Bioinformatics Resources
Exercise Sheet 6
discussed on June 15th and 18th

No-SQL
1. Discuss the differences between relational databases and NoSQL approaches in terms of: Scalability, structure of the data and changes thereof.
2. Explain what the CAP theorem is.
3. Database consistency models differ between NoSQL and relational database systems. Explain what ACID and BASE are and think of the trade-offs between both models.

neo4j
1. Download and install the neo4j graph database software. After installation you can use the neo4j browser at http://localhost:7474/browser/, to interact with the database. Familiarize yourself with neo4j and its query language cypher. There are several interactive tutorials available, directly from the browser. Next, load the cat breeds dataset from our website into your local database.
2. Give the cypher statements to answer the following questions
   a. The id of the cat with property 'special'
   b. All breeds of type LONGHAIR
   c. All males that are of type LONGHAIR but not of the breed Selkirk Rex Longhair (SRL)
   d. All descendants of the cat with id '9883'
   e. The number of cats with no parents
   f. The id and breed code of the cat that has fathered most children
3. Give the cypher query for retrieving the group of the cat with id 9883. Use the PROFILE command to see how neo4j executes the query. What is the most time-consuming step? Create an index to increase the performance of the query, and describe how neo4j executes the query under usage of the index. (Note: With so little data you are unlikely to notice an actual difference in execution speed)

mongodb
1. The next exercise will use mongodb. For now make sure you have a working server running either on your local machine or through for example mlab. Take a look at the GettingStarted guide with the shell before we transition to python next week.

Good luck,
J. Reeb & L. Richter