

APLICAÇÃO DE BUSINESS INTELLIGENCE NOS DADOS DE DIAGNÓSTICO DE UNIDADES FAMILIARES DE PRODUÇÃO NA PARAÍBA: UM ESTUDO DE CASO NA EMATER-PB

APPLICATION OF BUSINESS INTELLIGENCE IN THE DIAGNOSTIC DATA OF FAMILY PRODUCTION UNITS IN PARAÍBA: A CASE STUDY IN EMATER-PB

Nivaldo Mariano de Carvalho Junior¹
Thyago Maia Tavares de Farias²

RESUMO

O objetivo deste artigo é apresentar um estudo de caso no uso de técnicas de *Business Intelligence* (BI) para análise e mapeamento de bases de dados de diagnóstico de chamada pública realizados entre os anos de 2011 e 2017 pela Empresa de Assistência Técnica e Extensão Rural da Paraíba (EMATER-PB), com ênfase na ocupação de solo e indicadores sociais em Unidades Familiares de Produção (UFP), para que, através desse, possa-se obter um sistema de apoio à tomada de decisão em favor do planejamento de ações em ATER (Assistência Técnica e Extensão Rural) por parte da empresa em questão e no auxílio à definição de políticas públicas relacionadas à agricultura familiar por parte do Governo do Estado da Paraíba.

Palavras-chave: *Business Intelligence*. Agricultura Familiar. EMATER. Unidades de Produção Familiar. UFP.

ABSTRACT

The aim of this article is to present a case study on the use of Business Intelligence (BI), for analysis and mapping of public call databases, made between the years 2011 and 2017 by Empresa de Assistência Técnica e Extensão Rural da Paraíba (EMATER-PB), with emphasis on soil occupation and social indicators in Family Production Units (FPU), through this, a system can be created to support the decision-making in favor of the planning of actions in ATER by said company and aid in the definition of public policies related to Family Farming by the State Government of Paraíba.

¹ Graduado em Gestão de Tecnologia da Informação pelo Centro Universitário de João Pessoa (UNIPÊ). E-mail: nivaldo@vince.as

² Mestre em Informática pela Universidade Federal da Paraíba (UFPB). Docente do Centro Universitário de João Pessoa (UNIPÊ). E-mail: thyagomaia@gmail.com

Keywords: Business Intelligence. Familiar Agriculture. EMATER. Production Family Units. UFP.

INTRODUÇÃO

A EMATER-PB é uma Empresa de Assistência Técnica e Extensão Rural vinculada à Secretaria da Agricultura e Abastecimento do Estado, integrante do Sistema Brasileiro de Assistência Técnica e Extensão Rural – SIBRATER, criada pelo Decreto Estadual número 6.755, de 18 de dezembro de 1975, assinado pelo Governador Ivan Bichara Sobreira (GESTÃO UNIFICADA, 2017).

A empresa tem como objetivos básicos, segundo a lei que a criou, de colaborar com os órgãos competentes da Secretaria de Agricultura e Abastecimento do Estado da Paraíba e do Ministério da Agricultura do Governo Federal. O referido órgão formula e executa políticas de Assistência Técnica e Extensão Rural no Estado da Paraíba, visando a difusão de conhecimentos de natureza técnica, econômica e social, para aumento de produção e melhoria de condições de vida no meio rural (GESTÃO UNIFICADA, 2017).

O SIGATER (Sistema de Gerenciamento de ATER) (COSTA et al., 2011) é um sistema desenvolvido pela EMATER-PB no qual tem por objetivo o gerenciamento de informações e atividades de Assistência Técnica e Extensão Rural realizadas pelos técnicos da empresa na Paraíba.

Atualmente, o referido sistema conta apenas com módulos de inserção, edição, seleção e remoção de dados relativos a indicadores sociais e ocupação de solo relacionados à Unidades Familiares de Produção (UFP) na Paraíba. O SIGATER não conta com módulos de relatórios analíticos que permitam a filtragem e a manipulação de dados. Tais módulos poderiam gerar informações importantes a serem utilizadas na tomada de decisão por parte dos gestores da empresa, como as informações de práticas de convivência, da composição de renda familiar e dos indicadores sociais das UFPs. Devido à grande quantidade de dados que o SIGATER adquiriu nos últimos anos (entre 2011 e 2017), sua base de dados chegou ao nível no qual gestores e partes interessadas em tomada de decisão não conseguissem obter tais informações de uma forma mais eficiente.

Com a necessidade de geração de relatórios e monitoramento de atividades de ATER dos técnicos rurais, surgiu a oportunidade de criar um sistema analítico baseado em técnicas de *Business Intelligence* (BI) para o manuseio desses dados. Segundo a OLAP (2017), o termo BI se refere às tecnologias, aplicações, análises e a apresentação de informações de negócio. Os sistemas de BI fornecem visões históricas, atuais e preditivas das operações de negócios, geralmente usando dados coletados e ocasionalmente trabalhando a partir de dados operacionais (OLAP, 2017).

Para tal, foi dado ênfase no desenvolvimento de um Sistema de Apoio a

Decisão (SAD) que permitisse a exibição de painéis de relatórios de ocupação de solo e indicadores sociais em Unidades Familiares de Produção para que, a partir deles, fosse possível fornecer um sistema de apoio ao SIGATER em favor do planejamento das ações de ATER. Tais painéis possibilitariam ao Governo do Estado da Paraíba ter auxílio à definição de políticas públicas relacionadas a Agricultura Familiar, de forma a otimizar recursos empregados.

REVISÃO BIBLIOGRÁFICA

Neste tópico, será apresentada a revisão bibliográfica utilizada para a construção do projeto, como os conceitos de *Business Intelligence* e sua importância, as etapas necessárias para o tratamento dos dados obtidos de bases de dados de sistemas e, por fim, o que é um *Data Warehouse* e *Data Mart*, componentes essenciais na construção de uma ferramenta de BI.

Business Intelligence

Segundo Antonelli (2009), o termo *Business Intelligence* (BI) surgiu na década de 80, pela empresa Gartner Group. Desde a década de 90, até a atualidade, foi presenciada uma grande evolução em todos os setores, inclusive no setor de tecnologia da informação (TI) e ambiente empresarial. As ferramentas de software ganharam maior abrangência e importância, e o próprio termo BI ganhou reconhecimento mundial.

De acordo com Cunha (2015), os componentes de uma ferramenta de BI consistem na extração, transformação e carga dos dados (ETL), no armazenamento desses dados em *Data Warehouses* (DW) ou *Data Marts* (DM), e na análise das informações com um OLAP, no qual é uma interface que permite analisar dados de forma gráfica em múltiplas perspectivas. Esses, juntos, formam a arquitetura para a construção de uma solução de BI.

Os sistemas de BI, por já possuírem indicadores e métricas sobre a organização, permitem aos gestores confrontar os dados existentes em sistemas presentes na empresa, verificando em dados históricos várias informações que possam ajudar na tomada de decisão.

Segundo Antonelli (2009), as ferramentas de BI podem fornecer uma visão sistêmica do negócio e ajudar na distribuição uniforme dos dados entre usuários, sendo seu principal objetivo transformar grandes quantidades de dados em informação de qualidade para a tomada de decisão. Dessa forma, essas ferramentas podem auxiliar as partes interessadas da organização a entender os clientes, a relação entre eles, os produtos e serviços oferecidos e utilizar essa informação como estratégia para o negócio.

Extract, Transform and Load

Inicialmente, para a construção de um sistema de BI, é necessário organizar os dados dos sistemas que farão parte dos indicadores e gráficos presentes na ferramenta. O *Extract, Transform and Load* (ETL) é o primeiro passo para a idealização de um *Data Warehouse* (DW) ou *Data Mart* (DM), no qual os dados de outros sistemas serão tratados, organizados e transferidos.

O processo de ETL é um processo que exige esforço e a maior parte do tempo de construção de um DM/DW, por extrair os dados de fontes heterogêneas e ter de alimentar o mesmo de forma homogênea e concisa que irá servir de base para a geração de relatórios e gráficos de apoio à decisão para a gerência da corporação (DEVMEDIA, 2016).

São quatro as etapas principais de um processo ETL/BI, sendo a primeira a extração, onde é feita a coleta dos dados do sistema de origem, extraíndo-os e transferindo-os para o DW/DM. Após essa transferência, é feita a transformação, objetivando a limpeza e os ajustes nos dados, visando melhorar sua qualidade. Em seguida, é realizada a carga dos dados, ou seja, o carregamento dos dados para a camada de apresentação. Por último, é feito o gerenciamento por serviços que auxiliam o gerenciamento do DW/DM (RIBEIRO, 2011).

Data Warehouse e Data Mart

Segundo Dionizio (2017), “*Data Warehouse* (que no português significa, literalmente, armazém de dados) é um banco de dados orientado por assunto, integrado, não volátil, variável com o tempo, para apoiar decisões gerenciais”.

Um *Data Warehouse* (DW) consiste na junção de diversas tecnologias que permitem realizar consultas em bancos de dados, para posterior análise de informações. O DW possui apenas uma única fonte de dados, que facilita o trabalho do analista. Dados existentes em DWs são tratados previamente pelo ETL, garantindo a não redundância de informações (DIONIZIO, 2017).

De acordo com Kimball (2002, p. 3), existem seis objetivos fundamentais de um DW:

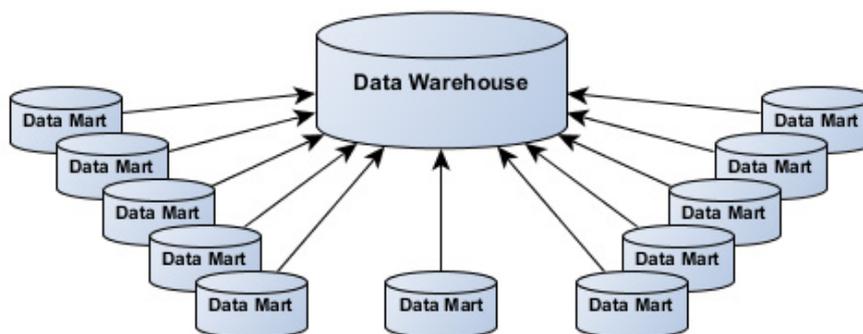
- O DW deve tornar as informações de uma organização facilmente acessíveis;
- O DW deve apresentar as informações de uma organização de forma consistente. Os dados devem ser confiáveis e serem obtidos de uma variedade de fontes da organização, serem consistentes e disponibilizados apenas quando aptos para o consumo dos usuários;
- O DW deve ser adaptativo e resiliente a mudanças. O DW deve ser concebido para lidar com mudanças inevitáveis, e as alterações no mesmo

devem ser feitas de forma a não invalidar dados existentes no DW e nem alterar dados das aplicações que o alimentaram;

- O DW deve ser uma base inabalável que proteja as informações ativas. As informações mais importantes de uma organização são armazenadas em um DW. Logo, o DW deve efetivamente controlar o acesso às informações confidenciais da organização;
- O DW deve servir de base para uma melhor tomada de decisão. O DW deve ter os dados corretos para suportar a tomada de decisão;
- A alta gestão da organização deve aceitar o DW. Se as partes interessadas não abraçarem o DW e não o utilizar de forma continuada, o projeto irá falhar.

Os *Data Marts* (DM) são estruturas moldadas pelos dados granulares encontrados no DW. Os DMs pertencem aos departamentos específicos dentro de uma empresa e são moldados pelos requerimentos dos departamentos (INMON, 2001, p. 11). Assim, os DMs podem ser vistos como as menores unidades que formam um DW. Na Figura 1, é possível visualizar a relação entre um DM e DW.

Figura 1: Data Mart vs. Data Warehouse



Fonte: Data on Focus (2017).

Segundo Souza (2003), a crescente popularidade dos DMs em relação aos grandes sistemas de DWs tem como base a redução no custo de implementação e manutenção de Sistemas de Apoio a Decisão (SAD), na facilidade de prototipação e na limitação de escopo.

METODOLOGIA

O Sistema de Apoio a Decisão (SAD) desenvolvido para o estudo de caso em questão utilizou como fonte de informações de diagnósticos de UFPs a base de dados do SIGATER (Sistema de Gerenciamento de Atividades em ATER), um sistema de processamento de informações em tempo real (*Online Transaction Processing*, ou OLTP) que permite o cadastramento de indicadores sociais, econômicos e ambientais de UFPs.

Devido ao grande volume de dados presente na referida base (3 gigabytes de dados), sua complexidade e a presença de inconsistências (pelo sistema aceitar a entrada de dados manuais ao invés de opções pré-definidas), foi necessária a construção de um repositório de informações integradas, ou *Data Mart* (DM). Na construção do DM, foram consideradas as seguintes etapas:

1. Sumarizar e agregar os dados do SIGATER – Nessa etapa, foi realizada uma análise de todos os dados contidos no SIGATER, a fim de compreender sua estrutura e detectar os dados e filtros desejados para a geração de relatórios;
2. Limpar e refinar dados – Em seguida, foi realizada a remoção de dados inconsistentes presentes no SIGATER, a fim de garantir a qualidade dos dados do *Data Mart*;
3. Importar os dados do SIGATER para o *Data Mart* em questão.

As etapas 2 e 3 descritas anteriormente compõem a etapa ETL (*Extraction, Transform and Load*) da proposta. Em outras palavras, são processos que viabilizaram a integração dos dados do SIGATER para o sistema SAD proposto. Trata-se do primeiro componente de um processo de BI. Para a execução dessa etapa, foi utilizado o software Pentaho Data Integration® (PENTAHO, 2017), no qual provê um ambiente de projeto intuitivo que elimina a codificação (graças a uma interface gráfica *drag-and-drop*) e a complexidade da execução da ETL.

Para armazenar esses dados e o *Data Mart* em questão, foi utilizado o banco de dados MySQL™ da Oracle. Trata-se de um banco de dados de código aberto, escalável, largamente utilizado para soluções baseadas na Web (ORACLE, 2017).

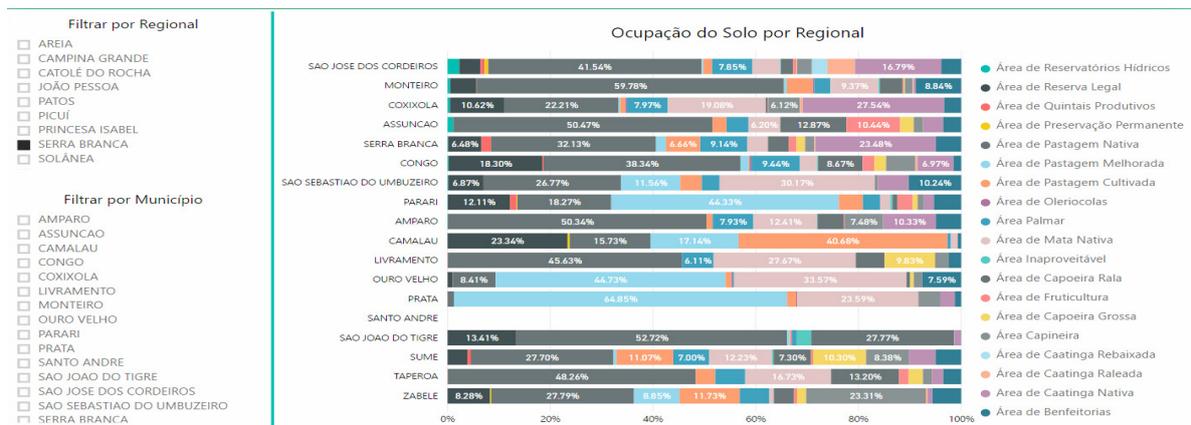
Após a etapa ETL da proposta, foram desenvolvidos os painéis de apresentação (ou *dashboards*) dos relatórios das unidades de produção envolvidas. No desenvolvimento dos *dashboards*, foi utilizado o software PowerBI®, da Microsoft. O PowerBI® é um pacote de ferramentas de análise de negócio para a produção e publicação de relatórios em ambiente Web (POWERBI, 2017). O PowerBI® foi escolhido para o desenvolvimento dos painéis de apresentação pela fácil integração com os principais bancos de dados existentes e pela possibilidade de publicação dos painéis e relatórios gerados diretamente na Internet. Os painéis publicados podem ser acessados por qualquer dispositivo com acesso à Internet, utilizando um navegador Web.

RESULTADOS E DISCUSSÃO

Para a proposta em questão foram definidos três *dashboards*. No primeiro, denominado Ocupação do Solo de uma UFP por município, é possível visualizar um gráfico empilhado horizontal que apresenta a porcentagem de ocupação e uso do solo em uma UFP por município da Paraíba.

Segundo o IBGE (2013), o levantamento da cobertura e do uso da terra indica a distribuição geográfica da tipologia de uso, identificada por meio de padrões homogêneos da cobertura terrestre, sendo de grande utilidade para o conhecimento atualizado das formas de uso e ocupação de espaço, constituindo importante ferramenta de planejamento e orientação na tomada de decisão. Na Figura 2, é apresentado um gráfico de ocupação de solo para os municípios da região de Serra Branca-PB gerado pelo sistema. Nela, é possível constatar que a maior parte da ocupação do solo nessa região é de área de pastagem nativa.

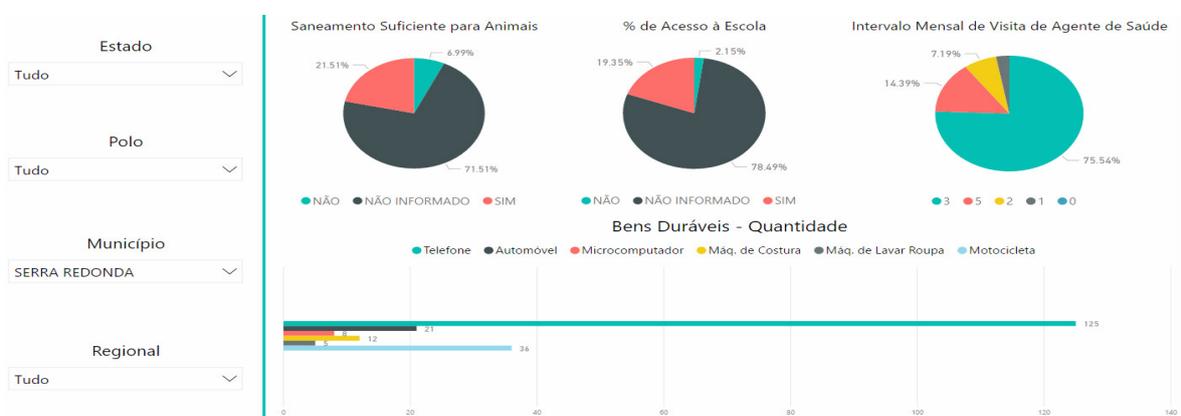
Figura 2: Painel de Ocupação de Solo para a região de Serra Branca-PB



Fonte: Elaborado pelo Autor

No segundo *dashboard*, de Indicadores Sociais de UFP por município, são apresentados gráficos de pizza e de barras horizontais com informações quantitativas sobre o saneamento rural, o acesso à escola e a posse de bens duráveis por parte dos membros da UFP, além do intervalo mensal de visita de agentes de saúde. A Figura 3 apresenta um painel de indicadores sociais do município de Serra Redonda-PB. Pode-se verificar que, no referido município, a maioria das unidades de produção familiar recebe a visita de agentes de saúde a cada três meses.

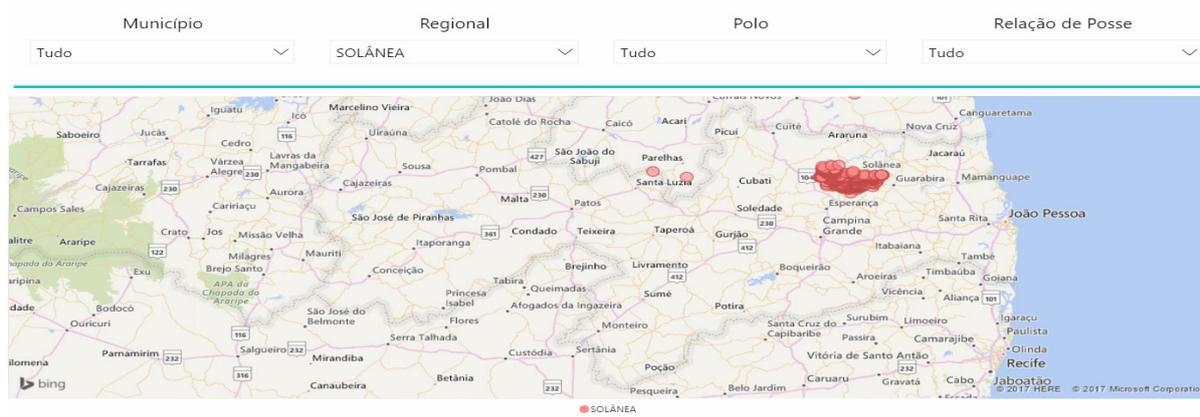
Figura 3: Painel de Indicadores Sociais de UFP para o município de Serra Redonda-PB



Fonte: Elaborado pelo Autor

O terceiro e último *dashboard*, intitulado Mapa de UFPs, é um painel que apresenta um mapa com a localização geográfica de todas as unidades de produção familiar presentes no *Data Mart*. Nesse, utilizando de ferramentas de controle de rolagem/zoom do próprio navegador ou dispositivo, é possível ampliar o mapa para uma região específica, e ao selecionar um dos marcadores do mapa, é exibida um balão com as informações básicas dessa UFP. A Figura 4 apresenta os municípios da região de Solânea-PB.

Figura 4: Painel de Mapas de UFPs da região de Solânea-PB



Fonte: Elaborado pelo Autor

Nesse *dashboard*, é possível também filtrar as unidades de produção familiar exibidas por relação de posse de terra. Ao selecionar quaisquer um dos filtros disponíveis na página, os marcadores das unidades de produção são atualizados com a quantidade de UFPs que obedecem ao critério estabelecido, exibindo também, ao selecionar o marcador, as informações detalhadas sobre a UFP filtrada.

CONSIDERAÇÕES FINAIS

O *Business Intelligence* (BI) é uma ferramenta de auxílio a tomada de decisão bastante importante para uma organização, por sua capacidade de sintetizar e fazer com que dados de diferentes sistemas ou bases de dados se tornem úteis para a estratégia empresarial.

Com o uso de técnicas de BI a partir dos dados do sistema SIGATER, a EMATER-PB poderá definir ações de extensão específicas para os tipos de culturas, plantio e terreno mais abundantes existentes no território paraibano, fazendo com que o atendimento realizado pelos técnicos da empresa seja mais preciso e focado nas necessidades dos agricultores familiares. De acordo com o *feedback* obtido a partir do sistema desenvolvido, foi possível constatar que a ocupação de solo predominante na Paraíba é a de área de pastagem nativa, tendo 31,20% de sua extensão com esse tipo de ocupação. A segunda maior ocupação de solo no Estado é de área de fruticultura. Além disso, foi constatado que a região com maior predominância de saneamento

suficiente para animais foi a de Solânea, com 36,02% das UFPs. A região com a menor predominância foi a de Serra Branca, com 18,26%. Em todas as UFPs, foi verificado que o intervalo mensal de visitas de agente de saúde é a cada três meses. Logo, projetos de ATER focados em tais características terão maiores chances de serem executados com sucesso, gerando um grande impacto na produção e na vida dos agricultores familiares.

Por fim, o sistema de BI desenvolvido facilitou a visualização e interpretação dos dados de unidades de produção familiar presentes no sistema SIGATER pelos gestores e partes interessadas da empresa. Os dados do sistema não eram apresentados de forma gráfica, devido à ausência de módulos de relatórios analíticos. Graças aos *dashboards* apresentados pela ferramenta de BI, os gestores da empresa passaram a compreender melhor as informações fornecidas pelo SIGATER, otimizando a análise de dados para a tomada de decisão.

Como trabalhos futuros e oportunidades de pesquisa, pode-se destacar a utilização do modelo de extração de dados definido nesse trabalho para outras áreas contempladas pelo sistema SIGATER não abordadas nesse estudo de caso, como informações de composição de renda familiar e de infraestrutura da UFP. As informações obtidas e tratadas nesse estudo também podem ser aplicadas em técnicas de *Machine Learning* e *Data Mining*, objetivando a busca por padrões nos dados de agricultura familiar e de unidades de produção familiar que auxiliem ou automatizem a tomada de decisão na EMATER-PB.

AGRADECIMENTOS

Agradecemos à EMATER-PB por gentilmente ter dado espaço e cedido informações, dados, bem como todo o apoio e infraestrutura necessária para a realização desse projeto.

REFERÊNCIAS

ANTONELLI, Ricardo Adriano. Conhecendo o Business Intelligence (BI); Uma Ferramenta de Auxílio à Tomada de Decisão. **Revista TECAP**, N° 3, Ano 3, Volume 3, 2009. Disponível em: <<http://revistas.utfpr.edu.br/pb/index.php/CAP/article/download/933/544>>. Acesso em 29 Set. 2017.

COSTA, G. M.; PENHA, J. L.; LUCENA, W.; FARIAS, T. M. T. **SIGATER – Sistema de Gerenciamento de Atividades em ATER**. 2011.

CUNHA, Juscelino C. Business Intelligence (BI); Conceitos, Técnicas, Sistemas e Ferramentas. **Revista Tecnologus**, Edição 09, 2015. Disponível em: <http://www.unibratesc.edu.br/tecnologus/wp-content/uploads/2015/12/tecnologus_edicao_09_artigo_01.pdf>. Acesso em 29 Set. 2017.

DATA ON FOCUS. **Data Mart vs Data Warehouse**. Disponível em: <<http://www.dataonfocus.com/data-mart-vs-data-warehouse/>>. Acesso em 29 Set. 2017.

DEVMEDIA. **Extract, Transformation and Load (ETL) – Ferramentas BI**. Disponível em: <<http://www.devmedia.com.br/extract-transformation-and-load-etl-ferramentas-bi/24408>>. Acesso em 29 Set. 2017.

DIONIZIO, Leandro. **Data Warehouse**. Disponível em: <<http://www.devmedia.com.br/data-warehouse/12609>>. Acesso em 29 Set. 2017.

GESTÃO UNIFICADA. **Histórico**. Disponível em: <<http://gestaounificada.pb.gov.br/emater-pb/historico>>. Acesso em: 27 Set. 2017.

IBGE. Manuais Técnicos em Geociências; **Manual Técnico de Uso da Terra**. N° 7. 3. ed. Disponível em: <<https://biblioteca.ibge.gov.br/visualizacao/livros/liv81615.pdf>>. Acesso em 30 Set. 2017.

INMON, W. H.; TERDEMAN, R. H.; IMHOFF, C. **Data Warehousing: Como transformar informações em oportunidades de negócios**. 3. ed. São Paulo: Berkeley, 2001.

KIMBALL, Ralph; ROSS, Margy. **The Data Warehouse Toolkit Second Edition**. New Jersey: John Wiley and Sons, Inc., 2002.

OLAP. **What is the Definition of OLAP? OLAP Definition**. Disponível em: <<http://olap.com/olap-definition/>>. Acesso em 30 Set. 2017.

ORACLE. **MySQL Oracle**. Disponível em: <<http://www.oracle.com/technetwork/database/mysql/index.html>>. Acesso em 26 Set. 2017.

PENTAHO. **Data Integration Platform and Software**. Disponível em: <<http://www.pentaho.com/product/data-integration>>. Acesso em 27 Maio 2017.

POWER BI. **O que é o Microsoft Power BI**. Disponível em: <<https://powerbi.microsoft.com/pt-br/what-is-power-bi/>>. Acesso em 27 Maio 2017.

RIBEIRO, Viviane. **O que é ETL?**. Disponível em: <<https://vivaneribeiro1.wordpress.com/2011/06/28/o-que-e-etl-2/>>. Acesso em 29 Set. 2017.

SOUZA, Michael de. **BI: Data Marts**. Disponível em: <<https://imasters.com.br/artigo/1612/gerencia-de-ti/bi-data-marts/>>. Acesso em 29 Set. 2017.

Recebido em: 01/10/2017.

Aceito em: 28/11/2017.