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

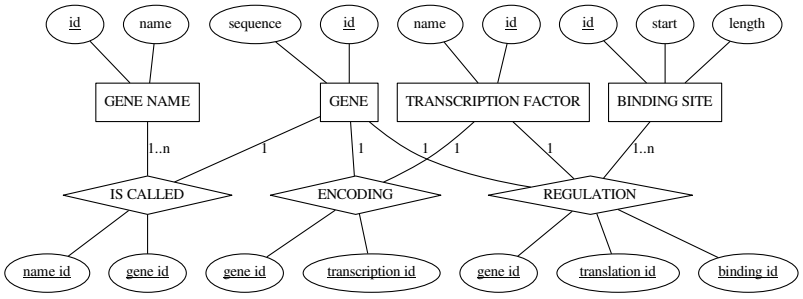
   - **Data Warehouse:** Is a central repository for all or significant parts of the data that an enterprise's various business systems collect. Typically, a data warehouse is housed on an enterprise mainframe server or increasingly, in the cloud.
   - **Federated Database:** Is a database management system, which transparently maps multiple autonomous database systems into a single federated database. This means that multiple databases appears as one database to the user.

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.*

   (a) A locking mechanism for concurrent access (If users access the same item in the database there is no RW or WR errors.)
   (b) The ability to swiftly recover from crashes and errors, including restartability and recoverability (If the System Crashes due to for example power failures the Logging system recovers know changes - hence reducing loss)
   (c) Robust data integrity capabilities (Data is the same from all access points).
   (d) Logging and auditing of activity (If changes are discarded before saving the system can re-roll.)

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:

   (a) A transcription factor is encoded by one gene.
   (b) 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.*

```
1  SELECT * FROM GENE WHERE
2    (
3    SELECT gene_id FROM TRANSCRIPTION
4    WHERE COUNT (gene_id) > 1
5    ) = id;
```

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

```
1  SELECT * FROM TRANSCRIPTION
2  ORDER BY TRANSCRIPTION.name, GENE.name, BINDING_SITE.id
```

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).*

   - **Advantage:** More flexible.
   - **Dis-advantage:** Hard to search due to structure. Might miss information.

   Which approach?

   - **Ontological:** If I would need a quick overview, where I don't know if there is any structure in the data or what the structure is.
   - **Standard:** Once a structure of the data has revealed itself.