

Cenário

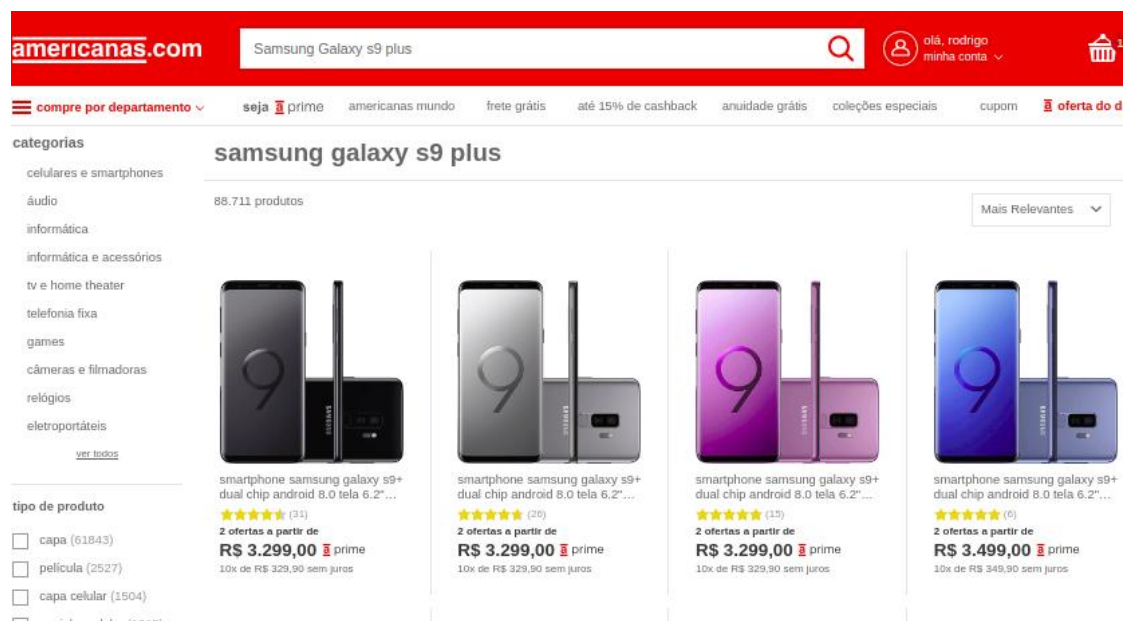
Fluxo de navegação nos e-commerces que fazem parte do grupo B2W.



1. Acesso ao site www.americanas.com.br (home)



2. Busca de produto (search)



3. Detalhes do produto (*product*)

americanas.com

tem tudo, pode procurar :)

Q

olá, rodrigo
minha conta

1

compre por departamento

seja prime

americanas mundo

frete grátis

até 15% de cashback

anuidade grátis

coleções especiais

cupom

oferta do dia

página inicial

>

celulares e smartphones

>

smartphone

>


samsung galaxy

>

galaxy s

>

galaxy s9 plus



Smartphone Samsung Galaxy S9+ Dual Chip Android 8.0 Tela 6.2" Octa-Core 2.8GHz 128GB 4G Câmera 12MP Dual Cam - Preto

(Cód.133274769) ★★★★★ (39)

Escolha uma loja abaixo e compre

americanas.com

☒

R\$ 3.299,00

Frete: -

Loja Samsung Oficial

☐

R\$ 4.399,00

Frete: -

vendido e entregue por americanas.com

R\$ 3.299,00 prime

10x de R\$ 329,90 s/ juros

comprar

adicionar à lista de casamento

R\$ 3.299,00 em até 12x de R\$ 274,91 s/ juros com Ame e receba R\$ 164,95 (5% de volta)

R\$ 2.969,10 em 1x no cartão americanas.com (10% de desconto) ou R\$ 3.299,00 em até 15x de R\$ 219,93 s/ juros

[formas de parcelamento](#)


4. Clique em comprar e adição no carrinho (*basket*)

produto

qtd.

entrega

preço



prime

Smartphone Samsung Galaxy S9+ - Preto

vendido e entregue por: americanas.com

☐ seguro roubo e furto + 1 ano protegido - [saiba mais](#)

1

▼

até 2 dias úteis

R\$ 3.299,00

remover

calcule frete e prazo

ok

não sei meu cep

☐ prime a jato até 1 dia útil - grátis - 1 restante

o que é o prime?

☐ prime até 2 dias úteis - grátis - ilimitado

o que é o prime?

☐ a jato até 1 dia útil - R\$ 9,99

resumo do pedido

subtotal (1 produto) R\$ 3.299,00

frete econômica R\$ 4,99

total R\$ 3.303,99

em até 15x s/ juros

R\$ 2.974,09 em 1x no cartão americanas.com

R\$ 2.974,09 em 1x no cartão de crédito

pague R\$ 3.303,99 com Ame e receba R\$ 164,95 de volta

continuar

possui cupom ou vale? você poderá usá-los na etapa de pagamento.

5. Escolha da forma de pagamento (payment)

americanas.com
a maior loja. os menores preços.

Minha cesta

Identificação

Pagamento

Obrigada

1 produto [visualizar produtos](#)

R\$ 3.299,00

frete para SÃO PAULO

R\$ 4,99

subtotal

R\$ 3.303,99

total a pagar

R\$ 2.974,09

[em 1x no cartão americanas.com](#)

R\$ 2.677,18

[em 1x no cartão de crédito](#)

R\$ 2.677,18

[pague R\\$ 2.974,09 com Ame e receba R\\$ 148,46 de volta](#)

endereço de entrega

[alterar endereço de entrega](#)

Quer comprar com frete grátis e entrega rápida em mais de 1 milhão de produtos? Escolha a entrega Prime!

opções de entrega

☐ **prime a jato**
até 1 dia útil
GRÁTIS - 2 restantes

☐ **prime**
até 2 dias úteis
GRÁTIS - ilimitado

☐ **a jato**
até 1 dia útil
R\$ 9,99

☒ **econômica**
até 2 dias úteis
R\$ 4,99

ame
selecione

americanas.com
selecione

cartão de crédito
selecione

boleto
selecione

pague na lotérica
selecione

débito
selecione

pague na loja
selecione

! **IMPORTANTE:** só emitimos boletos do Banco do Brasil. Confira os dados antes de pagá-lo.

Imprima o boleto e **pague no banco**

ou **pague pela internet** utilizando o código de barras do boleto

o prazo de validade do boleto é de **1 dia útil**

Total: R\$ 2.974,09

pagar com boleto

compra segura americanas.com

6. Confirmação da compra (*thankyou*)

americanas.com

baixe o app e receba
notificação sempre que
seu pedido for atualizado



Olá, rodrigo

Tudo certo com o pagamento!
Já estamos preparando a **entrega 01**, referente ao pedido **02-689104224**, para envio.

Entrega 01	americanas.com Vendido e entregue por Americanas.com				
	<table><tr><th>Qtd</th><th>Entrega</th></tr><tr><td>1</td><td>Até 15/02/2019</td></tr></table>	Qtd	Entrega	1	Até 15/02/2019
Qtd	Entrega				
1	Até 15/02/2019				

Características/detalhes do fluxo:

- as visitas/navegações nas páginas geram eventos (produtos consultados, produtos visitados, cliques em comprar, remover do carrinho, etc.) que são chamados de dados navegacionais;
- nas páginas **home**, **search** e **product** o usuário **não** precisa estar logado (identificado);
- nas páginas **basket**, **payment** e **thankyou** o usuário deve estar logado;
- as únicas páginas que o usuário consegue acessar diretamente (através da URL) são a **home**, a **search** e a **product**. Todas as outras são acessíveis somente através do fluxo navegacional (transição entre as telas);
- todos os eventos são diferenciados pelo nome de cada página, possuem informações de data e hora, dispositivo que acessou a página, sessão de usuário (caso o usuário esteja logado), entre outras informações (referentes a cada página);
- a página de **basket** deve ter pelo menos um produto (poderá ser adicionado mais de um produto);
- por motivos técnicos (falha na geração dos eventos, instabilidade do ambiente, volume de informações, etc.), algumas páginas podem não ter seus eventos gerados (**home**, **search** e **product**);
- as páginas **basket**, **payment** e **thankyou** sempre terão seus eventos gerados;
- a página **thankyou** é a ÚNICA confirmação de que a compra foi finalizada com sucesso.

O cenário acima foi detalhado em um fluxo de sucesso, ou seja, navegação entre todas as páginas, finalizando com a conclusão de compra.

Sabemos que no cenário real poucas visitas seguem esse fluxo.

Alguns usuários acessam diretamente a página **product** e podem ou não concluir a compra.

Outros usuários acessam pela página **home** e ficam navegando entre as páginas de **search**, **product**, **basket** e podem ou não concluir a compra, ou seja, sabemos que existem algumas opções de fluxos alternativos. Com isso, sabemos que podem haver buracos nos dados navegacionais, impossibilitando análises que dependem desse fluxo, como por exemplo a taxa de conversão¹.

¹ taxa de conversão: indicador que demonstra a proporção de visitas que concluíram em vendas em um espaço de tempo

Anexo temos os dados navegacionais conforme cenários descritos abaixo.

cenario 01: dados navegacionais do exemplo citado, fluxo completo de navegação.

cenario 02: dados navegacionais de um usuário que acessa a página de produto diretamente, navega entre a página de busca e de produtos, adiciona o produto no carrinho (efetuando login no site) e efetua a compra.

cenario 03: dados navegacionais de um usuário que acessa a página home (deslogado), adiciona o produto no carrinho (efetuando login no site) e efetua a compra. Analisando os registros, podemos afirmar que os eventos de busca e navegação do produto não foram gerados.

cenario 04: dados navegacionais de um usuário que navega mais entre as páginas **home**, **product**, **basket** e **payment**, porém, sem a finalização da compra.

Informações sobre os campos dos dados navegacionais (* indica que sempre terão informações):

*load_timestamp:	data e hora da navegação
*device_type:	informa se a navegação foi feita através de um dispositivo móvel, ou desktop, etc.
*page_type:	tipo da página
visit_id:	identificador único de cada visita (gerado quando o usuário efetua login)
url_location:	endereço (URL) da página
search_query:	caso tenha sido feito uma busca indica o termo buscado
product_id:	código do produto
site_department_id:	código do departamento do produto
product_unit_price:	preço unitário do produto
freight_delivery_time:	tempo de entrega do frete calculado
freight_value:	custo de entrega do frete calculado
cart_qty:	quantidade de itens em cada linha do carrinho
cart_total_value:	valor total dos itens de cada linha do carrinho (sem considerar o valor do frete)

Do desafio:

Esperamos a proposta de um pipeline de ingestão dos dados navegacionais. Tais dados devem ser

- ✓ consumidos: lidos de um arquivo (para o desafio), mas pense em uma solução escalável, pois essa informação deve ser fornecida em tempo real via streaming em ambiente de produção
- ✓ analisados: entenda os dados, veja os possíveis buracos que podem existir (dado os detalhes do fluxo)
- ✓ organizados: cubra os buracos que existem, facilite a relação entre as informações, deixe os dados o mais completo possível
- ✓ estruturados: salve em um ambiente big data, que possa ser consultado de forma fácil e rápida
- ✓ agregados: faça os dados gerarem informações úteis, nos informe a taxa de conversão para cada origem do site (navegações que partiram da **home** ou **product**) e outras métricas que achar importante (crie uma estrutura para os resultados)

Pré requisitos:

1. Propor um pipeline em nuvem, preferencialmente nas plataformas AWS ou GCP;
2. A solução proposta deve ser **escalável**;
3. A arquitetura, tecnologias, ferramentas são livres, de escolha do candidato;
4. Implementar uma pequena aplicação simulando a ingestão de eventos em um banco de dados.

Entregáveis:

1. A proposta do pipeline deve ser uma documentação descrevendo arquitetura, tecnologias e ferramentas;
2. A simulação da ingestão dos dados pode ser implementada em Java ou Python com os eventos sendo lidos do arquivo dados_navegacionais anexado e o resultado sendo escrito em um arquivo csv (não precisa subir um banco de dados).

Da solução:

1. Explique o porquê do modelo de arquitetura e de cada tecnologia/ferramenta adotada;
2. Descreva como a solução se comporta diante de grandes variações no volume de dados;
3. Inclua um arquivo README para sua implementação explicando claramente como instalar as dependências necessárias (se houver) e como executá-la;
4. A simulação não precisa seguir exatamente a proposta do pipeline, mas deve ser possível utilizá-la nele, com possíveis adaptações.