

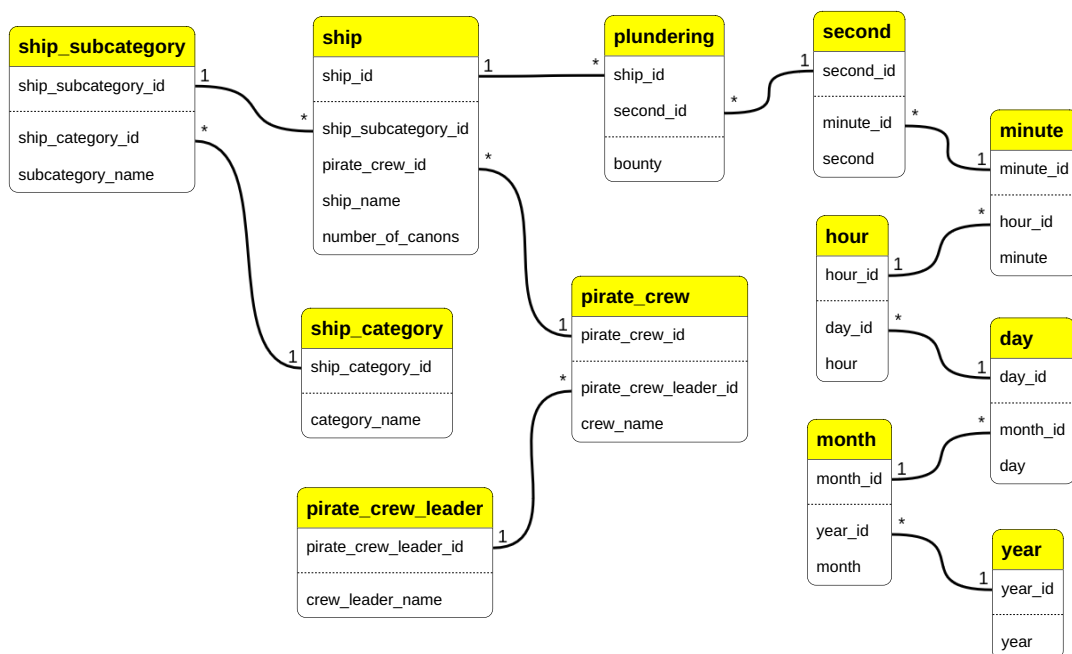
Relational Databases and Data Analysis

- Hand in your solution as PDF. The file must be named `solution-4.pdf`. Use `zipme.py` to create the archive to upload in ILIAS.

Exercise 1 *Multi-dimensional data model*

(1 + 1 + 2 + 2 + 1 + 1 Points)

A cutting-edge pirate crew wants to leverage relational databases to track their performance metrics in order to facilitate growth of their operational efficiency. They came up with the following schema.



- (a) For each table of the schema above, identify whether it is a fact table or a dimension table. Briefly explain how to differentiate between dimension tables and fact tables.

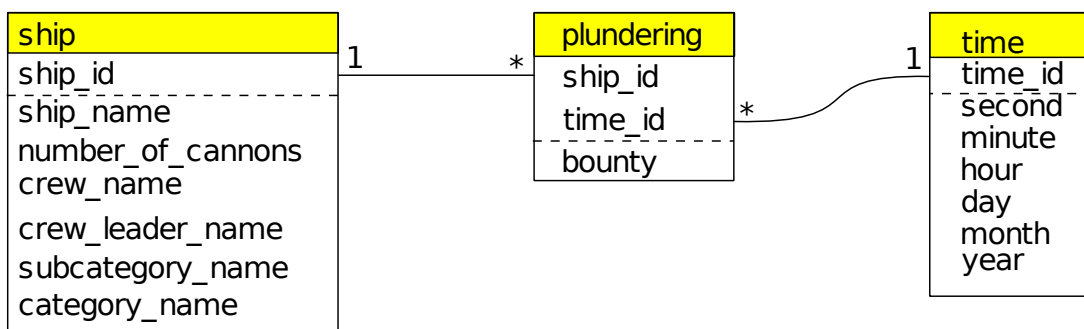
To identify the fact tables, look at the cardinality numbers (usually “*” and “1”). The tables which only have cardinality “*” are fact tables. The others are dimension tables. Alternatively, identify the tables where the collection of foreign keys are also the primary key. In our case, the only fact table is `plundering` and the other tables are dimension tables.

- (b) Is this a star schema, snowflake schema, galaxy schema or fact constellation schema? Briefly explain your reasoning!

- Star schema: single fact table, multiple referenced dimension tables
- Snowflake schema: single fact table, multiple dimension tables which themselves reference other tables
- Galaxy schema: multiple fact tables linking multiple dimension tables
- Fact constellation schema: like galaxy schema, but there are fact tables which contain aggregates.

We have a single fact table `plundering`, which rules out galaxy and fact constellation schema, which would have multiple fact tables. We also have transitive dimension tables, ruling out the star schema. This only leaves the snowflake schema, for which all conditions are fulfilled.

(c) Convert the previous schema into a star schema. Indicate cardinalities with “*” and “1” for each edge and separate the primary key with a dashed line.



(d) Give two reasons for preferring the star schema over the snowflake schema and two reasons for preferring the snowflake schema over the star schema.

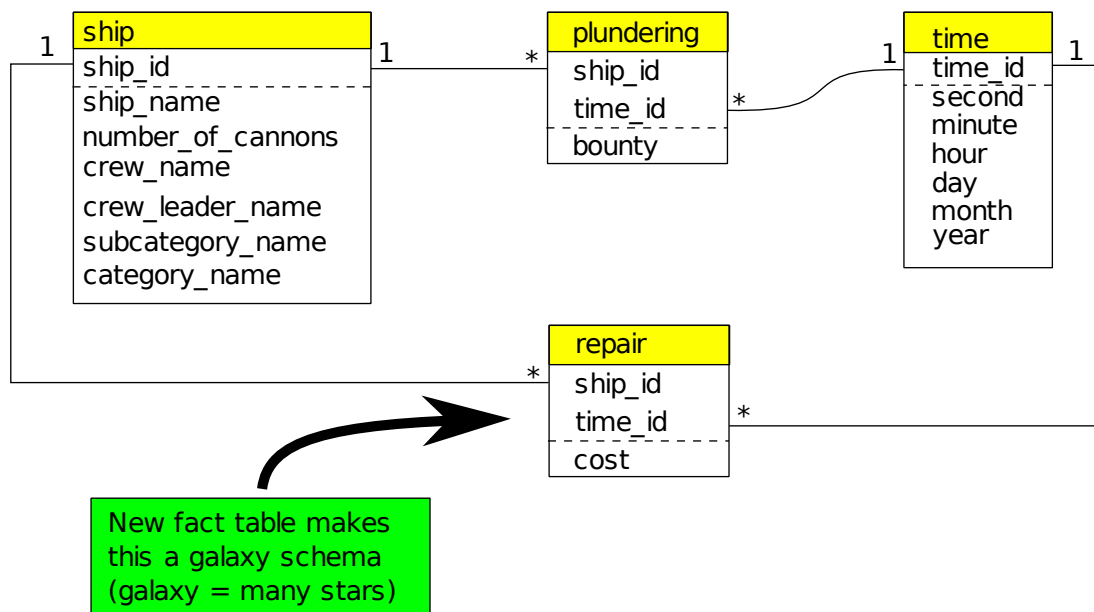
Advantages star schema / disadvantages snowflake schema:

- queries are easier to understand because of fewer joins, queries within dimension do not require joins at all
- queries might be faster due to fewer joins

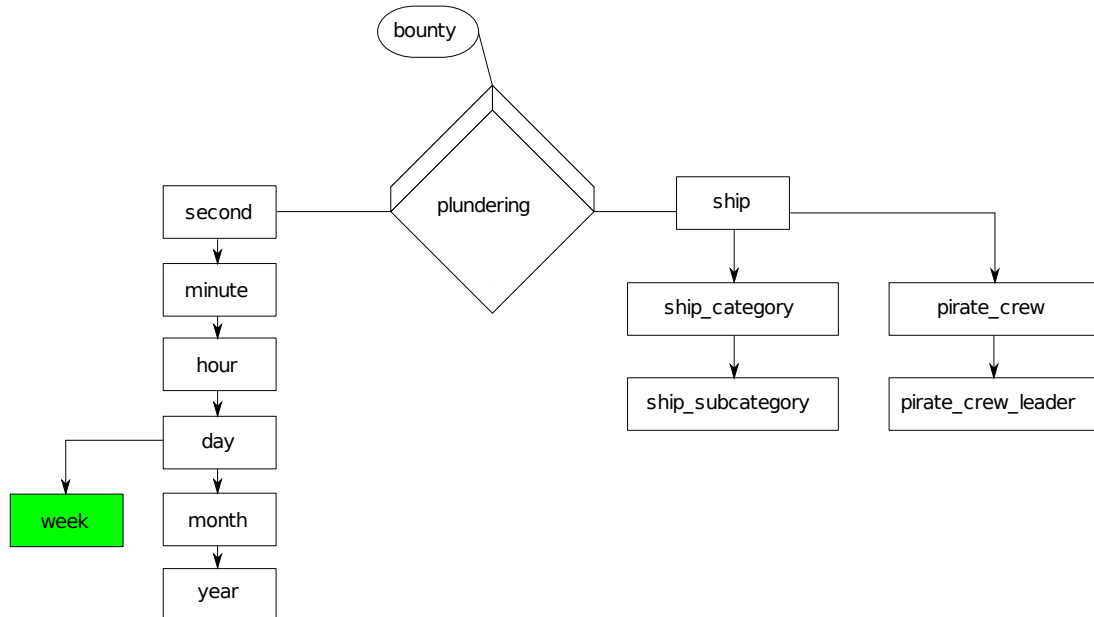
Advantages snowflake schema / disadvantages star schema:

- normalized data decreases risk of update anomalies
- can decrease storage cost if there are dimensions with large attributes or deep hierarchies

(e) Modify or extend the pirate’s schema to obtain an unmistakable galaxy schema. Briefly explain which aspect was important to make it so.



(f) Express the pirate's schema as a multi-dimensional Entity/Relationship model. Currently, it only contains simple hierarchies. Introduce an additional dimension such that there is a parallel hierarchy and highlight it in the diagram. Be advised that the shape of the nodes and the direction of arrow heads are relevant.



The newly introduced parallel hierarchy is highlighted in green.

Exercise 2 Classification Hierarchies

(2 Points)

Convert the following vertical classification hierarchy into a hybrid classification hierarchy like the one seen in the lecture.

Dimension_ID	Parent_ID
5801 S Ellis Ave	Chicago
Bavaria	Germany
Berlin	Berlin
Berlin	Germany
Chicago	Illinois
Christophstr. 1	Düsseldorf
Cologne	North Rhine-Westphalia
Düsseldorfer Str. 1	Cologne
Düsseldorf	North Rhine-Westphalia
Illinois	United States
Neuschwansteinstr. 20	Schwangau
North Rhine-Westphalia	Germany
Platz der Republik 1	Berlin
Schwangau	Bavaria
Universitätsstr. 1	Düsseldorf

Hybrid classification hierarchy

Dimension_ID	Level1_ID	Level2_ID	Level3_ID	Level
Universitätsstr. 1	Düsseldorf	North Rhine-Westphalia	Germany	0
Christophstr. 1	Düsseldorf	North Rhine-Westphalia	Germany	0
Düsseldorfer Str. 1	Cologne	North Rhine-Westphalia	Germany	0
Platz der Republik 1	Berlin	Berlin	Germany	0
Neuschwansteinstr. 20	Schwangau	Bavaria	Germany	0
5801 S Ellis Ave	Chicago	Illinois	United States	0
Düsseldorf	North Rhine-Westphalia	Germany	NULL	1
Cologne	North Rhine-Westphalia	Germany	NULL	1
Berlin	Berlin	Germany	NULL	1
Schwangau	Bavaria	Germany	NULL	1
Chicago	Illinois	United States	NULL	1
North Rhine-Westphalia	Germany	NULL	NULL	2
Berlin	Germany	NULL	NULL	2
Bavaria	Germany	NULL	NULL	2
Illinois	United States	NULL	NULL	2
Germany	NULL	NULL	NULL	3
United States	NULL	NULL	NULL	3