Grading criteria that were used for part II of the assignment

The number between brackets represents the relative importance of this criterion. For every criterion 0, 0.25, 0.5, 0.75, or 1 was given, depending on the severity of the error (if any). All issues were penalized under only one category; for instance, a problem with criterion 2 did not affect the evaluation of 6b, 6c, and 6d.

1. (3) Is the order of the dataflow tasks correct? That is, first dimension geography, data and time need to be filled, after that customer and branch, then account and last the fact table. This order is to guarantee referential integrity (foreign key dependencies)
2. (3) Are there special tuples introduced into the dimension tables that replace null-values? Failing to do so will lead to problems when browsing the cubes. Without the introduction of the special tuples the loading of factTransaction becomes much easier; hence a rather high penalty is applied here.
3. (2) Are data and time dimensions fully populated? Given that these dimensions do not change, they should be fully populated right from the initial load such that in incremental loads we do not need to check if a certain date or time already exists.
4. (1) MergeJoin operation should only be used when joining two large tables or when the join condition is not on a foreign key. In case the join is on a foreign key into a small table, the lookup operation is preferable.
5. (5) The script should run without errors (after properly setting all variables and connectors, of course). This includes NOT changing the source database!
6. Are all dimension tables and the fact table filled with the correct data?
    a. (2) dimGeography
    b. (3) dimCustomer (correct dateID)
    c. (1) dimBranch
    d. (4) dimAccount (correct dateID, CID)
    e. (4) factTransaction (correct dateID, AID, insert amount)
7. (3) Wherever possible bulk loading should be used and the use of the “ole db command” should be reduced to a strict minimum as it implies running a query for each tuple that arrives in this component. Doing this for many tuples is extremely time-consuming.
8. (2) Try to avoid the excessive use of scripting and SQL queries as they decrease readability, compatibility, and maintainability.