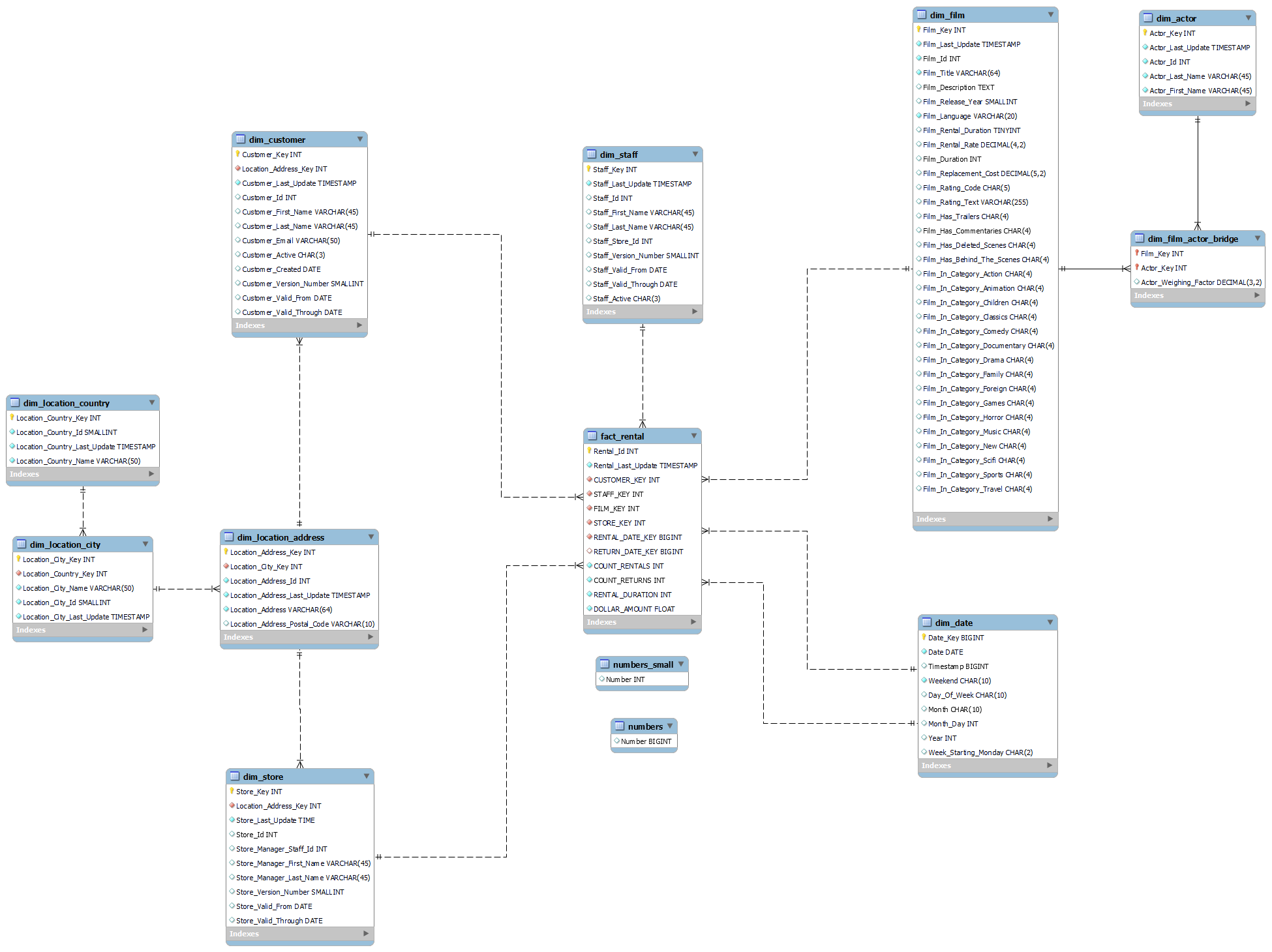
Assignment - 5 - ETL & Fact Table Creation

1. Final sakila\_snowflake EER diagram:



1. Areas Of Improvement:
   1. Are there any warnings that should be resolved? Describe them and recommend fixes.
      1. Script Warning: Updating all the dates without using a where is a very questionable action, as we should never update without a where clause.
      2. While inserting into the film table, instead of using case, an IF function would be easier to understand, as case ideally should only be used when there are multiple cases.
      3. Columns Version Number, Valid from and Valid Through are null as the last update column has been grandfathered in from the relational table. ( Applies to multiple tables)
   2. Are there any orphaned tables that could be safely removed from the data model?  Name them and explain why they can be safely dropped.
      1. Tables Numbers and Numbers\_Small can be safely removed as they were only useful when inserting dates between 2005-01-01 and 2017-01-01.
   3. Are there any changes to column names that would help make the data model more intuitive? Describe them and recommend changes.
      1. Column Dollar\_Amount could be renamed to payment\_amount to be more concise.
   4. Are there any metrics that you think should be added to sakila\_snowflake.fact\_rental? Describe them but do not implement.
      1. Payment Date Column can be added as a key of dim\_date.
   5. Describe any other areas of improvement you see.
      1. The Count\_Rentals and Count\_Returns columns are not very useful as the max value will always be 1 as rentals are always unique.
      2. We need to create a trigger to maintain the DIM\_DATE table for every new day. The trigger would have to calculate the values and insert every day at 12:00 AM to ensure that rentals for that day would not return null.
      3. Inserting all days between two limits is not wise, inserting as and when a new date is found is more time and space efficient.
      4. Film Special Features is retained as a column even after all its data is extrapolated into other columns. At minimum if there is a possibility for extra data to exist, then the redundant data should first be removed.