A Galaxy Interactive Environment for exploring the Neo4j Graph Database

Thoba Lose, Peter van Heusden, Alan Christoffels

thoba@sanbi.ac.za, pvh@sanbi.ac.za, alan@sanbi.ac.za
South African National Bioinformatics Institute
University of the Western Cape
South Africa

Introduction

• Storing biological data involves modeling and storing thousands of entities that are interrelated in complex ways.

• Relational databases meet very specific needs and are not designed to fit all scenarios.

• Graph databases, which focus on connections between entities, are seen as a natural fit to these complex relationships.

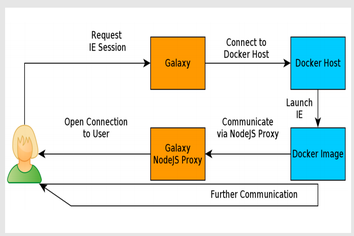
• Neo4j is a highly scalable graph database with a declarative query language called Cypher.

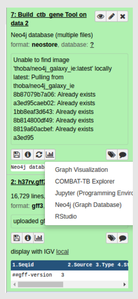
Implementation

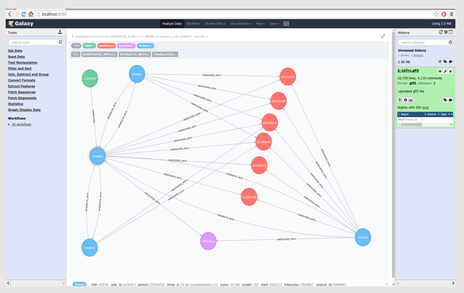
• We implemented a Galaxy Interactive Environment (GIE) to explore a Neo4J database that is stored as a Galaxy dataset.

•The GIE was built by linking a Galaxy plugin to a Docker container based on a modified version of the Neo4J:2.3 Docker image.

• The Interactive Environment is launched from the visualisation menu and is only available for Neo4J database (neostore) datatypes.



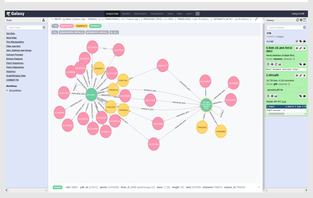




Results

A Neo4j GIE to explore the graph database using Cypher, a declarative query language desgined for expressing graph queries.

```
MATCH (g:Gene {locus_tag: "Rv0014c"})-[:TRANSCRIBED]->(t:Transcript)-[:PROCESSED_INTO]->
(c:CDS)-[:TRANSLATED]->(p:Protein) -[:INTERACTS_WITH]- (p:Protein)-[:ASSOCIATED_WITH]->
(term) RETURN p, term
```



```
MATCH (g:Gene {locus_tag: "Rv0014c"})-[:TRANSCRIBED]->(t:Transcript)-[:PROCESSED_INTO]->
(c:CDS)-[:TRANSLATED]->(p:Protein)-[:ASSOCIATED_WITH]->(term:GoTerm) WITH COLLECT(term) AS
terms UNWIND terms AS term
MATCH (gene:Gene) -[]->(:Transcript)-[]->(:CDS)-[]->(:Protein)-[]->(term:GoTerm) RETURN
term.name, COLLECT(gene.name)
```

Conclusion

The development of the Neo4J GIE bridges the gap between Galaxy and graph databases by allowing users to interrogate, using the Cypher declarative language, data without leaving Galaxy.

References