



Fundamentos do Bootcamp

Bootcamp: Desenvolvedor Business Intelligence

Daniel de Oliveira Capanema

2020

Fundamentos do Bootcamp

Bootcamp: Desenvolvedor Business Intelligence

Daniel de Oliveira Capanema

© Copyright do Instituto de Gestão e Tecnologia da Informação.

Todos os direitos reservados.

Sumário

Capítulo 1. Definição e Fundamentos de BI	7
Histórico	7
Definição	8
Arquitetura Básica.....	10
Armazém de Dados ou Data Warehouse	11
Características do Data Warehouse	11
Data Marts	13
Estrutura do Data Warehouse.....	13
Capítulo 2. Maturidade para BI.....	15
O modelo TDWI	15
Os cinco estágios do modelo de maturidade TDWI	16
Estágio 1 – Inexistente	16
O golfo	17
Estágio 2 – Preliminar	18
Estágio 3 – Repetido.....	18
O abismo.....	19
Estágio 4 – Gerenciado.....	20
Estágio 5 – Otimizado	21
As dimensões do modelo de maturidade TDWI.....	22
Escopo	22
Patrocinador.....	23
Apoio financeiro	24
Retorno	24
Arquitetura	24

Dados.....	25
Desenvolvimento.....	25
Entrega	26
Capítulo 3. Etapas do Projeto de BI	28
Justificativa	28
Planejamento	29
Análise	29
Design.....	30
Construção.....	31
Implantação	32
Capítulo 4. Planejamento do Projeto de BI.....	33
Gerenciamento do Projeto de BI.....	34
Definição do Projeto de BI	36
Metas e objetivos	36
Escopo do Projeto.....	37
Riscos de projeto	37
Restrições do projeto	38
Premissas	40
Procedimentos de Controle de Mudança	41
Procedimentos de gerenciamento de problemas	42
Planejamento do Projeto de BI	43
Atividades e tarefas.....	43
Técnicas de estimativa.....	44
Atribuição de recursos	45
Dependências de tarefas	46

Dependências de recursos	47
Método do Caminho Crítico	47
Horários do projeto.....	48
Capítulo 5. Ciclo Analítico da Inteligência de Negócios.....	50
Monitoramento de atividades	51
Identificação das exceções.....	51
Determinação dos fatores causais.....	51
Modelagem de alternativas	52
Atuação e acompanhamento dos resultados	52
Capítulo 6. Tipos de soluções de BI	53
Consultas de Acesso Direto	54
Relatórios padrão.....	55
BI Operacional	56
Aplicativos analíticos.....	56
Gestão do conhecimento	56
Inteligência competitiva.....	58
Data Mining.....	58
Dashboards.....	59
Balanced Scorecards.....	60
Capítulo 7. Modelagem Dimensional.....	63
Tabelas Fato e Dimensão	64
Hierarquia das dimensões	65
Enterprise Datawarehouse Bus Matrix.....	66
Operadores	67
Dados Operacionais e Informações.....	69

Data Warehouse e Data Marts.....	70
Capítulo 8. Construção do Modelo Dimensional.....	73
Abordagem Bill Inmon.....	73
Abordagem Kimball.....	74
O Ciclo Kimball e Projeto Físico.....	76
Seleção do processo de negócio	76
Definição do grão	77
Escolha das dimensões	77
Métricas das tabelas de fatos	78
Capítulo 9. Ferramentas de BI.....	79
Referências.....	86

Capítulo 1. Definição e Fundamentos de BI

Neste primeiro capítulo será feita uma introdução ao *Business Intelligence* (BI, ou Inteligência de Negócios, em português) além da apresentação de conceitos e definições fundamentais para o entendimento da disciplina.

Histórico

A análise de fatos passados para previsão de acontecimentos futuros ou para a tomada de decisões já é utilizada há milhares de anos. O livro “A arte da guerra”, escrito a mais de 2 mil anos, já dizia que para se prosperar em uma batalha, é necessário conhecer seus pontos fortes e fracos, assim como os do seu inimigo. Em nosso curso, no entanto, vamos focar no histórico de utilização do termo *Business Intelligence*, voltado para o apoio na tomada de decisões de empresas e instituições. A linha do tempo abaixo mostra alguns fatos marcantes da evolução desse termo, desde sua primeira utilização até os tempos atuais.

1865: o primeiro uso conhecido do termo "business intelligence". Richard Millar Devens descreveu o sucesso de um banqueiro em seu livro “Cyclopaedia of Commercial and Business Anecdotes”.

1958: pesquisador da IBM, Hans Peter Luhn, em seu artigo “A capacidade de apreender as inter-relações dos fatos apresentados de maneira a orientar a ação para um objetivo desejado”.

Anos 70 e 80: crescimento da popularidade dos sistemas de apoio à decisão (DSS) — infraestruturas informatizadas que organizam e analisam dados operacionais.

Década de 1990: ferramentas de Business Intelligence. O termo e as técnicas se tornam comercializáveis e atualizados através do uso de relatórios de processamento em lote. Surgimento de ERPs.

2000: Análise preditiva. Aprendizado de máquinas para prever mudanças no futuro dos negócios. Tecnologias da nuvem, comércio eletrônico e as redes sociais introduzem novas oportunidades para Business Intelligence.

2010 – 2020: Business Intelligence torna-se uma ferramenta padrão para todos, desde gigantes até PMEs. As aplicações contemporâneas cruzam vários dispositivos e usam análises visuais para aplicar raciocínio analítico aos dados, através de interfaces visuais interativas.

Definição

O termo Business Intelligence é citado de maneiras diferentes por diferentes autores, o que não significa que alguns estejam incorretos. No entanto, é preciso tomar cuidado com alguns fabricantes que apresentam suas ferramentas como um produto milagroso que irá realizar todo o processo de BI: o BI é bem mais que isso.

A seguir, são apresentadas algumas definições de autores conhecidos:

BARBIERE: Business Intelligence representa a habilidade de se estruturar, acessar e explorar informações, normalmente guardadas em Data Warehouses e Data Marts, com o objetivo de desenvolver percepções, entendimentos e conhecimentos, que podem produzir um melhor processo de tomada de decisão.

KIMBALL e ROSS: Business Intelligence é um termo genérico para descrever o levantamento de informações sobre os ativos internos e externos da organização, para que melhores decisões de negócios possam ser tomadas.

MOSS e ATRE: O BI não é um produto, nem um sistema. É uma arquitetura e uma coleção de aplicações e bancos de dados com acesso facilitado aos dados e que provê suporte à tomada de decisão.

Gartner Group: Business Intelligence é definido como a aplicação de um conjunto de metodologias e tecnologias (como por exemplo, J2EE, DOTNET,

Serviços da Web, XML, data warehouse, OLAP, Data Mining, tecnologias de representação, etc.) para melhorar a eficácia das operações da empresa e apoiar o gerenciamento/decisão para alcançar vantagens competitivas.

Todas as explicações do termo BI citadas anteriormente estão corretas e se complementam. Uma outra forma muito bem aceita de definição do BI é apontada por Howard Dresner em 1989. Ele o apresenta como um termo guarda-chuva, capaz de contemplar uma variedade de tecnologias, conhecimentos e estratégias usados para analisar e organizar os dados gerados durante as execuções dos processos inerentes a um determinado negócio.

O foco de soluções de BI é facilitar o entendimento do negócio das organizações, fornecendo a todos os níveis das organizações informações relevantes sobre suas operações internas e sobre o ambiente externo, incluindo clientes e competidores, parceiros e fornecedores. O ambiente externo inclui ainda variáveis independentes que possam impactar no negócio, como tecnologia, leis, economia mundial, entre outros.

As aplicações de BI podem auxiliar em vários segmentos das organizações, das quais podemos citar:

- Tendências de transformação do mercado.
- Alterações no comportamento de clientes e padrões de consumo.
- Preferências de clientes.
- Recursos das empresas.
- Condições de mercado.

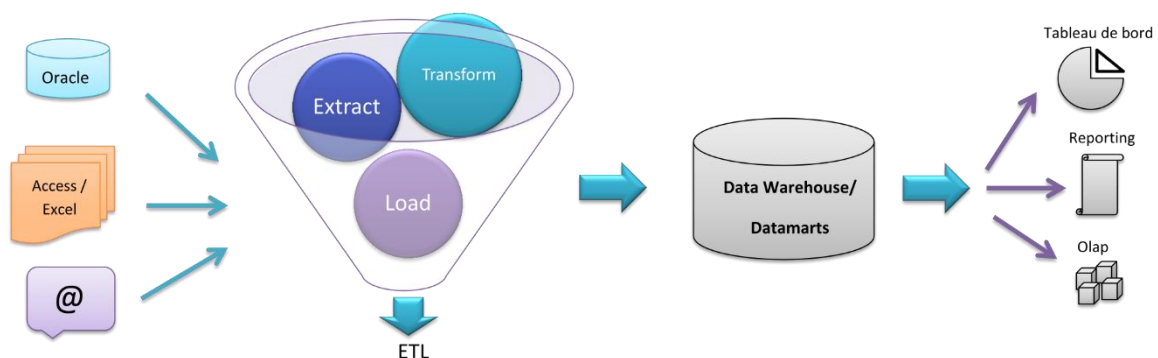
Arquitetura Básica

Como vimos anteriormente, a arquitetura de BI consiste em diversos processos, estratégias e técnicas para transformação de dados em informação. De maneira mais ampla, pode-se dividir a arquitetura de BI em três principais componentes:

- ETL (Extraction, Transformation and Loading): processo responsável por extrair os dados das bases operacionais (transacionais) da organização, por efetuar transformações a fim de gerar informações válidas para a análise e apoio ao processo decisório e, por último, por armazená-las em um repositório que facilite o acesso às informações.
- Repositório de dados analíticos: são representados pelos Data Warehouses (DW), repositórios de dados que utilizam modelagens (geralmente modelagem dimensional) que podem dispor os dados de maneira mais natural para a análise e o processo de decisão.
- O último componente é a camada de apresentação. Essa camada pode utilizar uma série de técnicas e/ou ferramentas para auxiliar o consumo e apresentação das informações armazenadas pelo DW.

A figura a seguir apresenta esta arquitetura com seus componentes e técnicas:

Figura 1 – Arquitetura básica de um processo de BI.



Fonte: wikimedia.org.

Na primeira camada da figura, podemos ver as diversas fontes de dados operacionais (transacionais) da organização. Nessa camada estão todas as bases que serão utilizadas, que podem vir de diferentes fontes e de diferentes setores.

O processo de ETL recupera os dados dispersos pelas várias bases operacionais das organizações e, após o processamento, os concentra na forma de informações analíticas no data warehouse da organização.

Os dados organizados no repositório podem ser consumidos pela camada de apresentação, que é representada pelas várias técnicas de ferramentas: relatórios, OLAP, Dashboards, Alertas, Scorecards, etc.

É importante ressaltar que não existe solução de BI genérica. Cada organização possui um cenário particular e, mais importante ainda, perguntas específicas que procuram respostas específicas. O há em comum em todas as organizações é a arquitetura geral tradicional de BI.

Armazém de Dados ou Data Warehouse

Um data warehouse (DW) é um conjunto de dados produzido para oferecer suporte à tomada de decisões. É também um repositório de dados atuais e históricos, disponível para os responsáveis pelas empresas ou instituições. Os dados normalmente são estruturados de modo a estarem disponíveis em um formato pronto para as atividades de processamento analítico (ex.: processamento analítico online [OLAP], data mining, consultas, geração de relatórios, outras aplicações de suporte à decisão). As características fundamentais de um DW serão apresentadas a seguir.

Características do Data Warehouse

Uma forma de apresentar o data warehouse é recorrer às suas características fundamentais:

- **Orientado por assunto:** os dados são organizados por assunto, como “vendas”, “produtos” ou “clientes”, e contém apenas as informações relevantes ao suporte à decisão. A orientação por assunto permite que os usuários determinem não só como está o desempenho de sua empresa, mas também o porquê desse desempenho. Um data warehouse difere de um banco de dados operacional porque esses, em sua maioria, são orientados por produto e ajustados para lidar com transações que atualizem o banco de dados. A orientação por assunto de um data warehouse proporciona uma visão mais abrangente da organização.
- **Integrado:** a integração está bastante ligada à orientação por assunto. Os data warehouses devem colocar os dados de diferentes fontes em um formato consistente. Para isso, devem enfrentar conflitos de nomenclaturas e discrepâncias entre unidades de medida. Espera-se que um data warehouse seja totalmente integrado.
- **Variável no tempo (série temporal):** um data warehouse mantém dados históricos. Eles detectam tendências, variações, relações, comparações e de longo prazo para previsão que auxiliam na tomada de decisões. Há uma qualidade temporal para cada data warehouse. O tempo é uma dimensão importante à qual todo data warehouse deve oferecer suporte. Os dados de análise vindos de diversas fontes contém diversos pontos de tempo (ex.: visualizações diárias, semanais, mensais).
- **Não-volátil:** após os dados serem inseridos em um data warehouse, os usuários não podem alterá-los ou atualizá-los. Os dados obsoletos são descartados e as alterações são registradas como dados novos. Isso permite que o data warehouse seja ajustado quase exclusivamente para o acesso a dados.

Data Marts

Um data warehouse une bancos de dados de toda uma empresa. Já um data mart normalmente é menor e se concentra em um assunto ou departamento específico. Um data mart é um subconjunto de um data warehouse que normalmente consiste em uma única área temática (ex.: marketing, operações) e pode ser **dependente** ou **independente**.

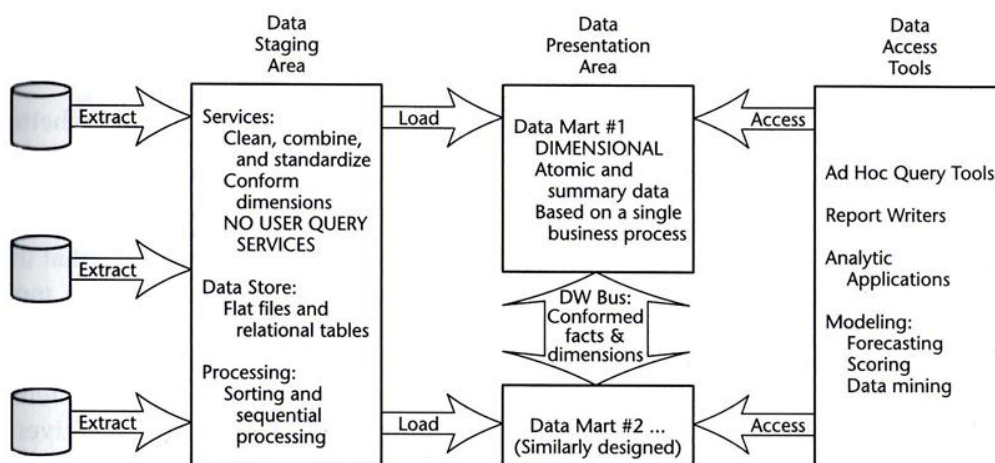
Um data mart **dependente** é um subconjunto criado diretamente a partir do data warehouse. Ele tem a vantagem de usar um modelo de dados consistente e de apresentar dados de qualidade. Os data marts dependentes suportam o conceito de um único modelo de dados em toda a empresa, mas o data warehouse deve ser construído antes. Eles garantem que o usuário final visualize a mesma versão de dados acessada por todos os outros usuários do data warehouse. O alto custo deste último limita seu uso às grandes empresas. Como alternativa, muitas empresas usam uma versão de data warehouse reduzida em custo e escala, denominada data mart independente.

Um data mart **independente** é um warehouse pequeno, projetado para uma unidade estratégica de negócios (UEN) ou um departamento.

Estrutura do Data Warehouse

A figura abaixo apresenta a estrutura do Data Warehouse, segundo Kimball Group:

Figura 2 – Estrutura de um Data Warehouse.



Fonte: Kimball Group.

Data Sources: é de onde buscamos e extraímos os dados necessários para análise dentro do sistema de BI. Uma solução de BI consiste em promover a integração de dados a partir de diversas fontes, tais como Excel, Access, SQL Server e outros.

Data Staging Area: parte do Data Warehouse responsável pelo armazenamento e pela execução de um conjunto de processos normalmente denominados como extração, transformação e Carga (ETL – extract, transformation, load) dos dados. A área de Staging encontra-se entre os sistemas operacionais e a camada de apresentação. É considerada a “cozinha do restaurante”, que está fora do acesso dos usuários.

Data Presentation Area: é onde os dados estão organizados, armazenados e disponíveis para responder às consultas dos usuários em um formato dimensional. A modelagem dimensional é uma técnica de modelagem de dados voltada especialmente para a implementação de um modelo que permita a visualização de dados de forma intuitiva e com altos índices de performance na extração desses. Normalmente nos referenciamos à camada de apresentação como uma série de data marts integrados orientados à processos de negócio.

Capítulo 2. Maturidade para BI

Atualmente, existem tecnologias e ferramentas de BI modernas e voltadas para o futuro. Porém, algumas empresas ainda possuem métodos de gestão voltados para o passado. Isso poderá acarretar o fracasso de projeto ou a subutilização do potencial do sistema. As ferramentas são apenas uma parte do processo; é preciso que elas recebam apenas o foco necessário, e que haja também uma cultura voltada para a decisão com base em informações e a utilização dos recursos disponíveis.

A maioria dos projetos de BI não obtém sucesso por não terem a maturidade necessária. De acordo com pesquisa do IDC (2011), 70% das aplicações de BI no Brasil ainda se referem à geração de relatórios, o que demonstra a falta de maturidade e a falta de foco dos programas de BI.

Moore (2003) define modelo de maturidade como sendo uma estrutura para caracterizar a evolução de um sistema de um estado menos ordenado e menos efetivo para um estado mais ordenado e altamente eficaz. Isso evidencia que há graus evolutivos de maturidade no processo.

As empresas necessitam, portanto, de medir qual seu nível de maturidade em diversos processos de trabalho para poder assim planejar as suas ações e obter sucesso nos processos de inovação. Existem vários modelos para analisar o nível de maturidade de uma empresa. Nesta disciplina vamos abordar o Modelo TDWI, que já é consolidado e largamente utilizado.

O modelo TDWI

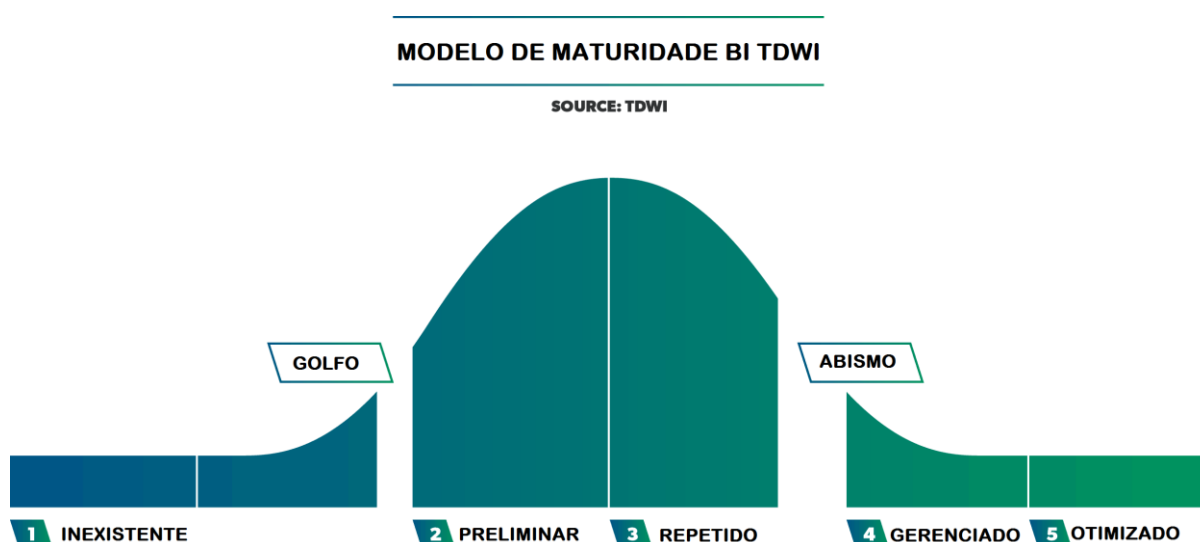
The Data Warehouse Institute (TDWI) foi fundado em 1995, e é o instituto educacional para BI e DW. Ele se dedica à educar profissionais e executivos de todo o mundo, interessados em informações sobre as estratégias, técnicas e ferramentas necessárias para projetar, construir e manter soluções de BI e DW. O modelo TDWI de maturidade em BI foi criado em 2004 com o objetivo de auxiliar os profissionais e

empresas a entender melhor os programas de BI, proporcionar a possibilidade de se comparar com outros negócios e traçar metas para melhorar a maturidade.

O modelo de avaliação TDWI mede a maturidade de um programa de BI de forma objetiva, representado em uma escala horizontal por onde passam as empresas de acordo com a evolução do seu domínio do processo de BI. Essa escala horizontal consiste em cinco estágios: inexistente, preliminar, repetido, gerenciado e otimizado. Além desses cinco estágios, o modelo apresenta dois grandes obstáculos que são o golfo e o abismo.

A figura abaixo apresenta estes estágios que serão detalhados posteriormente.

Figura 3 – Modelo de Maturidade BI TDWI.



Fonte: Guia Benchmark TDWI

Os cinco estágios do modelo de maturidade TDWI

Estágio 1 – Inexistente

Este estágio é formado por duas fases que estão interligadas: **relatórios operacionais** e **spreadmarts**.

A primeira fase, **relatórios operacionais**, é caracterizada por relatórios obtidos em sistemas operacionais. São relatórios estáticos e inflexíveis com uma variedade limitada de dados. Para se obter mais dados, é necessário solicitar customizações ao departamento de TI, que pode demorar para liberá-los. Nesse caso, a obtenção de informações para a tomada de decisões é difícil e limitada.

A segunda fase, **spreadmarts**, surge devido à falta de flexibilidade da primeira fase. Ela permite que os próprios usuários consigam resolver os problemas fazendo uso de aplicativos, geralmente planilhas do Excel, para coletar, tratar, transformar e exibir os dados. Realizam, assim, a função de um Data Mart ou Data Warehouse. O problema é que estes relatórios são desenvolvidos pelos analistas de negócio que desperdiçam muito tempo e definem os termos e as métricas de acordo com suas próprias visões e pontos de vista.

O golfo

O golfo é o primeiro obstáculo enfrentado. A maioria das organizações coloca uma base forte no estágio preliminar e constrói seu primeiro Data Mart. No entanto, uma combinação de falta de planejamento, problemas de qualidade de dados, resistência cultural e proliferação de spreadmarts impede a organização de fazer uma passagem limpa. Para superar o golfo, a organização precisa enfrentar os seguintes desafios:

- **Percepções executivas:** executivos nesse estágio igualam BI a relatórios operacionais e não se preocupam de gastar mais dinheiro na entrega de informações.
- **Financiamento adequado:** muitos projetos de BI são financiados por um chefe de departamento e são vulneráveis ao corte do orçamento, principalmente se o patrocinador muda de trabalho ou função.
- **Má qualidade dos dados:** muitos projetos de BI superestimam a qualidade dos dados do sistema de origem.

- **Aumento do escopo:** projetos iniciais de BI tendem a correr atrás do cronograma e acima do orçamento, em grande parte por causa dos problemas de qualidade dos dados acima mencionados. A chave é manter um rígido controle do projeto, mas interagir rapidamente para satisfazer as necessidades e mudanças do negócio.
- **A proliferação de spreadmarts:** o maior desafio no golfo são as pessoas. Como você afastará as pessoas de suas spreadmarts? Como você muda antigos hábitos das pessoas, formas de acesso e análise de informações e tomada de decisões? É necessário um robusto ambiente e ferramenta de BI e uma abordagem de várias frentes de gestão para proporcionar uma mudança organizacional e cultural.

Estágio 2 – Preliminar

O estágio preliminar é o momento em que a organização se insere em um ambiente de BI e DW. A organização adquire suas primeiras ferramentas de BI, principalmente de consultas, de relatórios e de ferramentas de processamento analítico (OLAP). Alguns usuários começam a usar as ferramentas para analisar tendências e dados históricos para ajudar a empresa a melhorar os planos e operações. A ênfase está na busca de insights, aumentando a consciência e compreensão de como o negócio foi executado no passado. A iniciativa é de âmbito departamental e geralmente um projeto único, sem precedentes ou processos estabelecidos com planejamento de projeto.

Estágio 3 – Repetido

Neste estágio, o trabalho passa a ser mais amplo, saindo de Data Marts não integrados para um modelo mais consolidado em um único DW. Isso gera economia, maior consistência nas informações utilizadas, maior análise do negócio e conseqüentemente maior compreensão dos dados. Neste estágio, é utilizado um

modelo padrão de projeto e desenvolvimento de metodologias que incorpora as melhores práticas adquiridas com iniciativas anteriores e consultores externos.

Também cresce o uso do BI entre usuários casuais, que utilizam de informações em conjuntos padrões de relatórios parametrizados, ou dashboards criados pelos usuários mais capacitados.

O abismo

Este é um obstáculo mais amplo e mais profundo do que o golfo e vários projetos acabam aqui. Desafios a serem superados:

- **Volatilidade do negócio:** mudanças estratégicas na empresa (como fundir-se à outra), novos executivos ou outras reestruturações podem comprometer o BI. Agilidade nos processos, arquitetura flexível e alinhamento de TI ao negócio são fundamentais para a gestão da volatilidade dos negócios.
- **Padronização:** conciliar termos, definições e regras em diversos Data Marts e DW é uma dificuldade política. Cada unidade de negócio e departamento vê o mundo através de sua própria ótica, e fica aos líderes empresariais a missão de definir regras para termos comuns e métricas. Essa é uma tarefa quase impossível sem um CEO que obriga todos a conciliarem seus diferentes pontos de vista para o interesse comum da organização.
- **Transição para TI corporativa:** é difícil fazer com que os departamentos entreguem suas soluções estimadas para uma área de TI corporativa. Os departamentos temem que a TI corporativa tente impor seus padrões e suas soluções, efetivamente matando momentos decisivos de seus projetos.
- **Caos de Relatórios:** muitos departamentos de TI acreditam que o objetivo do BI é capacitar os usuários para criar seus próprios relatórios, e com isso tirar a demanda de trabalho do TI na criação de relatórios. Isso pode levar ao caos de relatórios, pois usuários avançados e alguns casuais criam centenas (senão milhares) de relatórios. A maioria deles são variações sobre o mesmo assunto

e raramente são usados. O resultado é que os usuários não conseguem encontrar os relatórios certos e, frustrados, param de usar o recurso de BI completamente.

- **Evitar inflexibilidade de arquitetura:** diferente dos sistemas operacionais, as soluções de BI têm de se adaptar continuamente às mudanças. As solicitações dos usuários e as necessidades mudam de semana para semana e de mês para mês, devido ao ambiente dinâmico em que as empresas estão inseridas. Por isso, a arquitetura de BI precisa ser flexível para responder solicitações de mudanças dos usuários e para apoiar processos de desenvolvimento ágeis, que possam criar novos aplicativos rapidamente.

Estágio 4 – Gerenciado

As principais características das empresas que atingem este estágio são:

- **Arquitetura Unificada:** a base deste recurso é uma arquitetura unificada de DW que define o conjunto comum de regras, termos e métricas compartilhadas entre unidades de negócio e departamentos.
- **Totalmente carregado:** DW totalmente carregado com todos dados necessários para os trabalhos. Para atender novas demandas, simplesmente adaptam-se os dados existentes, sem a necessidade de se extrair e modelar novas fontes.
- **Flexível:** os desenvolvedores isolam a arquitetura, o que permite que a realização de mudanças em um componente não afete as outras partes.
- **Entrega *just in time*:** o DW integra-se com dados em tempo real para suportar aplicações operacionais que requerem uma entrega *just in time* da informação analítica.

- **Gerenciamento de desempenho:** o estágio gerenciado implementa scorecards em cascata para otimizar a execução da estratégia de negócios em todos os níveis da hierarquia corporativa.
- **Análise preditiva:** no estágio gerenciado as organizações começam a usar previsões mais sofisticadas e ferramentas de modelagem para antecipar, em vez de reagir aos movimentos do negócio.
- **Gestão centralizada:** organizações em fase adulta criam um grupo de gestão da informação que consolida toda a central de informações das áreas, tais como BI, DW, gerenciamento de conteúdo, análise preditiva e sistemas de informações geográficas.

Estágio 5 – Otimizado

Neste estágio, as atividades de desenvolvimento voltam para as unidades de negócio através de centros de excelência. As principais características deste estágio são:

- **Desenvolvimento federado:** para acelerar o desenvolvimento e se adaptar rapidamente às necessidades empresariais em constante mudança, o estágio otimizado redistribui algumas tarefas de desenvolvimento para as unidades de negócio e para os departamentos. Estes devem seguir processos, procedimentos e normas para a entrega de soluções de BI e DW, o que só é possível quando a redistribuição parte de um ambiente centralizado.

Assim, as equipes de BI precisam se centralizar para melhorar a sua maturidade, conforme visto nos estágios anteriores, e, no estágio otimizado, redistribuir algumas atribuições para as unidades.

- **Empresa ampliada:** uma das características principais do estágio. As organizações utilizam BI e DW para oferecer aos clientes e fornecedores personalização, relatórios interativos, dashboards e outros serviços de informação. O BI torna-se fundamental na geração de receitas e na obtenção

vantagem competitiva, consolidando relações com fornecedores. A empresa enxerga a importância de manter os níveis elevados de atendimento ao cliente, e o número de usuários interessados em utilizar a ferramenta dispara. Todos esses fatores fazem com que o programa de BI receba investimentos significativos para criar e manter uma solução forte de BI, sempre atendendo os níveis de serviço exigidos.

- **Serviço de BI:** no estágio otimizado, as organizações utilizam serviços orientados de arquitetura para acelerar o desenvolvimento de soluções de BI. Integrando funcionalidades e modelos de consulta com interface web service, os desenvolvedores podem fazer recursos de BI disponíveis para qualquer aplicação independentemente da plataforma em que é executada ou da linguagem de programação que utiliza. Então, desenvolvedores aprovados dentro ou fora da organização podem gravar aplicativos que utilizam vários componentes inseridos no BI ou no serviços de dados. Normalmente, estes aplicativos estão em um portal que exibe gráficos ou KPIs gerenciados por um servidor remoto de BI. Os aplicativos são mais sofisticados com previsões e alertas em tempo real. Combinam o melhor dos componentes operacionais e analíticos.

As dimensões do modelo de maturidade TDWI

Além dos cinco estágios evolutivos, o Modelo TDWI de maturidade em BI compreende oito dimensões que estão inseridas em cada um dos cinco estágios. Abaixo, temos uma breve abordagem de cada uma das dimensões.

Escopo

O escopo se refere à amplitude com que a equipe de BI da empresa atua e implementa os projetos. Barbieri (2011) assim classifica as áreas de escopo:

- Aplicações pessoais, com desenvolvimento para a própria área.

- Departamental local, com o desenvolvimento para um departamento dentro de uma unidade de negócios.
- Departamental empresa, com desenvolvimento para um departamento que atende a várias unidades de negócio.
- Unidade de negócio, desenvolvendo para todos os departamentos dentro de uma unidade de negócios.
- Corporativo, para todas para ou várias unidades de negócios.
- Intercorporativo, desenvolvendo para todas as unidades de negócios e mais clientes e fornecedores, ou seja, ultrapassando os limites da empresa e alcançando uma espécie de BI colaborativo, envolvendo parceiros de negócios.

O escopo de atuação da equipe dá sinais de maturidade através da aceitação e aderência por parte das unidades corporativas. Além disso, a forma pela qual os executivos percebem o núcleo de BI na empresa é fundamental, pois dá a ideia do retorno por parte dos usuários com relação ao papel desempenhado.

Patrocinador

Esta dimensão demonstra o comprometimento das áreas e executivos com o projeto. Conforme Barbieri (2011),

“Pode ser o CIO ou um diretor de TI, ou um patrocinador de uma unidade de negócio/departamento, ou múltiplos patrocinadores oriundos de várias unidades de negócios ou, finalmente, múltiplos níveis de comitês de supervisão pluridepartamental orientados a negócios.”

Esse pode ser considerado um dos fatores mais importantes na consolidação de uma área de BI. O patrocinador é a pessoa que demonstra o grau de comprometimento da empresa com relação aos resultados obtidos.

Apoio financeiro

Nesta dimensão se avalia qual é a disponibilidade de recursos que a empresa dispõe para um projeto de BI. Inicialmente, principalmente durante o entusiasmo inicial, os projetos de BI tendem a obter investimentos com facilidade, principalmente se houver um bom patrocinador. Porém, com o andamento do projeto, a tendência é que a disponibilidade de recursos para os projetos de BI se tornem mais difícil, principalmente se não são demonstrados resultados. Barbieri (2011) destaca o seguinte:

“Nesta dimensão, a empresa é avaliada na sua maturidade, através do tipo de orçamento que dispõe, como orçamento departamental ou de unidade de negócios, ou orçamento corporativo ou recursos próprios. Também a facilidade ou esforço com que o orçamento é obtido denotam o grau de maturidade da empresa em termos de BI.”

Retorno

O retorno representa os resultados que um programa de BI proporciona para a empresa através de valores quantificáveis ou retornos intangíveis. É medido pelo valor que os usuários percebem em relação ao programa e é graduado da seguinte forma por Barbieri (2011),

“Irrelevante para as suas funções, com valor relativamente marginal, pois o BI tangencia suas funções; relevante, pois são impactados pelas ações do BICC (BI Competency Center); crítico ou considerado fator-chave de sucesso, quando dependente em alto e muito alto grau respectivamente”.

Arquitetura

Para Barbieri (2011), a arquitetura representa a forma mais usual como as soluções de BI são implantadas. Na dimensão de arquitetura é onde aparecem as spreadmarts, criadas e mantidas por necessidades pontuais, sempre no nível inicial de maturidade de BI. Já para um nível um pouco mais avançado, temos os Data Marts não integrados, que são iniciativas isoladas, criados para atender a um único

departamento. Temos ainda na arquitetura o DW não integrado e o DW integrado, e por fim um serviço de BI, que usa uma interface comum para acessar via web services ou camada de EII (Enterprise Information Integrator) os recursos de informação.

Dados

Na dimensão de dados temos uma avaliação um pouco ampla, considerando desde a disponibilidade e confiabilidade dos dados até a facilidade para acessá-los. A acessibilidade dos dados pode ir de alta dificuldade até a alta acessibilidade, quando os usuários conseguem acessar facilmente todos os dados necessários.

A confiabilidade dos dados carregados no BI é também fator importante nessa dimensão. Conforme Barbieri (2011), a qualidade dos dados deverá, cada vez mais, ser considerada como elemento-chave nos insumos de dados usados nos processos de BI. É essa qualidade que garante a confiança dos usuários para a utilização das informações. Na avaliação de qualidade dos dados também é considerada a frequência de atualização das informações e com quais fontes de informações o BI está sincronizando. Este ponto avalia a extração, transformação e a carga dos dados no BI e no DW.

Desenvolvimento

Essa dimensão considera a forma como a empresa realiza o desenvolvimento das aplicações de BI. Barbieri (2011) define essa dimensão com a seguinte escala:

“A graduação vai de desenvolvimento isolado de aplicações ad hoc, usando as ferramentas necessárias, independentemente de estarem alinhadas com padrões definidos, ou de aplicações isoladas, porém usando o conjunto de ferramentas alinhadas com as definições do BICC, ou de portfólios de aplicações integradas usando os processos e as ferramentas padrão, alinhadas com a arquitetura definida, e, finalmente, na forma federada, permitindo que as unidades de

negócios desenvolvam suas aplicações desde que respeitando os padrões, os processos e a arquitetura corporativa definida”.

Nessa dimensão, também é avaliado o nível de acompanhamento e de gerenciamento dos projetos da equipe de desenvolvimento do BI, levando em conta o tempo de implantação, o número de projetos executados e o grau de padrões de desenvolvimento e de documentação.

Entrega

Pode-se dizer que essa dimensão avalia o produto do programa de BI, dando a ideia do grau de intensidade em que este é aplicado. Conforme Barbieri (2011),

“Pode variar entre a produção de relatórios para consumo organizacional ou a possibilidade de análise de tendências, ou a monitoração de eventos de negócios para possibilitar ações proativas, ou fazer previsão de resultados e modelar planos, e, finalmente, automatizar processos, incluindo interações com clientes”.

A entrega dos dados também pode ser avaliada segundo o perfil de usuários. A quantidade de usuários constantes ser maior do que os usuários casuais demonstram a maturidade de um programa de BI. Além disso, o aumento do número de usuários que utilizam o sistema acessando os dados ou mesmo recebendo relatórios padrões ou dashboards parametrizados demonstra a intensidade com a qual a empresa aplica o BI corporativo.

É importante destacar que as dimensões do modelo TDWI de maturidade em BI podem ser classificadas com instruções de valor para cada um dos estágios de maturidade. A figura abaixo apresenta um resumo dos estágios de maturidade para cada uma das dimensões.

Figura 4 – Instruções de valor do modelo de maturidade TDWI.

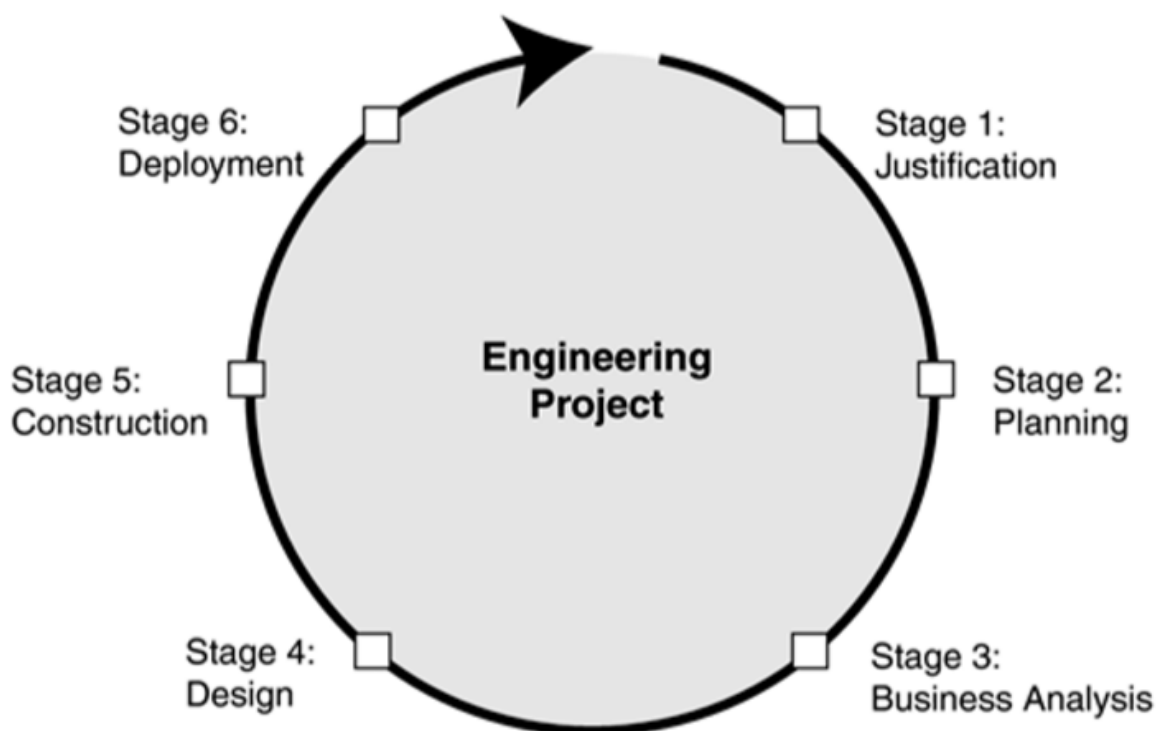
Categoria/Estágio	Inexistente	Preliminar	Repetido	Gerenciado	Otimizado
Escopo	Individual	Departamento	Divisão	Empresa	Inter-empresarial
Patrocinador	Inexistente ou descomprometido	↔	Um pouco comprometido e responsável	↔	Muito comprometido e responsável
Apoio Financeiro	Nenhum	Orçamento Departamental	Orçamento da Divisão	Orçamento de TI Corporativa	Auto-financiamento
Retorno	Centro de Custo	Tático	Missão Crítica	Estratégico	Diferencial competitivo
Arquitetura	Spreadmarts	Data Marts não integrados	Data Warehouses não integrados	Central DW com ou sem data marts	BI ou serviço de dados através de arquitetura orientada para serviços
Dados	Não Confiável, não oportuno, não completo	↔	Um pouco confiável, oportuno, e completo	↔	Totalmente confiável, oportuno, e completo
Desenvolvimento	Processos não padronizados	↔	Processos um pouco padronizados	↔	Processos totalmente padronizados
Entrega	Visualizar relatórios estáticos	Analisar tendências e problemas	Acompanhar os processos	Prever os resultados	Automatizar os processos

Fonte: Guia Benchmark TDWI

Capítulo 3. Etapas do Projeto de BI

Neste capítulo serão apresentadas as etapas de todo processo de BI, desde sua concepção até a entrega, que geralmente são divididos em 6 etapas iterativas, que como podemos verificar abaixo.

Figura 5 – Etapas do projeto de BI.



Fonte: Business Intelligence Roadmap.

Justificativa

Nesta fase, são avaliadas as necessidades de negócios que dão origem ao novo projeto.

O problema de negócio ou oportunidade de negócio é definido, e uma solução de BI é proposta. Cada versão do aplicativo de BI deve ser justificada por custos e deve definir claramente seus benefícios.

Planejamento

Nesta fase são desenvolvidos planos estratégicos e táticos, que estabelecem como o projeto de engenharia será realizado e implantado.

Os projetos de apoio a decisão de BI são extremamente dinâmicos. Mudanças no escopo, na equipe, no orçamento, na tecnologia, nos representantes empresariais e nos patrocinadores podem afetar severamente o sucesso de um projeto. Portanto, o planejamento do projeto deve ser detalhado, e o progresso real deve ser observado e relatado. Devido à importância deste tópico, ele será tratado com detalhes posteriormente.

Outro ponto que deve ser planejado é a infraestrutura corporativa. Alguns componentes de infraestrutura já podem estar em vigor antes do lançamento do primeiro projeto de BI. Outros componentes podem ter que ser desenvolvidos ao longo do tempo, como parte dos projetos de BI. Uma infraestrutura empresarial possui dois componentes:

- Infraestrutura técnica, que inclui hardware, software, middleware, sistemas de gerenciamento de banco de dados, sistemas operacionais, componentes de rede, repositórios de metadados, utilitários, e assim por diante.
- A infraestrutura não-técnica, que inclui padrões de metadados, metodologias, diretrizes, procedimentos de teste e de gerenciamento de problemas, processos de controle de mudanças, e assim por diante.

Análise

Nesta fase, são executadas análises detalhadas do problema ou da oportunidade de negócio para se obter uma sólida compreensão dos requisitos para uma solução potencial (produto).

Definir o escopo é uma das tarefas mais difíceis nos projetos de suporte à decisão de BI e um dos aspectos mais importantes da negociação dos requisitos para

cada entrega. As equipes devem esperar que esses requisitos mudem ao longo do ciclo de desenvolvimento, pois os empresários aprendem mais sobre as possibilidades e as limitações da tecnologia de BI durante o projeto.

Outro desafio para todos os projetos de suporte à decisão de BI é a análise da qualidade dos dados de origem. Maus hábitos podem ter sido desenvolvidos e são difíceis de quebrar, e os danos resultantes são muito caros, demorados e tediosos de se encontrar e se corrigir. Esse passo leva uma porcentagem significativa do tempo atribuído a todo o cronograma do projeto.

A análise dos resultados funcionais, que costumava ser chamada de análise do sistema, é melhor feita através da prototipagem para poder ser combinada com o design da aplicação. Novas ferramentas e linguagens de programação permitem aos desenvolvedores provar ou refutar de forma relativamente rápida um conceito ou uma ideia. A prototipagem também permite que os empresários vejam o potencial e os limites da tecnologia, o que lhes dá a oportunidade de ajustar os requisitos do projeto e suas expectativas.

Ter mais ferramentas significa ter mais metadados técnicos, além dos metadados do negócio, que geralmente são capturados em uma ferramenta de modelagem de engenharia de software auxiliada por computador. Os metadados técnicos precisam ser mapeados para os metadados do negócio, e todas as metadados devem ser armazenadas em um repositório de metadados.

Design

Nesta fase, é concebido um produto que resolva o problema ou permita a oportunidade de negócio em questão.

Um ou mais bancos de dados de destino BI armazenarão os dados comerciais de forma detalhada ou agregada, dependendo dos requisitos de relatórios da comunidade empresarial. Nem todos os requisitos de relatórios são estratégicos, e nem todos são multidimensionais.

O processo ETL é o processo mais complicado de todo o projeto de suporte à decisão de BI. As janelas de processamento ETL são tipicamente pequenas, e, no entanto, a má qualidade dos dados de origem geralmente faz com que muitas horas sejam necessárias para a execução dos programas de transformação de limpeza. Finalizar o processo ETL dentro da janela de tempo disponível é um desafio para a maioria das organizações.

Um repositório de metadados deverá ser concebido e o design deve atender aos requisitos do meta-modelo lógico.

Construção

Nesta fase é criado o produto, que deve fornecer um retorno sobre o investimento dentro de um período de tempo predefinido.

Muitas ferramentas estão disponíveis para o processo ETL, algumas sofisticados e algumas simples. Dependendo dos requisitos de limpeza e da transformação de dados verificados na fase de design, uma ferramenta ETL específica pode ou não ser a melhor solução. De qualquer forma, o processamento prévio das extensões de dados e de escrita é frequentemente necessário.

Uma vez que o esforço de prototipagem firmou os requisitos funcionais, o verdadeiro desenvolvimento da aplicação de acesso e de análise pode começar. Desenvolver o aplicativo pode ser uma questão simples de finalizar um protótipo operacional, ou pode ser um esforço de desenvolvimento mais envolvente que necessita de ferramentas de análise e de acesso diferentes, mais robustas.

Muitas organizações não utilizam ao máximo o seu ambiente de apoio à decisão. Os aplicativos de BI geralmente são limitados à relatórios pré-escritos, alguns dos quais nem sequer são novos tipos de relatórios, mas substituições de relatórios antigos. O retorno real vem da informação escondida nos dados da organização, que só pode ser descoberta com ferramentas de mineração de dados.

Se for decidido construir um repositório de metadados, ao invés de licenciar um, uma equipe separada geralmente é responsável pelo processo de desenvolvimento. Isso se torna um subprojeto considerável no projeto geral de BI.

Implantação

Nesta fase, é implementado ou vendido o produto finalizado. Em seguida, é medida sua efetividade para determinar se a solução atende, excede ou não consegue atender o retorno esperado.

Uma vez testados todos os componentes do aplicativo de BI, a equipe libera os bancos de dados e os aplicativos. O treinamento é agendado para a equipe de negócios e para outras partes interessadas que usarão o aplicativo de BI e os repositórios de metadados. Começam as funções de suporte, que incluem helpdesk, a manutenção dos bancos de dados de destino do BI, o agendamento e a execução de trabalhos de ETL, o monitoramento de desempenho, entre outros.

No contexto de lançamento de aplicativos, é muito importante aproveitar as lições aprendidas com os projetos anteriores. Quaisquer prazos perdidos, excessos de custos, disputas e suas respectivas resoluções devem ser examinados, e os ajustes do processo devem ser feitos antes da próxima versão começar. Todas as ferramentas, técnicas, diretrizes e processos que não foram úteis devem ser reavaliados e ajustados, ou até descartados.

As etapas do desenvolvimento de um projeto de BI são específicas em cada projeto e algumas podem ser realizadas em paralelos ou obter maior ou menor atenção, o que é definido pela equipe ou pelo responsável.

Capítulo 4. Planejamento do Projeto de BI

Os projetos de BI não são como outros projetos com um conjunto finito e estático de requisitos de uma pessoa de negócios ou de um departamento. O objetivo de um ambiente integrado de suporte à decisão de BI é fornecer capacidades de análise de negócios para todos os empresários e todos os departamentos da organização. Isso envolve uma variedade de novas tarefas, papéis e responsabilidades, e uma abordagem mais prática de gerenciamento de projetos.

Alguns pontos devem ser observados em um projeto de BI:

- **Envolvimento do Negócio**
 - Temos um bom patrocinador? Temos um patrocinador de negócios de backup?
 - Temos partes interessadas com quem precisamos nos comunicar regularmente?
 - Quanto tempo o representante da empresa compromete com este projeto? Está atribuído a este projeto em tempo integral, ou estará disponível somente mediante solicitação?
- **Escopo do projeto**
 - Recebemos um pedido formal para um projeto de BI?
 - O quão detalhados são os requisitos?
 - O que foi solicitado para ser entregue?
 - Podemos implementar o escopo solicitado, considerando o cronograma e os recursos disponíveis?
- **Análise de custo-benefício**
 - Já realizamos uma análise custo-benefício?

- Qual é o retorno esperado sobre o investimento (ROI)?
- Em que momento esperamos que o ROI se materialize?
- A infraestrutura
 - Revisamos nossos componentes de infraestrutura técnica e não técnica?
 - Nossa infraestrutura possui problemas?
 - Quais componentes de infraestrutura devemos trabalhar e entregar como parte do projeto de BI?
 - Quais componentes de infraestrutura técnica?
 - Quais componentes de infraestrutura não técnica?
- Pessoal e Habilidades
 - Já identificamos os membros da equipe?
 - Todos os membros da equipe têm as habilidades necessárias para desempenhar as responsabilidades de suas funções atribuídas?
 - Devemos agendar qualquer treinamento antes do lançamento do projeto?
 - O gerente de projeto é atribuído à este projeto em tempo integral? Ou tem outras responsabilidades administrativas?

Gerenciamento do Projeto de BI

O gerenciamento de projetos, na maioria das organizações, é tratado como uma função de relatório administrativo. O planejamento detalhado do projeto e o controle direto do projeto diário geralmente são minimizados, se não forem ignorados, especialmente quando as organizações tentam colocar vários aplicativos de BI

funcionando rapidamente. As organizações esquecem que um bom planejamento muitas vezes leva a ciclos de testes e implementação mais curtos e, portanto, a um tempo de entrega também mais curto.

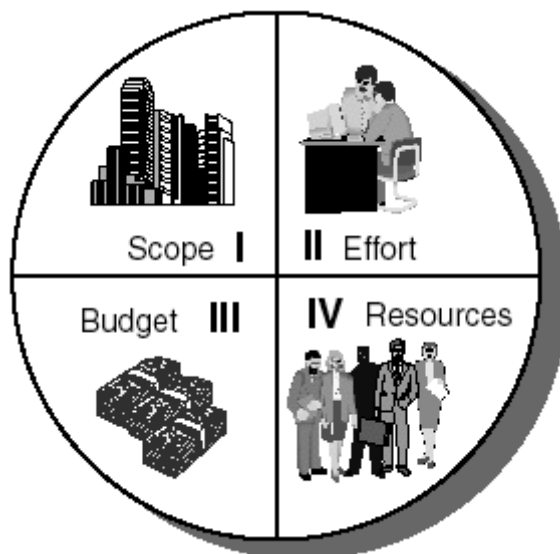
Difícilmente um projeto de BI é realizado sem algumas adaptações, e os atrasos são comuns. Pode faltar algum produto, pode haver trocar de pessoal, pode haver problemas com fornecedores, dentre outros.

Muitas organizações não planejam adequadamente esses tipos de atrasos e retrocessos, nem testam seus conceitos e estratégias de BI adequadamente. O planejamento de contratempos ajudará o gerenciamento a estabelecer datas de lançamento realistas para o projeto.

Descrevendo atividades de gerenciamento de projetos em termos mais simplistas, o objetivo é responder a quatro questões básicas. O que será entregue? Quando isso será feito? Quanto vai custar? Quem vai fazer isso?

Essas perguntas se traduzem, respectivamente, nas quatro principais restrições do projeto de escopo: esforço, tempo, orçamento e recursos. Antes que o gerente do projeto possa criar um plano para resolver essas restrições, ele deve passar algum tempo o definindo, para entender claramente os requisitos, riscos, restrições e pressupostos relacionados.

Figura 6 – Principais restrições de um projeto de BI (TDWI).



Fonte: Business Intelligence Roadmap.

Definição do Projeto de BI

O planejamento de projetos inclui a criação de uma carta de projeto, que define: metas e objetivos, escopo, riscos, restrições, premissas, procedimentos de controle de mudanças e procedimentos de gerenciamento de questões.

A carta do projeto é o acordo feito entre o patrocinador comercial e a equipe de TI para desenvolver o aplicativo de BI. Se qualquer de seus componentes muda, todo o projeto deve ser reavaliado e todas as suas restrições devem ser renegociadas. Cada fase será detalhada a seguir.

Metas e objetivos

Ao definir um projeto de BI, o primeiro levantamento a ser feito são as metas e objetivos. Qual o motivo da construção deste aplicativo de BI? Quanta dor (financeira) esse problema de negócios causa? Quais são os objetivos comerciais

estratégicos? Os objetivos do projeto de BI se enquadram nos objetivos de negócios estratégicos, ou o projeto é um desejo pessoal de alguém?

Os objetivos do projeto devem ser declarações mensuráveis, tais como: "Para aumentar a participação de mercado em 10% no próximo ano, o departamento de vendas deve ter acesso a dados de vendas mensais, bem como a dados de pipeline mesclados a dados de prospectos dentro de cinco dias úteis após a fim do ciclo contábil semanal". Os objetivos do projeto devem se relacionar com o ROI esperado. O representante de negócios terá que medir a eficácia do aplicativo de BI entregue e informar ao patrocinador da empresa se o projeto foi bem-sucedido ou não.

Escopo do Projeto

É impossível criar estimativas válidas para um projeto sem uma compreensão sólida do escopo. Nos projetos de BI, o escopo deve ser medido pelo número de elementos de dados que devem ser extraídos dos sistemas de origem, transformados, limpos e carregados nos bancos de dados de destino do BI.

O principal motivo para se concentrar em dados é que analisar e preparar dados de origem leva muito mais tempo do que fornecer acesso a estes e permitir a sua análise através de relatórios e consultas. Normalmente, a regra típica 80/20 aplica-se: 80 por cento de esforço para dados e 20 por cento de esforço para a funcionalidade.

Riscos de projeto

Todo projeto está sujeito a alguns riscos, e os riscos são inevitáveis. Eles podem afetar severamente o cronograma e os resultados do projeto, dependendo da probabilidade de que os estes se materializem e do impacto que terão. Portanto, a avaliação de risco deve ser revisada e expandida, se necessário. O gerente do projeto deve identificar disparadores para cada risco e incorporar um plano de mitigação, bem como um plano de contingência, no plano do projeto.

- Os disparadores são situações que sinalizam um potencial ou uma materialização iminente de um risco. Por exemplo, se a administração está revisando o orçamento para o projeto sem motivo aparente, isso indica um possível gatilho para a perda de suporte de gerenciamento para seu projeto de BI.
- O plano de mitigação especifica quais ações a equipe do projeto pode tomar para evitar que o risco se materialize. Continuando com o exemplo acima, você poderia solicitar o apoio do seu patrocinador comercial e promover a iniciativa de BI para outros executivos importantes da sua organização para manter o interesse da administração no projeto de BI. Caso o projeto tenha problemas, o risco de cancelamento será mitigado ou impedido.
- O plano de contingência especifica alternativas caso o risco se materialize. Por exemplo, se você perder o suporte de gerenciamento para o projeto de BI devido a um longo cronograma do projeto, planeje encurtar os ciclos de lançamento, fornecendo um escopo menor mais cedo. Se você perder o suporte de gerenciamento devido à saída do patrocinador comercial da organização, tenha um patrocinador alternativo pronto para se tornar o novo financiador do projeto de BI.

Alguns riscos de projetos comuns incluem o seguinte: falta de compromisso de gestão, perda do patrocinador, cronograma impensado e irreal, escopo irreal para o cronograma, expectativas irrealistas, orçamento irreal, equipe não treinada ou indisponível, mudanças constantes de negócios, gerenciamento de projeto ineficaz, escalabilidade limitada.

Restrições do projeto

Todos os projetos estão sujeitos às restrições mencionadas anteriormente: escopo, esforço, tempo, orçamento e recursos (pessoas capazes e disponíveis). Na realidade, existe uma quinta restrição: a qualidade. Embora a qualidade seja uma

qualitativa do projeto, ela também pode ser considerada uma restrição que deve ser equilibrada com as outras quatro restrições.

Enquanto todos no lado do negócio e no departamento de TI querem qualidade, raramente o tempo extra para alcançá-la é dado ou realizado. Maior qualidade requer mais esforço e, portanto, um maior tempo para a entrega. Uma vez que os fatores de tempo impulsionam a maioria das organizações, o esforço é a sua restrição número um (prioridade mais alta), seguida de escopo, orçamento e recursos (geralmente nessa ordem); a qualidade é empurrada para o fundo da pilha conforme mostrado na figura abaixo. As restrições do projeto de BI nunca devem estar nesta ordem.

Figura 7 – Ordem de prioridade não desejada das restrições.

Prioridade (Da maior para a menor)					
Restrição	1	2	3	4	5
Esforço (Tempo)	x				
Escopo		x			
Orçamento			x		
Recursos				x	
Qualidade					x

Fonte: Adaptação Business Intelligence Roadmap

Felizmente, as organizações têm controle total sobre mudar a prioridade das restrições do projeto. Insistir que o tempo e o escopo sejam as duas principais restrições é aceitável somente em projetos que tenham requisitos relacionados aos regulamentos impostos pelo governo. É recomendado que você retire qualidade do fundo da pilha e coloque o escopo lá, porque o escopo pode e continuamente será expandido e alterado à medida que o projeto avance. A figura abaixo mostra a ordem recomendada de restrições do projeto.

Figura 8 – Ordem de prioridade recomendada das restrições.

Prioridade (Da maior para a menor)					
Restrição	1	2	3	4	5
Qualidade	x				
Orçamento		x			
Recursos			x		
Esforço (Tempo)				x	
Escopo					x

Fonte: Adaptação Business Intelligence Roadmap.

Premissas

Premissas são hipóteses, condições que assumimos como verdade para alcançar um objetivo exclusivo. Para os processos de planejamento, sempre consideramos as premissas como certas, reais e seguras. Elas devem ser precisas e específicas.

Todas as premissas geram riscos ao projeto. Aqui está um exemplo de como duas premissas em um projeto falharam:

Premissa 1: "O fornecedor promete entregar um novo servidor de banco de dados em maio e, até o final de junho, a equipe de TI instalará e testará um novo produto de sistema de gerenciamento de banco de dados (SGBD) nesse servidor. Isso permite muito tempo antes do prazo final, que é 30 de setembro, o final do ano fiscal".

Premissa 2: "Jonas Maia será o administrador do banco de dados no projeto porque ele é a única pessoa em nossa organização que possui as habilidades específicas necessárias. Ele já se juntou à equipe do projeto".

Problemas: no dia 20 de junho (um mês atrasado), o novo servidor finalmente chega e, no dia 1 de julho, Jonas Maia encerra suas atividades na organização. O novo produto não é instalado e testado até o final de setembro.

Impacto: o projeto está atrasado três meses, com um aumento no orçamento de R\$ 60.000 (grande parte pago como honorários caros de consultoria para preencher a falta de Jonas Maia).

Premissas importantes devem ter riscos de contrapartida, caso alguma situação não saia como pensado, como no exemplo acima. Para cada risco de contrapartida, identifique triggers, um plano de mitigação e um plano de contingência.

Procedimentos de Controle de Mudança

Existe um modelo mental tradicionalista que prega que "A mudança é ruim - os empresários devem ser responsabilizados por suas decisões". Uma vez que os aplicativos de BI devem ser catalisadores para uma melhor tomada de decisão, o modelo mental deve mudar para "Mudança é boa - as pessoas de negócios devem refinar e melhorar suas decisões". No entanto, mudanças descontroladas ainda podem acabar com um projeto.

A solução é gerenciá-las. Muitas organizações acompanham suas solicitações de mudança registrando a data do pedido de alteração, o nome do solicitante, a mudança desejada, para quem foi atribuído e quando foi implementado. Essa é uma boa prática.

Para gerenciar uma mudança, você precisa começar com uma linha de base - o acordo entre o patrocinador comercial e a equipe de TI, conforme documentado na carta do projeto. Toda solicitação de alteração, uma vez registrada, passa por uma análise de impacto de custo-benefício para determinar seus efeitos no projeto. As mudanças, a menos que sejam minúsculas, terão sempre impacto nas restrições de esforço, tempo, escopo e qualidade. Algumas também afetam as outras duas

restrições (orçamento e recursos). Quando uma restrição muda, as restantes devem ser renegociadas.

Isso ocorre porque o representante comercial, o gerente do projeto e os principais membros da equipe acreditavam que poderiam completar o projeto sob as condições acordadas. Quando a restrição do escopo muda, por exemplo, o plano não é mais possível sem a adaptação de algumas outras restrições. Portanto, dependendo de quão crítico é o pedido de mudança, o representante comercial deve decidir entre:

- Reduzir o escopo atual, eliminando alguns dos dados e funcionalidades originalmente solicitados;
- Estender o prazo;
- Declarar a mudança solicitada como inviável no momento e adiá-la;
- Incorporar a alteração solicitada na próxima versão;
- Eliminar transformações complicadas, verificações e testes, o que afetará a qualidade do produto.

Procedimentos de gerenciamento de problemas

Problemas relacionados a questões comerciais ou técnicas sempre surgem durante os projetos. Semelhante aos pedidos de alteração, os problemas devem ser não apenas rastreados, mas também gerenciados. Todo problema deve ser atribuído a uma pessoa que tenha a responsabilidade pela sua resolução. Qualquer atividade relacionada a ele deve ser datada e descrita no registro. No final do projeto, todos os problemas devem ter uma resolução, mesmo esta seja um adiamento para uma futura versão de BI.

Alguns problemas são menores e podem ser resolvidos sem impacto no projeto. Outros podem se transformar em riscos ou alterar solicitações e devem ser

tratados. Portanto, o gerenciamento de problemas inclui análise de impacto e controle de mudanças.

Planejamento do Projeto de BI

O planejamento de projetos não é uma atividade única. Uma vez que um plano de projeto é baseado em estimativas, que muitas vezes não são mais do que as melhores suposições, estes devem ser ajustados constantemente. O primeiro sinal de que um projeto não está sendo propriamente gerenciado é um plano estático no qual as estimativas e os marcos nunca mudaram desde o primeiro dia do desenvolvimento.

Aqui está a sequência de atividades para preparar um plano de projeto:

- Crie uma estrutura de repartição do trabalho listando atividades, tarefas e subtarefas.
- Estime as horas de esforço para essas atividades, tarefas e subtarefas.
- Atribua recursos às atividades, tarefas e subtarefas.
- Determine as dependências da tarefa.
- Determine as dependências de recursos.
- Determine o caminho crítico com base nas dependências.
- Crie o plano detalhado do projeto.

Atividades e tarefas

Os projetos de BI são compostos por muitas atividades, cada uma com uma longa lista de tarefas. O gerente do projeto deve contar com uma lista pré-existente que seja abrangente das atividades mais necessárias. Naturalmente, nem todas as atividades devem ser realizadas em todos os projetos, e nem mesmo cada passo

deve ser realizado em todos os projetos. O gerente seleciona o número mínimo de etapas e atividades necessárias para produzir um produto aceitável sob as restrições impostas.

A abordagem de desenvolvimento aqui proposta não é tão linear nem tão rigorosa como a seguida nas metodologias tradicionais. É uma abordagem muito mais dinâmica para o desenvolvimento de aplicativos de BI, na qual algumas atividades podem aparecer em etapas diferentes para cada caso.

O plano do projeto deve refletir essa natureza dinâmica do desenvolvimento de aplicativos. Uma vez que as mudanças e contratempos são esperados, certas "atividades concluídas" terão de ser revisadas e retrabalhadas. O plano deve antecipar isso e refletir no cronograma. A maneira mais fácil de planejar essas iterações internas é usar o conceito de "looping" ou "refatoração", dividindo o projeto em múltiplos subprojetos pequenos, cada um com um produto, embora não completo. Em seguida, cada entrega é revisada, e são adicionados mais dados e mais funcionalidades até que o aplicativo de BI inteiro seja concluído com a entrega desejada.

Técnicas de estimativa

Uma vez que você selecionou as atividades e tarefas para o projeto e o organizou em subprojetos, você pode derivar as estimativas básicas usando um dos três métodos:

Histórico, com base em padrões aprendido em outros trabalhos;

Intuitivo, baseado na intuição e na experiência;

Fórmula, com base na média de possibilidades, que é calculada da seguinte forma: $[\text{Melhor_Estimativa} + \text{Pior_Estimativa} + (\text{Estimativa_Média} * 4)] / 6$.

Estimar as atividades do projeto de BI é muito mais difícil do que estimar projetos tradicionais porque nenhum BI é igual. Por exemplo, você pode usar uma

nova ferramenta, trabalhar com novos membros da equipe ou não ter experiência com um novo método de design.

As três técnicas de estimativa listadas acima esperam que você tenha alguma experiência anterior de projeto.

A técnica de estimativa histórica espera que você tenha estatísticas sobre quanto tempo os projetos similares foram realizados no passado, mas você talvez não tenha tido um projeto similar antes.

A técnica de estimativa intuitiva espera que você preveja, com base em sua experiência, quanto tempo demorará para completar uma atividade, mas possivelmente você nunca realizou uma atividade similar.

A técnica estimativa baseada em fórmulas espera que você conheça o tempo mais longo, o mais curto e o mais provável para realizar uma atividade, mas você não os saberá se nunca a tiver realizado antes.

Em todos esses casos, é melhor consultar outras pessoas (funcionários internos ou consultores externos) que já desenvolveram um aplicativo de BI similar, porque suas próprias estimativas não educadas podem ser subestimadas. Isso também demonstra a importância de rastrear o tempo real em projetos de BI: você precisará dessas informações para estimar seu próximo projeto.

Atribuição de recursos

As estimativas de esforço não podem ser concluídas até que as atividades e tarefas sejam atribuídas, já que devem levar em consideração as **habilidades** e os **conhecimentos** de cada membro da equipe, bem como os **fatores ambientais** que o afetam.

Habilidades: a capacidade de executar tarefas específicas. O membro da equipe fez esse tipo de trabalho antes?

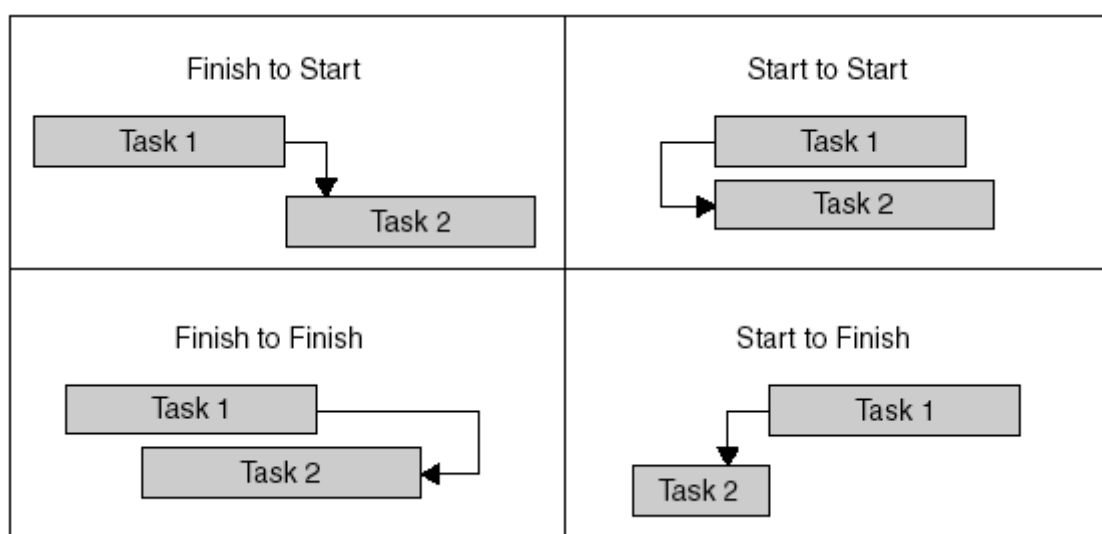
Conhecimentos: a posse de fatos ou conceitos sobre um assunto específico. O membro da equipe é um especialista nesta área de negócios?

Fatores ambientais: atividades administrativas e não relacionadas ao trabalho como férias, falta de computador, doença, reuniões, etc.

Dependências de tarefas

Nem todas as atividades e tarefas devem ser realizadas em série. Muitas podem ser realizadas em paralelo, desde que haja pessoal suficiente. O primeiro passo para determinar quais tarefas podem ser realizadas em paralelo é identificar as dependências entre elas e desenvolver o caminho crítico. A maioria das ferramentas de planejamento de projetos suporta quatro tipos de dependências de tarefas:

Figura 9 – Dependências entre tarefas.



Fonte: Business Intelligence Roadmap.

Término - Início: Nesta situação uma atividade é finalizada para outra ser iniciada. A Atividade 2 para ser iniciada depende que a Atividade 1 seja finalizada.

Início - Início: Este tipo de conexão é utilizada quando duas ou mais atividades iniciam ao mesmo tempo. Atividade 2 não depende da Atividade 1 para iniciar, e ambas serão iniciadas simultaneamente.

Término - Término: Nesta conexão as duas atividades são finalizadas igualmente. A Atividade 1 e a Atividade 2 dependem uma da outra para serem concluídas de maneira uniforme.

Início - Término: Neste caso uma atividade não poderá ser finalizada até que a outra atividade tenha sido começada. A Atividade 2 não pode ser finalizada até que a Atividade 1 tenha sido iniciada.

Quanto mais tarefas forem realizadas simultaneamente, mais rápido o projeto será feito. Para aproveitar as dependências da tarefa, você precisa do número certo de recursos com as habilidades corretas no momento certo.

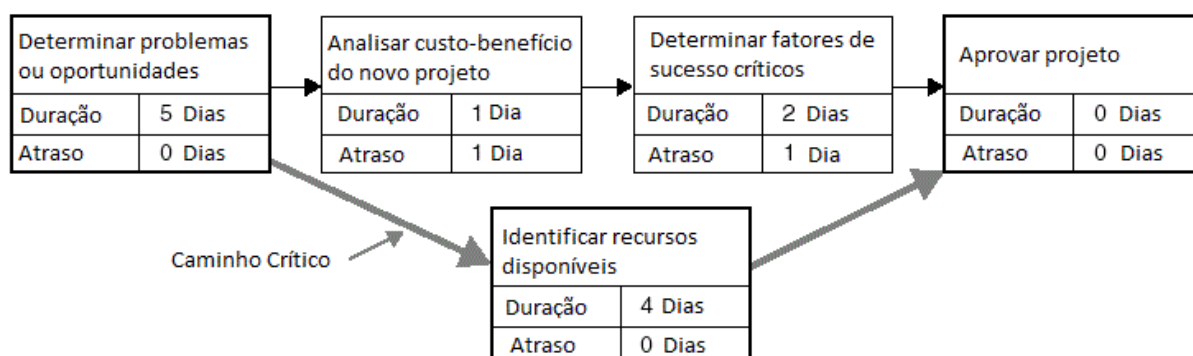
Dependências de recursos

A falta de pessoal pode reverter rapidamente os benefícios de ter poucas dependências de tarefas. Por exemplo, tarefas que poderiam ter sido executadas em paralelo, mas não podem ser atribuídas a vários membros da equipe devido a uma falta de pessoal, devem ser executadas em sequência.

Método do Caminho Crítico

Depois de identificar as dependências da tarefa e nivelar os recursos (ou seja, atribuir as tarefas e ajustar as dependências para os recursos disponíveis), use o método do caminho crítico (CPM) para descrever a duração da tarefa, indicando qualquer tempo de atraso para as tarefas que não estão no caminho crítico, conforme a figura abaixo. Isso fornece a visibilidade necessária para reatribuir recursos ou renegociar as restrições do projeto.

Figura 10 – Método do Caminho Crítico.



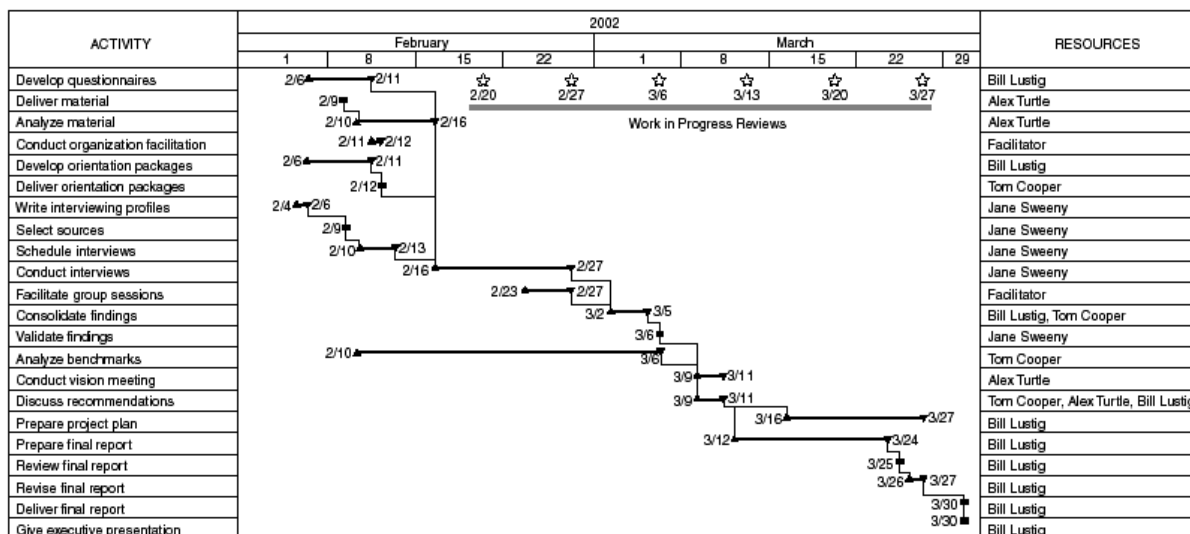
Fonte: Adaptado de Business Intelligence Roadmap.

No exemplo acima, a tarefa "Identificar recursos disponíveis" pode ser realizada em paralelo com as tarefas "Analisar custo-benefício do novo projeto" e "Determinar fatores de sucesso críticos". Uma vez que duração da tarefa "Identificar recursos disponíveis" é estimada em 4 dias, enquanto as das outras duas tarefas combinadas são estimadas em apenas 3 dias, a tarefa "Identificar recursos disponíveis" está no caminho crítico. Se esta tarefa demorasse 5 dias para se completar, ao invés de 4, atrasaria o marco "Aprovação do Projeto" por um dia. No entanto, se qualquer uma das outras duas tarefas fosse adiada por um dia, isso não afetaria o marco "Aprovação do projeto".

Horários do projeto

Depois de determinar todas as tarefas, recursos, dependências e estimativas, você pode agendar o projeto no calendário. A representação mais comum e mais familiar de um cronograma do projeto é um gráfico de Gantt. A figura abaixo mostra um exemplo.

Figura 11 – Gráfico de Gantt.



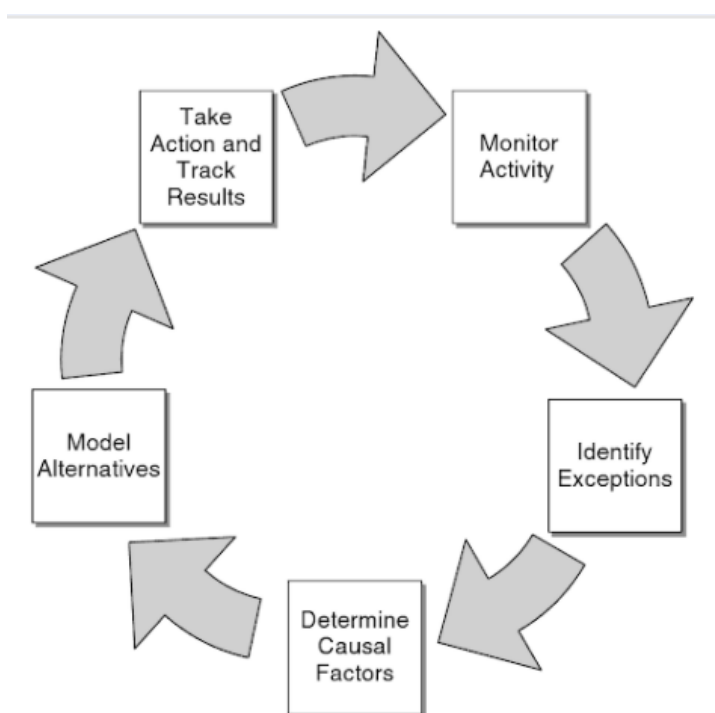
Fonte: Business Intelligence Roadmap.

Depois de inserir na ferramenta todos os componentes de planejamento (por exemplo, tarefas, estimativas, recursos, dependências), todos os ajustes que você fizer posteriormente automaticamente caem em cascata em todo o plano do projeto, atualizando todos os gráficos e relatórios. Embora os resultados ainda devam ser revisados e validados, um gerente de projeto experiente não precisa se tornar um escravo da ferramenta ou das atividades de planejamento do projeto.

Capítulo 5. Ciclo Analítico da Inteligência de Negócios

A análise de negócio segue um processo iterativo, que vai do monitoramento da atividade à identificação do problema ou da oportunidade, à determinação das ações a serem tomadas e, por fim, volta ao monitoramento dos resultados dessa ação.

Figura 12 – Ciclo Analítico.



Fonte: The Data Warehouse Lifecycle Toolkit.

O ciclo analítico de BI ajuda a entender como os usuários utilizarão os sistemas e quais as ferramentas a serem oferecidas para extraírem o máximo dos sistemas.

Esse ciclo deve ser explicitamente aplicado a cada aplicativo construído, sob pena de não se ter os resultados esperados caso isso não seja realizado.

Monitoramento de atividades

Consiste na análise das informações disponibilizadas pelos sistemas de inteligência de negócios. Através desses sistemas, é possível analisar o resultado instantâneo, bem como as condições do ecossistema no qual a empresa está inserida e os resultados históricos. É possível, assim, prever os impactos de uma determinada decisão.

O foco dessa etapa é totalmente orientado à camada de apresentação dos sistemas de BI, provendo as informações por meio de dashboards, portais de BI, balanced scorecards, entre outras ferramentas de visualização.

Um erro muito comum é acreditar que apenas essa fase é suficiente para uma solução de BI. Muitas vezes as implementações param nesse estágio e se julgam bem-sucedidas.

Identificação das exceções

Nessa etapa, identificam-se quais são e onde estão os problemas e as oportunidades a serem exploradas. Exceções ao comportamento normal podem tanto ser problemas como situações extraordinariamente positivas.

É importante dar ênfase aos pontos mais relevantes das informações. Mais importante do que mostrar todos os dados, nesse caso, é dar relevância aos que se comportam de maneira diferente.

As exceções podem gerar alarmes nos sistemas de BI, alertando seus usuários da condição identificada. Sistemas integrados a avisos por e-mail, SMS e *Push Notification* são bastante comuns nesses casos.

Determinação dos fatores causais

A determinação dos fatores causais é a busca pelos relacionamentos confiáveis das interações entre as variáveis determinantes no desempenho do

negócio. Tais relacionamentos são capazes de dizer aos analistas o “por que”, ou seja, a raiz causadora das exceções identificadas.

É importante integrar ao sistema de inteligência de negócios ferramentas estatísticas e de mineração de dados (especialmente regras de associação) para que seja possível determinar esse relacionamento de causa e efeito.

A integração dos dados dos sistemas internos da empresa à fontes externas que contenham indicadores que possam ter impacto no negócio, e o conhecimento do meio ambiente no qual se está inserido, é uma forma importante de não se equivocar na determinação das reais causas da exceção.

Modelagem de alternativas

A partir dos fatores causais identificados e dos dados históricos registrados no DW, desenvolvem-se alternativas de decisão. O objetivo final é poder realizar simulações dos efeitos oriundos de uma decisão em um cenário semelhante em ocasiões anteriores.

As ferramentas mais utilizadas nesse caso são a análise sensitiva, simulações de Monte Carlo e otimizações de busca de metas.

Atuação e acompanhamento dos resultados

Os resultados devem ser capturados e analisados para um refinamento contínuo dos processos, das regras de negócio e dos modelos analíticos.

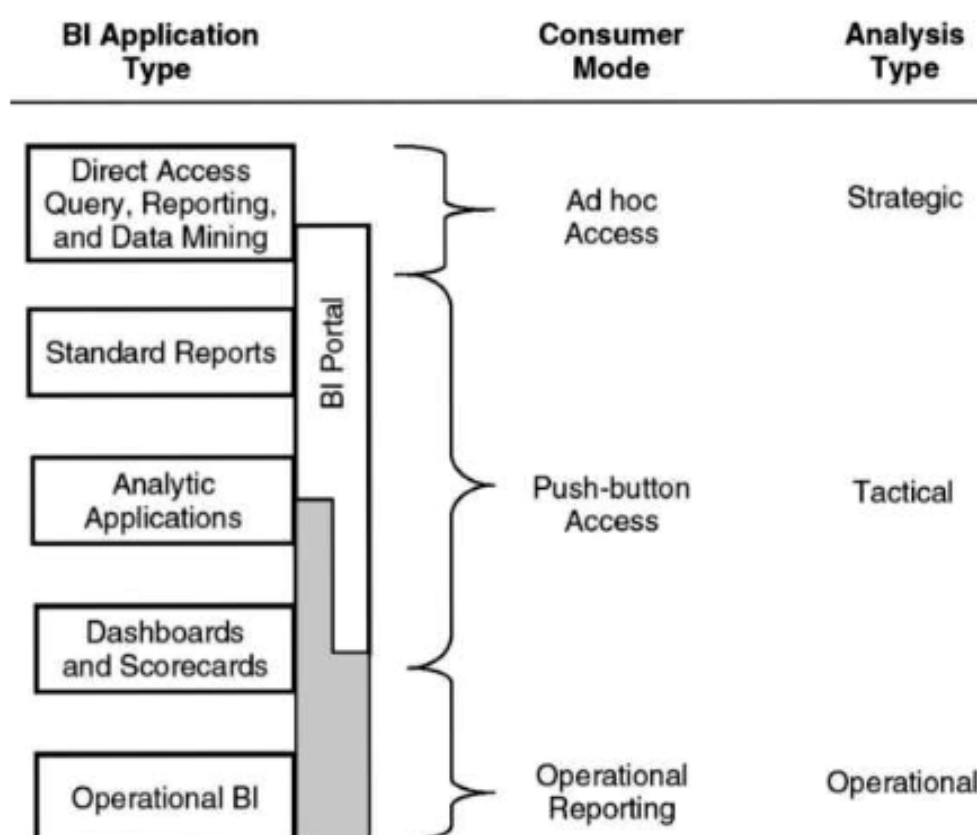
Para que isso seja possível, pode ser necessário que o DW seja capaz de acessar a linha operacional, de alterar os modelos dimensionais ou de criar bancos de acompanhamento de desempenho para que sejam guardadas as decisões de negócio tomadas e as métricas impactadas por elas para, então, serem definidas aquelas que tiveram um resultado satisfatório ou aquelas que não atingiram seus objetivos.

Capítulo 6. Tipos de soluções de BI

As aplicações de BI podem ser divididas em grupos que correspondem à forma de acesso e ao tipo de informação fornecida. O tipo de relatório está relacionado ao usuário, à frequência e ao valor da decisão a ser tomada.

Na figura abaixo, temos as formas mais comuns de apresentação das informações, divididas em forma de acesso e tipo de análise.

Figura 13 – Formas de acesso às informações em um projeto de BI.



Fonte: Adaptado de The Data Warehouse Lifecycle Toolkit.

Existem três tipos de análise com diferentes propósitos, cada um deles com sua devida importância.

O sistema de informação operacional é formado por operações rotineiras, e normalmente trabalha com um grande volume de operações de entrada e saída. A

maioria dos sistemas de informação estão neste nível e são caracterizados pela existência de muitos formulários de cadastros, relatórios e outras operações rotineiras. Geralmente são provenientes dos sistemas transacionais das empresas com relatórios estáticos e apoiam em tomadas de decisões de pequeno impacto, mas de grande frequência. Exemplos: formulários de cadastros, relatórios de conferência de dados, listagens, consultas e modificações de dados.

O sistema de informação tático é formado por operações de apoio na tomada de decisões. Tem função gerencial e trabalha com informações agrupadas. Este tipo de sistema usa as informações operacionais para criar mecanismos de gerenciamento das organizações. São caracterizados por relatórios onde o usuário pode selecionar alguns parâmetros e filtros, e apoiam decisões de médio impacto e média frequência. Exemplos: Projeção de vendas para o mês seguinte; Análise de clientes, produtos ou mercado; Lucro efetivo por produto; Comparativo de desempenho da empresa, segmentos ou produtos. Relatórios analíticos e sintéticos.

O sistema de informação estratégico é formado por operações estratégicas; apoia a alta diretoria filtrando as informações fundamentais e altamente estratégicas. São caracterizados por relatórios dinâmicos, onde o próprio usuário “monta” a informação a ser exibida através da junção de relatórios e do uso de cálculos e análises mais elaboradas. As decisões provenientes deste tipo de informação são de alto impacto e podem levar um longo tempo para ser tomada. Exemplos: Sistemas de controle de tráfego aéreo, sistemas de UTI e neonatal, controle de trânsito, sistemas que apoiam a alta direção das organizações e governos, entre outros.

Detalharemos agora alguns dos tipos de apresentação que constam na figura apresentada.

Consultas de Acesso Direto

Praticamente todos os sistemas de BI oferecem ferramentas que permitem aos usuários acessar diretamente o modelo dimensional através da camada de

metadados. Através desse tipo de acesso, pode-se ter desde resultados tabulares até mesmo complexos relatórios.

Essas ferramentas, em geral, exigem o conhecimento das linguagens de consulta disponíveis no sistema, bem como de uma visão geral do modelo lógico que estrutura os dados. Por isso, normalmente são utilizadas por usuários avançados, com um bom conhecimento das regras de negócio.

Algumas ferramentas permitem uma abordagem visual e interativa, na qual o usuário é capaz de arrastar e soltar colunas na área de relatório, definir restrições, transformações, cálculos e formatações a serem aplicadas aos dados ou mesmo exportar o resultado da consulta para uma ferramenta externa, como o Excel, para ter ainda mais flexibilidade na manipulação dos dados retornados.

A apresentação de resultados deve permitir que o usuário realize cálculos básicos, utilizando funções matemáticas, estatísticas, manipulação de *strings*, condicionais e relatórios de processamento sequencial. Os resultados devem ser pivotáveis, para que possam ser sumarizados, e deve ser possível acrescentar dados adicionais que aumentem o detalhamento do relatório e não existam no repositório de metadados. Deve ser possível, ainda, realizar formatações complexas nos dados exibidos e filtrá-los por parâmetros configuráveis pelo usuário, bem como gerar gráficos a partir deles.

Relatórios padrão

Entrega básica de qualquer sistema de BI, oferecem informações relevantes a respeito de uma área monitorada no negócio. A facilidade de acesso às informações faz deles o um produto bastante utilizado por usuários não técnicos.

Apesar de serem padronizados, os relatórios, em geral, oferecem suporte à entrada de parâmetros para seleção e restrição dos dados que serão apresentados no relatório.

BI Operacional

Enquanto o BI para o nível estratégico se preocupa em analisar tendências com o acúmulo de dados em meses e/ou anos para auxílio de decisões estratégicas, o BI de nível operacional está concentrado no gerenciamento e na otimização diária das operações do negócio, entregando a informação certa, no tempo certo, para usuários de negócio corretos.

Como o próprio nome diz, atua na parte operacional do processo e se caracteriza por relatórios estáticos, muitas vezes advindos dos sistemas transacionais das empresas, como os ERPs. As decisões tomadas a partir destes relatórios são rápidas e de baixa complexidade.

Aplicativos analíticos

Desenvolvidos para resolver um problema específico, costumam ser concebidos tendo em vista o responsável pela solução do problema.

Seu uso costuma ser mais complexo que os relatórios padrão. Aplicativos analíticos podem incluir algoritmos ou modelos de mineração de dados para encontrarem informações ocultas nos dados.

Gestão do conhecimento

Um bom projeto de BI deve aplicar processos de gestão do conhecimento para obter apoio para análises e tomadas de decisão. O termo “Gestão do Conhecimento” provém do inglês “Knowledge Management”, e trata de uma área de atuação transversal entre as diversas disciplinas relacionadas, sobretudo, à gestão estratégica, teoria das organizações, a sistema de informação, gestão da tecnologia, e à áreas mais tradicionais (como a economia, sociologia, psicologia, marketing, entre outras.).

Para serem mais competitivas, as empresas precisam conhecer melhor seus processos. Conhecer é conquistar, elaborar e praticar melhor; porém, saber muito, por si só, não significa melhor nível de competitividade.

O conhecimento, necessita de gestão, de suporte, de um processo de armazenagem, de zelo na guarda de suas informações de canais para a sua disseminação. O conhecimento abrange o capital intelectual, o capital humano, a capacidade de pesquisar e inovar e a inteligência empresarial.

A gestão de conhecimento é necessária pois muito dos processos das empresas estão apenas nas cabeças das pessoas, e essa informação precisa ser estruturada e gerenciada. A organização pode coletar as informações por meio de seu próprio conhecimento adquirido e desenvolvimento e a partir do conhecimento colhido no ambiente externo (experiência de concorrentes, influências culturais, inovações tecnológicas, etc). Essa gestão se preocupa com as condições organizacionais e com a localização, geração e partilha do conhecimento. Se preocupa também com as ferramentas a serem utilizadas na comunicação e organização de determinado conteúdo.

A gestão de conhecimento amplia a vantagem competitiva e concorrencial da empresa. Reduz custos com P&D (Planejamento e Desenvolvimento) e possibilita geração de novos modelos de negócio, melhor aproveitamento e desenvolvimento do capital intelectual da empresa, suporte às tomadas de decisão e melhorias na produção e na prestação de serviços.

Em suma, é uma modalidade de gestão que facilita o controle e o acesso às informações relevantes num processo de trabalho e a administração de seus meios. O conhecimento parte de uma informação, pesquisa ou experiência e produz impactos positivos ou negativos na sociedade e em determinada organização, dependendo de como esse conhecimento é filtrado, analisado e gerido.

Inteligência competitiva

A inteligência competitiva é saber se antecipar às tendências do mercado utilizando informações privilegiadas, sejam essas coletadas dentro da empresa, obtidas através de fontes externas ou analisando a concorrência.

Uma boa abordagem da inteligência de negócios engloba processos como obtenção do mercado em constante atualização, análise dos concorrentes para saber sobre novos produtos e poder se posicionar para riscos e oportunidades, planejamentos do negócio utilizando aprendizados obtidos através de erros e acertos do passado e bastante coleta de dados.

O processo de inteligência de negócios ajuda a empresa a se capitalizar e ser mais competitiva pois utiliza informações para tomar decisões e ser mais competitiva. A inteligência de negócios pode se utilizar de informações externas e também de informações obtidas através de dados do processo da gestão do conhecimento.

Data Mining

O Data mining, ou mineração de dados, é uma classe de análise de informações baseada em bancos de dados, que procura padrões ocultos em uma coleção de dados que podem ser usados para prever comportamentos futuros. Entretanto, o termo é comumente usado de forma equivocada para descrever software que apresentam dados de novas maneiras. O verdadeiro software de data mining não muda apenas a apresentação: ele de fato descobre relações antes desconhecidas entre os dados, e este conhecimento é aplicado para se alcançar metas de negócios específicas.

As ferramentas de data mining são usadas para substituir ou aprimorar a inteligência humana devido à sua capacidade de verificar enormes armazéns de dados. Desta forma, elas descobrem novas e significativas correlações, padrões e tendências através de tecnologias de reconhecimento de padrões e métodos estatísticos avançados.

As principais categorias das tarefas de mineração de dados são:

- Modelagem preditiva: criação de modelos para variável alvo como função das variáveis independentes. Podemos prever, por exemplo, que o consumidor X irá gastar 8000 este ano.
- Análise associativa: descoberta de padrões descritores de características fortemente associadas nos dados analisados. Por exemplo, o consumidor que comprou o produto X, também irá gostar do produto Y.
- Análise de clusters: identificação de grupos de registros que apresentem uma similaridade significativa entre si. Por exemplo, consumidor A é do tipo X, e consumidor B é do tipo Y.
- Detecção de anomalias: identificação de registros que apresentem uma diferença significativa do restante dos dados. Por exemplo, uma fraude.

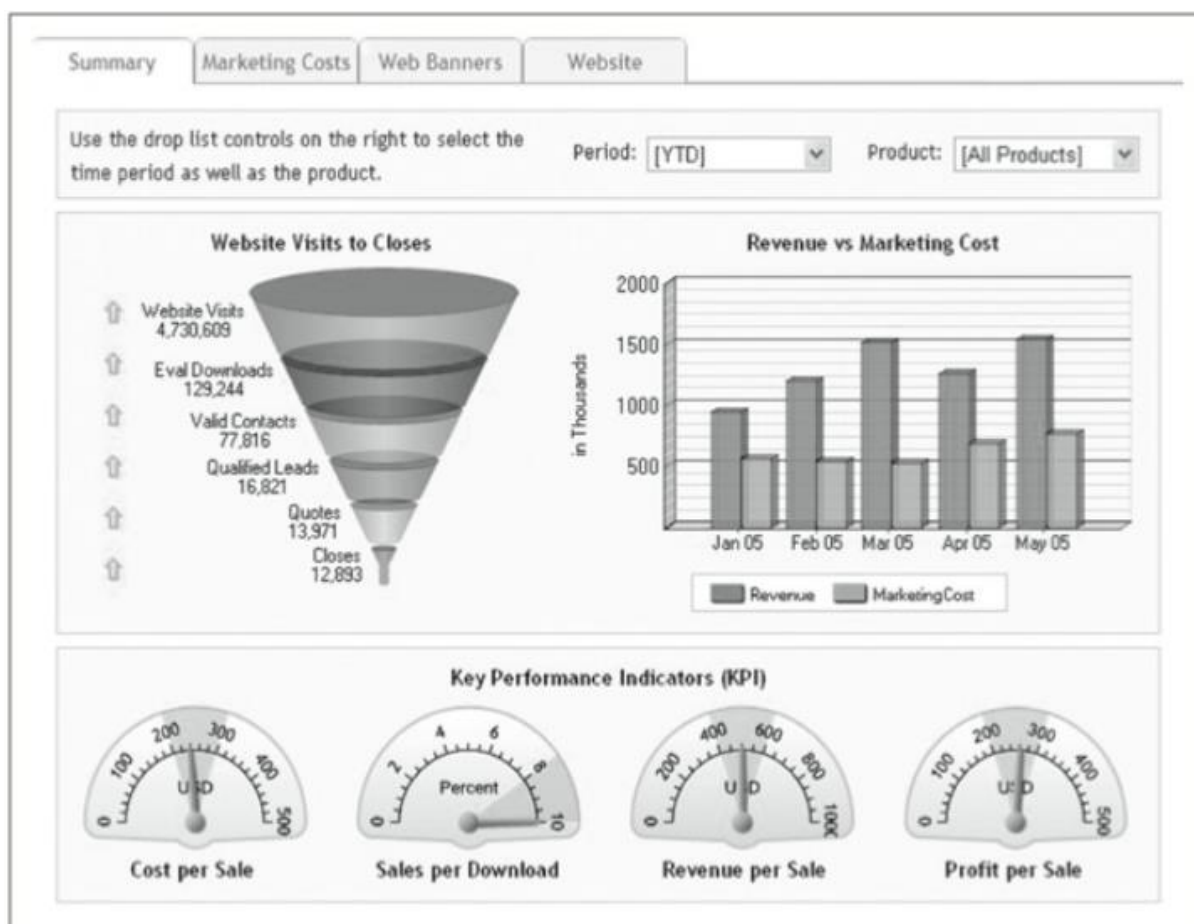
Dashboards

Dashboards são painéis que mostram métricas e indicadores importantes para alcançar objetivos e metas traçadas de forma visual, facilitando a compreensão das informações geradas. Existem diversos tipos de geração de informação, mas o formato visual ganha muitos pontos quando se trata de entender como estão indo resultados, possibilitando que qualquer pessoa consiga consumir a informação de forma fácil e rápida.

O objetivo dos dashboards é possibilitar o monitoramento dos resultados de uma empresa distribuídos em diversos indicadores. É preciso saber exatamente o que se quer apresentar e a finalidade dos painéis, do contrário eles serão apenas gráficos bonitos. Para isso, é necessário responder perguntas básicas para definir se um determinado dashboard deve ser construído e apresentado. Também é necessário compreender as necessidades da empresa e definir as métricas de KPIs.

De uma maneira geral, as perguntas a serem respondidas são: Qual informação quero evidenciar? Qual a melhor forma para receber a informação? Quanto tempo eu demoro para explicar a informação? Que decisão eu posso tomar com esta informação?

Figura 14 – Dashboards.



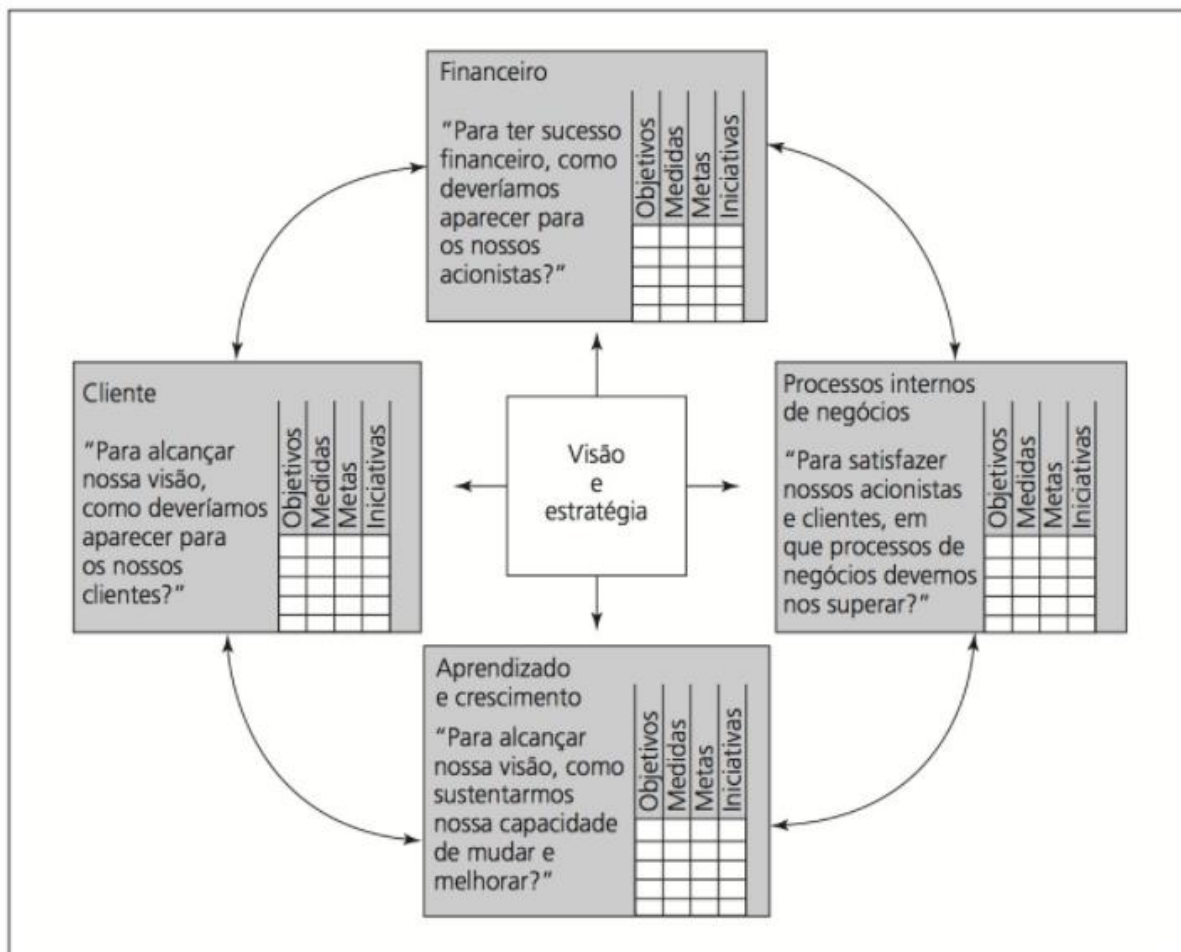
Fonte: Business Intelligence: Um enfoque gerencial.

Balanced Scorecards

O balanced scorecard é relativamente novo, surgiu em 1992, mas talvez seja o sistema de gerenciamento de desempenho mais bem conhecido e amplamente utilizado. Ele consiste em analisar os objetivos e metas da empresa através de quatro perspectivas: financeira, dos clientes, dos processos internos e de aprendizado e

crescimento. Portanto, ao planejar alguma ação ou estratégia para a empresa, elas devem perpassar por estes quatro níveis, e aí que está o balanceamento do processo.

Figura 15 – Perspectivas do balanced scorecard.



Fonte: Business Intelligence: Um enfoque gerencial.

Como podemos ver na figura anterior, para cada perspectiva a ser analisada é preciso definir os seguintes pontos:

- **Objetivos:** Definir o que a empresa deseja alcançar em cada perspectiva estratégica.
- **Indicadores:** Indicam o desempenho da empresa referente a cada objetivo definido.

- **Metas:** Em função dos indicadores, qual o nível de performance esperado que se deve atingir.
- **Projetos estratégicos:** As ações, iniciativas e intervenções que devem ser tomadas para que as metas de desempenho determinadas sejam atingidas.

Dessa forma, podemos criar um mapa estratégico para cada necessidade da empresa, abordando as quatro perspectivas e seus pontos a serem analisados, conforme exemplo abaixo de uma estratégia para uma joalheria:

- **Perspectiva Financeira** - Objetivos: Aumentar a lucratividade. Metas: Aumentar em 15% a lucratividade da empresa. Indicadores: Demonstrativos financeiros. Iniciativas: Negociar parcerias de parcelamento com empresas de cartões de crédito.
- **Perspectiva do Cliente** - Objetivos: Ter uma loja mais atraente para os clientes. Metas: Aumentar em 20% a média de visitas diárias à loja. Indicadores: Contagem de clientes. Iniciativas: Melhorar a exposição das joias nas vitrines e investir em mídias sociais.
- **Perspectiva de processos internos** - Objetivos: Ser referência em atendimento. Metas: Aumentar o número de elogios em 15% e diminuir as reclamações no SAC em 80% . Indicadores: Análise estatística dos relatórios do SAC. Iniciativas: Redesenhar processo de atendimento.
- **Perspectiva de aprendizado e crescimento** - Objetivos: Ter uma força de vendas com profissionais experientes. Metas: Substituir 30% dos vendedores. Indicadores: Número de contratações x demissões. Iniciativas: Iniciar processo de recrutamento e seleção.

Capítulo 7. Modelagem Dimensional

Durante muitos anos a TI concentrou-se em construir sistemas essenciais que permitissem principalmente o processamento de transações corporativas. Tais sistemas deveriam ser praticamente tolerantes a falhas e oferecer execução eficiente e resposta rápida. O processamento de transações online (OLTP) ofereceu uma solução eficaz, que gira entorno de atividades repetitivas e de rotina, usando um ambiente de banco de dados relacional. Os últimos desdobramentos nesta área são o uso de softwares de ERP e de gestão da cadeia de fornecimento (SCM) para tarefas de processamento de transações, aplicações de gerenciamento de relacionamento com o cliente (CRM) e integração com tecnologias baseadas na web e intranets.

Tanto as aplicações de OLTP quanto de sistemas de suporte a decisões, constantemente necessitam de acesso aos dados. Infelizmente, tentar servir os dois tipos de solicitação pode ser uma tarefa problemática. Assim, o ideal é separar os sistemas de informações em tipos OLTP e tipos OLAP.

O OLTP é voltado para o processamento de transações repetitivas em grandes quantidades e manipulações simples. O OLAP envolve o exame de muitos itens de dados (constantemente milhares ou até milhões) em relacionamentos complexos. Além de responder às consultas dos usuários, o OLAP consegue analisar esses relacionamentos e buscar padrões, tendências e exceções. Em outras palavras, o OLAP é um método direto de suporte à decisão.

Uma consulta de OLAP pode acessar um banco de dados com gigabytes ou terabytes de informações sobre vários anos de vendas, a fim de encontrar todas as vendas de produtos em cada região para cada tipo de produto. Com essas informações, os usuários podem filtrar dados, fazer comparações e análises. A forma mais usual de organizar os dados em um modelo OLAP é através de tabelas fato e dimensão.

Tabelas Fato e Dimensão

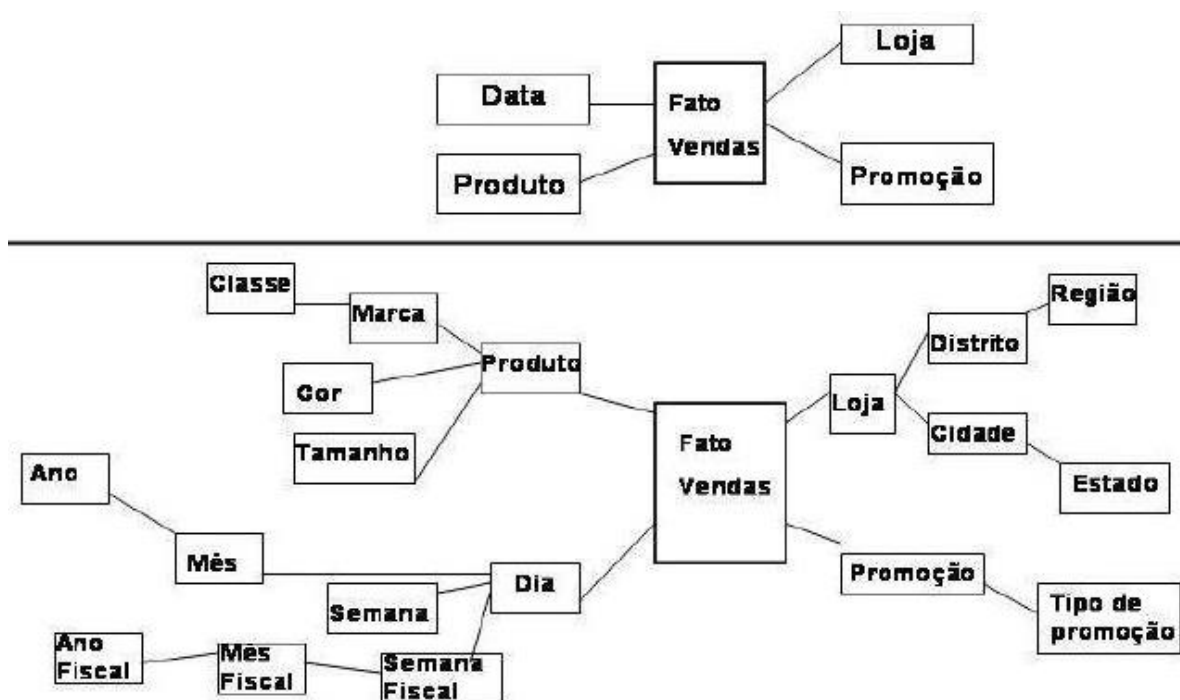
O modelo dimensional possui características diferente do conhecido modelo relacional. No modelo dimensional, utilizado pelas ferramentas OLAP, os dados são estruturados de forma que facilite o entendimento e visualização de problemas típicos de tomada de decisão, deixando o processamento mais intuitivo. Os dados são modelados através de tabelas fato e dimensão.

As tabelas fato são as tabelas centrais e possuem dados sobre medidas numéricas que mantêm informações sobre o desempenho do negócio. Cada linha desta tabela corresponde ao valor de uma medida dentro de algumas dimensões. As tabelas fatos possuem chaves estrangeiras que referencia informações nas tabelas dimensão.

As tabelas dimensão são tabelas periféricas que armazenam informações descritivas acerca do negócio. São utilizadas para filtros, agrupamento e rótulos. Estas tabelas normalmente são desnormalizadas, não preocupando com redundância de informações e sim com a performance das consultas.

As tabelas dimensão podem ser modeladas de duas formas: modelo estrela e modelo floco de neves. O modelo estrela é o mais utilizado e sua principal característica é desnormalização dos dados, onde toda informação de um mesmo contexto fica armazenada na mesma tabela, enquanto no modelo floco de neve os dados estão mais normalizados, muitas vezes aproveitando as junções existentes nos bancos dos sistemas transacionais.

Figura 16 – Modelo estrela (acima) e modelo floco de neve (abaixo).



Fonte: Cetax consultoria.

Hierarquia das dimensões

As dimensões podem e devem possuir dados hierárquicos, pois essas informações permitem que, ao criar cubos, esses dados auxiliem em visualizações em diferentes níveis e em filtros. Podemos destacar informações geográficas como país, estado, cidade, bairros etc. Outro tipo de informação que permite a divisão hierárquica são os campos de datas, esses são de extrema importância em aplicativos de BI e devem ser estruturados hierarquicamente, dividindo a informação em dia, mês, ano, dia da semana, feriado, períodos fiscais etc.

É importante que os dados hierárquicos sejam armazenados na mesma tabela para facilitar e agilizar a visualização dos dados.

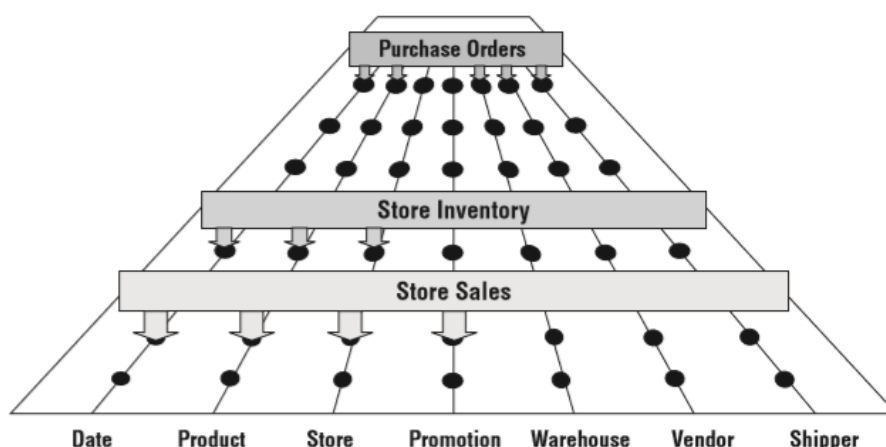
Enterprise Datawarehouse Bus Matrix

Os processos de negócios da empresa possuem suas tabelas fatos e dimensões específicas, porém, elas podem possuir tabelas dimensões em comum que podem ser compartilhadas entre si. Para organizar este compartilhamento de tabelas entre os processos, é utilizado o Data Warehouse Bus Matrix (DWEBM). O processo é incremental, possibilita que várias equipes de desenvolvimento trabalhem de maneira independente e assíncrona, além de padronizar tabelas dimensão para mais processos de negócio.

No diagrama bus matrix, as dimensões conformadas são o barramento e são determinadas pelo seguinte processo de design dimensional:

- Escolha do processo de negócio.
- Definição da granularidade.
- Identificação das dimensões já contempladas no Enterprise Data Warehouse Bus.
- Inclusão das dimensões não contempladas no barramento.
- Identificação dos fatos.

Figura 17 – Data Warehouse Bus Matrix.



Fonte: The Data Warehouse Lifecycle Toolkit.

Outra forma de apresentar as tabelas no bus matrix é através de uma matrix onde as linhas são fatos e as colunas são dimensões. Este é um dos processos mais importantes na construção de um DW, pois através deste artefato podemos determinar as dimensões que merecem atenção especial por estarem presentes em mais tabelas fatos.

Figura 18 – Data Warehouse Bus Matrix – Apresentação alternativa.

BUSINESS PROCESSES	COMMON DIMENSIONS						
	Date	Product	Warehouse	Store	Promotion	Customer	Employee
Issue Purchase Orders	X	X	X				
Receive Warehouse Deliveries	X	X	X				X
Warehouse Inventory	X	X	X				
Receive Store Deliveries	X	X	X	X			X
Store Inventory	X	X		X			
Retail Sales	X	X		X	X	X	X
Retail Sales Forecast	X	X		X			
Retail Promotion Tracking	X	X		X	X		
Customer Returns	X	X		X	X	X	X
Returns to Vendor	X	X		X			X
Frequent Shopper Sign-Ups	X			X		X	X

Fonte: The Data Warehouse Lifecycle Toolkit.

Operadores

O modelo dimensional, como o próprio nome diz, organiza os dados em múltiplas dimensões, que muitas das vezes seguem uma determinada hierarquia, como visto anteriormente. Este formato permite a exibição dos dados em diferentes perspectivas. Uma forma muito usual é a exibição através de cubos. Os cubos são

representações multidimensionais, onde os valores contidos nas tabelas fato são apresentados de acordo com as informações selecionadas das tabelas dimensão.

Existem diversas operações que usualmente são utilizadas na maior parte das ferramentas de BI, são elas:

- **Drill down:** ocorre quando o usuário aumenta o nível de detalhe da informação, diminuindo o nível de granularidade.
- **Roll up:** ocorre quando o usuário aumenta o nível de granularidade, diminuindo o nível de detalhe da informação.
- **Drill Across:** ocorre quando o usuário pula um nível intermediário dentro de uma mesma dimensão.
- **Drill through:** ocorre quando um usuário passa de uma informação contida em uma dimensão para uma outra.
- **Slice:** é a operação que corta o cubo, mas mantém a mesma perspectiva de visualização dos dados.
- **Dice:** é o slice porém em mais dimensões.
- **Pivot:** é o ângulo pelo qual os dados são vistos. Na prática, corresponde à modificação das dimensões em um gráfico ou troca de linhas por colunas em uma tabela.

Outras ações também são possíveis mas menos usuais, nas ferramentas de exibição de cubos:

- **Ranking:** classifica determinada informação baseada nos melhores indicadores.
- **Last-week:** mostra os valores relacionados à semana anterior tendo base a semana atual.
- **Prior-week:** mostra os valores relacionados aos últimos sete dias.

- **Year-to-date:** com os dados do ano corrente até a data atual.
- **Nest-unnest:** redução das dimensões.

Dados Operacionais e Informações

Os dados podem ser estruturados de duas maneiras, de acordo com os propósitos a serem utilizados. Operacionais (OLTP), quando destinados a sistemas transacionais, normalizados, com dados voláteis (inserções, deleções, alterações) e utilizados em sistemas ERP, CRM e outros. Informacionais (OLAP), quando destinados a sistemas analíticos, não normalizados, que facilitam a consulta e muito utilizado para sistemas de apoio à tomada de decisão. Abaixo temos um comparativo com características de cada natureza de dados.

Figura 19 – Dados Operacionais X Dados Informacionais.

CARACTERÍSTICAS	DADOS OPERACIONAIS	DADOS INFORMACIONAIS
1. Conteúdo	Valores correntes	Valores sumarizados, calculados, integrados de várias fontes
2. Organização dos dados	Por aplicação/sistema de informação	Por assuntos/negócios
3. Natureza dos dados	Dinâmica	Estática até o <i>refreshment</i> dos dados, de tempos em tempos
4. Formato das estruturas	Relacional, próprio para computação transacional	Dimensional, simplificado, próprio para atividades analíticas
5. Atualização dos dados	Atualização campo a campo	Acesso granular ou agregado, normalmente sem <i>update</i> direto
6. Uso	Altamente estruturado em tabelas, processamento repetitivo	Estruturado em fatos e dimensões, com processamento analítico/preditivo
7. Tempo de resposta	Otimizado para faixas abaixo de 1 seg	Análises mais complexas, com tempos de respostas maiores

Fonte: BI2: Business Intelligence.

Data Warehouse e Data Marts

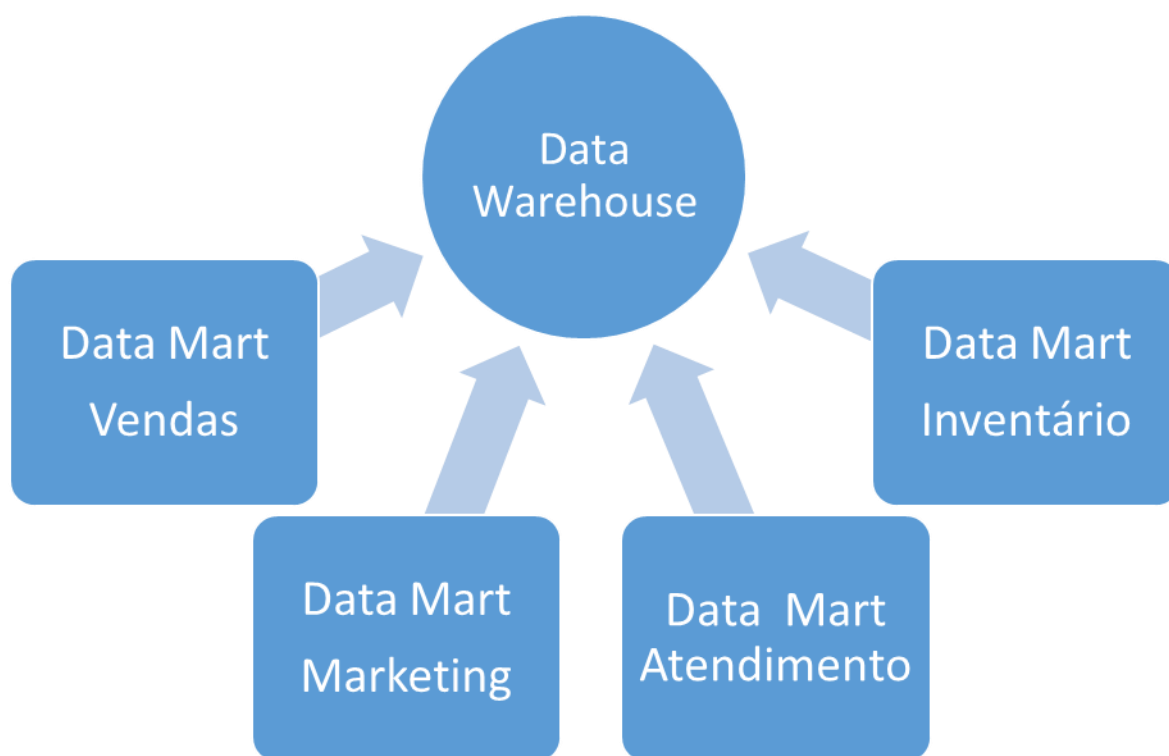
Um Data Warehouse (DW), ou armazém de dados, é um banco de dados com dados temporais usados para análise e decisões das mais diversas perguntas realizadas por executivos. Os dados contidos nos DW são sumarizados, periódicos e descritivos. Com a manipulação desses dados os executivos podem tomar decisões baseadas em fatos e não em intuições e especulações. O banco de dados de um Data Warehouse deve ser projetado para processamento analítico on-line (OLAP), onde caracteriza-se pela ênfase na performance da recuperação das informações. Dentre as vantagens de se utilizar um DW, podemos destacar:

- Melhora a qualidade dos dados, criando uma padronização de códigos e descrições e identificando e corrigindo dados ruins.
- Apresenta as informações da organização de forma consistente.
- Fornece um único modelo de dados para toda a organização, independentemente da fonte.
- Reestrutura os dados de modo a satisfazer as necessidades dos usuários do negócio.
- Reestrutura os dados para melhorar o desempenho de consulta, mesmo para consultas analíticas complexas, sem afetar os sistemas em operação.
- Agrega valor às aplicações de negócio operacional, principalmente a gestão de relacionamento com clientes (CRM).

Um Data Mart é um sub Data Warehouse, ou seja, um pequeno armazenamento de dados que fornece suporte à decisão de um pequeno grupo de pessoas. Os Data Marts atendem às necessidades de unidades específicas de negócios, ao invés da corporação como um todo. Eles podem ser apropriados e gerenciados por pessoal de fora do departamento de informática das corporações. A crescente popularidade dos Data Marts em cima da popularidade dos grandes sistemas de Data Warehouses corporativos é baseada em bons motivos:

- Os Data Marts têm diminuído drasticamente o custo de implementação e manutenção de sistemas de apoio à decisões, colocando-os ao alcance de um número muito maior de corporações.
- Os Data Marts têm o escopo mais limitado e são mais identificados com grupos de necessidades dos usuários, o que se traduz em esforço/equipe concentrados.
- Os departamentos autônomos e as pequenas unidades de negócios frequentemente preferem construir o seu próprio sistema de apoio à decisões via Data Marts.

Figura 20 – Data Warehouse e Data Marts.



A diferença entre um DW e um DM basicamente consiste no volume de dados, abrangência e foco. Enquanto o DW foca na organização como um todo, os DMs focam em um determinado departamento ou conjunto específico de usuário, por exemplo.

A construção do armazém pode acontecer de duas formas, Top Down, onde primeiro se constrói o DW e posteriormente se especializa construindo os DM, ou Bottom Up, onde se inicializa construindo cada DM e expandindo para um DW. Cada abordagem tem seus prós e contras, as circunstâncias e particularidades de cada projeto é que determinarão qual utilizar.

Capítulo 8. Construção do Modelo Dimensional

Uma organização tem à sua escolha um grande conjunto de ferramentas de design, armazenamento e manutenção, comerciais ou de software livre, para implementar o DW. Nem todas as ferramentas são compatíveis entre si e nem todas são apropriadas para a metodologia de desenvolvimento escolhida. Apesar do grande universo de escolha nas ferramentas, a modelagem das mesmas está geralmente baseada numa de duas metodologias: Inmon ou Kimball, que abordam os modelos Top Down e Bottom Up, respectivamente.

Abordagem Bill Inmon

Em 1990, Bill Inmon ganhou o apelido “Pai do Data Warehouse”, apresentando o termo Data Warehouse na publicação Building the Data Warehouse. O modelo proposto por Bill Inmon é a Top Down, onde primeiro cria-se o DW para depois especializá-lo em DMs. Sua abordagem foi amplamente utilizada e evoluiu ao longo do tempo. Suas principais características são:

- Arquitetura lógica para extrair os dados de BDs operacionais dispersos.
- Dados são transformados e organizados temporalmente num único DW.
- Parte destes dados são então extraídos para BDs menores, criando BDs departamentais denominadas Data Mart (DM), onde os utilizadores finais exploram os dados e criam relatórios.

Vantagens dos modelo top-down:

- Melhor definição estratégica do projeto.
- DW Corporativo (DWC) modelizado segundo um modelo normalizado:
 - Simplificação nos procedimentos de ETL.
 - Menores taxas de crescimento do volume de dados.

- Proporciona um recenseamento integral dos sistemas fonte e conteúdos existentes na organização.
- Desenvolve uma abordagem sistematizada e completa sobre os processos de integração.

Desvantagens do modelo top-down:

- Abordagem top-down centrada nos dados, mais morosa e dispendiosa.
- Maiores custos iniciais em TI.
- Abordagem excessivamente centrada nos dados (todo o processo de desenvolvimento depende da prévia conclusão do modelo corporativo dos dados):
 - Inviabiliza o envolvimento dos utilizadores no projeto.
 - Prolonga o período de ausência de resultados.
 - Transfere para segundo plano a identificação das reais necessidades de informação dos utilizadores.
- Modelos normalizados, que resulta em pior desempenho analítico.

Abordagem Kimball

O modelo proposto por Kimball é o bottom up, onde são criados os DMs e expandindo para o DW. Este é o modelo mais indicado e amplamente utilizado atualmente.

Vantagens do modelo bottom up:

- Infraestrutura mais adequada às exigências de um sistema de apoio à decisão.
- DWC modelizados segundo modelo desnormalizado (esquemas estrela):

- Estrutura mais flexível, comportando mais facilmente as alterações nos sistemas fonte.
 - Desenvolvimento de modelos mais intuitivos e com melhor desempenho.
- Abordagem Iterativa centrada nas necessidades de informação.
- Permite antecipar a entrega de resultados.
- Garante o maior envolvimento dos utilizadores.
- Permite fasear os custos de investimento em infraestrutura.
- Proporciona um melhor time to market e retorno sobre o investimento.
- Abordagem de implementação totalmente integrada.

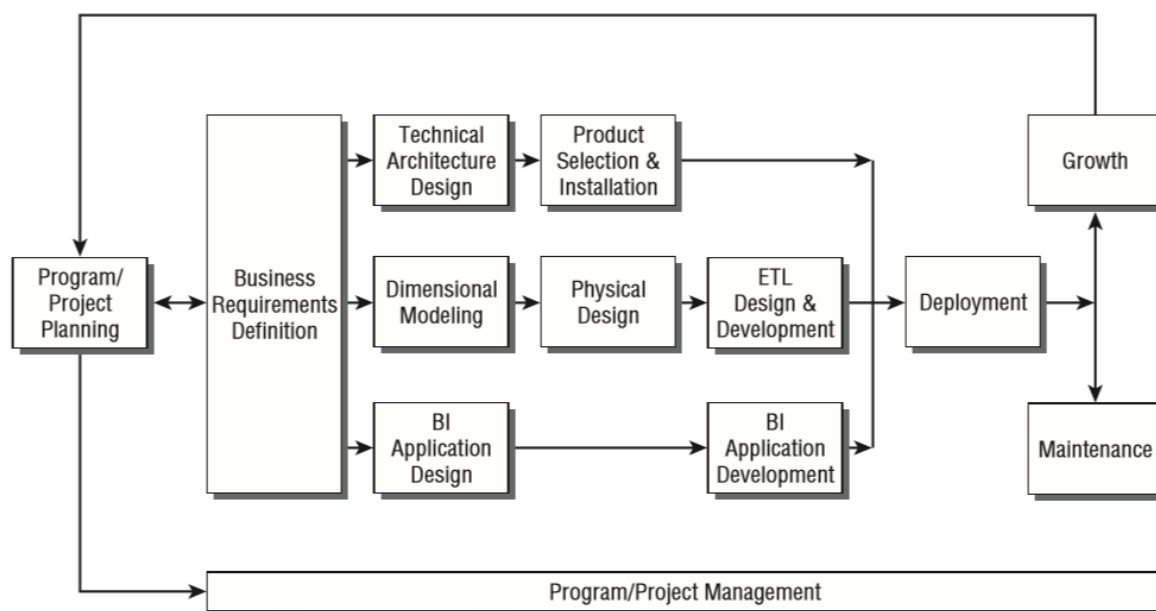
Desvantagens do modelo bottom up:

- Dificuldade em definir as dimensões e fatos.
- Esquemas estrela do DWC -> vertiginoso crescimento do volume de dados armazenado.
- Conduz à obtenção de procedimentos de ETL, mais complexos:
 - Modelos dimensionais requerem operações adicionais de transformação e agregação dos dados dos sistemas operacionais (usualmente representados em modelos normalizados).
 - Alterações ao nível dos sistemas operacionais implicam alterações em procedimentos dedicados a diferentes esquemas, em estrelas de diferentes granularidades.

O Ciclo Kimball e Projeto Físico

Kimball define uma série de passos e processos para a criação de um aplicativo de BI, conforme é apresentado na figura abaixo. Vários processos já foram vistos em capítulos anteriores como o planejamento, a definição da infraestrutura, os aplicativos utilizados na camada de aplicação, o acompanhamento e gerenciamento de um projeto, dentre outros. Neste capítulo iremos abordar os quatro passos para o processo de design do modelo dimensional. Os passos são selecionar o processo de negócio, definir o grão, escolher as dimensões e identificar as métricas das tabelas de fatos.

Figura 21 – Ciclo de vida de um aplicativo de BI.



Fonte: The Data Warehouse Toolkit.

Seleção do processo de negócio

Os processos de negócios são as atividades operacionais realizadas por sua organização, como fazer um pedido, processar uma reivindicação de seguro, registrar alunos para uma aula ou fazer snapshotting de cada conta a cada mês. Os eventos de processos de negócios geram ou capturam métricas de desempenho que se

traduzem em fatos, e esses fatos em uma tabela de fatos. A maioria das tabelas de fatos se concentram nos resultados de um único processo comercial. Escolher o processo é importante porque define um alvo de design específico e permite que os grãos, dimensões e fatos, sejam declarados. Cada processo comercial corresponde a uma linha na matriz do barramento do data warehouse da empresa.

Definição do grão

Definir o grão é um passo fundamental em um design dimensional. O grão estabelece exatamente o que representa uma única linha da tabela de fatos. A declaração de grãos torna-se um contrato vinculativo sobre o projeto. O grão deve ser declarado antes de escolher dimensões ou fatos, porque cada dimensão ou fato do candidato deve ser consistente com o grão. Esta consistência reforça uma uniformidade em todos os projetos dimensionais que são críticos para o desempenho da aplicação de BI e facilidade de uso. O grão atômico refere-se ao nível mais baixo, no qual os dados são capturados por um determinado processo comercial e essa deve ser a abordagem utilizada, ou seja, escolher um grão que possua um único propósito.

Cada grão de tabela de fatos proposto resulta em uma tabela física separada. Grãos diferentes não devem ser misturados na mesma tabela de fatos.

Escolha das dimensões

As dimensões fornecem a “quem, o que, onde, quando, por que e como” o contexto que envolve um evento de processos de negócios. As tabelas de dimensões contêm os atributos descritivos usados por aplicativos de BI para filtrar e agrupar os fatos. Com o grão de uma tabela de fatos em mente, todas as dimensões possíveis podem ser identificadas.

Sempre que possível, uma dimensão deve ser de valor único quando associada a uma determinada linha de fato. As tabelas de dimensão às vezes são

chamadas de “alma” do data warehouse, porque contêm pontos de entrada e rótulos descritivos que permitem que o sistema DW / BI seja alavancado para análise de negócios. Uma quantidade desproporcional de eficiência é colocada na governança de dados e no desenvolvimento de tabelas de dimensões, porque elas são as direções da experiência de BI do usuário.

Métricas das tabelas de fatos

Os fatos são as medidas resultantes de um evento de processos de negócios e quase sempre são numéricas. Uma única linha da tabela de fatos tem uma relação de um para um, para um evento de medição, conforme descrito pelo grão da tabela de fato. Assim, uma tabela de fatos corresponde a um evento observável fisicamente, e não às exigências de um relatório específico.

Dentro de uma tabela de fatos, apenas os fatos consistentes com o grão declarado são permitidos. Por exemplo, em uma transação de venda a quantidade de um produto vendido e seu preço unitário são bons fatos, enquanto que o salário do gerente da loja não é incluído.

Capítulo 9. Ferramentas de BI

Conforme discutido durante toda a apostila, o projeto de BI é um processo que envolve diversas ações e atividades, e não pode ser resumido em apenas uma ferramenta. Porém, atualmente existem diversas opções de ferramentas voltadas para processos do BI que auxiliam muito o trabalho.

Para exemplificar a utilização destas ferramentas, será apresentada uma situação hipotética na qual serão realizadas integrações de fontes de dados, tratamentos e exibições de dados de uma maneira bastante sucinta e simplificada, apenas para introduzir os alunos à essas ferramentas. Construções de data warehouses, definições de métricas, KPIs, tratamentos mais complexos e painéis são apresentados de forma mais aprofundada em outras disciplinas do IGTI.

Para este exemplo, serão criadas duas fontes de dados fictícias que representarão uma base de dados em Mysql, que apresenta a estrutura de dados de um sistema onde estão cadastrados alunos e cursos do IGTI e um arquivo CSV, que representa um documento que foi exportado de um sistema de eventos onde alguns destes alunos participaram de uma semana de palestras. Assim, será feita a integração e tratamento destes dados para posteriormente serem analisados.

Para essa simulação, serão utilizadas três ferramentas: Xampp, que irá armazenar a base de dados Mysql; PDI do Pentaho, que irá realizar o ETL; e o Power BI desktop da Microsoft, que fará a exibição dos dados. Tanto o Pentaho quanto o Power BI realizam ambos trabalhos, porém serão utilizadas duas ferramentas devido à popularidade e a facilidade destas ferramentas para cada tarefa e para que os alunos tenham contato com mais opções.

A instalação foi realizada no sistema operacional Windows 10. Foram realizados os downloads dos respectivos arquivos .exe e a instalação padrão dos aplicativos Xampp e Power BI. Já o Pentaho PDI, o download foi realizado da pasta compactada, que foi extraída na pasta [c:/Pentaho](#), e para executá-lo é necessário executar o arquivo spon.bat.

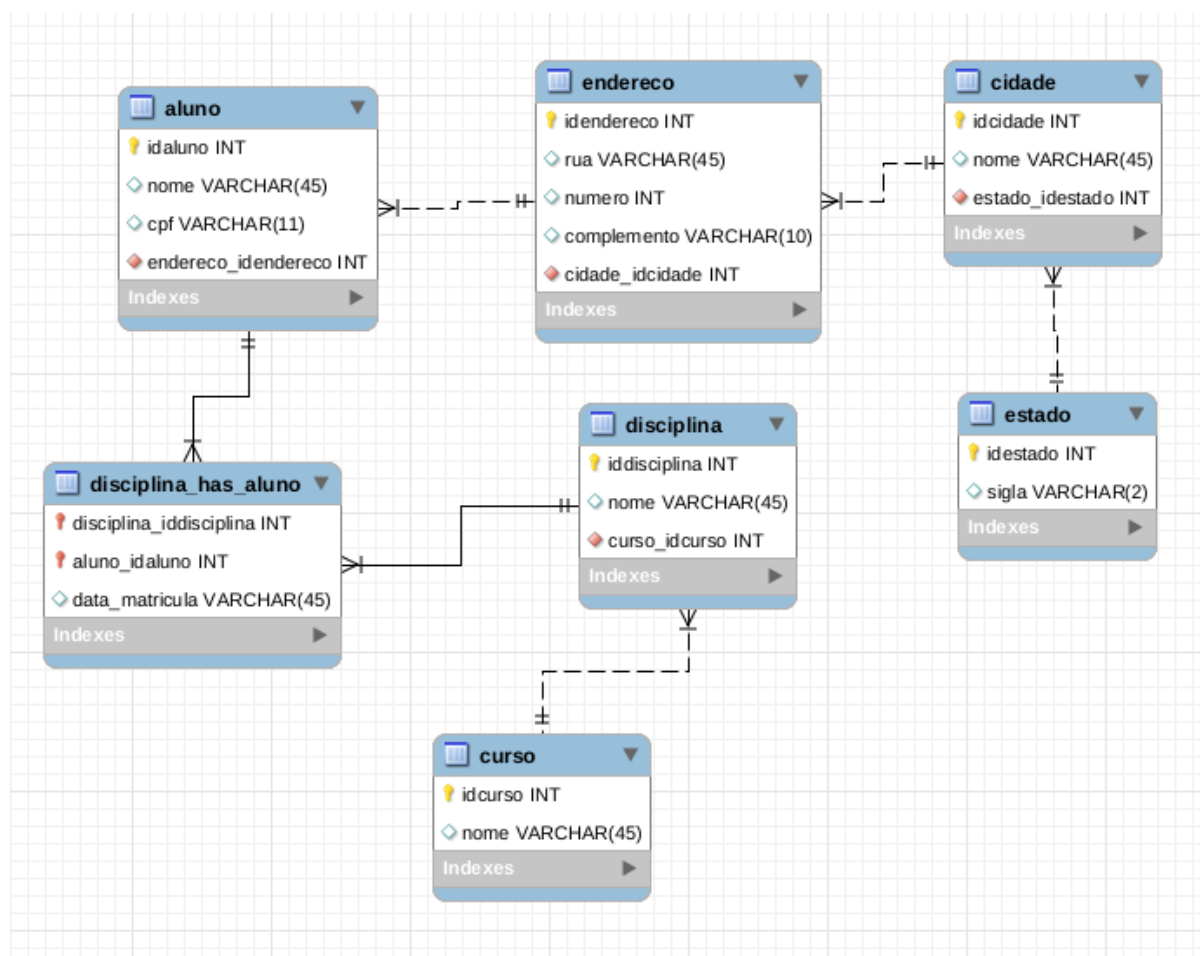
As fontes de dados utilizadas nesta simulação estão disponíveis para download no link abaixo. Suas estruturas são apresentadas nas figuras a seguir.

- <https://drive.google.com/drive/folders/1UEMm6fCMykWbULchpPp64yQGERTzmpb0?usp=sharing>.

Figura 22 – Arquivo CSV dos alunos participantes no evento.

```
cpf,data
62872683860,02/03/2020
10836995941,02/03/2020
85705166913,02/03/2020
94331268644,02/03/2020
66549644806,02/03/2020
90046114758,02/03/2020
82507144540,02/03/2020
11151573719,02/03/2020
20937875852,02/03/2020
06325672058,02/03/2020
42212877452,02/03/2020
56002978375,02/03/2020
68506018313,02/03/2020
44822334466,02/03/2020
29728284462,02/03/2020
95146826922,02/03/2020
46895736780,02/03/2020
42058732503,02/03/2020
05052164280,02/03/2020
90330445311,02/03/2020
00000000000,02/03/2020
```


Figura 23 – Estrutura da base de dados Mysql dos alunos e cursos.

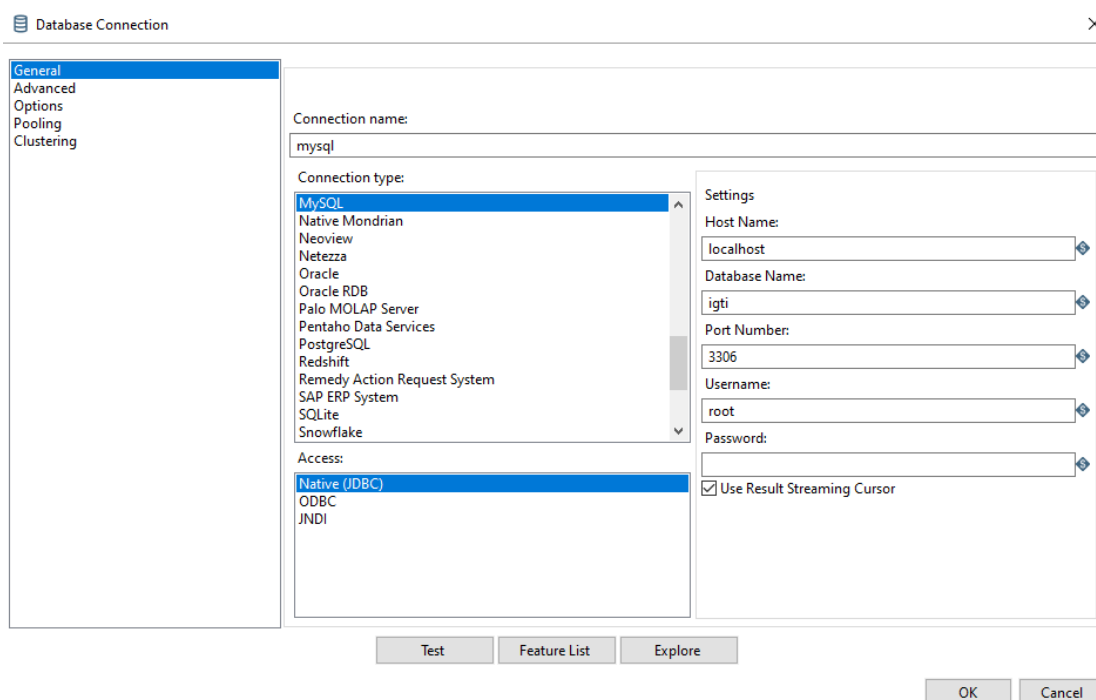


Primeiramente foi iniciado o xampp, e no navegador de internet foi acessado a página localhost/phpmyadmin, na qual foi criada uma nova base de dados chamada igti e executado o arquivo igti.sql, que já contém a estrutura e dados a serem utilizados.

No Spoon do Pentaho existem duas atividades principais: Transformations, onde são realizadas as integrações, padronizações, limpezas, exportações etc., e Job, onde são realizadas tarefas como agendamento de execuções, envio de e-mails etc. Nestas simulações iremos criar uma guia de transformations, onde iremos acessar as fontes na pasta de inputs, que é utilizada para selecionar as entradas de dados. Para o arquivo CSV basta escolher o tipo “CSV file input” e escolher a localização do arquivo. Para a base em Mysql, é necessário primeiramente baixar o arquivo de conexão Mysql jdbc connection e salvar o arquivo .jar na pasta lib, que se

encontra na raiz das pastas do Pentaho. Feito isso, é necessário escolher a opção “Table input” da pasta inputs e realizar a conexão com a base de dados que, caso esteja utilizando o xampp em suas configurações padrões, ficará como na imagem a seguir:

Figura 24 – Conexão com a base Mysql



Após a conexão, poderá ser inserido no campo SQL a consulta com os campos desejados, que neste caso será:

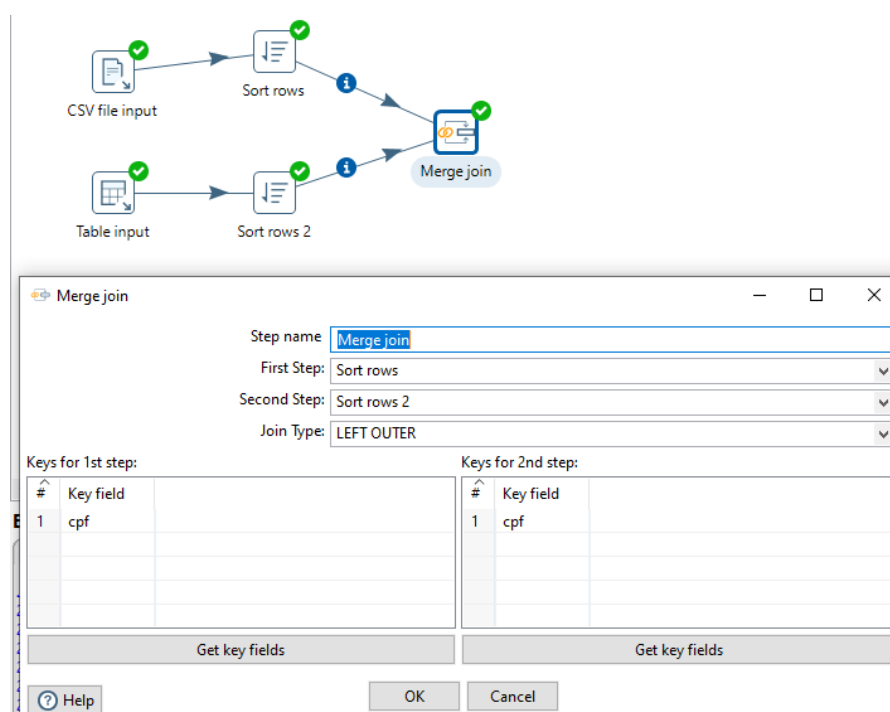
```
select aluno.cpf, aluno.nome as aluno, disciplina.nome as disciplina, curso.nome as curso
from aluno inner JOIN disciplina_has_aluno on
aluno.idaluno=disciplina_has_aluno.aluno_idaluno
inner join disciplina on disciplina_has_aluno.disciplina_iddisciplina=disciplina.iddisciplina
inner join curso on disciplina.curso_idcurso=curso.idcurso;
```

Para cada uma das entradas iremos ordenar pelo CPF, utilizando o recurso “Sort rows” da pasta transform, para posteriormente realizar a junção dos dados através do recurso “Merge Join” da pasta Joins. Essa ordenação é fundamental para que a junção através do Merge funcione corretamente e em nosso caso ambas foram

ordenadas pelo CPF. Sempre que quiser interligar um recurso ao outro, basta segurar a tecla shift e arrastar com o mouse entre os dois recursos desejados.

Clique duas vezes no recurso Merge Join para realizar a configuração da junção conforme apresentado na Figura 26. Veja que devemos escolher os dois Sorts como primeiro e segundo passo. Como estamos interessados nas presenças dos alunos nos eventos, vamos ordenar pela esquerda (left outer), pois assim ele se orientará pelo first step, que é o arquivo CSV. Nas chaves, devemos escolher os campos que serão comparados para a junção, que neste caso é o CPF.

Figura 25 – Junção das entradas.

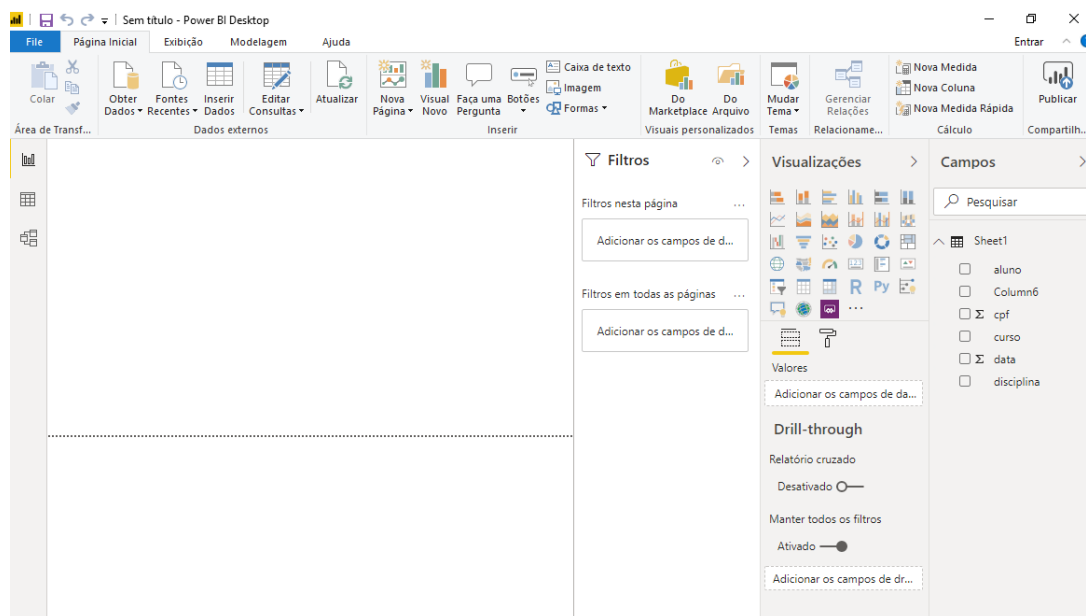


Agora só falta criar o arquivo de saída com o resultado da junção. Neste caso iremos utilizar um arquivo do Excel, necessitando apenas de escolher o recurso “Microsoft Excel writer” na pasta Output, criando uma ligação entre o recurso merge e escolhendo o local para os arquivos serem salvos. Agora basta rodar a aplicação que todo processo será executado.

Com o arquivo do excel já preparado com os dados das duas fontes integradas, iremos utilizar o Power BI para realizar as análises. Ao entrar no sistema,

clique em obter dados e escolher a fonte excel, para logo após selecionar o arquivo desejado.

Figura 26 – Tela principal do Power BI.



Como podemos ver na Figura 27, a tela principal do Power BI apresenta diversas guias e botões para realizar diversas tarefas. Ao lado esquerdo temos um quadro branco onde iremos montar os painéis, gráficos, tabelas e demais visualizações que irão compor nosso dashboard. Ao lado direito temos três colunas: Filtros, para selecionar quais dados serão exibidos; Visualizações, onde iremos selecionar o tipo de visualização, fazer parametrizações, formatações etc.; e a coluna Campos, onde iremos selecionar quais dados serão utilizados em cada visualização.

Neste exemplo, iremos criar um gráfico e uma matriz para apresentar as participações dos alunos nos eventos baseado nos cursos que eles fazem. Primeiramente, selecionamos os campos curso e CPF na coluna campos e o gráfico de colunas em visualizações; nesta mesma coluna no campo valor, selecione para o campo CPF apresentar no formato contagem, para que seja exibida quantas pessoas fizeram os cursos. Veja que para todos tipos de visualização existe um botão Formato, onde pode ser escolhido tamanho de fonte, títulos, cores etc.

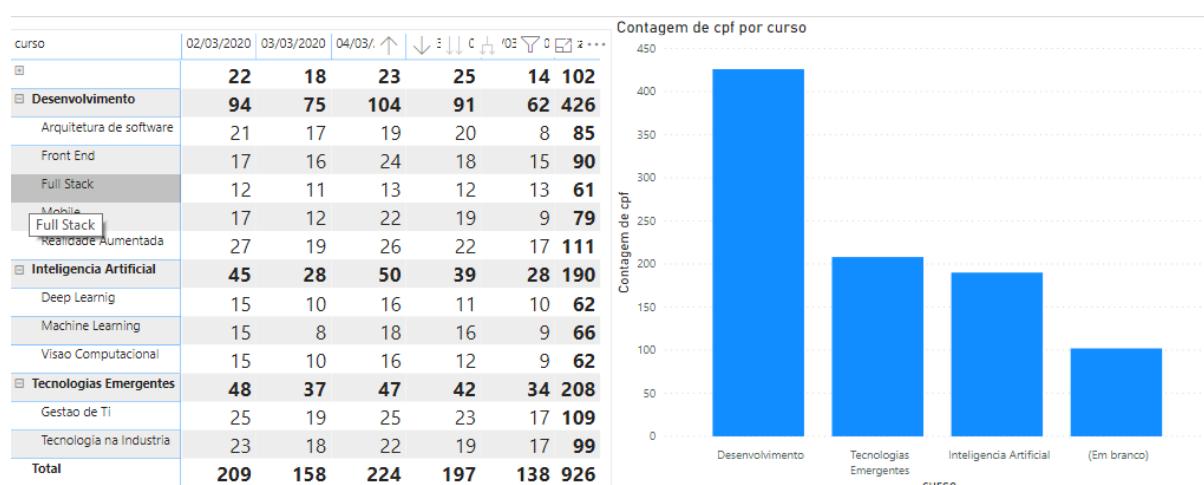
Você pode ajustar este gráfico no quadro branco para posicioná-lo e movê-lo para o local mais adequado. Clique fora do gráfico para criarmos agora uma tabela, contendo a quantidade detalhada de alunos por curso e disciplinas nos eventos.

Selecione os campos curso, disciplina, data e CPF (também no formato contagem), selecione a visualização Matriz, onde iremos montar um cubo com os dados que queremos. Para a visualização ficar mais adequada, colocamos como linhas os campos curso e disciplina; como colunas o campo data; e como valores o campo CPF. O campo data foi importado no formato inteiro, mas isso é facilmente corrigido acessando a guia modelagem na barra superior e alterando o formato do campo.

Após poucas formatações e ajustes nos tamanhos das visualizações, o resultado ficou como apresentado na Figura 28. Este dashboard pode ser publicado e acessado por outras contas, colocados em algum monitor etc., basta criar uma conta no power bi.

Esta foi uma pequena demonstração de algumas ferramentas para auxílio nos projetos de BI, com uma simples apresentação dos dados em formato excel e sem criações de datamarts e outros recursos mais avançados, que são vistos com detalhes em outras disciplinas do IGTI.

Figura 27 – Painel com apresentação dos dados no Power BI.



Referências

CECI, Flávio. *Business intelligence*. Palhoça: UnisulVirtual, 2012. Disponível em: <https://docplayer.com.br/529955-Business-intelligence-flavio-ceci.html>. Acesso em: 11 nov. 2020.

CETAX CONSULTORIA. Disponível em: < <https://www.cetax.com.br/> >. Acesso em: 11 nov. 2020.

GARTNER GROUP. Disponível em: < <https://www.gartner.com/en> >. Acesso em: 11 nov. 2020.

KIMBALL et al. *The Data Warehouse Lifecycle Toolkit*. 2. ed. Indianapolis: John Wiley & Sons Inc., 2008.

KIMBALL, Ralph; ROSS, Margy. *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*, Third Edition. Indianapolis: John Wiley & Sons Inc., 2013.

MOSS, Larissa; ATRE, Shaku. *Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications*. Addison-Wesley Professional, 2003.

PASTORI, Eduardo. *Nível de maturidade em business intelligence*. 2012. Dissertação (MBA) - Controladoria e Finanças, Universidade do Vale do Rio dos Sinos, Porto Alegre, 2012.

SELL, Denilson. *Uma arquitetura para business intelligence baseada em tecnologias semânticas para suporte a aplicações analíticas*. 2006. Tese (Doutorado) - Centro Tecnológico, Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Engenharia de Produção, Florianópolis, 2006.

SILVA, Dhiogo Cardoso da. *Uma arquitetura de business intelligence para processamento analítico baseado em tecnologias semânticas e em linguagem natural*. 2011. Dissertação (Mestrado) – Universidade Federal de Santa Catarina, Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, Florianópolis, 2011

TURBAN, E.; SHARDA, R.; ARONSON, E.; KING, David. *Business Intelligence: Um Enfoque Gerencial*. Bookman, 2009.

XAVIER, Fabrício; PEREIRA, L. *SQL dos conceitos*. Rio de Janeiro: Ciência Moderna, 2009.