

Search

Under the Hood

Agenda

- explain how search works
- the work needed to prepare the search index

What is ElasticSearch?

Document

- denormalized



Indexed

- queries known before hand

Implication

- queries known before hand



Overall Architecture



Responsibilities: Data Team

- ingest the data
-



Responsibilities: Product Team

 **design the index**



candidate_profile						
	full_name	summary	location_lat	location_lon	is_likely_to_jump	job_exp
	Olaf Oldtimer	An oldtimer in the IT Industry	40.71	74.00	TRUE	15
	Marco Midd	Worked as a Software Engineer for quite a while	40.71	74.00	TRUE	10
	Nero Newbie	A newbie in the IT Industry	40.71	74.00	FALSE	3

candidate_job_history						
	cand_id	job_title	industry_id	is_current	years_exp	
0001	CN_10001	Sr. Software Engineer	IND_100001	TRUE	10	
0002	CN_10001	Jr. Tech Support	IND_100002	FALSE	5	
0003	CN_10002	Sr. Software Engineer	IND_100001	TRUE	5	
0004	CN_10002	Jr. Software Engineer	IND_100001	FALSE	4	
0005	CN_10002	Intern	IND_100001	FALSE	1	
0006	CN_10003	Jr. Software Engineer	IND_100001	TRUE	2	
0007	CN_10003	Intern	IND_100001	FALSE	1	

ref_industry	
id	name
IND_100001	IT Engineering
IND_100002	IT Support

Responsibilities: Product Team

- ☛ **unload the data from the Data Warehouse (Redshift) to AWS S3**



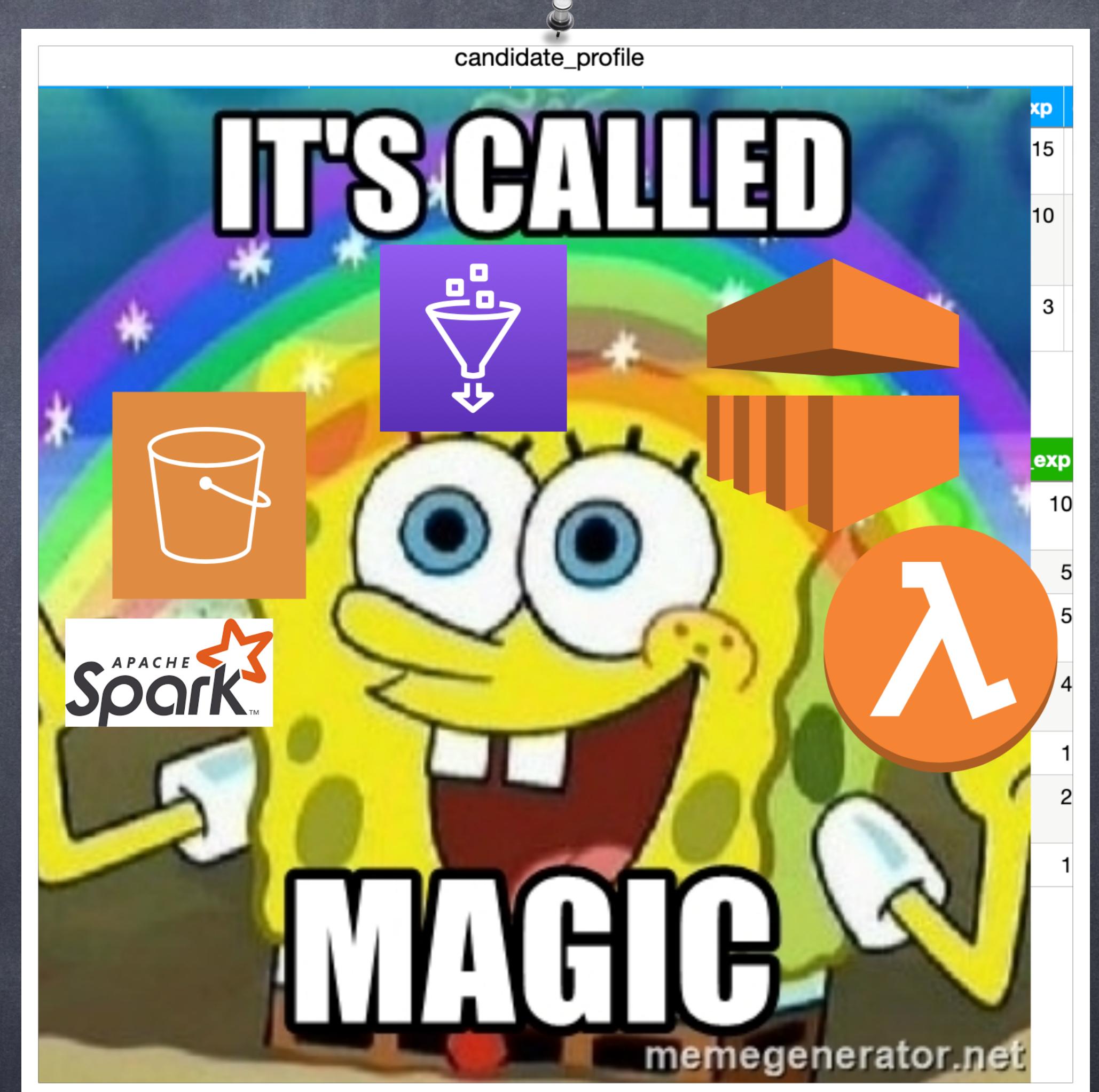
candidate_profile						
	full_name	summary	location_lat	location_lon	is_likely_to_jump	job_exp
	Olaf Oldtimer	An oldtimer in the IT Industry	40.71	74.00	TRUE	15
	Marco Midd	Worked as a Software Engineer for quite a while	40.71	74.00	TRUE	10
	Nero Newbie	A newbie in the IT Industry	40.71	74.00	FALSE	3

candidate_job_history					
	cand_id	job_title	industry_id	is_current	years_exp
0001	CN_10001	Sr. Software Engineer	IND_100001	TRUE	10
0002	CN_10001	Jr. Tech Support	IND_100002	FALSE	5
0003	CN_10002	Sr. Software Engineer	IND_100001	TRUE	5
0004	CN_10002	Jr. Software Engineer	IND_100001	FALSE	4
0005	CN_10002	Intern	IND_100001	FALSE	1
0006	CN_10003	Jr. Software Engineer	IND_100001	TRUE	2
0007	CN_10003	Intern	IND_100001	FALSE	1

ref_industry	
id	name
IND_100001	IT Engineering
IND_100002	IT Support

Responsibilities: Product Team

- ☛ **unload the data from the Data Warehouse (Redshift) to AWS S3**
- ☛ **process the data using Spark (AWS Glue or EMR works)**



Responsibilities: Product Team

- ☛ **unload the data from the Data Warehouse (Redshift) to AWS S3**
- ☛ **process the data using Spark (AWS Glue or EMR works)**
- ☛ **insert into Elasticsearch Service**

```
141 PUT candidates_demo/_doc/CN_10003
142 {
143   "fullName": "Nero Newbie",
144   "summary": "A newbie in the IT Industry",
145   "location": [40.71, 74.00],
146   "currentFunction": "FN_100001",
147   "likelyToJump": false,
148   "yearsOfExperience": 3,
149   "jobHistories": [
150     {
151       "jobTitle": "Jr. Software Engineer",
152       "industry": "IND_100001",
153       "industryName": "IT Engineering",
154       "yearsOfExperience": 2,
155       "current": true
156     },
157     {
158       "jobTitle": "Intern",
159       "industry": "IND_100001",
160       "industryName": "IT Engineering",
161       "yearsOfExperience": 1,
162       "current": false
163     }
164   ]
165 }
```

