

# THE GAP BETWEEN BUSINESS INTELLIGENCE AND ADVANCED ANALYTICS

Vânia Besouro<sup>1</sup>, Pedro Arroz Serra<sup>2</sup>

Autor<sup>1</sup>

Orientador<sup>2</sup>

Complementos de Engenharia de Software

Universidade Lusófona de Humanidades e Tecnologia

## Resume

Inteligência de negócio, refere-se ao processo de extração, organização, análise, carregamento e monitorização dos dados. *Analytics* é a descoberta, interpretação e comunicação de padrões significativos nos dados. As organizações acumulam informação que são geradas e armazenam-nas em data *warehouses* ou data marts.

Todas as empresas produzem milhares de dados, alguns estruturados outros não.

## [Palavras-chave:]

Análise Avançada de Dados; Inteligência de Negócio; Dados;

## Abstract

Business intelligence, refers to the process of extracting, organizing, analyzing, loading and monitoring data. Analytics is the discovery, interpretation and communication of significant patterns in the data. Organizations accumulate information that is generated and store it in data warehouses or data marts.

All companies produce thousands of data, some structured and others not.

## [Keywords:]

Advanced Data Analysis; Business Intelligence; Data;

## Introdução

No mundo competitivo de dados e negócios uma empresa depende de ferramentas que a possam ajudar na tomada de decisão.

Este documento visa indicar alguns problemas na implementação bem como arquitetura, explicando a metodologia que é implementado num sistema de inteligência de negócio.

Este documento é um artigo para uma unidade curricular para o curso de Mestrado de Engenharia Informática e Sistemas de Informação. É o ponto de partida para algo que possa ser a continuado em Tese. Está organizado por capítulos como: Conceitos, Metodologias, Análise de Dados sem descurar os clássicos capítulos, Introdução, Conclusão e Referencias.

## Conceitos

Inteligência de negócio, (em inglês *business intelligence*, do acrónimo B.I.), refere-se ao processo de extração, organização, análise, carregamento e monitorização dos dados. Este processo tem como base oferecer suporte na tomada de decisão empresarial. Para poder usufruir desta tecnologia há a necessidade de ter grandes volumes de dados. Uma vez que estes dados existem o principal objetivo é usufruir deles. Com os dados, as tecnologias de B.I. podem fornecer a visão do histórico das operações do negócio assim como a visão atual ou até mesmo possíveis previsões. Porém, estas não são as únicas funcionalidades que B.I. pode oferecer.

## Metodologia

Na década de 90, numa determinada empresa, surgiu o conceito de inteligência de negócio com a necessidade de explorar informações e recursos financeiros em proveito dos diretores.

A evolução da tecnologia e consequentemente da informação com as direções que levou ao longo dos anos atendeu a muitas necessidades, e particularmente às necessidades associadas à inteligência de negócio. Estas informações, alienado à estrutura de informações que cada vez mais é variada integrado à tecnologia para aprimorar a sua utilidade e significado com os dados que são recolhidos de várias fontes. As necessidades empresariais levaram a cabo a necessidade de utilizar as mais recentes tecnologias para que se deixasse de fazer *querys* extensas e obtivessem informações de forma tão rápida quanto “um estalar de dedos”.

Nos dias de hoje, as organizações acumulam informação que são geradas e armazenam-nas em *Data Warehouses* ou *Data Marts*. Esta é uma realidade atual e é um conceito que é descendente de uma aplicação do processamento de transações OLTP que é migrado por um sistema de B.I..

### Modelação de Dados e B.I. Aplicada ao Desenvolvimento

*Structured Query Language* é uma linguagem, normalmente com permissões CRUD, associada exclusivamente a banco de dados relacionais. Assim como bancos de dados são simplesmente coleções consideráveis de dados relacionados, SQL é uma linguagem declarativa.

O ciclo de vida de desenvolvimento de *software* representa um conceito abstrato. O método *Waterfall* é uma implementação de um *Software Development Life Cycle* (SDLC - processo usado na indústria de desenvolvimento de software para desenvolver e testar a qualidade do software), geralmente categorizado por cinco a sete estágios através dos quais é criado o software aplicado ao negócio.

Ainda na implementação alguns problemas são exibidos num plano de projeto. Estes problemas incluem, a título de exemplo, as pessoas que reúnem toda a experiência em aplicações, negócios e conhecimento para criar o modelo de dados,

não são as mesmas pessoas que criam a programação necessária para aceder ao modelo de dados; Os programadores são as pessoas mais habilitadas para encontrar colunas e outros erros no modelo de dados, no entanto, quando o trabalho é concluído em certos momentos estes são libertados. Ou até mesmo quando os programadores contam com os modeladores para fazer alterações o que pode criar dependências e por consequência erros maiores.

Na fase de teste do sistema, novos erros são descobertos à medida que se torna conhecido que os programadores interpretaram mal o modelo de dados, juntando-se incorretamente outras atividades similares.

Normalmente, e certamente sem tudo estar concluído, o desempenho da aplicação também é dos últimos elementos a serem testados após a migração de dados do sistema existentes.

Além da modelagem de dados para B.I., o processo de modelagem de dados pode ser integrado num ciclo de vida de desenvolvimento de software, porém, este não é o foco.

### Ambiente de Informação

Deve ficar evidente que B.I. tem tudo a ver com B.I. mas não só. A residência, por assim dizer, para este conceito é um armazém e é criado para uma vantagem específica. Esta vantagem nos dias de hoje é vista como um requisito comercial. Este requisito é um ambiente de informação estruturado e organizado. Esta estrutura tem vários tipos de organizações de dados diferentes. Na figura seguinte está representado, em forma de pirâmide, a estrutura desta informação.

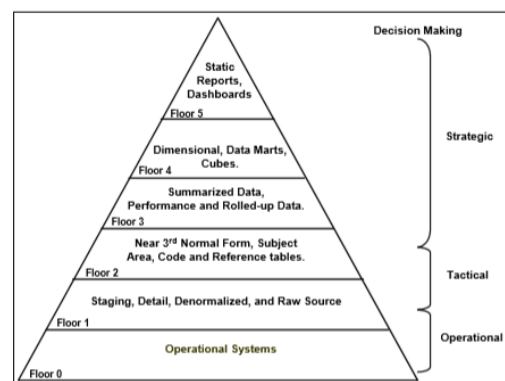


Figura 1 - Estrutura da Informação

A organização de tecnologia da informação tem tradicionalmente visto

diferentes tipos de dados na pirâmide de informação como camadas separadas e exigia que tipos de dados na pirâmide de informação como camadas separadas e exigia que dados a serem copiados para outra. No entanto quando se olha para tipos diferentes de dados como visões diferentes dos mesmos dados, com diferentes características necessárias para realizar um trabalho específico. Para enfatizar isto, é apresentado por pisos de informação. Para mover e copiar os dados entre os dados (geralmente da parte inferior para a parte superior) não é a única opção disponível. Há um número variado de abordagens disponíveis. Estas abordagens permitem a integração dos dados na empresa e existem ferramentas que permitem essas abordagens. Os dados de cada andar têm características diferentes, como volume, estrutura e método de acesso. Com este modelo podemos escolher a melhor forma de armazenar e organizar fisicamente os dados nos diferentes andares. Este tipo de sistema deve ser mapeado para as camadas numa arquitetura de *data warehouse* existente. No entanto, só deve ser utilizado para completar a compreensão e a subsequente migração dos dados. O ideal seria uma fonte corporativa integrada de dados para a tomada de decisão e por sua vez a exibição atual e ou até mesmo em tempo real.

## B.I. em Tempo Real

A tomada de decisões táticas e estratégicas, estão cada vez mais a possuir alinhamento à medida que os negócios e a tecnologia continuam mudando e melhorando. Existe uma quantidade vantajosa de requisitos e uma necessidade de informações atuais por todos. A alimentação desta necessidade e aceleração em direção a outra chave é chamada de *closed-loop analytics*, representado na imagem seguinte.

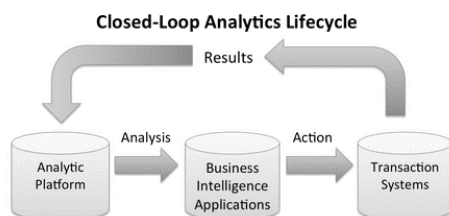


Figura 2 - *Closed-loop Analytics*

Este ciclo tem a particularidade de mostrar que os resultados dos dados

armazenados as análises estão a ser devolvidas ao sistema operativo. Após este os eventos podem ser realizados quase imediatamente, em vez de simplesmente tentar minimizar o seu impacto, este é um salto significativo em frente e agora pode ser realizado com capacidade de suporte em tempo real.

A Internet teve um papel fundamental neste movimento. Os investidores têm acesso à informação como nunca antes tinham, com a segurança que antes para se conseguir demorava tempo. A necessidade de velocidade tornou-se um requisito importante. As empresas querem informação aqui, e agora!

Os processos que fornecem dados podem funcionar independentemente ou em conjunto.

Um exemplo de implementação de B.I. em tempo real envolve a integração de várias atividades. Essas atividades são necessárias em qualquer mecanismo *Extract Transform Load (ETL)* paralelos. Os dados devem ser extraídos do sistema operativo, limpos e transformados antes de serem colocados no armazém para ser utilizados no processamento e análise de consultas. Estes processos de extração, carregamento, transformação têm sido historicamente orientados a porções. Nesta fase eles podem ser alterados para serem executados num único processo contínuo. A execução múltipla desses processos em paralelo é outra alternativa possível. Outra abordagem é extrair e carregar os dados e executar tão rapidamente quando possível. Este processo pode também ser feito num período o menor possível.

Outro exemplo de implementação de B.I. em tempo real são replicação do *WebSphere Information Integrator – Replication Edition* fornece essa capacidade – duplicação. O componente de captura de duplicação leva os dados necessários dos *logs* do sistema, o que elimina a necessidade de desenvolvimento de aplicações para satisfazer alguns requisitos. Este é um exemplo de como um sistema pode fornecer a capacidade de independência humana.

Estas não são as únicas ferramentas que existem para B.I. em tempo real, mas o facto essencial é que e sempre foi essencial adaptar as necessidades das empresas e verificar sempre as ferramentas existentes e como podemos tirar o melhor partido destas.

## Análise de Dados

Todas as empresas produzem milhares de dados, alguns estruturados outros não. A verdade é que os dados que não são estruturados fazem a maioria dos dados.

A consulta dessa informação, após toda a estrutura armazenada em *data warehouse* ou até mesmo em *data marts* tem um processo desde a consulta até que é transformado num relatório. A definição de consulta é o processo de responder a uma pergunta ou hipótese comercial e traduzindo-o para um formato que possa ser usado para um suporte de decisão específico. Quando a consulta é executada, a ferramenta gera o idioma apropriado para aceder e recuperar os dados solicitados, retornando no que normalmente é chamado de conjunto de respostas. O analista de dados então executa os cálculos e manipulações no conjunto de respostas para alcançar os resultados desejados. Esses resultados são então formatados para facilitar a compreensão pelo utilizador.

O processo de consulta desta forma está simplificado na figura seguinte.

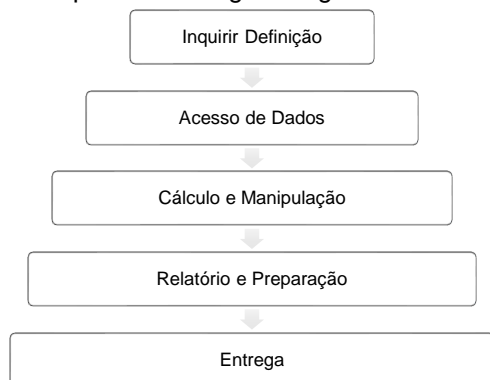


Figura 3 - Processo de Consulta

A análise é a descoberta, interpretação e comunicação de padrões significativos nos dados. Para que esta análise seja avançada implica que a aplicação de padrões de dados seja tomada em conta na tomada de decisão. A análise pode ser entendida como o tecido conjuntivo entre os dados e a tomada de decisão eficaz dentro de uma organização. Este fator é valioso em áreas ricas como estatística, programação e pesquisa operacional. Pode servir para quantificar desempenho, mas não só.

As organizações podem aplicar análises aos dados para descrever, prever e melhorar o desempenho. Especificamente

as áreas da análise incluem análise preditiva, análise prescritiva, gestão de decisões, análise descritiva, cognitiva, *Big Data Analytics*, variedade e otimizações, modelagem, dimensionamento e optimização, análise de risco, força e, entre outras áreas, análise de fraude.

A análise exige computação extensiva, os algoritmos utilizam os métodos mais atuais em ciências da computação, estatística e matemática.

A análise é focada na compreensão do passado. O que aconteceu e porque aconteceu. A análise desde dados é um campo multidisciplinar. As informações dos dados são usadas para recomendar ações. Existe uma tendência acentuada para usar análises de texto, bem como técnicas e *machine learning* com redes neurais artificiais, árvores de decisão, lógica de regressão, análise de regressão linear a múltipla, classificação de modelagem preditiva, mas também análise de cluster, análise de componentes principais e análise de associação.

## Conclusão

Alguns são os desafios mais comuns que as empresas enfrentam nos dias de hoje nas empresas.

Na era da informação, as organizações com visão de futuro estão a alcançar novos níveis arrebatadores, aproveitando os recursos das soluções inovadoras que a tecnologia tem para oferecer. Porém ainda existem algumas dificuldades.

As principais dificuldades ainda são o facto de algumas tecnologias serem caras, a falta de entendimento por todos nas empresas, o facto de as fontes de dados serem ainda muito diferentes ou até mesmo as empresas ainda não estarem despertas e não estarem a medir os indicadores certos.

Na falta de estratégias para lidar com estas dificuldades recomenda-se soluções práticas e simples.

## Referências

BALLARD, C. (2006). DIMENSIONAL MODELING. [UNITED STATES]: IBM INTERNATIONAL TECHNICAL SUPPORT ORGANIZATION.

KOHAVI, R., ROTHLEDER, N., & SIMOUDIS, E. (2002). EMERGING TRENDS IN BUSINESS ANALYTICS. COMMUNICATIONS OF THE

ACM, 45(8). DOI:  
10.1145/545151.545177

MARKET RESEARCH MATTERS: TOOLS AND  
TECHNIQUES FOR ALIGNING YOUR  
BUSINESS. (2000). CHOICE REVIEWS  
ONLINE, 38(01), 38-0388-38-0388. DOI:  
10.5860/CHOICE.38-0388