

Nobody Wants To Wait

Case by **R0CCAI**





9:41

89%

Boldspilleren

Boldspilleren elsker konkurrence og teamwork. Uanset om det er fodbold, basketball eller håndbold, trives de med at være en del af et hold, hvor strategi og samarbejde er nøglen til succes.

[Se Aktiviteter](#)

75%

Ketchermesteren

Ketchermesteren har præcision, hurtighed og taktisk sans. Uanset om det er tennis,

Theory

Theory

- Optimize for what?

Flexibility

Development Cost

Training Time

Memory

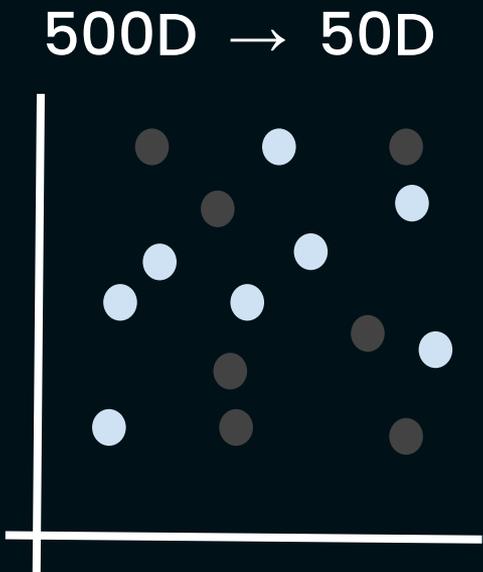
Inference Time

Accuracy

Production Cost

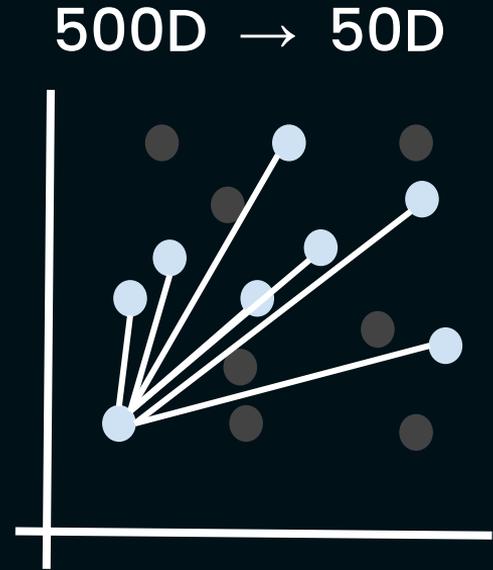
Theory

- Optimize for what?
- Minimize the Search Space



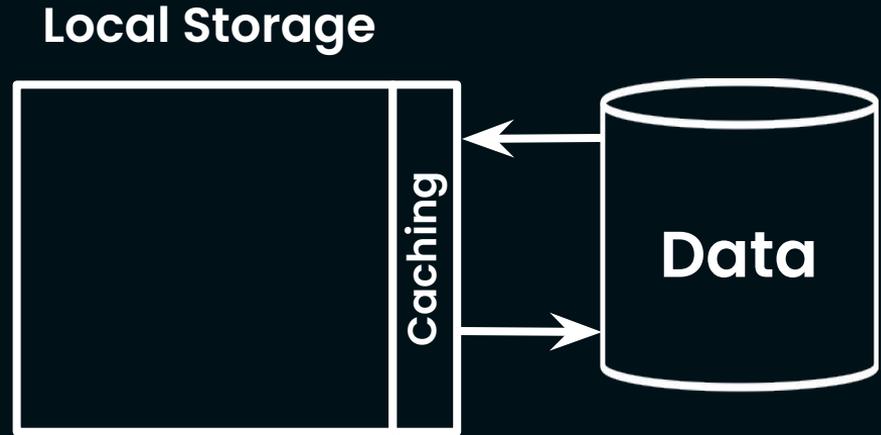
Theory

- Optimize for what?
- Minimize the Search Space
- Precalculate and reuse



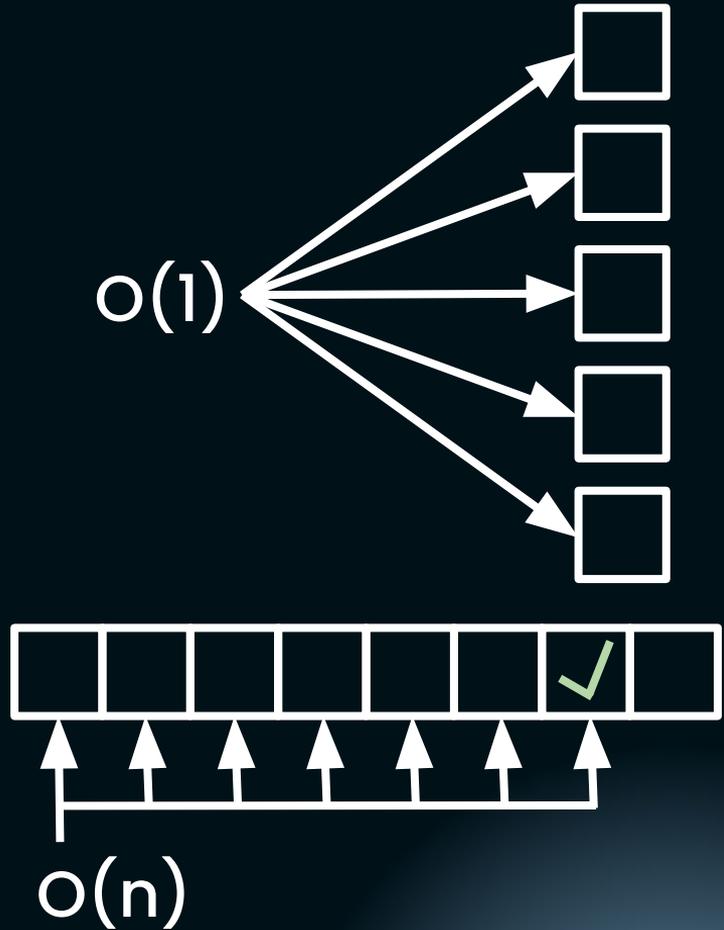
Theory

- Optimize for what?
- Minimize the Search Space
- Precalculate and reuse
- Move Data / Caching



Theory

- Optimize for what?
- Minimize the Search Space
- Precalculate and reuse
- Move Data / Caching
- Data Structures
- O-Notations



Theory

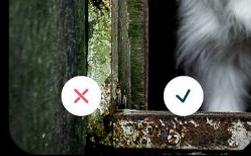
- Optimize for what?
- Minimize the Search Space
- Precalculate and reuse
- Move Data / Caching
- Data Structures
- O-Notations
- Profiling
- Experience of time (UX)

The answer to life, the universe, and everything?



The answer to life, the universe, and everything is a famous phrase from Douglas Adams' science fiction series "The Hitchhiker's Guide to the Galaxy." In the story, a supercomputer named Deep Thought is built

Real World Examples



9:41

89%

Boldspilleren

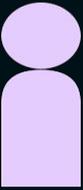
Boldspilleren elsker konkurrence og teamwork. Uanset om det er fodbold, basketball eller håndbold, trives de med at være en del af et hold, hvor strategi og samarbejde er nøglen til succes.

[Se Aktiviteter](#)

75%

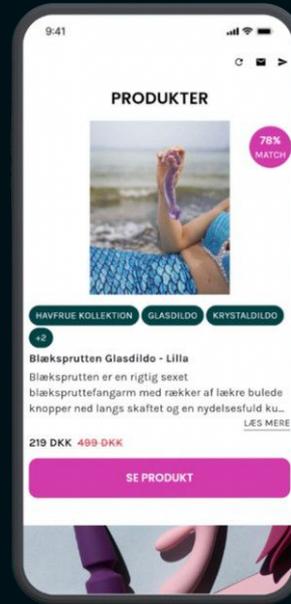
Ketchermesteren

Ketchermesteren har præcision, hurtighed og taktisk sans. Uanset om det er tennis,

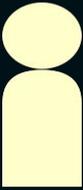

Customer



Guide



Recommendations


End-User

Prior



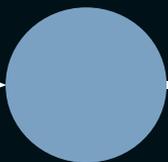
End-User

Create Cards

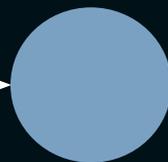
Import
Products



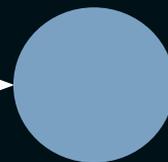
Analyze
Products



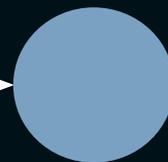
Analyze
Cards



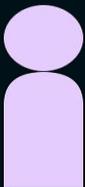
Card
Sequence



Relevant
Products



Live



Customer



Prior



End-User

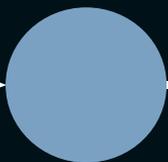
Create Cards

Import
Products

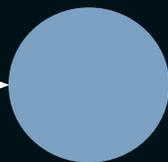


2 min.

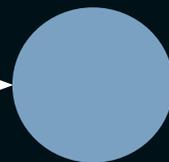
Analyze
Products



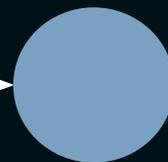
Analyze
Cards



Card
Sequence



Relevant
Products



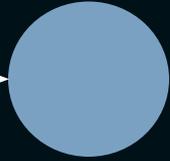
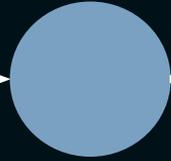
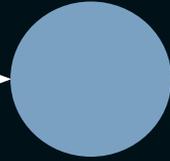
Live



Customer



Get Text



Word
Frequency

Word
Relevance

Filter
Words

Prior



End-User

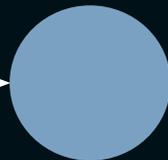
Create Cards

Import
Products



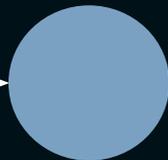
2 min.

3 min.



Analyze
Products

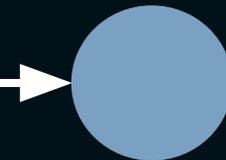
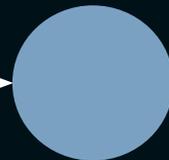
Analyze
Cards



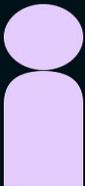
1 min.

Card
Sequence

Relevant
Products

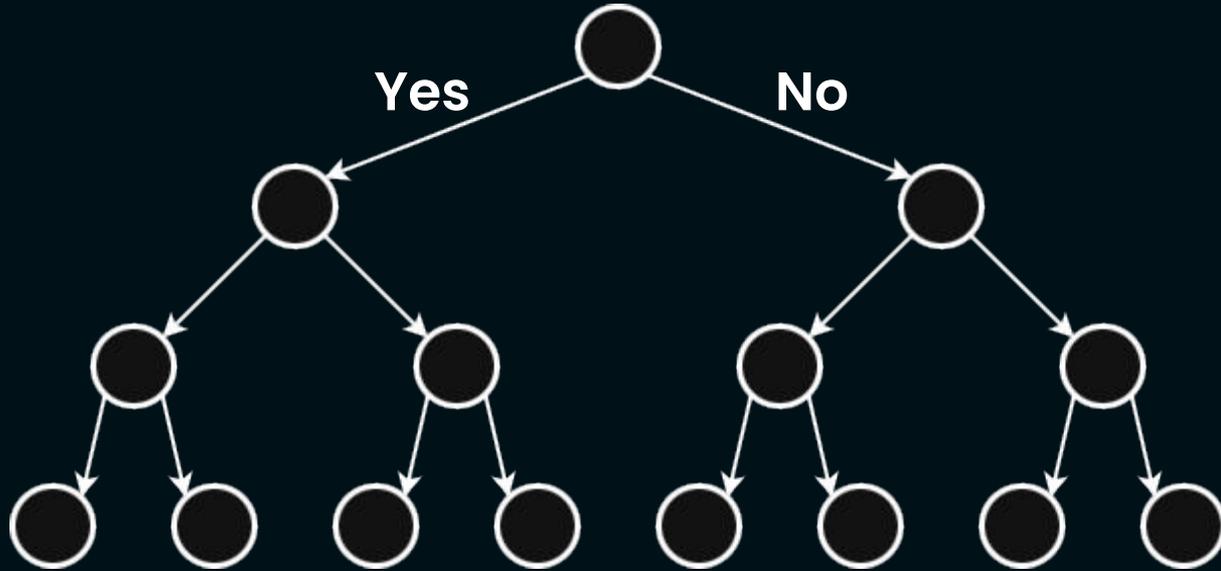


Live



Customer

Decision Tree



Prior



End-User

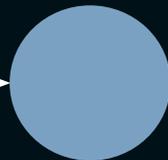
Create Cards

Import Products



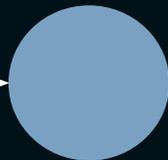
2 min.

3 min.



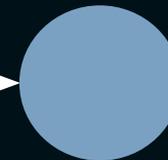
Analyze Products

Analyze Cards



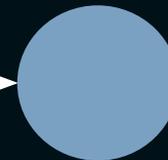
1 min.

10 min.

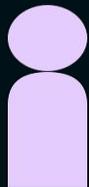


Card Sequence

Relevant Products



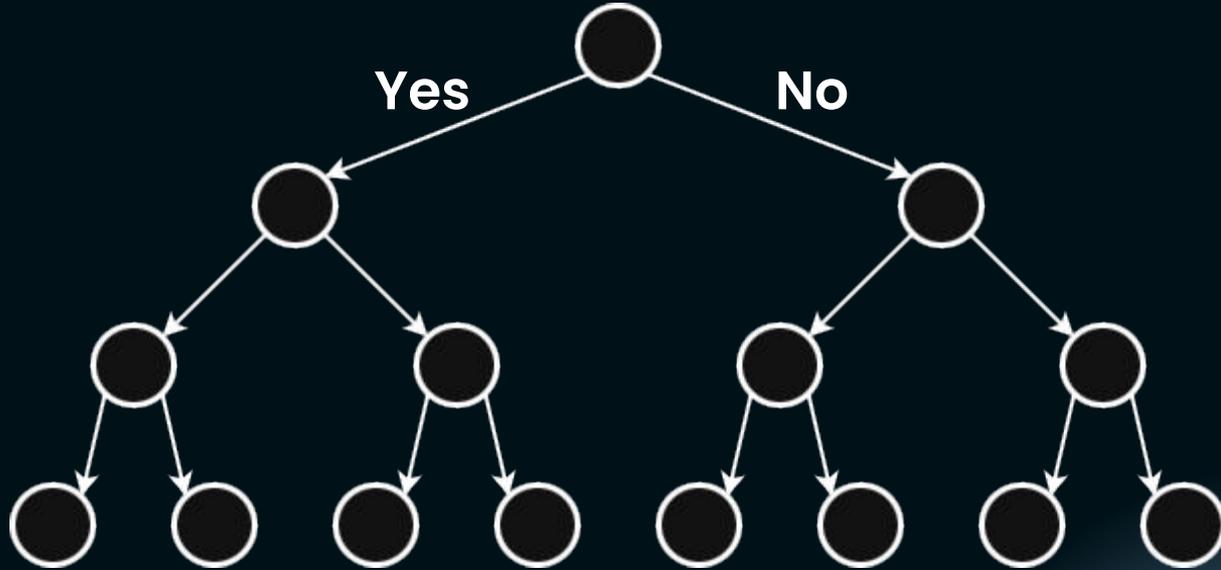
Live



Customer

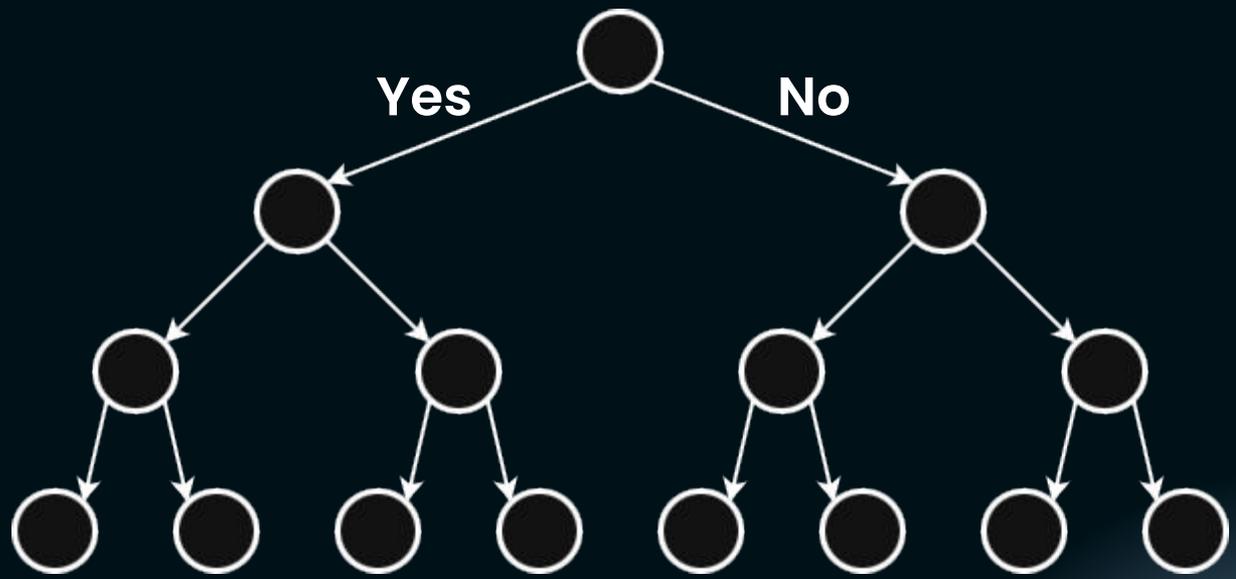
Decision Tree

16 x 11 x 3 x 11 x 11 x 3 x 3 x 3 x 3 x 3



Decision Tree

≈ 15 million → 10 million



Prior



End-User

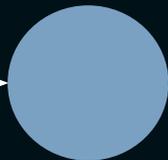
Create Cards

Import
Products



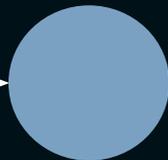
2 min.

3 min.



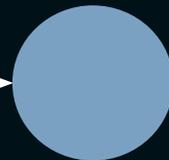
Analyze
Products

Analyze
Cards



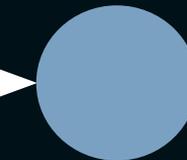
1 min.

10 min.



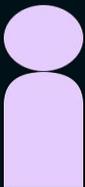
Card
Sequence

Relevant
Products



10 min.

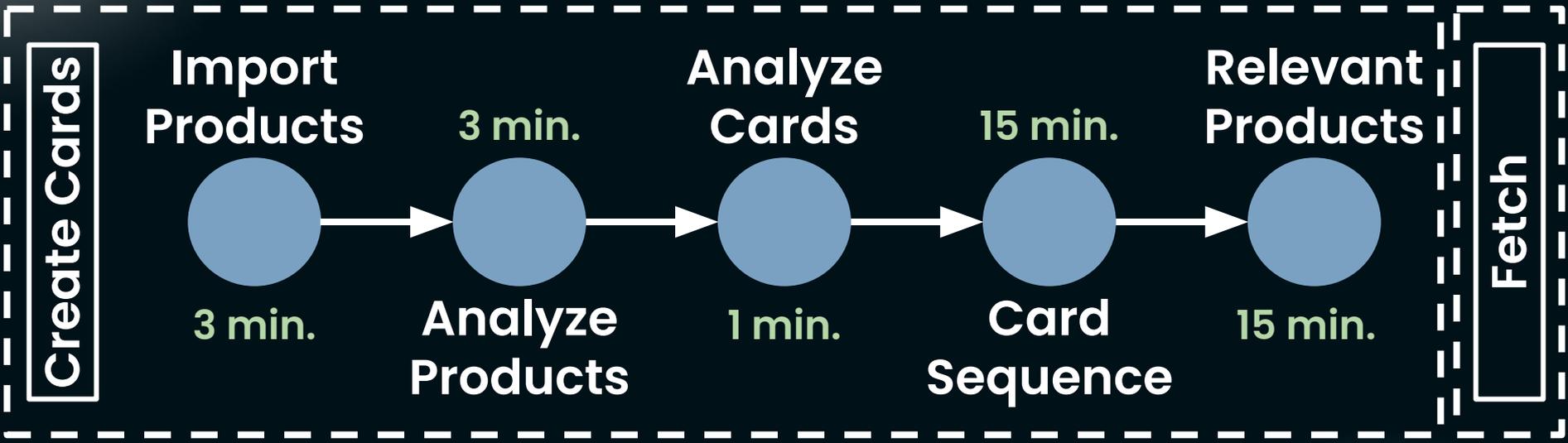
Live



Customer

Pre-Calculate

Prior

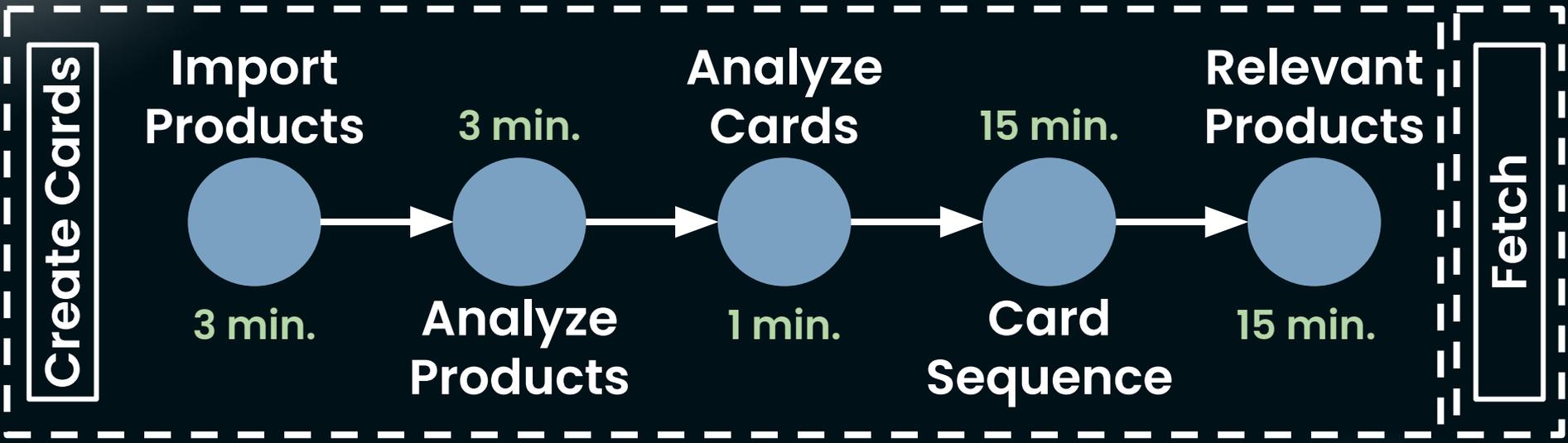


Prepare Cost

- 10 million node + 10 million leafs = 20 million documents
- $(20.000.000 \text{ Docs}/100.000) \times \$0.18 \times 6.7 = 241 \text{ DKK}$

Pre-Calculate

Prior



0.1 sec.

Live

Prior



End-User

Create Cards

Import Products



3 min.

3 min.



Analyze Products

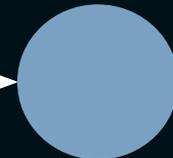
Analyze Cards



1 min.

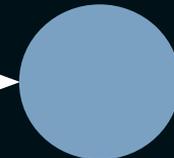
Fetch

10 min.



Card Sequence

Relevant Products



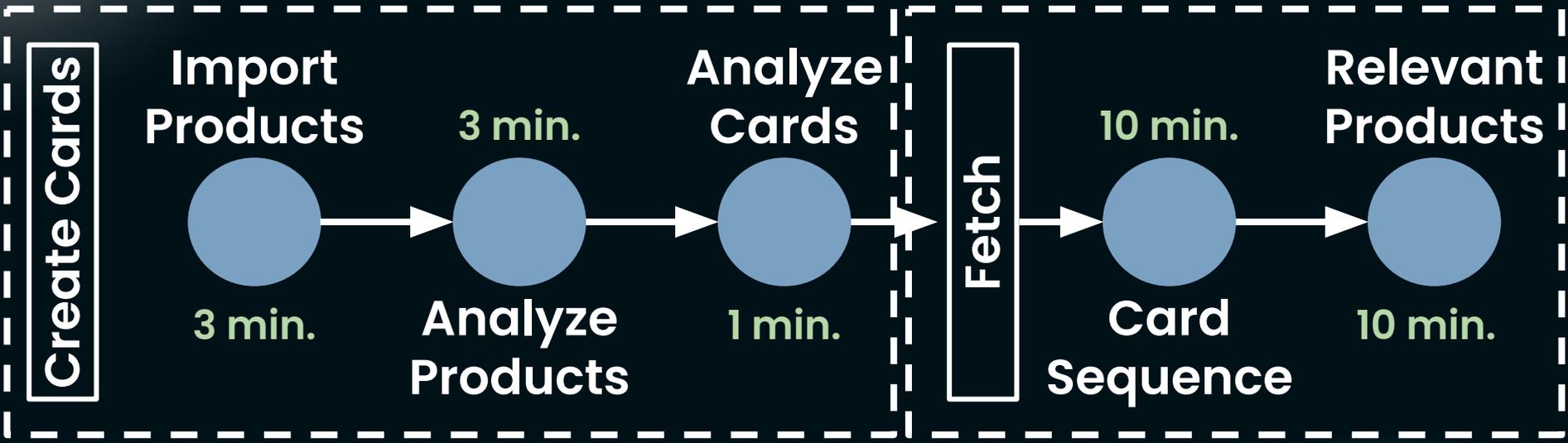
10 min.

2 sec.

Live



Customer



Search Space



End-User

Prior

Create Cards

Import Products

3 min.

3 min.

Analyze Products

Analyze Cards

1 min.

Fetch

2 sec.

Card-Product Relationships

2 min.

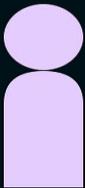
10 Sec.

Card Sequence

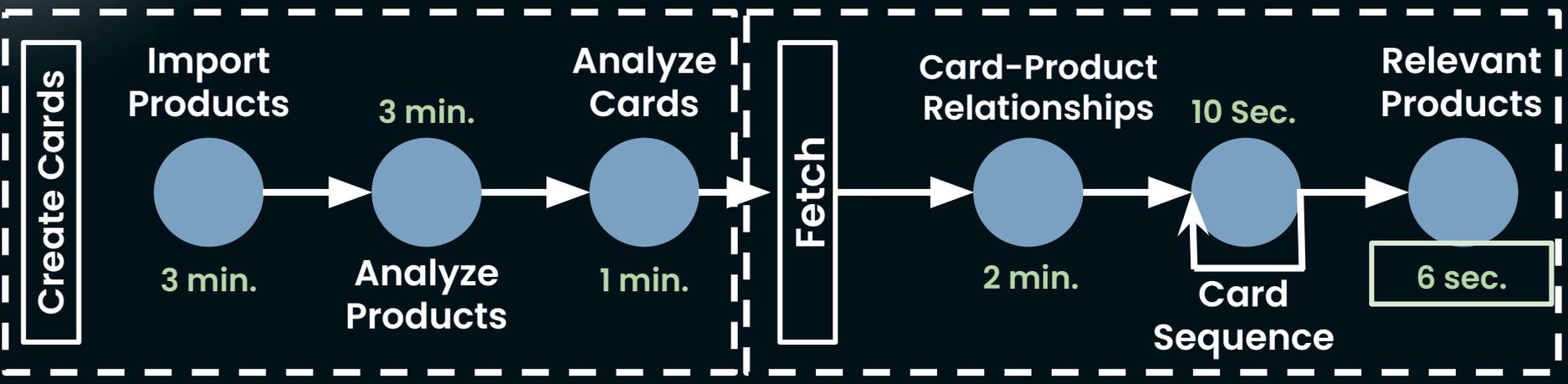
Relevant Products

6 sec.

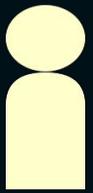
Live



Customer



Pre-Calculate



End-User

Prior

Create Cards

Import Products



3 min.

3 min.

Analyze Products



Analyze Cards



1 min.

Card-Product Relationships

3 min.



Fetch

1 sec.

10 Sec.

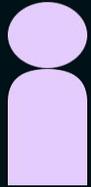
Card Sequence

Relevant Products



6 sec.

Live



Customer

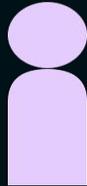
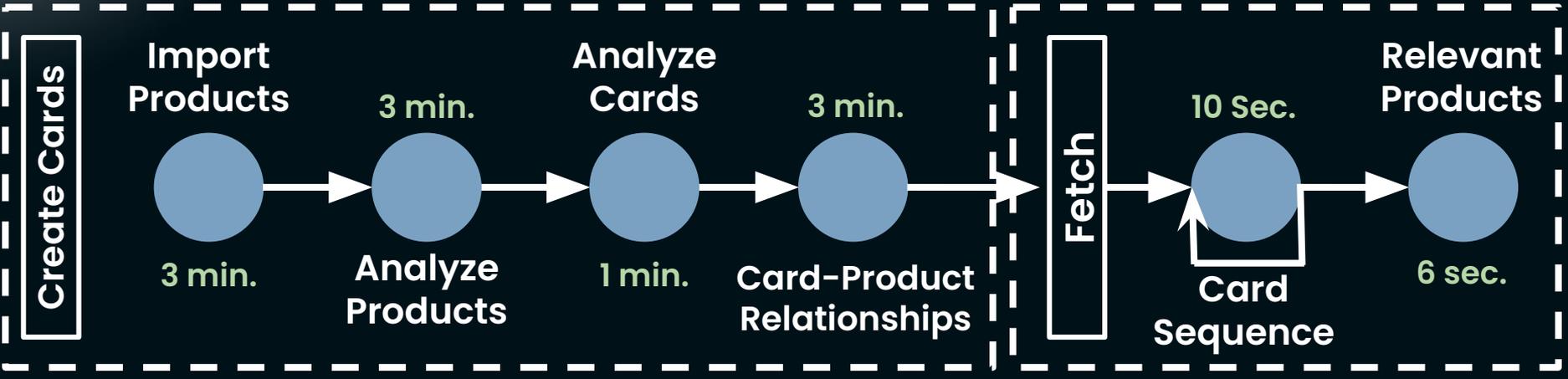
Move less Data



End-User

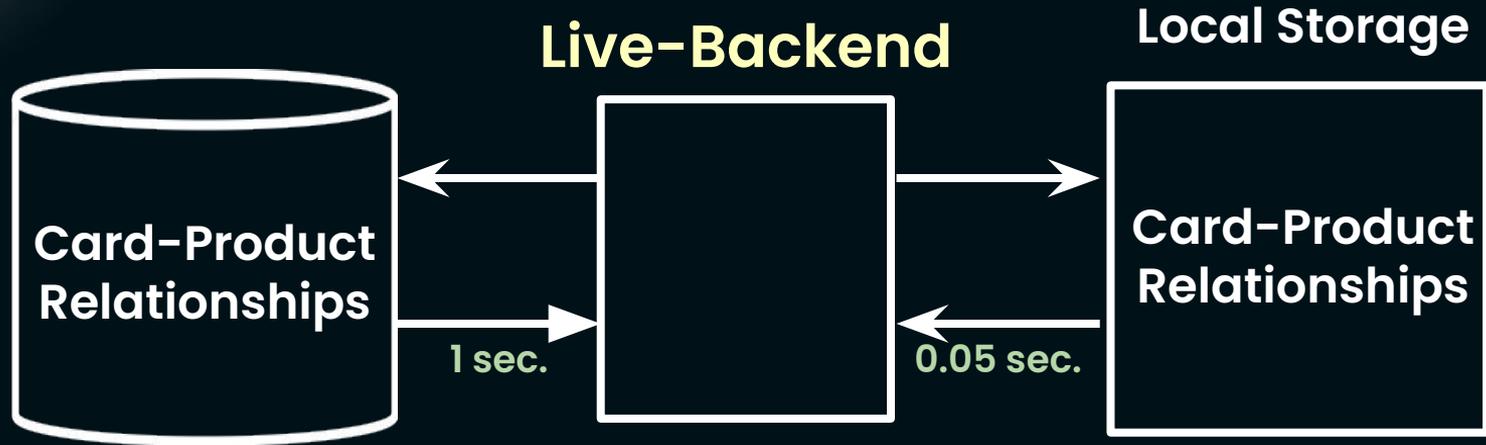
Prior

Prior-Backend Live-Backend



Customer

Live



Move less Data



End-User

Prior

Prior-Backend

Live-Backend

Create Cards

Import Products



3 min.

3 min.

Analyze Products



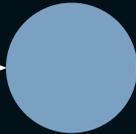
Analyze Cards



1 min.

Card-Product Relationships

3 min.



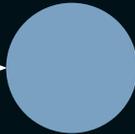
Fetch

10 Sec.



Card Sequence

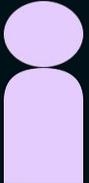
Relevant Products



6 sec.

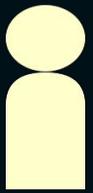
0.05 sec.

Live



Customer

Profiling



End-User

Prior

Prior-Backend

Live-Backend

Create Cards

Import Products

3 min.

3 min.

Analyze Products

Analyze Cards

1 min.

Card-Product Relationships

3 min.

Fetch

0.05 sec.

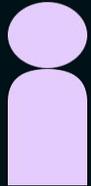
10 Sec.

Card Sequence

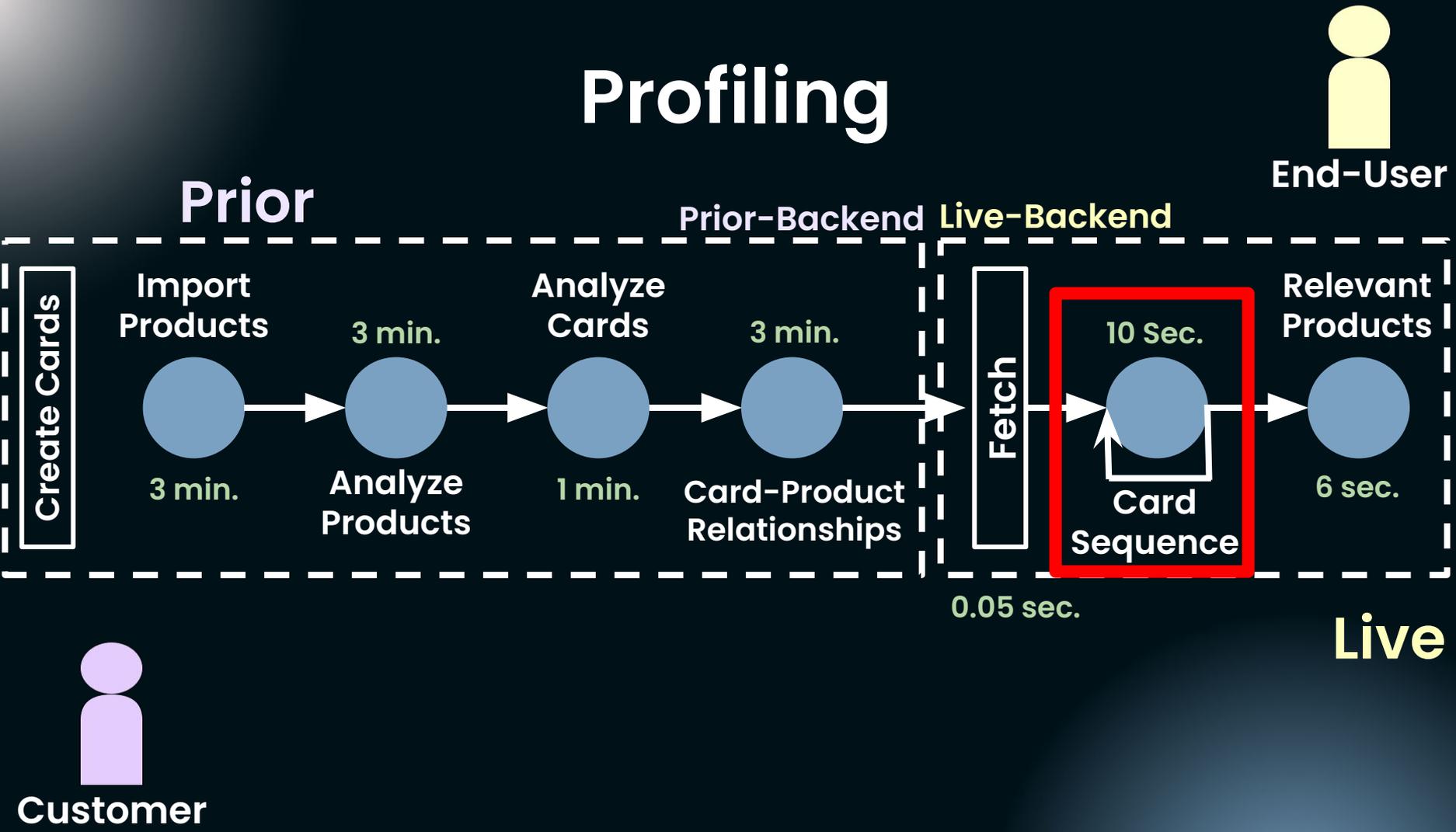
Relevant Products

6 sec.

Live



Customer

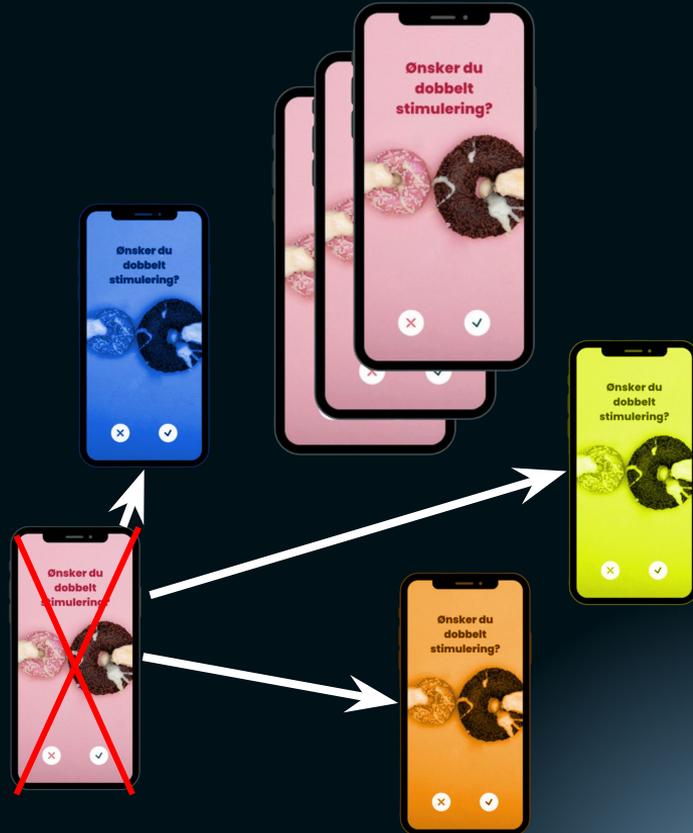
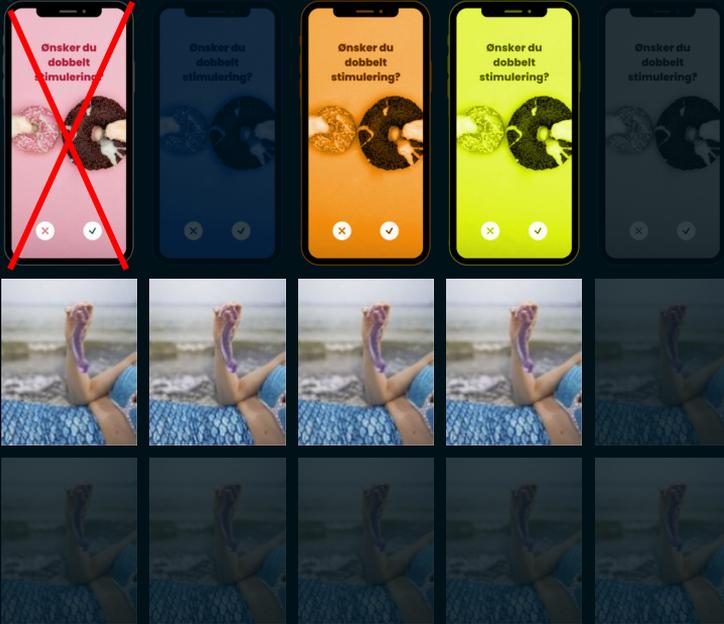


Card Sequence Profiling

- 3 Card into the future → 1-2 Card into the future
- Consider All Cards and All Products for each node
 - Minimize the Search Space

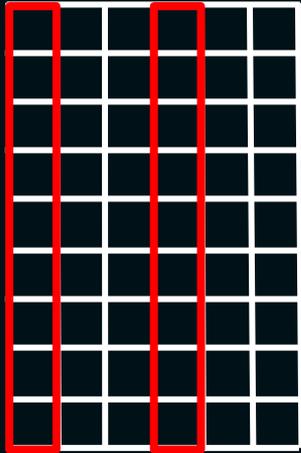
Subsequent Cards

- Less Products
- Less Cards



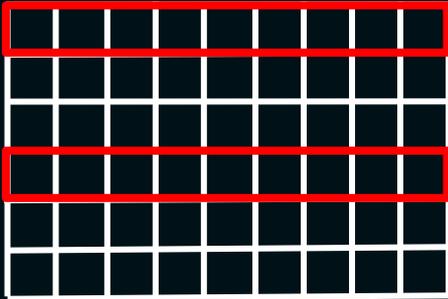
Subsequent Cards

- Less Products
- Less Cards
- Vector Access



Subsequent Cards

- Less Products
- Less Cards
- Vector Access



Profiling



End-User

Prior

Prior-Backend

Live-Backend

Create Cards

Import Products



3 min.

3 min.

Analyze Products



Analyze Cards



1 min.

Card-Product Relationships

3 min.



Fetch

0.7 Sec.

Card Sequence



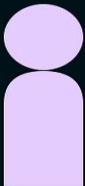
Relevant Products



6 sec.

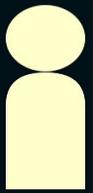
0.05 sec.

Live



Customer

Profiling



End-User

Prior

Prior-Backend

Live-Backend

Create Cards

Import Products

3 min.

3 min.

Analyze Products

Analyze Cards

1 min.

Card-Product Relationships

3 min.

Fetch

0.05 sec.

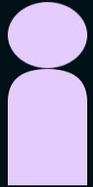
0.7 Sec.

Card Sequence

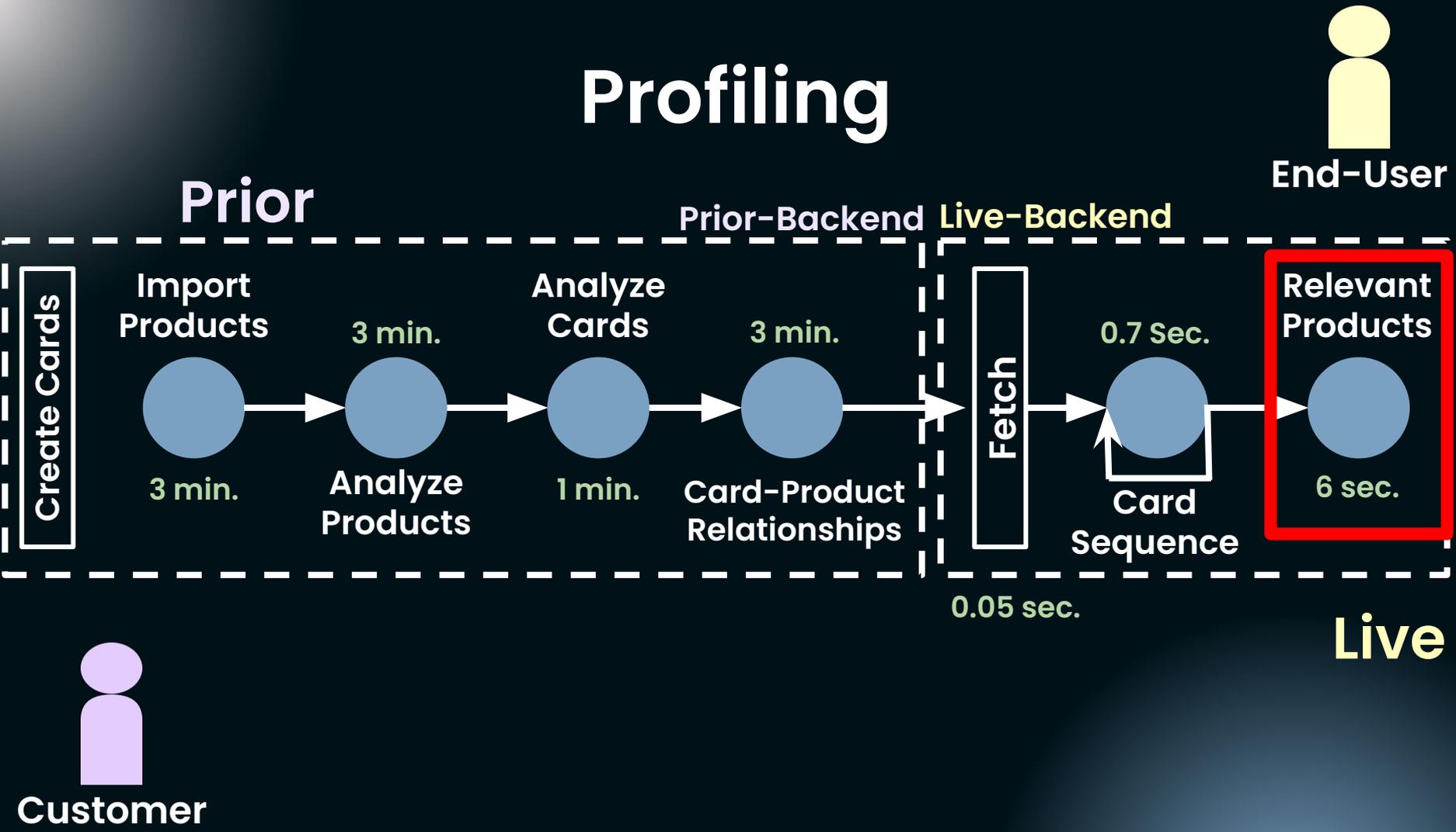
Relevant Products

6 sec.

Live



Customer



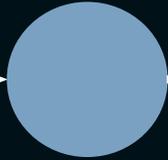
Product Recommendation Profiling



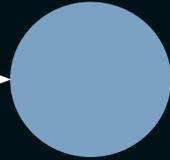
Get
Decisions



Explainable
AI



Rank
Products



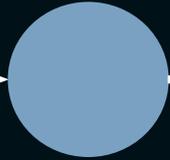
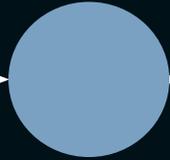
Product Recommendation Profiling

- O-Notation
- Explainable AI
 - $O(n^2)$

Product Recommendation Profiling



Get
Decisions



Calc.
Likelihood

Explainable
AI

Rank
Products

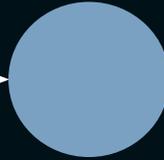
Product Recommendation Profiling

- O-Notation
- Explainable AI
 - $O(n^2)$
- Calc. Likelihood
 - $O(n)$

Product Recommendation Profiling



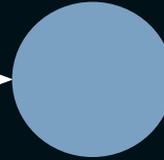
Get
Decisions



Calc.
Likelihood



Top 1000



Explainable
AI



Rank
Products

Profiling



End-User

Prior

Prior-Backend

Live-Backend

Create Cards

Import Products

3 min.

3 min.

Analyze Products

Analyze Cards

1 min.

Card-Product Relationships

3 min.

Fetch

0.05 sec.

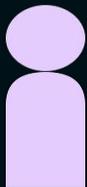
0.7 Sec.

Card Sequence

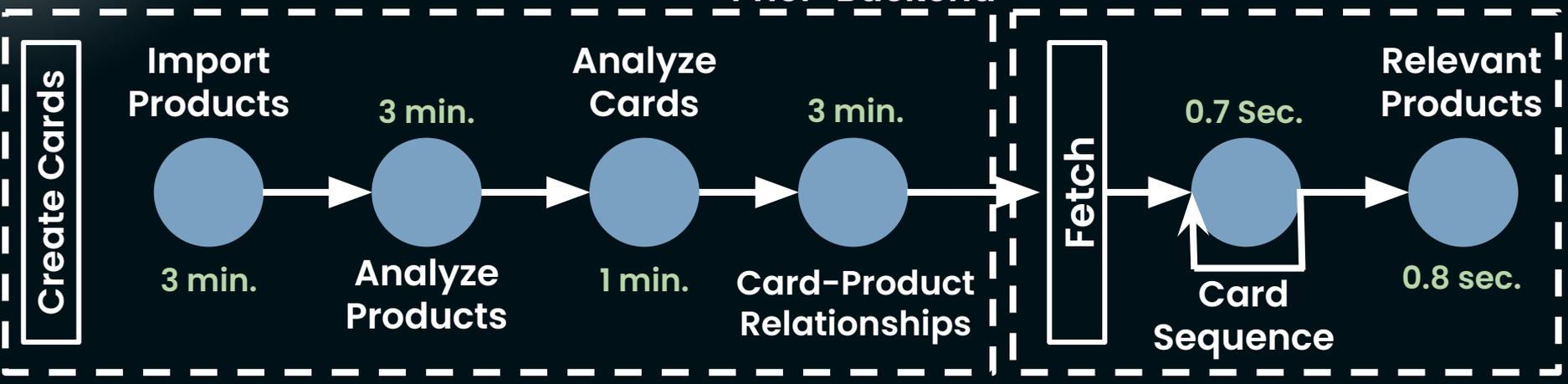
Relevant Products

0.8 sec.

Live



Customer



Summary

- Different types of optimizations
- Minimize the Search Space
- Precalculate and reuse
- Move Data / Caching
- Data Structures
- O-Notations
- Profiling
- Experience of time (UX)



Thank you!

HEAD OF AI, ROCCAI

Mathias Aamand Grønne

Mathias0909@gmail.com

+45 42450602