

# BIG DATA

Fernando Santos

## ÍNDICE

Definição de Big Data

Os três Vs do Big Data

A história do Big Data

Benefícios do Big Data e da análise de dados

Exemplos de uso do Big Data

Desafios do Big Data

Como funciona o Big Data (principais ações)

Diretrizes para construir uma base de Big Data

### Definição de Big Data

É conjuntos de dados muito grande e complexos que o software de processamento tradicional não consegue gerir e que podem ser usados para resolver problemas que não o poderiam ser de outra forma.

### Os três Vs do Big Data (\*)

**Volume** - refere-se ao grande volume de dados de baixa densidade e não estruturados, para serem processados. Os dados podem ser de valor desconhecido, como fluxos de cliques em uma página da web ou aplicativo móvel, dados provenientes de sensores de um equipamento, etc..

**Velocidade** - traduz a rapidez com que os dados são recebidos e tratados. Quando a velocidade é alta os dados fluem diretamente para a memória e, com velocidades mais baixas, são gravados em disco. Alguns produtos inteligentes habilitados para Internet operam em tempo real ou quase e requerem avaliação e ação igualmente em tempo real.

**Variiedade** - traduz os vários tipos de dados disponíveis. Os dados tradicionais apresentam-se estruturados e integrados num banco de dados relacional mas, com o surgimento do Big Data, apresentam-se não estruturados. Os dados não estruturados e semiestruturados, como texto, áudio e vídeo, requerem pré-processamento adicional para terem significado e originar metadados.

### A história do Big Data

O conceito de big data é relativamente novo, mas a origem de grandes conjuntos de dados para serem tratados remonta às décadas de 1960 - 1970, quando apareceram os primeiros “data center” e se iniciou o desenvolvimento do banco de dados relacional.

Por volta de 2005 o volume de dados gerados pelos serviços online (Facebook, YouTube e outros) conduziu ao desenvolvimento de uma estrutura de código aberto (Hadoop) específica para armazenar e analisar conjuntos de big data; o NoSQL também começou a ganhar popularidade nessa época.

O desenvolvimento de estruturas de código aberto, como hadoop, cassandra e, mais recentemente, o Spark, foi essencial para o crescimento do big data porque torna-o mais fácil de trabalhar e mais barato para armazenar.

Nos anos subsequentes o volume de big data disparou, pois os utilizadores, a IoT, etc., tem gerado grandes quantidades de dados. Inicialmente os dados eram introduzidos manualmente nas bases de dados, sendo depois obtidos do registo da atividade humana (dados gerados por humanos) e, posteriormente, com o aparecimentos das IoT, diretamente do desempenho de dispositivos inanimados (dados gerados por máquinas). (\*)

Com o aparecimento da **Internet das Coisas (IoT)**, mais objetos e dispositivos são conectados à Internet, reunindo dados sobre o desempenho de produtos e de padrões de uso de clientes. O surgimento da aprendizagem pelas máquinas produziu ainda mais dados.

Tendo o big data avançado muito a sua utilidade está apenas começando, pois a computação em nuvem expandiu as suas possibilidades; a nuvem oferece escalabilidade verdadeiramente elástica, onde os utilizadores podem ativar clusters “ad hoc” para testar um subconjunto de dados.

### **Benefícios do Big Data e da análise de dados**

Permitir obter respostas mais completas porque se tem mais informação; dar respostas mais completas significa mais confiança nos dados, o que permite uma abordagem diferente para resolver os problemas.

### Exemplos de uso do Big Data

#### Desenvolvimento de produtos

Utilização do big data para **antecipar as necessidades dos clientes**, construindo modelos preditivos para novos produtos e serviços baseados na classificação dos principais atributos dos produtos ou serviços passados e atuais e modelando a relação entre esses atributos e o sucesso comercial das ofertas.

Utilização de dados e análises de grupos específicos, mídia social, mercados de teste e lançamentos iniciais de lojas para planejar, produzir e lançar novos produtos.

#### Manutenção preditiva

Previsão de possíveis falhas mecânicas que podem estar relacionadas com dados estruturados antigos, como o ano, marca e modelo do equipamento, bem como com dados não estruturados que abrangem milhões de entradas de registro, dados de sensores, mensagens de erro, etc..

Ao analisar as indicações de potenciais problemas antes que estes aconteçam, as organizações podem proceder à manutenção dos equipamentos e maximizar o tempo de disponibilidade dos seus componentes.

### **Aprendizagem da máquina**

A aprendizagem das máquinas já é uma realidade pois é possível ensinar as máquinas em vez de programá-las; a disponibilidade de big data para treinar modelos de aprendizagem de máquina torna isso possível.

### **Experiência do cliente**

A visão mais clara da experiência do cliente está cada vez mais acessível, pois com o big data reúnem-se dados de mídia social, visitas na web, registros de chamadas e outras fontes, o que aumenta a experiência da interação e maximiza-se o valor do produto, permitindo desenvolver ofertas personalizadas, reduzindo-se a rotatividade de clientes e tratando os problemas de maneira proativa.

### **Impulsionamento da inovação**

O big data ao estudar as interdependências entre humanos, instituições, entidades e processos, determina novas formas de usar esses “insights” ajudando a inovar, melhorando as decisões sobre considerações financeiras e de planejamento. Examinam-se tendências, o que os clientes desejam oferecer, novos produtos e serviços e implementam-se preços dinâmicos abrindo-se, assim, infinitas possibilidades.

### Desafios do Big Data

O grande volume de dados é o principal desafio pois, embora as novas tecnologias tenham sido desenvolvidas para armazenar muitos dados, o seu volume duplica de dois em dois anos, pelo que, para serem valiosos devem ser “limpos” e organizados (utilizados) os mais relevantes para permitir uma análise objetiva de cada situação. Os cientistas de dados gastam 50 a 80 % do tempo preparando os dados antes que possam ser usados.

### Como funciona o Big Data (principiais ações)

**Integrar** - Os big data integram dados de muitas fontes e de distintas aplicações, pelo que os mecanismos tradicionais como ETL (extrair, transformar e carregar) não estão à altura da tarefa, sendo necessárias novas estratégias e tecnologias para analisar conjuntos de big data à escala de terabyte ou mesmo petabyte.

Durante a integração é necessário que os dados sejam processados e formatados para estarem disponíveis para que os analistas os possam utilizar.

**Gerir** - Para gerir dados é necessário que eles estejam armazenados na nuvem, localmente ou em ambos. Pode-se armazenar conjuntos de dados no formato que se desejar e utilizar os requisitos e mecanismos de processamento necessários para esses conjuntos.

A escolha da solução de armazenamento deve estar de acordo com o local onde os dados estão armazenados. A nuvem está gradualmente a ganhar popularidade porque oferece suporte aos requisitos de computação atuais e permite que se acione os recursos conforme o necessário.

**Analisar** - O investimento em big data compensa quando se analisa e atua com base em dados, obtendo-se uma maior clareza com a apresentação visual da sua análise. Esta pode ser utilizada para fazer novas descobertas, compartilhando-as com outras pessoas e para criar modelos de dados provenientes da aprendizagem de máquinas e da inteligência artificial.

### **Diretrizes para construir uma base de Big Data**

#### **Alinhar o Big Data com objetivos específicos**

Filtrar os logs da web para entender o comportamento do comércio eletrônico derivando o sentimento dos mídia social e interações de suporte ao cliente e compreender os métodos de correlação estatística e sua relevância para os dados dos clientes, produtos, manufatura e engenharia.

#### **Facilitar a falta de conhecimentos com padrões e governança**

Um dos maiores obstáculos para se beneficiar do investimento em big data é a falta de conhecimentos que podem ser mitigadas garantindo que as tecnologias, considerações e decisões de big data sejam adicionadas ao programa de governança das TI. Padronizar a abordagem permite gerir custos e aproveitar recursos.

As organizações que implementam soluções e estratégias de big data devem avaliar os seus requisitos de conhecimento com frequência e antecedência e devem identificar proativamente potenciais lacunas, que podem ser resolvidas cruzando os recursos existentes, contratando novos recursos ou mediante consultoria.

### **Otimizar a transferência de conhecimento com um centro de excelência**

Utilizar um centro de excelência para compartilhar conhecimento, controlar a supervisão e gerir as comunicações do projeto. Quer o big data seja um investimento novo ou em expansão, os custos de software e de hardware podem ser compartilhados por toda a empresa, ajudando a aumentar os recursos de big data e a maturidade geral da arquitetura de informação de uma forma mais estruturada e sistemática.

### **Retorno do alinhamento de dados não estruturados com dados estruturados**

É importante analisar o big data isoladamente mas pode-se incrementar o conhecimento de negócios ligando e integrando big data de baixa densidade com os dados estruturados que já estão a ser usados.

Ao captar clientes, produtos, equipamentos ou big data ambientais, o objetivo é adicionar dados relevantes aos principais resumos analíticos, levando a melhores conclusões; os processos e modelos analíticos de big data podem ser baseados em humanos e em máquinas.

Os recursos analíticos de big data incluem estatística, análise espacial, semântica, descoberta interativa e visualização. Usando modelos analíticos pode-se correlacionar diferentes tipos e fontes de dados para fazer associações e descobertas significativas.

Dados estruturados estão formatados (organizados em tabelas, ex. as folhas de cálculo), os semiestruturados (NoSQL), ex. textos, os não estruturados resultam dos Serviços Suplementares(USSD), ex. dos “Quik Codes”.

### **Planeamento de laboratório para descoberta de desempenho**

Descobrir o significado dos dados nem sempre é simples pois, por vezes, nem se sabe o que se está à procura; a gestão e as TI tem de dar suporte a essa "falta de direção" ou "falta de requisitos claros".

É importante que os analistas e cientistas de dados trabalhem em estreita colaboração com as empresas para compreender as principais lacunas e requisitos de conhecimento do negócio. Para acomodar a exploração interativa de dados e a experimentação de algoritmos estatísticos é preciso áreas de trabalho de alto desempenho e certificar-se de que o ambiente da área restrita tem o suporte de que precisa e que seja administrada adequadamente.

### **Alinhamento com o modelo operacional da nuvem**

Os processos e os utilizadores de big data requerem acesso a uma ampla gama de recursos para experimentação iterativa e execução de tarefas. Uma solução de big data inclui todos os domínios de dados, incluindo transações, dados mestre, dados de referência e dados resumidos.

A gestão de recursos é fundamental para garantir o controle de todo o fluxo de dados, incluindo pré e pós-processamento, integração, sumarização no banco de dados e modelagem analítica. Uma estratégia de segurança e provisionamento de nuvem pública e privada bem planeada desempenha um papel fundamental no suporte a esses requisitos em mudança.

## Departamento de Agronomia

[https://video.search.yahoo.com/search/video?fr=mcafee\\_uninternational&p=big+data+video#id=6&vid=3e663e68f699e9e15dee3be5ccfa672b&action=view](https://video.search.yahoo.com/search/video?fr=mcafee_uninternational&p=big+data+video#id=6&vid=3e663e68f699e9e15dee3be5ccfa672b&action=view)

[https://video.search.yahoo.com/search/video?fr=mcafee\\_uninternational&p=big+data+video#id=5&vid=2af8cf0d050ccbce5c6d335fe809fc44&action=view](https://video.search.yahoo.com/search/video?fr=mcafee_uninternational&p=big+data+video#id=5&vid=2af8cf0d050ccbce5c6d335fe809fc44&action=view)

[https://video.search.yahoo.com/search/video?fr=mcafee\\_uninternational&p=big+data+video#id=13&vid=fa17a1776ffb848840d780993df540f1&action=view](https://video.search.yahoo.com/search/video?fr=mcafee_uninternational&p=big+data+video#id=13&vid=fa17a1776ffb848840d780993df540f1&action=view)

[https://video.search.yahoo.com/search/video?fr=mcafee\\_uninternational&p=big+data+video#id=2&vid=fa5f0d5134d66540571f336e4eb164fa&action=view](https://video.search.yahoo.com/search/video?fr=mcafee_uninternational&p=big+data+video#id=2&vid=fa5f0d5134d66540571f336e4eb164fa&action=view)

[https://video.search.yahoo.com/search/video?fr=mcafee\\_uninternational&p=big+data+video#id=4&vid=b1aa432ab37e2b0e179d2f786ba5389a&action=view](https://video.search.yahoo.com/search/video?fr=mcafee_uninternational&p=big+data+video#id=4&vid=b1aa432ab37e2b0e179d2f786ba5389a&action=view)

[https://video.search.yahoo.com/search/video?fr=mcafee\\_uninternational&p=big+data+video#id=12&vid=2dbca295a9c5910faf93b8df960a1861&action=view](https://video.search.yahoo.com/search/video?fr=mcafee_uninternational&p=big+data+video#id=12&vid=2dbca295a9c5910faf93b8df960a1861&action=view)

[https://video.search.yahoo.com/search/video?fr=mcafee\\_uninternational&p=big+data+video#id=15&vid=51ecaac2497c8955b4e5d63e6b5c0eb6&action=view](https://video.search.yahoo.com/search/video?fr=mcafee_uninternational&p=big+data+video#id=15&vid=51ecaac2497c8955b4e5d63e6b5c0eb6&action=view)

[https://video.search.yahoo.com/search/video?fr=mcafee\\_uninternational&p=big+data+video#id=17&vid=167535d9dab03f3b805b6e17ff59385a&action=view](https://video.search.yahoo.com/search/video?fr=mcafee_uninternational&p=big+data+video#id=17&vid=167535d9dab03f3b805b6e17ff59385a&action=view)

[https://video.search.yahoo.com/search/video?fr=mcafee\\_uninternational&p=big+data+video#id=19&vid=c0d4a82434aaaa327023f816931c5528&action=view](https://video.search.yahoo.com/search/video?fr=mcafee_uninternational&p=big+data+video#id=19&vid=c0d4a82434aaaa327023f816931c5528&action=view)

[https://video.search.yahoo.com/search/video?fr=mcafee\\_uninternational&p=big+data+video#id=21&vid=bc389eb570c9dfbfec8e0fff7c7915aa&action=view](https://video.search.yahoo.com/search/video?fr=mcafee_uninternational&p=big+data+video#id=21&vid=bc389eb570c9dfbfec8e0fff7c7915aa&action=view)

[https://video.search.yahoo.com/search/video?fr=mcafee\\_uninternational&p=big+data+video#id=22&vid=3169273958b42ac2ae67eedcdef4d8f5&action=view](https://video.search.yahoo.com/search/video?fr=mcafee_uninternational&p=big+data+video#id=22&vid=3169273958b42ac2ae67eedcdef4d8f5&action=view)

[https://video.search.yahoo.com/search/video?fr=mcafee\\_uninternational&p=big+data+video#id=28&vid=20e63744d5d10ac7a370845787869d3e&action=view](https://video.search.yahoo.com/search/video?fr=mcafee_uninternational&p=big+data+video#id=28&vid=20e63744d5d10ac7a370845787869d3e&action=view) (tutorial do Hadoop)

[https://video.search.yahoo.com/search/video?fr=mcafee\\_uninternational&p=big+data+video#id=8&vid=b35fa34995c904906d3a568ef6bd67a2&action=view](https://video.search.yahoo.com/search/video?fr=mcafee_uninternational&p=big+data+video#id=8&vid=b35fa34995c904906d3a568ef6bd67a2&action=view)

[https://video.search.yahoo.com/search/video?fr=mcafee\\_uninternational&p=big+data+video#id=14&vid=c0d4a82434aaaa327023f816931c5528&action=view](https://video.search.yahoo.com/search/video?fr=mcafee_uninternational&p=big+data+video#id=14&vid=c0d4a82434aaaa327023f816931c5528&action=view)