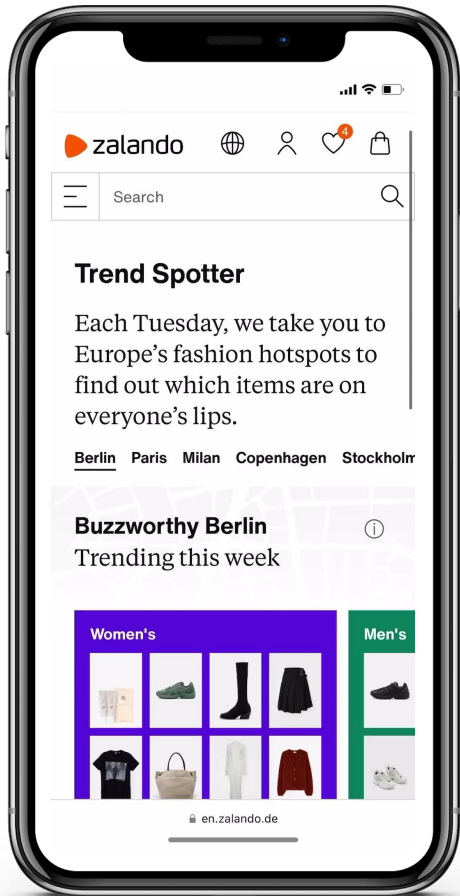




Enhancing Fashion E-Commerce with Cutting-Edge AI, Powered by AWS

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

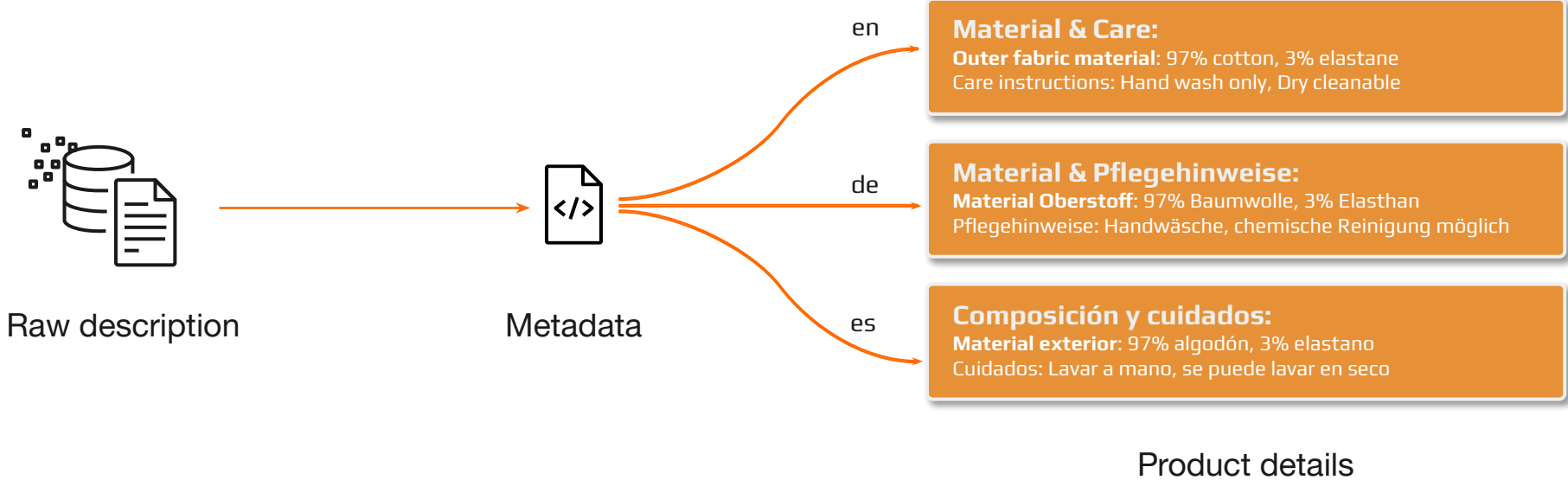
Sleeve length: 84.45 cm

Total length: Size 40

Detailed product information allow customers to make the right purchase decisions



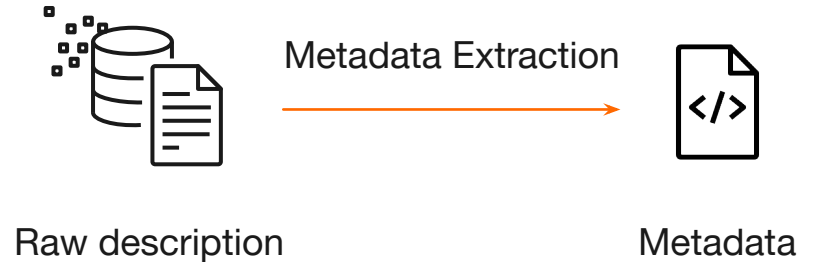
Metadata



Challenges

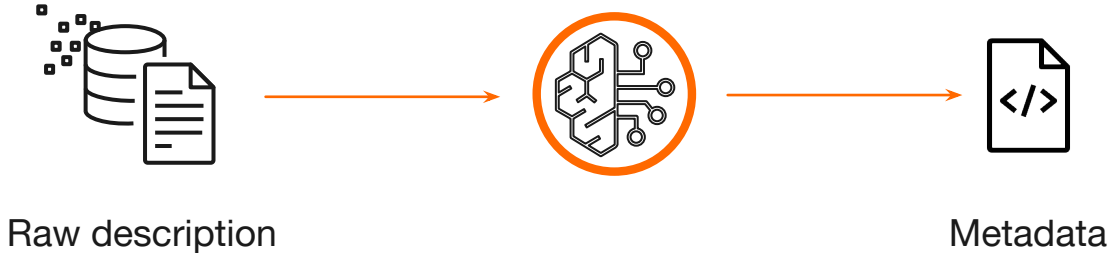
Product descriptions come from different data sources

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



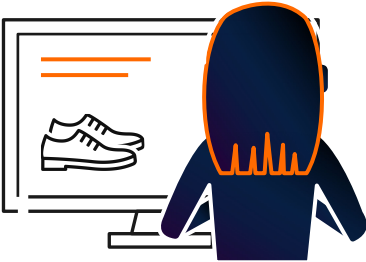
Metadata extraction was partially done **manually**

Better Metadata with GenAI

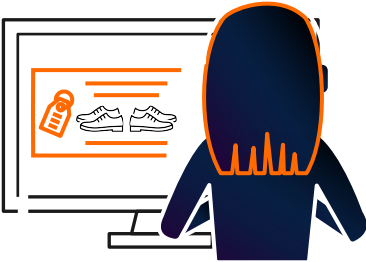


Use GenAI to Automate

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)

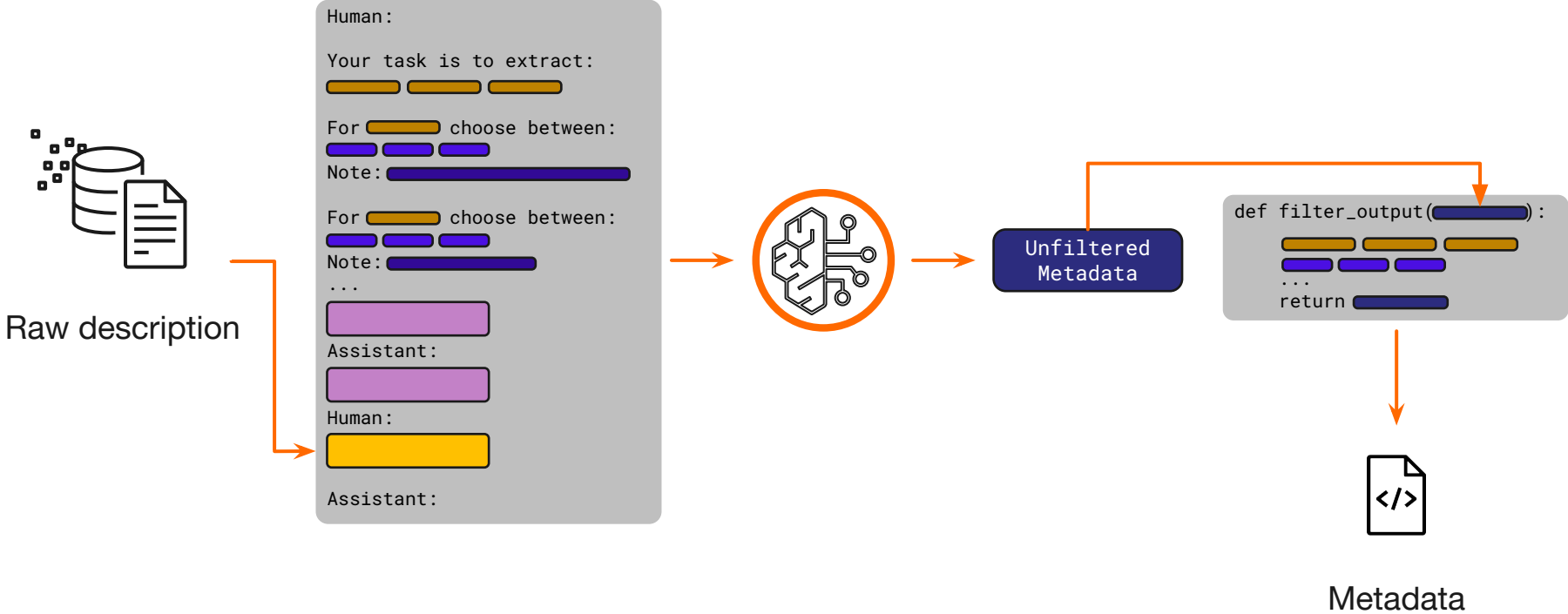
Total length: 27.0" (Size M)

Accurate metadata will **improve personalization**

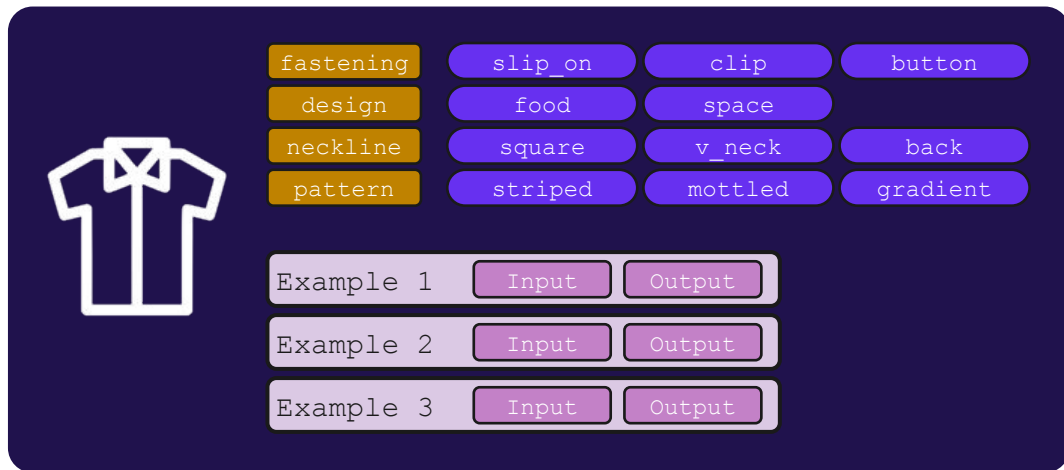
Better understanding of how it would fit will **reduce returns**

Less repackaging, less pollution, faster delivery, **lower costs**

How It Works



Prompts and Filters



A diagram illustrating a product taxonomy for shirts. On the left is a white outline of a shirt. To its right is a grid of 16 colored buttons representing different attributes: fastening (orange), slip_on (purple), clip (purple), button (purple), design (orange), food (purple), space (purple), neckline (orange), square (purple), v_neck (purple), back (purple), pattern (orange), striped (purple), mottled (purple), and gradient (purple). Below the grid are three example rows, each with a text input field and two buttons labeled 'Input' and 'Output'. A legend at the bottom shows an orange box for 'input' and a purple box for 'output'.

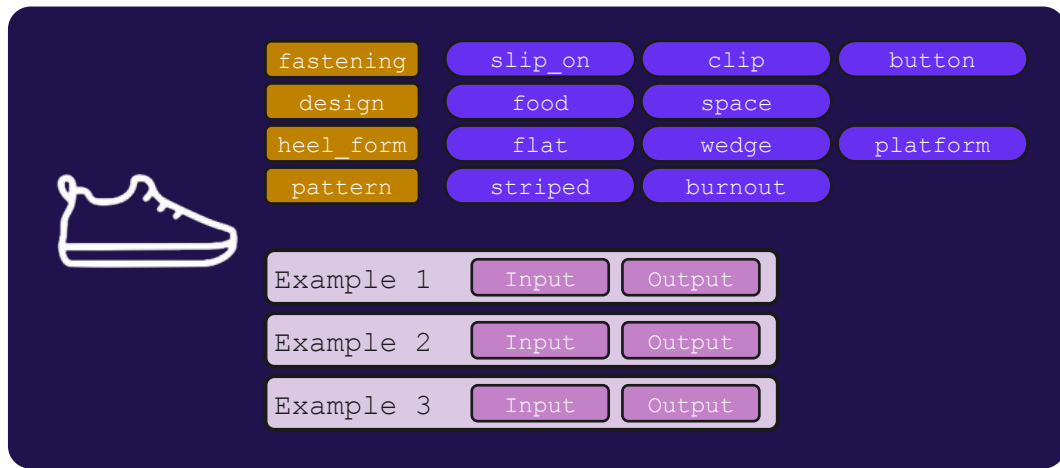
input output

Human:
Your task is to extract:
[orange] [orange] [orange]
For [orange] choose between:
[purple] [purple]
Note: [purple]
For [orange] choose between:
[purple] [purple]
Note: [purple]
...
[purple]
Assistant:
[purple]
Human:
[orange]
Assistant:

```
def filter_output([purple]):  
    [orange] [orange] [orange]  
    [purple] [purple]  
    ...  
    return [purple]
```

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

Prompts and Filters



A diagram illustrating a product taxonomy for shoes. On the left is a white line-art icon of a sneaker. To its right is a grid of 16 colored buttons representing different attributes: fastening (orange), slip_on (purple), clip (purple), button (purple), design (orange), food (purple), space (purple), heel_form (orange), flat (purple), wedge (purple), platform (purple), and pattern (orange), striped (purple), burnout (purple). Below the grid are three example rows, each with an 'Input' and 'Output' field. A legend at the bottom shows an orange box for 'input' and a purple box for 'output'.

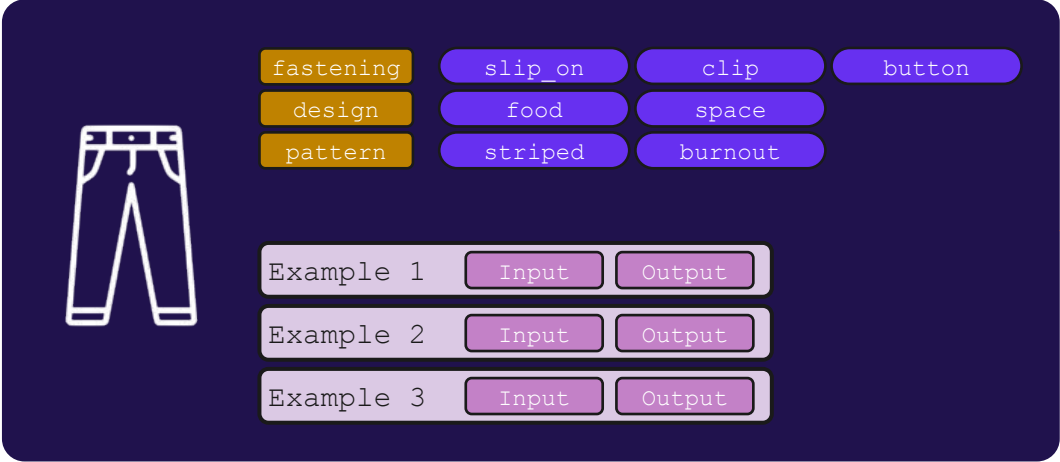
input output

Human:
Your task is to extract:
[orange] [orange]
For [orange] choose between:
[purple] [purple] [purple] [purple]
Note: [purple]
For [orange] choose between:
[purple] [purple]
Note: [purple]
...
[purple]
Assistant:
[purple]
Human:
[orange]
Assistant:

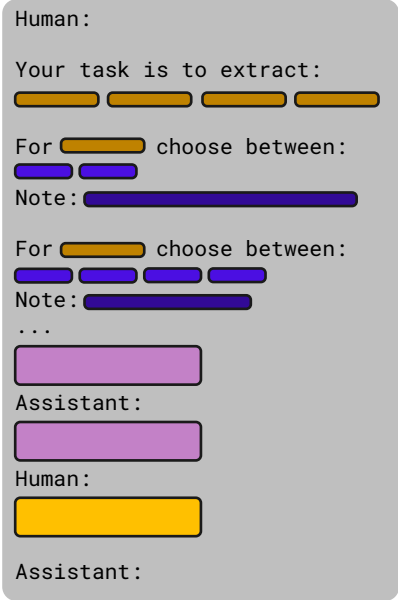
```
def filter_output([purple]):  
    [orange] [orange] [orange]  
    [purple] [purple] [purple]  
    ...  
    return [purple]
```

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

Prompts and Filters



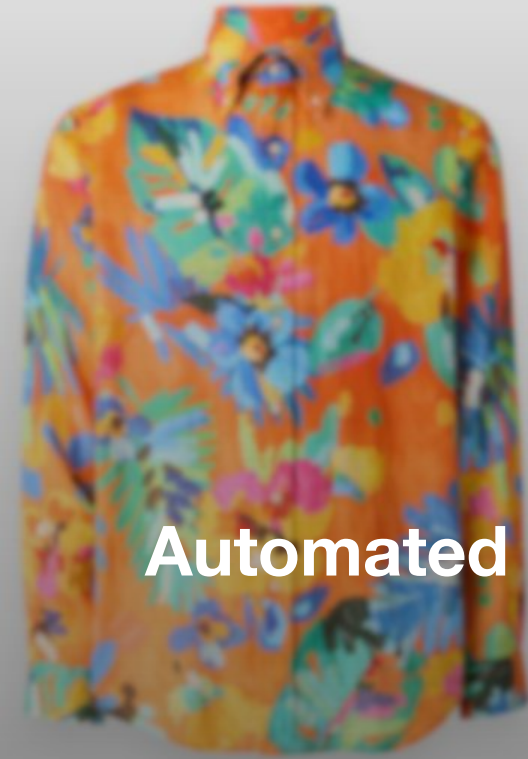
A dark blue rounded rectangle contains a white line-art icon of a pair of pants on the left. To the right of the icon are several colored boxes representing prompt components: 'fastening' (yellow), 'design' (yellow), and 'pattern' (yellow) are stacked vertically; 'slip on', 'food', and 'striped' are in a row; 'clip', 'space', and 'burnout' are in a row; and 'button' is in a separate box to the right. Below these are three example rows, each with 'Example 1', 'Example 2', or 'Example 3' followed by 'Input' and 'Output' boxes. At the bottom of the rectangle are two boxes labeled 'input' (yellow) and 'output' (dark blue).



A light gray rounded rectangle shows a chat conversation structure. It starts with 'Human:' followed by 'Your task is to extract:' and four yellow boxes. Then 'For [yellow box] choose between:' followed by two blue boxes. Then 'Note:' followed by a long blue box. This pattern repeats once more. There are three purple boxes representing assistant responses, with an ellipsis between the first and second. The conversation ends with 'Assistant:' followed by a purple box, 'Human:' followed by a yellow box, and 'Assistant:'.

```
def filter_output([blue box]):  
    [yellow box] [yellow box] [yellow box]  
    [blue box] [blue box] [blue box]  
    ...  
    return [blue box]
```

The prompt and the filter functions differ for every silhouette



Automated Extraction

With the use of GenAI



**Accuracy Improvement
70% → 90%**





Algorithmic red prices: Introduction



Nike Sportswear MANOA 17 UNISEX - High-top trainers

41,45 € VAT included
Originally: 54,95 € -25%

★★★★☆ 1

Colour: wheat/black



Choose your size ∨

Add to bag ♡



1-2 working days



Algorithmic red prices: Introduction

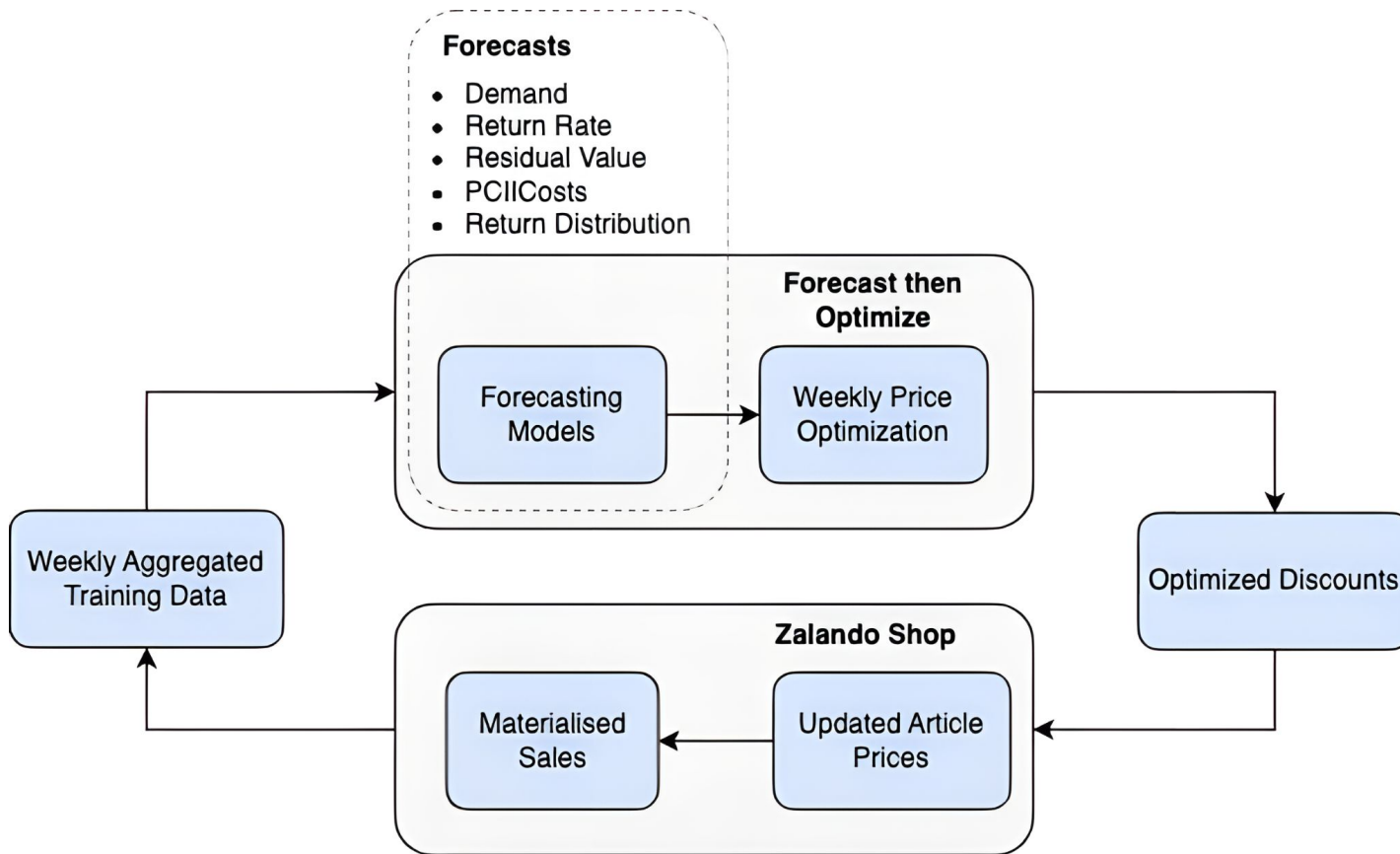


Goal

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

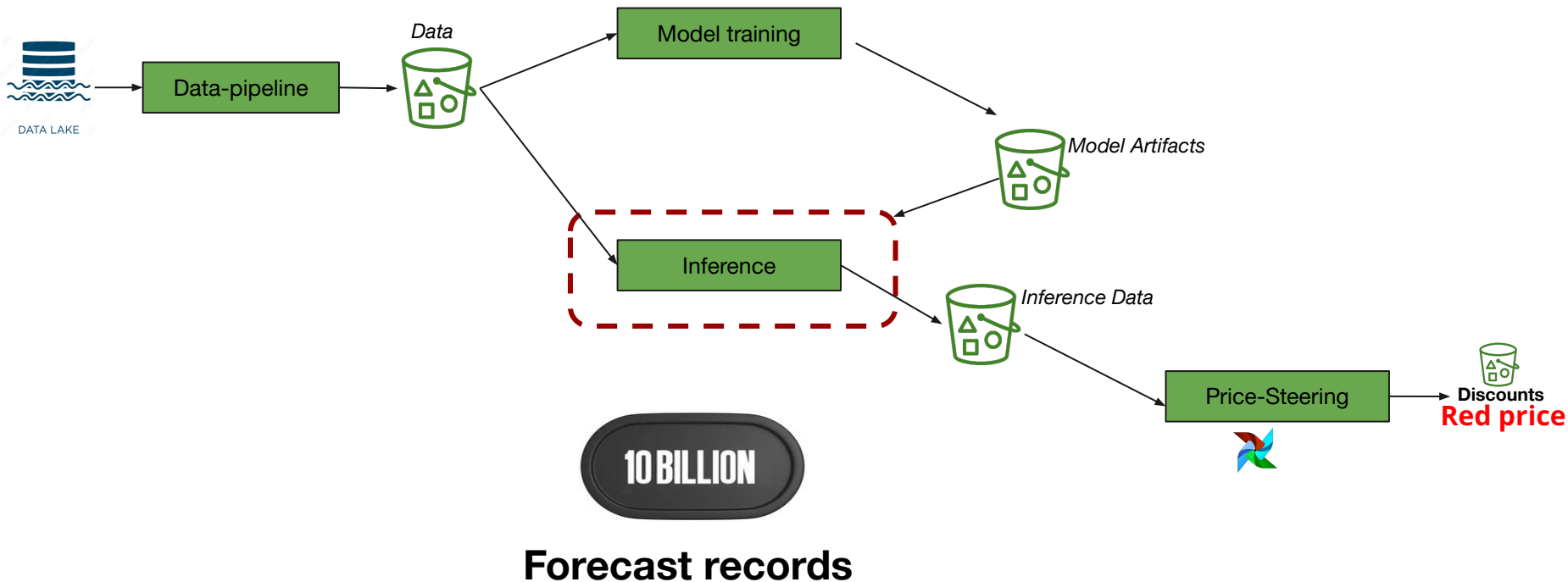


Algorithmic red prices: The role of ML forecasts



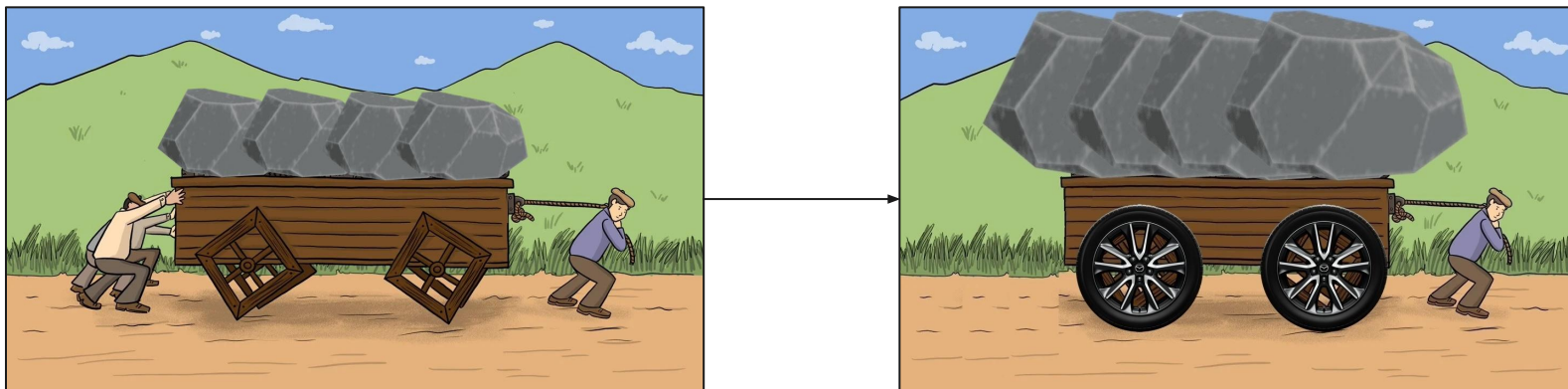


Algorithmic red prices: Forecasting Infrastructure





High-level outcome: Heavily improved inference pipeline



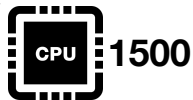
Inference running **4x faster**, **18x cheaper**, scaling to more than **3x more data load***

*Work done in collaboration with external consultants from AWS ProServe



Key Challenges

Slow & Resource-Intensive pipeline execution



Impact: SLA



Lack of Scalability

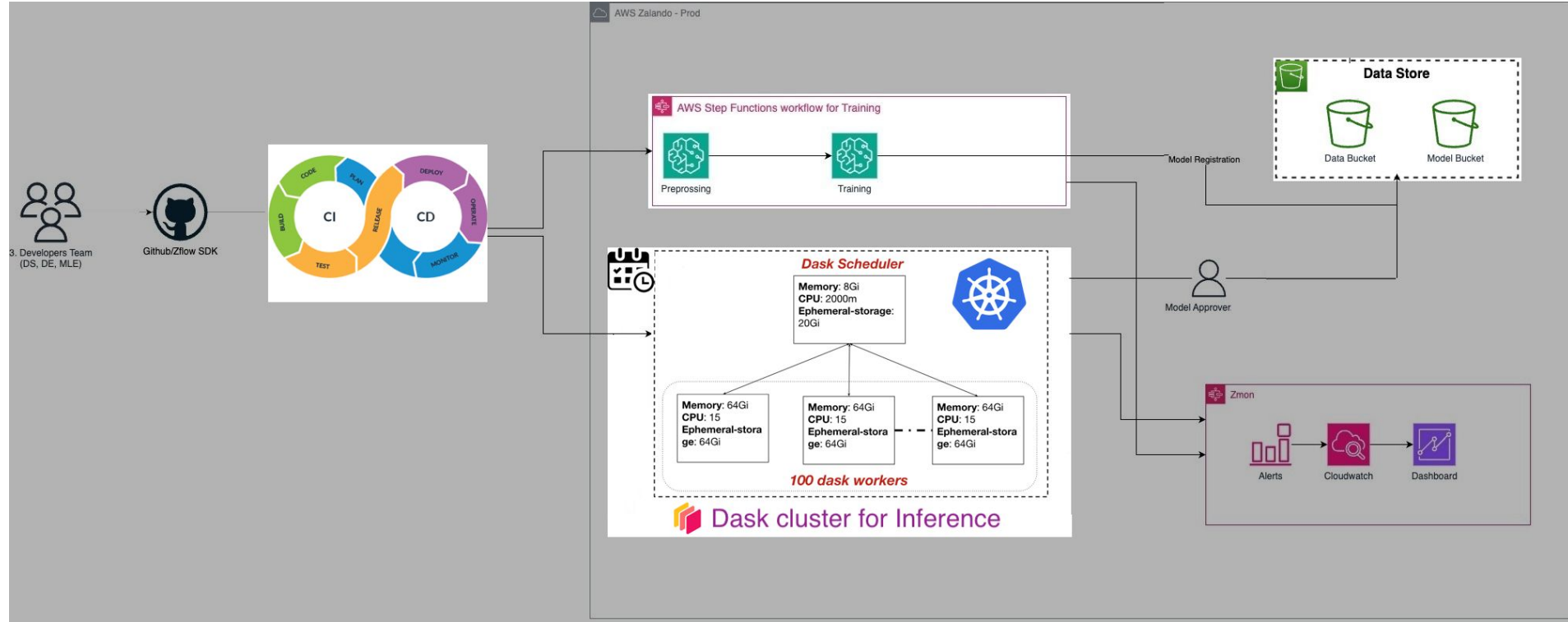


Impact: Cluster slowdown & unresponsive



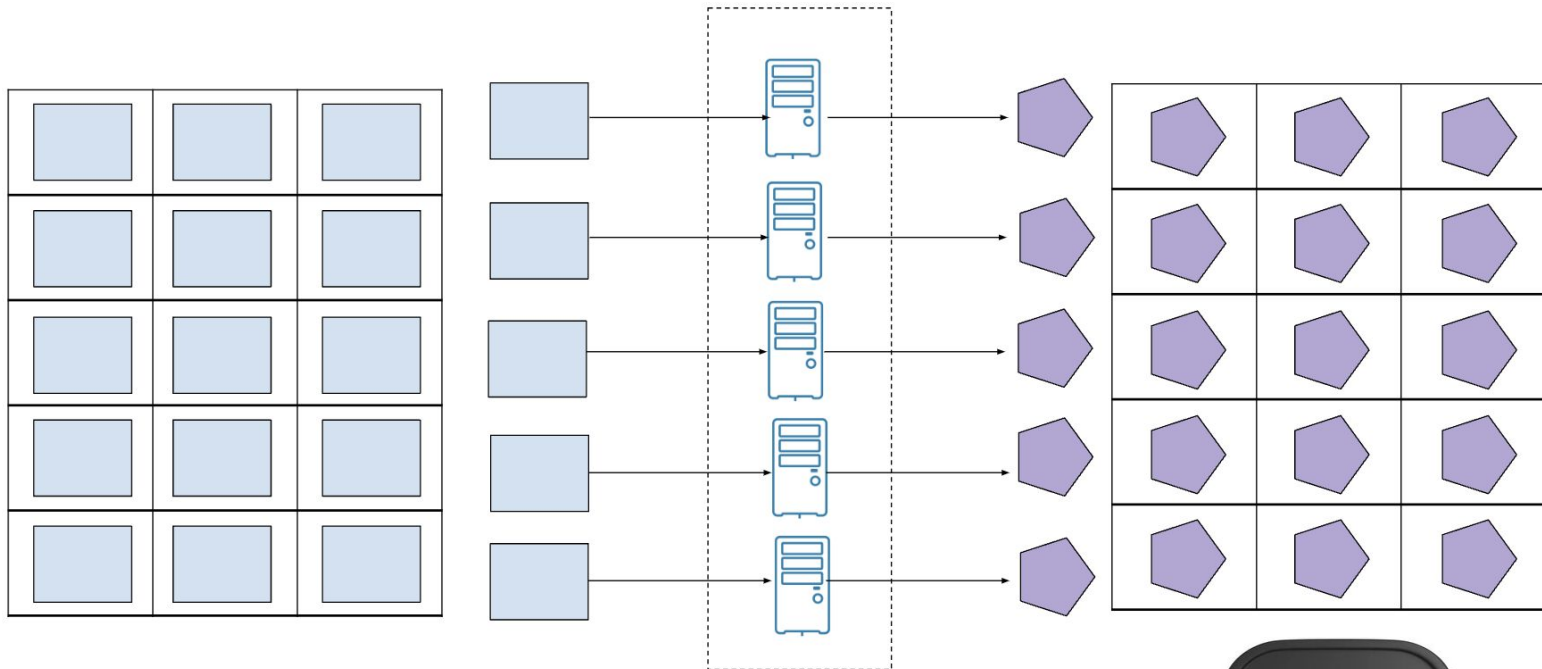


Old architecture





Addressing Challenges: Data & Operation



Data with multiple partitions

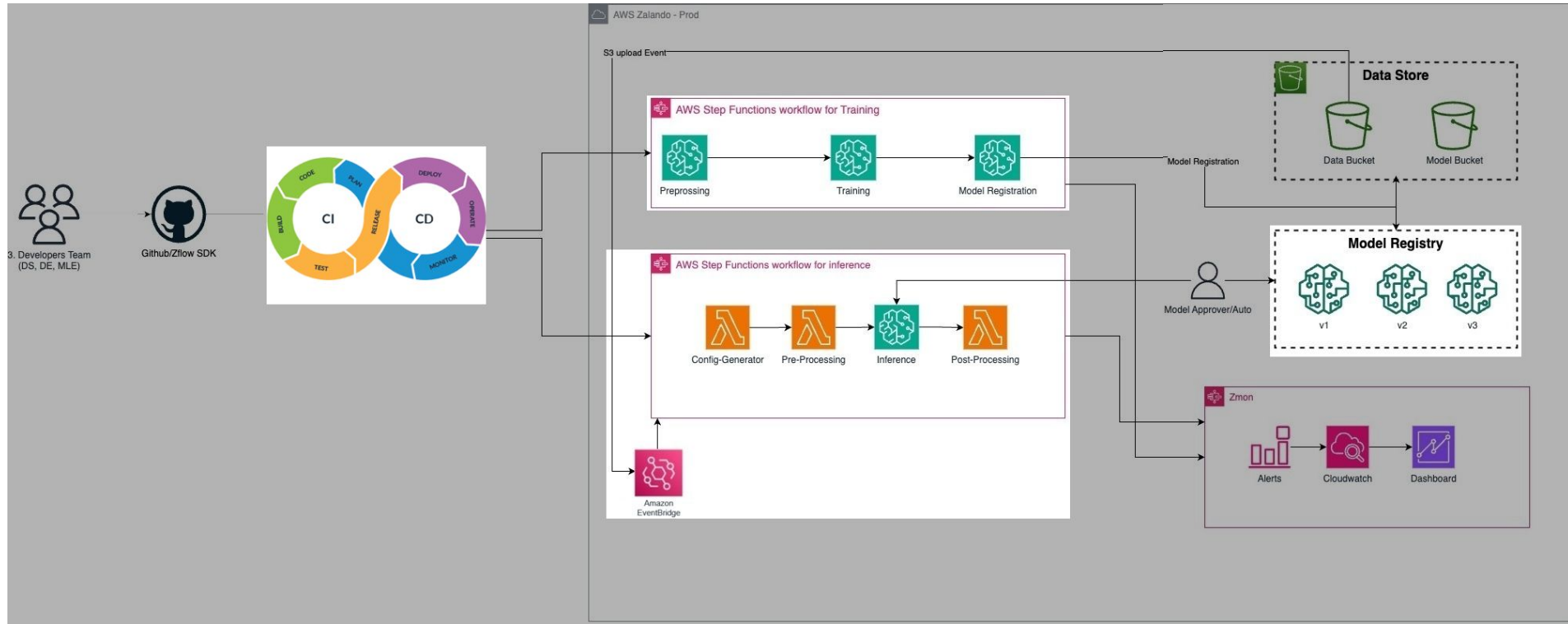
- *Embarrassingly parallel operations*
- *Operations don't require coordination between nodes*

10 BILLION

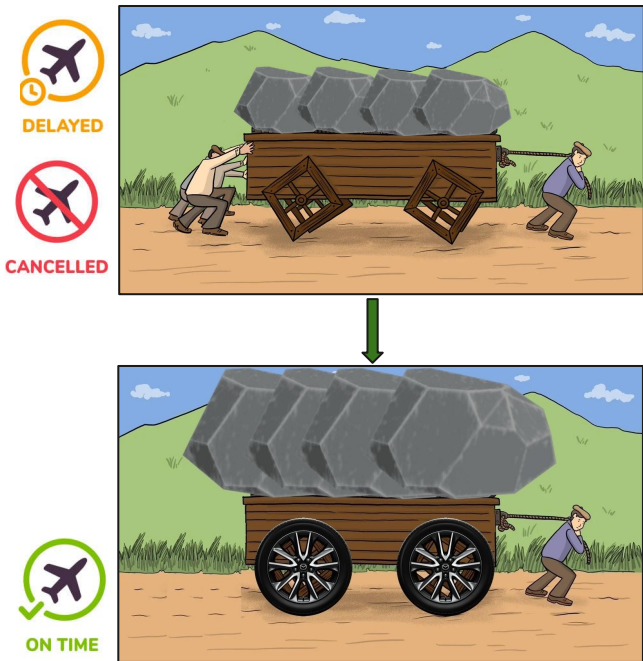
Forecast records



New architecture



Summary: Improvements with new Infrastructure



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

	Old	New
	1500	160
	~ 6TB	~ 700 GB

