Big data

Desafio que motivou o uso de Big Data:

Imagine que um hospital recebe milhares de pacientes por mês, cada um com históricos médicos diferentes, exames, prescrições e sintomas registrados em sistemas variados. Antes do uso de Big Data, era difícil cruzar essas informações para detectar padrões, prever surtos de doenças ou oferecer tratamentos personalizados.  
Com Big Data, é possível analisar rapidamente todos esses dados, identificar riscos de complicações em tempo real, prever a propagação de doenças e até sugerir intervenções médicas baseadas em casos similares. Isso torna o atendimento mais eficiente e salva vidas.

Como foi resolvido:

Big Data resolve os desafios de analisar grandes volumes de dados ao usar tecnologias como Hadoop e Spark para processar dados em paralelo em múltiplos servidores, permitindo lidar com dados estruturados e não estruturados. Isso possibilita a análise em tempo real de informações. Além disso, algoritmos de machine learning ajudam a identificar padrões e prever complicações, enquanto ferramentas de visualização de dados permitem que médicos tomem decisões rápidas.

Casos de uso:

Experiência do cliente – É possível ter uma visão clara de como está sendo a experiência do cliente, agora que com big data podemos reunir dados de mídias sociais, visitas da web, registro de chamadas e outras fontes para aprimorar a experiência de interação. Com isso é possível oferecer ofertas personalizadas, conteúdos específicos para cada usuário, ou seja, os tão ditos algoritmos da internet.

Machine Learning – Machine learning está crescendo rapidamente, parte disso é graças à big data. Usando grande quantidade de informações de dados para aprimorar as decisões tomadas pelo modelo. Examine as tendências e o que os clientes desejam para oferecer novos produtos e serviços. Implemente um sistema de preços dinâmico. Há infinitas possibilidades.

Fraude e conformidade - Big data ajuda você a identificar padrões em dados que indicam fraudes e agregar grandes volumes de informações para tornar os relatórios regulamentares muito mais rápidos.

Big Data é o termo utilizado para designar um grande conjunto de dados gerados no ambiente digital. Nesse contexto, refere-se à imensa quantidade de informações que é produzida constantemente por dispositivos, redes sociais, transações online e outras fontes digitais. A expressão abrange não apenas a quantidade volumosa de dados, mas também a variedade de formatos e a velocidade com que essas informações são geradas.

E esse crescente aumento da presença online resulta em um considerável acúmulo de dados. Nesse cenário, não apenas a análise, mas também a organização desses dados é crucial, afinal, o verdadeiro potencial do Big Data se manifesta quando conseguimos atribuir significado a esses dados.

Tipos de dados:

Estruturados: Armazenados em bancos de dados tradicionais, em tabelas.

Semiestruturados: Podem seguir diversos padrões, de forma heterogênea.

Não-Estruturados: Mescla de dados oriundos de várias fontes distinta**s**, como vídeo, texto, áudio, imagens, XML.

Big data é um grande volume de dados e com muita variedade, o que requer mais velocidade de processamento, definido com os três V’s (+ 2 V’s).

Volume - Lidando com big data, você terá que processar grandes volumes de dados não estruturados de baixa densidade. Podem ser dados de valor desconhecido, como feeds de dados do Twitter, fluxos de cliques em uma página da web. Para algumas empresas, isso pode utilizar dezenas de terabytes de dados. Para outras, podem ser centenas de petabytes.

Velocidade - Alguns produtos inteligentes habilitados para internet operam em tempo real ou quase em tempo real e exigem avaliação e ação em tempo real. Portanto a análise de big data precisa ser rápida para processar todo o grande volume de dados complexos em questão.

Variedade – Big data lida com diversos tipos de dados. Os tipos tradicionais foram estruturados e se adequam perfeitamente a um banco de dados, porém com o aumento de big data, os dados vêm em novos tipos de dados não estruturados. Tipos de dados não estruturados e semiestruturados, como texto, áudio e vídeo, exigem um pré-processamento adicional para obter significado e dar suporte a metadados.

Nos últimos anos mais dois V’s foram adicionados: valor e veracidade.

Valor - Dados possuem valor intrínseco. Mas isso é inútil até que esse valor seja descoberto.

Veracidade – Seus dados apontam fatos descobertos após uma análise, o quanto você pode confiar neles?

Breve história:

Por volta de 2005, as pessoas começaram a perceber a quantidade de usuários de dados gerados pelo Facebook, YouTube e outros serviços online. O Hadoop (uma estrutura de código aberto criada especificamente para armazenar e analisar grandes conjuntos de dados) foi desenvolvido no mesmo ano.

O desenvolvimento de estruturas de código aberto, como o Hadoop, (e, mais recentemente, o Spark) foi essencial para o crescimento de big data, porque elas tornaram o trabalho com big data mais fácil e seu armazenamento mais barato. Nos anos seguintes, o volume de big data disparou.

Com o desenvolvimento e crescimento de IOT (Internet das Coisas), que reúne dados sobre padrão de uso do cliente e desempenho do produto, assim como o surgimento de machine learning, aumentaram a quantidade de dados produzidos.

A computação em nuvem expandiu ainda mais as possibilidades de big data. A nuvem oferece uma escalabilidade verdadeiramente elástica, na qual os desenvolvedores podem simplesmente criar clusters ad hoc para testar um subconjunto de dados.

Dificuldades enfrentas:

Mesmo com big data facilitando muito o desenvolvimento de diversas coisas, ele continua sendo BIG data, ou seja, ele é muito grande. A tecnologia de armazenamento de dados está se desenvolvendo, mas os dados gerados estão dobrando a cada dois anos. Portanto está sendo um desafio a capacidade de armazenamento acompanhar o volume de dados gerados com eficiência.

Além de armazenar dados é necessário limpá-los, em alguns casos você pode receber informações para serem analisadas semelhantes:

Nome: Donzela Trevisan Wendy

Idade: 13

Filme preferido: Encanto (2021)

Considerando que existem mais dados de mais crianças e seus filmes favoritos, para analisar a média de idades de crianças que tem “Encanto (2021)” como seu filme favorito o nome da criança não é necessário, portanto, podemos descartá-lo.

Por fim, a tecnologia de big data está mudando em ritmo acelerado. Há alguns anos, o Apache Hadoop era a tecnologia popular usada para lidar com big data. Em seguida, o Apache Spark foi introduzido em 2014. Hoje, uma combinação das duas estruturas parece ser a melhor abordagem. Manter-se atualizado com a tecnologia de big data é um desafio contínuo.

Os três passos de big data

**1.** Integrar - Big data reúne dados de diversas fontes e aplicativos diferentes. Os mecanismos tradicionais de integração de dados, como extração, transformação e carregamento (ETL), geralmente não estão aptos à tarefa. Isso requer novas estratégias e tecnologias para analisar conjuntos de big data em terabytes ou até mesmo em escala de petabytes.

Durante a integração, você precisa inserir os dados, processá-los e verificar se estão formatados e disponíveis de forma que seus analistas de negócios possam começar a utilizá-los.

**2.** Gerenciar - Big data exige armazenamento. Sua solução de armazenamento pode estar na nuvem, no local ou em ambos. Você pode armazenar seus dados da forma que desejar e trazer os requisitos de processamento desejados e os mecanismos de processos necessários para esses conjuntos de dados sob demanda. Muitas pessoas escolhem a solução de armazenamento de acordo com a localização atual dos dados. A nuvem está gradualmente ganhando popularidade porque é compatível com as suas necessidades atuais de computação e permite que você crie recursos conforme necessário.

**3.** Analisar - Seu investimento em big data é compensado quando você analisa seus dados e age. Obtenha mais clareza com uma análise visual dos seus conjuntos de dados variados. Explore ainda mais os dados para fazer novas descobertas. Compartilhe suas descobertas com os outros. Crie modelos de dados com machine learning e inteligência artificial. Faça seus dados funcionarem.

Benefícios:

Big data te retorna respostas mais completas, uma vez que podem trabalhar com uma quantidade maior de informações.

Com respostas mais completas a Veracidade dos dados cresce também, agora com uma análise mais detalhada chegamos a respostas mais concretas