

# O grande livro de Big Data

Um guia prático para tirar o  
seu primeiro projeto de Big Data  
do papel.

# Conteúdo

<b>Introdução</b>	<b>3</b>
-------------------	----------



## Parte A: Como se preparar

4

### O que você precisa saber

5

Por que a maioria das empresas implementa projetos de Big Data

6

Por que os projetos de Big Data falham

7

Como fazer o seu projeto de Big Data funcionar

10

### Escolha do projeto certo

12

Como seria o projeto certo

13

Considere o impacto

15

Projetos táticos de Big Data: Alguns exemplos

17

A jornada básica do Big Data

19



## Parte B: Sua estratégia

23

### Defina as suas metas

24

As metas de negócios

25

Metas de TI

27

### Defina as necessidades de dados

29

Quais são os dados que você precisa?

30

Cinco considerações essenciais em relação aos dados

33



## Parte C: Sua cadeia de suprimentos de Big Data descomplicada

36

### Sua equipe

37

Cinco lições essenciais para a criação da equipe

38

Implemente a governança de dados

42

As habilidades que você tem e aquelas de que precisa

45

### Suas ferramentas

47

Noções básicas sobre as ferramentas de Big Data

48

### Seus processos

52

Os oito Big Data

53

### Sua arquitetura

56

Primeiros passos: Sua área restrita

57

A arquitetura ideal do Big Data

59

### Seu plano de projeto

60

Seu plano de projeto

61

### Seguindo em frente

63

### Próximos passos

64

### Sobre a Informatica

65

**Dica: Clique para acessar a seção**

# Como ser grande

Poucas tendências tecnológicas foram tão comentadas quanto o Big Data.

E, ao mesmo tempo, poucas tendências tecnológicas ofereceram às empresas tanto potencial de transformação. Desde que o software começou a envolver processos de negócios completos, na virada do século, uma coisa ficou clara: os dados mudam o modo como trabalhamos.

É claro que, com grandes expectativas, também vêm grandes decepções. E, no caso do Big Data, com certeza já vimos bons conselhos e informações erradas difundidos na mesma medida. Infelizmente, como este admirável mundo novo de dados infinitos ainda é bem recente, todo esse barulho deixou muita gente confusa.

Este livro tem o objetivo de desfazer essa confusão.

O segredo é blindar a sua estratégia e executá-la de forma pragmática. Este livro servirá como um guia prático na sua jornada, esteja você começando uma iniciativa específica e tática ou planejando um esforço mais estrutural para toda a empresa.

**Vamos nos aprofundar.**



# Parte A: Como se preparar

O livro está dividido em três partes.  
Nesta primeira parte, tentaremos aprimorar a sua  
visão para que você possa escolher o projeto certo.

[Voltar para Conteúdo](#)







# O que você precisa saber

Antes de falarmos sobre as especificidades do seu próprio projeto, veja algumas questões que a maioria dos profissionais que estão trabalhando com Big Data gostaria de ter ficado sabendo antes de começarem seus projetos.





O que você  
precisa saber

# Por que a maioria das organizações implementa projetos de Big Data

Quando as empresas decidem que vão se dedicar ao Big Data, o motivo geralmente é um dos descritos a seguir.

**Ao tentar melhorar suas análises,** as empresas percebem que, para isso, precisarão analisar uma quantidade de dados significativamente maior. Geralmente, uma unidade de negócio (por exemplo, marketing) começa essas iniciativas.

**O processo torna evidente que é possível envolver os produtos em uma camada de serviços,** fornecendo uma análise (quase sempre em tempo real) capaz de ajudar os clientes a usar os produtos de maneira mais eficiente e infalível.

**As empresas querem fazer as coisas de maneira mais rápida, melhor e mais barata** usando o Big Data para comunicar todas as decisões relacionadas a uma unidade de negócio ou a um processo específico.

**Percebem que o Big Data é essencial para todas as unidades de negócio** da organização e tentam estabelecer as bases para uma visão centralizada em dados, em todo o seu universo.

**Elas sabem que precisam começar um projeto de Big Data antes que seja tarde demais,** mas ainda não sabem realmente o que fazer. O objetivo é aprender e experimentar.

Todos esses aspectos são ótimos motivos para se interessar por Big Data. Mas, se quiser garantir que os seus projetos resistam ao teste do tempo (e da análise multidepartamental), você terá que ser muito claro sobre qual desses motivos melhor representa o seu interesse.



O que você  
precisa saber

# Por que os projetos de Big Data falham

Uma pesquisa<sup>1</sup> descobriu que 55% de todos os projetos de Big Data não são concluídos e muitos outros ficam aquém dos objetivos. Embora essa taxa de desistência não seja atípica em um estágio tão inicial de uma tendência de tecnologia, seria insensato não aprender as lições que esses projetos podem ensinar.

Vamos analisar os quatro principais motivos para o fracasso de projetos de Big Data.

1

## Metas vagas

O motivo de falha mais citado na pesquisa foi o “escopo impreciso” do projeto. Muitas empresas têm como meta projetos ambíguos, (ambíguos demais, vale ressaltar), com objetivos não muito claros, e que falham quando elas precisam fazer escolhas difíceis sobre o que é ou não é importante.

Se envolver com Big Data só para ter um projeto de Big Data é a receita para o desastre. As complexidades desses projetos exigem um compromisso firme com a obtenção de um determinado resultado. Sem a certeza das metas, isso não é possível.



O que você  
precisa saber

# Por que os projetos de Big Data falham

2

## Expectativas mal administradas

Todo o entusiasmo em torno do Big Data gera algumas expectativas muito perigosas sobre o que o projeto pode proporcionar. Por mais que seja tentador fazer promessas a curto prazo, é importante manter uma visão realista do que se pode esperar do projeto, quanto tempo isso vai levar e a quantidade de esforços necessários para chegar lá.

Quando a expectativa de impacto e insights é muito alta, você se pega navegando em terabytes de anônimos e desconhecidos à procura de ouro. Quando a expectativa de resultados não é realista, você acaba buscando prazos e orçamentos que não são justos.

3

## Atrasos e demoras nos projetos

Apesar dessa disciplina ser bem nova para as empresas, não é novidade que a maioria dos projetos de Big Data acaba custando caro ou demorando muito. Normalmente, isso é resultado de uma mistura de expectativas mal administradas e de visões inadequadas sobre como construir uma arquitetura dimensionável.

Quando raros e caros desenvolvedores de Hadoop Java são contratados e encarregados de implementações gigantescas com codificação manual, as empresas logo percebem que é impossível sair do ambiente de área restrita sem erros. Como resultado, os projetos de Big Data acabam definhando como um experimento científico e nunca saem do laboratório.





O que você  
precisa saber

# Por que os projetos de Big Data falham

4

## Incapacidade de dimensionar

Já é difícil encontrar cinco ótimos desenvolvedores Hadoop Java. Mas, quando os projetos aumentam e precisam se ajustar a 30 desenvolvedores Java em um único ano, as coisas podem chegar ao limite. O pior não é o custo da oportunidade de clusters Hadoop não usados; é o custo da oportunidade e do tempo perdidos.

Muito frequentemente, as empresas buscam a conveniência do curto prazo em vez da sustentabilidade a longo prazo. Embora seja precipitado de nossa parte sugerir que é sempre possível evitar a troca de um pelo outro, nunca é demais enfatizar a importância da visão de longo prazo. Para que os seus dados sejam protegidos e gerenciados adequadamente, você precisa sempre supervisionar as implicações de longo prazo do seu projeto.

Os quatro motivos do fracasso do Big Data são preocupantes e muito comuns. Por isso, vamos falar sobre como é possível evitá-los e construir uma implementação duradoura.



O que você  
precisa saber

# Como fazer o seu projeto de Big Data funcionar

Como a maioria dos projetos de Big Data falha por falta de clareza e devido à incapacidade de demonstrar a funcionalidade da iniciativa, você mesmo deve se responsabilizar por isso para trazer foco e comprovação ao seu projeto. Seguem três dicas úteis para garantir que o seu projeto já comece dando certo e continue em operação.

1

## Defina objetivos claros e administre as expectativas

Se você não sabe qual deve ser a meta do seu projeto, pense nos objetivos que definiu para a infraestrutura de dados existente.

Se a sua organização já precisa de dados para determinados processos de negócios (como detecção de fraudes ou análise de mercado), pense em como o Big Data pode melhorar ou valorizar esses processos. Em vez de abordar um problema totalmente novo, você deve ter como alvo melhorar um processo ou projeto já existente.

Sem um foco claro e um valor perceptível aos usuários de negócios, o projeto estará condenado ao fracasso.

2

## Defina as métricas que comprovam o valor do projeto

Métricas claramente definidas e que se ajustem aos objetivos podem evitar uma grande quantidade de problemas. Ao definir metas realistas que podem ser medidas, todos ao seu redor conseguirão ver o progresso que você está obtendo.

O mais importante: eles saberão que você tem como objetivo o longo prazo. Pergunte a si mesmo como você pode medir o impacto do seu projeto no contexto das suas metas.

Isso é crucial, pois os usuários de negócios precisarão de ajuda para racionalizar compromissos de curto prazo, e as metas mensuráveis ajudam a provar que você está proporcionando mais valor do que eles imaginam.



# Como fazer o seu projeto de Big Data funcionar

## 3

### Seja estratégico sobre ferramentas e codificação manual

Evite a tentação de codificar tudo manualmente diretamente no Hadoop. Lembre-se de que o objetivo aqui não é construir uma implementação de trabalho manual a partir do zero; o objetivo é entregar o valor do Big Data para a sua organização.

Em vez de tentar codificar manualmente cada integração, limpar cada conjunto de dados e codificar manualmente todas as análises, você deve analisar ferramentas e automação para ajudar a acelerar esses processos.

O mais importante é não cair na armadilha de desperdiçar talentos de desenvolvimento de Java, raros e caros, em aspectos que não podem ser dimensionados ou transferidos para outros funcionários. A sua função é tomar decisões estratégicas sobre a implantação de recursos escassos, de forma a atingir seus objetivos.

Adote ferramentas que possam aumentar a produtividade da equipe de desenvolvimento por meio do aproveitamento das habilidades e do conhecimento da qualidade dos seus dados de ETL existentes e de especialistas em Business Intelligence, enquanto libera os especialistas em Java para trabalhar na lógica específica para a qual não há ferramentas disponíveis.

Além disso, como as tecnologias do tipo Hadoop evoluem a cada dia, vale a pena considerar uma camada de abstração que possa lhe proteger das constantes mudanças nas especificações das tecnologias subjacentes.

Acima de tudo, lembre-se de que as habilidades necessárias são escassas, mas as ferramentas estão sempre disponíveis.



# Escolha do projeto certo

Sob a perspectiva dos desafios que você enfrentará, vamos analisar agora o que você deve fazer para escolher o projeto certo para a sua organização.



# Como seria o projeto certo

Se a sua organização está ansiosa por mudanças e já aceitou que será preciso ter uma estrutura abrangente de governança de dados para melhorar a forma como eles trabalham, provavelmente você poderá ignorar esta seção.

Por outro lado, se você está pensando em um projeto tático e específico, que pode ser adaptado posteriormente para a empresa como um todo, continue lendo.

O projeto certo tem estes quatro componentes.

1

## Valor demonstrável

O projeto certo é aquele em que o valor é compartilhado de maneira igual entre a TI e a unidade de negócio que você está tentando ajudar. Isso significa proporcionar um valor claro para um departamento, uma unidade de negócio ou um grupo, de forma que eles possam ver.

2

## Patrocínio

Os executivos que aderirem à sua visão serão essenciais para o sucesso do projeto. Os projetos de Big Data precisam de defensores e patrocinadores em altos cargos, que estejam dispostos a defender o trabalho que você está fazendo.

Portanto, se você sabe que pode construir análises excelentes para a logística, mas o único executivo que aderiu foi o gerente de fábrica, pense novamente. Se o marketing é o seu defensor, trabalhe para apoiar os requisitos de análise de marketing. Não dá para forçar ninguém a mudar. Siga a influência e extraia o máximo de valor possível.





# Como seria o projeto certo

3

## Um efeito de jogo de boliche

A importância estratégica do seu primeiro projeto tático é vital. Além de provar, sem sombra de dúvida, que o Big Data pode ajudar a unidade de negócio que você está apoiando, é bom se certificar de que o valor possa ser comunicado facilmente para a empresa em geral.

Portanto, ao escolher o seu primeiro projeto, faça isso de maneira estratégica.

Depois de ter demonstrado o valor do Big Data para o seu departamento de marketing, por exemplo, será mais fácil conquistar a adesão das equipes de logística que talvez estejam reticentes.

4

## Habilidades transferíveis

Como dissemos no último ponto, é preciso que o valor do seu primeiro projeto ajude a convencer os outros departamentos da empresa. Para isso, é preciso se certificar de que você pode aprender as habilidades, as capacidades e as lições certas do seu primeiro projeto. Mais enfaticamente, é preciso garantir que tudo isso seja documentado para que você possa transferir conhecimento para o próximo projeto. Lembre-se, se a sua meta é o sucesso, então você já está se preparando para futuros projetos.

Portanto, prepare-se para dimensionar, de modo que você possa lidar com mais projetos no futuro. Não se trata apenas de dimensionar o seu cluster. Trata-se de dimensionar habilidades e operações. Você vai precisar encontrar mais especialistas de Java/Hadoop ou encontrar maneiras de extrair mais dos recursos que já tem.





# Considere o impacto

Quando for escolher qual será o seu próximo projeto, você também terá que pensar em como isso afetará a organização.

Há três aspectos gerais que devem influenciar a sua decisão de buscar o projeto certo de Big Data.

## 1

### Custo e interrupção

No nível mais básico, o custo do projeto se baseia no tempo e no dinheiro que serão necessários para torná-lo realidade. Na verdade, é preciso considerar também a possível interrupção que isso causará.

Às vezes, a interrupção é processual – por exemplo, quando as unidades de negócio são usadas para armazenar os dados e não se sentem à vontade em abrir mão do controle deles para uma estrutura centralizada de governança de dados.

Outras vezes, está relacionada com a tecnologia e as habilidades – ou seja, quando você precisa integrar novas tecnologias à infraestrutura existente e reorganizar ou atualizar as capacidades para fazer isso.

Seja qual for o caso, você deve prever, reconhecer e certificar-se de que quer minimizar a interrupção ou informar por que isso é importante.



# Considere o impacto

2

## Cronograma de benefícios e impacto

Ao considerar o início de diferentes projetos, a tendência é que você se volte aos que podem oferecer o máximo de impacto e melhoria aos negócios. Mas também é importante considerar a natureza do impacto nos negócios. O projeto proporcionará a maior parte do valor no curto ou no longo prazo?

Mais importante: quando os usuários de negócios verão os resultados nos negócios? Por exemplo, você poderia introduzir o gerenciamento de dados mestres no data warehouse e melhorar radicalmente a eficiência de Business Intelligence. Mas esse valor só será sentido quando os analistas de negócios perceberem que não precisarão limpar os dados financeiros novamente.

3

## Recursos e restrições

Tendo em conta a análise dos dois últimos fatores, considere os recursos à sua disposição. Nós abordaremos isso com mais detalhes posteriormente. Por enquanto, basta lembrar que, naturalmente, você quer que o seu projeto ofereça mais retorno do que o dinheiro investido.

Atingir essa meta funciona nos dois sentidos. Por um lado, você tem como meta o máximo impacto nos negócios. Mas também é preciso ser estratégico na hora de gastar o orçamento. Embora seja tentador construir uma equipe de cientistas de dados similar à equipe do Google, há realmente recursos suficientes para tanto? Fazer escolhas inteligentes sobre as ferramentas e o número de funcionários será fundamental para o sucesso do seu projeto.



# Projetos táticos de Big Data: Alguns exemplos

Há uma grande variedade de aplicativos para Big Data. Por mais emocionante que seja, isso também se torna um pouco assustador para as pessoas que não têm certeza quanto ao projeto certo para começar. Aqui está uma lista de projetos táticos de Big Data que vimos nossos clientes empreenderem.

Se você ainda não tem certeza sobre qual projeto deve ser iniciado primeiro pela organização, considere os seguintes exemplos para ter uma ideia melhor do que o Big Data pode oferecer para a sua empresa.

## Financeiro

- Análise de risco e portfólio
- Recomendações de investimento

## Varejo

- Envolvimento de clientes proativos
- Serviços baseados em localização

## Mídia

- Rastreamento de comportamento em jogo
- Opções de vendas cruzadas e adicionais

## Fabricação

- Programas de veículos conectados
- Manutenção preditiva

## Assistência médica

- Previsões de resultados de pacientes
- Custo total do tratamento
- Descoberta de medicamentos

## Setor público

- Seguro saúde
- Intercâmbios
- Otimização tributária
- Detecção de fraudes



# Projetos táticos de Big Data: Alguns exemplos

## O que alguns de nossos clientes buscam

Confira o modo como alguns de nossos clientes descrevem seus esforços específicos. Este é o tipo de foco que você deve buscar.

- Uma grande empresa de tecnologia no Vale do Silício tem como objetivo economizar mais de US\$ 10 milhões em custos do data warehouse que não para de crescer utilizando uma combinação de Hadoop e uma tecnologia tradicional de data warehouse para diminuir o crescimento do custo total por terabyte.
- Um grande fabricante na área de transportes está buscando reduzir a taxa de consumo de combustível de seus veículos em 1% ao longo dos próximos 10 anos. Ele também pretende reduzir as emissões tóxicas de carbono aumentando os períodos de manutenção em 10% e melhorando a quilometragem em 1%.
- Um fabricante envolvido com locomotivas pretende liberar uma milha adicional por hora em rotas diárias para que seus clientes possam economizar até US\$ 200 milhões por ano.
- Uma empresa global de serviços de pagamentos está tentando aumentar os negócios digitais em 30% através do aumento da personalização do cliente, tudo parte de uma estratégia de Big Data denominada “otimização omni-channel de varejo”.

Essas são algumas grandes vitórias para qualquer equipe de Big Data.



# A jornada básica do Big Data

Se você estiver pronto para construir as bases para uma abordagem de Big Data para toda a empresa, os três passos seguintes serão essenciais para a sua jornada.

Na verdade, mesmo que esteja planejando apenas alguns projetos táticos de Big Data, é necessário considerar estes três passos. Cada um é essencial para a integridade fundamental da organização centrada em dados. Na verdade, para obter o máximo custo/benefício, siga estas etapas na ordem.

1

## Otimização do data warehouse

Isso implica optar pelo armazenamento e processamento de dados na plataforma mais econômica. Muitas vezes, isso começa com a retirada de dados brutos ou pouco usados e de cargas de trabalho de ETL do hardware caro de data warehouse.

O objetivo é evitar atualizações onerosas do seu data warehouse e começar a usar hardware e estruturas de computação distribuída mais econômicos, como o Hadoop, para que você esteja preparado para lidar com o volume, a variedade e a velocidade do Big Data.



# A jornada básica do Big Data

2

## Um lago de dados gerenciados

Um lago de dados gerenciados é um local único para gerenciar a oferta e a demanda de todos os seus dados. A palavra-chave aqui é “gerenciar”. O objetivo é converter a confusão multiestruturada em informações confiáveis e seguras, adequadas à sua finalidade.

Isso significa criar um lago de dados que refine, governe e domine os dados. Porém, é preciso ter uma visão antecipada muito boa para chegar lá, pois é necessário incorporar políticas de governança de dados e processos estratégicos rigorosos. Sem eles, o lago corre o risco de se transformar basicamente em um pântano de dados.

3

## Inteligência operacional em tempo real

Aqui, você cria as tecnologias (interfaces de envolvimento, aplicativos que consomem muitos dados e análises) necessárias para acessar, analisar e entregar todos os dados. Os aplicativos criados aqui devem ser fáceis de usar e entregar as informações que os usuários precisam.

Pode ser uma interface para os representantes de atendimento ao cliente que monitoram o comportamento dos clientes através de vários canais e identificam aqueles com maior probabilidade de rotatividade nas próximas duas semanas.





**Escolha do  
projeto certo**

# Uma jornada de três passos

Como já dissemos, para obter a melhor relação custo/benefício, recomendamos seguir três passos, na ordem apresentada.

## Otimização do data warehouse

Reduza os custos de infraestrutura e reforce a arquitetura da empresa.

## Lago de dados gerenciados

Crie um único local para gerenciar a demanda e a oferta de dados.

## Análise em tempo real

Entregue aplicativos avançados que forneçam as informações de que as pessoas precisam.



# A jornada básica do Big Data

## Como nossos clientes definem seus objetivos fundamentais

Até mesmo projetos básicos devem ser específicos sobre o que estão tentando construir. Embora a especificidade aqui talvez não seja relacionada aos custos e ao tempo economizados, ela se aplica aos limites do que está sendo exatamente construído. Considere os seguintes exemplos de projetos de infraestrutura de Big Data dos nossos clientes.

- Uma organização global que conduz centenas de milhões de transações financeiras em centenas de países construiu um hub de dados de toda a empresa. O objetivo é fazer a análise do Big Data e identificar as principais macrotendências e padrões na interação com os clientes.
- Uma grande empresa de tecnologia construiu uma análise em nuvem de toda a empresa para reduzir o tempo de introdução no mercado de produtos orientados por dados, incluindo novos conjuntos de dados em análises a serem usadas nas unidades de negócio.
- Uma organização de consultoria financeira global criou uma infraestrutura lógica de data warehouse para garantir a disponibilização de informações consistentes em todas as plataformas padrão (incluindo Hadoop, bancos de dados operacionais e data warehouses tradicionais) que estão sendo usadas pela organização.

**Em resumo: Grandes jogadas causam grandes impactos, mas exigem as bases certas.**



# Parte B: Sua estratégia

Agora vamos ser práticos e analisar os requisitos específicos para o seu próximo (ou primeiro) projeto de Big Data.

[Voltar para Conteúdo](#)







# Defina as suas metas

Pegue o lápis. Como já dissemos, a causa número um de falhas em projetos de Big Data é a falta de objetivos claros. Agora, precisamos garantir que o projeto que você tem em mente não se perca em ambiguidades.

598  
55mph

276m  
70mph

101m  
75mph

501m  
69mph

411m  
67mph

136m  
72mph





# As metas de negócios

Vamos começar com os negócios, porque esses objetivos deverão ter precedência sobre as metas de TI se você quiser total apoio para o seu projeto.

Seja o mais específico possível ao estabelecer as metas de negócios a serem alcançadas pelo projeto. E lembre-se de estabelecer metas cujo impacto seja mensurável.

Por exemplo, no caso da interface de serviço ao cliente que prevê a rotatividade de clientes, as metas para esse projeto não devem ser listadas de forma vaga, como, por exemplo, “melhorar a experiência do cliente”.

Quanto mais claras forem suas metas, maiores serão suas chances de atingi-las. Uma meta superfocada vale mais que cinco metas vagas.



# As metas de negócios

**Liste, em ordem de importância, as metas do seu projeto de Big Data que digam respeito aos negócios e aos usuários de negócios. Fique à vontade para inserir menos ou mais metas.**

Por exemplo, reduzir a rotatividade de clientes

---

---

---

---

---

**Defina intervalos de tempo mínimo e máximo para cada meta a ser alcançada.**

Por exemplo, de três a seis meses

---

---

---

---

---

**Agora, para cada meta, escreva uma medida de sucesso que possa ser usada para determinar se a meta foi alcançada. Idealmente, essas medidas devem ser métricas disponíveis ou seus respectivos cálculos.**

Por exemplo, reduzir a taxa de rotatividade média mensal em X%

---

---

---

---

---

## Qual deve ser a duração do seu projeto de Big Data?

Seu projeto de Big Data deve demorar o tempo que for necessário para que seja plenamente efetivo. Em nossa experiência, o âmbito do projeto determina o horizonte de tempo.

Nós trabalhamos com clientes que entregaram projetos táticos em menos de três meses. E trabalhamos com clientes que passaram três anos produzindo programas básicos.

Para projetos mais longos, lembre-se de ter como objetivo demonstrar o valor do projeto a cada seis meses. Se adotar uma abordagem ágil, é recomendável mostrar as diferentes fases e marcos como projetos menores.

Uma coisa é certa: você não deve tentar adivinhar esse prazo. Estime o tempo de entrega com base em sua experiência e na experiência de outras pessoas que tenham realizado anteriormente projetos semelhantes. Se não souber para quem pedir orientação, você sempre poderá entrar em contato conosco.





Defina as suas  
metas

# Metas de TI

Agora vamos analisar as metas de TI relacionadas ao seu projeto.

(Lembre-se: se o objetivo do seu projeto for melhorar as condições de trabalho ou agilizá-lo, será bem difícil sensibilizar os usuários de negócios. Por isso, as metas de TI devem ser comunicadas junto com as metas que os usuários de negócios estão esperando.)

**Liste, em ordem de importância, as metas do seu projeto de Big Data que digam respeito à TI. Fique à vontade para inserir menos ou mais metas.**

Por exemplo, estabelecer processos para coleta, limpeza, criação de dados mestres e armazenamento de dados de clientes globais em tempo real, dados de utilização de cartões de crédito, dados de gráfico sociais e indicadores de taxa de rotatividade

---

---

---

---

---

## Pare, colabore e escute

Escrevemos este livro para que você possa iniciar seu projeto de Big Data, não importando se você trabalha na área de negócios ou em TI. Em ambos os casos, não use adivinhações para suas metas. Se precisar de orientações específicas sobre quais metas escolher, procure um parceiro com a experiência necessária e comece a trabalhar em colaboração agora.

O sucesso do seu projeto depende de colaborações estratégicas.



# Metas de TI

**Defina intervalos de tempo mínimo e máximo para cada meta a ser alcançada.**

Por exemplo, de dois a quatro meses

---

---

---

---

---

**Agora, para cada meta, escreva uma medida de sucesso que possa ser usada para determinar se a meta foi atingida. Idealmente, essas medidas devem ser métricas disponíveis ou seus respectivos cálculos.**

Por exemplo, a taxa de previsão de rotatividade exata de X%

---

---

---

---

---



# Defina as necessidades de dados

Agora que já definimos as metas específicas dos seus esforços de Big Data, vamos direto ao núcleo do projeto: os dados em si. Seja qual for o projeto, você terá que pensar de maneira estratégica sobre quais informações são necessárias, quais conjuntos de dados atendem a essa necessidade, como obter esses dados e como usá-los.



# Quais são os dados que você precisa?

Primeiro, vamos analisar a finalidade mais básica do seu projeto de Big Data: as informações que você está tentando fornecer à organização. Responda às seguintes perguntas o mais especificamente possível.

**Para atingir as metas de negócios descritas anteriormente, o que os usuários de negócios precisam saber para tomar uma decisão fundamentada?**

Por exemplo, quais dos clientes mais valiosos são suscetíveis à rotatividade e quais comportamentos se correlacionam à rotatividade

---

---

---

---

---

**Para proporcionar esse conhecimento, quais dados podem ser usados?**

Por exemplo, o histórico de compras do cliente, os dados de revisão, a taxa de compras, a taxa de desistência, a taxa de rejeição, a qualidade do atendimento ao cliente

---

---

---

---

---



# Quais são os dados que você precisa?

## Quais sistemas de origem contêm esses conjuntos de dados?

Por exemplo, registros de atendimento ao cliente,  
métricas de desempenho do produto, banco de  
dados de atividade do cliente, gerenciamento de  
dados mestres de clientes

## Além dos dados já mencionados, há outras informações que podem proporcionar contexto ou valor adicional às suas análises?

Por exemplo, dados de pesquisa de atendimento  
ao cliente, análise de concorrentes, dados  
meteorológicos, dados sociais

---

---

---

---

---

---

---

---

---

---



# Quais são os dados que você precisa?

**Entre os conjuntos de dados que não posso acessar atualmente, quais deles podem conter dados contextuais adicionais?**

Por exemplo, dados sociais de terceiros, dados de mercado de terceiros, dados meteorológicos

---

---

---

---

---

## A busca por dados obscuros

Ao considerar os conjuntos de dados aos quais você não tem acesso, não se limite aos dados externos à sua organização. O Gartner descobriu que a maioria das empresas usa apenas 15% dos seus dados<sup>2</sup>. A Appfluent, uma empresa que faz análise estatística sobre a utilização de data warehouse, descobriu que entre 30% a 70% dos dados de um data warehouse estão inativos.

O restante fica escondido em silos de difícil alcance e localização, cujos custos de utilização são altos, em arquivos legados e em repositórios de dados. Isso não seria um problema se você já não estivesse pagando para armazenar todos esses dados.

Ao procurar os dados de que precisa, vale a pena começar por aqueles que a sua organização já tem.





# Cinco considerações essenciais em relação aos dados

Após esboçar os dados que vai procurar, você terá uma visão mais clara dos desafios específicos do Big Data. Em especial, há cinco elementos-chave que devem ser considerados antes de avançar muito, pois eles determinarão o que precisará ser feito para cada conjunto de dados, bem como para o conjunto de dados de Big Data.

1

## Prepare-se para o volume

Você terá que se preparar para lidar com a “enormidade” dos dados que serão necessários. Usando dimensões, classifique os dados com base nos seus respectivos valores (por exemplo, transações de clientes), uso (frequência de acesso), tamanho (gigabytes, terabytes), complexidade (dados de máquinas, dados relacionais, vídeo etc.) e quem tem permissão para acessá-los (apenas seus especialistas de dados ou qualquer usuário ocasional de negócios).

Um inventário completo e organizado dos seus dados ajudará a determinar como gerenciar tudo isso. Avalie sua capacidade atual de armazenamento e processamento e procure as formas mais econômicas e eficientes de torná-la dimensionável.



# Cinco considerações essenciais em relação aos dados

2

## Responsabilize-se pela variedade

O aspecto mais desafiador do Big Data é a grande variedade de formatos e estruturas que você terá que conciliar em suas análises. Será preciso integrar várias fontes se quiser incluir novos tipos e estruturas de dados (sociais, sensores, vídeo) às fontes com as quais você já está acostumado (relacionais, mainframes legados).

Tentar codificar manualmente cada integração é tão complicado que poderia consumir todo o tempo e os recursos disponíveis. Aproveite ao máximo as ferramentas de integração e qualidade de dados disponíveis para acelerar o processo para tarefas mais importantes.

3

## Administre a velocidade

A combinação dos dados de streaming em tempo real e dos seus dados históricos geralmente aumenta a capacidade de previsão da análise. Assim, alguns dos dados que você deseja só poderão ser úteis se estiverem constantemente fluindo nos sistemas.

Na verdade, a maioria das análises em tempo real precisa se basear em dados de streaming – muitas vezes, a partir de diferentes fontes, em diferentes formatos. Prepare seu projeto com uma tecnologia analítica de streaming e uma infraestrutura lógica para gerenciar todos os dados.



# Cinco considerações essenciais em relação aos dados

4

## Garanta a veracidade

Não importa o grau de importância de suas análises, elas não valerão nada se as pessoas não puderem confiar razoavelmente nos dados que conseguirem. Quanto mais dados você analisar, mais importante será manter um alto nível de qualidade dos dados.

Para que os dados sejam adequados à finalidade, você precisa conhecer essa finalidade. Se um especialista de dados estiver procurando padrões em dados agregados do cliente, a preparação necessária será mínima. Por outro lado, os dados de relatórios financeiros e da cadeia de suprimentos deverão ser altamente editados, limpos e certificados para obter precisão e conformidade.

Crie categorias baseadas na quantidade de preparação necessária, o que abrange desde dados brutos até um repositório de dados mestres altamente editado, contendo dados limpos, confiáveis e fidedignos.

5

## Pense na conformidade

Os vários conjuntos de dados com os quais você lidará virão com diferentes condições e requisitos de segurança. Para cada conjunto de dados, é preciso considerar o que será necessário para torná-los anônimos com base em políticas de segurança.

Quantidades imensas de dados irão se proliferar por toda a empresa em centenas de repositórios de dados. Entenda onde os dados confidenciais residem e lembre-se de protegê-los na fonte por meio de criptografia e, em seguida, controle quem tem acesso a eles.

Além do arquivamento seguro e inteligente de dados confidenciais, faça o mascaramento dos dados com regras predefinidas sempre que migrá-los ou inseri-los nos ambientes de desenvolvimento e teste.

Aplique essas cinco considerações a cada conjunto de dados que administrar e, assim, você estará preparado mais realistamente para o grande desafio relacionado aos dados.



# Parte C: Sua cadeia de suprimentos simples de Big Data

Os métodos tradicionais de Business Intelligence e data warehouse não dimensionáveis o suficiente para atender às necessidades das iniciativas de Big Data. Agora, veremos como você pode dimensionar a equipe, os processos e a infraestrutura.

[Voltar para Conteúdo](#)

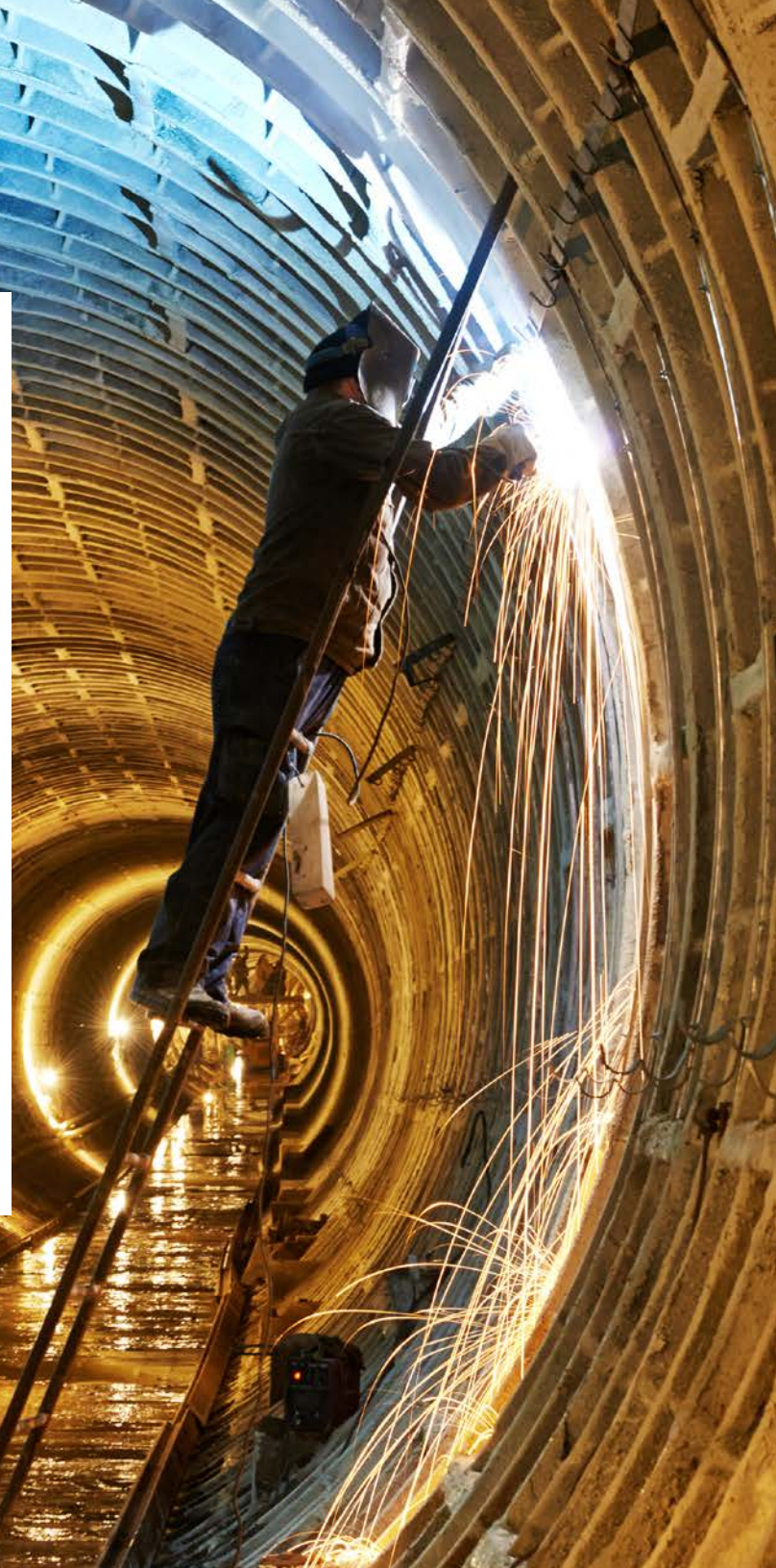






# Sua equipe

Sua equipe de Big Data representa o seu maior desafio e a sua maior oportunidade. Você precisa de um bom equilíbrio entre pessoas que entendem as metas de negócios e pessoas que podem executar suas exigências técnicas.







# Cinco lições essenciais para a criação de equipes

A maioria das organizações subestima o nível de habilidade necessário à aplicação bem-sucedida de uma nova tecnologia como o Hadoop.

As estruturas de dados distribuídos são muito difíceis de gerenciar. Das habilidades Java necessárias para desenvolver em Hadoop às novas habilidades em ciências de dados a serem contratadas, será necessário obter muitas novas habilidades para que o projeto realmente decole.<sup>3</sup>

Quando começar a montar sua equipe, lembre-se de incorporar as seguintes lições à estratégia de contratação.



# Cinco lições essenciais para a criação de equipes

1

## Use as habilidades dos profissionais que você contratou

Um dos maiores erros que as empresas cometem quando contratam especialistas de dados e analistas quantitativos é pedir que façam o trabalho pesado. Quando os seus recursos mais qualificados gastam todo o tempo em integrações de dados que exigem codificação manual e na limpeza de dados, você não só os deixa frustrados, como também não aproveita as habilidades que foram tão difíceis de encontrar.

Concentre as habilidades raras em tarefas que realmente precisam delas. Você não quer que os seus melhores profissionais peçam demissão e certamente não quer que eles percam tempo com um trabalho que pode muito bem ser feito com ferramentas.

2

## Pense estrategicamente sobre a composição da equipe

Se tudo der certo, o projeto crescerá em escopo e recursos. Pense de maneira estratégica agora e poupe-se da difícil conscientização de que você não pode executar determinados processos com a rapidez necessária porque há um número limitado de pessoas com as habilidades necessárias, até mesmo no Vale do Silício.

Se o seu projeto crescer em escopo, quais são as competências que você pode razoavelmente esperar encontrar a tempo de atender às suas necessidades? Por exemplo, especialistas de dados são infinitamente mais difíceis de encontrar, treinar e contratar do que desenvolvedores.<sup>4</sup>

O equilíbrio da sua equipe é fundamental. Você está procurando a combinação certa de experiência em gerenciamento de dados conquistada com trabalho duro e do entusiasmo para aprender novas ferramentas. Além disso, é preciso encontrar o equilíbrio entre pessoas com habilidades técnicas e aquelas com a experiência de domínio necessária para construir os modelos certos.

<sup>4</sup>"Big Data's High-Priests of Algorithms" (Big Data - Os bispos dos algoritmos), Wall Street Journal, 8 de agosto de 2014: <http://online.wsj.com/articles/academic-researchers-find-lucrative-work-as-big-data-scientists-1407543088>

# Cinco lições essenciais para a criação de equipes

3

## Alinhe as metas do seu projeto no início e depois as comunique

Um dos erros mais comuns que as empresas cometem ao contratar uma nova equipe é esquecer de comunicar as verdadeiras metas do projeto. Da primeira entrevista até o trabalho em si, é preciso ficar claro o que você está tentando oferecer aos usuários de negócios. Potencialize o apoio dos executivos para divulgar a missão e compartilhar histórias de sucesso, bem como problemas.

Sem um controle firme do valor comercial do seu projeto, os novos contratados correm o risco de achar que só precisam pensar nas metas de TI do projeto.

4

## Quando a sua equipe aumenta, a necessidade de gerenciá-la também aumenta

Diferentemente de uma nova tecnologia que pode ser implantada, implementada e integrada de forma objetiva, o novo pessoal precisa se acostumar ao local de trabalho, ao trabalho que estão fazendo e às razões por que estão fazendo esse trabalho. Não importa se é você ou outra pessoa: alguém precisa abraçar o desafio do gerenciamento que uma nova equipe exige.

Elementos como cultura e coerência não podem ser subestimados. Pense muito e com calma sobre como integrar as novas contratações em seus processos. Talvez você não consiga treiná-los para as habilidades, mas com certeza pode ajudá-los a serem melhores integrantes da equipe.





# Cinco lições essenciais para a criação de equipes

5

## Sua equipe não pode ficar parada

As tecnologias de Big Data estão surgindo todos os dias. E as que já existem estão evoluindo rapidamente. Este é um momento bastante emocionante para as empresas que são corajosas o suficiente para adotar as práticas recomendadas anteriormente. Mas representa também o desafio definitivo de começar antes dos concorrentes.

As habilidades da equipe precisam evoluir com a mesma velocidade das mudanças que estão ocorrendo no mundo. A boa notícia é que o que mais motiva os melhores profissionais é o desafio de chegar primeiro. O desafio está em oferecer o treinamento e o debate necessários para continuar aumentando as habilidades dessas pessoas e as suas próprias.

## A importância de ser estratégico

Uma escolha importante que você fará várias vezes é a possibilidade de criar seus recursos usando ferramentas automatizadas ou integrações manuais.

A codificação manual lhe oferece um controle completo e preciso sobre o que você está construindo. Muitas vezes, isso é inestimável e necessário; por exemplo, quando você escreve um script complexo para extrair metadados de uma forma que ainda não é possível.

As ferramentas, por outro lado, oferecem mais agilidade e a capacidade de repetir o mesmo processo de forma sustentável. Para tarefas como a integração e a qualidade dos dados, isso é essencial, porque significa que você não está forçando os seus analistas e especialistas superinteligentes para fazer o trabalho pesado.

Seja realista sobre os seus recursos. Se não pode montar uma equipe tão grande e genial como a do Google, não desperdice seus escassos recursos tentando.

# Implemente a governança de dados

Se (e quando) estiver montando um trabalho de Big Data mais básico, coloque em prática a estrutura processual para governança de dados. Na verdade, mesmo que o seu projeto de Big Data tenha como meta entregar valor para um único departamento, considere a criação de uma pequena equipe de governança de dados para poder aprender a lidar com os desafios exclusivos desse tipo de grupo.

Essencialmente, a equipe de governança de dados é o quadro formal de executivos responsável por supervisionar a abordagem de dados da empresa. Mas também deve incluir responsáveis pelo gerenciamento de dados – pessoas de departamentos específicos que tenham a tarefa de gerenciar os dados provenientes de uma unidade de negócio específica.

De fato, alguns de nossos clientes atribuem funções de responsabilidade pelo gerenciamento de dados com base no domínio de dados. Isso significa que uma pessoa fica responsável pelos dados de produtos, enquanto outra fica responsável pelos dados de clientes e assim por diante.



# Implemente a governança de dados

Tenha como meta criar processos capazes de garantir que a sua estrutura de governança de dados seja mais uma ajuda do que uma dificuldade. Trabalhe ativamente para assegurar que isso não se transformará em sobrecarga burocrática – para tanto, garanta que todos estejam empenhados em atingir os mesmos objetivos, e nos mesmos prazos.

Sua estrutura de governança de dados deve conter estas cinco características.

1

## Interfuncional

Um conselho de governança de dados formado por pessoas diferentes com funções semelhantes será ineficaz. O objetivo é criar um conselho que possa representar as opiniões e as necessidades exclusivas de cada unidade de negócios que o seu projeto de Big Data deverá apoiar.

2

## Comunicativa

Sem uma boa comunicação entre as funções, os departamentos e os domínios, o projeto pode se perder na burocracia e nos mal-entendidos. Isso acontece com mais frequência do que se imagina. Certifique-se de que todas as preocupações sejam resolvidas ou tratadas adequadamente.

# Implemente a governança de dados

3

## Eficiente

O seu processo interfuncional não deve ser uma barreira. Será preciso ter bastante agilidade para que o projeto de Big Data tenha êxito. Portanto, crie regras de relatórios de automação e exceção sempre que possível e adote ferramentas de colaboração para manter os canais de comunicação abertos e adequados.

4

## Comprometida

Lembre-se de comunicar com eficiência as metas principais do seu projeto e certifique-se de que todos os envolvidos na estrutura de governança de dados estejam comprometidos com essas metas. Metas em comum ajudam a guiar o seu pensamento de governança e as suas decisões.

5

## Centralizada

O maior desafio de uma estrutura de governança de dados é quando você precisa priorizar as metas de uma unidade de negócios sobre as outras que estão sendo representadas no conselho. Garanta que as suas decisões contemplem o benefício a longo prazo de todo o conselho, mesmo que isso signifique benefícios de curto prazo para uma única unidade de negócio.

# As habilidades que você tem e aquelas de que precisa

Hora de pegar o lápis novamente. Agora que você conhece as várias ciladas e oportunidades subjetivas que a nova equipe apresentará, vamos analisar como é essa equipe na realidade.

A página a seguir lista funções de Big Data com base em cargos para os quais já vimos nossos clientes contratarem. Considerando o pessoal realmente disponível e a duração esperada para o projeto (conforme inserido na seção iniciada na página 24), liste quantas pessoas você precisa contratar.



A função	Alguém já pode executar essa função?	Eu preciso contratar alguém para essa função	Com base na quantidade de tempo disponível, eu preciso contratar X pessoas
Especialista de dados	✓ ou ✕	✓ ou ✕	
Especialista em domínios			
Analista de negócios			
Analista de dados			
Engenheiro de dados			
Administrador de banco de dados			
Arquiteto corporativo			
Arquiteto de solução de negócios			
Arquiteto de dados			
Responsável pelo gerenciamento de dados			
Desenvolvedor de ETL (integração de dados)			
Desenvolvedor de aplicativos			
Desenvolvedor de painéis de controle			
Modelador de estatísticas			
Outro			
Outro			
Outro			
Outro			
Outro			

### A necessidade de um pensamento integrado

Quando sair em busca de novos integrantes para a equipe, não se limite a pessoas com as qualificações adequadas. Não se engane: encontrar pessoas com as qualificações adequadas já é um desafio em si. Mas você também precisa buscar pessoas que tenham vontade de sintetizar as metas de negócios e as capacidades técnicas.

Cada vez mais, ouvimos de clientes sobre como é importante que as pessoas que se unem aos seus projetos de Big Data possam ser capazes de entender a realidade dos negócios e trabalhar com uma ciência de dados complexos. Esse tipo de pensamento integrado é importantíssimo e difícil de encontrar. Vale a pena investir em treinamento.





# Suas ferramentas

Como já discutimos várias vezes, as ferramentas utilizadas desempenham uma função estratégica na execução do seu projeto de Big Data. Nesta seção, vamos analisar as ferramentas que você tem e aquelas de que precisa.



# Noções básicas sobre as ferramentas de Big Data

Em nossa experiência, as seguintes ferramentas são essenciais para a arquitetura necessária aos projetos de Big Data (discutiremos a arquitetura em mais detalhes posteriormente). Claro que as suas metas e os seus recursos devem determinar a combinação das tecnologias necessárias ao seu projeto específico.

**Veja esta lista de ferramentas e coloque um ✕ para as mais importantes e mais estrategicamente relevantes ao seu projeto específico.**

## Ingestão de dados

○ processo de consumir os dados necessários de forma adequada, eficiente e metódica.

### ○ Carga de lotes

Você acessa todos os tipos de dados de que precisa e dimensiona com eficiência o desempenho da carga de lotes nos repositórios de dados?

### ○ Captura de dados de mudanças

Você captura as alterações feitas nos dados dos sistemas de origem sem afetá-los?

### ○ Fluxo de dados

Você coleta dados de forma confiável em tempo real e os transmite nos repositórios de dados?

### ○ Arquivamento

Você arquiva e compacta dados que não são usados com frequência, garantindo fácil acesso aos dados arquivados, se necessário?



# Noções básicas sobre as ferramentas de Big Data

Veja esta lista de ferramentas e coloque um ✕ para as mais importantes e mais estrategicamente relevantes ao seu projeto específico.

## Gerenciamento de dados

Todas as políticas, processos e práticas necessárias para gerenciar adequadamente a eficácia, a precisão, a confiabilidade e a disponibilidade dos dados.

### ○ Integração de dados

Você consegue preparar e consolidar várias estruturas e fontes em um conjunto de dados consistente para análise?

### ○ Qualidade de dados

Você consegue, de forma confiável, limpar os dados, deduplicar e remover os erros?

### ○ Segurança de dados

Você descobre e protege os dados em todos os repositórios de dados, atribuindo para isso regras sobre uso, acesso e permissões?

### ○ Máquina virtual de dados

Você consegue criar uma camada de abstração para os dados que combine vagamente o processamento de dados no ambiente de implantação subjacente?

### ○ Gerenciamento de dados mestres

Você armazena uma versão consolidada, completa e fidedigna da verdade para vários domínios de dados?

### ○ Estrutura distribuída de dados

Você usa tecnologias como o Hadoop para dimensionar de forma econômica suas necessidades de armazenamento e processamento?

### ○ Data warehouse

Você tem uma tecnologia de data warehouse que possa lidar com os requisitos de desempenho, uso e dimensionamento para análises e integrações de Big Data com infraestruturas Hadoop?



# Noções básicas sobre as ferramentas de Big Data

Veja esta lista de ferramentas e coloque um ✕ para as mais importantes e mais estrategicamente relevantes ao seu projeto específico.

## Entrega de dados

Processo de envio dos dados necessários aos sistemas e aplicativos que precisam deles.

### ○ Carga de lotes

Você dimensiona com eficiência a carga de lotes de dados entre os sistemas de origem, análise e back-end operacionais?

### ○ Fluxo de dados em tempo real

Você oferece fluxo de dados em tempo real para sistemas de aplicativos, análises e back-end que precisam disso?

### ○ Hub de integração de dados

Você disponibiliza dados usando uma abordagem como o modelo de publicação e assinatura para evitar a proliferação de integrações ponto a ponto?

### ○ Visualização de dados

Você entrega dados dos seus sistemas sem sobrecarregá-los?

### ○ Processamento baseado em eventos

Você é capaz de detectar, analisar e reagir a ameaças, oportunidades e outros eventos críticos de negócios em tempo real?



# Noções básicas sobre as ferramentas de Big Data

Veja esta lista de ferramentas e coloque um ☒ para as mais importantes e mais estrategicamente relevantes ao seu projeto específico.

## Análise

As ferramentas e os processos que transformam dados brutos em insights, padrões, previsões e cálculos sobre o domínio que você está analisando.

### ○ Visualização

Você pode apresentar dados e resultados de maneira a facilitar a compreensão e o entendimento?

### ○ Análise avançada

Você consegue aplicar algoritmos analíticos avançados aos seus conjuntos de dados para fazer cálculos complexos?

### ○ Aprendizado de máquina

Você consegue aplicar sofisticados algoritmos de aprendizado de máquina para identificar padrões e fazer previsões em um nível para o qual não tem a largura de banda manual compatível?

Entre essas ferramentas e tecnologias, algumas ferramentas – como a integração de dados, a qualidade de dados e o gerenciamento de dados mestres – são tão essenciais à sua jornada de Big Data que realmente não vale a pena recriá-las. O prazo e os recursos necessários à criação dessas ferramentas com suas próprias mãos não vale as preciosas habilidades e horas de trabalho do seu projeto de Big Data.

Lembre-se das metas do seu projeto; elas não incluem criar tudo de maneira personalizada.





# Seus processos

Vamos nos aprofundar nos processos reais de que você precisará para lidar com o Big Data. Seus processos específicos serão exclusivos para suas metas e necessidades, mas esta seção pode dar uma visão geral sobre o que esperar e o que você aprenderá.





# Os oito Big Data

Por experiência, podemos dizer que as metodologias ágeis são uma excelente abordagem para projetos de Big Data. Elas garantem que você gerencie as expectativas, aprenda com os erros e faça a iteração do caminho para os melhores processos. Mesmo assim, a abordagem do projeto depende inteiramente de você e da sua situação.

De qualquer maneira, os oito passos a seguir se mostrarão essenciais à cadeia de suprimentos do Big Data. Não importa como você cuide disso. Você e a sua equipe devem estabelecer processos eficazes para estas etapas.

1

## Acessar os dados

Seu primeiro desafio será adquirir todos os dados de que precisa. Em alguns casos, isso implicará a captura de dados de streaming e, em outros, significará extrair dados de um banco de dados. Configure processos gerenciáveis e que possam ser repetidos para garantir o armazenamento desses processos de acordo com o uso que fará deles.

2

## Integrar os dados

O desafio mais complexo do Big Data envolve a variedade de estruturas e formatos de dados. Para que a sua análise seja conduzida de forma sustentável, você precisará configurar um processo de integração e normalização de todos esses dados. Idealmente, isso deve usar um processamento que seja o menos manual possível.

# Os oito Big Data

3

## Limpar os dados

Para que a sua análise seja confiável, é preciso garantir a limpeza dos dados a fim de eliminar duplicações, erros, imprecisões e dados incompletos. O processo deve garantir que analistas e especialistas mais qualificados não fiquem o tempo todo fazendo o trabalho errado.

4

## Criar dados mestres

Uma forma de manter uma fonte confiável de dados limpos e integrados é estabelecer um processo de criação de dados mestres. O objetivo é criar uma coleção avançada de dados consolidados, organizada por domínio (como produtos, clientes etc.) e repleta de insights de Big Data, e que, assim, possa alimentar todos os outros sistemas.

5

## Proteger os dados

Aqui, você estabelecerá dois processos básicos. O primeiro será sobre como definir as regras e as práticas de segurança que cada conjunto de dados exige. O segundo será sobre como detectar dados confidenciais e mascará-los de forma persistente ou dinâmica para garantir que essas regras e práticas recomendadas sejam aplicadas de maneira consistente.



# Os oito Big Data

6

## Analisar os dados

O processo de análise dependerá dos analistas, das ferramentas de análise e das necessidades relacionadas às metas. A mentalidade de descoberta iterativa e melhoria contínua desempenhará um papel crucial nesse caso, pois esse processo deverá ficar melhor, mais rápido, mais barato e mais dimensionável ao longo do tempo e com uma experiência maior.

7

## Analisar as necessidades de negócios

Esta etapa é crítica, mas quase sempre esquecida. Configure um processo claro para a análise das necessidades dos negócios, mesmo quando estiver analisando os dados. Isso é muito importante porque, caso você desacelere o ritmos dos negócios, poderá isolar seus esforços e minimizar o impacto nos próprios negócios.

8

## Operacionalizar o insight

Como já discutimos no início deste livro, o impacto do seu projeto de Big Data nos negócios precisa ser sentido. Crie pipelines automatizados para as respostas que você encontrar e os entregue aos usuários de negócios que mais precisam deles. Por exemplo, os dados sobre os clientes com maior probabilidade de rotatividade devem ser disponibilizados aos agentes de atendimento ao cliente por meio de um painel de controle. Lembre-se de incorporar um ciclo de feedback também, assim poderá ver como o insight é recebido.

## A importância da documentação

Tenha como meta dominar esses oito passos e o seu projeto de Big Data avançará na direção certa. O objetivo é estabelecer processos claros, repetíveis, dimensionáveis e com melhorias contínuas. Para isso, a documentação desses processos e as melhorias contínuas são vitais para a equipe.

As habilidades, os recursos e as lições do seu projeto de Big Data devem ser transferíveis e comunicadas com frequência.





# Sua arquitetura

Para que a cadeia de suprimentos de Big Data seja enxuta e eficaz, é preciso garantir que a arquitetura seja sólida e construída de maneira estratégica. Nesta seção, vamos discutir como deve ser uma arquitetura de Big Data ideal e como implantá-la a sua em uma abordagem em fases.





# Primeiros passos: Sua área restrita

Para começar a criar a arquitetura do seu projeto de Big Data, o ponto de partida mais lógico é a criação de um ambiente de desenvolvimento em uma área restrita, na qual você possa usar dados de teste para garantir a viabilidade da arquitetura. Ao fazer isso, lembre-se de levar em consideração as seguintes lições.

## Comece pequeno

Começando com uma área restrita bem definida sobre a qual tenha total controle, você conseguirá fazer a interação do caminho para uma implementação mais bem-sucedida. Comece a trabalhar o mais rapidamente possível e a documentar as lições aprendidas em cada iteração.

## Tamanho é importante

A principal diferença entre a área restrita e a implementação real é o ambiente de produção, que será muito maior. Isso exigirá um tratamento automatizado para ingerir, integrar, limpar e distribuir a saída. Portanto, será preciso ter uma estrutura muito mais robusta, além de componentes e processos comprovada e verdadeiramente confiáveis e flexíveis em um ambiente de produção dinâmico.

# Primeiros passos: Sua área restrita

## Faça o mascaramento antes de testar

Quando as organizações usam dados de teste, elas geralmente usam uma variante dos dados de produção dinâmicos para garantir que os formatos e as estruturas representem o ambiente dinâmico. Infelizmente, se esses dados não forem mascarados de forma adequada, os dados confidenciais poderão ficar expostos em um ambiente de teste totalmente inseguro.

## Não se perca na tradução

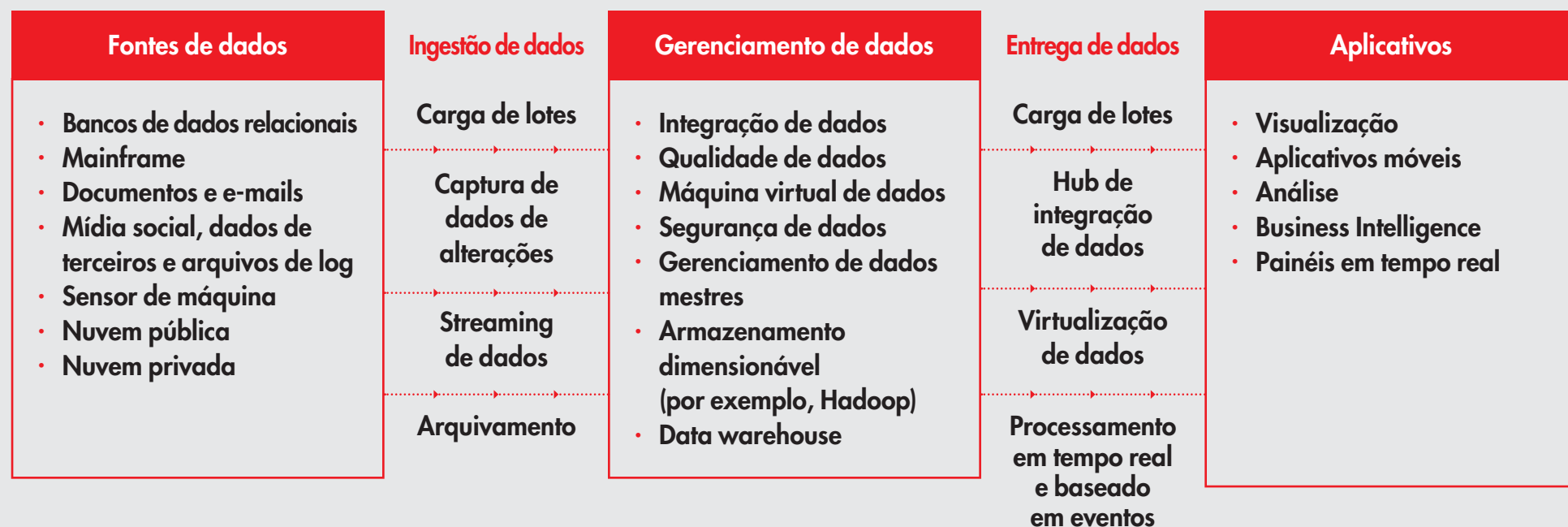
Uma das fontes mais comuns de atrasos e desvios orçamentários onerosos dos projetos de Big Data decorre do fato de que os erros de codificação manual (que se perderam na área restrita) voltam a assombrar a equipe quando a arquitetura é ativada. Assim, se você codificar manualmente partes significativas da arquitetura, espere o refatoramento de vários códigos para atender aos requisitos no nível de produção e administre as expectativas adequadamente. Como alternativa, use ferramentas de produtividade e automação para evitar a necessidade de refatorar o código e os erros.





# A arquitetura ideal do Big Data

O diagrama a seguir representa nossa recomendação para criar a arquitetura ideal de tecnologia e processos de Big Data.







# Seu plano de projeto

Agora, nós já analisamos todos os aspectos da sua jornada de Big Data. O próximo passo é usar este plano de projeto como um guia estrutural para gerenciar um projeto de Big Data, do lançamento à implementação.



# Seu plano de projeto

Use este modelo de plano de projeto como uma estrutura para documentar os detalhes e os diferentes elementos do seu projeto de Big Data. Depois, use o documento compilado como uma maneira de obter a adesão necessária do restante da sua organização. Ele também será útil quando você abordar parceiros externos.

## Estágio 1: A estratégia

**Identificar as metas de TI e de negócios**

**Definir as medidas de sucesso**

## Estágio 2: Os dados

**Identificar as informações necessárias**

**Identificar os dados e as origens para entregá-los**





# Seu plano de projeto

## Estágio 3: A cadeia de suprimentos

### As pessoas

- Avaliar as habilidades necessárias
- Avaliar as habilidades existentes

### O processo

- Acessar os dados
- Integrar os dados
- Limpar os dados
- Criar dados mestres
- Proteger os dados
- Analisar os dados
- Analisar as necessidades de negócios

### As ferramentas

- Computação distribuída (por exemplo, Hadoop)
- Qualidade de dados
- Integração de dados
- Gerenciamento de dados mestres
- Mascaramento de dados
- Visualização
- Análise de streaming
- Análise
- Aprendizado de máquina

## Estágio 4: Operacionalizar o insight

### Desenvolver painéis

### Automatizar processos para entrega de dados

### Definir um processo de feedback

# Seguindo em frente

Use as listas de verificação, os princípios e as diretrizes que descrevemos neste livro para trazer o potencial do Big Data para a sua organização. Seja qual for o tamanho do seu projeto, agora temos certeza de que você estará melhor equipado para lidar com os vários desafios deste projeto.

Lembre-se: seja estratégico em relação aos seus recursos e mantenha um foco preciso no desenvolvimento de processos e habilidades que sejam transferíveis, dimensionáveis e estejam em constante aperfeiçoamento. Se você mantiver uma visão de longo prazo ao realizar esse projeto, poderá preparar a sua organização para análises aprimoradas e decisões mais fundamentadas por um longo tempo.

De muitas maneiras, seu primeiro projeto de Big Data será inesquecível. A partir dos erros que poderão ocorrer para criar uma equipe, você estará pronto para uma jornada de imenso valor estratégico para a sua empresa.

Ao navegar e evitar as muitas ciladas que discutimos, e se mantiver um forte compromisso com a sua visão desse projeto, você será capaz de mudar a forma como a organização funciona.

**Isso vai ser o máximo!**

# Próximos passos

Pronto para começar a colocar em prática o que aprendeu?



Se você é um desenvolvedor da Informatica, poderá ser um desenvolvedor de Hadoop. Nossos testes de software, conectores e serviços de Big Data colocarão você no caminho certo.

# Sobre a Informatica

Ajudamos as empresas a gerenciar dados para que elas possam extrair deles um valor de negócios mensurável. E estamos ajudando algumas das maiores empresas do mundo a encontrar os erros mais comuns de gerenciamento de dados para que alcancem o sucesso em projetos repetíveis e dimensionáveis de Big Data.

**Vamos conversar.**

