Big Data Workshop FAQ

Workshop Questions

- Recordings – there will not be a recording of this event. Prior events from a different series can be found at https://www.youtube.com/XSEDETraining

- Slides – the slides are linked into the agenda on the workshop webpage:

  https://www.psc.edu/resources/training/hpc-workshop-big-data-july-2024/

Setup – I missed the start, what do I do:

  Log on to apr.psc.edu and set an initial password if you have not.
  Log on to Bridges-2.
  ssh username@bridges2.psc.edu
  Run the setup script that will copy over the BigData directory we will all use..
  ~training/Setup
  Edit a file to make sure you can do so. Use emacs, vi or nano for beginners.
  Start an interactive session.
  interact

  You can find these instructions on the last slide of the first slide deck:
  https://www.psc.edu/wp-content/uploads/2024/07/Big-Data-Welcome.pdf

Spark Technical Questions

- py4j errors ○
  If you are getting py4j errors usually this means either you have a typo in your file name or you're not in the right directory. You don't get an error when you run the sc.textFile() because it doesn't attempt to read the file until you perform an action later in your code (usually rdd.count()). The solution is to make sure you're in the right directory (usually ~/BigData/Shakespeare) and that you have spelled the file name correctly.

- sc not identified ○
  These errors usually occur when you attempt to run spark on the login node instead of a compute node. Exit pyspark and run

interact.   # (wait for session to start)
module load spark pyspark

- Batch Submission for python spark jobs
Please see the Bridges2 User Guide for information about running Spark jobs in noninteractive mode - https://www.psc.edu/resources/software/spark/

Account Issues

- My password doesn't work ○ Reset your password by going to https://apr.psc.edu
- I don't have an account ○ Check your email to see if you received a message from PSC indicating your Bridges-2 userid. You can also check at the bottom of this page:
    https://allocations.access-ci.org/profile