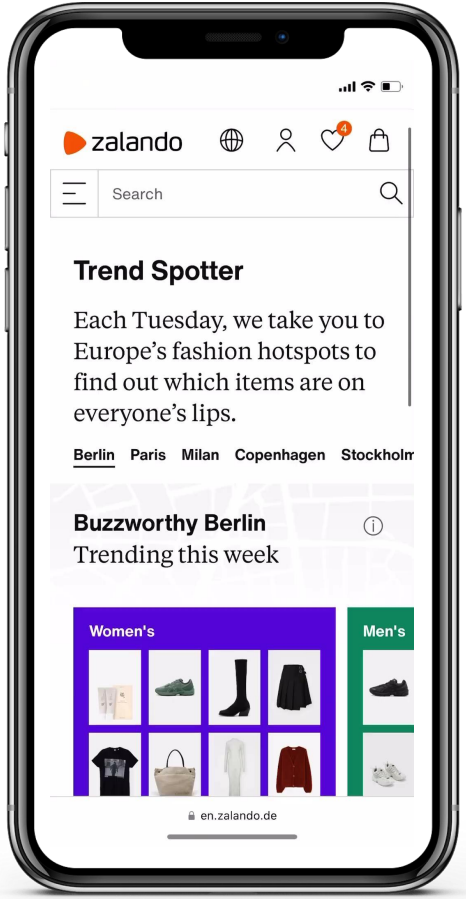
Enhancing Fashion E-Commerce with Cutting-Edge AI, Powered by AWS

Ravi Sharma, Mones Raslan, Weiwei Cheng








long trench coat

Product Details

Material & Care

Outer fabric material: 97% cotton, 3% elastane
Padding type: No lining
Care instructions: Hand wash only, Dry cleanable

Details

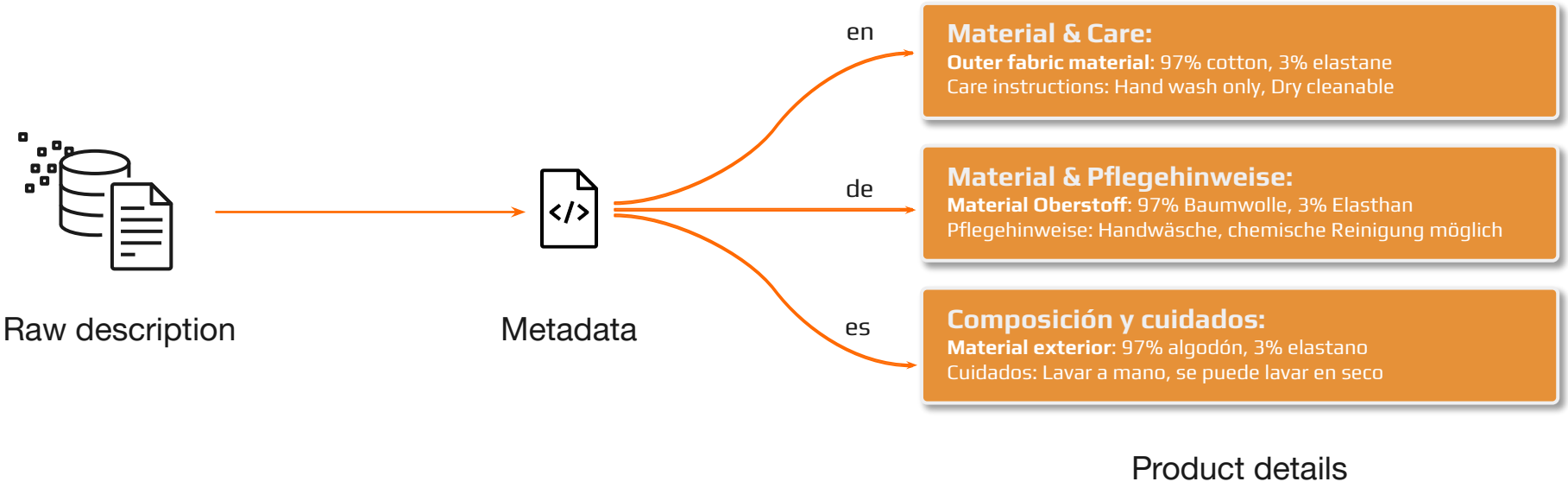
Collar: Lapel collar
Fastening: Button
Pockets: Inseam pockets

Size & Fit

Fit: Regular fit
Shape: Fitted
Sleeve length: 84.45 cm
Total length: Size 40

Detailed product information allow customers to make the right purchase decisions

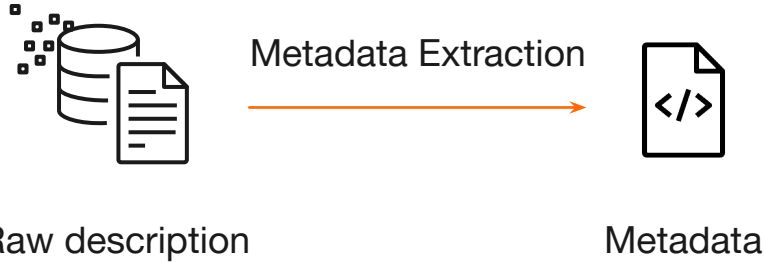
Metadata


en
de
es
Material & Care:
Outer fabric material: 97% cotton, 3% elastane
Care instructions: Hand wash only, Dry cleanable
Material & Pflegehinweise:
Material Oberstoff: 97% Baumwolle, 3% Elasthan
Pflegehinweise: Handwäsche, chemische Reinigung möglich
Composición y cuidados:
Material exterior: 97% algodón, 3% elastano
Cuidados: Lavar a mano, se puede lavar en seco
Raw description
Metadata
Product details

Challenges

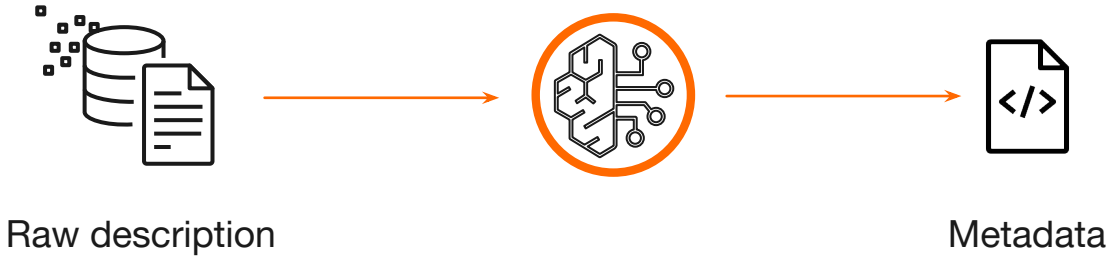
Product descriptions come from different data sources

- The descriptions use various languages
- The data structure is unique to every data provider
- The naming conventions vary from one manufacturer to the other



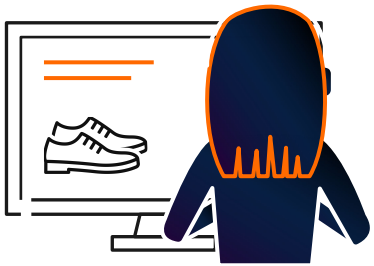
Metadata extraction was partially done manually

Better Metadata with GenAI

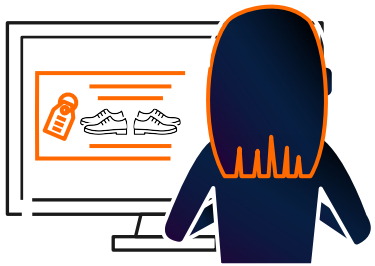


Use GenAI to Automate

Impact on Search



Detailed metadata leads to more accurate search results



Easier to find → Easier to buy

Improved overall shopping experience

Impact on Personalization

Size & Fit

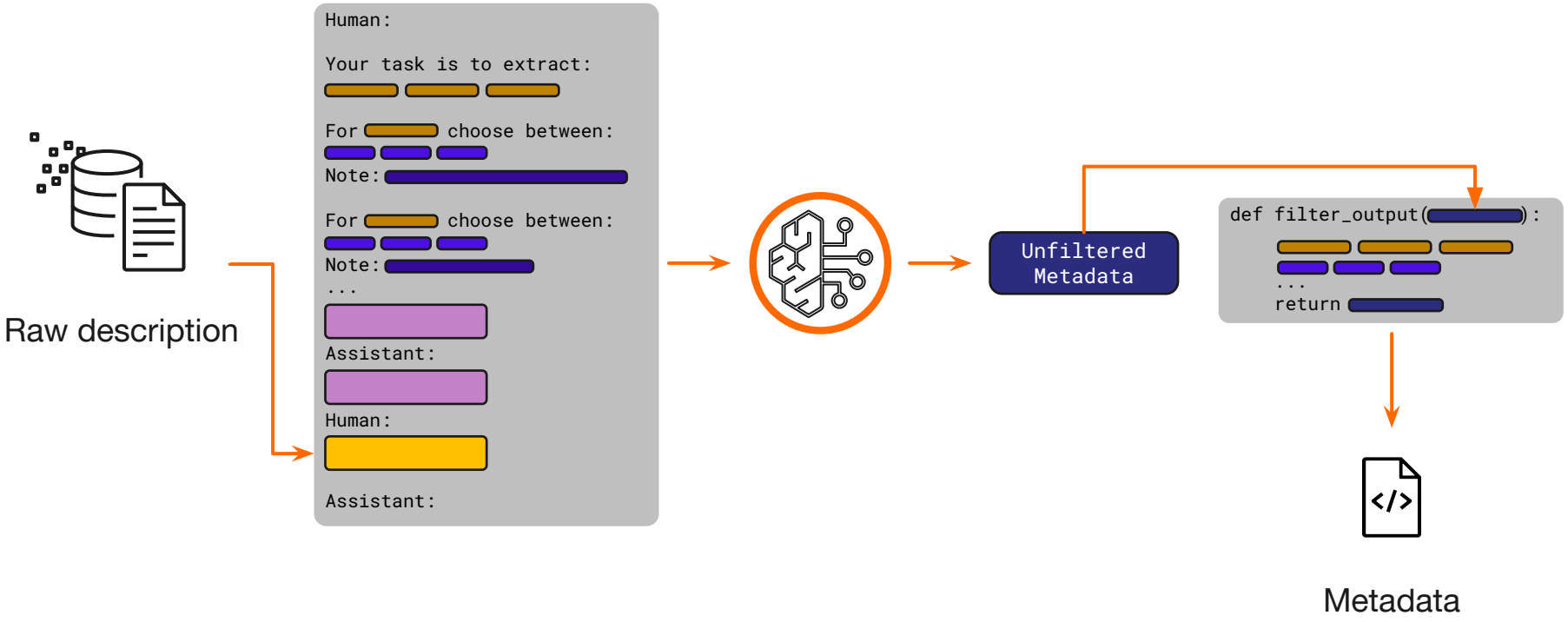
Model's height: Our model is 6'1" tall and is wearing size M
Fit: Regular fit
Shape: Straight
Sleeve length: 25.5" (Size M)
Total length: 27.0" (Size M)

Accurate metadata will improve personalization

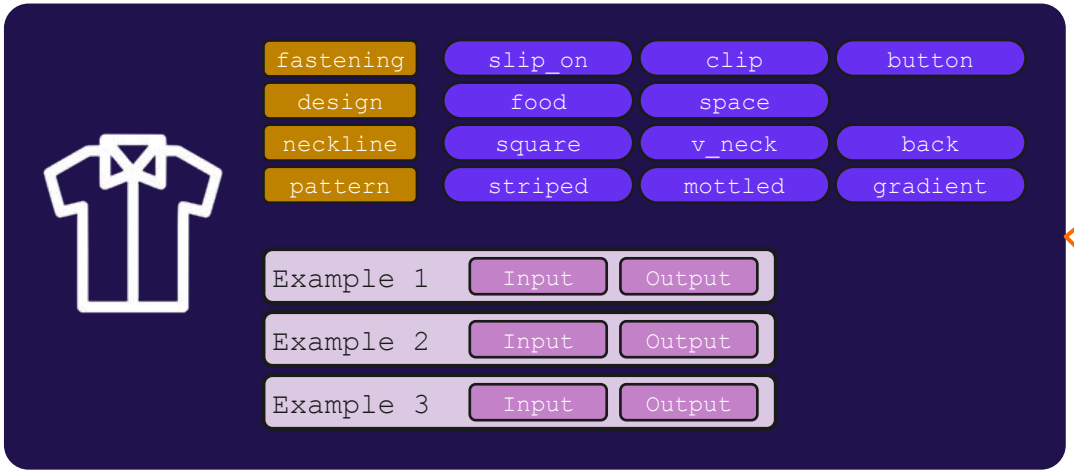
Better understanding of how it would fit will reduce returns

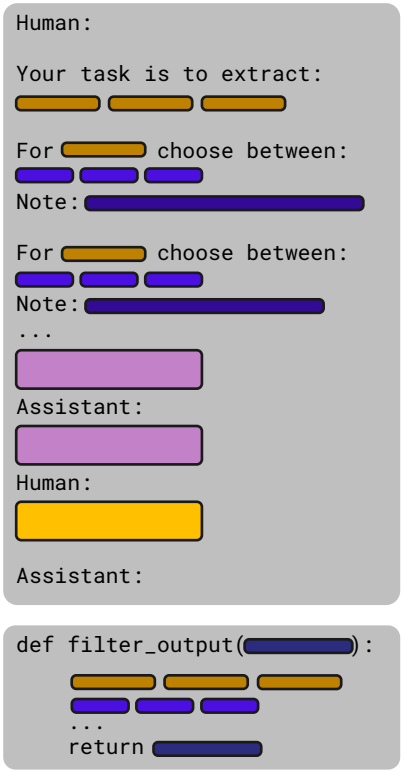
Less repackaging, less pollution, faster delivery, lower costs

How It Works


Raw description
Human:
Your task is to extract:
For choose between:
Note:
For choose between:
Note:
...
Assistant:
Human:
Assistant:
Unfiltered Metadata
def filter_output():
...
return
Metadata

Prompts and Filters

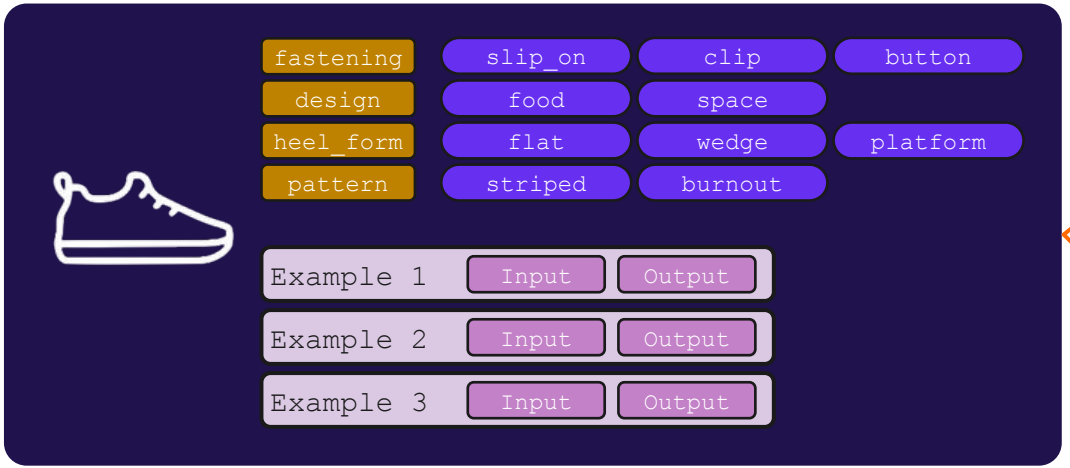


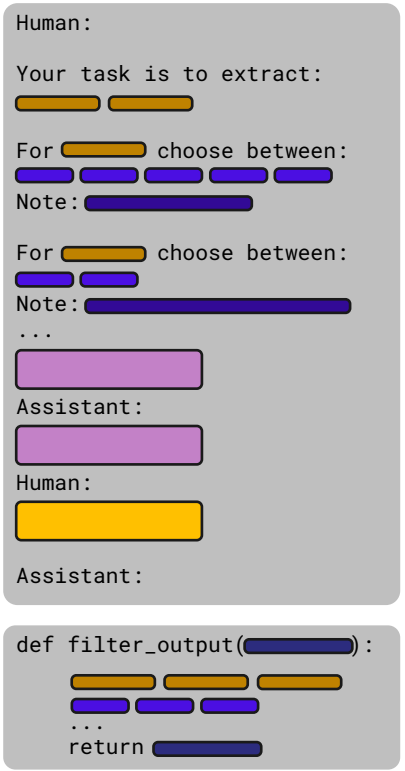


input output

Given a product description and the related taxonomy, we create a custom prompt and a filter function

Prompts and Filters

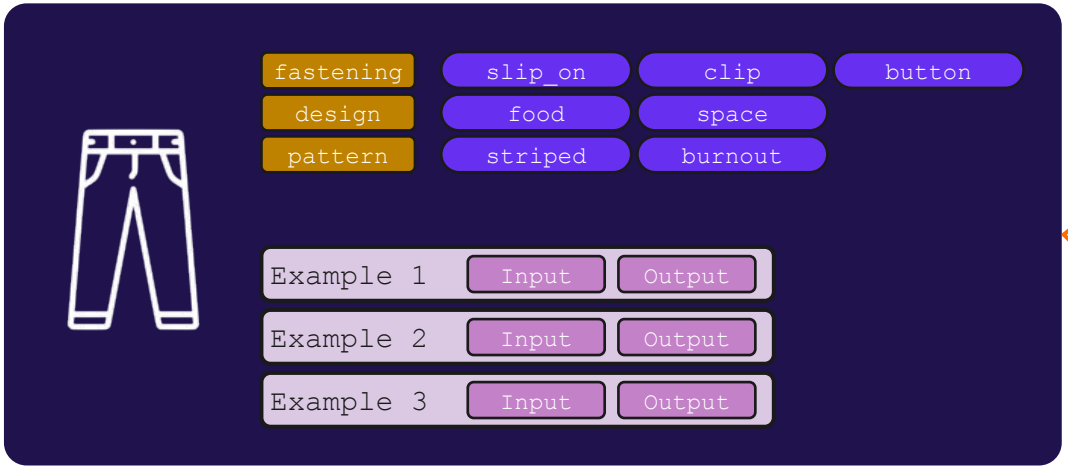


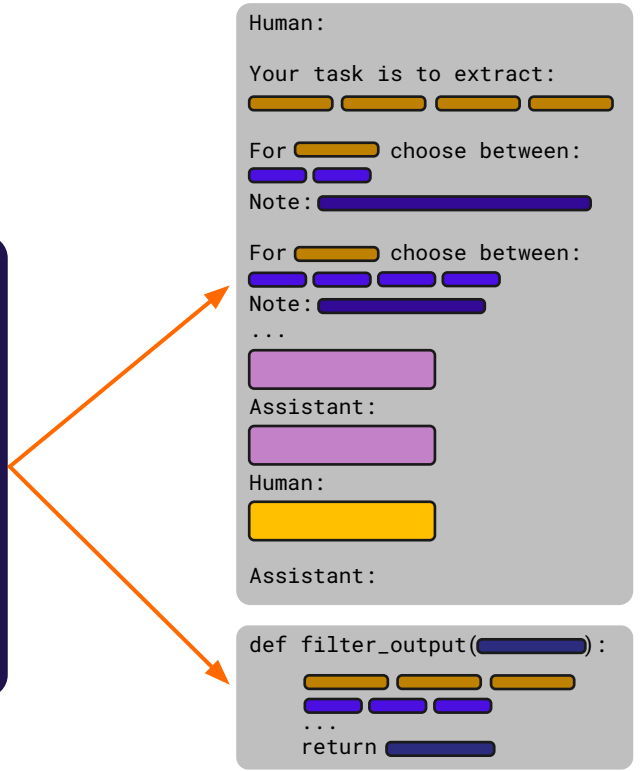


input output

Given a product description and the related taxonomy, we create a custom prompt and a filter function

Prompts and Filters





input output

The prompt and the filter functions differ for every silhouette


With the use of GenAI
Automated Extraction
Accuracy Improvement
70% → 90%

Algorithmic red prices: Introduction

Nike Sportswear

MANOA 17 UNISEX - High-top trainers

41,45 € VAT included

Originally: ~~54,95 €~~ -25%

★★★★½ 1

Colour: **wheat/black**



Choose your size

Add to bag

1-2 working days

Algorithmic red prices: Introduction

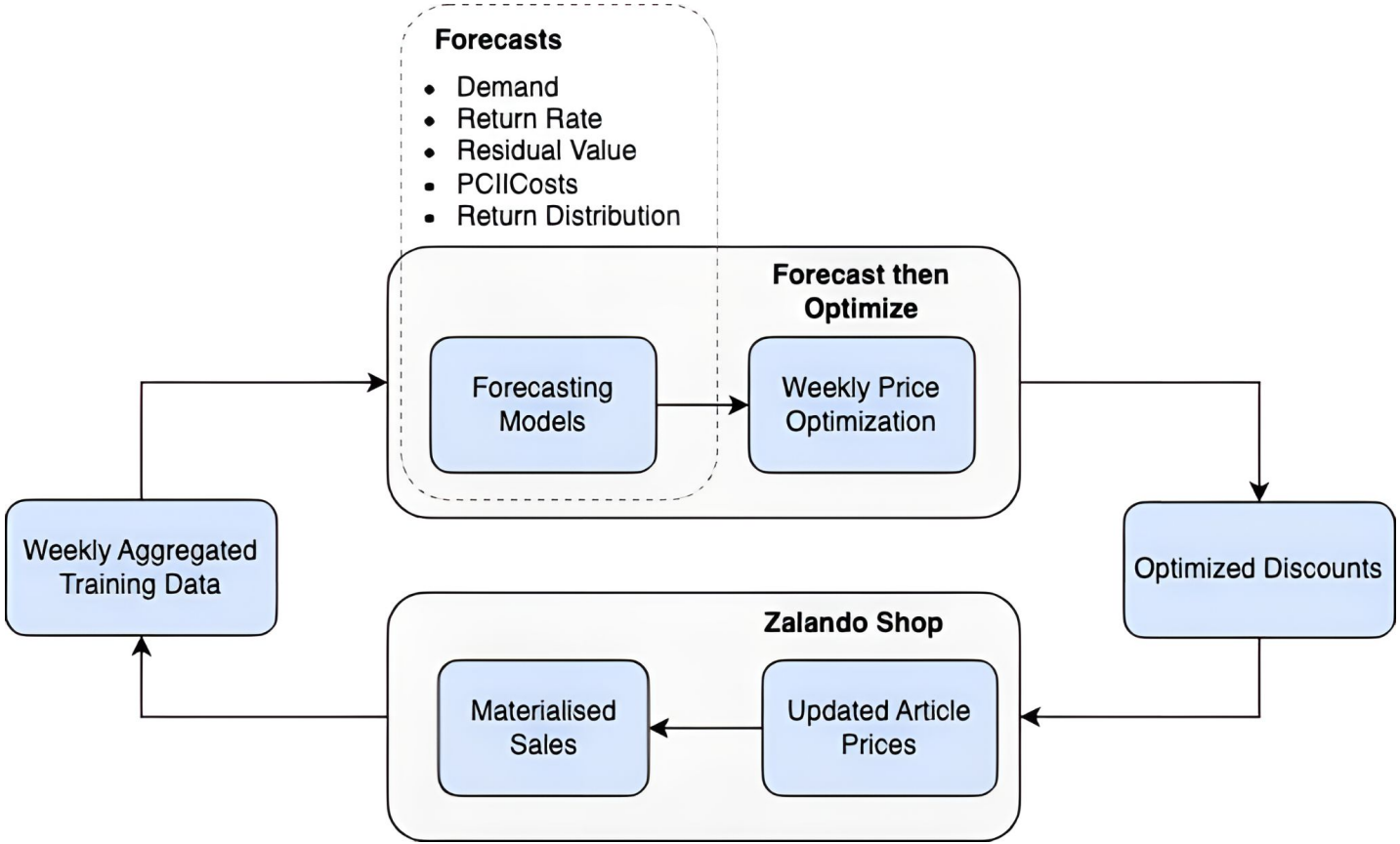
“Red price” (discounted price, adjustable)

“Black price” (recommended retail price, usually set once)

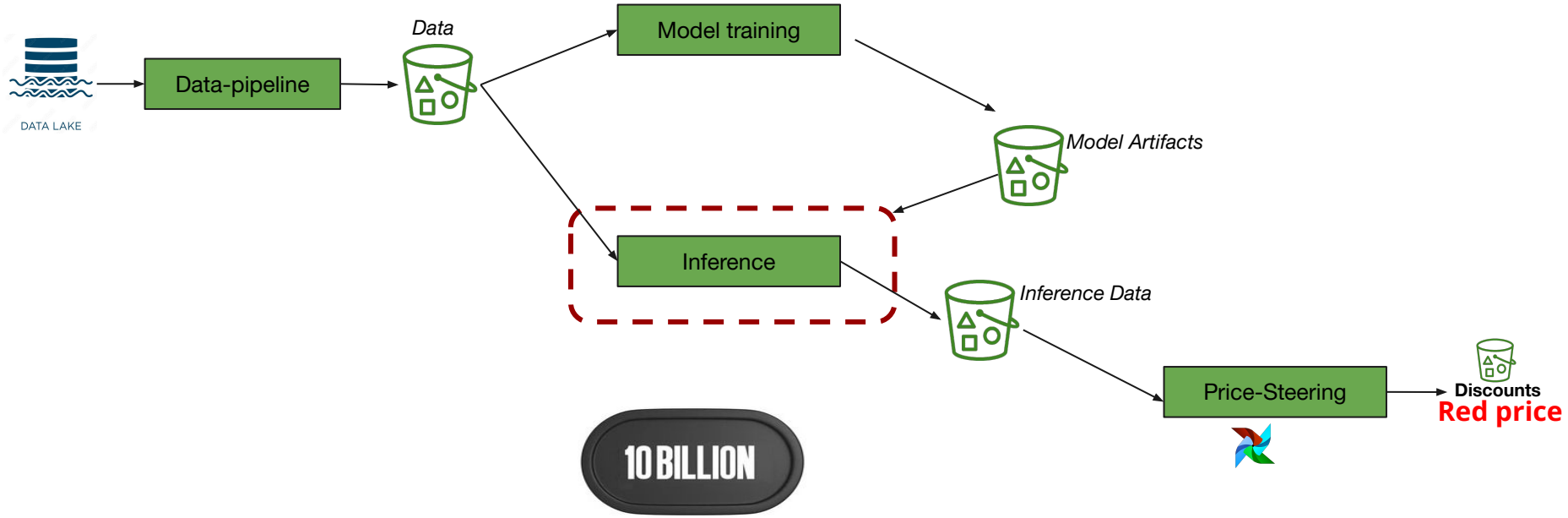
Goal

“On article-level, regularly (daily, weekly,...) recommend profit-optimal discounts for ~900K articles such that business constraints are fulfilled.”

Algorithmic red prices: The role of ML forecasts

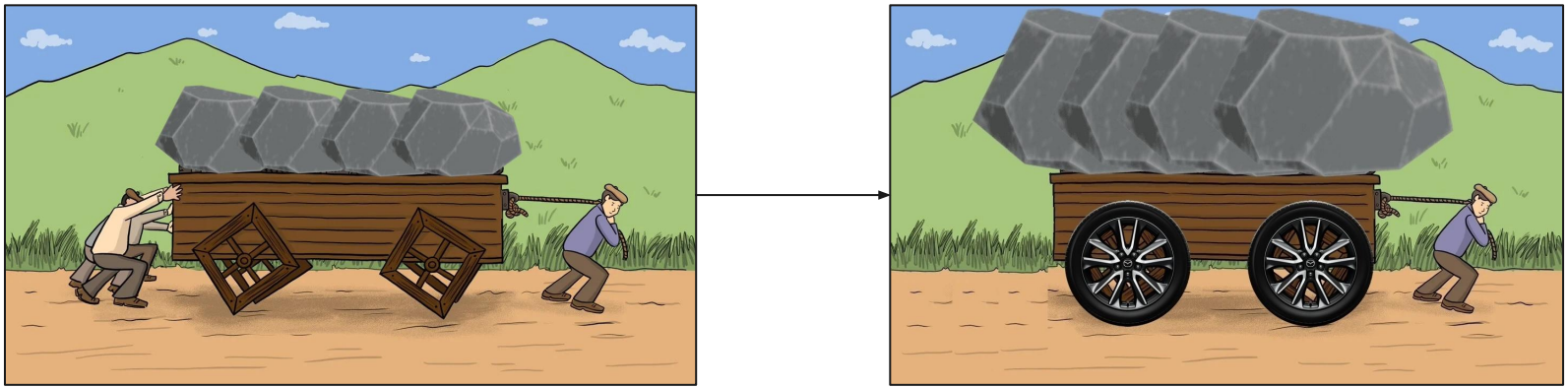


Algorithmic red prices: Forecasting Infrastructure



Forecast records

High-level outcome: Heavily improved inference pipeline

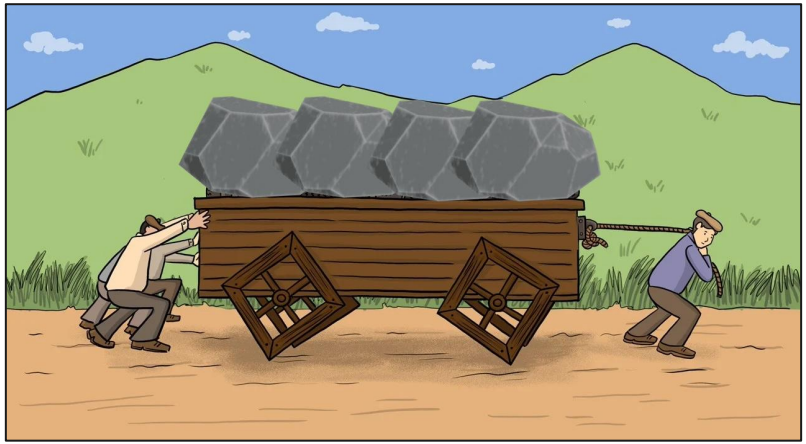


Inference running **4x faster**, **18x cheaper**, scaling to more than **3x more data load***

*Work done in collaboration with external consultants from AWS ProServe

Key Challenges

Slow & Resource-Intensive pipeline execution



Impact: SLA

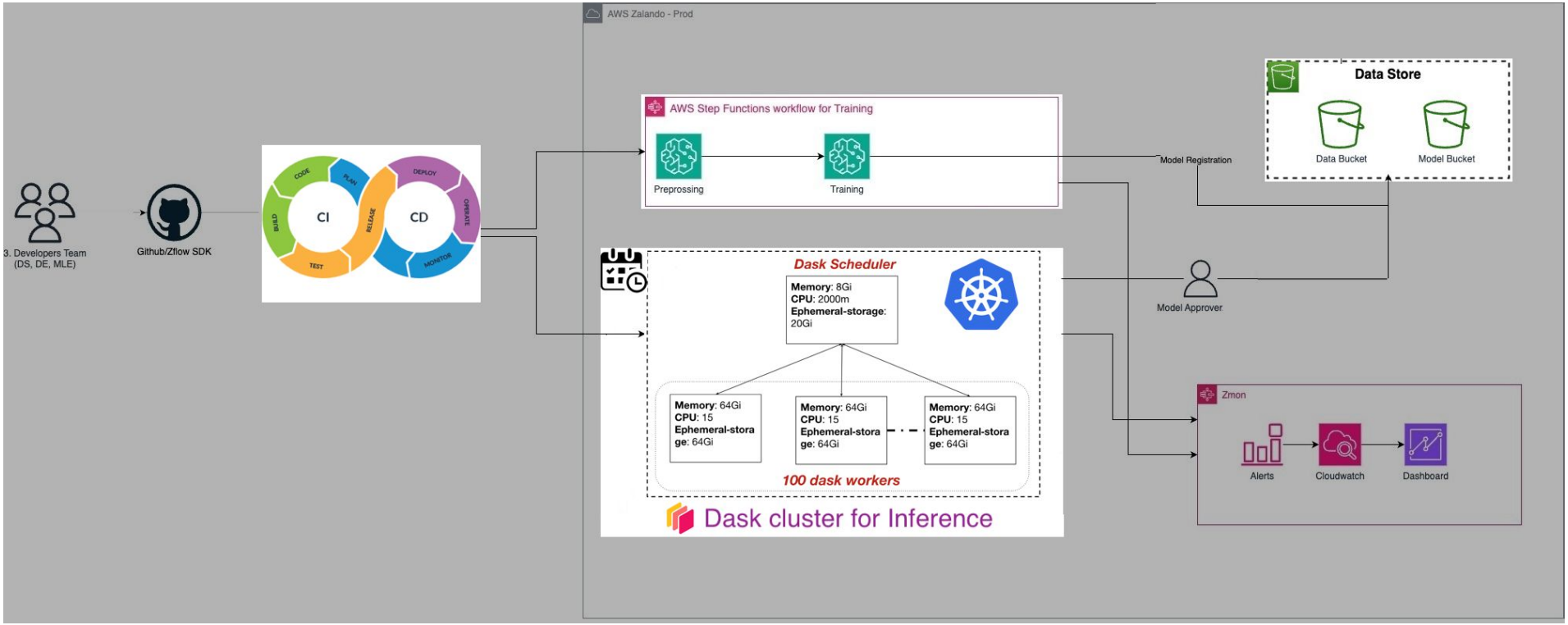
DELAYED

Lack of Scalability

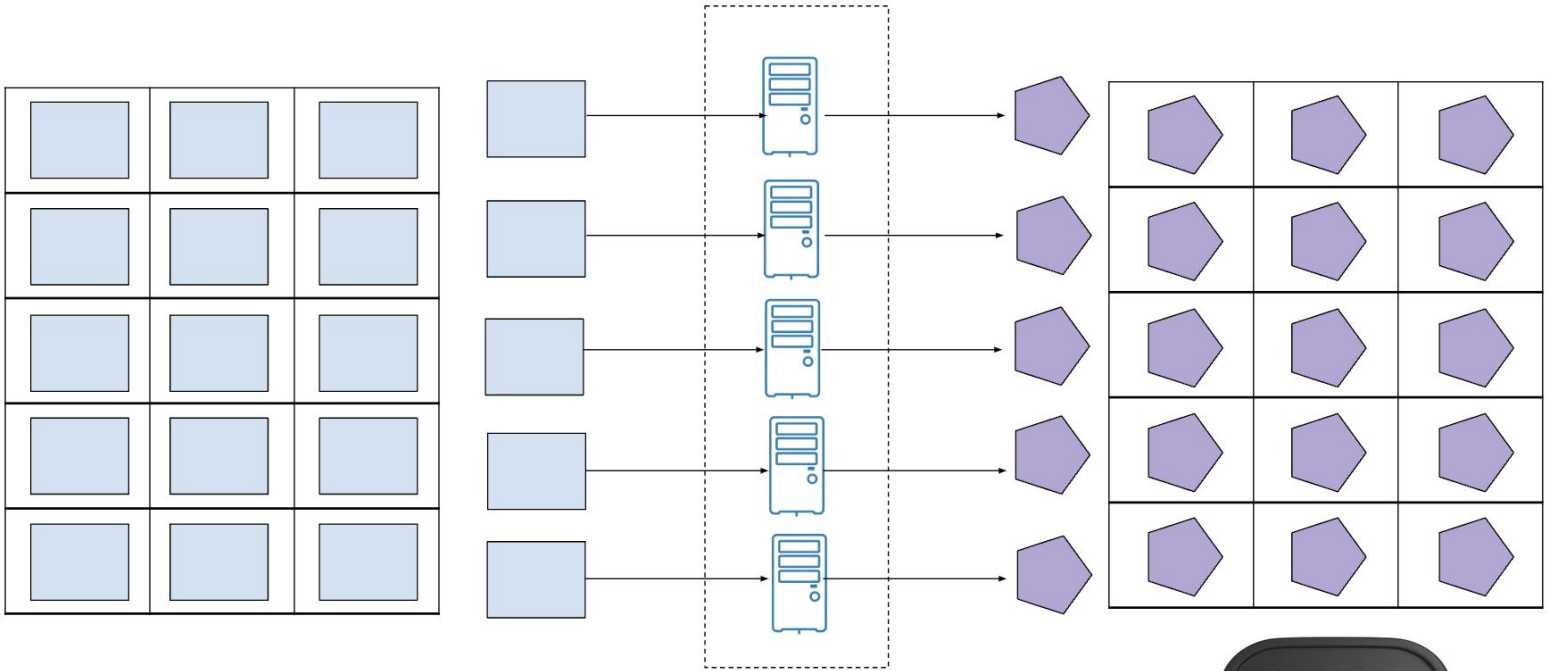
Impact: Cluster slowdown & unresponsive

CANCELLED

Old architecture



Addressing Challenges: Data & Operation



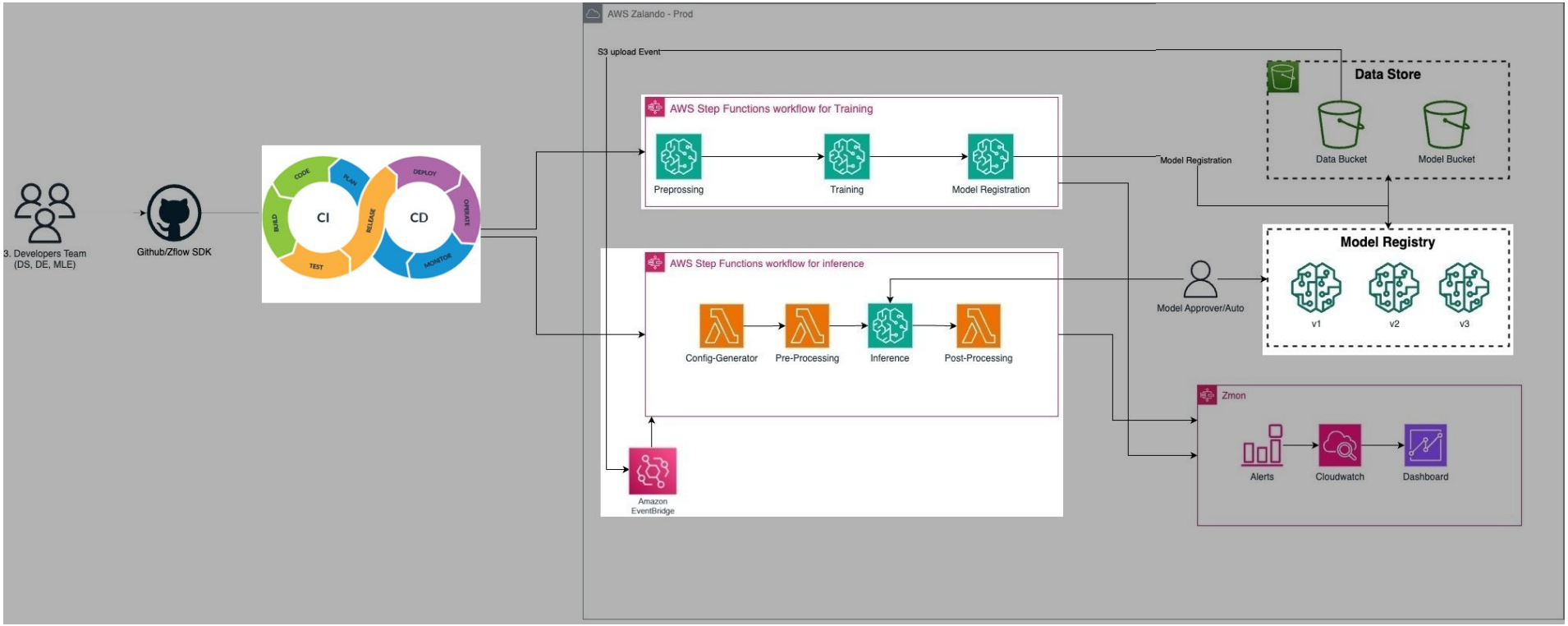
Data with multiple partitions

- *Embarrassingly parallel operations*
- *Operations don't require coordination between nodes*

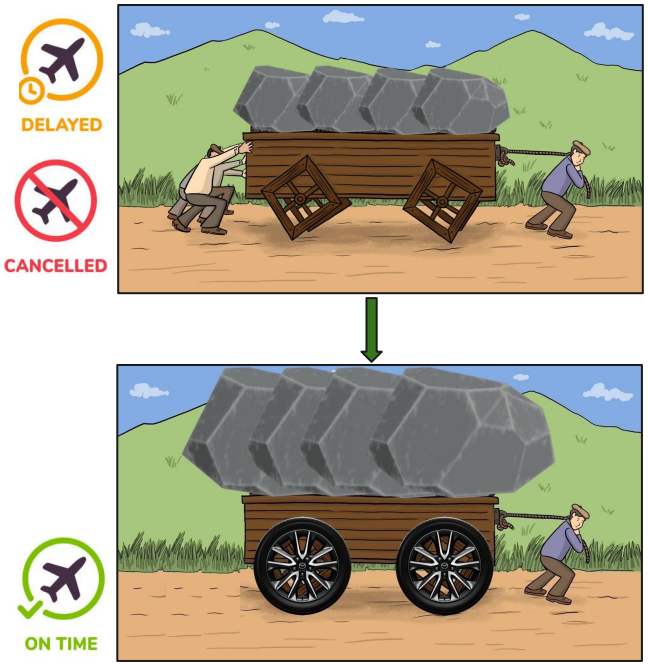
10 BILLION

Forecast records

New architecture



Summary: Improvements with new Infrastructure



SageMaker Model Registry

Amazon S3

Amazon EventBridge

AWS Step Functions

Amazon SageMaker

AWS Lambda

- **Inference** running **4x faster**, **18x cheaper**
- **Inference** scaling to more than **3x more data load**

	Old	New
	30 mins	8 mins
CPU	1500	160
RAM	~ 6TB	~ 700 GB