

Class of 04/15/2019

Debugging Apache Beam

- Most of these are Python errors, not specifically from Beam. It's just, mistakes made by the programmer may trigger these errors and it's common for the following scenarios in Beam.
- `HttpUnauthorizedError`
 - client is not authorized to connect and perform a certain task
 - likely results from a permission issue (check service account roles, roles of the email you're accessing GCP with)
- `RuntimeError: Transform "<transform name>" does not have a stable unique label`
 - All transforms need to be labeled with a unique label - check transform labels and make sure they're all unique
- `IndexError: list index out of range [while running '<transform name>']`
 - Broadly, we attempted to access a value in a list or tuple (or "indexable class") that is bigger than the size of the list or tuple
- `NameError: global name '<variable name>' is not defined [while running '<transform name>']`
 - Broadly, a variable is not defined as it says
 - In Beam, could stem from the distributed nature of Dataflow, if that runner is used
 - keep in mind that when using Dataflow runners, *only transform code* is executed in each computer so everything you want used that's not passed in like the `PCollection` values themselves must be declared within the transform code.