Neo4j: A Graph Database

Presenter: Jinliang Wei

CMU CSD

November 13, 2013
Why Graph Database?

Relational and NoSQL databases lack support for relationships.

- Relational databases: expensive joins.
- NoSQL databases: embed related keys in value.
Graph Database

- Designed for handling relationships.
- Process query by graph traversal from individual elements. Graph traversal is efficient.
- Avoid expensive joins.
- Query processing time is proportional to how much of the graph that query explores instead of the size of data stored.
- Example query: who are friends of James Bond’s friends?
Neo4j

- Most popular graph database.
- Used by: Adobe, Cisco, Glassdoor, Huawei, HP...
Neo4j Data Model

- Graph model: property graph.
- Nodes.
- Relationships (edges) connect nodes.
- Both have properties.
Features of Neo4j

- ACID transactions.
- Distributed.
- Supports API in many languages: Java, Python, Ruby, JS...
MATCH (john {name: 'John'})-[[:friend]->()]-[[:friend]->(fof)]
RETURN john, fof
## Performance: Neo4j vs. RDBMS

Table 2-1. Finding extended friends in a relational database versus efficient finding in Neo4j

<table>
<thead>
<tr>
<th>Depth</th>
<th>RDBMS execution time (s)</th>
<th>Neo4j execution time (s)</th>
<th>Records returned</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.016</td>
<td>0.01</td>
<td>~2500</td>
</tr>
<tr>
<td>3</td>
<td>30.267</td>
<td>0.168</td>
<td>~110,000</td>
</tr>
<tr>
<td>4</td>
<td>1543.505</td>
<td>1.359</td>
<td>~600,000</td>
</tr>
<tr>
<td>5</td>
<td>Unfinished</td>
<td>2.132</td>
<td>~800,000</td>
</tr>
</tbody>
</table>

1 reference: Graph Databases, p20