Lighthouse

Large-scale graph pattern matching on the Pregel model

Peter Boncz
Claudio Martella
Spyros Voulgaris

Once upon a time

MapReduce was created,

and large-scale analytics
on commodity clusters
was suddenly easy.

We got greedy, and we asked for more (or should I say *less* programming).

And there was Hive and Pig (and all the others…)

In other terms, write a query in a high-level language, and get it automatically translated into a series of optimised M/R jobs (instead of lots of Java code).

And everybody was happy *again*.

Later came *massive* graphs (e.g. Social Networks, the Web, road maps, protein-protein networks…).

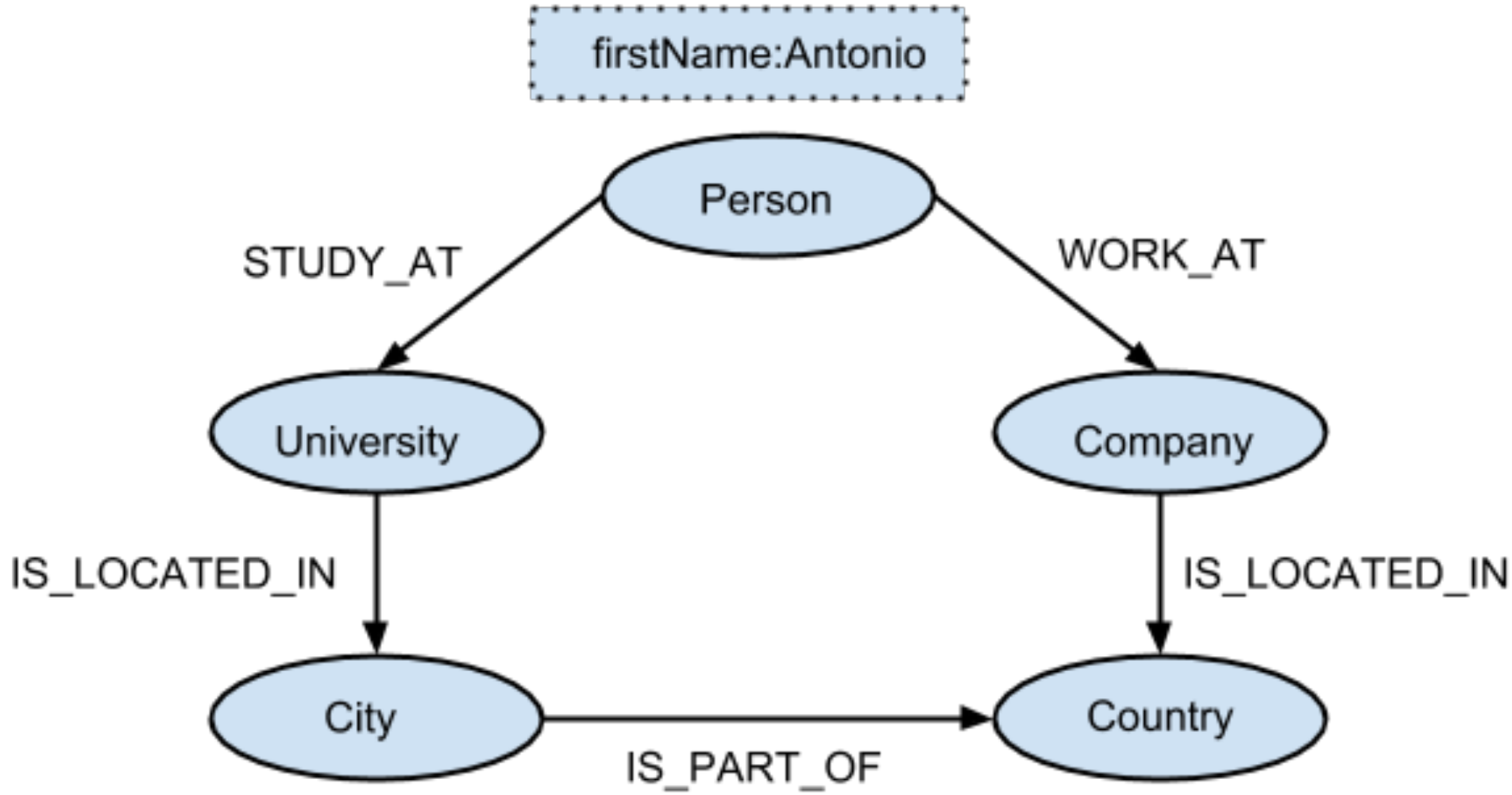
We created Pregel,

and large-scale graph
processing was easy too.

But we got greedy again,
and Pregel was not
enough too.

Lighthouse allows you to express graph analytics in a high-level language, and it executes it on Pregel (Giraph).

Lighthouse currently runs a subset of Cypher (Neo4J's query language).


firstName:Antonio
Person
STUDY_AT
WORK_AT
University
Company
IS_LOCATED_IN
IS_LOCATED_IN
City
Country
IS_PART_OF

```
MATCH (person:Person {firstName:"Antonio"})
-[:WORK_AT]-> (company) -[:IS_LOCATED_IN]->
(country)
<-[:IS_PART_OF]- (city) <-[IS_LOCATED_IN]-
(university) <-[:STUDY_AT]- (person)
RETURN company.id, country.id, person.id,
university.id, city.id, country.id;
```

We want more.

- Better automatic query planning
- Wider support of Cypher (aggregations, path-queries, etc.)
- Efficient memory and network-optimised engine (e.g. pipelining, etc.)

Do you want to work
on the next big thing?

Join us.

Peter Boncz p.a.boncz@vu.nl
Claudio Martella c.martella@vu.nl
Spyros Voulgaris s.voulgaris@vu.nl