Lecturer: Jan Baumbach Term: Winter 2015/2016

Class: Introduction to Bioinformatics

Exercise sheet – Bioinformatics Databases

1. What is the difference between a "Data Warehouse" and a "Federated Database"?

2. Name and describe at least four typical advantages of DBMS-based data storage over file system based data management. Give one real world example for each.

3. Design and paint an Entity Relationship (ER) diagram for a small database dedicated to store gene regulatory interactions of a bacterium. The following entities are necessary: genes (with ID and one or more (!) names), transcription factors (with ID and one name) and binding sites (with binding sequence of 10…30 bp length). Include the following relations: (1) A transcription factor is encoded by one gene. (2) A transcription factor may regulate a gene by binging to one or more binding sites.

4. Download and install MySQL. Design and write DDL code for MySQL that implements the in exercise 3 designed database model. Afterwards, provide DML code that fills the database with arbitrary data. Finally, give the corresponding SQL code for the following two queries:

(a) A list of all genes that are regulated by two ore more transcription factors. Hint: Use the MySQL online manual. You might need the "group" and "count" clauses, for instance.

(b) A list of all gene regulations. Output style: transcription factor, target gene, binding site

5. Name and describe one major advantage and one major disadvantage of an ontological, generalized data structure over "standard", explicit data structures. When would you use which approach. Explain why (give example if necessary).