

Tasklist for National Archive Case Study

1. Download all appropriate Excel files. (This completes the 'Extract' phase.)
2. Consolidate the Excel files into a single Excel file with a common set of column names.
3. Use the column names to design a star schema data warehouse for the data.
4. Add any necessary metadata and aggregate data tables, together with any further master data tables that are needed.
5. Draw the data warehouse schema in UMLet.
6. Export the data warehouse schema diagram as a .PNG ready to put into your ETL document.
7. Complete the ETL document using the ETL template from Moodle.
 - a. Include the data warehouse schema diagram at the start of the ETL document.
 - b. Identify where the data for each table in the data warehouse comes from (hint: you only have one data source – the Excel spreadsheet).
 - c. For each data warehouse table in turn, identify where the data for each attribute comes from (hint: most will be columns in the Excel spreadsheet although some will be automatically generated).
8. Return to the Excel spreadsheet. Make sure that there are no blank spaces in the named columns of the spreadsheet. Fill any spaces with the value "Unknown" if it is a text field or 0 if it is a numerical field.
9. For each column of text that will become a dimension, copy it to a new worksheet and remove any duplicate values. (This completes the 'Transform' phase.)
10. Copy the main spreadsheet and paste into Access as a new fact table.
11. Copy each of the worksheets with the dimension data and past into Access as a new dimension table.
12. Make sure that each of the dimension tables has an autonumber primary key.
13. Using an UPDATE query, replace the text of a dimension field in the fact table with the corresponding primary key value of the dimension table (hint: the query will have the form `UPDATE fact.dimensionname=dimension.primarykey WHERE fact.dimensionname=dimension.dimensionname`).
14. For each dimension field in the fact table, change the field type from TEXT to NUMBER.
15. Fill any extra data needed in the master data tables. N.B. You do NOT need to create the database triggers used to automatically update the metadata and aggregate data tables.
16. Go to the Relationships tool and create the PK/FK relationships between the tables so that the relationship model matches the schema you created in step 4. (This completes the 'Load' phase.)