

A Data Warehouse é um dos principais componentes de um sistema de BI que contém as fontes de dados.

As Data Warehouse destinam-se a trabalhar com dados utilizados para processamentos OLAP.

Algumas organizações referem-se ao processo de priorização de projetos de BI como sendo governança do BI.

A analytics descritiva ajuda os gestores a compreenderem os acontecimentos atuais das organizações, incluindo causas, tendências e padrões.

O desejo que um cliente tem para adaptar (personalizar) um produto, faz parte dos factores do ambiente de negócios na categoria de exigências de consumidor.

Aplicações baseadas em sensores e em dados de localização e que contribuem para o crescimento exponencial dos dados recolhidos chamam-se RFID.

À medida que aumenta o número de potenciais aplicações de BI surge a necessidade de as justificar e priorizar. Esta não é tarefa fácil devido ao elevado número de intangíveis.

Business Intelligence é um termo genérico que combina arquiteturas, ferramentas, bases de dados, ferramentas analíticas, aplicações e metodologias.

A analytics prescritiva ajuda os gestores a tomarem decisões de modo a poderem alcançar um melhor desempenho no futuro.

A analytics preditiva ajuda os gestores a compreenderem resultados futuros prováveis.

A Sabre utiliza dashboards para apresentar medidas de desempenho de uma forma concisa aos seus gestores.

A interface é um dos principais componentes de um sistema de BI que se baseia habitualmente num browser que muitas vezes apresenta um portal ou um dashboard.

A existência de uma força de trabalho, faz parte dos fatores do ambiente de negócios na categoria social.

As organizações que utilizam BI estão habitualmente à procura de fechar a lacuna entre o desempenho existente na organização e o desempenho desejado.

Muitos consultores e profissionais de BI envolvidos em iniciativas de BI bem-sucedidas aconselham que uma estrutura de planeamento é uma pré-condição necessária.

Com fluxos de dados em tempo real, os gestores podem ver o estado atual dos seus negócios e identificar problemas rapidamente.

O modelo de Kimball, também conhecido como a abordagem de data mart, é uma abordagem bottom-up, “planejar grande, construir pequeno”.

Os metadados descrevem a estrutura e o significado dos dados, contribuindo para o seu uso efetivo.

A integração das aplicações da empresa é um mecanismo que integra as funcionalidades das aplicações e as partilhas em todo o sistema, permitindo assim flexibilidade e reutilização.

A integração da informação da empresa é um mecanismo para extrair dados de sistemas fonte de modo a satisfazer um pedido de informação. É um espaço de ferramentas em evolução que promete integração de dados em tempo real a partir de uma variedade de fontes tais como bases de dados relacionais ou multidimensionais, serviços web, etc.

O modelo de Inmon, também conhecido como abordagem EDW, enfatiza o desenvolvimento top-down.

OLTP é um termo utilizado para um sistema de transações que é antes de mais responsável pela captura e armazenamento de dados relacionados com as funções do dia a dia das organizações.

Grande parte das data warehouses são construídas utilizando sistemas de gestão de bases de dados relacionais para controlar e gerir os dados.

A modelação dimensional é um sistema de extração de informação que suporta acessos com consultas que envolvem grandes volumes de dados.

O papel responsável pela gestão e administração de uma data warehouse com sucesso é o administrador da data warehouse que deve estar familiarizado com software, hardware e tecnologias de rede de alto desempenho, bem como possuir uma sólida compreensão do negócio.

Uma arquitetura de data warehouse federada envolve a integração (lógica/física) de sistemas díspares e de recursos analíticos de múltiplas fontes, de modo a corresponder às necessidades em mudança ou às condições do negócio.

No estudo de caso das agências estatais do Michigan, a abordagem empresarial é preferida à abordagem de desenvolver plataformas de BI/DW separadas para cada área de negócio ou agência estatal.

Podemos dizer que os 3 principais tipos de data warehouse são: data marts, operational data stores e enterprise data warehouse.

Executar ETL exaustivamente para mover os dados para a data warehouse, pode ser sinal de uma pobre gestão de dados e da falta de uma estratégia de gestão dos mesmos.

Uma arquitetura de data warehouse hub-and-spoke é utilizada para construir uma infraestrutura escalável e sustentável, a qual inclui uma data warehouse centralizada e vários data marts dependentes.

Podemos dizer que OLAP é a técnica analítica mais habitual em ambientes de data warehouse.

Em termos de conteúdo e formato, há apenas algumas categorias de relatórios do negócio: informal, formal e pequeno.

A Travel and Transport criou um sistema de BI self-service que permite aos clientes aceder diretamente à informação.

No estudo de caso da Mace, a aplicação IBM Cognos permitiu a criação e integração rápida de relatórios de 60 países, substituindo a utilização de folhas de cálculo grandes e complexas.

Os gráficos, diagramas e outros elementos visuais típicos utilizados em aplicações baseadas em visualização envolvem duas dimensões.

No estudo de caso da Blastrac, a aplicação de analytics da Tableau foi utilizada para substituir a utilização massiva de folhas de cálculo que eram carregadas a partir de múltiplos sistemas ERP.

Os dashboards apresentam representações visuais de informação importante que são consolidadas e mostradas num só ecrã.

No estudo de caso da DeltaLloyd Group, a etapa final é a etapa do processo de criação de relatórios na qual os dados consolidados são citados, formatados, e descreve a forma final do texto do relatório.

A gestão do desempenho do negócio (BPM) compreende um ciclo fechado de processos que liga a estratégia de execução com o objectivo de otimizar o desempenho do negócio.

Os dashboards de desempenho permitem operações de drill-down ou drill-through que permitem aos utilizadores aceder às fontes de dados que estão “por baixo” e obter mais detalhes.

Relatórios de métricas de gestão são utilizados para gerir o desempenho do negócio através resultados orientados a métricas.

Uma métrica alinhada com a estratégia é normalmente conhecida como indicador chave de desempenho,

No estudo de caso da empresa Saudi Telecom, as aplicações de visualização de informação permitem aos gestores ver tendências e corrigir alguns aspetos antes que se tornem problemas.

Um relatório de negócio é artefacto de comunicação, que diz respeito a assuntos do negócio, preparado com a intenção específica de transmitir informação com bom aspecto.

Com os dashboards, a camada de informação que utiliza dados gráficos, abstratos para manter separadores contendo métricas chave de desempenho é a camada de monitorização.

A perspetiva dos processos internos do negócio da organização sugere que o balanced scorecard se foca nos processos de negócio e como eles decorrerem. Diagramas PERT ou diagramas de rede mostram relações de precedência entre tarefas/atividades de projetos.

Os mapas são habitualmente utilizados em conjunto com outros gráficos e diagramas, a eles sobrepostos, e mostram códigos postais, nomes de países, etc.

O analytics visual é amplamente reconhecido como uma combinação de visualização com analytics preditiva.

Gráficos de barras são úteis para mostrar dados nominais ou numéricos que se dividem bem em diferentes categorias de forma a poder comparar rapidamente resultados e tendências.

Tal como é descrito no estudo de caso da 2degrees, um problema comum na indústria das telecomunicações móveis é definido como churn, o que significa que o cliente abandona a operadora.

Devido ao seu sucesso na aplicação em problemas de negócio na indústria de retalho, a mineração com regras de associação é também chamada como market-basket analysis,

Extracção de conhecimento, análise de padrões, descoberta de conhecimento, KDD, pesquisa de padrões são nomes alternativos para mineração de dados.

No estudo de caso da aula, a Cabela utiliza a ferramenta de mineração de dados SAS para criar modelos preditivos para otimizar a selecção de clientes dos seus contactos. O algoritmo Apriori é aquele que é mais utilizado para descobrir regras de associação. Os dados estão muitas vezes profundamente escondidos em grandes bases de dados, as quais por vezes contêm dados com vários anos.

Tem havido um aumento na utilização da mineração de dados para lidar com a competição global e com as necessidades mais sofisticadas dos clientes.

O estudo de caso da mineração de dados na investigação do cancro explica que os métodos da mineração de dados são capazes de extrair padrões e relações escondidas profundamente em grandes e complexas bases de dados médicas.

Há muitos séculos que os humanos têm vindo a extrair manualmente padrões dos dados, mas o volume crescente dos dados nos tempos modernos, criou a necessidade de uma abordagem mais automatizada.

A ideia básica das árvores de decisão é a de recursivamente dividir o conjunto de treino até que a divisão consista completa ou principalmente em dados pertencentes a uma só classe.

Fayyad et al. (1996) definiu descoberta de conhecimento em bases de dados como o processo de utilizar métodos de mineração de dados para encontrar informação e padrões nos dados.

Os dados categóricos representam etiquetas de várias classes utilizadas para dividir a variável em grupos. São exemplos raça, sexo e nível de educação.

Na k-Fold Cross Validation, o conjunto de dados é dividido de forma aleatória em conjuntos mutuamente exclusivos de tamanho aproximadamente igual e é testado múltiplas vezes em cada um dos conjuntos obtidos, utilizando os outros como conjunto de treino.

Enquanto que a estatística começa com um conjunto de proposições e hipóteses bem definidas, a mineração de dados começa com uma afirmação genérica.

A preparação dos dados, a terceira fase da metodologia CRISP-DM, é também conhecida por preprocessamento dos dados.

No estudo de caso dos fundos a terroristas, um desvio observado no preço pode estar relacionado com evasão fiscal, lavagem de dinheiro ou financiamento a atividades terroristas.

O processo de mineração de texto pode ser dividido em três tarefas consecutivas, a primeira das quais é definir o corpus.

Quando uma palavra tem mais de um significado, seleccionar o significado com mais sentido, apenas pode ser alcançado tendo em conta o contexto no qual a palavra está a ser utilizada. Este conceito é conhecido como desambiguação do sentido da palavra.

O Watson da IBM utiliza uma arquitectura computacional massivamente paralela, focado em mineração de texto, baseada em evidências probabilísticas chamada DeepQA.

As páginas Web contêm quer informação não estruturada, quer hiperligações, que são conexões com outras páginas Web.

Os Web crawlers/spiders são utilizados para lerem automaticamente o conteúdo de sítios Web.

Sentiment Analysis é uma técnica utilizada para detectar opiniões favoráveis e desfavoráveis sobre produtos e serviços determinados, utilizando grandes quantidades de fontes de dados textuais.

Dado que a matriz de termos por documento é muitas vezes esparsa, um importante passo na optimização é reduzir a dimensionalidade da matriz.

A maioria dos Big Data é gerada automaticamente por máquinas.

O Hadoop é, antes de mais, um sistema distribuído de ficheiros ao qual faltam as capacidades que associamos aos Sistemas de Gestão de Bases de Dados.

No mundo dos Big data, MapReduce ajuda no processamento e na análise de grandes volumes de dados multi-estruturados.

O rastreador de tarefas é um nó num cluster Hadoop que inicia e coordena as tarefas MapReduce, ou o processamento dos dados.

Veracidade refere-se à conformidade dos factos: exactidão, qualidade, veracidade, confiança dos dados.

O analytics em movimento é muitas vezes negligenciado no mundo do BI e do Big data.

O Big Data emprega técnicas de processamento paralelo e capacidades de armazenamento de dados não relacionais, de modo a processar dados estruturados e semi-estruturados.

À medida que a complexidade dos sistemas analíticos aumenta, a necessidade de sistemas analíticos mais eficientes também aumenta de modo a obter um melhor desempenho.

Todas as afirmações seguintes sobre mineração de dados são verdadeiras EXCEPTO a mineração de dados deve ser um processo de um só passo para chegar aos resultados.

A arquitectura do sistema que suporta o Watson (do estudo de caso) utiliza todos os seguintes, EXCEPTO

c) um mecanismo central que pode ser operado ininterruptamente noutros domínios sem alterações.

O Business Intelligence (BI) pode ser caracterizado por uma transformação de

b) dados em processamento, em informações, em acções.

A analytics visual pretende responder a quais dos seguintes tipos de questões?

b) Porque é que aconteceu?

Em que fase da extracção, transformação e carregamento de uma data warehouse, os dados são agregados?

c) transformação

O campo de dados “grupo étnico” pode ser melhor descrito com

b) dados nominais.

Quando se representam os dados de uma data warehouse, utilizando diversas tabelas dimensão ligadas a uma tabela de factos, qual a representação que estamos a utilizar?  
c) esquema em estrela.

Que tipo de data warehouse é criada separadamente da data warehouse da empresa, por um departamento, não dependendo de actualizações?

a) data mart independente.

Os Big data muitas vezes envolvem uma forma de armazenamento e processamento distribuídos utilizando Hadoop e MapReduce. Uma das razões para isso é:

d) a capacidade de processamento necessária para o modelo centralizado iria sobrecarregar um processador utilizado de forma isolada.

No estudo de caso da Magpie Sensing, a recolha automática de dados sobre a temperatura e a humidade dos produtos embalados apoiou vários tipos de analytics.

Qual dos seguintes é um exemplo de analytics preditiva?

a) avisos sobre a possibilidade de abertura de um selo.

Qual a área da mineração de dados que analisa os dados, formando regras que permitem distinguir entre diferentes classes?

d) classificação.

Qual das metodologias/processo de mineração de dados é a/o mais popular e completo, de acordo com o ranking da Kdnugget.com?

b) CRISP-DM

Todas as afirmações seguintes sobre metadados são verdadeiras EXCEPTO:

a) para a maioria das organizações, os metadados da datawarehouse são uma despesa desnecessária.

Qual é o desafio fundamental no desenho de um dashboard?

d) assegurar que toda a informação necessária é mostrada claramente num único ecrã.

Mineração do conteúdo da Web diz respeito a analisar o conteúdo não estruturado das páginas da Web. Verdadeiro

A analytics de texto é um subconjunto da mineração de texto que lida com extracção de informação, mais mineração de texto.

a) falso.

Os sistemas OLTP lidam com as tarefas rotineiras das organizações. Pelo contrário, uma data warehouse normalmente é

d) um sistema distinto que fornece armazenamento para os dados que serão depois utilizados para análise.

Qual dos componentes de um sistema de relatórios do negócio contém passos que detalham como é que as transacções armazenadas são convertidas em métricas e em dashboards?

b) lógica do negócio.

Qual é a arquitectura de data warehouse que utiliza uma data warehouse normalizada que alimenta múltiplas data marts?

b) arquitectura de data warehouse hub-and-spoke.

Qual dos seguintes NÃO é um exemplo que se situa nas quatro categorias dos factores do ambiente de negócios das organizações de hoje em dia?

a) menos regulamentações governamentais.