ele próprio, e montar uma análise visual baseada nas informações cruzadas e publicar em um portal de informações corporativas.
- Automatização do BI: a plataforma tem a funcionalidade de automatizar processos analíticos diferenciados como descoberta de informações e sugestões de ações a serem tomadas.
- Integração com fontes de informação: As ferramentas oferecem uma interface amigável para o usuário final se conectar às fontes de informações oficiais e também outras fontes desejáveis, como fontes de dados não estruturadas (através de APIs), planilhas pessoais e serviços de dados nas nuvens.

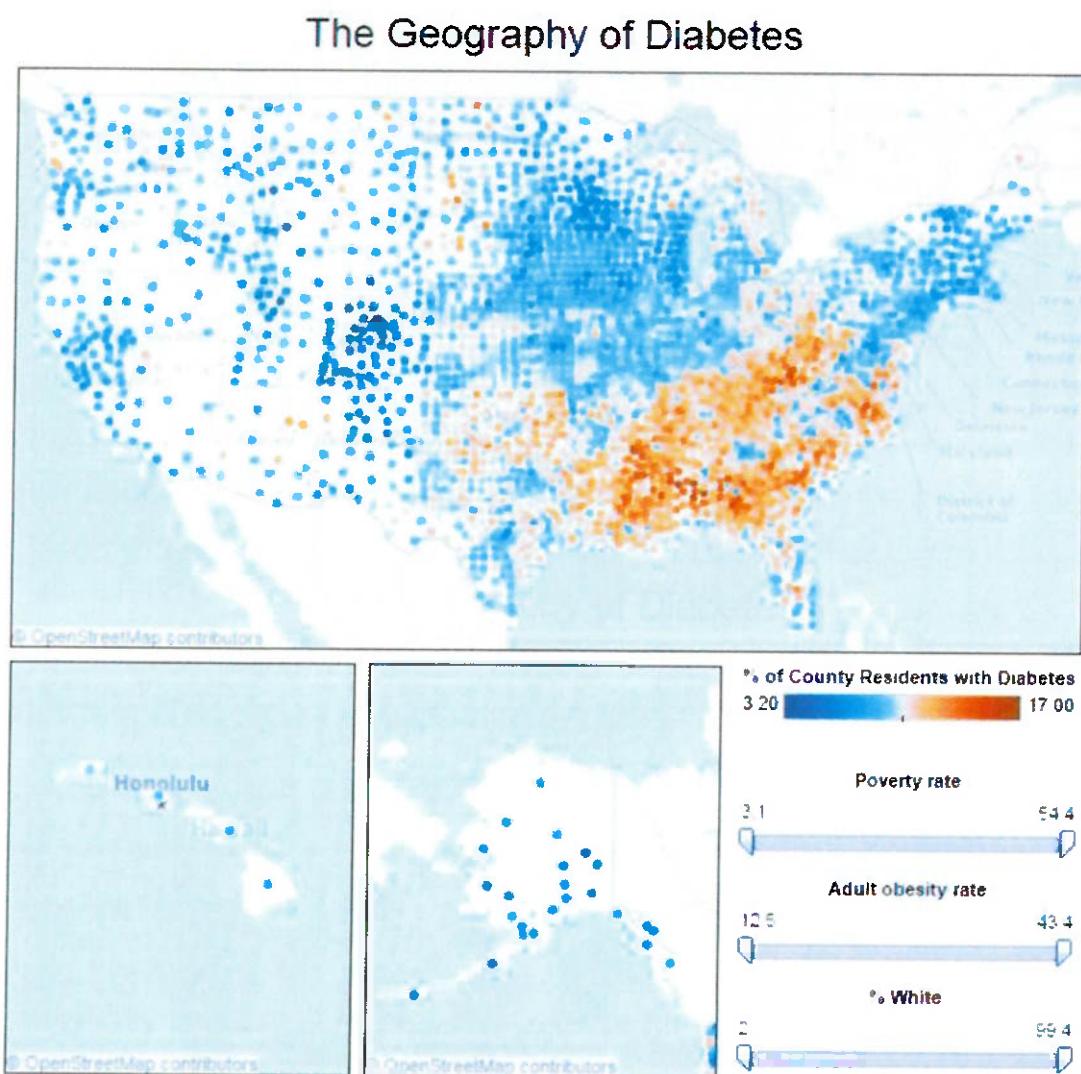
- Tipos de visualizações avançadas: capacidade de criação de diversos tipos de gráficos e visualizações dinâmicas, baseadas em consultas às fontes de informações, sem a necessidade de codificar ou de ter conhecimento específico para a formulação das consultas. A diferenciação entre as ferramentas nesse quesito se apresenta em tipos de visualização, como por exemplo: infográficos, representações baseadas em geolocalização, entre outros. As Figuras 4, 5, 6 e 7 a seguir apresentam quatro recursos de visualização de informação feitas em ferramentas de data Discovery.

Figura 4: Representação gráfica de um painel construído na ferramenta QlikView.



A Figura 4 representa gráficos tradicionais de colunas, pizza, linha, colunas empilhadas e outros. A partir das informações pode-se montar painéis com gráficos interligados, ou seja, ao selecionar ou filtrar parte de um deles acaba-se refletindo nos demais.

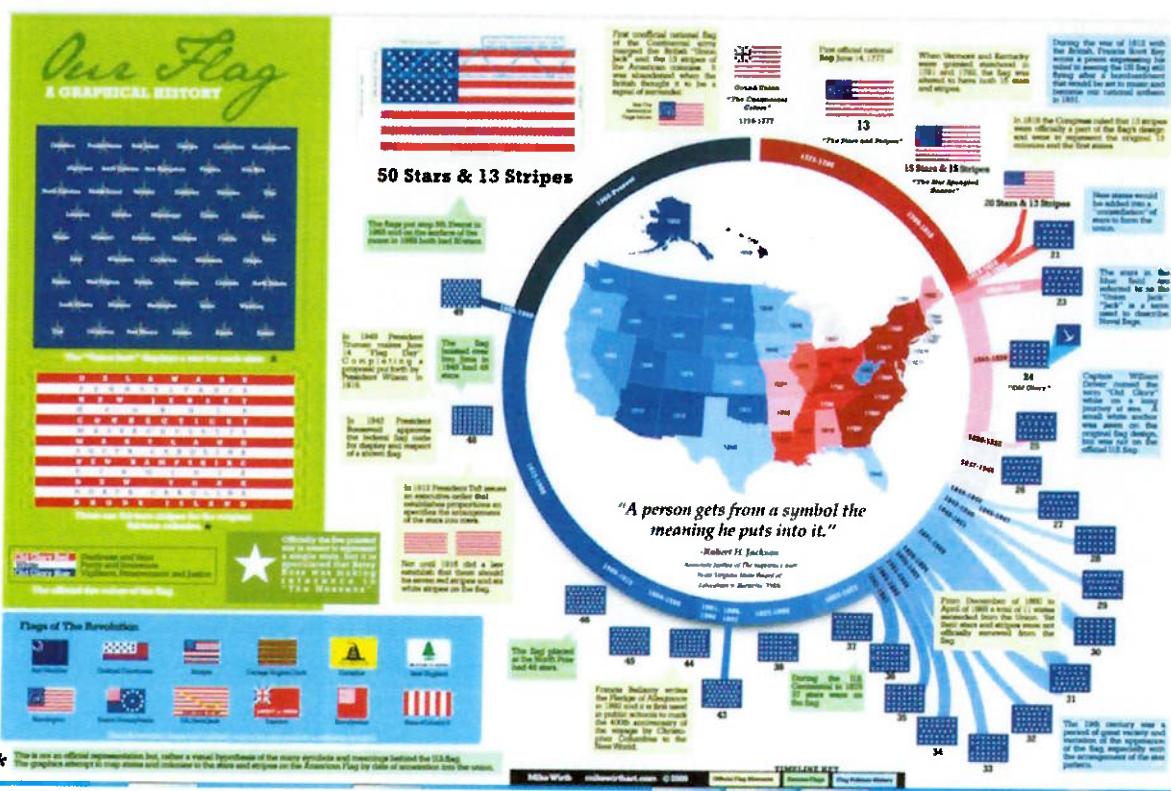
Figura 5: Representação de dados sobre diabetes nos Estados Unidos, construída na ferramenta Tableau.



A figura 5, apresenta as capacidades de utilização de dados de geolocalização dos dados. A partir de informações de cidade, estado, bairro, códigos

postais, latitude e longitude pode-se montar representação espaciais utilizando mapas. No caso em questão, foram utilizadas as informações para apresentar quais cidades possuem maior incidência de diabetes.

Figura 6: Infográfico sobre os estados americanos, construído na ferramenta Spotfire.



A figura 6, mostra outra capacidade gráfica presente nas ferramentas de Data Discovery, a criação de infográficos. Os infográficos são pôsteres interativos que explicam e detalham determinado assunto. Pode-se filtrar e selecionar partes do gráfico para detalhar partes específicas. No caso da Figura 6 está sendo apresentado informações sobre cada um dos estados dos Estados Unidos.

Figura 7: Exemplo de nuvem de palavras.



As nuvens de palavras, exemplificado na Figura 7, é um recurso muito utilizado nos sites atuais e consegue mostrar a relevância de cada palavra de acordo com o tamanho.

2.3.1 BI Tradicional x Data Discovery

Os atuais sistemas de BI possuem uma arquitetura tradicional onde os dados provenientes de diferentes sistemas estruturados de dados, como por exemplo vendas, cadastro, transacional e legados, são estruturados, armazenados e organizados em um repositório de dados. Esse repositório de dados centralizado é a origem única para as ferramentas analíticas, que ao se conectar pode consolidar as informações da maneira mais apropriada para as análises. Esse modelo de dados centralizados garante a integridade das informações e que diferentes setores estão analisando as mesmas informações.

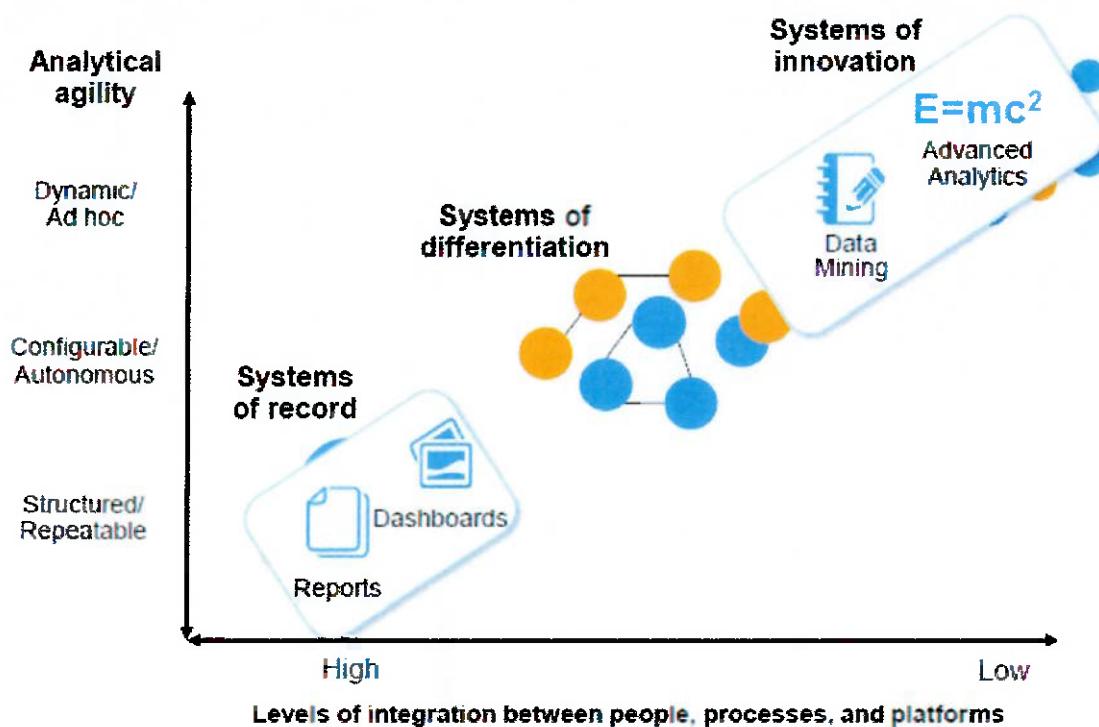
No entanto a velocidade com que o negócio vem sofrendo alterações está aumentando, e cada vez mais a identificação do que aconteceu ou o que está

acontecendo no momento é necessário para evitar crises ou ganhar vantagens estratégicas. O problema é que esse sistema atual não possui uma interface que permita ao usuário de negócio customizar as análises na velocidade desejada. Geralmente o ciclo de desenvolvimento dura algumas semanas para o departamento de TI, e muitas vezes esse tempo é insuficiente para uma reação efetiva embora seus aprendizados possam ser utilizados futuramente.

As soluções de BI tradicionais promovem muito bem a elaboração de relatórios operacionais e dashboards analíticos, desde que sejam bem definidos e estruturados. E também promovem análises estatísticas avançadas permitindo planejamentos estratégicos bastante assertivos ao utilizarem modelos estatísticos implementado nessas soluções.

Existe, todavia, um intervalo entre as análises descritivas e as análises avançadas citadas que as soluções não conseguem suprir adequadamente, como mostra a Figura 5 nos sistemas de diferenciação. No caso de análises pontuais e não planejadas inicialmente é necessário um novo desenvolvimento que possui um custo elevado e tempo de desenvolvimento. Uma opção a um novo desenvolvimento é a extração de dados sob demanda para a utilização de planilhas, porém esse processo assume riscos tanto na extração quanto na elaboração das planilhas.

Figura 8: Representação gráfica dos tipos de análises dos sistemas de BI tradicionais.

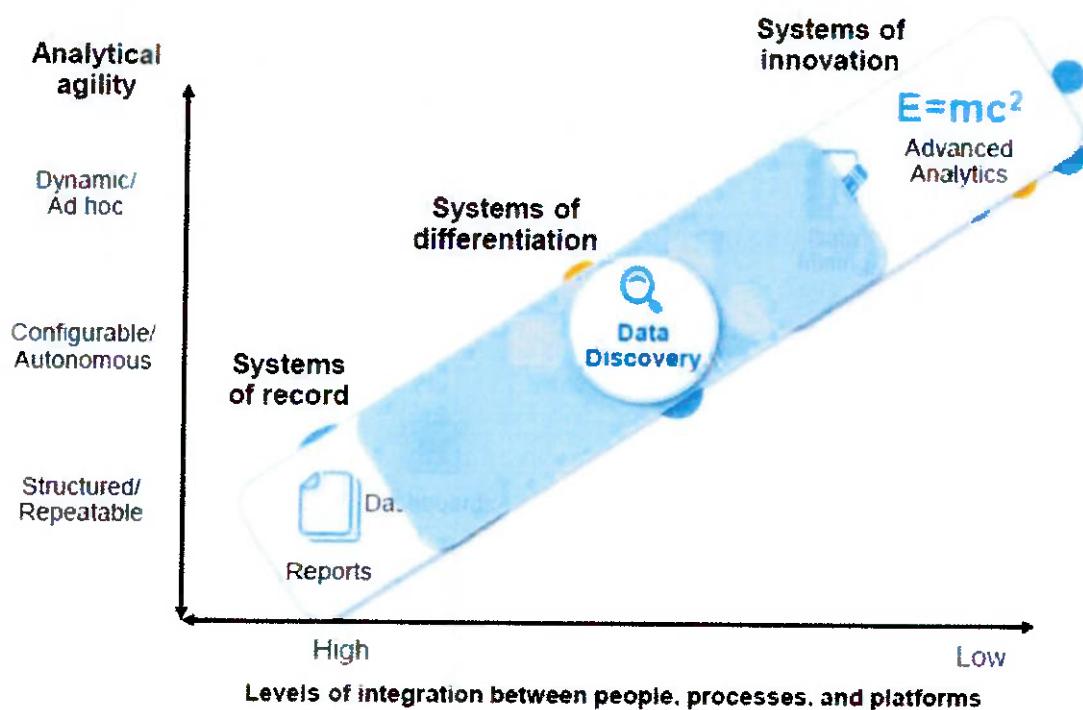


(Fonte: Gartner Advanced Analytics Summit 2014, São Paulo)

As ferramentas de Data Discovery apesar de não possuírem as capacidades analíticas das soluções de BI avançadas com técnicas de mineração de dados e simulações conseguem realizar análises adequadas para extrair rapidamente valor das informações. Esta facilidade é provida por uma interface amigável e voltada para o usuário de negócios, que permite a exploração e a apresentação de dados de forma visual. As ferramentas de Data Discovery não são substitutas das soluções analíticas atuais, mas são complementares e podem conviver no mesmo ambiente.

A Figura 9 representa como a lacuna entre as análises descritivas e as análises avançadas consegue ser suprida ao utilizar-se ferramentas de Data Discovery.

Figura 9: representação gráfica dos tipos de análises dos sistemas de BI tradicionais sobreposto pelas ferramentas de Data Discovery.



(Gartner Advanced Analytics Summit 2014, São Paulo)

2.3.2 Data Discovery e estratégia Big Data

Os acontecimentos no mundo atual dos negócios e mercado provocam a necessidade de reação cada vez mais rápida por parte das organizações, para dessa forma conseguir vantagens estratégicas sobre a concorrência e tomar as melhores decisões. Os investimentos em tecnologias de Big Data vêm crescendo nos últimos anos e faz parte da estratégia de tecnologia da maior parte das corporações.

Caminhando lado a lado com o processamento, o volume e a variedade dos novos dados, existe a necessidade de conseguir informações relevantes para os

negócios e gerar conhecimento e benefícios a partir deles. Assim sendo, para justificar o investimento em uma nova tecnologia de Big Data é necessário comprovar que o retorno do investimento, que não é pequeno, pode ser viabilizado.

De forma complementar às ferramentas de processamento e aquisição de dados não estruturados, as ferramentas de Data Discovery se mostram como opções para a camada de análise e apresentação desses dados, permitindo aos usuários de negócio fazer a exploração desses dados, conseguir tirar análises e disseminar esse conhecimento adquirido no tempo correto para não perder uma oportunidade de negócio.

Portanto, as ferramentas de Data Discovery se introduzidas dentro da estratégia de tecnologia Big Data atuam de forma a melhorar a produtividade dos usuários de negócio e consegue, ainda, auxiliar a viabilizar o investimento necessário.

2.3.3 Mercado e principais fornecedores de ferramentas de Data Discovery

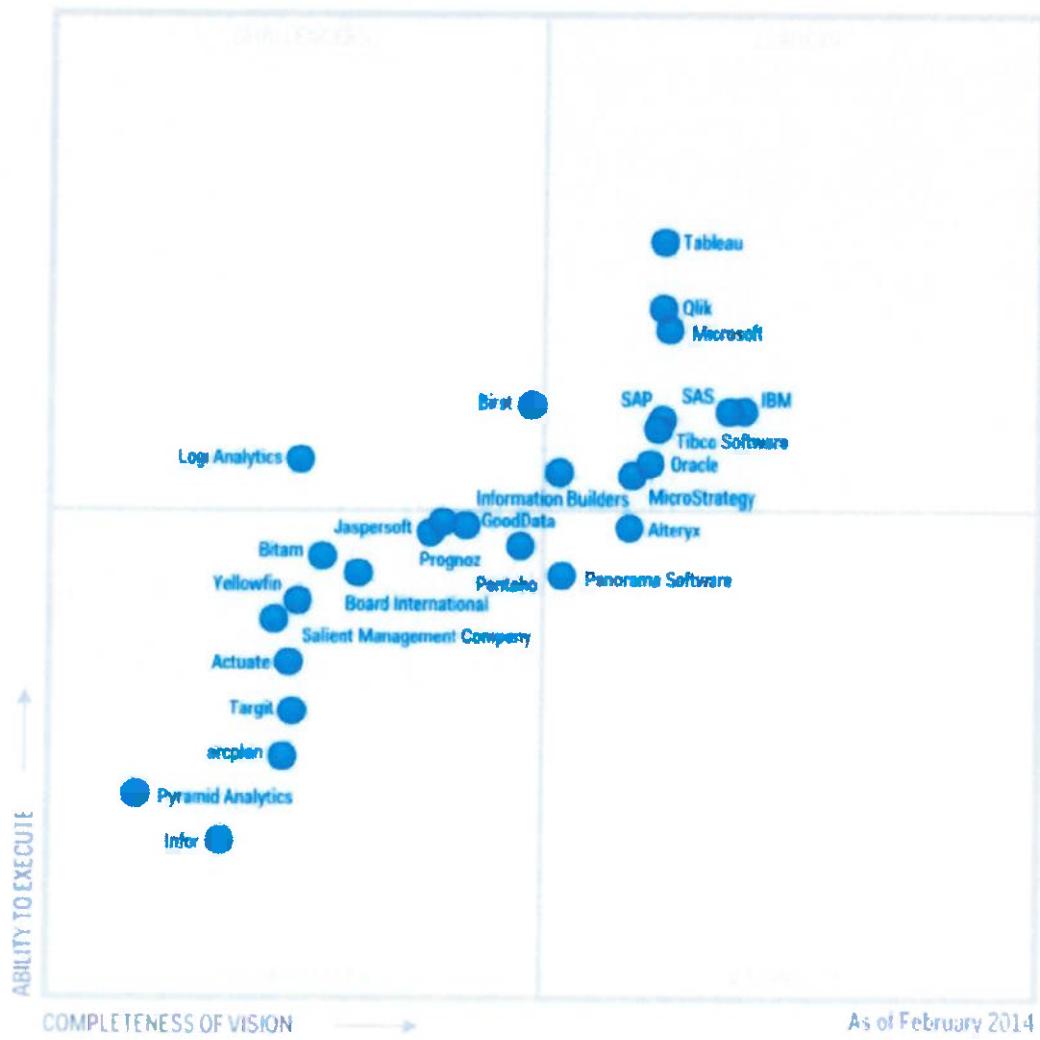
O mercado de BI está em constantemente em modificação e evolução, as demandas dos negócios cada vez mais ágeis provocam a necessidade de ferramentas igualmente ágeis e que não tenham um ciclo de desenvolvimento longo para a geração de informação que se traduzam em vantagens estratégicas sobre os concorrentes, permitindo assim alavancar os negócios mais rapidamente.

O instituto de pesquisas Gartner anualmente divulga uma análise comparada sobre os fornecedores de tecnologia de BI. Considerando as soluções de Data Discovery tal análise é feita considerando um gráfico dividido em 4 quadrantes (Figura 10) composto por: Competidores, Líderes, Fornecedores de Nicho e Visionários. Os principais fornecedores de soluções de ferramentas de Data Discovery são distribuídos nesses quadrantes de acordo com a qualificação resultante da análise. O eixo Y representa a facilidade de utilização da plataforma, em termos da velocidade de aprendizado e utilização da ferramenta no dia a dia. Já o eixo X representa as capacidades analíticas das soluções, ou seja, se é

necessário a utilização de outras ferramentas estatísticas para chegar nos resultados desejados. A divisão em diferentes quadrantes, Figura 10, está associada a classificação resultante da análise comparada segundo o Gatner, como segue:

- Competidores (*Challengers*): as soluções que tem como destaque a facilidade de utilização, porém a parte analítica ainda tem baixa aceitação/uso no mercado quando comparadas a outras concorrentes.
- Líderes (*Leaders*): as soluções agregam a facilidade de uso e a capacidade analítica. Nesta categoria tem-se novas soluções como QLIK, TABLEAU e TIBCO, além das tradicionais como SAS, Microsoft, IBM, Oracle e SAP.
- Fornecedores de Nicho (*Niche Players*): as soluções são voltadas para segmentos específicos. Em geral, tem limitações quanto a capacidade analítica. Exemplos são soluções de planejamento de demandas logísticas e soluções de cálculos de risco de crédito.
- Visionários (*Visionaries*): soluções que possuem diferenciais analíticos em relação às outras, porém não são vistas como ferramentas simples de operar.

Figura 10: Representação dos quadrantes mágicos do Gartner para ferramentas de DataDiscovery.



(Fonte: Gartner Advanced Analytics Summit 2014, São Paulo)

2.4 TRABALHOS CORRELATOS

Nesta seção discorre-se sobre alguns artigos relacionados ao tema deste trabalho, os quais foram selecionados por apresentarem dados diretamente relacionados às ferramentas de análise de dados.

(WEIKUN, 2013) apresenta um panorama geral sobre a situação das ferramentas de Data Discovery e o contexto de suas origens e atual relevância. Também mostra como as ferramentas de exploração de dados visuais serão úteis na sociedade e nos negócios em um prazo de dez anos, e os desafios e oportunidades atuais para que a previsão se torne realidade.

(KAMBATLA, KOLLIAS, KUMMAR, GRAMA, 2014) apresenta uma previsão de evolução de sistemas analíticos de big data para suportar a quantidade de dados que serão gerados. Segundo os autores, a infraestrutura das plataformas atuais sofrerá mudanças tanto nos aspectos de processamento e memória quanto em redes e sustentabilidade, espera-se o aumento de desempenho de processamento e maior utilização de tecnologia de virtualização. O artigo mostra ainda um panorama de como a tecnologia de soluções analíticas de big data podem e serão aplicadas nas mais diversas áreas da sociedade e negócios.

(PHILIP CHEN, ZHANG, 2014) apresenta uma visão detalhada sobre Big Data, desde aplicações até dificuldades e como lidar com problemas utilizando metodologias que foram bem sucedidas. Inicialmente o artigo contextualiza o leitor e mostra um breve histórico do surgimento e os problemas que só puderam ser solucionados através dessa tecnologia. Logo a abordagem tem foco maior nos desafios de arquitetura das soluções como um todo. A apresentação das ferramentas e técnicas para análise dos dados adquiridos e apresentação das informações é a seção mais importante para o presente trabalho.

(EBNER, BÜHNEN, URBACH, 2014) apresenta as diferentes estratégias para Big Data que é adotada no meio corporativo atual. Apresenta um procedimento para auxiliar na identificação de qual dessas estratégias é a mais adequada para cada tipo de necessidade. Ao levar em conta os custos, o tipo de análise e os profissionais existentes nas empresas pode-se indicar a melhor estratégia de big data para ser promovida.

(AIGNER, 2013) tem como objetivo confirmar algumas hipóteses sobre a análise visual dos dados. Para isso, foram entrevistadas pessoas relevantes na área

analítica, de acordo com a análise dos resultados, as hipóteses foram ou não confirmadas. Questionamento sobre a importância da interatividade dos usuários com os gráficos em *dashboards* para os usuários finais e a relevância da representação visual dos dados para encontrar *insights* são respondidos pela pesquisa.

(EVELSON, 2014) é uma publicação do instituto de pesquisa FORRESTER sobre as plataformas de BI ágeis, apresenta comparações entre os principais fornecedores de tecnologia. O artigo ressalta a necessidade das empresas investirem nesse tipo de tecnologia e os possíveis ganhos estratégicos que podem conseguir utilizando os recursos analíticos das ferramentas. Ainda apresenta um perfil de cada fornecedor analisado e a situação de cada um deles no cenário atual.

(COTA, CASTRO, DOMINGUEZ, 2013) ressalta a importância da utilização de ferramentas visuais para apoiar as tomadas de decisões nas empresas. Apresenta também as novas ferramentas de visualização de dados como um diferencial para as decisões estratégicas e como isso deve ser praticado nas corporações para virar parte da cultura da empresa e parte do processo natural dos sistemas de suporte a decisões.

(BAI, WHITE, SUNDARAM, 2013) informa que os sistemas de BI atuais já estão integrados às ferramentas de visualização de informação, as quais permitem extrair as informações necessárias para gerar conhecimento. Porém, destaca que a parte visual ainda precisa ser melhorada. Assim sendo, propõe uma forma para implementar esse tipo de solução de forma que as corporações aproveitem a totalidade das funcionalidades dessas ferramentas em um contexto adaptativo.

(SACHA, STOFFEL, STOFFEL, KWON, 2014) é um dos artigos mais interessantes estudados para o presente trabalho. O artigo apresenta diversas metodologias para extrair conhecimento dos dados através das ferramentas visuais. Além de apresentar os processos atuais para extrair valor dos dados, apresenta também como deve ser o ciclo de aprendizado e aplicação dos conhecimentos

adquiridos. Mostra ainda como a interação com as ferramentas visuais podem acelerar esse processo e ainda criar outros métodos de aprendizados.

3. EXPERIMENTO

Como já comentado no capítulo 1, o estudo é feito no contexto de uma organização que atua no mercado de cosméticos. No último ano identificou-se a necessidade de melhorar as capacidades analíticas para extrair e explorar os dados da empresa de uma forma mais rápida e assertiva para a geração de conhecimento e tomada de decisões.

Foi realizado, portanto, um experimento para testar algumas ferramentas de Data Discovery com o objetivo de dar mais autonomia para as áreas de negócio fazerem análises com mais agilidade e segurança nos resultados.

Primeiramente foram realizados estudos em diversas áreas de negócio para identificar os possíveis cenários de testes para a realização do experimento. Como parte do experimento também avaliou-se quais seriam os impactos financeiros e de aprendizados para o negócio.

O interesse das áreas foi imediato, devido principalmente à vantagem da obtenção mais rápida da informação e ao impacto visual que as ferramentas causam.

A avaliação do impacto financeiro foi determinante para a escolha da área que foi escolhida como caso de uso para o experimento. Para tal avaliação considerou-se o valor do investimento no produto contra os benefícios obtidos com o mesmo, de forma a justificar ou não a compra do produto. Assim, a área selecionada para ser utilizada no experimento foi definida com base no valor comprovado que os aprendizados gerados pela nova solução, ou produto, poderiam trazer. Dessa forma, o caso de uso utilizado teria condições de ser utilizado como porta de entrada da ferramenta na empresa, uma vez que o retorno do investimento já estaria comprovado.

O projeto piloto (experimento escolhido) foi um projeto de automatização de sete análises gráficas relacionadas às vendas de produtos por categoria, bastante recorrentes na área comercial.

Atualmente, cada estudo realizado pela área comercial consome uma parcela grande do tempo dos profissionais da empresa. Tais profissionais preparam o ambiente para a extração e consolidação das informações para serem analisadas, gerando conhecimento que pode ser aplicado futuramente e traduzido em receita.

Dessa forma, a automatização das análises mais frequente nos estudos impacta a velocidade, qualidade e profundidade dos estudos que são realizados. Ainda, essas análises em conjunto conseguem derrubar hipóteses e prevenir que estudos sem impactos sejam realizados, enquanto outras podem antecipar resultados e tomadas de decisões estratégicas, minimizando prejuízos.

Graças à automatização das análises, os usuários de negócio não precisarão mais ficar reféns do departamento de TI (tecnologia da informação) da empresa, e podem ter maior produtividade ao gastar menos tempo na preparação e mais tempo analisando as informações. O departamento de TI por sua vez terá menos demandas e pode despender o tempo em outras atividades, como melhorias nos sistemas e qualidade de dados.

Uma vez definido o caso de uso para os testes, foram escolhidas as ferramentas de Data Discovery para serem avaliadas. Os fornecedores TABLEAU, QLIK e TIBCO, estão apresentados na Figura 7 (quadrante superior direito), e são reconhecidos no mercado, pelo Gartner, como os líderes nas soluções de Data Discovery. Por esta razão foram selecionadas para o experimento.

Outro aspecto considerado para a escolha das ferramentas foi a usabilidade e a capacidade estatística observadas. A Tabela 1 mostra os produtos selecionados para avaliação no projeto piloto.

Tabela 1 - Especificação dos produtos de Data Discovery selecionados para avaliação

Produto	Empresa fabricante
TABLEAU Versão 8.3	TABLEAU Software
qlik sense Versão 1.0	qlik
SPOTFIRE Versão 6.5	TIBCO

Os produtos apresentados na Tabela 1 possuem propósitos similares relativamente a capacidades gráficas, análise e publicação de painéis para os usuários finais.

O processo de avaliação de cada produto teve a duração de aproximadamente 3 semanas. A área de negócios envolvida no caso de uso escolhido teve participação na avaliação das capacidades analíticas das soluções, identificado os pontos fracos e forte de cada ferramenta e ajudando na pontuação para cada quesito.

3.1 Definição dos Critérios para a Avaliação das Ferramentas

O processo de definição dos critérios para análise das ferramentas é um tópico fundamental deste trabalho. Assim buscou-se suporte em artigos científicos relacionados ao assunto (EVELSON, 2014), TRUSTRADIUS, (COTA, et al, 2014).

Por meio dessa revisão foi possível formular os seguintes critérios, que foram divididos em 4 grupos principais:

- 1. Funcionalidades:** o grupo de funcionalidades tem o objetivo de analisar as capacidades funcionais das soluções. E foi dividido nos subgrupos a seguir.

- a. **Acesso:** avalia a facilidade e qualidade do acesso à interface de desenvolvimento e visualização da ferramenta.
 - b. **Conectividade:** avalia a variedade de conectores, ou, a capacidade de a ferramenta conectar-se, de forma nativa, a diferentes fontes de dados tanto para dados estruturado quanto não-estruturados.
 - c. **Desenvolvimento de análises:** avalia funcionalmente as ferramentas quanto ao processo de desenvolvimento de análises.
 - d. **Integração:** avalia como a ferramenta se integra com os outros ambientes e ferramentas existentes na corporação
2. **Exploração:** o grupo de exploração é voltado para a área de negócios e tem o objetivo de avaliar as capacidades analíticas de cada solução e ainda avaliar a visão do usuário final sobre a qualidade gráfica e usabilidade ferramenta. Foi divido nos subgrupos a seguir.
- a. **Visualizações:** avalia a quantidade de tipos de gráficos e visualizações disponíveis nas ferramentas. Capacidade de produzir infográficos, efeito de animação temporal, *storytelling*, nuvens de palavras e utilização de mapas e geolocalização são exemplos de tipos de gráficos avaliados, exemplificados na Fig. 4.
 - b. **Funcionalidades Analíticas:** avalia a capacidade do analista de negócios customizar a análise desejada sem a necessidade de utilizar outra ferramenta para isso. Por exemplo, capacidade de tratamento de exceções, criação de fórmulas e indicadores customizados, criação de hierarquias para fazer “drill-downs” e outras.
 - c. **Criação de Dashboards:** avalia a capacidade de criar painéis interativos para divulgar os resultados e análises. Facilidade de desenvolvimento e publicação pelo usuário final e a capacidade de interação com o painel são exemplos de pontos avaliados.

- d. **Análises automáticas e descobertas de padrões:** avalia a capacidade de deixar um processo de análise automatizado para encontrar padrões de forma contínua.
 - e. **Dispositivos móveis:** avalia como as ferramentas se adaptam nos dispositivos móveis, como tablets e smartphones.
 - f. **Desempenho e aprendizado:** avalia a qualidade de desempenho das ferramentas como o tempo despendido para a criação da análise, desde a definição até a publicação, além da curva de aprendizado da ferramenta.
3. **Colaboração:** o grupo de colaboração tem como foco a capacidades de disseminação do conhecimento através da plataforma das ferramentas. Foi dividido nos subgrupos a seguir.
- a. **Colaboração:** avalia a capacidade de mais de um usuário ajudar na construção de uma análise ou nos painéis. Por exemplo fóruns, anotações, comentários, alertas e notificações.
 - b. **Recomendações:** avalia se a ferramenta consegue sugerir outras análises ou criar listas de painéis favoritos.
4. **Administração:** O grupo de administração tem como objetivo avaliar a capacidade e desempenho das ferramentas relativamente à administração do ambiente. Avalia-se ainda a facilidade da implementação da ferramenta, além da periodicidade de atualizações e facilidade de aplicar as mesmas. Foi dividido nos subgrupos a seguir.
- a. **Segurança e Administração:** avalia os pontos relacionados à segurança, autenticação de usuários, atribuição de perfis e restrição de acessos às informações confidenciais. Também avalia a qualidade da plataforma de administração da ferramenta.

- b. **Instalação e implementação:** avalia a facilidade de instalação da plataforma, incluindo a aplicação de possíveis atualizações disponibilizadas. A existência de suporte e a qualidade deste serviço.
- c. **Suprimento de idiomas:** avalia a capacidade de incluir novos idiomas e caracteres dos dados.
- d. **Maturidade e uso no mercado:** avalia o posicionamento dos fornecedores no mercado pelo Gartner, maturidade da ferramenta (tempo de mercado e clientes), quantidade de fornecedores e representantes oficiais no Brasil, além da procura de profissionais experientes na utilização das ferramentas avaliadas pelo mercado.

Para a avaliação definiu-se que as 3 ferramentas serão avaliadas por 2 quesitos: um grau de importância e um valor. Assim, considerando o projeto piloto, cada um dos critérios foi avaliado relativamente à importância do critério e à um valor que mede a adequação ou não pela ferramenta. A Tabela 2 mostra como os critérios foram valorados no processo de avaliação das ferramentas de Data Discovery.

Tabela 2 - Avaliação das ferramentas

Grau de importância do critério	Valor da adequação da ferramenta
0 – Sem Importância	0 – Não Disponível
1 – Desejável, mas não fundamental (nice to have)	1 – Atende minimamente/ parcialmente os requisitos
2 – Necessário	2 – Adequado
3 – Importante	3 – Excepcional ou excede as expectativas
4 – Crucial	

A nota para cada critério variou de zero a três e esse valor foi multiplicado pelo grau de importância, com variação de zero a quatro, para o cálculo do valor final. Cada subgrupo, composto por diversos critérios, teve um valor final atribuído consolidado a partir da somatória resultante. Ainda, para cada subgrupo foi atribuído um peso, que variava de zero a seis, de acordo com a relevância de cada subgrupo no entendimento das áreas envolvidas.

Os subgrupos considerados de maior relevância, com valor igual a seis, foram: Conectividade e Desenvolvimento, pertencentes ao grupo de Funcionalidades; Desempenho e Aprendizado pertencentes ao grupo de Exploração; e Instalação e Implementação, Maturidade e Uso no Mercado ligados ao grupo Administração.

Dessa forma foi possível obter uma tabela de resultados consolidada por subgrupos e grupos. A tabela de resultados foi preenchida com valores percentuais em relação ao maior valor possível que cada subgrupo poderia atingir, ou seja, se cada critério dentro de cada subgrupo obtivesse nota três.

O cálculo de uma nota final para cada ferramenta foi definido da seguinte forma: as notas de cada subgrupo foram multiplicadas pelo valor da relevância associado a cada um, somados e ponderados da mesma forma para chegar a um valor percentual relativo à maior nota possível.

A fim de consolidar os resultados, foi atribuído um peso a somatória dos pontos obtidos de cada grupo. O peso variou de zero a seis, de acordo com a relevância de cada grupo. Quanto maior o peso, maior a importância do grupo.

A Figura 11 representa um exemplo de como o trabalho anteriormente descrito – de atribuição de notas – foi realizado, para cada critério dentro do subgrupo de Visualizações.

Figura 11: Exemplificação da atribuição de notas para aos critérios pré-definidos.

click for instructions		Criteria	Importance	Sisoffie		Tableau		QlikSense	
Visualizações				Availability	Score	Availability	Score	Availability	Score
Infográficos	Capacidade de produzir infográficos e apresentações visuais de dados InfoGraphic -Visual Presentation of data & insights	2-Needed	2-Proper	4	2-Proper	4	2-Proper	4	2-Proper
	Efeitos de animação temporal			1-Need to Have	1-Minimal	1	2-Proper	2	0-Not Available
	Capacidade de armazenar e utilizar um banco de imagens para os infográficos.			1-Need to Have	1-Minimal	1	1-Minimal	1	1-Minimal
Gráficos	Formatação do tipo do Gráfico e customização dos mesmos	3-Important	2-Proper	6	2-Proper	6	1-Minimal	3	3-Important
	BOXPLOT			3-Important	2-Proper	6	3-Exceptional	9	2-Proper
	Mapas e Geolocalização			3-Important	2-Proper	6	2-Proper	6	2-Proper
	Drill Down			3-Important	2-Proper	6	2-Proper	6	3-Exceptional
	Filtros e métricas customizáveis			3-Important	2-Proper	6	2-Proper	6	2-Proper
	Nuvem de palavras			3-Important	2-Proper	6	2-Proper	6	2-Proper
KPIs	Velocímetros, faróis de trânsito e outros tipos de gráficos de indicadores Visual animated gauges, traffic lights and more	2-Needed	3-Exceptional	6	0-Not Available	0	1-Minimal	2	0-Not Available
	Customização dos gráficos e cores			1-Need to Have	2-Proper	2	2-Proper	2	2-Proper
	Customizable graphics and colors			1-Need to Have	2-Proper	2	2-Proper	2	2-Proper
Outras Visualizações	StoryTelling	3-Important	2-Proper	6	2-Proper	6	2-Proper	6	2-Proper
		0-Not Available	0-Minimal	0	1-Minimal	0	1-Minimal	0	1-Minimal

(Fonte: Elaborado pelo autor).

3.2 ANÁLISE COMPARATIVA E RESULTADOS

A Tabela 3 ilustra para cada uma das 3 ferramentas de Data Discovery, a pontuação resultante, considerando os critérios de análises e que foram apresentados na seção 3.1 do presente trabalho. A pontuação dos candidatos no consolidado de cada subgrupo é ponderado com o máximo valor de avaliação calculado a partir da importância e dos pesos dos critérios (Tabela 2).

Em verde está destacado qual o fornecedor obteve melhor rendimento nos subgrupos escolhidos, e ao final o total ponderado. Em uma análise visual rápida, pode-se perceber que o TABLEAU obteve maior rendimento na maior parte dos quesitos, isso se refletiu na avaliação final, onde obteve 64% dos pontos possíveis, enquanto o QLIKSENSE e o SPOTFIRE obtiveram 55% e 52% respectivamente.

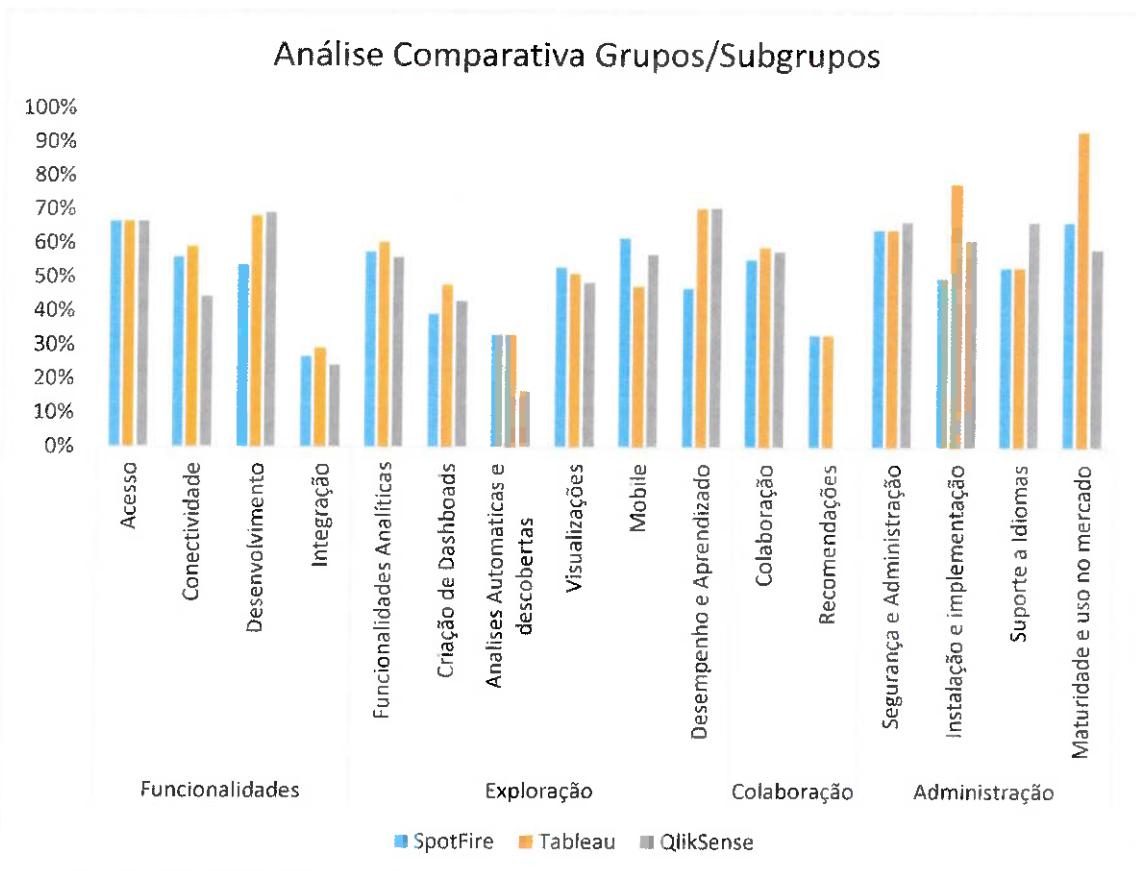
Tabela 3 - Pontuação das ferramentas de Data Discovery testadas.

Grupo	SubGrupo	SpotFire	Tableau	QlikSense	Pesos dos Critérios
Funcionalidades	Acesso	67%	67%	67%	1
	Conectividade	56%	59%	44%	6
	Desenvolvimento	54%	68%	69%	6
	Integração	27%	29%	24%	3
Exploração	Funcionalidades Analíticas	58%	61%	56%	3
	Criação de Dashboards	40%	48%	43%	3
	Analises Automáticas e descobertas	33%	33%	17%	1
	Visualizações	53%	51%	49%	3
	Mobile	62%	48%	57%	1
	Desempenho e Aprendizado	47%	71%	71%	6
Colaboração	Colaboração	56%	59%	58%	1
	Recomendações	33%	33%	0%	1
Administração	Segurança e Administração	64%	64%	67%	2
	Instalação e implementação	50%	78%	61%	6
	Supporte a Idiomas	53%	53%	67%	1
	Maturidade e uso no mercado	67%	94%	59%	6
TOTAL PONDERADO		52%	64%	55%	

Apesar da melhor nota geral e a análise visual, na grande maioria dos critérios a nota da ferramenta TABLEAU não foi muito superior às dos concorrentes, ou seja, a diferença ocorreu pelo destaque em subgrupos com pesos muito relevantes, como Maturidade e uso no mercado e instalação e implementação.

O Gráfico 4, mostra em barras as diferenças entre os subgrupos e pode-se comprovar que a disparidade não é tão evidente quanto aparenta na tabela 1, somente nos subgrupos de instalação e implementação e maturidade e uso no mercado podemos visualmente destacar a solução do TABLEAU. Esses dois subgrupos de destaque possuem o peso máximo nos critérios de avaliação e por conta disso a disparidade no total ponderado.

Gráfico 1 - Gráfico de pontuação das ferramentas analisadas.



O grupo de funcionalidades foi avaliado durante todo o experimento, e o SPOTFIRE levou uma leve desvantagem em relação aos outros dois concorrentes no subgrupo de desenvolvimento, enquanto o QLIKSENSE não conseguiu estabelecer as conexões com os bancos de dados escolhidos com a mesma facilidade dos outros dois.

O grupo de exploração foi avaliado com a ajuda da área de negócios e o resultado foi bastante interessante. A ferramenta da TIBCO se mostrou a mais completa em relação as capacidades analíticas, com características de uma ferramenta estatística e com a maior capacidade de customização entre as três, mas também com a interface menos amigável para o usuário final construir os painéis.

O grupo de administração foi onde a ferramenta TABLEAU se destacou dos concorrentes, A pontuação relacionada a instalação e a implementação da solução foi bastante superior aos demais, assim como a maturidade da solução no mercado.

4. CONCLUSÕES E CONSIDERAÇÕES FINAIS

As ferramentas de Data Discovery, de acordo com as referências analisadas, se mostram como uma tendência, posto que fornecedores tradicionais agora estão correndo atrás para reconquistar a liderança do mercado nesse tipo de solução.

Essas ferramentas inovadoras trazem muitos benefícios para os negócios, pode-se destacar a redução do tempo de atendimentos das demandas pela área de TI. Ao oferecer ao usuário final a capacidade de exploração dos dados, desenvolvimento e publicação de análises, promove um aumento na produtividade da organização, ao encontrar oportunidades e gerar conhecimento na velocidade adequada para aproveitá-las. Dessa forma o papel do departamento de tecnologia da informação fica mais concentrado na governança dos dados do que na produção de informações para os negócios.

A análise comparativa feita entre 3 fornecedores, dentro de um cenário analítico específico, mostrou que o TABLEAU é a ferramenta que mais se adequou às necessidades do negócio analisado, porém deve-se ressaltar que a disparidade entre os concorrentes se deve principalmente a dois quesitos com pesos muito relevantes. O primeiro deles relacionado a maturidade da solução no mercado e o segundo relacionado a facilidade na instalação e implementação da solução.

Vale ressaltar que a ferramenta SPOTFIRE se mostrou muito melhor que dos concorrentes com relação as capacidades analíticas, conseguindo fazer simulações de modelos estatísticos na própria ferramenta. Se o caso de uso escolhido fosse voltado para uma área de negócios com foco em simulações, certamente seria a mais adequada.

A ferramenta QLIKSENSE apresentou capacidades de customização e desempenho excelentes. O fato da ferramenta estar na versão inicial e não possuir tantos clientes no mercado foi determinante na nota de maturidade da ferramenta, e isso culminou na escolha da outra ferramenta.

É possível apontar uma fragilidade da pesquisa realizada no trabalho, que consiste no reduzido número de pessoas consultadas para a atribuição das notas aos quesitos especificados. Assim sendo, encoraja-se a continuação dos estudos no sentido de ampliar a quantia de pessoas envolvidas na atribuição de tais notas a fim de conferir maior credibilidade à análise.

Outro ponto que oferece campo para trabalhos futuros é a introdução da análise de negócio ao estudo técnico já realizado. A utilização de dados econômicos e financeiros, intimamente ligados à cada ferramenta, pode representar uma condicionante de peso, influenciando no resultado final.

REFERENCIAS

- WEIKUN, G.; 2013. **Data Discovery**, Data Science Journal, July. Volume 12.
- KAMBATLA, K.; KOLLIAS, G.; KUMMAR, V.; GRAMA, A.; 2014. **Trends in big data analytics**. Journal Parallel Distrib. Comput. 74 2561–2573 [2]
- PHILIP CHEN, C. L.; ZHANG, C.; 2014. **Data-intensive applications, challenges, techniques and technologies: A survey on Big Data**, Information Sciences, n 275, pp. 314–347. [3]
- EBNER, K.; BÜHNEN, T.; URBACH, N.; 2014. **Think Big with Big Data: Identifying Suitable Big Data Strategies in Corporate Environments**. Proceedings of 47th Hawaii International Conference on System Science. [4]
- AIGNER,W.;2013. **Current Work Practice and User's Perspectives on Visualization and Interactivity in Business Intelligence**. Proceedings of 17th International Conference on Information Visualization [5]
- EVELSON, B.; 2014. **The Forrester Wave: Agile Business Intelligence Platforms, Q3 2014**. For Application Development & Delivery Professionals [6]
- COTA, M.; CASTRO, M.; DOMINGUEZ, J.; 2014. **Importance of visualization usage in enterprise decision-making environments**. I Conference on nformation Systems and Technologies (CISTI), 2014 9th Iberian. 18-21 June 2014, pp. 1-7.
DOI: [10.1109/CISTI.2014.6876973](https://doi.org/10.1109/CISTI.2014.6876973) [7]
- BAI, X.; WHITE, D.; SUNDARAM, D.; 2013. **Context Adaptative Visualization for Effective Business Intelligence**. Proceedings of ICCT2013. [8]
- SACHA, D.; STOFFEL, A.; STOFFEL, F. KWON, B. C.; 2014. **Knowledge Generation Model for Visual Analytics**. IEEE Transactions on visualization and computer graphics, Vol. 20. NO. 12. December 2014.

SHARDA, R.; ASAMOAH, D. A.; PONNA, N.; 2013. **Business Analytics: Research and Teaching Perspectives**. Proceedings of 35th International Conference on Information Technology Interfaces (ITI).

CHUAH, M.; 2010. An enterprise business intelligence maturity model (EBIMM): Conceptual framework. Proceedings of Fifth International Conference on Digital Information Management (ICDIM)

MOHAMADINA, A. A.; 2012. **Business Intelligence: Concepts, Issues and Current System**. Proceedings of International Conference on Advanced Computer Science Applications and Technologies.

HANS, R. T.; 2013. **Modeling Software Engineering Projects as a Business: A Business Intelligence Perspective**. AFRICON

TIEN, J. M.; 2013 **Big Data: Unleashing Information**. J Syst Sci Syst Eng. 22(2): 127-151

SOMMER, D.; 2013. **Market Trends: The Collision of Data Discovery and Business Intelligence will cause Destruction**. Gartner (G00229826)

TRUSTRADIUS (<https://www.trustradius.com/compare-products/tableau-desktop-vs-tibco-spotfire-vs-qlikview>) acessada em 15/10/2014