Bioinformatics Resources
Exercise Sheet 6
due on June 26th, 9 a.m.
Total number of points: 13P

Task 1
Name and explain in one sentence three different NoSQL data storage approaches. (1P)

Task 2
Discuss the differences between relational databases and NoSQL approaches in terms of: Scalability, structure of the data and changes thereof. (1P)

Task 3
Explain what the CAP theorem is. (1P)

Task 4
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 it's 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. (1P)

Task 5 (6P in total, 1 for each sub-task)
Give the cypher statements to answer the following questions:
1. The number of cats with no parents.
2. All breeds of type LONGHAIR
3. All males that are of type LONGHAIR but not of the breed Selkirk Rex Longhair (SRL)
4. The id of the cat with property 'special'
5. The id and breed code of the cat that has fathered most children
6. All descendants of the cat with id '9883'

Task 6
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 a difference in execution speed) (3P)

Good luck,
J. Reeb & L. Richter