



Ferramentas e Tecnologias para Big Data Analytics

Com o intuito de ajudar as empresas a trilhar o caminho da inteligência ao navegar em grandes volumes de dados, a indústria tecnológica se movimenta e enxerga o imenso potencial do Big Data. As ferramentas e tecnologias para Big Data Analytics desempenham um papel fundamental na extração de valor e insights significativos a partir de grandes volumes de dados. À medida que a quantidade de dados disponíveis continua a crescer exponencialmente, torna-se cada vez mais crucial contar com ferramentas adequadas para lidar com esse desafio.

A Análise de Dados e a Automação

Milhões de planilhas preparadas manualmente são usadas em diversos setores, incluindo finanças, ciência e economia. No entanto, de acordo com ZDNet, 90% de todas as planilhas contêm erros que afetam seus resultados. Problemas de corte e colagem, células ocultas e outros erros custam milhões de dólares para as empresas.

As soluções e os processos analíticos tradicionais também podem causar atrasos no fornecimento das informações necessárias às empresas para tomar decisões oportunas. Automatizar processos de análise e colocar os processos na nuvem pode ser um divisor de águas para empresas de todos os tamanhos e setores.





Os Três Componentes Principais do Big Data

Uso de Dados

Identificando oportunidades e construindo confiança. As empresas devem criar uma cultura que encoraje experiências e apoie um processo de ideação voltado para os dados. Também precisam focar a confiança, não apenas construindo-a com os consumidores, mas a exercendo como arma competitiva.

Motor de Dados

Estabelecendo uma base técnica e moldando a organização. Plataformas técnicas rápidas, escaláveis e flexíveis o bastante para atender a diferentes tipos de aplicações são fundamentais, assim como o conjunto de habilidades necessário para construí-las e administrá-las.

Ecosistema de Dados

Participando de um ecossistema de big data e criando relações. O big data gera oportunidades que muitas vezes não estão nos negócios ou mercados tradicionais de uma empresa. Parcerias serão cada vez mais necessárias para obter dados, conhecimento, capacidades ou clientes.



Competências para Materializar o Big Data

1. Identificar Oportunidades

O big data pode agregar valor ao negócio de várias maneiras, mas as oportunidades de inovar mais com dados – potencialmente mais lucrativas – provavelmente não serão muito aparentes. As empresas precisam criar um ambiente no qual novas aplicações de dados possam ser rapidamente identificadas e desenvolvidas.

1 Encorajar Ideias Não Tradicionais

A exploração de novas aplicações de dados tem de ser encorajada em todos os níveis da organização, dando aos colaboradores tempo e recursos para que desenvolvam suas ideias.

2 Fomentar a Colaboração

A ampla variedade de conhecimentos necessários para identificar e desenvolver aplicações raramente será adquirida por um único indivíduo. Tais esforços tendem a exigir habilidades de muitas pessoas.

3 Adotar Abordagem "Teste e Aprenda"

Velocidade e agilidade são essenciais para a criação de aplicações de big data. A regra deve ser ciclos curtos, desenvolvimento iterativo e pilotos frequentes.



Construindo Confiança com os Dados

O acesso à informação, em grande parte de natureza pessoal, é fundamental para extrair valor de aplicações de big data. No entanto, os consumidores se preocupam cada vez mais com o modo como as empresas usarão as informações que acessam sobre eles.

7%

Confiança dos Consumidores

Apenas 7% dos consumidores se sentem confortáveis com a possibilidade de seus dados serem utilizados fora do propósito para o qual foram coletados.

5-10x

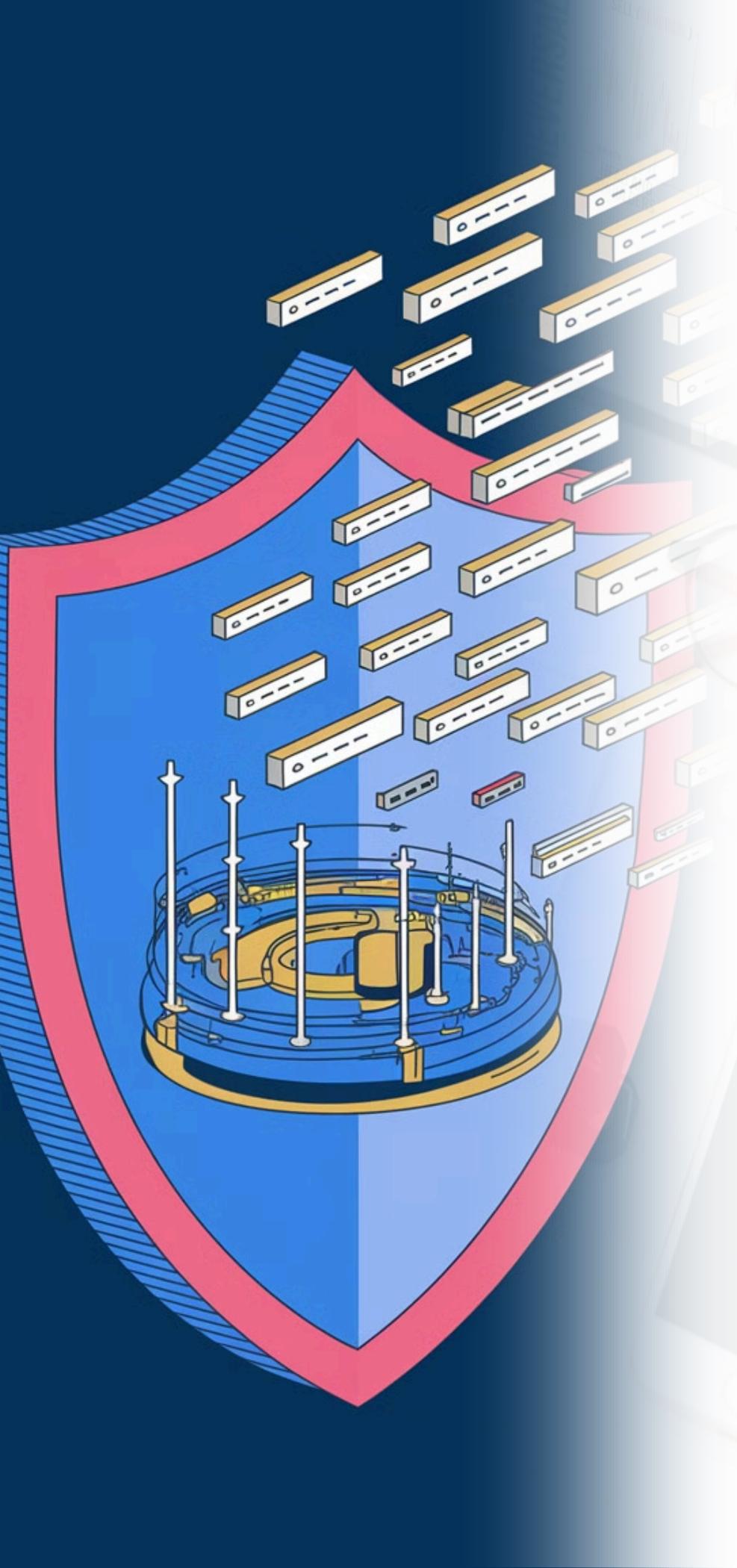
Vantagem da Confiança

Organizações que administram bem a confiança são recompensadas com cinco a dez vezes mais acesso aos consumidores na maioria dos países.

3

Práticas Essenciais

Informar com clareza como os dados serão usados, oferecer escolhas e controle, e enfatizar os benefícios do uso de dados.





Estabelecendo a Plataforma Tecnológica

O relacionamento com dados é antigo nas empresas, e esse passado muitas vezes pode se voltar contra elas no caso do big data. A experiência acumulada ensina que a infraestrutura de TI deve ser grande, rígida e cara, formada por sistemas complexos customizados para uma tarefa particular.

Ocorre que o big data é uma experiência muito diferente disso, com tecnologias, requisitos e possibilidades diversos. A infraestrutura de dados tradicional, que depende de servidores de dados centralizados e altamente estruturados, não é mais a única opção.

Ferramentas novas como o Hadoop podem ser melhores do que as tradicionais tecnologias SQL



Orientações para Construir a Plataforma Ideal

1

Usar uma Plataforma Escalável e Polivalente

Implementar uma plataforma de abrangência corporativa ajuda a evitar o problema da "anarquia de dados", na qual unidades de negócios diferentes utilizam fontes de dados duplicadas ou conflitantes.

2

Não Eliminar os Sistemas Existentes

Apesar de as tecnologias SQL não oferecerem tanta flexibilidade como as ferramentas mais recentes, elas são maduras e trabalham bem com dados essenciais.

3

Ajustar Processos Operacionais

Embora o big data costume gerar vários insights valiosos, estes não significarão muito se não forem alavancados de maneira oportuna – cada vez mais, em tempo real ou perto disso.



Moldando a Organização para o Big Data

As plataformas de big data mais bem-sucedidas alavancam não apenas novas tecnologias, mas também novas estruturas corporativas. Centralizar recursos-chaves (cientistas de dados e analistas, por exemplo) em uma unidade isolada ajudará as empresas a atrair e conservar os talentos de que precisam.

1 Criar um Centro de Excelência

Criar um centro de excelência com talentos na área permite que o conhecimento seja construído de maneira rápida, expondo esses profissionais a uma série de problemas e soluções.

2 Obter Apoio do Nível Sênior

Iniciativas de big data precisam de um patrocinador na organização – um executivo sênior dinâmico com prestígio suficiente para se certificar de que as coisas estão sendo feitas.

3 Atrair e Reter Habilidades Fundamentais

Provavelmente, novos conjuntos de habilidades serão exigidos, e os profissionais que os possuem vão querer trabalhar em ambientes não tradicionais.



Participando de um Ecossistema de Dados

O big data está transformando não apenas a maneira como as empresas fazem negócios, mas também com quem negociam. Ao criarem a necessidade de parcerias, as novas aplicações de dados muitas vezes borrarão as fronteiras do setor de atividade.

Entender a Oportunidade Econômica

A empresa deve dar uma olhada nos produtos e serviços existentes: que dados eles geram? Que dados adicionais podem melhorá-los?

1

2

3

Identificar Parceiros Estratégicos

Em uma aliança de dados bem-sucedida, os parceiros oferecem recursos complementares e conhecimento: dados, competências e ativos.

Começar de Baixo e Escalar

Uma abordagem iterativa e exploradora do big data é melhor do que uma estratégia de três anos elaborada nos mínimos detalhes.



Fazendo os Relacionamentos Funcionarem

As parcerias que o big data estimula têm de ser administradas e mantidas. Os termos da empresa precisam ser construídos de maneira que todos possam prosperar e receber um incentivo para trocar informações complementares.

Construir Competências para Parcerias

Ecosistemas de dados mudam a praxe da maioria das empresas, que é a de criar coisas sozinhas; várias organizações têm de trabalhar em conjunto para oferecer novos produtos e serviços aos clientes.

Criar Cláusulas Contratuais Benéficas

Impor cláusulas contratuais restritivas para parceiros pode afastá-los. Ceder muito, de outro lado, pode encolher desnecessariamente o lucro potencial.

Assegurar Integração Tecnológica

Parceiros de ecossistema precisam compartilhar dados rápida e facilmente. Uma empresa, então, deve permitir o acesso de terceiros a essas plataformas de dados.



NoSQL: Flexibilidade para Dados Não Estruturados

NoSQL é um banco de dados, não uma técnica de organização de dados ou documentos. São bancos de dados que trabalham com dados semi e não estruturados e possuem altíssima flexibilidade na forma como trabalham os dados.

Em bancos de dados NoSQL, os dados podem ser armazenados sem definir o esquema com antecedência—o que significa que você tem a capacidade de se mover e iterar rapidamente, definindo o modelo de dados à medida que avança.

Armazenamento por Chave-Valor

Dados armazenados como pares de chave-valor, permitindo recuperação rápida de informações.

Armazenamento Colunar

Dados organizados por colunas em vez de linhas, otimizando consultas analíticas.

Orientado a Documentos

Armazena documentos completos em formatos como JSON, ideal para conteúdo web.

Orientado a Grafos

Foca nas relações entre dados, perfeito para redes sociais e recomendações.



Hadoop: Plataforma para Big Data

O Hadoop não é uma técnica, mas uma plataforma de código aberto que armazena e processa de forma distribuída grandes conjuntos de dados. Trata-se de um software open source escrito em Java para suportar grandes quantidades de dados em um ambiente de computação distribuída, concebido pela Apache Software Foundation.

O ecossistema Hadoop trabalha com duas estruturas, uma de armazenamento e outra de processamento. O sistema funciona dividindo grandes quantidades de dados em blocos que são distribuídos nos clusters para então serem processados.



Computação Distribuída

Trabalha com clusters, que são grupos de computadores ou servidores trabalhando em conjunto em uma arquitetura de sistema capaz de unir o processamento dessas máquinas.



Tolerância a Falhas

Se um dos nós deixar de funcionar, outro assume o seu lugar, pois o Hadoop distribui os arquivos de dados em todos os nós, cada um possuindo uma réplica dos arquivos.



Processamento Rápido

A distribuição de dados entre múltiplos nós permite processamento paralelo, tornando a análise de grandes volumes de dados muito mais eficiente.



Data Lake: Armazenamento Flexível

Data Lake é um armazém de dados que, diferentemente do Data Warehouse, tem como característica armazenar todo tipo de dados: estruturados, semiestruturados e não estruturados em grandes volumes.

Armazenamento Original

No Data Lake os diferentes dados são acessados e armazenados em sua forma original, permitindo buscar correlações e insights diretamente, sem processamento prévio.

Schema on Read

Os modelos de dados não são impostos antecipadamente, mas emergem à medida que trabalhamos com os próprios dados, seguindo o conceito de "late binding".

Complemento ao Data Warehouse

Pode-se gerar o tradicional Data Warehouse a partir do Data Lake para tratar dados estruturados, oferecendo o melhor dos dois mundos.

Diferentemente do Data Warehouse, onde o modelo de dados ou schema deve ser previamente definido, o Data Lake permite maior flexibilidade e adaptação às necessidades analíticas emergentes.



Internet das Coisas (IoT)

O termo IoT (Internet of Things) foi utilizado primeiramente em 1999 por Kevin Ashton, e abrange todo tipo de "coisas" que podem estar conectadas, não necessariamente à Internet. Essas coisas podem estar ligadas a outros dispositivos através de uma rede dedicada ou até mesmo através de Bluetooth ou infravermelho.



Dispositivos Domésticos

Pulseiras digitais, smartphones, Smart TVs, eletrodomésticos inteligentes como geladeiras, lava-louças, termostatos e lâmpadas, todos capazes de se conectar e compartilhar dados.



Aplicações Industriais

Sistemas inteligentes de ar condicionado, sensores industriais e dispositivos com reconhecimento ambiental que ajudam empresas a monitorar e otimizar operações.



Cidades Inteligentes

Câmeras de segurança, sistemas de tráfego e dispositivos que ajudam governos a responder mais rapidamente a desastres naturais, como terremotos e incêndios florestais.



O Futuro do Big Data e Seus Desafios

Para controlar o big data, as empresas terão de esquecer muito do que sabem sobre trabalhar com dados. Precisarão adotar novas mentalidades, novas tecnologias e, sobretudo, novas competências. E terão de fazer isso rapidamente, porque o big data oferece riscos – organizações tradicionais podem se ver vulneráveis a novos entrantes do mercado que sobressaiam nessas competências.

- 1 **Transformação Cultural**
Adoção de mentalidade orientada a dados
- 2 **Capacitação Técnica**
Desenvolvimento de habilidades e ferramentas
- 3 **Infraestrutura Escalável**
Plataformas flexíveis e adaptáveis
- 4 **Governança de Dados**
Políticas de segurança e privacidade
- 5 **Parcerias Estratégicas**
Ecossistema colaborativo de dados

O sucesso na era do Big Data não depende apenas de investimento em ferramentas, mas principalmente da capacidade da organização em criar uma cultura de dados, desenvolver expertise interna, lidar com a complexidade tecnológica, fomentar soluções inovadoras e promover o trabalho em equipe multidisciplinar.