Correlations

In addition to the well known correlations, such as analysis of objects close to other objects, we will likely need to support a more general usecase: correlating any two attributes. Example: within some selected sample of objects (based say on color redshift and magnitude ranges), is there a correlation between two attributes (for example shape and surface brightness). It is expected no more than ~1-2 billion objects will need to be correlated per query, pairs of attributes at a time [per Tony].

First thoughts:

- Allow such correlations for Object table only (not for Sources, not for ForcedSources)
- Maintain a dedicated server used only for such correlations, with lots of very fast storage (ideally RAM, possibly SSD). Lots = say 64 GB, which is fairly modest given we are talking about 2018 - 2028 time scale
- How to:
    - use shared scan to select a sample that needs to be correlated. Extract only the columns used for the correlation, plus objectId, that is ~16 bytes of data per row, ~32GB of data for 2 billion row sample, plus indexes, should comfortably fit on ~64 GB.
    - use the entire machine and all available fast storage, allow only one correlation at any given time
    - build indexes on the columns to be correlated
    - run the correlation in memory
    - note that the last two steps could be done outside of database if it is more convenient.