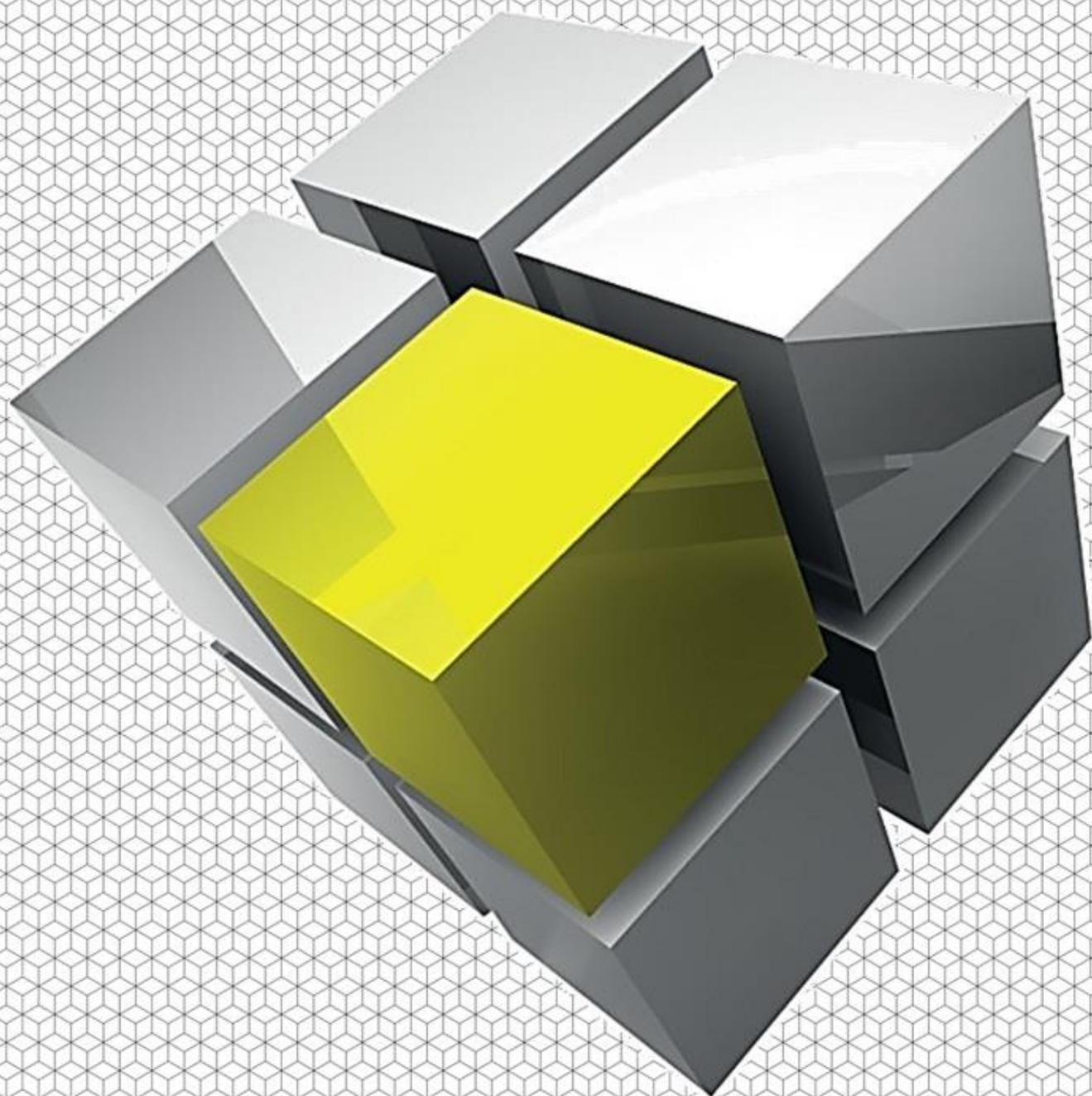


DARK ANALYTICS

PROSPECÇÃO DE INSIGHTS ESTRATÉGICOS



FUNDAMENTALS GUIDE

João Roberto Peres



“Nossa memória virou um depósito de informações inúteis.”

Augusto Cury



“A maneira como você coleta, gerencia e utiliza as informações determina se você vai vencer ou perder”

Bill Gates

Este e-book compreende a síntese do conteúdo apresentado durante o IT Fórum Expo 2018, no dia 17/10 - no Palco "High Tech", sobre o tema "Dark Analytics - prospecção de insights estratégicos".

O objetivo é explanar de forma simples o significado tático do "Dark Analytics" e o seu grande impacto para os negócios, advindos das tecnologias aplicadas. Em poucas páginas ideio aclarar o tema aos executivos e interessados.

Boa leitura!*

**João Roberto Peres é pesquisador sobre inovações nos temas ligados à Ciência da Computação e Segurança Cibernética. Atua nos últimos 35 anos como professor, palestrante, empresário e consultor (KOMP/FGV). Possui formação superior em Engenharia Elétrica e Ciência da Computação, entre outras, com cursos e certificações internacionais ligadas a área de Segurança Empresarial.*

É especialista em Governança, Riscos e Compliance (GRC), Prevenção de Fraudes e Corrupção, Perícia Computacional, Direito Digital, investigação e defesa de Ciber Crimes, IoT, DS, Big Data & Analytics. É autor de diversas publicações, artigos e e-books, coautor do Guia de Referência "Segurança Corporativa", da Comissão Direito Eletrônico e Crimes de Alta Tecnologia – OAB/SP. É membro do Grupo de Estudos Temáticos de Direito Digital e Compliance da FIESP e da OAB-SP, participante do estudo "Bytes de IoT – Internet das Coisas: um plano de ação para o Brasil" e membro da comunidade "Tudo sobre IoT", entre outros. - <http://www.tudosobreiot.com.br/> - <https://www.linkedin.com/in/jo%C3%A3o-peres-a0521017/>

Sumário e Indicações

O conteúdo apresentado neste **micro-eBook** é objetivo (entre as páginas 06 até 33) e inclui complementos anexos, para a leitura dos interessados. Desejo aos leitores uma rápida absorção das informações e o melhor aproveitamento possível dos conceitos em suas atividades laborais. Estarei á disposição para maiores elucidações.

João Roberto Peres

Conteúdo Temático indexável

- [Introdução – Do que estamos Falando](#)
- [Dark Data o que é?](#)
- [Onde estão nossos dados?](#)
- [Como tratar o Lixo Dark Data?](#)
- [Dark Data na Deep Web](#)
- [O que se espera do Dark Analytics?](#)
- [Aplicações do DA](#)
- [Potencial de Mercado do Dark Analytics.](#)
- [Impactos e Riscos do DA](#)
- [Considerações Finais](#)
- [Anexos Complementares](#)

Esta publicação é uma versão preliminar e gratuita - SEM REVISÃO – será comercializada a partir de janeiro de 2019 e publicada com as retificações ortográficas necessárias e formatos adequados aos diversos públicos. Não descartamos a possibilidade de expandir o conteúdo em abrangência na versão comercial. Contamos com a colaboração dos leitores; em apontar divergências, erros de grafia e gramaticais ou colaborar com conteúdos adicionais.



JOÃO ROBERTO PERES

DARK ANALYTICS

PROSPECÇÃO DE INSIGHTS ESTRATÉGICOS

FUNDAMENTALS GUIDE

1ª edição

São Paulo
Edição do Autor
2018

Do que estamos falando?

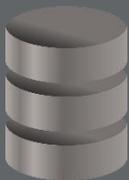
Em 2018 o universo de dados armazenados pela humanidade pode chegar ao exorbitante número de **32 ZB** (Zetabytes - 1 ZB = 1000⁷ bytes, ou 32 ZB = 32.000.000.000.000.000.000.000 bytes), mas a previsão é que em **2025** possamos ter próximo de **163 ZB**, segundo a IDC*, na pesquisa encomendada pela fabricante de Hard Disk SEAGATE*.

O problema é que cerca de **90%** desse volume de dados armazenados, não servem no momento, para nada, e são considerados **lixo Digital** ou **Dark Data!**



Dark Data o que é?

A Gartner Inc. descreve os dados escuros como *"recursos de dados e informação que as organizações coletam, processam e armazenam no decorrer de suas atividades comerciais regulares, mas não conseguem usar para outros fins"*



De acordo com o especialista em tecnologia de pesquisa Lucidworks, as empresas tipicamente apenas analisam cerca de 10% dos dados que coletam. Segundo o estudo, 60% dos dados do planeta estão nos países mais desenvolvidos, mas a proporção deve se inverter com o tempo. Em 2020, 60% dos dados estarão em países emergentes como Brasil, China, Índia e México.

Dark Data ou Dados Escuros

Dark Data compreendem parte dos dados estruturados armazenados por empresas, incluindo os dados desestruturados, também armazenados, como arquivos de texto, tabelas, figuras, logs, e-mails, SMS, Apps Data, áudios, imagens, vídeos, IoT Data, backups, ZIP data, PDF data, Geolocation Data, (HTTPS) Referral Data, dados da Surface Web, Deep e Dark Web, entre outros.

A International Data Corporation (IDC) declarou que 90% dos dados não estruturados nunca são analisados.

Ela estima que até 2020, cerca de 37% do universo digital "dark" conterá informações que podem ser valiosas, se analisadas.



Onde estão nossos dados?

Dados Gerados e Armazenados

- Nas empresas – estruturados e desestruturados;
- Nas Redes em trânsito;
- Armazenados nas Nuvens privadas e públicas – em (HDC- Hyperscale Data Centers);
- Armazenados em ambientes Fog Computing, Edge Computing, Devices IoT, CPS...
- Armazenados em desktops, laptops, tablets, smartphones, Routers, DVRs, ARUs...
- Armazenados em SSDs, HDs externos, CDs/DVDs, pen drives, cartões de memória...;
- Armazenados em Firmwares, Stock ROM / RAM, NVRAM, ...
- Armazenados e distribuídos nos hosts das várias camadas da WWW....
- ...



Como tratar o Lixo Dark Data?

Através da “**Análise de Dados**” ciência que já possui mais de 50 anos nos meios acadêmicos e que a partir de 2015 foi denominada de “Ciência de Dados” ou “**Data Science**”.

De que forma?

Projetando e desenvolvendo pacotes **ETL** (Extract Transform and Load) escaláveis, extraíndo dados e criando “**Big Dark Data**”, aplicando técnicas de **Dark Analytics** ou “Predictive / Prescriptive / Cognitive Analytics”, sobre dados brutos selecionados, produzindo estatísticas e informações seletas para obter através da “**Decision Science**” > insights e soluções para promover ações e decisões assertivas e seguras para o futuro.

Hoje, “Dark Analytics”, opera com foco em Big Dark Data, entre outras técnicas, linguagens de programação e novas ferramentas de “Advanced Analytics”, que incluem AI (Artificial Intelligence / Machine Learning...), que permitem extrair informações de grande valor para as organizações.



Principais locais para obter Dark Data

Algumas plataformas Dark Analytics:
[DeepDive](#) - [Dark Vision](#)
- [Snorkel](#)



Dark Data na Deep Web

Deep Web ou **Web profunda** compreende as diversas redes que não são indexadas por mecanismos de busca como o Google, portanto, ficavam ocultas para a maioria dos usuários da Internet. De forma similar aos governos, **as empresas** poderão monitorar o volume e o contexto das atividades na Web profunda para identificar **possíveis ameaças**, ampliar a **inteligência competitiva**, etc. usando uma variedade de ferramentas de pesquisas emergentes, projetadas para ajudar os usuários na busca de **dados estratégicos** e referências sobre seus **negócios**.

Já existem alguns mecanismos de busca (search engines) para a **Deep Web**:

- **Shodan** (www.shodan.io) IPv4 – The Search Engine for the Internet of Things.
 - **Zmap** (www.zmap.io) IPv4
 - **Censys** (www.censys.io) IPv4
 - **MrLooquer** (<https://mrlooter.com/>) IPv6
- ... Use com CUIDADO!

Busque informações adicionais antes de navegar.



O que se espera do Dark Analytics?

O uso de “**dark data**” passa a ser estratégico para as organizações de todos os portes e ramo de atividades, pois estes dados obscuros, podem oferecer inúmeras oportunidades de **insights** para “**NOVOS NEGÓCIOS**” e potencialização do “**market share**”, além de permitir ampliar o **nível de segurança e “reduzir custos**” para as organizações.

O **DA (Dark Analytics)** permite:

- Explorar o universo de dados “internos e externos” da organização, ampliando a **Inteligência Competitiva** na busca de **correlações estratégicas**, que possam alcançar visões e entendimentos sobre **o mercado e a própria organização**, até pouco tempo atrás, indisponíveis;
- Reduzir os “**Custos de Oportunidades**”, com a exploração sistemática de fontes abertas de dados, nas Redes Sociais, na Surface e na Deep Web, agora possível;
- Reduzir os “**Riscos de Conformidade**”, avaliando dados **DARK** que possam levar a exposição indevida e envolver responsabilidade legal e financeira da empresa;
- Reduzir os “**Riscos de Inteligência e de Reputação**” por dados e informações abertas em dados **DARK**, que possam prejudicar ou desabonar a empresa.
- ...



Aplicações do DA

Algumas opções:

- Análise de dados e informações de Redes **IoT – M2M, CPS, MA (Mobile All)**, ...
- Vendas Atacado – Análise e **Gestão Concorrencial**.
- Negócios de Varejo – Análise e Gestão dos Consumidores (**Hypermarket**).
- Serviços **Bancários e Financeiros** – Análise e Gestão
- Serviços de Saúde – Análise e Gestão – **subsídios para pesquisas avançadas**.
- Recursos Humanos – Análise e Gestão (suporte em **Controles e Compliance**).
- Indústria da Educação – **Análise comportamental** e evolutiva do aprendizado
- Todos os ramos Industriais – **pesquisa, inovação, redução de custos e perdas...**
- Orçamento Empresarial – **Análise** e subsídios estratégicos
- Pesquisas Médicas e Científicas – Exploração de **variáveis ocultas** e novos insights
- Segurança Patrimonial – Análise de evidências, **imagens, sons e Gestão melhorada**
- Segurança da Informação – **Evolução e efetividade** dos sistemas SIEM (gestão de eventos e evidências)
- **Conformidade e Programas de Integridade – Análise e Gestão – apuração de desvios e fraudes**
- Negócios em Geral – potencializar a **IC** (inteligência Competitiva) e a proteção estratégica
- Governo – **Comportamento Social** – Análise e Gestão Populacional
 - ...



Potencial de Mercado do Dark Analytics.

Importante fonte* indica que o Mercado das Análises Escuras vale até **867,7 milhões de dólares até 2023** – atuando principalmente em (Soluções, Serviços de Análise), Modo de Implantação (Cloud, Local), Aplicativos de Negócios (Marketing, Operações, Finanças, Recursos Humanos), Vertical (BFSI (Banking, Financial Services and Insurance), IoT Industry, Healthcare)... **MarketsandMarkets™**

Principais players poderosos do mercado internacional em DA: IBM, SAP, Amazon Web Services, Micro Focus, Dell EMC, AvePoint, Teradata, Komprise, Zoomdata, Globanet, SAS, Microsoft, Commvault, Symantec, Datameer, Datumize, Veritas, IRI, Formcept, SynerScope, Esdebe, Quantta Analytics, Levi, Ray & Shoup, VMware Inc., ...

■ ...



*Fonte: < <https://www.marketsandmarkets.com/> > disponível em 03/09/2018.



Potencial de Mercado do Dark Analytics.

Nas operações de negócios, a **análise de dados Dark** oferece a **competência imprescindível**, sem a qual, qualquer empresa **não pode** atingir suas metas e permanecer na corrida com a concorrência. Os **insights** de dados obscuros ajudam na tomada de decisões, visando novos caminhos para o futuro, dando novas oportunidades, reduzindo riscos e aumentando o **ROI**. O potencial para todo porte de organização é altamente significativo – Observe:



*Maximize Market Research provides B2B and B2C research on 20,000 high growth emerging opportunities & technologies as well as threats to the companies across the Healthcare, Pharmaceuticals, Electronics & Communications, Internet of Things, Food and Beverages, Aerospace and Defense and other manufacturing sectors.

<<https://www.maximizemarketresearch.com/market-report/dark-analytics-market/11694>> disponível em 03/09/2018.



Impactos e Riscos do DA



Especificar as vantagens, os impactos positivos e negativos, bem como, os riscos do “Dark Analytics”, não é um exercício fácil, pois, como toda nova “Tecnologia para Negócios”, possui seus prós e contras:

- Usar DA será inevitável, pois é uma tecnologia complementar para o “Big Data & Analytics”, cujas as vantagens e desvantagens são amplamente discutidas no mercado – o uso do DA segue o mesmo caminho.
- Usar bem o DA é mandatário, tomando-se os devidos cuidados na criteriosa extração de dados, das diversas fontes, considerando a qualidade, a precisão dos dados e os riscos sobre o uso de dados confidenciais e pessoais (GDPR/LGPD...).
- Proteger dados extraídos por DA é fundamental. Criptografar e Armazenar de forma segura os “Big Dark Data Sets” selecionados, pode garantir privacidade e reduzir impactos negativos por vazamento de informações.
- Usar os resultados do DA para os devidos fins, exige empregar outras tecnologias da “Decision Science”, para confiar nos insights e soluções identificadas, que garantam decisões assertivas e seguras para o futuro.

Considerações Finais

Dark Analytics "**DA**", como tecnologia para os negócios, promove uma nova e empolgante era, **de fluxo infinito**, de reflexão e análise consistente de dados, que produzem informações relevantes, empregando de forma integrada, avançadas tecnologias em **benefício da humanidade**.

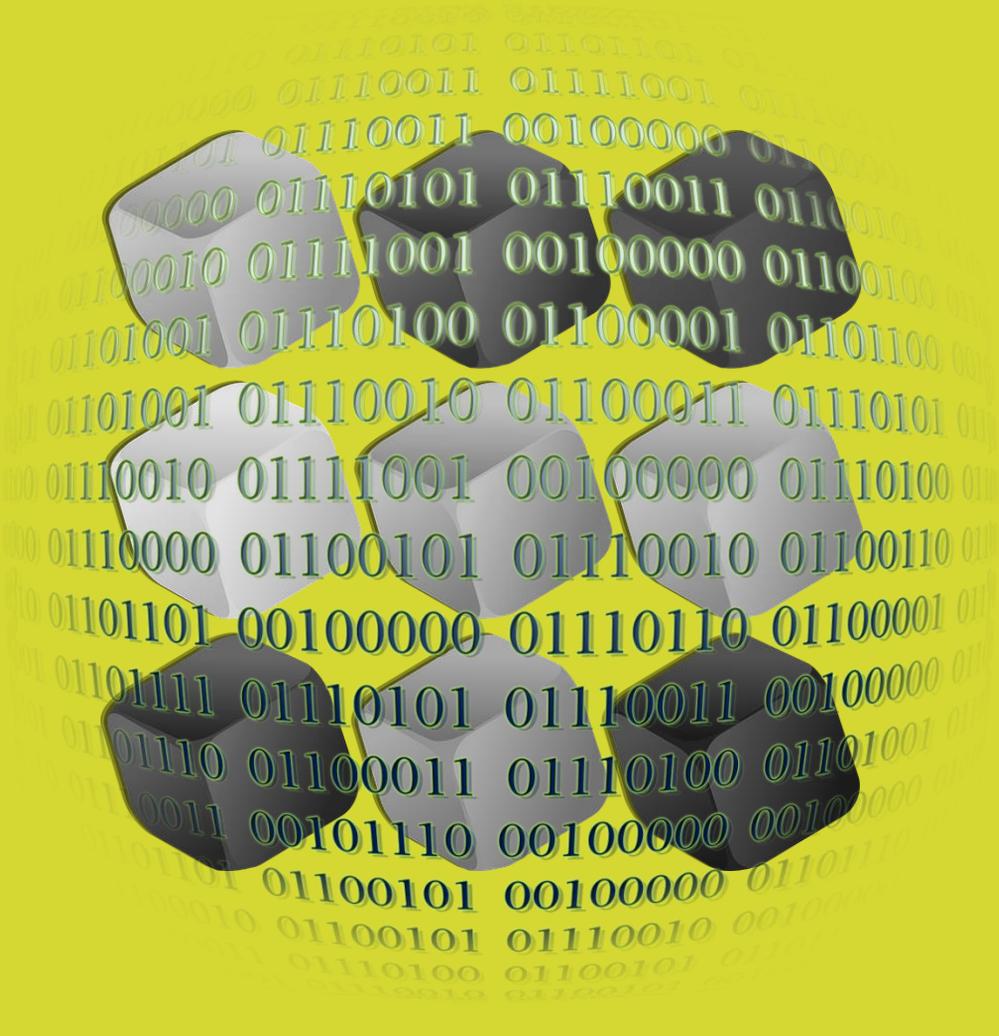
Hoje, enfrentamos ainda alguns aspectos críticos, mais perfeitamente superáveis para viabilizar o uso do **DA**, entre eles; a qualificação da mão-de-obra, o poder computacional para armazenar e processar grandes bases de dados, ferramentas mais amigáveis para extração, análise, validação, entre outros.

Para atender as demandas do **DA**, felizmente, já se conta com plataformas tipo "**DeepDive**", "**Massively Parallel Processing (MPP)**", "**Apache Hadoop**", "**Spark**", "**NoSQLNLP**", "", linguagens como "**Python**", "**R**", "**PCA**"..., sistemas de processamento paralelo de alto desempenho e baixo custo, com tecnologias "**GPU CUDA**" (Compute Unified Device Architecture), bem como as potentes "**FPGAs**" (Field Programmable Gate Arrays) que podem atender às demandas de desempenho da Inteligência Artificial Analítica e Big Dark Data, ambos em crescimento global exponencial.

Grato por sua leitura!

João Roberto Peres





Anexos Complementares

- 1 - Glossário de Termos**
- 2 - Sobre a Participação no IT FORUM X 2018**
- 3 - Corporate Friends Connection**



GLOSSÁRIO DE TERMOS TÉCNICOS



Glossário Big Data – Analytics & Data Science...

O conteúdo a seguir corresponde ao resultado de pesquisas e traduções livres que o autor realizou, sintetizando e adaptando para o português, conceitos e fundamentos, sobre os principais termos aplicados para as tecnologias envolvidas nas atividades de pesquisa de "dados escuros" utilizando a "Ciência de Dados" em Dark Analytics.

Acesso a dados

O ato ou método de visualizar ou recuperar dados armazenados.

ACID Teste

Um teste aplicado a dados para atomicidade, consistência, isolamento e durabilidade.

Administrador de banco de dados (DBA)

Uma pessoa, muitas vezes certificada, que é responsável por suportar e manter a integridade da estrutura e do conteúdo de um banco de dados.

Administrador de dados

Uma pessoa responsável pelos dados armazenados em um storage de dados ou em nuvem.

Agregação

Processo de busca, coleta e apresentação de dados.

Agregação de dados

O ato de coletar dados de várias fontes para fins de relatório ou análise.

Algoritmo

Uma fórmula matemática colocada em um software que executa uma análise em um conjunto de dados selecionados.

Análise de agrupamento

É o processo de identificar objetos que são semelhantes entre si e agrupá-los para entender as diferenças, bem como as semelhanças nos dados.

Análise no Banco de Dados

Compreende a integração da análise de dados no Data Warehouse sobre um Data Set.

Análise do Fluxo de Cliques

A análise da atividade da Web dos usuários por meio dos itens em que eles clicam em uma página.

Análise de causa raiz

Compreende o processo para determinar a causa principal de um evento ou problema.

Análise de classificação

Um processo sistemático para obter informações importantes e relevantes sobre os dados, também chamados metadados; dados sobre dados.

Análise de Cluster

Compreende o processo de identificação de objetos que são semelhantes entre si e agrupados para compreender as diferenças, bem como as semelhanças nos dados.

Análise Comportamental

Tipo de avaliação usada sobre dados do comportamento das pessoas para entender a intenção e prever ações futuras.

Análise Comparativa

Assegura um procedimento passo a passo de comparações e cálculos para detectar padrões em conjuntos de dados muito grandes.

Glossário

Análise de Dados Paralelos

Rompendo um problema analítico em componentes menores e executando algoritmos em cada um desses componentes ao mesmo tempo. A análise de dados paralelos pode ocorrer no mesmo sistema ou em vários sistemas.

Análise de Dark Data

Compreende análise sistemática dados desconhecidos, desestruturados ou escuros, que podem ficar perdidos ou armazenados como "lixo", sem a possibilidade de serem acessados ou analisados convencionalmente. A análise é realizada através de técnicas, programação e ferramentas especiais em Big Dark Data.

Análise Espacial

Refere-se à análise de dados espaciais como dados geográficos ou dados topológicos para identificar e compreender padrões e regularidades dentro de dados distribuídos no espaço geográfico.

Análise de Eventos

Mostra a série de etapas que levaram a uma ação.

Análise Exploratória

Permite localizar padrões dentro de dados sem procedimentos ou métodos padrão. É um meio de descobrir os dados e encontrar as principais características dos conjuntos de dados.

Análise de Imagem Baseada em Objeto

A análise de imagens digitais pode ser realizada com dados de pixels individuais, enquanto a análise de imagens baseada em objetos usa dados de uma seleção de pixels relacionados, chamados objetos ou objetos de imagem.

Análise de Localização

A análise de localização traz análises mapeadas e orientadas por mapas para sistemas de negócios corporativos e data warehouses. Ela permite associar informações geoespaciais a conjuntos de dados.

Análise Preditiva

Forma de usar funções estatísticas em um ou mais conjuntos de dados para prever tendências ou eventos futuros.

Análise de Rede

Análise que permite visualizar relacionamentos entre os nós em termos de rede ou teoria de grafos, ou seja, analisando conexões entre nós em uma rede e a força dos laços.

Análise de Risco dos Dados

A aplicação de métodos estatísticos em um ou mais conjuntos de dados para determinar o risco provável de um projeto, ação ou decisão – é fundamental para a aplicar na Decision Science.

Análise de Roteamento

Encontrar o roteamento otimizado usando muitas variáveis diferentes para um determinado meio de transporte, a fim de diminuir os custos de combustível e aumentar a eficiência.

Análise de Sentimentos

A aplicação de funções estatísticas nos comentários que as pessoas fazem na Web e nas redes sociais para determinar como se sentem em relação a um produto ou empresa.

Análise de Texto

A aplicação de técnicas estatísticas, linguísticas e de aprendizado de máquina em fontes baseadas em texto para obter significado ou insight.

Analítica entre Canais

Avaliação de Canais que pode atribuir vendas, mostrar o valor médio do pedido ou o valor da vida útil.

Anonimização

O corte de links entre pessoas em um banco de dados e seus registros para impedir a descoberta da origem dos registros. Fundamental para atender Regulamentos e Leis como GDPR/LGPD... entre outras.

Glossário

API (Application Program Interface)

Um conjunto de padrões de programação e instruções para acessar ou construir aplicativos de software baseados na web ou em outras plataformas.

Aplicação (APP)

Software que permite que um computador ou dispositivo móvel (tipo smartphone) execute uma determinada tarefa.

Aprendizado de máquina

O uso de algoritmos para permitir que um computador analise dados com a finalidade de "aprender" que ação executar quando um padrão ou evento específico ocorre.

Armazenamento de dados operacionais (ODS)

Um local especial para coletar e armazenar dados de várias fontes, para que mais operações possam ser executadas antes de enviar para o data warehouse para geração de relatórios.

Arquitetura e design de dados

Forma como os dados corporativos são estruturados. A estrutura ou projeto real varia dependendo do eventual resultado final requerido. A arquitetura de dados possui três estágios ou processos: 1) representação conceitual de entidades de negócios. 2) A representação lógica dos relacionamentos entre essas entidades e, 3) a construção física do sistema para suportar a funcionalidade.

Arquivo de log

Um arquivo que um computador, rede ou aplicativo cria automaticamente para registrar eventos que ocorrem durante a operação - por exemplo, a hora em que um arquivo é acessado.

Avro Apache

Avro é um sistema de serialização de dados que permite codificar o esquema de arquivos do Hadoop. É especialista em analisar dados e executar chamadas de procedimentos remotos.

Balanceamento de carga

O processo de distribuição da carga de trabalho em uma rede de computadores ou cluster de computador para otimizar o desempenho.

Banco de Dados

Uma coleção digital de dados e a estrutura em torno da qual os dados são organizados. Os dados são normalmente inseridos e acessados através de um sistema de gerenciamento de banco de dados (DBMS).

Bancos de Dados de Armazenamento de Documentos

Um banco de dados orientado a documentos, especialmente projetado para armazenar, gerenciar e recuperar documentos, também conhecido como dados semiestruturados.

Banco de Dados Colunar ou banco de dados orientado por coluna

Um banco de dados que armazena dados por coluna em vez de por linha. Em um banco de dados com base em linha, uma linha pode conter um nome, endereço e número de telefone. Em um banco de dados orientado por colunas, todos os nomes estão em uma coluna, endereços em outro e assim por diante. Uma das principais vantagens de um banco de dados colunar é o acesso mais rápido ao disco rígido.

Banco de Dados Dark

É a agregação seletiva de dados semiestruturados, desestruturados, de texto, vídeo, áudio, entre outros, que podem compor um Big Dark Data.

Glossário

Banco de Dados de Gráficos

Eles usam estruturas gráficas (um conjunto finito de pares ordenados ou certas entidades), com arestas, propriedades e nós para armazenamento de dados. Ele fornece adjacência livre de índice, o que significa que cada elemento está diretamente ligado ao seu elemento vizinho.

Bancos de dados KeyValue

Eles armazenam dados com uma chave primária, um registro exclusivamente identificável, que facilita e agiliza a consulta. Os dados armazenados em um KeyValue normalmente são algum tipo de primitivo da linguagem de programação.

Banco de dados na Memória

Qualquer sistema de banco de dados que se baseia em memória para armazenamento de dados.

Banco de dados MPP

Um banco de dados otimizado para funcionar em um ambiente de processamento maciçamente paralelo.

Bancos de dados Multidimensionais

Um banco de dados otimizado para aplicativos de processamento analítico online de dados (OLAP) e para data warehousing.

Bancos de Dados MultiValue

Eles são um tipo de NoSQL e bancos de dados multidimensionais que entendem dados tridimensionais diretamente. Eles são primariamente cadeias gigantes que são perfeitas para manipular sequências de caracteres HTML e XML diretamente

Bancos de Dados de Objetos

Eles armazenam dados na forma de objetos, conforme usados pela programação orientada a objetos. Eles são diferentes dos bancos de dados relacionais ou de gráficos e a maioria deles oferece uma linguagem de consulta que permite que o objeto seja encontrado com uma abordagem de programação declarativa.

Banco de Dados como um Serviço (DaaS)

Um banco de dados hospedado na nuvem e vendido em uma base medida. Exemplos incluem o Heroku Postgres e o Amazon Relational Database Service.

Bancos de Dados XML

Os Bancos de **Dados XML** permitem que os dados sejam armazenados em formato XML. Os bancos de dados XML costumam estar vinculados a bancos de dados orientados a documentos. Os dados armazenados em um banco de dados XML podem ser consultados, exportados e serializados em qualquer formato necessário.

Biometria

A biometria implica o uso de análise e tecnologia na identificação de pessoas por uma ou mais das suas características físicas, como reconhecimento de impressões digitais, reconhecimento facial, reconhecimento de íris, etc. É mais comumente usado em smartphones modernos.

Business Intelligence (BI)

O termo geral usado para a identificação, extração e análise de dados em benefício dos negócios empresariais.

Glossário

Cache Distribuído

Um **cache de dados** que é distribuído em vários sistemas, mas funciona como um. É usado para melhorar o desempenho.

Cascading

Cascading fornece um nível mais alto de abstração para o Hadoop, permitindo que os desenvolvedores criem tarefas complexas de maneira rápida, fácil e em vários idiomas diferentes executados na JVM, incluindo Ruby, Scala e muito mais. Na verdade, isso destruiu a barreira das habilidades, permitindo que o Twitter usasse o Hadoop de maneira mais ampla.

Cassandra

O Cassandra é um banco de dados distribuído e de código aberto. Projetado para lidar com grandes quantidades de dados distribuídos em servidores comuns, fornecendo um serviço altamente disponível. É uma solução NoSQL que foi inicialmente desenvolvida pelo Facebook. Está estruturado na forma de valor-chave.

CDRs de análise de registro detalhado da chamada

(CDR) contêm dados que uma empresa de telecomunicações coleta sobre chamadas telefônicas, como tempo e duração da chamada. Esses dados podem ser usados em qualquer número de aplicativos analíticos.

Chukwa

Chukwa é um subprojeto do Hadoop dedicado à coleta e análise de registros em larga escala. O Chukwa é construído sobre a estrutura do Hadoop distributed filesystem (**HDFS**) e **MapReduce** e herda a escalabilidade e robustez do Hadoop. O Chukwa também inclui um kit de ferramentas flexível e poderosa para exibir os resultados de monitoramento e análise, a fim de fazer o melhor uso desses dados coletados.

Cientista de Big Data

Alguém que é capaz de desenvolver os algoritmos através de linguagens de programação adequadas, para entender grandes volumes de dados.

Cientista de Dados

Um praticante de ciência de dados.

Ciência de dados

Um termo recente que tem várias definições, mas geralmente aceito como uma disciplina que incorpora estatísticas, visualização de dados, programação de computadores, mineração de dados, Inteligência Artificial, aprendizado de máquina e engenharia de banco de dados para resolver problemas complexos.

Clojure

Clojure é uma linguagem de programação dinâmica baseada no LISP que usa a Java Virtual Machine (JVM). É bem adequada para processamento de dados paralelos.

Coleta de Dados

Qualquer processo que captura qualquer tipo de dados.

Glossário

Comparadores

Duas maneiras pelas quais você pode comparar suas chaves é implementando a interface ou implementando a interface RawComparator. Na primeira abordagem, você irá comparar objetos (desserializados), mas na última abordagem, você comparará as chaves usando seus bytes brutos correspondentes.

Computação em nuvem

Um sistema de computação distribuída em uma rede usada para armazenar dados fora do local.

Confabulação

O ato de tomar uma decisão baseada na intuição, mas que pareça ser baseada em Análise de dados.

Conjunto de Dados

Uma coleção de dados, geralmente em forma de tabela.

Consulta

Solicitação de informações para responder a uma determinada pergunta.

Consulta Paralela

Uma consulta executada em vários encadeamentos do sistema para desempenho mais rápido.

Dados Climáticos

Os dados climáticos em tempo real agora estão amplamente disponíveis para as organizações usarem de várias maneiras. Por exemplo, uma empresa de logística pode monitorar as condições climáticas locais para otimizar o transporte de mercadorias. Uma empresa de serviços públicos pode ajustar a distribuição de energia em tempo real.

Dados Demográficos

Dados relativos às características de uma população humana.

Dados Estruturados

Dados organizados por uma estrutura predeterminada.

Dados Externos

Dados que existem fora de um sistema fechado ou proprietário.

Dados Frios

Armazenamento de dados antigos que dificilmente são usados em servidores de baixa potência computacional. Recuperar os dados levará mais tempo.

Dados Gerados por Máquina

Quaisquer dados criados automaticamente a partir de um processo de computador, aplicativo ou outra fonte não humana.

Dados de Localização

Dados que descrevem uma localização geográfica – coordenadas geo.

Dados Machine2Machine

Dados gerados por duas ou mais máquinas que estão se comunicando entre si de forma independente.

Dados Não-estruturados

São os dados que não tem nenhuma estrutura identificável - por exemplo, o texto das mensagens de e-mail.

Dados de Pesquisa

Dados agregados sobre termos de pesquisa usados ao longo do tempo.

Glossário

Dados Públicos

Dados públicos ou Informações públicas ou conjuntos de dados criados com financiamento público

Dados Semiestruturados

Dados que não são estruturados por um modelo de dados formal, mas fornecem outros meios para descrever os dados e hierarquias.

Dados de Telefones

celulares - Os telefones celulares geram uma enorme quantidade de dados, e grande parte deles está disponível para uso com aplicativos analíticos.

Dados em Tempo Real

Dados criados, processados, armazenados, analisados e visualizados em milissegundos, nas redes ou cache de processamento, etc.

Dados Transacionais

Dados que mudam de forma imprevisível. Exemplos incluem dados de contas a pagar e receber, ou dados sobre remessas de produtos, etc.

Dados Vinculados

Conforme descrito pelo inventor da World Wide Web, Tim Berners-Lee, "Coleta seletiva de atributos ou linguagens comuns para identificar conexões ou relacionamentos entre fontes diferentes de dados".

Data Center

Uma instalação física que abriga um grande número de servidores e dispositivos de armazenamento de dados. Os datacenters podem pertencer a uma única organização ou vender seus serviços a muitas organizações.

Data Mart

É a camada de acesso de um data warehouse usada para fornecer dados aos usuários.

Data de Referência

Dados que descrevem um objeto e suas propriedades. O objeto pode ser físico ou virtual.

Data Warehouse

Um local para armazenar dados para fins de produzir relatórios e análises.

De-identificação

O ato de remover todos os dados que vinculam uma pessoa a uma determinada informação.

Deep Thunder

O serviço de previsão de tempo da IBM que fornece dados meteorológicos para organizações como utilitários, que usam os dados para otimizar a distribuição de energia.

Drill

Um sistema distribuído de código aberto para executar análises interativas em conjuntos de dados de grande escala. É semelhante ao Dremel do Google e é gerenciado pelo Apache.

Glossário

Elasticsearch

Um mecanismo de pesquisa de código aberto criado no Apache Lucene.

Escalabilidade

A capacidade de um sistema ou processo de manter níveis de desempenho aceitáveis à medida que a carga de trabalho ou o escopo aumentam.

Esquema

A estrutura que define a organização dos dados em um sistema de banco de dados.

Estrutura de dados

Uma maneira específica de armazenar e organizar dados.

Exabyte

Um milhão de terabytes, ou 1 bilhão de gigabytes de informação.

Exaustão de Dados

Os dados que uma pessoa cria como um subproduto de uma atividade comum - por exemplo, um registro de chamada de celular ou um histórico de pesquisa da web.

Extrair, transformar e carregar (ETL)

Um processo usado no data warehousing para preparar dados para uso em relatórios ou análises.

Failover

Quando a comutação automática para outro computador ou nó falhar.

Feed de Dados

Um meio para uma pessoa receber um fluxo de dados. Exemplos de mecanismos de feed de dados incluem RSS ou Twitter.

Flume

O Flume é uma estrutura para preencher o Hadoop com dados. Os agentes são preenchidos em toda a infraestrutura de TI - em servidores da Web, servidores de aplicativos e dispositivos móveis, por exemplo - para coletar dados e integrá-los ao Hadoop.

Fonte de dados

Qualquer provedor de dados - por exemplo, um banco de dados ou um fluxo de dados.

Gerenciamento de Documentos

É a prática de rastrear e armazenar documentos eletrônicos e imagens digitalizadas de documentos em papel.

Gerenciamento de Dados

Um conjunto de processos ou regras que garantem a integridade dos dados e que as práticas recomendadas de gerenciamento de dados são atendidas.

Governança de Dados

Atividade chave organizacional na motivação das pessoas em atenderem a classificação de dados e as regras do gerenciamento de dados estabelecidas no tratamento, armazenamento e acesso. A Governança de Dados, governa as pessoas nos temas ligados a gestão de dados empresariais, sendo parte da Governança de TIC.

Glossário

Grid Computing

Garante o desempenho de funções computacionais usando recursos de múltiplos sistemas distribuídos. A computação em grade geralmente envolve arquivos grandes e é usada com mais frequência em vários aplicativos. Os sistemas que compõem uma rede de computação em grade não precisam ser semelhantes em design ou na mesma localização geográfica.

Guardião de dados

Uma pessoa responsável pela estrutura do banco de dados e pelo ambiente técnico, incluindo o armazenamento de dados.

Hadoop

Um projeto de biblioteca de software de código aberto administrado pela Apache Software Foundation. O Apache define o Hadoop como "uma estrutura que permite o processamento distribuído de grandes conjuntos de dados em clusters de computadores usando um modelo de programação simples" – vem sendo usado para processar Big Dark Data..

Hama

Hama é uma estrutura de computação distribuída baseada em técnicas de computação paralela em massa para cálculos científicos maciços, como matriz, algoritmos gráficos e de rede. É um projeto de alto nível sob a Apache Software Foundation.

HANA

Uma plataforma de computação in-memory de software / hardware da SAP projetada para transações de alto volume e análises em tempo real.

HBase

HBase é um banco de dados não relacional que permite pesquisas rápidas de baixa latência no Hadoop. Ele adiciona recursos transacionais ao Hadoop, permitindo que os usuários conduzam atualizações, inserções e exclusões. EBay e Facebook usam o HBase pesadamente.

HCatalog

HCatalog é um serviço centralizado de gerenciamento e compartilhamento de metadados para o Apache Hadoop. Ele permite uma visão unificada de todos os dados nos clusters do Hadoop e permite que diversas ferramentas, incluindo Pig e Hive, processem qualquer elemento de dados sem precisar saber fisicamente onde no cluster os dados estão armazenados.

HDFS (sistema de arquivos distribuídos Hadoop)

HDFS (Hadoop Distributed File System) a camada de armazenamento do Hadoop, é um sistema de arquivos distribuído, escalável e baseado em Java, adepto de armazenar grandes volumes de dados não estruturados (Dark).

Hive

Hive é um framework semelhante ao Hadoop, baseado em data warehousing, originalmente desenvolvido pelo Facebook. Ele permite que os usuários escrevam consultas em uma linguagem semelhante a SQL chamada HiveQL, que são então convertidas em **MapReduce**. Isso permite que os programadores SQL sem experiência MapReduce para usar o armazém e torna mais fácil a integração com ferramentas de inteligência de negócios e de visualização, como Microstrategy, Tableau, Revolutions Analytics, etc.

Glossário

Hue

O Hue (Hadoop User Experience) é uma interface baseada na Web de código aberto para facilitar o uso do Apache Hadoop. Ele possui um navegador de arquivos para HDFS, um aplicativo Oozie para criação de fluxos de trabalho e coordenadores, um designer / navegador de tarefas para o MapReduce, uma interface do usuário Hive e Impala, um Shell, uma coleção de APIs do Hadoop e muito mais.

Impala

Impala (By Cloudera) fornece consultas SQL rápidas e interativas diretamente nos dados do Apache Hadoop armazenados em HDFS ou HBase usando os mesmos metadados, sintaxe SQL (Hive SQL), driver ODBC e interface do usuário (Hue Beeswax) como Apache Hive. Isso fornece uma plataforma familiar e unificada para consultas em tempo real ou orientadas por lote.

Integração de dados

O processo de combinar dados de diferentes fontes e apresentá-los em uma única exibição.

Integridade de dados

É a medida de confiança que uma organização tem na exatidão, integridade, pontualidade e validade dos dados.

Inteligência artificial

Processo para desenvolver máquinas e softwares de inteligência capazes de perceber o ambiente e tomar ações correspondentes quando necessário e até aprender com essas ações.

Identificação e captura automática (AIDC)

Qualquer método de identificar e coletar automaticamente dados em itens e, em seguida, armazenar os dados em um sistema de computador. Por exemplo, um scanner pode coletar dados sobre um produto enviado por meio de um chip RFID.

Internet das Coisas

Dispositivos comuns conectados à Internet ou a redes a qualquer momento, em qualquer lugar, agregado de sensores e interface de comunicação de dados.

Invocação de Método Paralelo (PMI)

Permite que o código de programação chame múltiplas funções em paralelo.

Kafka

Kafka (desenvolvido pelo LinkedIn) é um sistema de mensagens de publicação distribuída que oferece uma solução capaz de lidar com toda a atividade de fluxo de dados e processar esses dados. Ex.: site do consumidor. Esse tipo de dados (exibições de páginas, pesquisas e outras ações do usuário) é um ingrediente-chave na Web social atual.

Glossário

Latência

É qualquer atraso em uma resposta ou entrega de dados de um ponto para outro.

Limpeza de Dados

É o ato de revisar e revisar dados para remover entradas duplicadas, corrigir erros ortográficos, adicionar dados ausentes e fornecer mais consistência.

Linguagem de Consulta Estruturada (SQL)

Uma linguagem de programação projetada especificamente para gerenciar e recuperar dados de um sistema de banco de dados relacional.

Mahout

O Mahout é uma biblioteca de mineração de dados. Ele usa os algoritmos de mineração de dados mais populares para realizar clustering, testes de regressão e modelagem estatística e os implementa usando o modelo Map Reduce.

MapReduce

MapReduce é uma estrutura de software que serve como a camada de computação do Hadoop. As tarefas do MapReduce são divididas em duas partes (obviamente nomeadas). A função "Mapear" divide uma consulta em várias partes e processa os dados no nível do nó. A função "Reduzir" agrega os resultados da função "Mapear" para determinar a "resposta" à consulta.

Mashup

O processo de combinar diferentes conjuntos de dados em um único aplicativo para aprimorar a saída, por exemplo, combinando dados demográficos com listagens de imóveis.

Mecanismo de recomendação

É um algoritmo que analisa as compras e ações de um cliente em um site de comércio eletrônico e, em seguida, usa esses dados para recomendar produtos complementares.

Metadados

Dados sobre dados; fornece informações sobre o que são os dados.

Migração de dados

O processo de mover dados entre diferentes tipos ou formatos de armazenamento ou entre diferentes sistemas de computador.

Mineração de Dados

O processo de prospecção de dados, com base em padrões ou conhecimentos, sobre grandes conjuntos de dados, para explorar, descobrir regras, identificar fatores e tendências-chave e relacionamentos ocultos.

Modelo de Dados, modelagem de dados

Um modelo de dados define a estrutura dos dados para fins de comunicação entre pessoas funcionais e técnicas para mostrar dados necessários para processos de negócios ou para comunicar um plano para desenvolver como os dados são armazenados e acessados entre os membros da equipe de desenvolvimento de aplicativos.

Modelagem Preditiva

O processo de desenvolvimento de um modelo que provavelmente irá prever uma tendência ou resultado.

Glossário

MongoDB

O MongoDB é um banco de dados NoSQL orientado a documentos, desenvolvido sob o conceito de código aberto. Ele salva estruturas de dados em documentos JSON com um esquema dinâmico (chamado formato MongoDB BSON), tornando a integração dos dados em certos aplicativos mais fácil e ágil.

NewSQL

É um sistema de banco de dados elegante e bem definido, mais fácil de aprender e melhor que o SQL. É ainda mais recente que o NoSQL

NoSQL

O NoSQL (comumente interpretado como "não apenas SQL") é uma ampla classe de sistemas de gerenciamento de banco de dados identificados pela não-aderência ao modelo de sistema de gerenciamento de banco de dados relacional amplamente utilizado. Bancos de dados NoSQL não são construídos principalmente em tabelas e geralmente não usam SQL para manipulação de dados.

Nuvem

Um termo amplo que se refere a qualquer aplicativo ou serviço baseado na grande rede Internet hospedado remotamente.

Objeto Distribuído

É um módulo de software projetado para trabalhar com outros objetos distribuídos armazenados em outros computadores.

OpenDremel

A versão open source do código java Big Query do Google. Está sendo integrado ao Apache Drill.

Open Data Center Alliance (ODCA)

Um consórcio de organizações globais de TI cujo objetivo é acelerar a migração da computação em nuvem.

Oozie

Oozie é um sistema de processamento de fluxo de trabalho que permite aos usuários definir uma série de trabalhos escritos em vários idiomas - como Map Reduce, Pig e Hive - e, em seguida, vinculá-los de forma inteligente uns aos outros. O Oozie permite que os usuários especifiquem, por exemplo, que uma determinada consulta só deve ser iniciada após a conclusão de tarefas anteriores especificadas, nas quais ela confia que os dados foram concluídos.

Glossário

Painel

Uma representação gráfica das análises realizadas por algoritmos.

Pentaho

A Pentaho oferece um conjunto de produtos BI (Business Intelligence) de código aberto chamado Pentaho Business Analytics, fornecendo integração de dados, serviços OLAP, relatórios, painéis, mineração de dados e capacidades de ETL.

Perfil de dados

É o processo de coleta de estatísticas e informações sobre dados em uma fonte existente.

Ponto de dados

Um item individual em um gráfico ou matriz.

Porco

O Pig Latin é uma linguagem baseada no Hadoop desenvolvida pelo Yahoo. É relativamente fácil de aprender e é adepto de pipelines de dados muito longos e muito profundos (uma limitação do SQL).

Processamento Analítico on-line (OLAP)

O processo de análise de dados multidimensionais usando três operações: consolidação (a agregação de dados disponíveis), detalhamento (a capacidade de os usuários verem os detalhes subjacentes) e fatia e dados (a capacidade de os usuários selecionarem subconjuntos e visualizá-los de diferentes perspectivas).

Processamento Distribuído

A execução de um processo em vários computadores conectados por uma rede de computadores.

Processamento de Eventos Complexos (CEP)

CEP é o processo de monitorar e analisar todos os eventos nos sistemas de uma organização e agir sobre eles quando necessário, em tempo real.

Processamento Paralelo

A capacidade de executar várias tarefas ao mesmo tempo.

Processamento Transacional on-line (OLTP)

O processo de fornecer aos usuários acesso a grandes quantidades de dados transacionais de forma que eles possam derivar significado deles.

Petabyte

Um milhão de gigabytes ou 1.024 terabytes.

Qualidade de Dados

Qualidade dos dados e a "avaliação dos dados" para determinar seu valor para a tomada de decisões, planejamento ou operações.

Glossário

R

"R" é uma linguagem e ambiente para computação estatística e gráfica. É um projeto GNU semelhante à linguagem "S".

O "R" fornece uma ampla variedade de estatística (modelagem linear e não linear, testes estatísticos clássicos, análise de séries temporais, classificação, agrupamento, ...) e técnicas gráficas, e é altamente extensível.

Reconhecimento de padrões

É a classificação ou rotulagem de um padrão identificado no processo de aprendizado de máquina.

Re-identificação

Combinação de vários conjuntos de dados para localizar uma determinada pessoa em dados anônimos Dados.

Replicação de dados

O processo de compartilhar informações para garantir a consistência entre fontes redundantes.

Repositório de dados

É a localização dos dados armazenados permanentemente.

Segurança de dados

É a prática de proteger dados contra adulteração, destruição ou acesso não autorizado.

Servidor

Um computador físico ou virtual que atende e processa as solicitações de um aplicativo de software e entrega essas solicitações em uma rede.

Sistema de Arquivos Distribuídos

Sistemas que oferecem acesso simplificado e altamente disponível ao armazenamento, análise e processamento de dados.

Sistema de Gerenciamento de Banco de Dados (DBMS)

Software que coleta e fornece acesso a dados em um formato estruturado.

Software como serviço (SaaS)

Software de aplicativo usado pela web por um thin client ou navegador da web. O Salesforce é um exemplo bem conhecido de SaaS.

SQL

É uma linguagem de programação para recuperar dados de um banco de dados relacional.

Sqoop

Sqoop é uma ferramenta de conectividade para mover dados de armazenamentos de dados não-Hadoop - como bancos de dados relacionais e data warehouses - para o Hadoop. Ele permite que os usuários especifiquem o local de destino dentro do Hadoop e instruem o Sqoop a mover dados do Oracle, do Teradata ou de outros bancos de dados relacionais para o destino.

Storm

Storm é um sistema de computação distribuída em tempo real, de código aberto e gratuito, nascido no Twitter. O Storm facilita o processamento confiável de fluxos de dados não estruturados no campo de processamento em tempo real, o que tornou o Hadoop para processamento em lote.

Glossário

Thrift

"O Thrift é uma estrutura de software para o desenvolvimento escalonável de serviços entre linguagens. Ele combina uma pilha de software com um motor de geração de código para construir serviços que funcionam de forma eficiente e sem problemas entre C ++, Java, Python, PHP, Ruby, Erlang, Perl, Haskell, C #, Cacau, Smalltalk, e OCaml."

Tomada de decisão direcionada por Dados

É o uso de dados analisados para apoiar a tomada de decisões cruciais.

Valor dos Dados

Todos os dados disponíveis criarão muito valor para organizações, sociedades e consumidores. Big Data selecionado e tratado contem informações comerciais valiosas que podem ser comercializadas.

Valor-chave de Armazenamento (Key Value Stores)

Os principais armazenamentos de valores permitem que o aplicativo armazene seus dados de maneira sem esquema. Os dados podem ser armazenados em um tipo de dados de uma linguagem de programação ou um objeto. Por isso, não há necessidade de um modelo de dados fixo.

Visualização

Uma abstração visual de dados projetados com a finalidade de obter significado ou comunicar informações de maneira mais eficaz.

Visualização de dados

Similar a Visualização, porém com seus detalhes.

Volume

A quantidade de dados armazenados ou em transito, variando de megabytes a brontobytes.

WebHDFS Apache Hadoop

WebHDFS O Apache Hadoop fornece bibliotecas nativas para acessar o HDFS. No entanto, os usuários preferem usar o HDFS remotamente nas bibliotecas nativas do lado do cliente. Por exemplo, alguns aplicativos precisam carregar dados dentro e fora do cluster ou interagir externamente com os dados do HDFS. O WebHDFS resolve esses problemas fornecendo uma API REST HTTP totalmente funcional para acessar o HDFS.

Zetabyte

Zettabyte (ZB) É uma medida de armazenamento que corresponde a 2^{70} bytes. Equivale a 1.024 Exabytes, ou 1.048.576 Petabytes, ou ainda a 1.073.741.800 Terabytes – e para ser exato, 1,180,591,620,717,411,303,424 bytes – um sextilhão de bytes - unidade seguida de 21 zeros 10^{21} .

ZooKeeper

ZooKeeper é um projeto de software da Apache Software Foundation, um serviço que fornece configuração centralizada e registro de nome de código aberto para grandes sistemas distribuídos. O ZooKeeper é um subprojeto do Hadoop.zz

...

Sobre a Participação no IT FORUM X 2018



IT Forum X

A história do amanhã!

17/18
OUT-2018

Transamerica
Expo Center
SP



Sophia

Dr. Hanson

PALESTRA - Dark Analytics - prospecção de insights estratégicos

Fundamentação: Os dados digitais globais armazenados dobram de quantidade a cada ano e devem ultrapassar 44 trilhões de gigabytes, em 2020. Todos utilizam, produzem e guardam dados, desde um usuário da Internet, uma pequena empresa ou uma grande corporação. Boa parte do retido pode ser considerado “lixo digital”, pois são dados desestruturados e praticamente não utilizados ou correlacionados. Esse grande volume de “Lixo Digital”(arquivos de imagem, áudio, vídeo e dados inexplorados da grande “deep web”,etc.) podem ser denominados de “Dark Data” ou dados obscuros e ou não tratados. A partir do aparente Lixo Digital surge uma oportunidade imperdível o “Dark Analytics”, atual e irreversível tendência na busca de novos insights para negócios e ampliação do market share, que já agita seguimentos estratégicos como o de “Educação”, na formação de novos profissionais “Cientistas de Dados-DS” e o da “Indústria de Software” na geração de novas ferramentas de ECM, Big Data Analytics, AI-Analytics, linguagens especializadas, etc. Venha entender as convergências, os campos de aplicação, o potencial das tecnologias, as vantagens, impactos negativos e riscos do Dark Analytics.

João Roberto Peres



Dark Analytics é tendência irreversível na tecnologia

Por [Jorge Junior](#) - 17/10/2018



90% dos dados digitais armazenados podem ser considerados lixo digital e têm potencial de mercado, afirma o especialista João Roberto Peres.

Saber detectar, analisar e utilizar o grande volume de “lixo digital” armazenado é uma tendência irreversível para empresas. Essa é a opinião de João Roberto Peres, que proferiu a palestra Dark Analytics – Prospecção de Insights Estratégicos, hoje, durante o IT Forum Expo.

Continua na próxima página...

Atualmente, os dados digitais globais armazenados dobram de quantidade a cada ano e devem ultrapassar 32 zetabytes (sextilhões de bytes) em 2020 e 163 zetabytes em 2025. Desse total, 90% são o chamado "lixo digital", dados desestruturados ou desperdiçados, que não são utilizados em benefício das organizações, como arquivos de imagem, áudio e vídeo, e dados inexplorados da Deep Web (Internet Profunda).

Esses dados obscuros, ou não tratados, são chamados de Dark Data. Daí surge o Dark Analytics, que é a análise desses dados, produzidos tanto por usuários simples da internet como por grandes corporações. "O uso de Dark Data passa a ser estratégico para as organizações de todos os portes e ramos de atividades, pois pode oferecer inúmeras oportunidades de insights para novos negócios e potencialização do market share, além de permitir ampliar o nível de segurança e reduzir custos para empresas", afirmou Peres.

Potencial infinito

Para o especialista, os campos de aplicação do Dark Analytics têm potencial infinito. "Projeções sobre o mercado global das análises escuras apontam movimentação de quase US\$ 1 trilhão. É uma grande oportunidade de negócios em diversas áreas, como educação, desenvolvimento de softwares, serviços de consultoria, criptografia e backup."

Divulgação do Tema – Dark Analytics – antes da palestra no "IT Fórum X"

Principais Links (URLs):

- Dark Analytics – IT FÓRUM –
< <https://www.itforumexpo.com.br/palestrante/joao/> >
- Dark data, robótica na gestão e cultura hacker são temas do IT Fórum Expo – Terra Notícias –
< <https://www.terra.com.br/noticias/dino/dark-data-robotica-na-gestao-e-cultura-hacker-sao-temas-do-it-forum-expo,3fbbe6a564fea395aaaa145ae408ba3al30hzgv9.html> >
- Como a ciência de dados com dark analytics pode transformar mercados é tema no IT FORUM X 2018 – Portal SEGS -
< <https://www.segs.com.br/info-ti/120713-como-a-ciencia-de-dados-com-darkanalytics-pode-transformar-mercados-e-tema-no-it-forum-x-2018> >
- IT FORUM X 2018 - Prof. João Roberto Peres da Komp Consultoria de Gestão e Segurança/FGV – YouTube –
< <https://www.youtube.com/watch?v=OgZPdGHiN3g> >

Acessos disponíveis em 20/10/2018.



Corporate Friends Connection



Friends



Homenagem

Paulo Roberto Vampré Hummel

Mestre, Conselheiro e Amigo!



Para conhecermos os amigos é necessário passar juntos pelo sucesso e pela adversidade.

No sucesso, verificamos a quantidade e, na adversidade, a qualidade.

Confúcio

Paulo é nota máxima!

Paulo R. V. Hummel



PAULO ROBERTO VAMPRE HUMMEL

Graduado em Engenharia Mecânica pelo Instituto Tecnológico da Aeronáutica (ITA), em 1966. Mestre em Administração de Empresas pela Escola de Administração de Empresas de São Paulo (EAESP/FGV), onde foi professor do Departamento de Produção e Operações Industriais desde 1971.

Durante os últimos 40 anos exerceu diversas atividades em cargos executivos e como consultor em diversas empresas privadas e estatais.

Especialização: Modelagem e Avaliação de Empresas, Racionalização e Otimização de Empresas, Planejamento Estratégico, Gerenciamento de Projetos e Planejamento de Mudança. Vice Presidente da **GVconsult** desde 1996/2002 (hoje FGV Projetos). Diretor de Consultoria em projetos de reorganização e em sistemas financeiros / operacionais / administrativos, em empresas nacionais, multinacionais e estatais como Advocacia Geral da União, Anatel / CPqD e Secretaria de Minas e Metalurgia. Executou modelagem e Avaliação de Empresas para Privatização no Governo do Estado de Pernambuco. Vice-Presidente de empresa de informática por 8 anos, responsável por Marketing e Finanças.

Presidente da Darumatec (empresa da área de Telecomunicações), reestruturando estrategicamente a empresa. Foi Diretor Administrativo e Financeiro em empresa nacional de médio porte, na área de telecomunicações, tendo efetuado seu saneamento financeiro e sua consolidação administrativa em 12 meses. Vivência de 9 anos como primeiro executivo financeiro em empresa industrial multinacional.

Resumindo

40 anos de atividades profissionais como: Diretor de Consultoria em projetos de reorganização e em sistemas financeiros / operacionais / administrativos, em empresas nacionais, multinacionais e estatais.

Autor dos Livros:

- "Análise e Decisão sobre Investimentos e Financiamentos" (Engenharia Econômica - Teoria e Prática), Editora Atlas, São Paulo, 1995 - 4ª Edição.
- "Matemática Financeira e Engenharia Econômica: A Teoria e a Prática da Análise de Projetos de Investimentos. Editora Pioneira – Thompson Learning – 2008.



Paulo hoje é nosso parceiro e mentor, oferecendo "Insights Estratégicos", promovendo subsídios para novas oportunidades e diretivas de trabalhos avançados. Evoluímos com ele !!!

Atividades Conjuntas

Paulo Hummel e eu, João Peres, trabalhamos juntos na FGV Projetos por mais de 10 anos, onde como Diretor de Projetos, Paulo sempre foi um líder exemplar, respeitando e valorizando o conhecimento dos seus coordenados. Desenvolvemos diversos projetos para empresas públicas e privadas de médio a grande porte, bem como, internamente desenvolvemos em P&D diversos padrões metodológicos, convertidos em produtos de Consultoria - entre eles destaque, conforme evidência a seguir:

Segue reportagem do "Jornal do Comercio do Rio Grande do Sul":

O jornal de economia e negócios do RS

Jornal do Comércio

GESTÃO - Notícia da edição impressa de 22/10/2012

Incidência de casos de fraudes preocupa empresas – por Rafael Vigna

Escândalos como o do mensalão reforçam a necessidade de implementação de controles mais rígidos.

A segunda fase do julgamento do mensalão, dedicada à análise do crime de gestão fraudulenta por parte dos diretores do Banco Rural, e a retomada da ação no Ministério Público contra executivos do banco Panamericano trazem à tona um risco constante às empresas que atuam no País. Corrupção, suborno, desvios de ativos e conflitos de interesses, além de estarem entre as principais preocupações das corporações, têm em comum o fato de serem problemas diretamente relacionados ao comportamento das pessoas nas organizações.

De acordo com uma pesquisa da KPMG, realizada em faixas de faturamento entre R\$ 250 milhões e R\$ 3 bilhões, quase 70% dos entrevistados revelaram ter sofrido danos financeiros nos últimos dois anos. Entre as organizações consultadas, 77% tiveram prejuízos inferiores a R\$ 1 milhão e 5% perderam mais de R\$ 10 milhões.

Apesar do cenário, de acordo com o diretor da Squadra - empresa que atua em consultoria para as áreas de inteligência e segurança -, Leandro Longhi, a maioria dos casos ainda permanecem não revelados, por critérios de proteção às marcas no mercado. Entretanto, a opção de omitir os fatos, "para prevenir eventuais escândalos" pode prejudicar o processo de produção de provas para o ingresso de ações judiciais contra os envolvidos.

Atividades Conjuntas

“O nível de preocupação no Brasil costuma ser muito baixo. A maioria tem carência de controles internos e auditoria. Muitas desconhecem os riscos. Nas pequenas, há a vantagem de uma maior proximidade com os donos. Mas nas médias, as ações de prevenção são praticamente nulas. É uma falta de cultura das empresas não possuir uma área específica de gestão de risco”, analisa o especialista. O primeiro passo para ampliar o grau de controle, segundo ele, é diagnosticar as fragilidades, as ameaças e vulnerabilidades. Com isso, comenta Longhi, é possível compor um indicador de risco, adaptável a cada segmento e área de atuação dos negócios.

Atentos ao problema, os coordenadores de Projetos da Fundação Getulio Vargas (FGV) **João Roberto Peres e Paulo Hummel** lançaram um programa de prevenção de fraudes corporativas. O objetivo, explica Hummel, é demonstrar que a união de processos pode viabilizar uma mudança da cultura organizacional e antecipar as características mais comuns das fraudes. Segundo ele, há uma tendência em focar as ações apenas após a ocorrência.

Ele destaca algumas falhas que as empresas cometem antes mesmo de contratar seus funcionários, como a falta de atenção às referências. Outra vulnerabilidade muito comum são os tradicionais “presentes” recebidos por fornecedores ou clientes, que podem gerar conflitos de interesses. “Se a empresa não definir seus limites, o profissional não saberá como agir corretamente. A falta de uma política pode abrir uma brecha para a corrupção e fraudes”, aponta Hummel.

Outro aspecto que permeia a incidência dos casos está diretamente ligado ao fator humano. A análise de alguns casos, assegura Peres, identificou a localização de desvios comportamentais em um determinado perfil de profissionais. Por isso, o **programa de prevenção de risco** deve envolver, além de áreas como **auditoria e compliance**, o departamento de recursos humanos, um dos principais setores na metodologia da FGV.

O próprio mercado tem cobrado maior comprometimento das empresas. Exemplo disso são as ações da BM&FBovespa e da CVM, que passaram, a partir de uma norma lançada em maio, a exigir das companhias de capital aberto a elaboração de códigos de conduta, que, diferentemente dos códigos de ética, possuem ingerência sobre os processos.



Atividades Conjuntas

Metodologia amplia envolvimento de setores para evitar desvios

A metodologia da FGV Projetos prevê 12 políticas, 12 controles e 8 programas complementares de prevenção de fraudes. Um exemplo é a criação de um **programa de delação segura**. Embora ainda raro na maioria das companhias, o sistema pressupõe a existência de um ouvidor ou de um escritório independente e transparente para receber denúncias de funcionários, sem que haja riscos ou prejuízos pessoais com a identificação dos prejudicados.

Com a prevenção, diz um dos coordenadores de Projetos da Fundação Getúlio Vargas (FGV) João Roberto Peres, existe maior ênfase nos **códigos de conduta** e é possível aumentar a incidência de atos de menor magnitude, que não aparecem nas auditorias contábeis.

Por isso, o programa da FGV contempla aspectos como RH seguro, código de conduta, política de segurança, antifraudes, antilavagem, auditoria interna focada em fraudes, controles extras focados na prevenção de fraudes, contratação de agentes departamentais, **processos forenses corporativos**, instrumentos de delação segura e programas de **valorização do trabalho**.

Brechas constitucionais dificultam a punição dos envolvidos

Diferentes tipificações quanto aos crimes financeiros e de ordem tributária dificultam a produção de provas contra os envolvidos em fraudes corporativas. Em dois dos casos mais comuns, conforme explica o advogado criminal e professor de direito penal da Pucrs, Marcelo Bertolucci, as punições são de média intensidade. A lei específica do controle dos crimes contra o sistema financeiro prevê reclusão de 2 a 5 anos. Já na lavagem de dinheiro, de 3 a 8 anos.

Mesmo assim, o jurista considera a legislação brasileira suficiente. “O problema não está na nossa lei. As dificuldades estão em uma série de questões que envolvem a necessidade de averiguação e residem na falta de provas de casos de fraude. Isso decorre de consequências como a postura da empresa, que não se esforça na produção das provas”, analisa.

Obs.: Este artigo foi publicado em 22/10/2012, abordando temas como fraudes, corrupção, riscos, compliance, etc. – no entanto, a Lei Anticorrupção brasileira LEI Nº 12.846, só foi sancionada em 1º de agosto de 2013. Estávamos e continuamos “hoje”, a frente das necessidades atuais (KOMP – FGV)!

Atividades Conjuntas

O coordenador da FGV Projetos, João Roberto Peres, concorda com uma parcela da análise. Ele afirma que cabe às empresas assumirem as eventuais brechas constitucionais. "Pelo fato de não termos uma Justiça atuante, é preciso definir políticas claras de prevenção, que substituam a sensação de impunidade. Só assim as pessoas pensarão duas vezes antes de cometer o ilícito", defende.

↳ Perfil dos fraudadores

■ Homens	78%
■ Faixa etária entre 26 e 40 anos	65%
■ Salários entre R\$ 1 mil e R\$ 3 mil	51%
■ Com média de 2 a 5 anos de organização	41%
■ Presidência	21%
■ Chefia	21%
■ Gerência	22%
■ Staff	53%

↳ Causas apontadas para o aumento dos casos

■ Perdas de valores morais	63%
■ Insuficiência de sistemas de controles	56%
■ Impunidade	53%
■ Problemas econômicos	11%

↳ Ações tomadas após a descoberta do ilícito

■ Demissão dos envolvidos	55%
■ Denúncia Criminal	31%
■ Investigação por auditoria interna	33%
■ Revisão da eficácia dos procedimentos antifraude	30%

FONTE: PESQUISA KPMG - A FRAUDE NO BRASIL 2010

AURACEBIO PEREIRA/ARTE/JC

NELSON JR./SCO/STF/DIVULGAÇÃO/JC

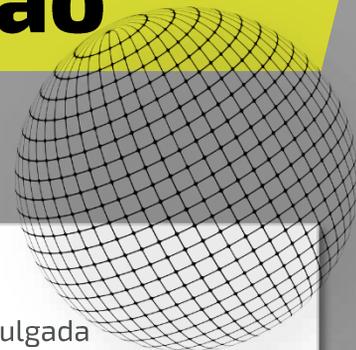
Acesso Jornal do Comércio RS < <https://www.jornaldocomercio.com/site/noticia.php?codn=106644> >

Disponível em 20/10/2018.

Vide também INSTITUTO DE PESQUISA GIANELLI MARTINS (IPGM):

< http://www.ipgm.org.br/web/noticias_det.php?id=1813 > Disponível em 20/10/2018.

Declaração Sobre a Publicação



Este é um e-book gratuito, de uso livre e irrestrito, podendo o conteúdo ser copiado, modificado e ampliado, desde de que seja divulgada a fonte, seu autor e endereço WEB de origem, suas partes mínimas copiadas e as modificações realizadas. O autor possui o direito de reivindicar judicialmente a autoria dos textos e imagens copiados sem a devida referência. A publicação tem ISBN oficial.

A publicação se destina a reforçar os conhecimentos prévios dos leitores interessados nos temas correlatos a Big Data & Analytics – em especial Dark Analytics, bem como, no atendimento das necessidades pessoais e empresariais. Não é uma publicação comercial, nem abrangente, e não esgota o assunto. Não se concorre com a grandes empresas de Consultoria na divulgação de temas. Da mesma forma, usamos linguagem coloquial, não técnica ou jornalística. O conteúdo não deve ser considerado como aconselhamento profissional, pois este, deve ser buscado por leitores para atender suas necessidades específicas. O autor, cocriadores e envolvidos não assumem responsabilidades sobre o uso das informações disponibilizadas.

As definições, afirmações e conceitos, abrangem unicamente a convicção do autor e apoio dos cocriadores, não representando a opinião integral dos apoiadores, patrocinadores e divulgadores.

Toda a produção gráfica e textual do conteúdo é artesanal e foram fundamentadas em informações universais de fontes abertas, URLs web, banco de imagens e figuras de domínio público, sendo as imagens tratadas especificamente, onde as referências autorais e direitos de terceiros, sempre foram respeitados e citados. Da mesma forma, a validação textual de plágio por coincidência, foi legitimada e aprovada por sistema automático de alta reputação e aceitação internacional.

Estamos abertos a qualquer colaboração, seja crítica ou construtiva – faça contato: kontato@komp.com.br

Fontes de Imagens e figuras: < <https://pixabay.com/pt/> > ou através do site de figuras livres < <http://br.freepik.com/> > .

Validação: "Full Grammarly's Grammer and Plagiarism Checker": disponível no endereço URL < <https://www.grammarly.com/plagiarism-checker> >." acessos em 20/10/2018.

Publicação ofertada por NewKOMP, apoiada por: NTSC-BR, KOMP SECURITY BRAZIL, Komp Tecnologia, JDPERES, ...

Publicação Registrada no ISBN nº **978-85-923158-3-2**





Contato:

kontato@komp.com.br

Endereço:

Av. Das Nações Unidas, 8341 – 11º
Bloco B – Pinheiros – São Paulo -SP

<http://www.komp.com.br>