Job Recommendations at XING

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

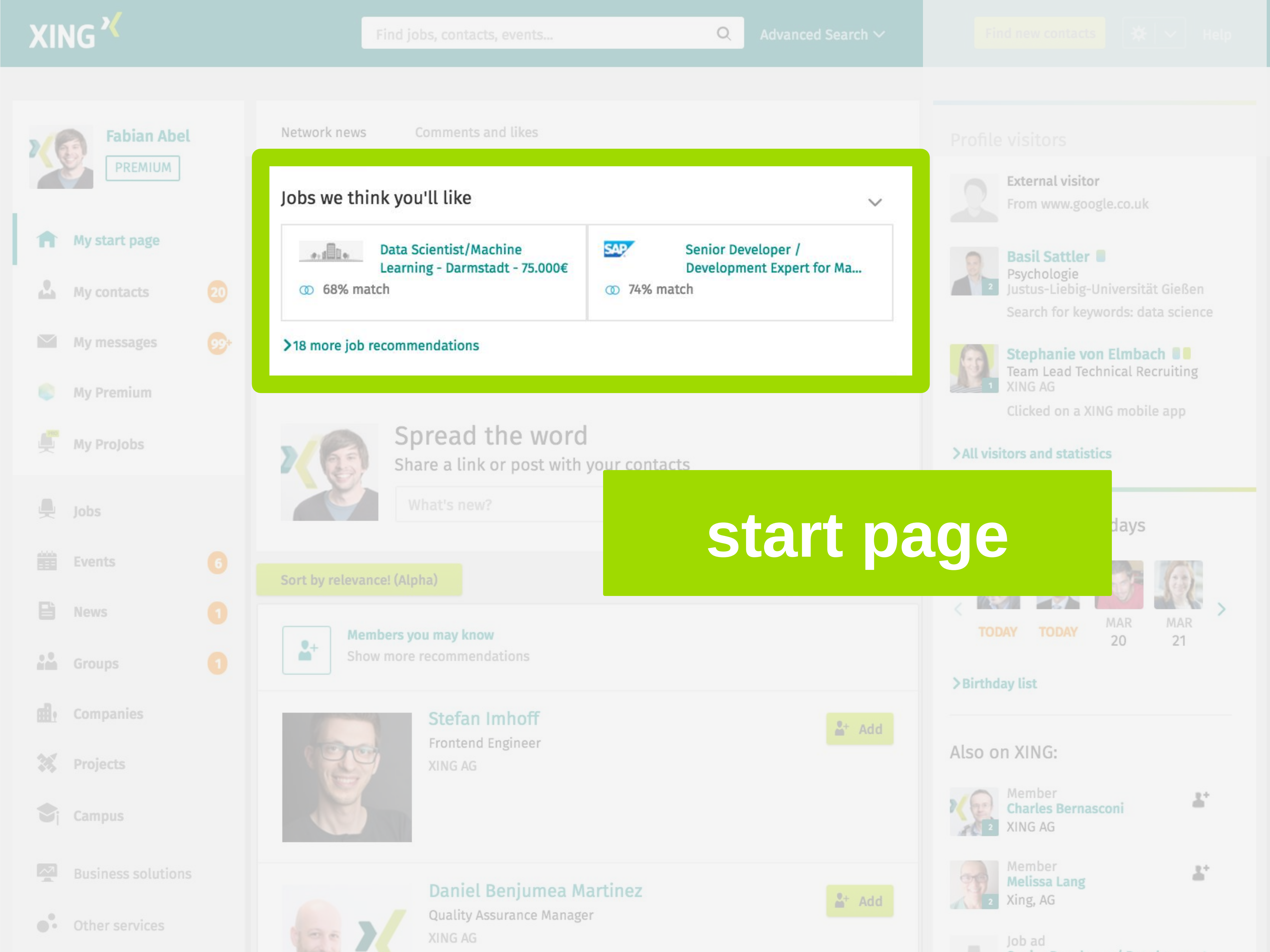
Katja Niemann

XING

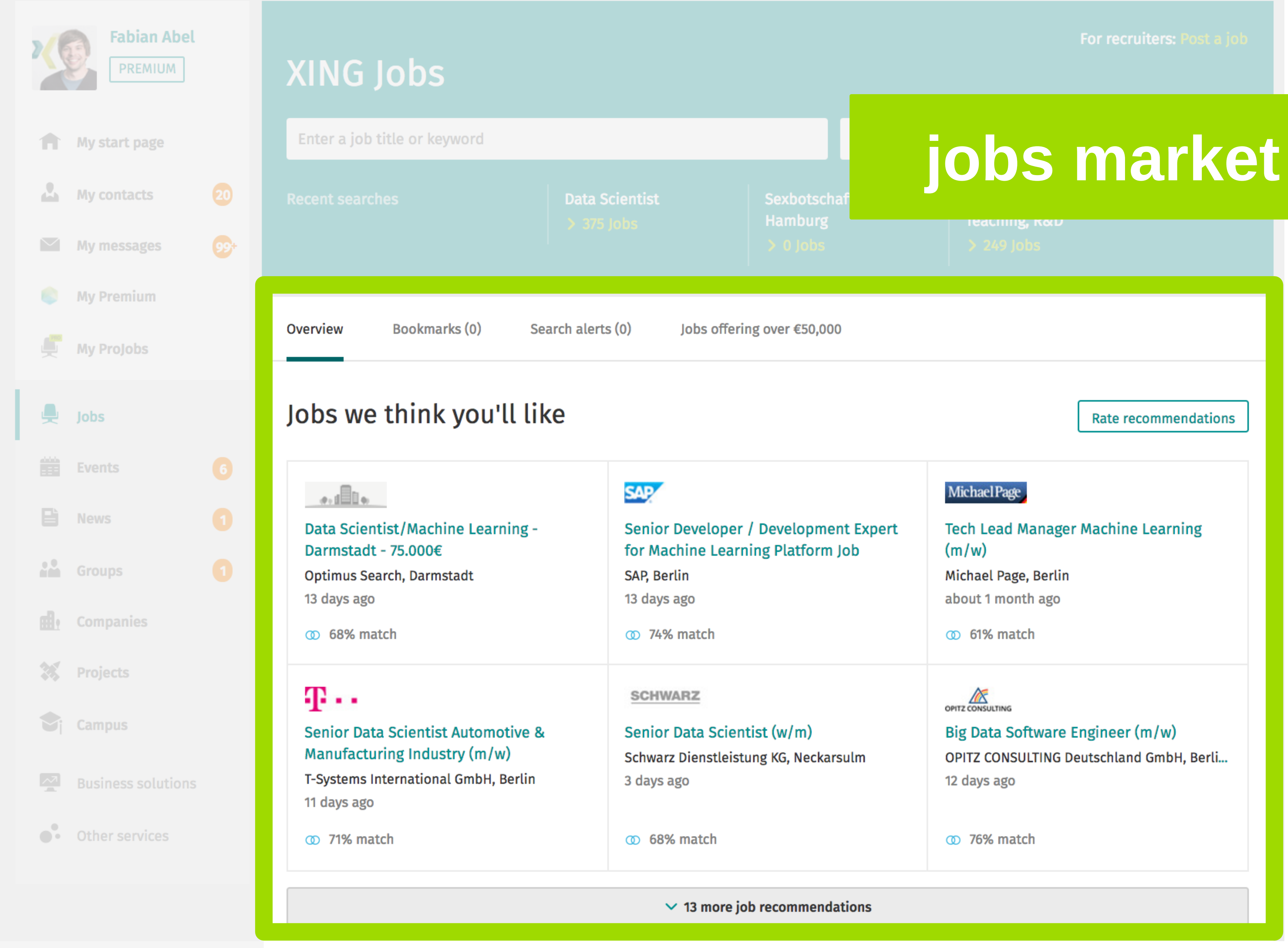
Outline

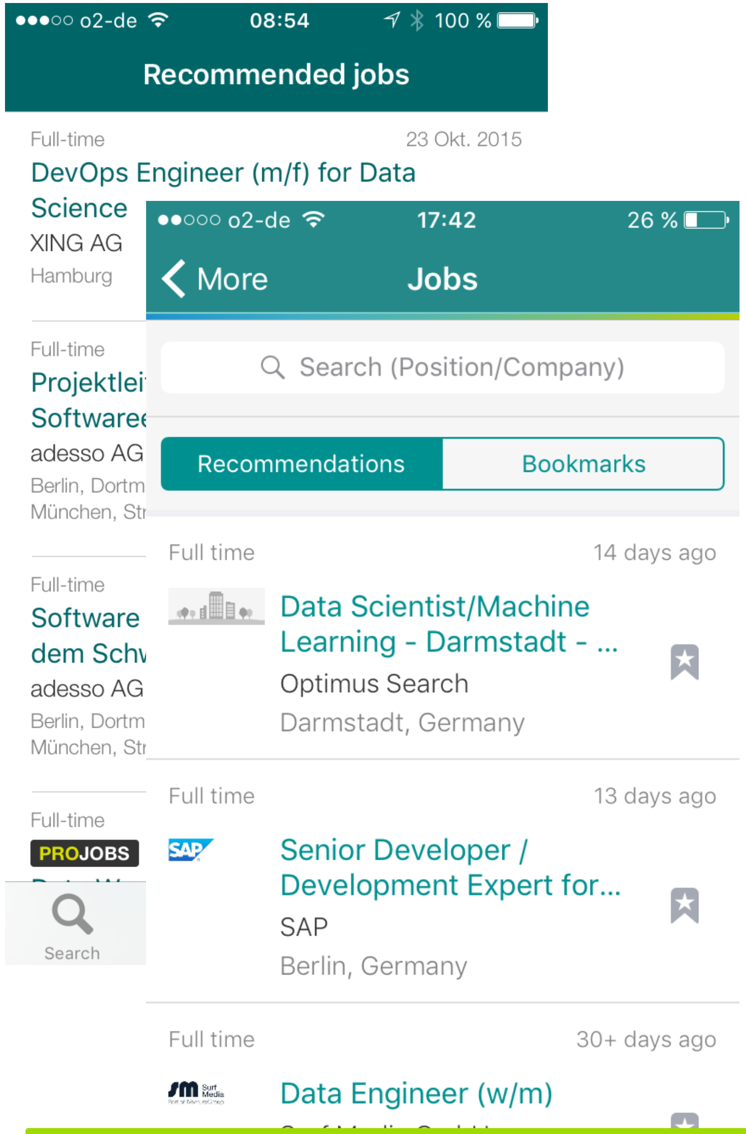
- Challenges and Algorithms
- Building Recommender Systems
- Deployment of RecSys Models
- Q&A

RecSys Challenges and Algorithms

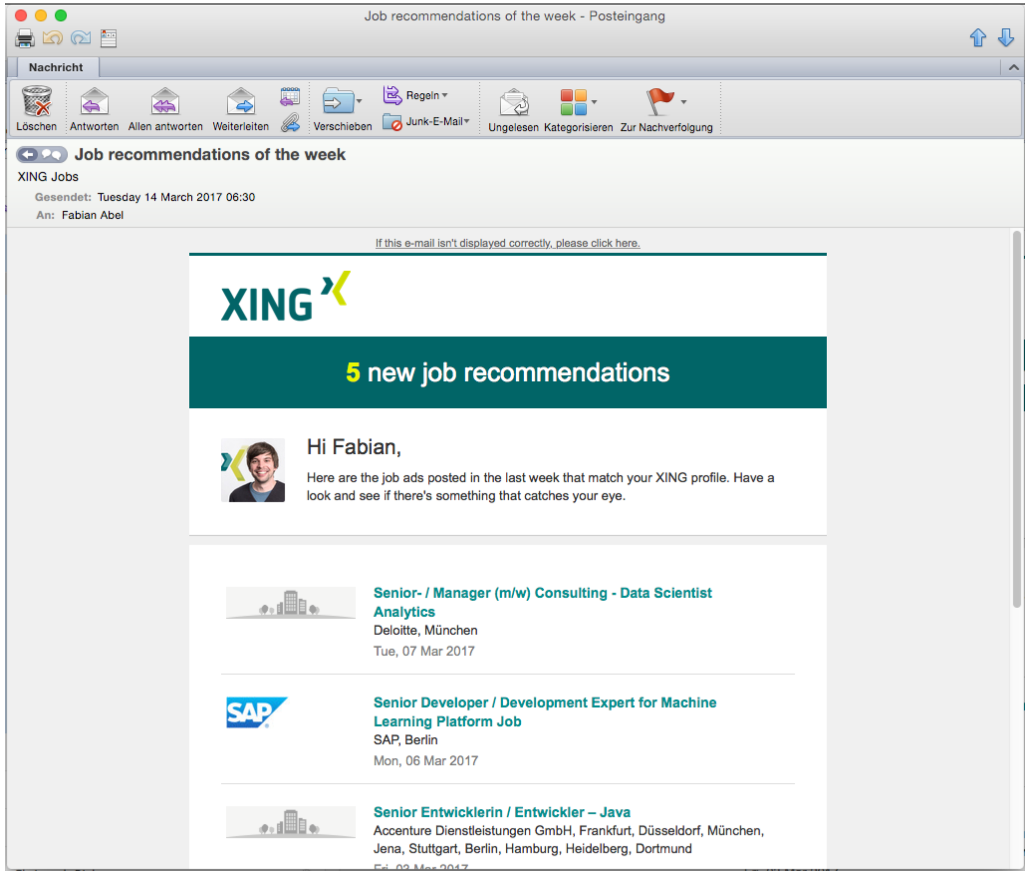

XING
Find jobs, contacts, events...
Advanced Search
Find new contacts
Help
Fabian Abel
PREMIUM
My start page
My contacts
20
My messages
99+
My Premium
My ProJobs
Jobs
Events
6
News
1
Groups
1
Companies
Projects
Campus
Business solutions
Other services
Network news
Comments and likes
Jobs we think you'll like
Data Scientist/Machine Learning - Darmstadt - 75.000€
68% match
SAP
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
days
TODAY
TODAY
MAR 20
MAR 21
Birthday list
Also on XING:
Member
Charles Bernasconi
XING AG
Member
Melissa Lang
Xing, AG
Job ad
start page

jobs market


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
For recruiters: Post a job
XING Jobs
Enter a job title or keyword
Recent searches
Data Scientist
> 375 Jobs
Sexbotschaf
Hamburg
> 0 Jobs
> 249 Jobs
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



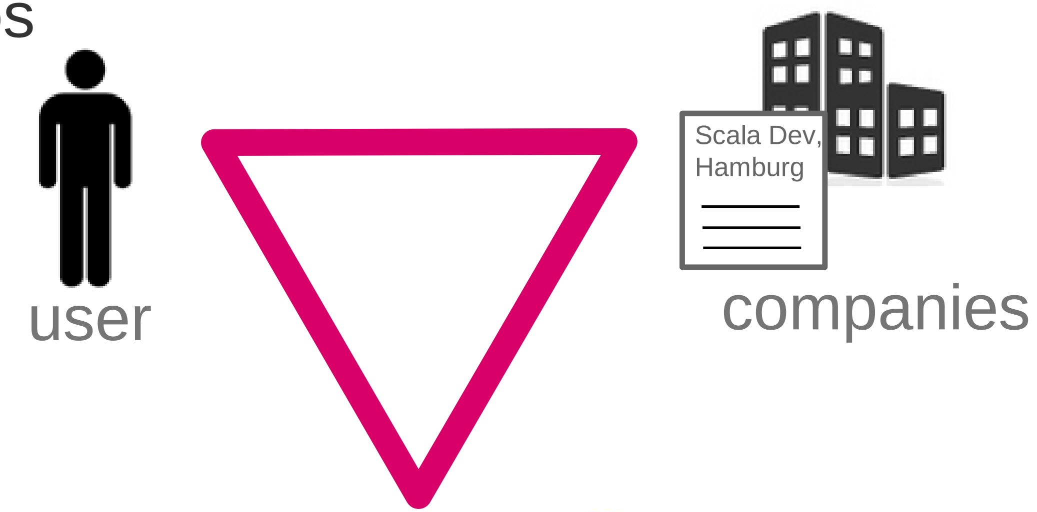
mobile



email

Goals / Triangle of contradiction

- relevant recos
- no spam



- relevant candidates
- high reach

XING

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

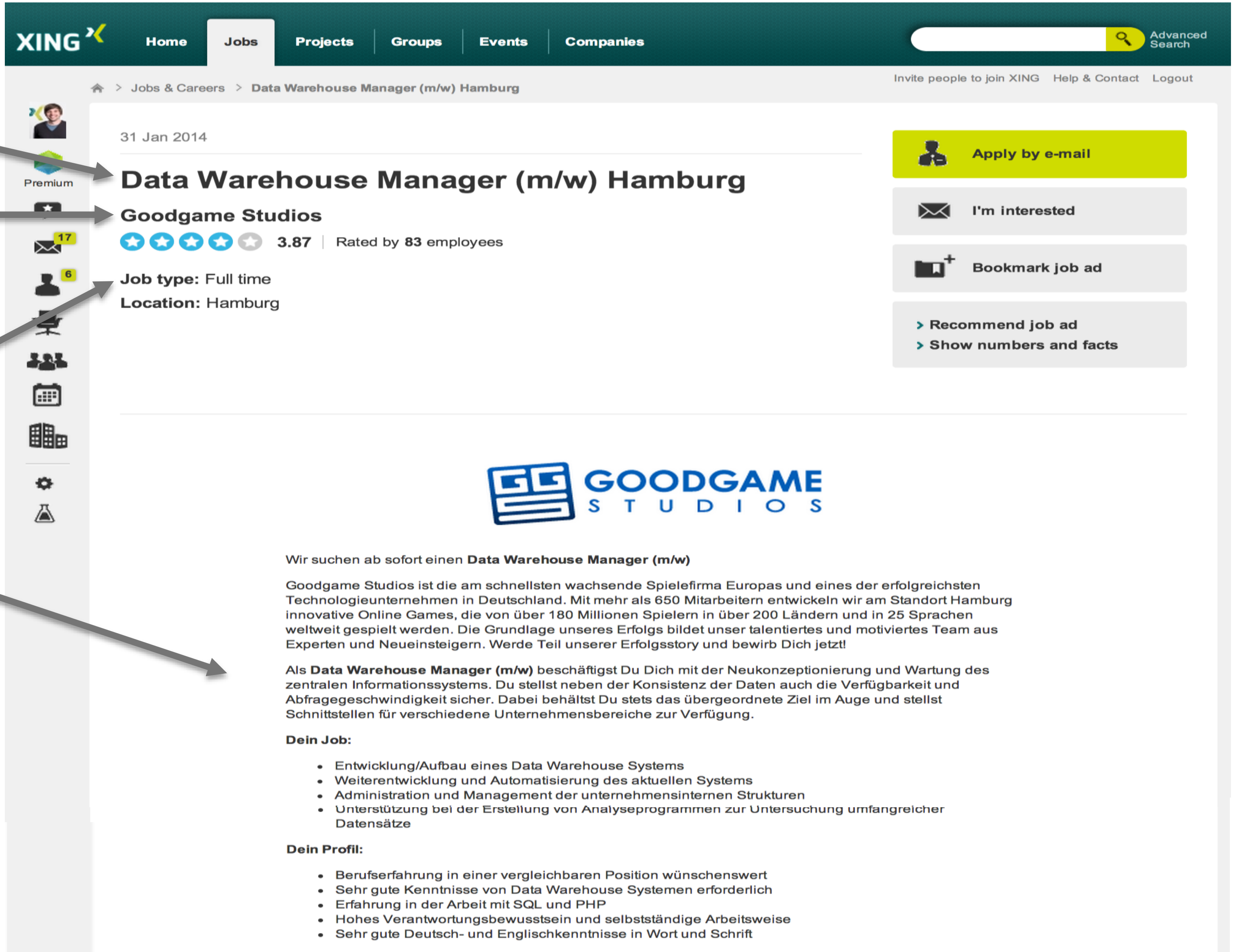
Key properties of a job posting

Title

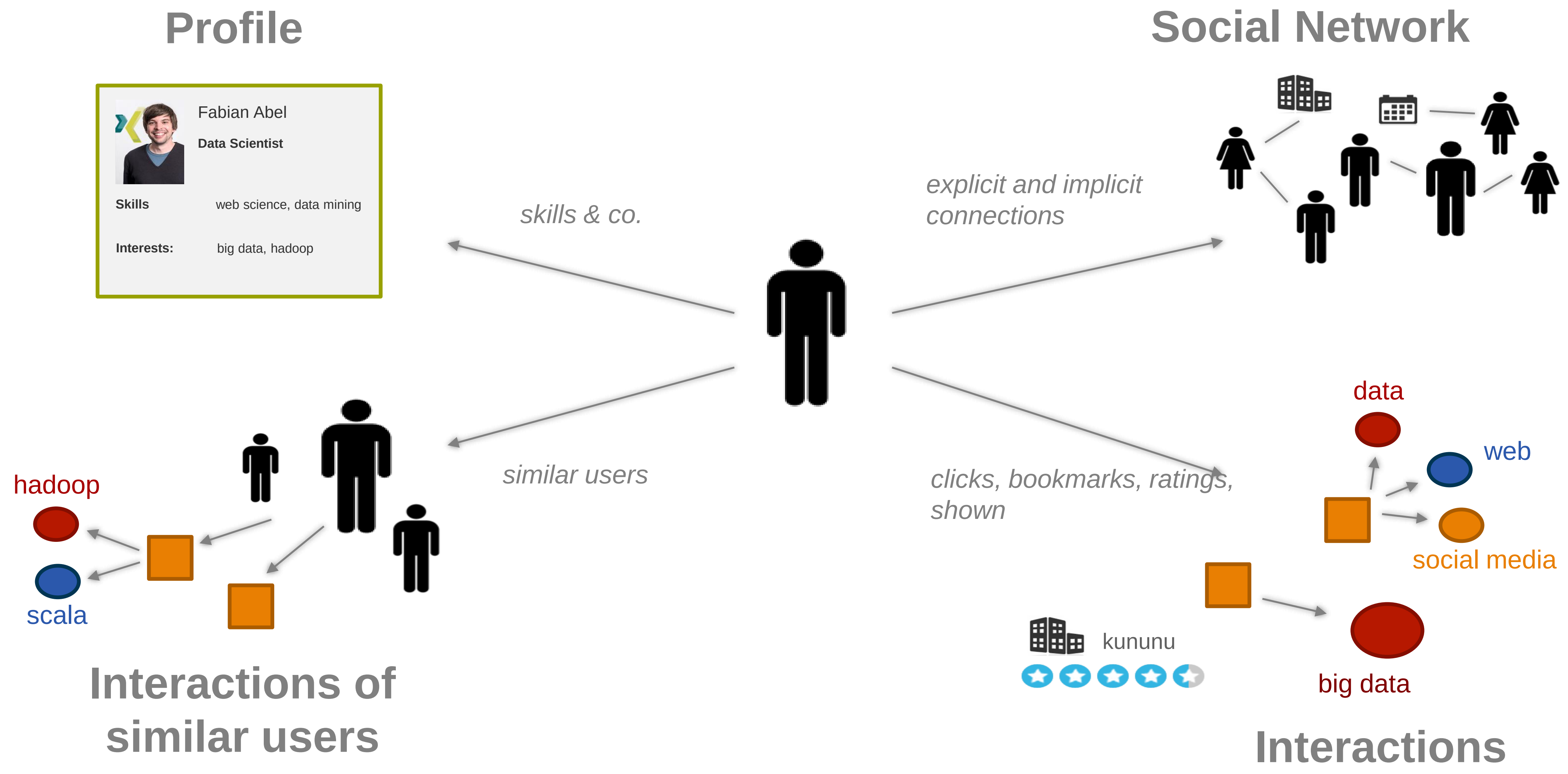
Company

Employment type and career level

Full-text description



Key sources for understanding user demands



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 fitering (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

ontology

[jobrole]

6940

Data Scientist

Synonyms: Data Mining Expert, Data Mining Specialist, …

[skills]

162

Hadoop

Synonyms: Apache Hadoop, …

263

JEE

Synonyms: J2EE, Java Enterprise, …

473

Machine Learning

Synonyms: Maschinelles Lernen, …

[field of studies]

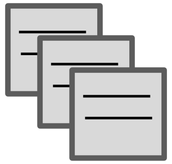
370

Computer Science

Synonyms: Informatik, Comp. Sci., CS, …

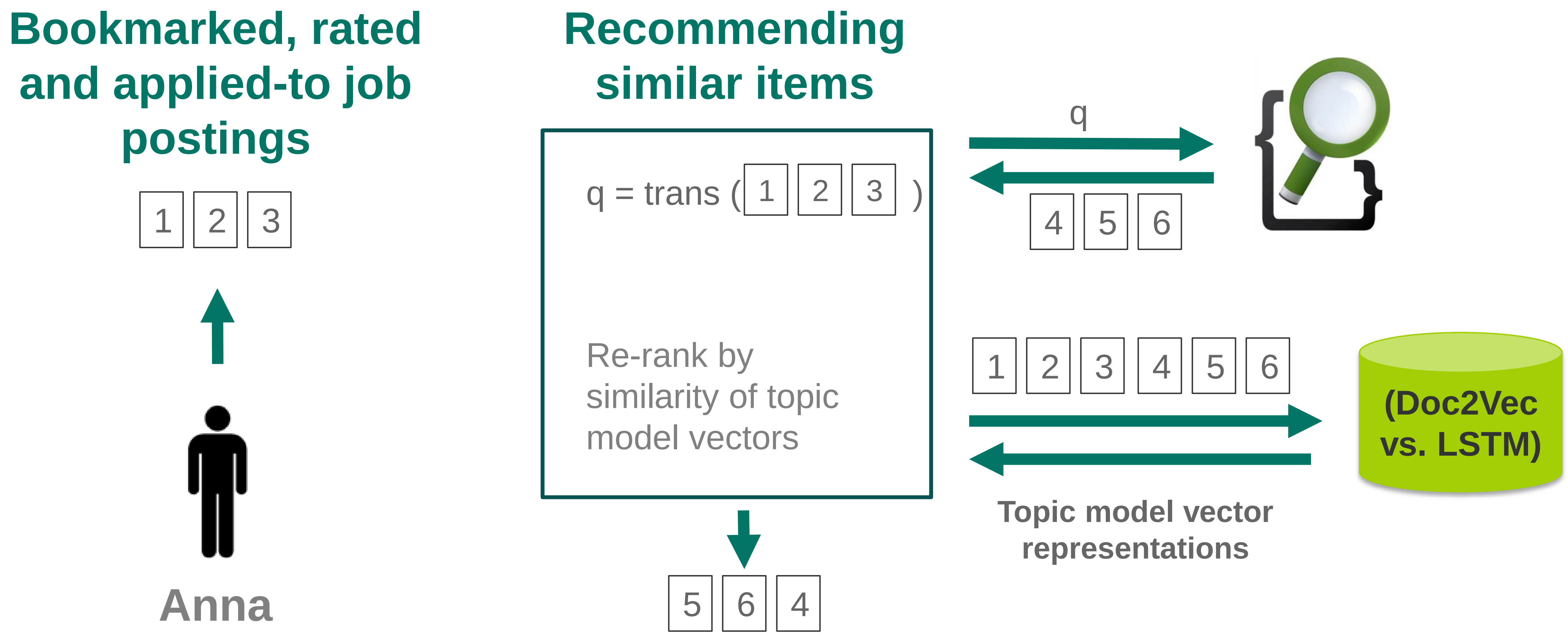
ES query

TFxIDF



Content-based filtering

Example: More-like-this component



People freak out if we recommend something wrong!

→ Try to eliminate freakommendations (outliers)

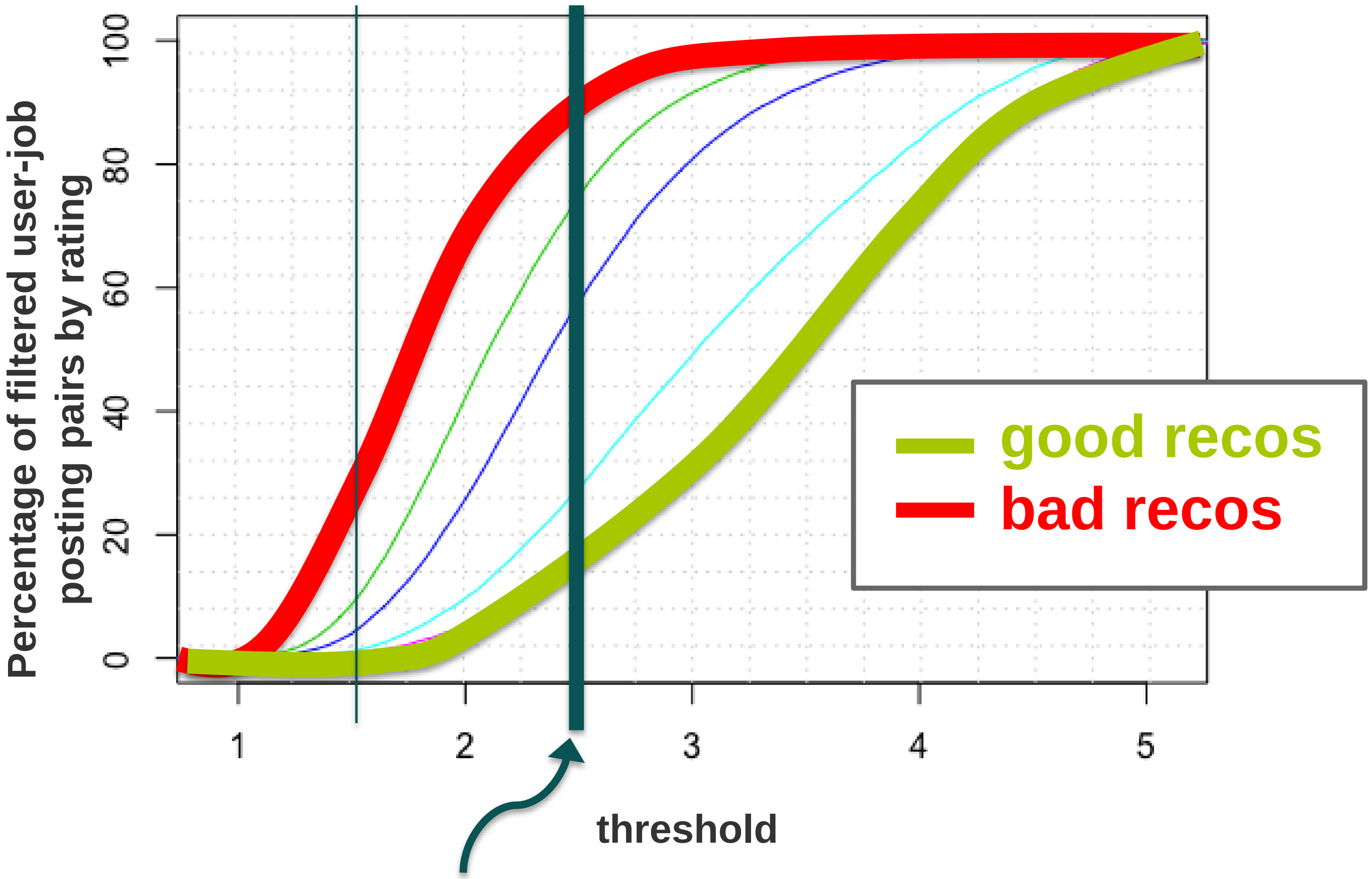
Filtering and Re-ranking

- 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

Bulding Recommender Systems

elastic

hadoop

HIVE

Scala

cassandra

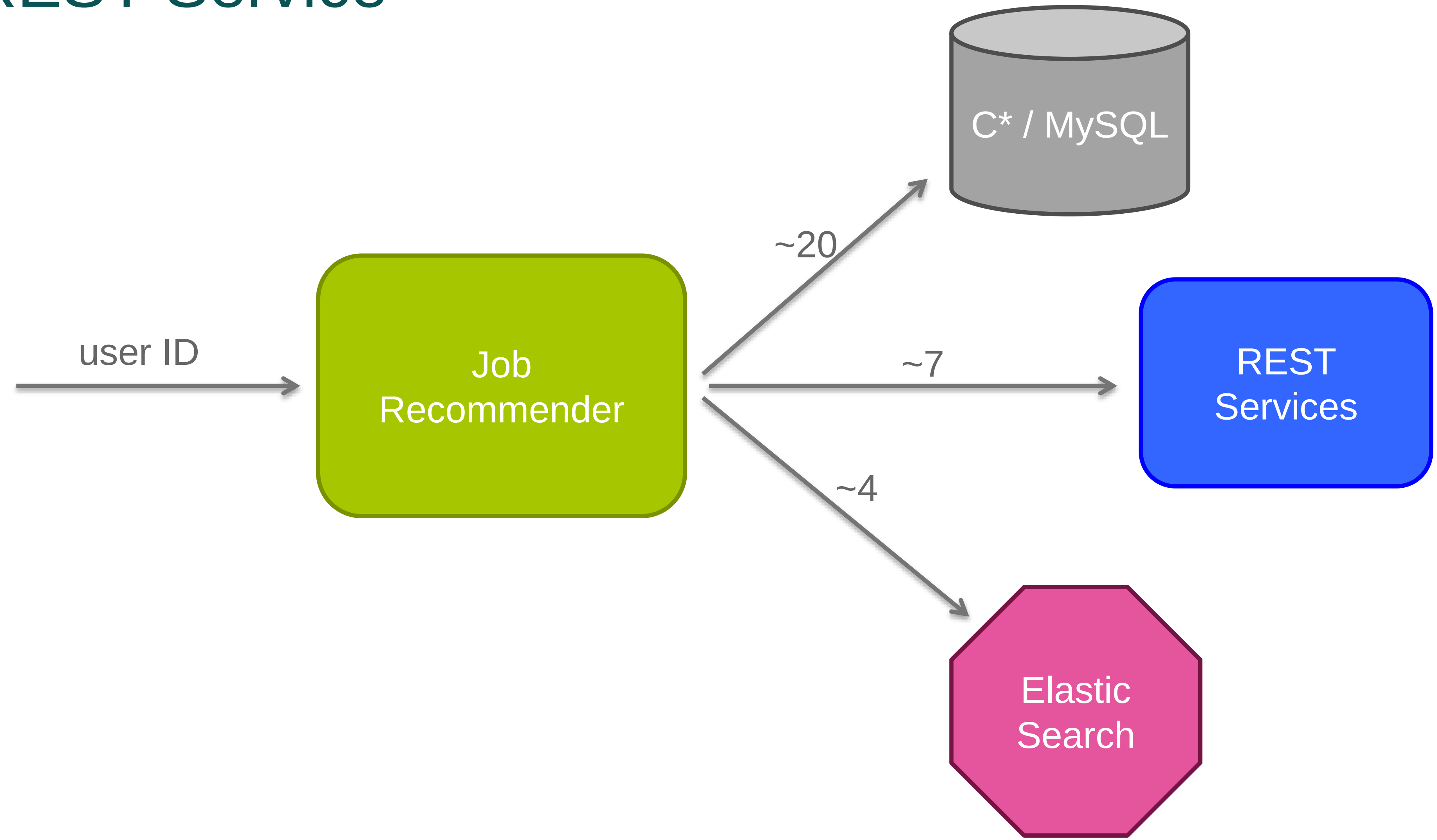
play

MySQL

kafka

akka

REST Service

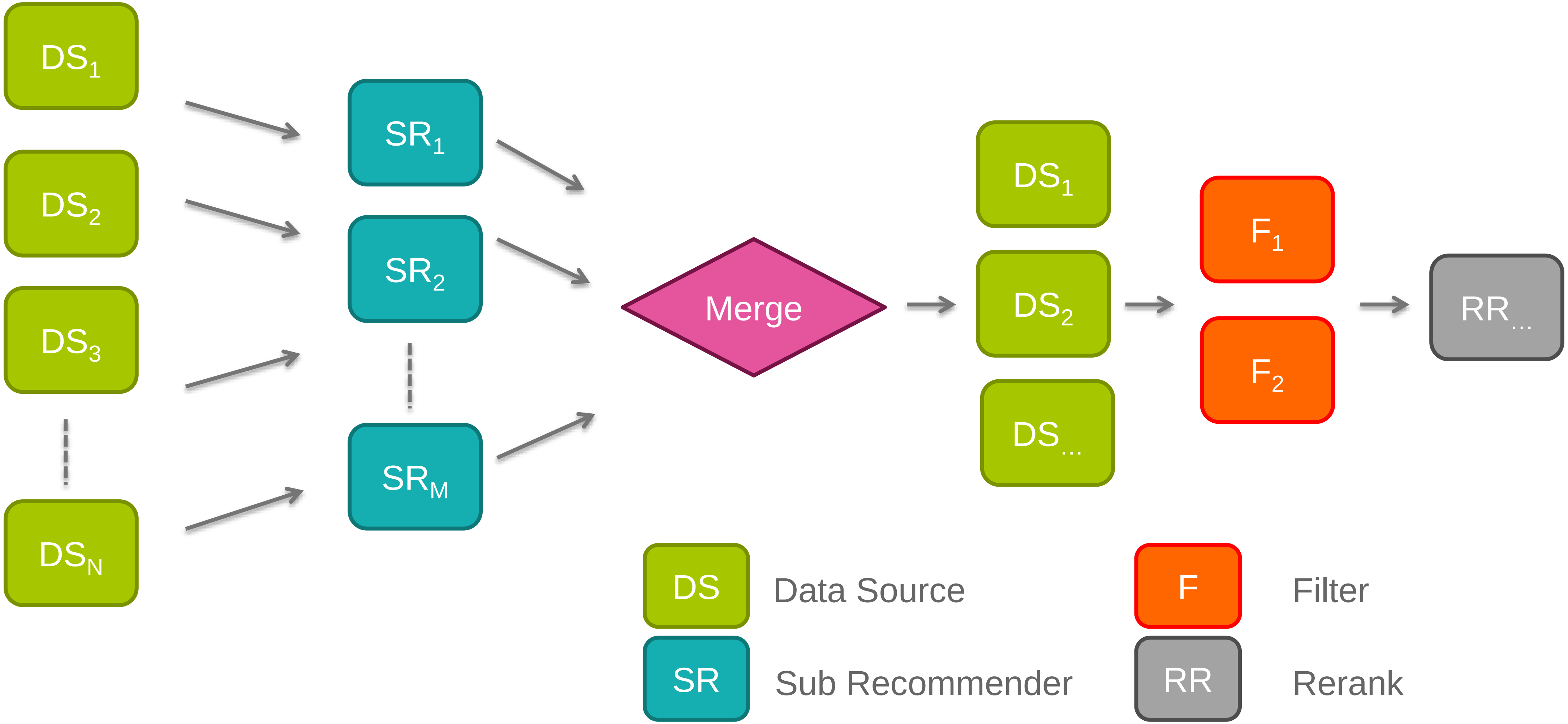


Single threaded this would take up to 12 seconds!

How to make it faster?

Do computations in parallel!

Internal recommender structure



Mixture of Online & Batch Processing

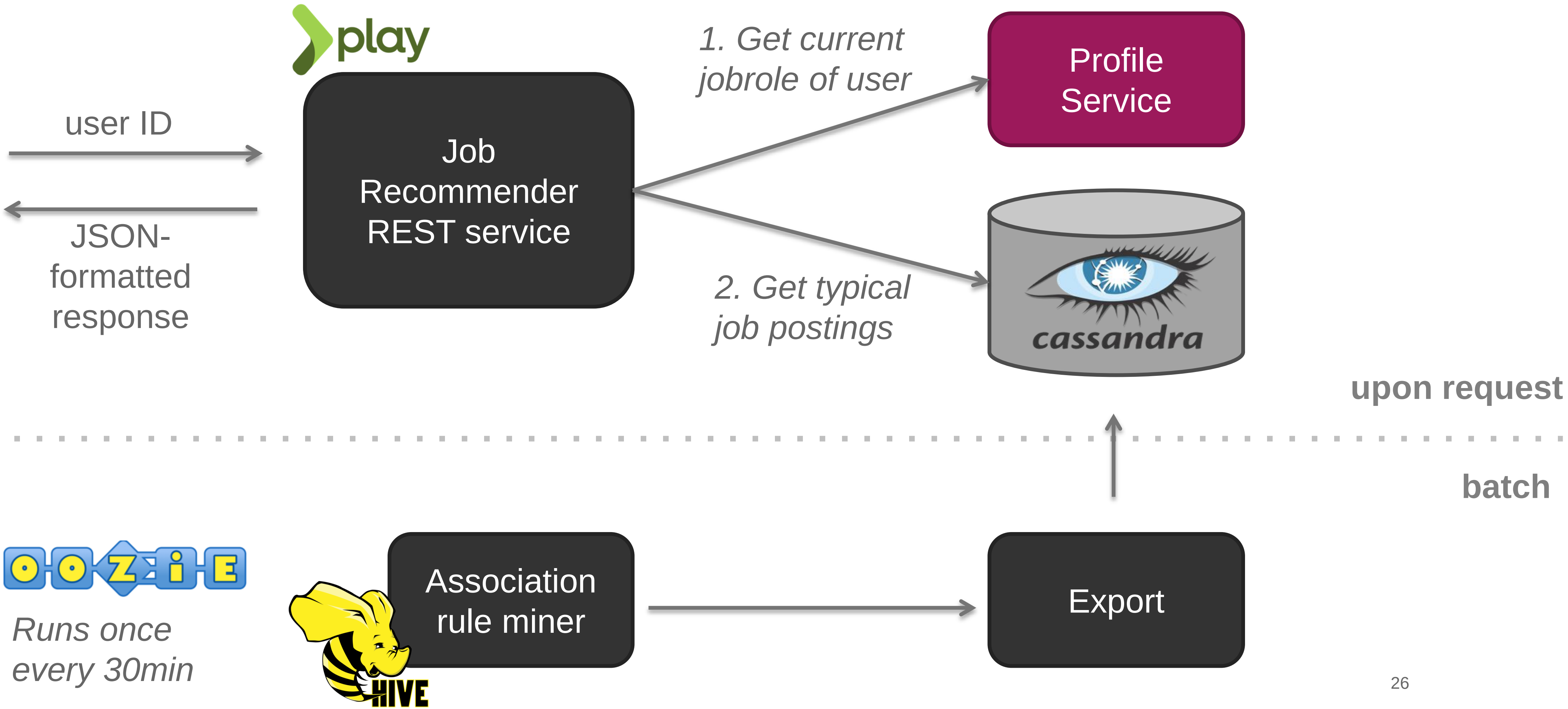
Online data

- 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

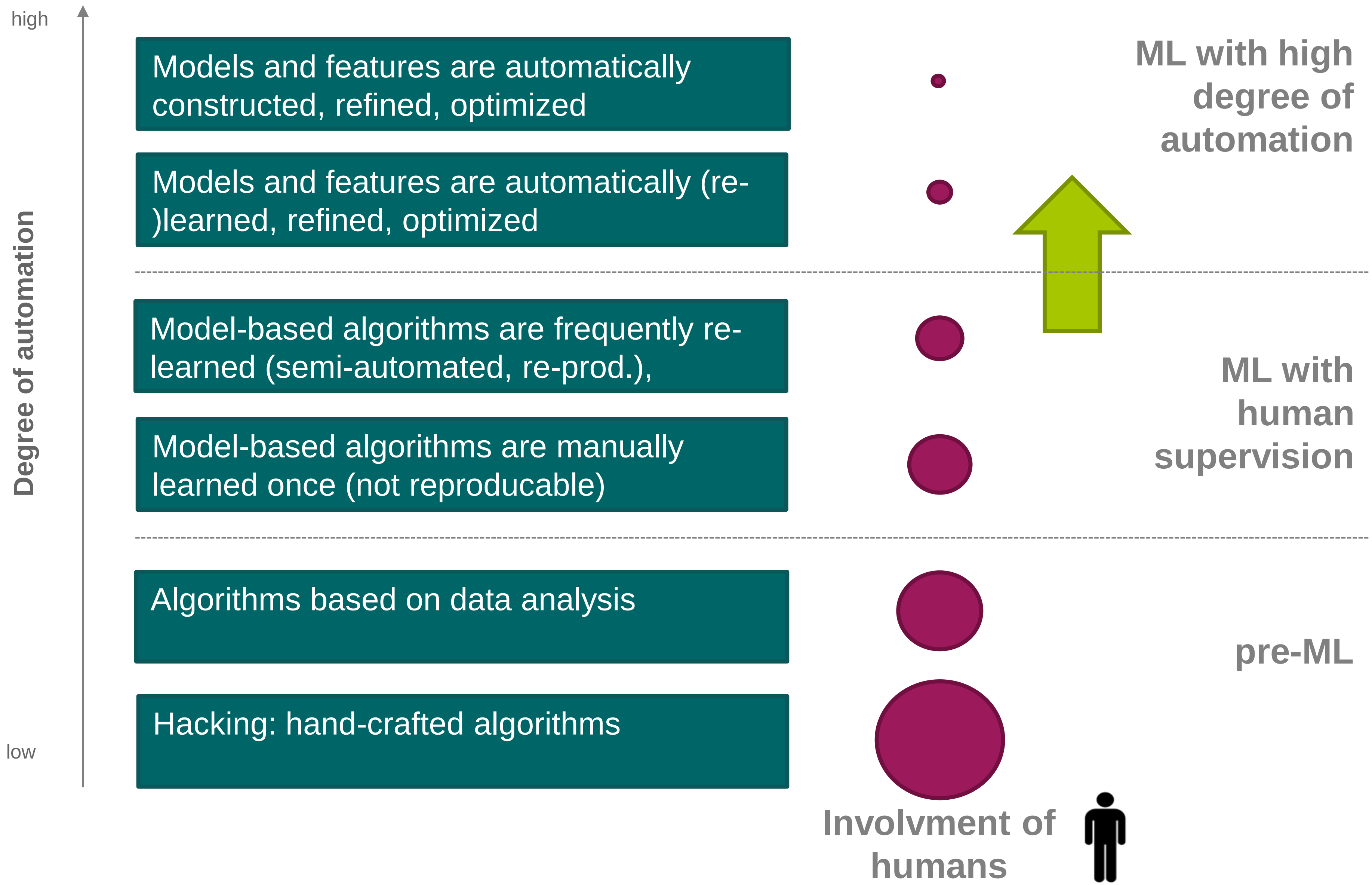


Deployment of RecSys Models

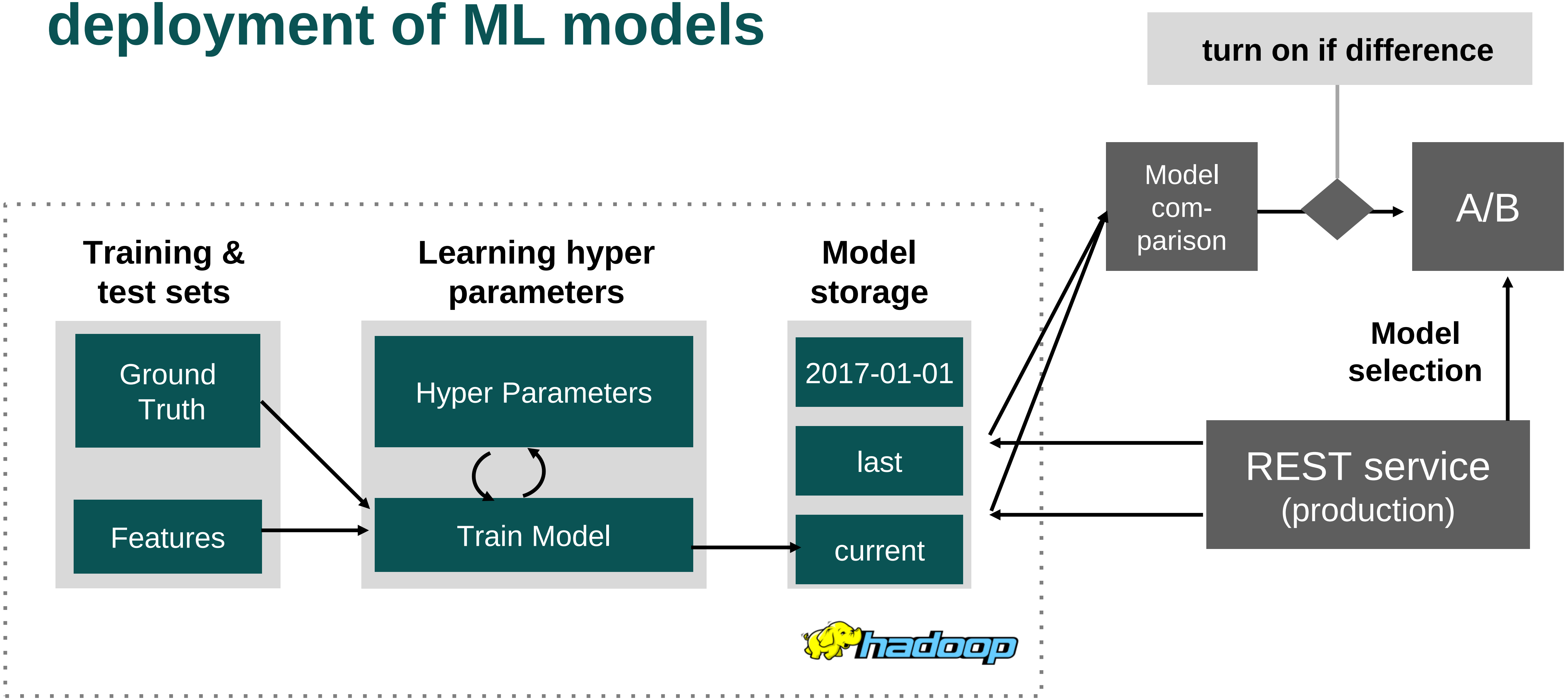
Examples of models used in production

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 is hiring

Data Engineers

and

Data Scientists