You will access the following datasets in S3, throught pig (using the LOAD command -- see example.pig)

s3n://uw-cse-344-oregon.aws.amazon.com/cse344-test-file -- 250KB. This is used in example.pig. Always use this file for debugging your scripts first!

s3n://uw-cse-344-oregon.aws.amazon.com/btc-2010-chunk-000 -- 2GB. You will use this dataset in questions 1, 2, 3..

s3n://uw-cse-344-oregon.aws.amazon.com -- 0.5TB. This directory contains 251 files btc-2010-chunk-000 to btc-2010-chunk-317 (since only 251 of the original 318 files uploaded correctly). You will use this in problem 4.