Cypher for Apache Spark

Max Kießling
CAPS - The Spark SQL for graphs

Spark SQL

```scala
spark.read.parquet("hdfs://...").createTempView("persons")
spark.sql("SELECT * from persons").show()
```

<table>
<thead>
<tr>
<th>id</th>
<th>name</th>
<th>age</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Alice</td>
<td>25</td>
</tr>
<tr>
<td>1</td>
<td>Bob</td>
<td>23</td>
</tr>
</tbody>
</table>

(2 rows)

Cypher for Apache Spark

```scala
val graph = GraphSources.fs("hdfs://...").parquet.graph("myGraph")
graph.cypher("MATCH (p:Person)-[:KNOWS]->(p2:Person)
|RETURN p1, p2").show
```

<table>
<thead>
<tr>
<th>n</th>
<th>n:Person</th>
<th>n_name</th>
<th>m</th>
<th>m:Person</th>
<th>m_name</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>true</td>
<td>Alice</td>
<td>1</td>
<td>true</td>
<td>Bob</td>
</tr>
</tbody>
</table>

(1 row)
Graph Catalog
  Manages available graph source and graph views

Multiple Graph Queries
  Query and produce multiple graphs in a single Cypher Query

Graph Data Source - Neo4j, SQL, Filesystem
  Read and Write graphs to different data sources including Neo4j, RDBMS and Filesystem
MATCH (n:Person)-[:LOVES]->(s:System)
WHERE n.name = 'Alice'
RETURN n.name, s.name
(:Thank)-[: ]->(:You)

github.com/openCypher/cypher-for-apache-spark