

Job Recommendations at XING

Data Science Portugal Meet-up,
Porto, 14th May 2018

Katja Niemann

XING 

Outline

- Challenges and Algorithms
- Building Recommender Systems
- Deployment of RecSys Models
- Q&A



RecSys Challenges and Algorithms



Fabian Abel

PREMIUM

My start page

My contacts

20

My messages

99+

My Premium

My ProJobs

Jobs

Events

6

News

1

Groups

1

Companies

Projects

Campus

Business solutions

Other services

Network news

Comments and likes

Jobs we think you'll like ▾

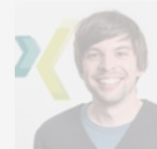
Data Scientist/Machine
Learning - Darmstadt - 75.000€

68% match

Senior Developer /
Development Expert for Ma...

74% match

>18 more job recommendations



Spread the word

Share a link or post with your contacts

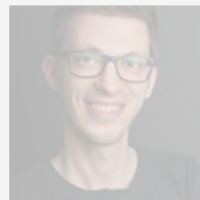
What's new?

Sort by relevance! (Alpha)



Members you may know

Show more recommendations



Stefan Imhoff

Frontend Engineer
XING AG

Add



Daniel Benjumea Martinez

Quality Assurance Manager
XING AG

Add

Profile visitors



External visitor

From www.google.co.uk



Basil Sattler

Psychologie
Justus-Liebig-Universität Gießen
Search for keywords: data science

Stephanie von Elmbach

Team Lead Technical Recruiting
XING AG

Clicked on a XING mobile app

>All visitors and statistics

start page

days



TODAY



TODAY

MAR
20MAR
21

>Birthday list

Also on XING:

Member
Charles Bernasconi
XING AGMember
Melissa Lang
Xing, AG

Job ad



Fabian Abel

PREMIUM

My start page

My contacts

20

My messages

99+

My Premium

My ProJobs

Jobs

Events

6

News

1

Groups

1

Companies

Projects

Campus

Business solutions

Other services

XING Jobs

For recruiters: [Post a job](#)

Enter a job title or keyword

Recent searches

Data Scientist

> 375 Jobs

Sexbotschaft

Hamburg

> 0 Jobs

teaching, R&D

> 249 Jobs

jobs market

Overview

Bookmarks (0)

Search alerts (0)

Jobs offering over €50,000

Jobs we think you'll like

[Rate recommendations](#)



**Data Scientist/Machine Learning -
Darmstadt - 75.000€**

Optimus Search, Darmstadt

13 days ago

68% match



**Senior Developer / Development Expert
for Machine Learning Platform Job**

SAP, Berlin

13 days ago

74% match



**Tech Lead Manager Machine Learning
(m/w)**

Michael Page, Berlin

about 1 month ago

61% match



**Senior Data Scientist Automotive &
Manufacturing Industry (m/w)**

T-Systems International GmbH, Berlin

11 days ago

71% match



Senior Data Scientist (w/m)

Schwarz Dienstleistung KG, Neckarsulm

3 days ago

68% match



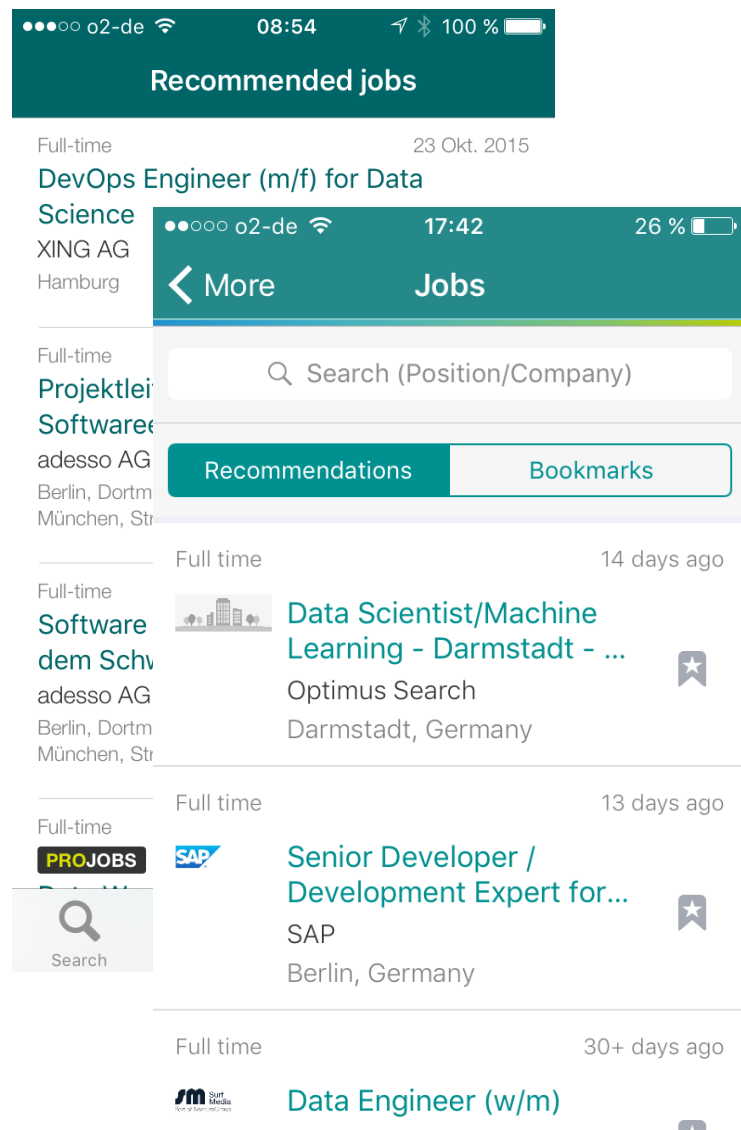
Big Data Software Engineer (m/w)

OPITZ CONSULTING Deutschland GmbH, Berli...

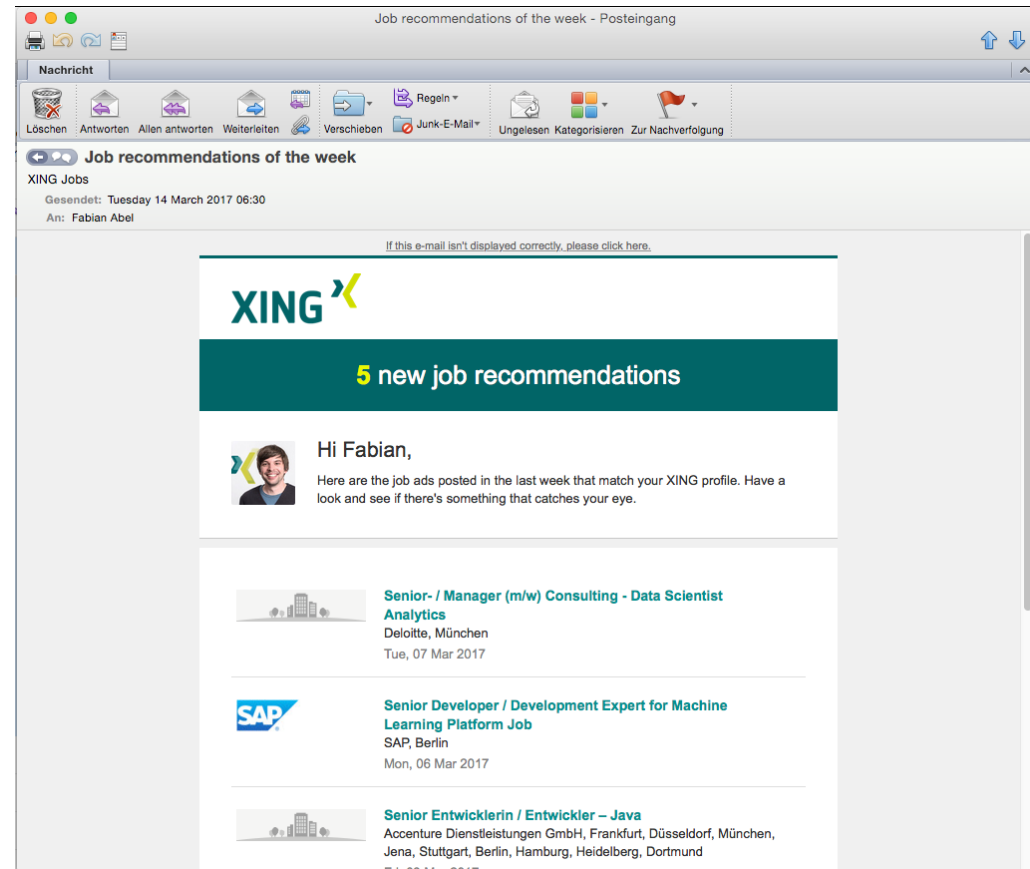
12 days ago

76% match

13 more job recommendations



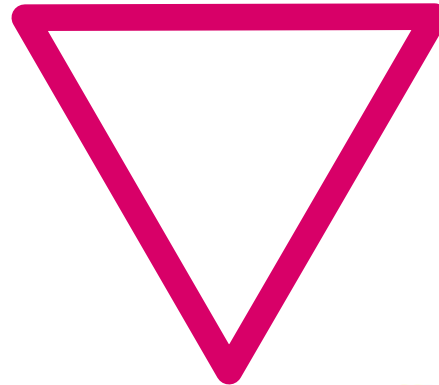
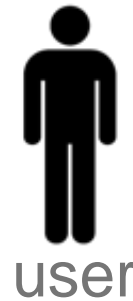
mobile



email

Goals / Triangle of contradiction

- relevant recos
- no spam



companies

- relevant candidates
- high reach

XING 

- high revenue (e.g. many clicks on paid content)
- happy customers

Key properties of a job posting



Title

Company

Employment type and career level

Full-text description

The screenshot shows a XING job posting page. The header includes the XING logo and navigation tabs: Home, Jobs, Projects, Groups, Events, and Companies. A search bar is on the right. The main content area displays the job title 'Data Warehouse Manager (m/w) Hamburg' with a date of '31 Jan 2014'. Below the title is the company name 'Goodgame Studios' with a rating of 3.87 and 'Rated by 83 employees'. The job type is 'Full time' and the location is 'Hamburg'. A sidebar on the left contains icons for various job categories. On the right, there are buttons for 'Apply by e-mail', 'I'm interested', and 'Bookmark job ad', along with links to 'Recommend job ad' and 'Show numbers and facts'. The bottom section features the Goodgame Studios logo and a detailed job description in German.

Data Warehouse Manager (m/w) Hamburg

Goodgame Studios

3.87 | Rated by 83 employees

Job type: Full time
Location: Hamburg

GOODGAME STUDIOS

Wir suchen ab sofort einen **Data Warehouse Manager (m/w)**

Goodgame Studios ist die am schnellsten wachsende Spielefirma Europas und eines der erfolgreichsten Technologieunternehmen in Deutschland. Mit mehr als 650 Mitarbeitern entwickeln wir am Standort Hamburg innovative Online Games, die von über 180 Millionen Spielern in über 200 Ländern und in 25 Sprachen weltweit gespielt werden. Die Grundlage unseres Erfolgs bildet unser talentiertes und motiviertes Team aus Experten und Neueinsteigern. Werde Teil unserer Erfolgsstory und bewirb Dich jetzt!

Als **Data Warehouse Manager (m/w)** beschäftigst Du Dich mit der Neukonzeptionierung und Wartung des zentralen Informationssystems. Du stellst neben der Konsistenz der Daten auch die Verfügbarkeit und Abfragegeschwindigkeit sicher. Dabei behältst Du stets das übergeordnete Ziel im Auge und stellst Schnittstellen für verschiedene Unternehmensbereiche zur Verfügung.

Dein Job:

- Entwicklung/Aufbau eines Data Warehouse Systems
- Weiterentwicklung und Automatisierung des aktuellen Systems
- Administration und Management der unternehmensinternen Strukturen
- Unterstützung bei der Erstellung von Analyseprogrammen zur Untersuchung umfangreicher Datensätze

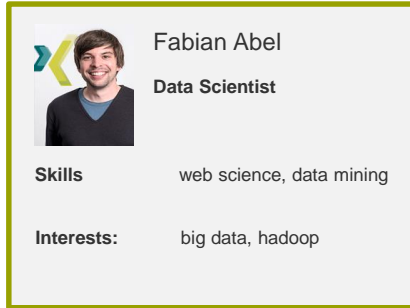
Dein Profil:

- Berufserfahrung in einer vergleichbaren Position wünschenswert
- Sehr gute Kenntnisse von Data Warehouse Systemen erforderlich
- Erfahrung in der Arbeit mit SQL und PHP
- Hohes Verantwortungsbewusstsein und selbstständige Arbeitsweise
- Sehr gute Deutsch- und Englischkenntnisse in Wort und Schrift

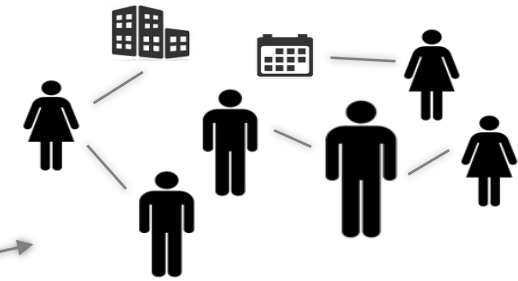
Key sources for understanding user demands



Profile



Social Network

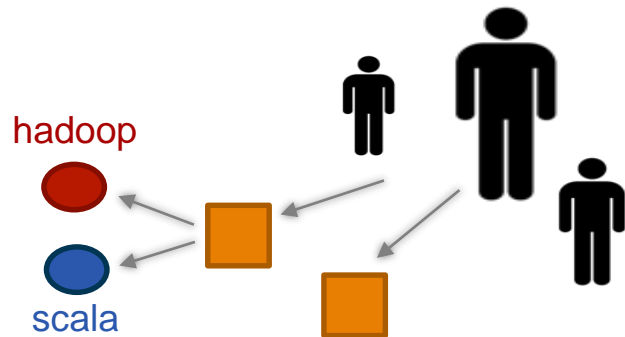


skills & co.

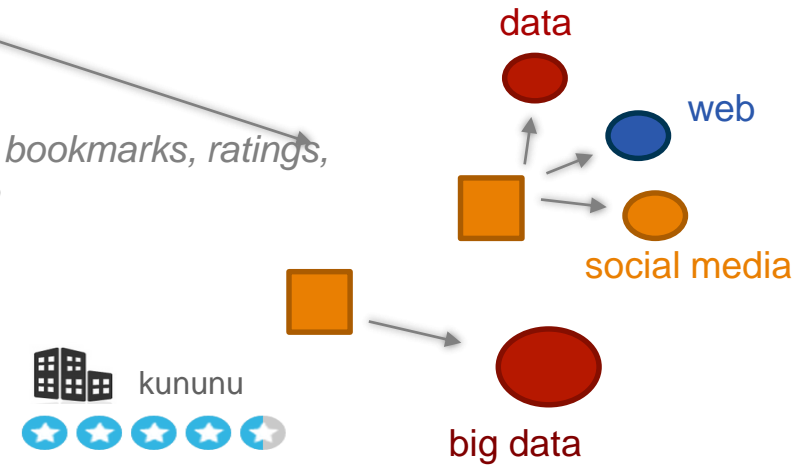
explicit and implicit connections

similar users

clicks, bookmarks, ratings, shown



Interactions of
similar users



Interactions










Recommender strategies

- Content-based filtering
 - explicit user profile
 - implicitly given interest profile / inferred attributes
- Item-to-item recommendations (more like this)
- (Pseudo) collaborative filtering

Collaborative filtering

Theory: User-based and Item-based CF

User-Item-Rating Matrix

	Java D. 	SAP Co 	Data En 	Data Sc 	BI Dev 
Anna 	3	-	4	-	2
Julia 	2	-	5	4	1
Tim 	4	3	-	5	1
John 	-	4	5	4	-

User-based CF:










- Compare users based on their ratings (e.g. cosine sim.)
- Use the n most similar users to predict a rating on an item

Item-based CF:

- Compare items based on their ratings (e.g. cosine sim.)
- Use the n most similar items to predict a rating from a user (simple weight average)

Collaborative filtering

Reality: Ultra sparse User-Item Matrix and primarily implicit feedback

	Java D. 	SAP Co 	Data En 	Data Sc 	BI Dev 
 Anna	-	-	1	-	-
 Julia	-	-	-	-	-
 Tim	-	-	-	-	-
 John	1	-	-	-	-

High level of sparsity: classical collaborative filtering (incl. matrix factorization) does not work

Collaborative filtering

Reality: Ultra sparse User-Item Matrix and primarily implicit feedback

		Java D. 	SAP Co 	Data En 	Data Sc 	BI Dev
	 Anna	-	-	1	-	-
Data Scientists 		-	-	32	18	-
Skilled in Java 		524	3	1	-	-
BI Dev 		-	-	2	4	-

Pseudo CF:

- Cluster users based on...
 - jobrole
 - skills
 - field of study
- Recommend items that similar users (= clusters) interacted with

New item problem remains...

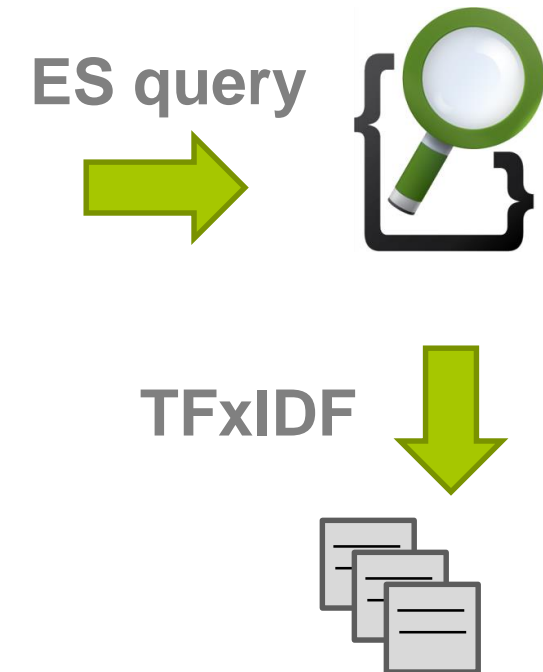
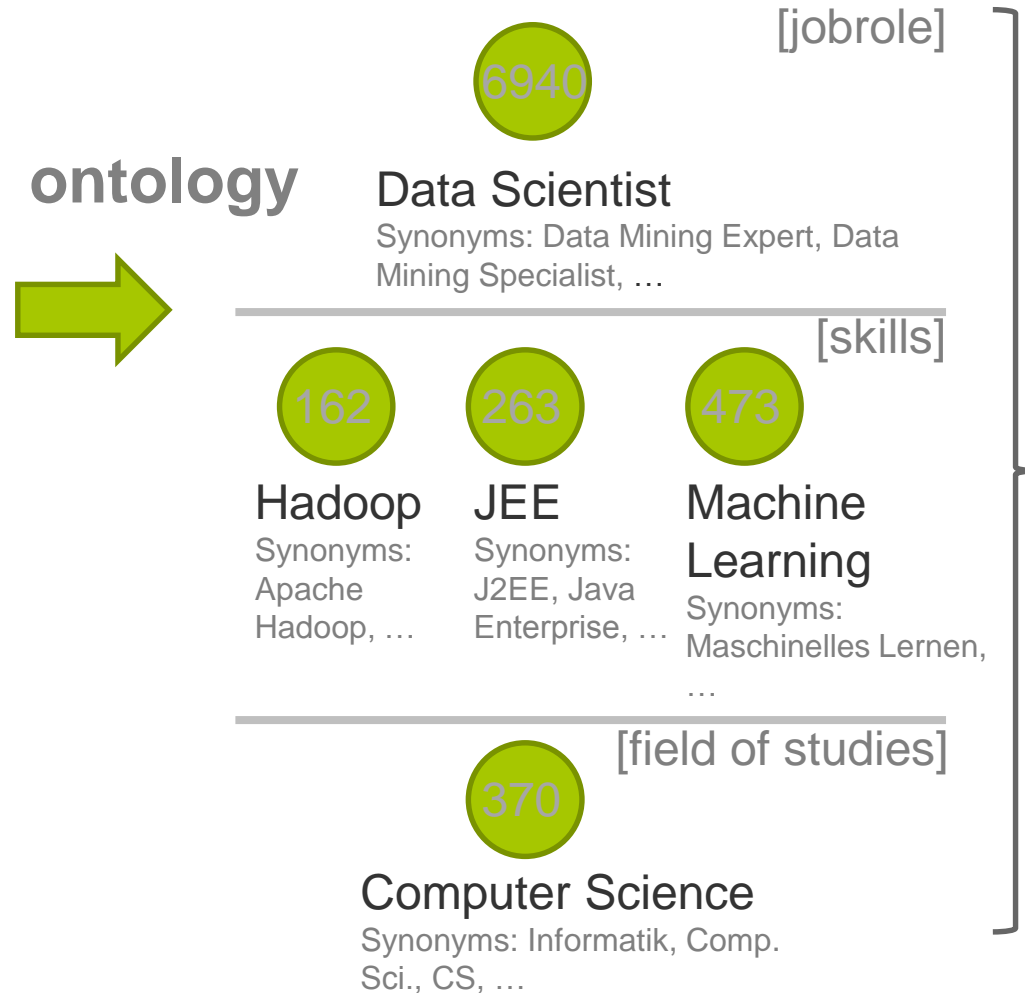
Content-based filtering

Example: semantic search



Fabian Abel
Data Mining Expert

Skills: ML, j2ee
Interests: Hadoop
Education: Informatik



Content-based filtering

Example: More-like-this component

**Bookmarked, rated
and applied-to job
postings**

1 2 3



Anna

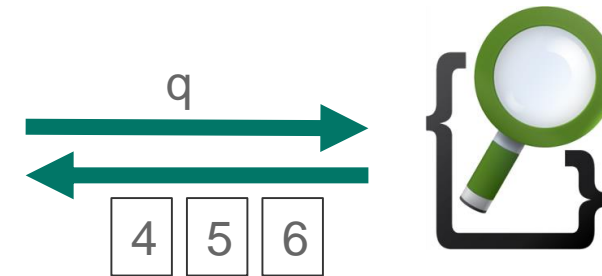
**Recommending
similar items**

$q = \text{trans} \left(\begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \right)$

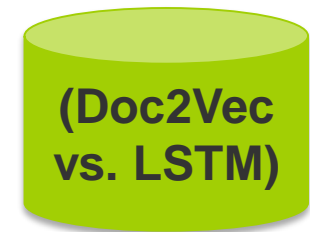
Re-rank by
similarity of topic
model vectors



5 6 4



Topic model vector
representations



People freak out if
we recommend
something wrong!

→ Try to eliminate
freakommendations
(outliers)



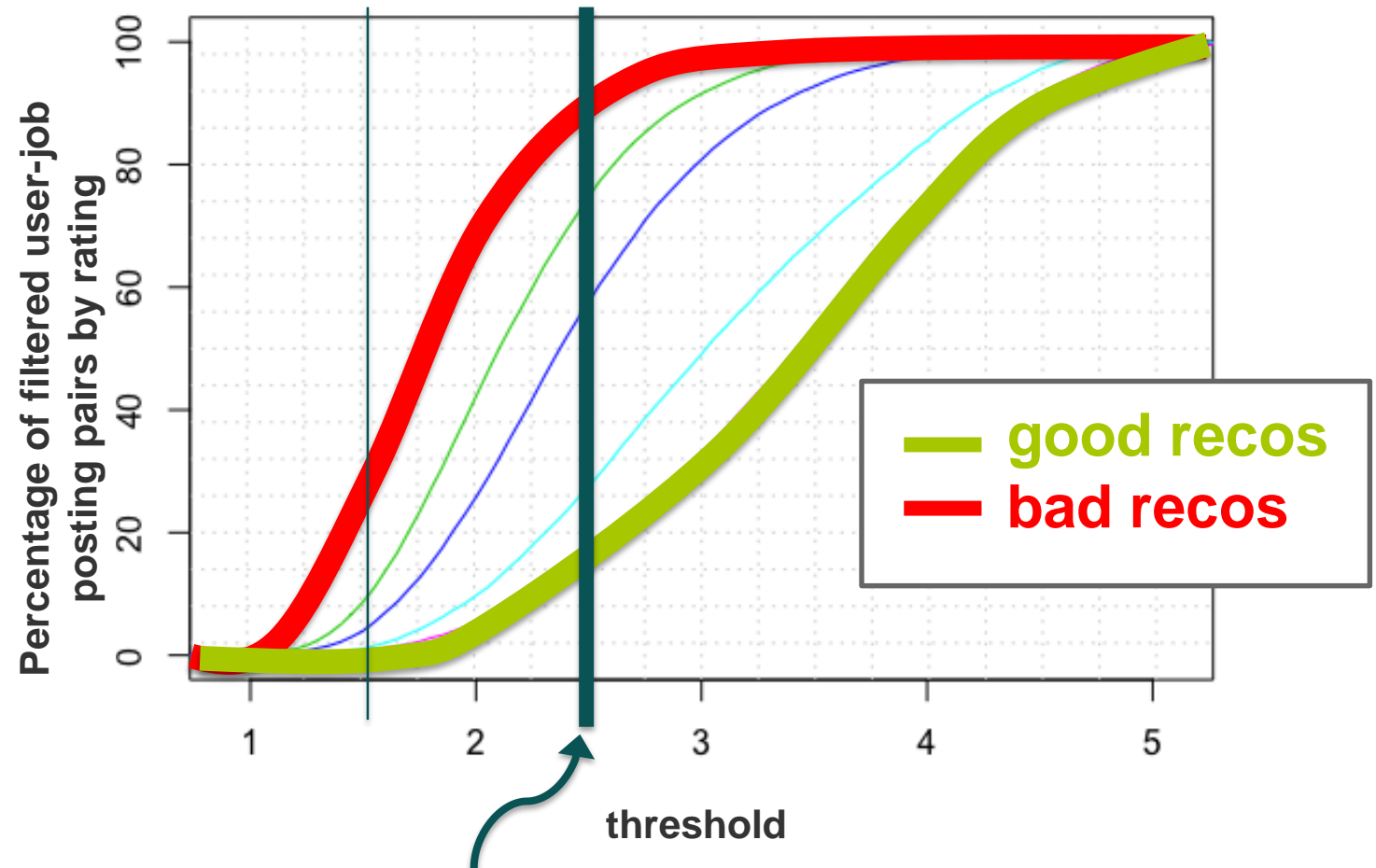
Filtering and Re-ranking

- career level filtering
- less-like-this
- same company filtering
- boosting jobs from cities where people have contacts
- boosting paid content
- outlier filtering
- ...

Outlier Filtering: Trade-off between killing bad recommendations and keeping good ones

Approach:

1. Predict a rating for each user and its top-x recommendations
2. Remove recos with a rating below the threshold



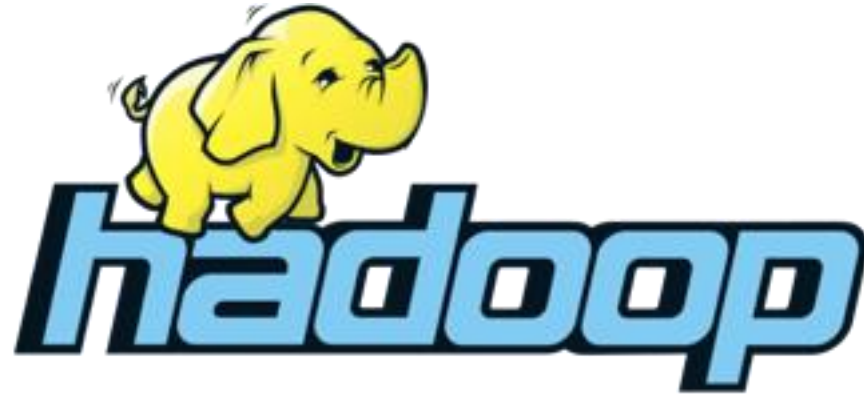
Example: with a threshold of 2.5 we kill 86% of the bad and 18% of the good recos



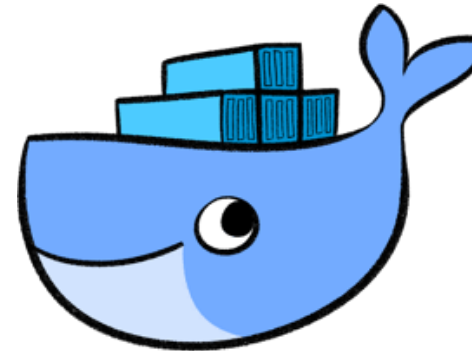
Building Recommender Systems



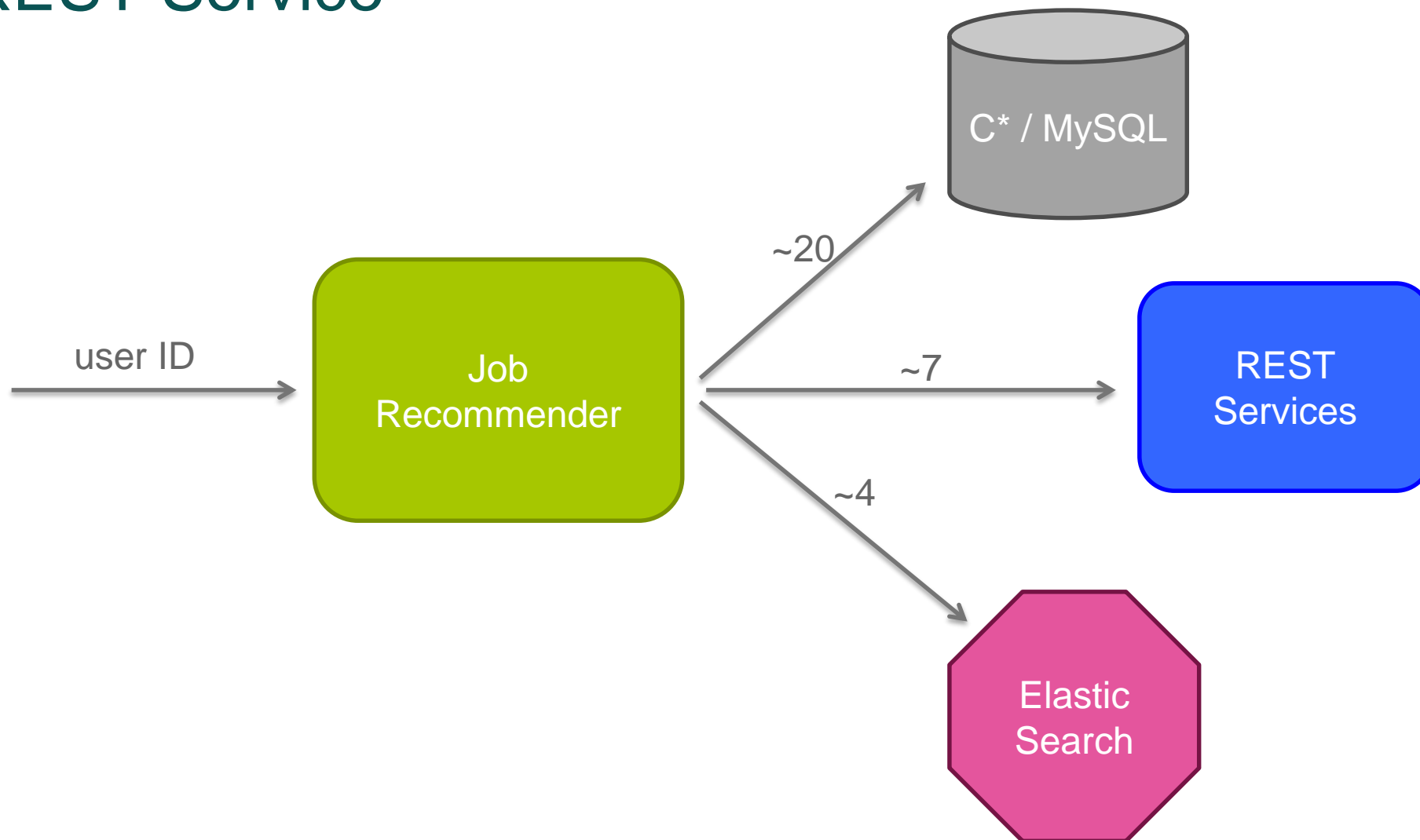
elastic



Scala



REST Service



Single threaded
this would take up
to 12
seconds!

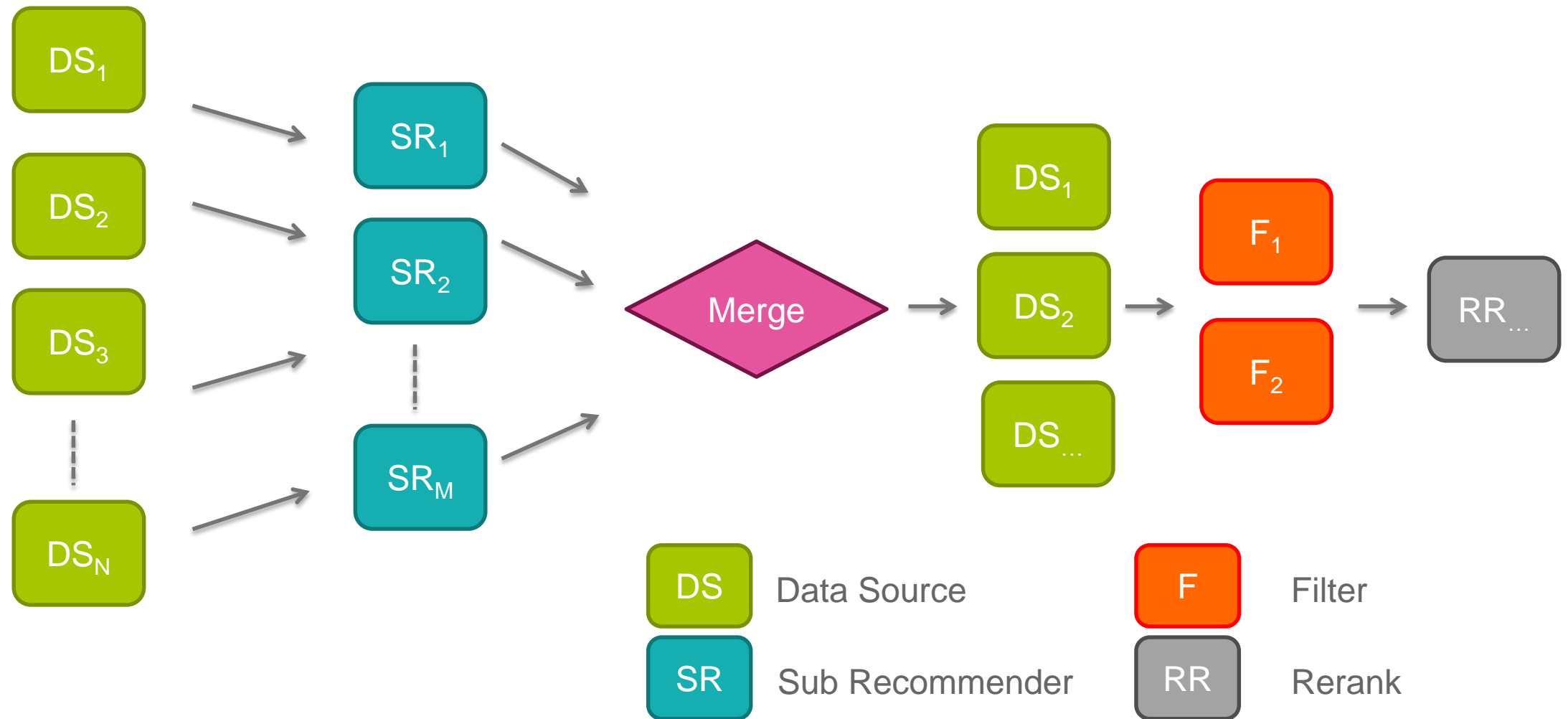


How to make it
faster?

Do computations
in parallel!



Internal recommender structure



Mixture of Online & Batch Processing

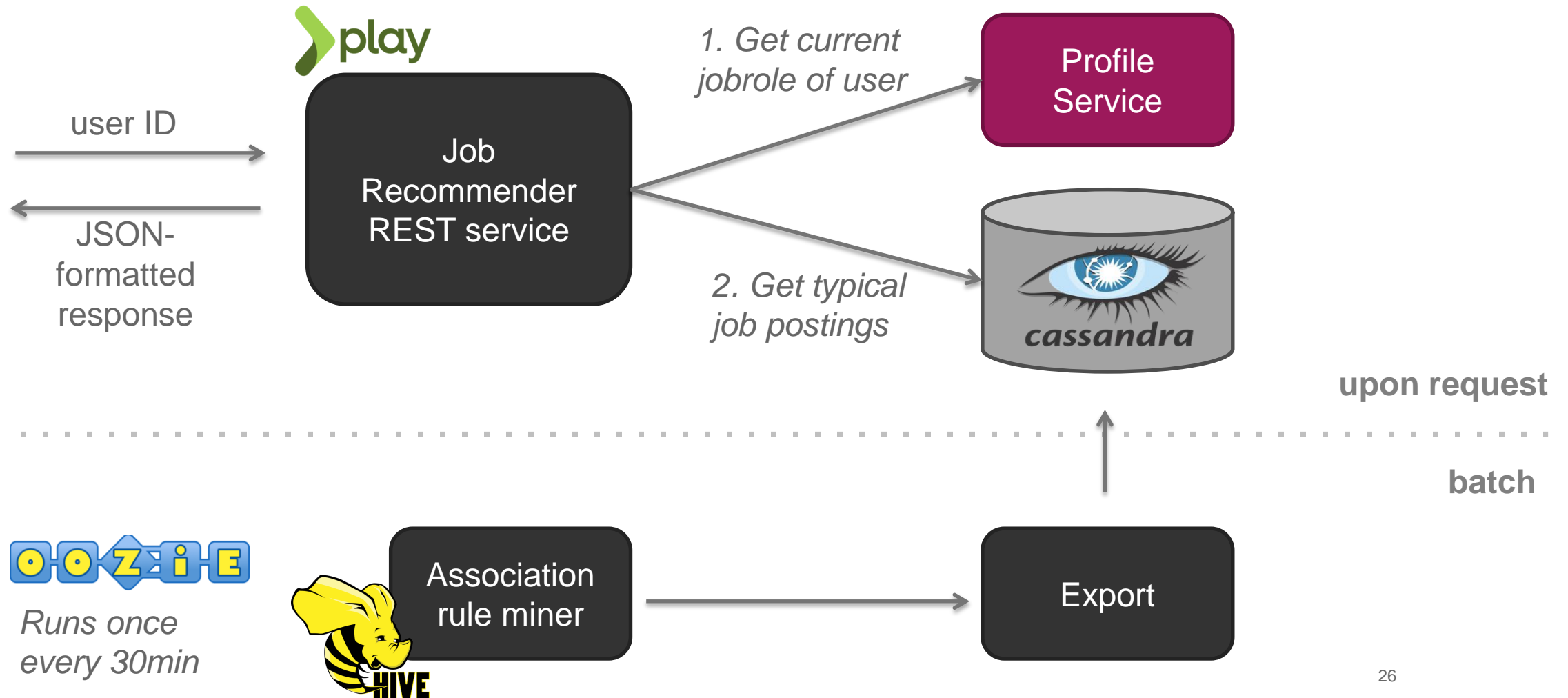
Online data

- Lookup or computation takes milliseconds
- Often required when data must be up-to-date (e.g. contextual user info)
- Examples:
 - User profile data
 - Contacts of the user
 - Negative feedback (e.g. ratings)

Pre-computed data (batch)

- Takes minutes or hours to compute
- Often required when a “complete picture” (also about other users) is necessary
- Examples:
 - Interaction-based profiles
 - Association rules
 - Topic modeling (Doc2Vec, LSTM)

Example: Users in jobrole X typically click on



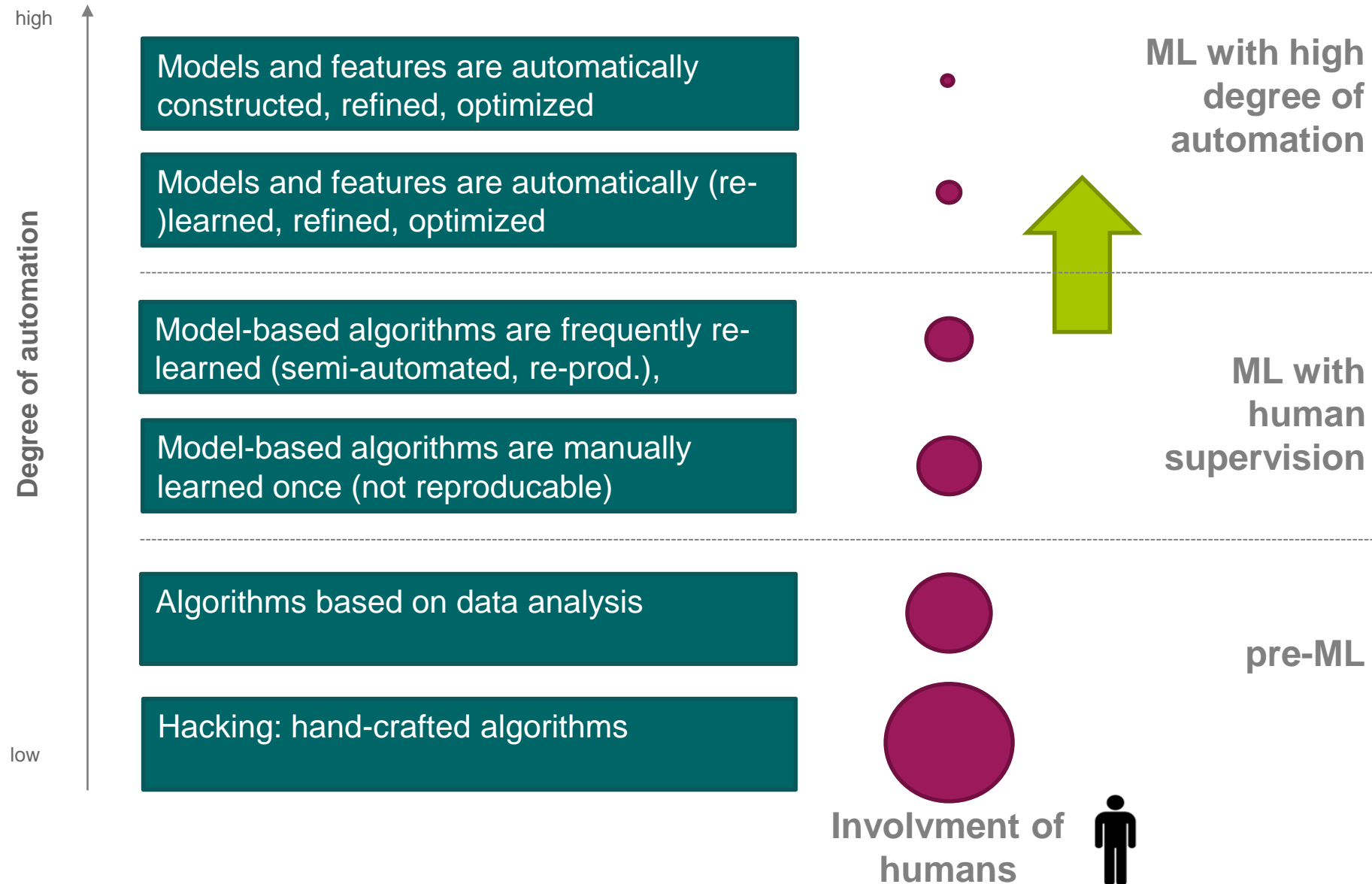


Deployment of RecSys Models

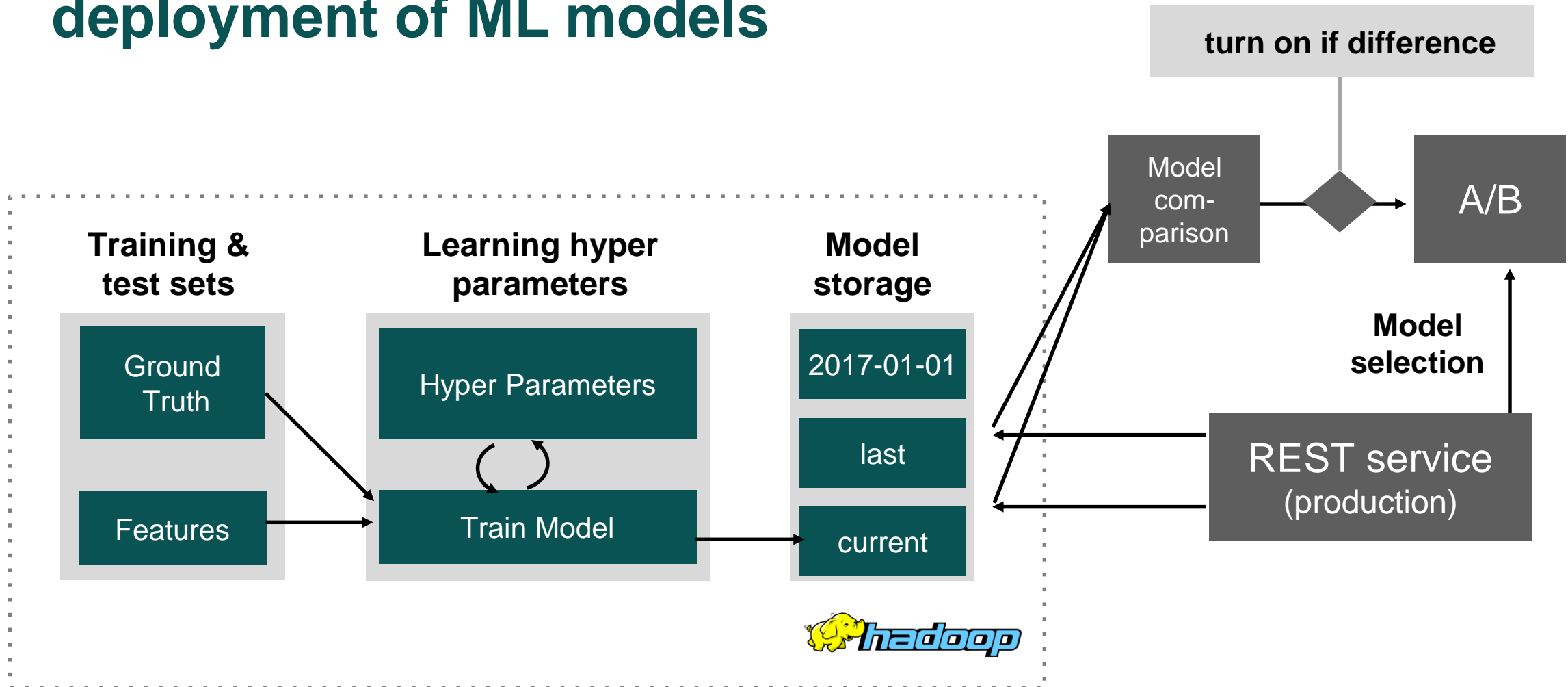
Examples of models used in production

Use Case	Type of model
Categorizing job posting	Logistic regression (>100k features)
Enrichment of job postings	Hierarchical Clustering (hierarchical K-Means)
Identifying similar job postings	Doc2Vec-based topic modeling; LSTM
Pseudo-CF	Association rule mining
Outlier filtering	XGBoost (~100 features, optimizing RMSE)
Core Ranking of job recos	XGBoost (~140 features, optimizing pair-wise loss)
Estimating willingness to change jobs	XGBoost (~90 features, optimizing cross entropy)
Estimating user preferences	Naïve Bayes
...	...

Degree of automation in ML



Automated learning & deployment of ML models



Thank you

Contact: katja.niemann@xing.com

<http://bit.ly/data-science-team>

XING 

XING is hiring

Data Engineers

and

Data Scientists

