



# Talent Analytic Studio:

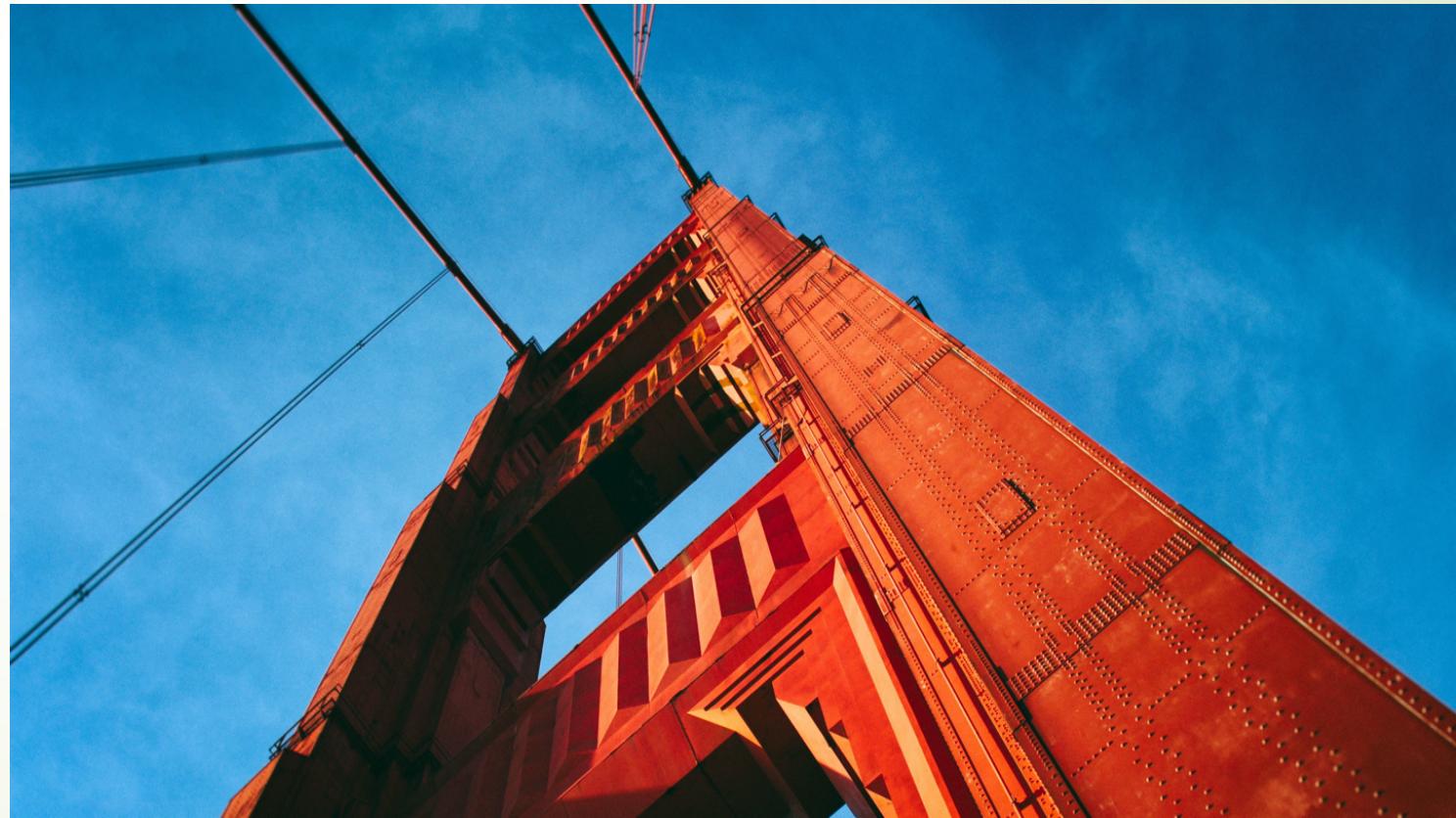
Recommendation Engine for  
Employee Training Courses at Orange - Silicon  
Valley

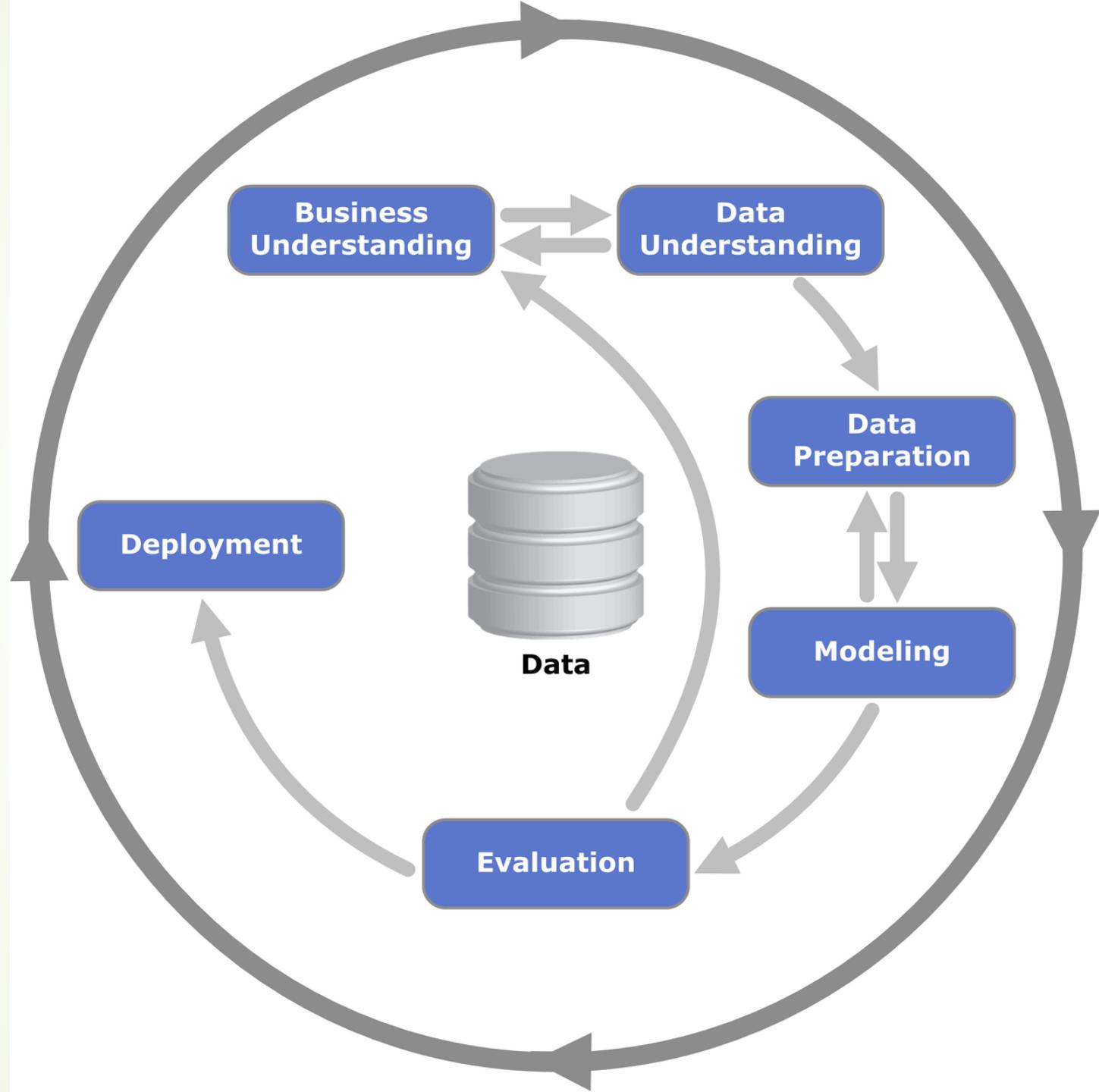
Justin Wang

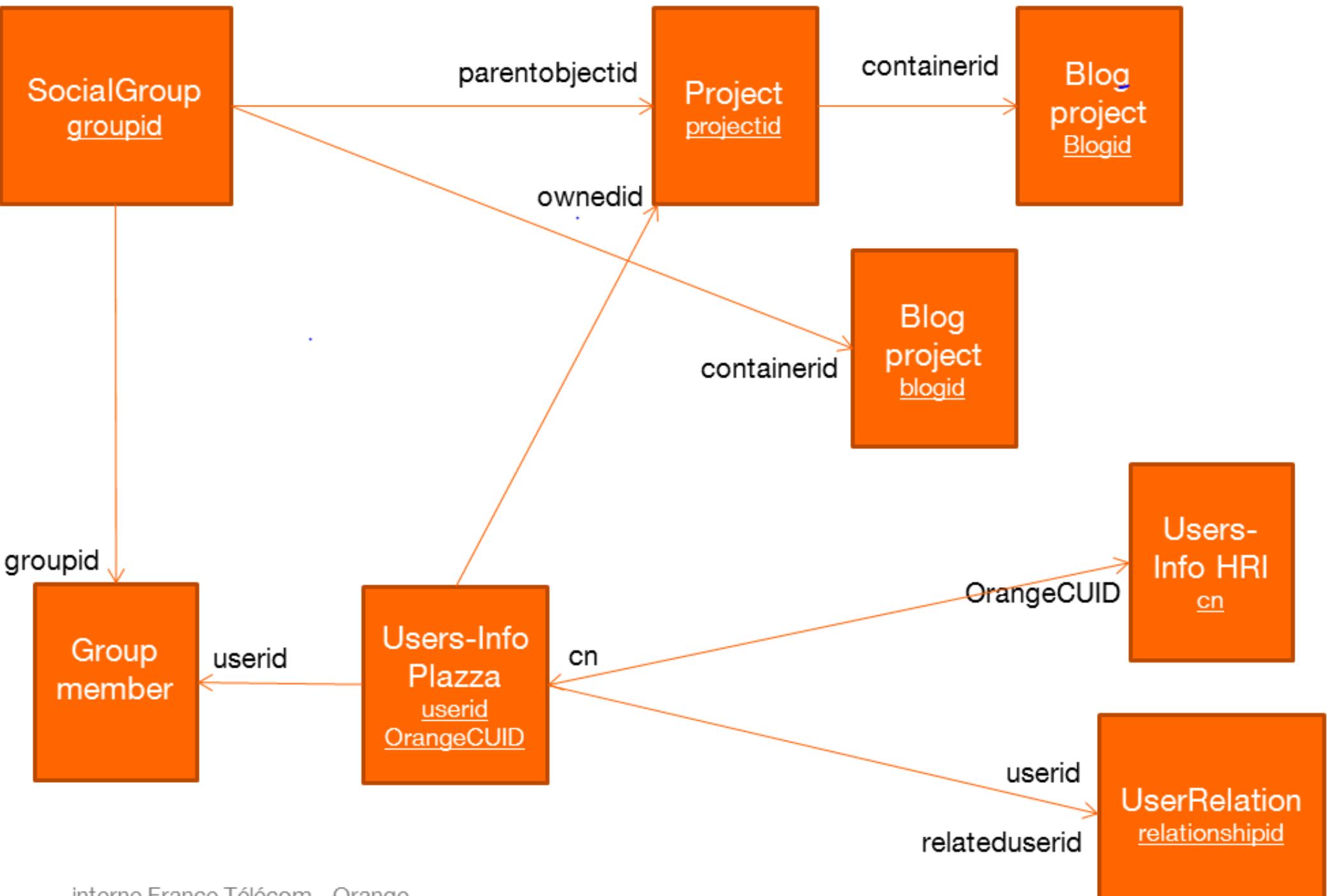
## Personalized HR Experience: Outcomes and Impacts

increase employee training course activity, retention, and completion rates

engaged employees who produce higher quality work, offer diverse skillsets, and provide excellent customer experiences

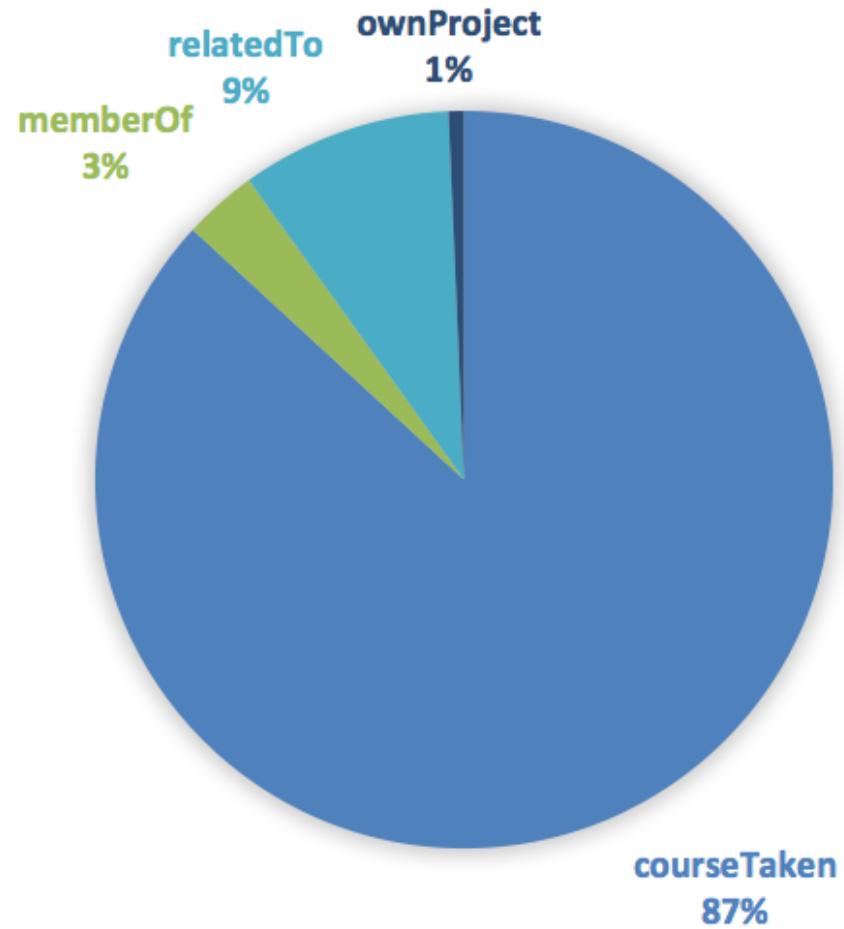






	event count	unique users	unique items	events per user	events per item
courseTaken	129,193	12,637	17,374	10.2234	7.4360
memberOf	4,816	474	1,671	10.1603	2.8821
relatedTo	13,747	286	316	48.0664	43.5032
ownProject	1,000	83	329	12.0482	3.0395

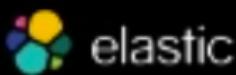
## DISTRIBUTION OF EVENT TYPES



# TECH STACK



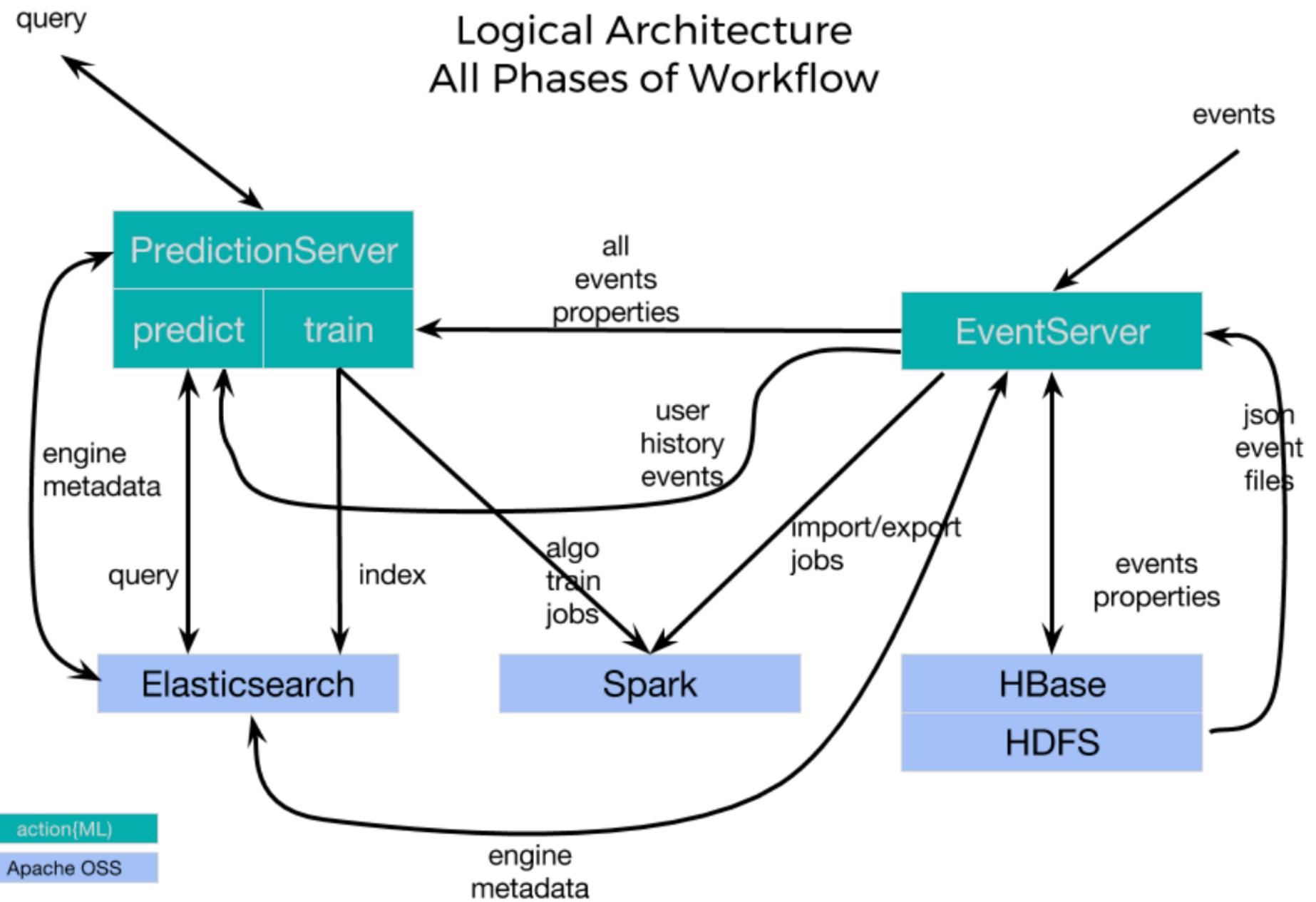
- Hbase 1.X
  - Postgres, MySQL, or other JDBC possible
- Spark 1.6.X
  - Fast, massively scalable, seems like the "winner"
- HDFS 2.6—Hadoop Distributed File System
  - Reliable, massively scalable, the defacto standard
- Spray
  - Supplies REST endpoints, multi-threaded via Akka actors
- Elasticsearch 1.7.X or 2.X
  - Reliable, massively scalable, fast
- Scala & Java 8
  - Fits functional and oop programming style for productivity
- Stable, Scalable, High Availability, Well Supported



# The Universal Recommender in PredictionIO

Logical Architecture

All Phases of Workflow

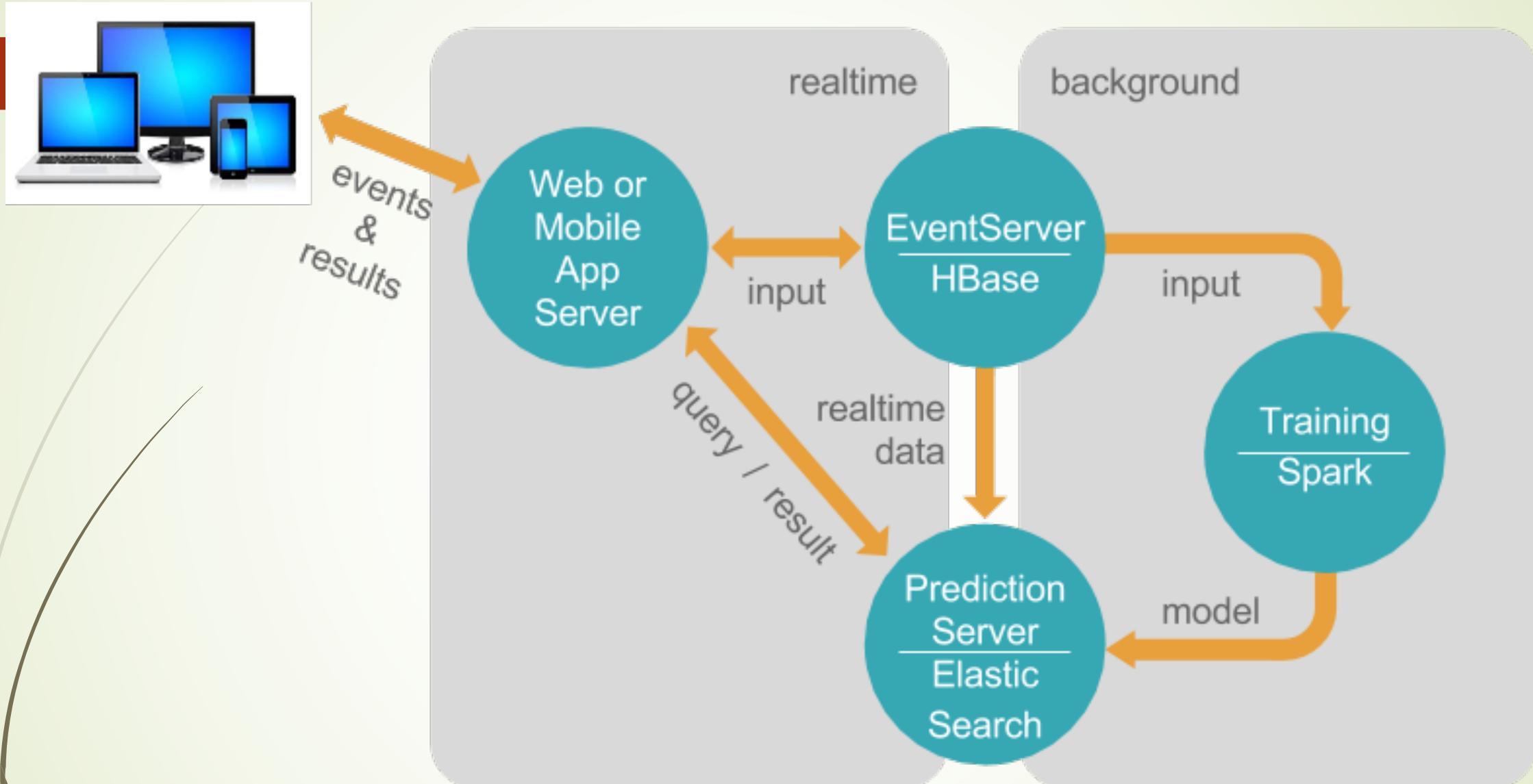


# Single Term Collaborative Filtering

recommendation scores:  $r = (C^t C)h_c$

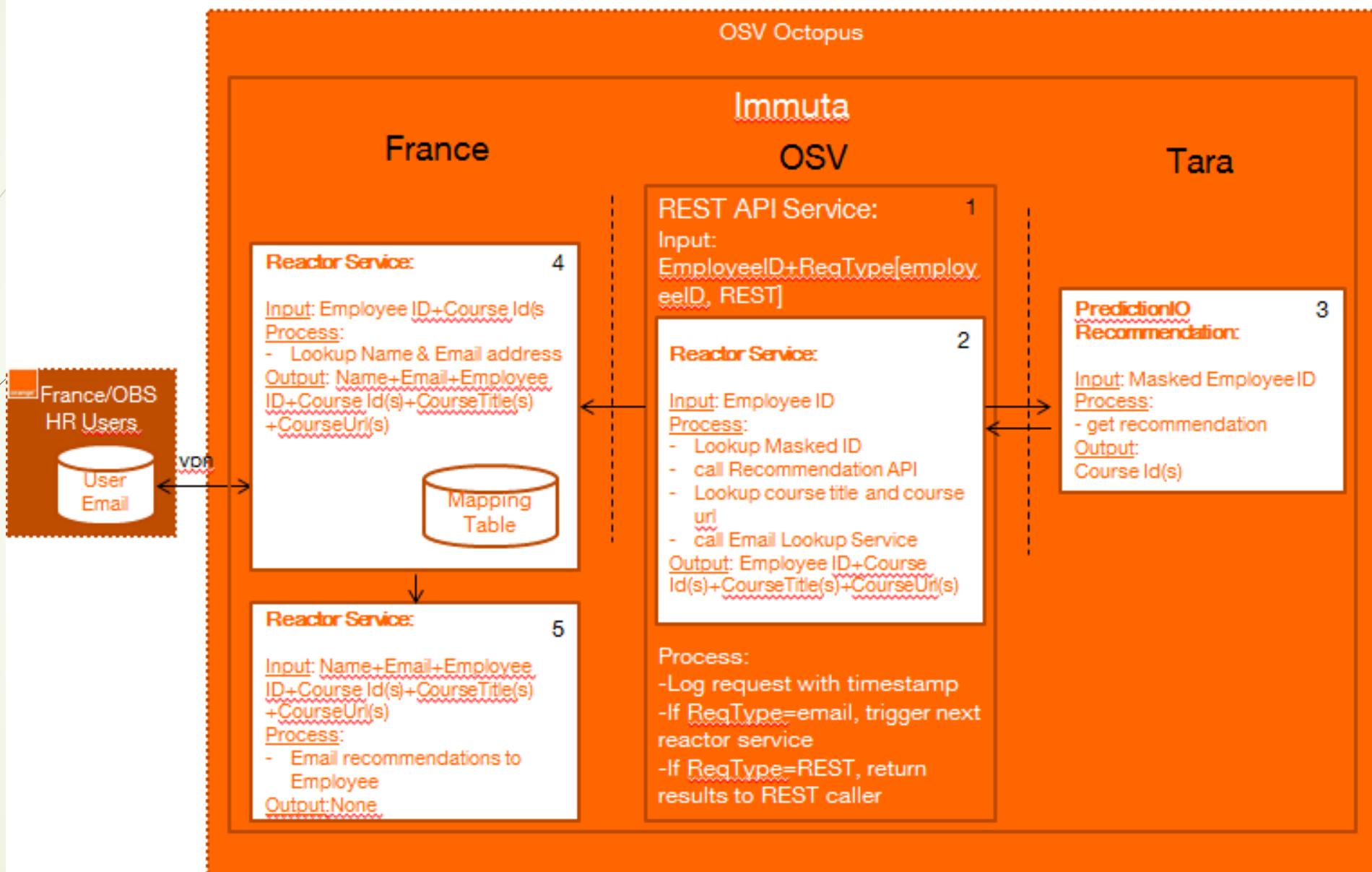
## Correlated Cross-Occurrence

$$r = (C^t C) h_c + (C^t P) h_p + (C^t S) h_s \quad (1)$$



```
oot@f186347626b2: /Tara/universalRecommenderNewroot@f186347626b2:/Tara/universalRecommenderNew# curl -H "Content-Type: application/json" \
> -d '{ "user": "de22719ee50124fdb8c09882bc19fde8", "num": 4 }' http://0.0.0.0:8000/queries.json
[INFO] [URAlgorithm] Query received, user id: Some(de22719ee50124fdb8c09882bc19fde8), item id: None
[INFO] [URAlgorithm] Misconfigured date information, either your engine.json date settings or your query's dateRange is incorrect.
Ignoring date information for this query.
[INFO] [URAlgorithm] Query:
{"size":4,"query":{"bool":{"should":[{"terms":{"courseTaken":["_addon_books_001","116671_eng","233057_eng","233061_eng","50527","58863","78129","83855_eng","84476","84871_eng","ib_itil_a01_it_ru","ib_ovvw_a02_it_enus","sp_cptf_a01_it_enus","sp_cptf_a05_it_enus","sp_cptf_a08_it_enus","sp_etih_a06_it_enus","TP1D0470_eng","TP1D0510_eng"]}],{"terms":{"memberOf":[]}}, {"terms":{"ownProject":[]}}, {"terms":{"relatedTo":[]}}, {"constant_score":{"filter":{"match_all":{}}}, "boost":0}], "must":[], "must_not":{"ids":{"values":["TP1D0510_eng","TP1D0470_eng","sp_etih_a06_it_enus","sp_cptf_a08_it_enus","sp_cptf_a05_it_enus","sp_cptf_a01_it_enus","ib_ovvw_a02_it_enus","ib_itil_a01_it_ru","84871_eng","84476","83855_eng","78129","58863","50527","233061_eng","233057_eng","116671_eng","_addon_books_001"]}}, "boost":0}}, {"minimum_should_match":1}}, "sort":[{"_score":{"order":"desc"}}, {"popRank":{"unmapped_type":"double", "order":"desc"}}]}}
[INFO] [esClient$] Results: 4 retrieved of a possible 73335
{"recommendedCourses":[{"course":"sp_cisp_a01_tp_enus", "score":1.0131391286849976}, {"course":"46519", "score":1.0008630752563477}, {"course":"sp_etih_a07_it_enus", "score":0.8782050013542175}, {"course":"sp_cptf_a06_it_enus", "score":0.716497004032135}]}oot@f186347626b2: /Tara/universalRecommenderNewroot@f186347626b2:/Tara/universalRecommenderNew# █
```

# Project Amazon Architecture





RMSE  
MAE  
precision@k  
MAP@k

```
  "preparatorParams":{  
    "":{}  
  },  
  "algorithmParamsList":[{  
    "als":{  
      "rank":10,  
      "numIterations":10,  
      "lambda":0.01,  
      "seed":3  
    }  
  }],  
  "servingParams":{  
    "":{}  
  }  
}  
Metrics:  
  Precision@K (k=10, threshold=4.0): 0.14810405643738975  
  PositiveCount (threshold=4.0): 5.66013071895425  
  Precision@K (k=10, threshold=2.0): 0.15317200954455856  
  PositiveCount (threshold=2.0): 6.718954248366012  
  Precision@K (k=10, threshold=1.0): 0.14865649963689181  
  PositiveCount (threshold=1.0): 9.830065359477125  
[INFO] [CoreWorkflow$] runEvaluation completed
```

formula for average precision at n is:

$$ap@n = \frac{\sum_{k=1}^n P(k)}{min(m, n)} \quad (2)$$

, where  $m$  is the number of relevant courses and  $n$  is the number of predicted courses. If the denominator is zero,  $P(k)/min(m, n)$  is set to zero.

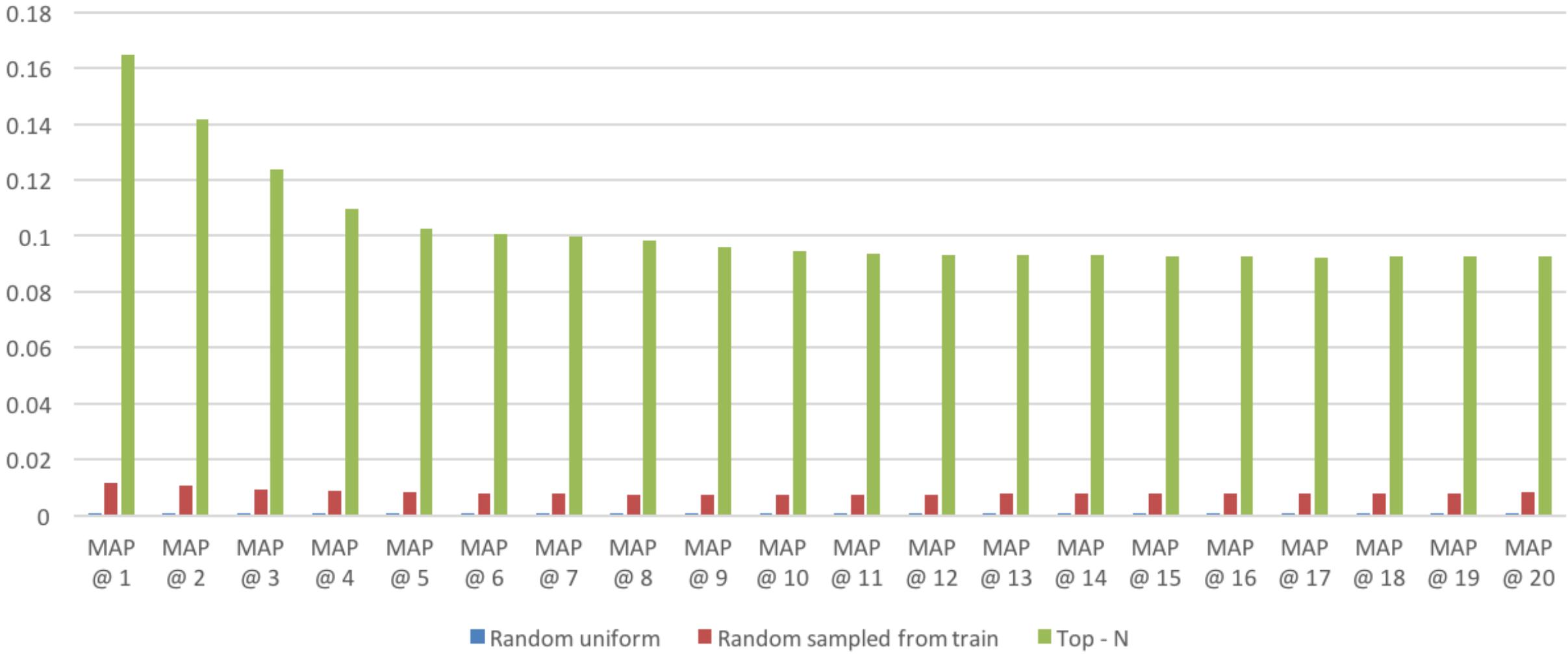


Mean Average Precision at k (MAP@k) is the average of the average precision of each user:

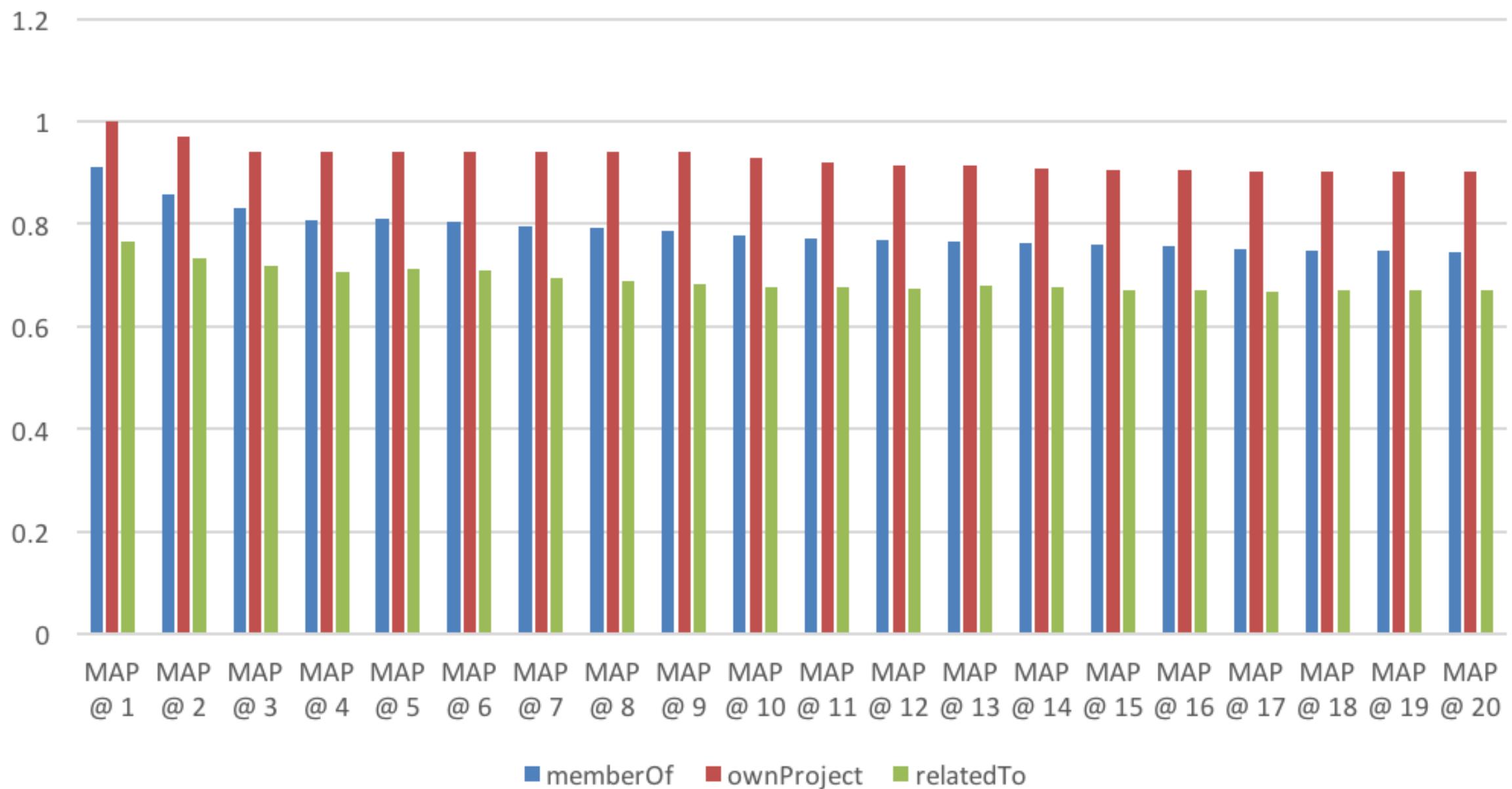
$$MAP@k = \frac{\sum_{i=1}^N ap@k_i}{N} \quad (3)$$

, where N is the number of users.

## MAP@k of Recommendations Compared to Random Sample



## MAP@k of Indicators, Isolated



Here is an example of an impressive match result from doc2vec. These two course descriptions were converted to vectors using doc2vec, and they had a cosine similarity of 0.999690:

**Exhibit A: Documentation and Criteria Used for Business Analysis**

["Business analysts must develop a repository of common language to facilitate communication and strategically align activities and goals. In this course, you'll learn about a number of business analysis techniques included in the categories of documentation, business and user cases, and setting metrics and criteria ..."]

**Exhibit B: Business Analysis and Solution Evaluation**

["After a solution has been partially or wholly implemented, a business analyst measures its effectiveness and ability to deliver the expected value to stakeholders. This involves measuring performance and identifying limitations or constraints that are keeping the solution from reaching its full value potential. The business analyst then recommends actions for overcoming any limitations..."]

*FIG. 4.3. Highest Scoring Cosine Similarity Pairing for these Course Descriptions*