# 2 EMBASAMENTO TEÓRICO

Todas as fontes, trabalhos, artigos e autores utilizados para o embasamento desse trabalho são descritas abaixo.

## 2.1 Business Intelligence

Richard Millar Devens, em 1865, foi quem introduziu o termo “*Business Intelligence*" (Inteligência de Negócios) em seu livro *Cyclopædia of Commercial and Business Anecdotes*. Lá ele conta sobre um bancário, Sir Henry Furnese, que conseguia atuar antes da competição reunindo informações e conseguindo lucrar com elas.

Em 1958, Hans Peter, um cientista da computação da IBM, publicou um artigo que foi um marco no assunto, na qual descrevia o quão potencial o *Business Intelligence* (BI) seria com o uso da tecnologia. O artigo intitulado “*A Business Intelligence System*” descrevia "*An automatic system is being developed to disseminate information to the various sections of any industrial, scientific or government organization. This intelligence system will utilize data-processing machines for auto-abstracting and auto-encoding of documents* […]". (LUHN, 1958, p. 314).

 Após a Segunda Guerra Mundial, houve a necessidade de simplificar e organizar o grande e rápido crescimento dos dados tecnológicos e científicos. Hoje, Luhn é popularmente conhecido como o pai do *Business Intelligence*.



Figura 1 – Hans Peter Luhn

Fonte: (IEEE Spectrum, 2018)

A partir dos anos 60, houve o surgimento de novas formas de armazenamento como os DBMS (*Database Management Systems*), tendo uma evolução no modo de gerenciar grandes volumes de dados, no final da década de 70, nasce o modelo relacional no DBMS. A partir desse momento, todas as informações eram apresentadas e armazenadas em formato digital, fazendo com que fosse possível a concretização do Business Intelligence nas próximas décadas.

Também nessa mesma época, os CPD (Centros de Processamento de Dados), estavam se consolidando, se transformando no meio-termo da tecnologia da informação com os negócios. Os CPD eram focados totalmente em dados, diferente da TI que o centro focava em software, hardware e redes.

Com o surgimento do conceito de sistemas de informações executivas (EIS) há uma grande disseminação do assunto, sendo um dos maiores aliados aos sistemas de BI. Segundo Turban (2009, p. 27), "Esse conceito expandiu o suporte computadorizado aos gerentes e executivos de nível superior. Alguns dos recursos introduzidos foram sistemas de geração de relatórios dinâmicos multidimensionais [...]”.

Na década de 80, alguns fabricantes de softwares voltados ao campo do BI começaram a ganhar terreno. Softwares como *MicroStrategy*, *Business Objects* e *Crystal Reports* começaram a ser populares nas empresas que começaram a usar realmente computadores na época.

Em 1988 houve outro marco importante, com o intuito de simplificar as análises em BI a conferência internacional em Roma, organizada pelo *Multiway Data Analysis Consortium*, dá início a era moderna do Business Intelligence, sendo o termo popularizado pelo analista do *Gartner Group*, Howard Dresner, em 1989.

Na década de 90, se populariza o conceito de Data Warehouse, como um sistema dedicado a auxiliar o BI, também separando em momentos distintos os processamentos OLAP (*Online Analytical Processing*) e OLTP (*Online Transaction Processing*), sendo o OLAP usado ao lado do BI para a montagem de relatórios e posteriormente painéis em inúmeras visões diferentes, e o OLTP sendo utilizado geralmente para explorações estatísticas. Ao mesmo tempo, por consequência, o conceito de ETL (*Extraction, Transformation and Loading*) é incorporado ao *Data Warehouse,* com o intuito de fornecer dados relevantes e fornecer uma extração focada. Os sistemas ERP (*Enterprise Resource Planning*) é um software voltado para a gestão das empresas, garantindo a entrada de dados essencial para que os sistemas de BI, sendo possível reportar aos gestores análises em pontos específicos. Consolidados todos os sistemas envolvidos no BI, quais sejam, Sistemas de Informações Executivas (EIS), ERP, *Data Warehouse* para armazenamento, OLTP, ETL e OLAP nasce o Business Intelligence 1.0. (KUMAR, 2017).

É importante ter em mente sobre a construção e organização do BI é que, este processo é sem fim, sempre haverá novos requisitos a serem cumpridos mesmo tendo terminado o trabalho, sendo necessário refazer todas as etapas novamente.

## 2.2 Sistemas de Informação OLAP/OLTP

De acordo com Primak (2008), o objetivo de uma ferramenta OLAP é permitir a análise multidimensional dinâmica dos dados, apoiando os usuários finais em suas atividades, oferecendo várias perspectivas, onde o próprio usuário cria suas consultas dependendo de suas necessidades, fazendo cruzamento dos dados de formas diferenciadas, auxiliando na busca pelas respostas desejadas.

Para oferecer as várias perspectivas o método mais comumente usados é o MOLAP onde se usa um banco de dados multidimensional com tabelas que mais parecem um cubo, que por esse motivo originou a denominação “dados cúbicos”. Com essas tabelas de múltiplas dimensões é possível cruzar informações que antes não seria possível por uma pessoa, fazendo assim necessário o uso da ferramenta. Existem também outros métodos como o ROLAP que utiliza um banco de dados relacional para seus dados e possui um tempo maior para resposta.

Para Thomsen (2002) o OLAP precisa dos requisitos abaixo para funcionar:

·       Uma estrutura dimensional

·       Especificação eficiente das dimensões e cálculos

·       Separação da estrutura e representação

·       Flexibilidade

·       Velocidade suficiente para suportar as análises ad hoc

·       Suporte para multiusuários

Segundo Prasad (2007) o processo OLAP se diferencia do OLTP (*Online Transaction Processing* ou Processamento de Transações em Tempo Real) que foca em processar transações repetitivas em alta quantidade e manipulação simples. Já o OLAP envolve uma análise de vários itens de dados em relacionamentos complexos, que busca padrões, tendências e exceções.

O foco principal do OLTP é transações online como o nome já diz, suas consultas são simples e curtas, portanto não precisa de tanto tempo de processamento, também utiliza pouco espaço. O banco de dados é atualizado frequentemente, pode acontecer no momento da transação e também é normalizado. Um exemplo muito utilizado ao se explicar o OLTP é o ATM (do inglês, caixa automático) onde cada transação modifica a conta do usuário.

## 2.3 Data Warehouse

    As aplicações de *Business Intelligence* necessitam de um repositório específico (devido diversas particularidades) para buscar os dados que serão usados para a operação, sejam eles de que tipo for. O local onde serão centralizadas essas informações é chamado de *Data Warehouse* (DW). Segundo Inmon (2005, p. 29), “Data Warehouse (que no português significa literalmente armazém de dados) é um deposito de dados orientado por assunto, integrado, não volátil, variável com o tempo, para apoiar as decisões gerenciais”. Como podemos ler na definição de Inmon, a necessidade de um repositório específico para as informações é definida pelos seguintes itens:

* Orientado por assunto: Significa que os dados ali presentes têm contexto direto com as atividades da empresa, ou seja, as informações contidas são, unicamente, as necessárias para definir as operações.
* Integrado: Todas as atividades do DW devem estar conectadas ao ambiente operacional.
* Não volátil: Os dados que estão especificadamente no Warehouse não podem sofrer mudanças, tal como alterações e inclusões, (já que precisamos uma informação concreta que não se altere para tomar decisões) podendo ser apenas consultados ou excluídos.
* Variável com o Tempo: Está ligado ao conceito da Não-Volatilidade, mas aqui com um foco maior no tempo dessas informações. Já que os dados dentro do DW não podem ser alterados, o horário das informações também não pode mudar. Por isso precisamos ter a certeza do tempo em que elas estão armazenadas.

     Enquanto o Data Warehouse agrega todos os dados, definições e relacionamentos entre eles, o *Data Mart* (DM) é um pequeno DW dentro do contexto maior que se preocupa em adquirir apenas uma parte dessas informações para uma operação específica. Podemos fazer uma analogia com um comerciante que possui uma loja e um armazém, ao comprar novos produtos para o seu comércio, ele, primeiramente, vai armazenar tudo o que for possível nesse armazém, para, posteriormente, pegar certas quantidades de determinados produtos e apresentá-los na sua loja para vendê-los. Isso é feito para que os relatórios gerados utilizem uma única base, a específica para adquirir as informações.

## 2.4 O Método ETL

As informações que serão utilizadas no processo de Business Intelligence são adquiridas por meio de um processo chamado ETL (*Extract, Transform and Load)*. Esse processo tem como função "transportar" e transformar os dados para todas as instâncias do BI, seja na aquisição dos dados das fontes, inserção desses dados no Banco de Dados e transformação dos dados nos tipos necessários para a realização da análise. Esse processo já foi conhecido como ETLM, em que o "M" significava *Maintenance* (Manutenção), mas esse último já caiu em desuso (BRAGHITTONI, 2017). Cada uma das suas ações será explicada adiante.

E -  *Extract (Extração*): A primeira parte do processo de transporte é a extração periódica das informações para que sejam posteriormente carregadas. Elas podem ser extraídas de diversas fontes, sejam arquivos CSV, arquivos texto (TXT), arquivos  *mainframe* , sites e até arquivos em diferentes fontes de dados (CARVALHAES e ALVES, 2015). A extração deve estar preparada para reconhecer esses dados nas mais variadas fontes possíveis.

T -  *Transform (*Transformação): Após os dados extraídos, eles precisam passar por uma verificação para depois serem "carregados" no *Data Warehouse.* Como Inmon (p.29, 2005) afirma que os dados dentro do DW não podem ser modificados (propriedade Não Volátil), é recomendado o uso de uma instância intermediária ao *Warehouse para* que eles sejam transformados segundo as regras de negócio. Essa instância pode ser um outro BD chamado Staging Areaou a própria memória do Servidor/Computador (CARVALHAES e ALVES, 2015).

L -  *Load* (Carga): O último processo é da carga propriamente dita no Data *Warehouse* e nos seus Data *Marts.* No final desse processo os dados já estão prontos para o uso de alguma ferramenta de BI, transformando eles em algum tipo de visualização gráfica (*Dashboards).* Essa carga é feita de forma incremental, já que, como mencionado anteriormente por Inmon, esses dados não podem sofrer mudanças (BRAGHITTONI, 2017).

## 2.5 Pentaho Data Integrator

Com o desejo de alcançar uma mudança positiva no mercado de análise de negócio dominada por grandes vendedores que oferecem produtos baseados em plataformas com custo elevado, foi-se criado o *Pentaho,* uma ferramenta de gerenciamento de *Business Intelligence (*BI), desenvolvido para fins de recolher o máximo de dados diversificados e não estruturados a partir de diversas fontes e analisa-los para encontrar novos padrões, indicadores de tendências e base de dados para inovação.

Por ser uma tecnologia *Open Source,* o *Pentaho* possui uma versão funcional extremamente potente em que não há custo algum com licenças. É a ferramenta de código aberto mais utilizada do mundo, contendo um ambiente de desenvolvimento integrado e bastante poderoso.

O *Pentaho* possui diversas funções a fim de entender e analisar o negócio empresarial. Algumas delas permitem que o usuário possa acessar e preparar fontes de dados para análise, mineração e geração de relatórios *on-line,* via *web.* Também vêm integrado com um ambiente de desenvolvimento, baseado no *Eclipse* (API), que visa a resolução de soluções mais complexas de BI.

## 2.6 Microsoft Power BI

Com o crescimento de negócios das empresas, uma grande massa de dados que entram nessas empresas também aumentou e por causa disso, ficou difícil organizar, analisar, monitorar e compartilhar essa quantidade de informações. Para resolver esse problema, foi necessário criar ferramentas que pudessem atender a esses requisitos.

Dentre várias ferramentas que foram criadas, têm-se o *Power BI.* É um pacote de ferramentas de análise de negócios que tem como objetivo a visualização, organização e análise de dados. O *Power BI* é uma ferramenta baseada na nuvem, tornando possível, ao usuário, a conexão de seus dados em tempo real, ou seja, em qualquer lugar que eles estejam, com rapidez, eficiência e compreensão. Além disso, ele permite a criação de painéis de simples visualização, fornecimento de relatórios interativos e o compartilhamento desses dados obtidos.

Sendo uma ferramenta de *Business Intelligence,* o *Power BI* não é apenas para a inserção de dados e criação de relatórios. Ela transforma os dados brutos em informações significativas e uteis a fim de analisar o negócio, permite uma fácil interpretação do grande volume de dados e entrega análises que ajudam na decisão de uma grande variedade de negócios, variando do operacional ao estratégico. Também é capaz de limpar os dados formatados de forma irregular garantindo que se tenha só os dados interessados ao negócio e modela os dados quem vêm de diversas fontes para que se consiga fazer com que esses dados trabalhem de forma eficiente.

## 2.7 MEC e INEP

MEC

A história do MEC é antiga, começando em 1930 quando foi criado no governo de Getúlio Vargas, inicialmente era chamado de Ministério dos Negócios da Educação e Saúde Pública. Como podemos ver pelo nome, a educação não era o único foco de atividade. Apenas em 1995, no governo de Fernando Henrique Cardoso, a educação ficou exclusiva ao ministério. A sigla MEC surgiu em 1953, quando se criou o Ministério da Educação e Cultura.

Até 1960 de acordo com o MEC (2015), ele era centralizado, um modelo que era seguido por todos os municípios e estados. A primeira Lei de Diretrizes e Bases da Educação (LDB), que fora aprovada em 1961, diminuiu a centralização do MEC, fazendo assim os órgãos municipais e estaduais ganharem mais autonomia.

Em 2015 a primeira versão da BNCC (Base Nacional Comum Curricular) é disponibilizada, uma pauta muito debatida e importante para nossa educação, somente em 2017 ela foi homologada para Ensino Básico e um ano depois, para o Ensino Médio.

A Base é um documento normativo da maior importância, porque define o conjunto de aprendizagens essenciais que todos os alunos devem desenvolver ao longo da Educação Básica e do Ensino Médio, e orientar as propostas pedagógicas de todas as escolas públicas e privadas de Educação Infantil, Ensino Fundamental e Ensino Médio, em todo o Brasil. (MEC, 2015).

INEP

O Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) é vinculado ao MEC e engloba várias áreas educacionais, desde ensino básico até a graduação. A respeito de sua origem, é preciso considerar que:

Chamado inicialmente de Instituto Nacional de Pedagogia, o Inep foi criado, por lei, em 13 de janeiro de 1937, no Rio de Janeiro. Foi em 1938, entretanto, que o órgão iniciou, de fato, seus trabalhos. A publicação do Decreto-Lei nº 580 regulamentou a organização e a estrutura da instituição, além de modificar sua denominação para Instituto Nacional de Estudos Pedagógicos. [...] em 1952, assumiu a direção do Instituto o professor Anísio Teixeira, que passou a dar maior ênfase ao trabalho de pesquisa. Seu objetivo era estabelecer centros de pesquisa como um meio de fundar em bases científicas a reconstrução educacional do Brasil. A ideia foi concretizada com a criação do Centro Brasileiro de Pesquisas Educacionais (CBPE), com sede no Rio de Janeiro, e dos Centros Regionais, nas cidades de Recife, Salvador, Belo Horizonte, São Paulo e Porto Alegre. Tanto o CBPE como os Centros Regionais estavam vinculados à nova estrutura do Inep. (INEP, 2015)

Na década de 70, com a sede sendo transferida para Brasília e o CBPE sendo extinto, fez com que o modelo que fora idealizado por Anísio Teixeira fosse finalizado, que causou um reconhecimento ao INEP tanto nacionalmente quanto internacionalmente. Nos anos 80, passou por uma reforma institucional após um período de dificuldades que passou, e tinha dois objetivos, que eram reorientação das políticas de apoio a pesquisas educacionais e o reforço do processo de disseminação de informações educacionais.

O INEP que conhecemos hoje é devido à incorporação do Serviço de Estatística da Educação e Cultura (Seec) à Secretaria de Avaliação e Informação Educacional (Sediae) que de acordo com o (INEP, 2015) a partir de 1997 um único órgão encarregado das avaliações, pesquisas e levantamentos estatísticos educacionais no âmbito do governo federal passou a existir através de uma integração do Sediae ao INEP. Nesse mesmo ano, o Inep foi transformado em autarquia federal.

Como os dados são coletados e disponibilizados

Os dados oferecidos pelo MEC, INEP e outros órgãos do governo são dados abertos, o que significa que estão disponíveis para todos usarem e também redistribuírem como quiserem, sem restrição de patentes, licenças ou parecido. Cada órgão disponibiliza os seus dados de acordo com seu plano de dados abertos (*PDA*) e é responsável pela catalogação do mesmo.

De acordo com o Portal Brasileiro de Dados Abertos (2017) em 2007 um grupo de 30 pessoas se reuniu na Califórnia - EUA, para definir os princípios dos Dados Abertos Governamentais. Eles criaram oito princípios, que são:

1. **Completos**: Todos os dados públicos são disponibilizados, que no caso, não são sujeitos a limitações válidas de privacidade, segurança ou controle de acesso, reguladas por estatutos.
2. **Primários**: Os dados são publicados do mesmo jeito que fora coletada na fonte, e não de forma agregada ou transformada.
3. **Atuais**: Os dados são disponibilizados o mais rápido possível para preservar o seu valor.
4. **Acessíveis**: Os dados são público para o maior público possível.
5. **Processáveis** **por máquina**: Os dados são estruturados para possibilitar o seu processamento automático.
6. **Acesso não discriminatório**: Os dados estão disponíveis para todos sem necessidade de identificação.
7. **Formatos não proprietários**: Os dados estão disponíveis sem que nenhum ente tenha exclusivamente um controle.
8. **Licenças livres**: Os dados não estão sujeitos a restrições como dito no parágrafo acima.