How To Run the Holden and Jacobsen Code on the WRDS Cloud

Overview: The Holden and Jacobsen code provides two SAS programs to access the: (1) Daily TAQ (DTAQ) data and (2) Monthly TAQ (MTAQ) data. Specifically, the programs allow you to select the data that you want from either DTAQ or MTAQ in the WRDS Cloud, cleans the data, computes the National Best Bid and Offer (NBBO), computes standard liquidity measures, and writes output files to your WRDS cloud folder. This code is based on our published paper:

Holden, C. and S. Jacobsen, 2014, Liquidity Measurement Problems in Fast, Competitive Markets: Expensive and Cheap Solutions, *Journal of Finance* 69, 1747-1785.

Our original research code has been adapted to work with the WRDS Cloud and the DTAQ code has been updated for the switch to nanosecond timestamps. Our MTAQ code implements our three recommended MTAQ solutions (adjusting for withdrawn quotes, using our "interpolated time" technique, deleting periods where the NBBO appears to be locked or crossed) for fixing liquidity measurement problems. Our MTAQ code implements our recommended DTAQ procedures (combining data from both the NBBO file and quote file to construct the official, complete NBBO and adjusting for withdrawn quotes). Both programs run entirely on the WRDS Cloud servers. This approach runs much faster than downloading large DTAQ or MTAQ raw data files to your PC and doing the your computations there.

Monthly TAQ spans 1/1/1993 to 12/31/2014 with trades and quotes timestamped to the second. Daily TAQ spans 9/10/2003 to the present with trades, quotes, and NBBOs originally timestamped to the millisecond (one-thousandth of a second). On 7/25/2015 the Daily TAQ timestamps were switched to the microsecond (one-millionth of a second) and on 10/24/2016 the Daily TAQ timestamps were switched to the nanosecond (one-billionth of a second). Our DTAQ code uses nanosecond timestamps throughout even though some of the trailing digits will be zeros during the millisecond and microsecond eras.

Needed code and programs:

1. The Holden and Jacobsen Daily TAQ and Monthly TAQ SAS code can be downloaded from either Craig Holden's web site www.kelley.iu.edu/cholden or Stacey Jacobsen's web site www.smu.edu/Cox/Departments/FacultyDirectory/JacobsenStacey. You will obtain the zip file: "**Holden-and-Jacobsen-Daily-TAQ-and-Monthly-TAQ-Code-2018-03-16.zip**". Then you want to extract the need files from the zip file. You can do this on a PC by opening Windows Explorer, right-clicking on the zip file, click on **Extract All**, and click on **Extract**. You will obtain three files: (1) our DTAQ SAS code, (2) our MTAQ SAS code, and (3) a PDF file with instructions on how to run the Holden and Jacobsen code on the WRDS Cloud (i.e, the instructions you are reading right now).

2. Download and install the free, open-source programs "**PuTTY**" and "**WinSCP.**" You can obtain PuTTY at http://www.putty.org/ and obtain WinSCP at https://winscp.net/eng/download.php.

Steps:

1. **Configure PuTTY and then login to WRDS Cloud.** See Appendix 1 for step-by-step instructions.

2. **Configure WinSCP and then login to WRDS Cloud.** See Appendix 2 for step-by-step instructions.

3. **Make Changes to the Daily TAQ Code and Then Run It.** Launch SAS on your PC, open the Holden-and-Jacobsen-Daily-TAQ-Code-2018-03-16.sas, and make the following changes:

   - **Adapt to your WRDS account.** The first line of code is:
     ```
     libname project '/home/institution/username';
     ```
     Replace "institution" with your WRDS institution name (e.g., "indiana" or "smu") and "username" with your WRDS username.

   - **Select dates.** The DTAQ data is organized into separate data files for each day, because the amount of data is so large. Below the **data** DailyNBBO step, there are requested data files in "nbbo.nbbom_YYYYMMDD" format: `set nbbo.nbbom_20161207 nbbo.nbbom_20161208;` Similarly, below the **data** DailyQuote step, there are requested data files in the "cq.cqm_YYYYMMDD" format: `set cq.cqm_20161207 cq.cqm_20161208;` Similarly, Below the **data** DailyTrade step, there are requested data files in the format "ct.ctm_YYYYMMDD" format: `set ct.ctm_20161207 ct.ctm_20161208;` Replace the three current lists of data files with three similarly formatted lists of data files with the YYYYMMDD trading dates that you want.

   - **Select common stocks.** There is a list of requested common stock symbols in three locations (below **data** DailyNBBO, below **data** DailyQuote, and below **data** DailyTrade):
     ```
     where sym_root in ('AAPL','IBM')
     ```
     Replace the current list ('AAPL','IBM') with a similarly formatted list of common stock symbols that you want. By default, the program selects common stocks (i.e., screens out non-common stocks) with the code line:
     ```
     and sym_suffix = ''
     ```

   - **Alternatively, select other types of securities.** Alternatively, you can select dual shares, preferred shares, etc. by deleting the two SAS code lines mentioned above, enabling the two SAS code lines below (i.e., removing the comment out symbols "/*" and "*/"), and replacing "sym_root" with "symbol" throughout the code below that point.
     ```
     space='';symbol=catx(sym_root,space,sym_suffix);format symbol $17.;
     where symbol in ('AAPL','IBM','BRK A','BRK B') */
     ```
     The new "symbol" variable in the "NYSE symbol format" with up to symbol root of up to 6 characters maximum, then one space, and then a symbol root with up to 10 characters maximum. For details on permitted suffixes, see the latest DTAQ documentation, which as NYSE stock symbol suffixes in appendix B and NASDAQ stock symbol suffixes in appendix C.

   - **Select time-of-day filters.** By default, the data is filtered down to "regular trading hours." Specifically, the NBBO and quote data is restricted to 9:00 am to 4:00 pm:
     ```
     and (("9:00:00.000000000"t) <= time_m <= ("16:00:00.000000000"t));
     ```

and the trade data is restricted to 9:30 am to 4:00 pm:

```
and (("9:30:00.000000000"t) <= time_m <= ("16:00:00.000000000"t));
```

The NBBO and quote data is retrieved 30 minutes before the regular market open in order to be sure that official, complete NBBO quotes are available for the initial trades at the beginning of the regular trading day. However, DTAQ data is available for 16 hours per trading day from 4:00 am to 8:00 pm. By changing the three time-of-day filters you could choose to include "pre-open" data from 4:00 am to 9:30 am and "post-close" data from 4:00 pm to 8:00 pm.

- **Save the updated file.**

4. **Upload the SAS file and run it.** In WinSCP, drag the updated SAS file from your local computer folder to your WRDS account folder. In PuTTY, type "LS" and you will see your SAS file. Also in PuTTY, type "**nohup qsas yourfile.sas**". To explain, "nohup" prevents the command from being aborted if you log out or exit the shell and "qsas" submits the sas job. You will see a notification that you job has been submitted and a job number. If you type "LS" again, you will see a new SAS log file "yourfile.log". In WinSCP, open and view this file to make sure that everything ran correctly.

5. **Examine the results.** Once the job has finished running, type "LS" and you will see three new output files containing standard liquidity measures:
   - "QuotedSpreadsandDepths" contains Quoted Spreads and Depths
   - "EffectiveSpreads" contains Effective Spreads
   - "RealizedSpreadsandPriceImpacts" contains Realized Spreads and Price Impacts that aggregated based on three conventions:

     Ave = simple average, DW = dollar-weighted, SW = share-weighted

   In WinSCP, drag these three files from your WRDS account folder to your local computer folder and examine them in PC SAS.

Additional Notes:

1. **The MTAQ code works in the same manner as the DTAQ code explained above.** One key difference is that the MTAQ code uses the single variable "symbol" to select any type of stock, rather than the two separate variable "sym_root" and "sym_suffix", because the MTAQ database differs from DTAQ on this point. So for example, if you request symbol roots with no suffix (i.e., `where symbol in ('AAPL','IBM')`), then you will get common stocks. By contrast, if you request a symbol root and symbol suffix combination with no space in between (i.e., `where symbol in ('BRKA','BRKB')`), the you will get non-common stocks, such as the dual shares of Berkshire Hathaway "A" shares and "B" shares in this case.

2. **You can access the raw data files and intermediate data files.** At the very end of the Holden and Jacobsen codes are commented out lines to copy six other files from the SAS temporary WORK folder to your WRDS account folder. All six are relatively large files. If you remove the comment outs, then you will get the following:

Three raw data files containing DTAQ data for the firms and dates you selected:
"DailyNBBO" contains NBBO data
"DailyQuote" contains quote data
"DailyTrade" contains trade data

Importantly, the "DailyNBBO" file does NOT contain the complete NBBO. When one exchange has both the best bid and best offer it is only noted in the "DailyQuote" file, not the "DailyNBBO" file. Our code combines data from both files to construct the official complete NBBO (see file below).

Three intermediate data files:
"OfficialCompleteNBBO" contains the official complete NBBO
"TradesPriorNBBOandQuotes" contains trades, prior
    microsecond NBBO, and quotes
"BuySellIndicators" adds buy/sell indicators based on three
    conventions: LR = Lee & Ready (1991), EMO = Ellis, Michaely &
    O'Hara (2000), CLNV =Chakrabarty, Li, Nguyen, & Van Ness (2006)

3. **What to do if your program runs out of space in the WRDS Cloud.** DTAQ and MTAQ files are relatively large. If you select a large number of firms and long time period, then it is possible for SAS to run out of available work space in the WRDS Cloud and your program will crash. If this happens to you, cut it into a series of smaller programs. For example, you might divide a program using a year of data into four programs using a quarter worth of data each or twelve programs using a month worth of data each. The number of trades and quotes is growing about 30% per year compounded year after year. So a program using data from recent years will have much bigger files than a program using data from earlier years.

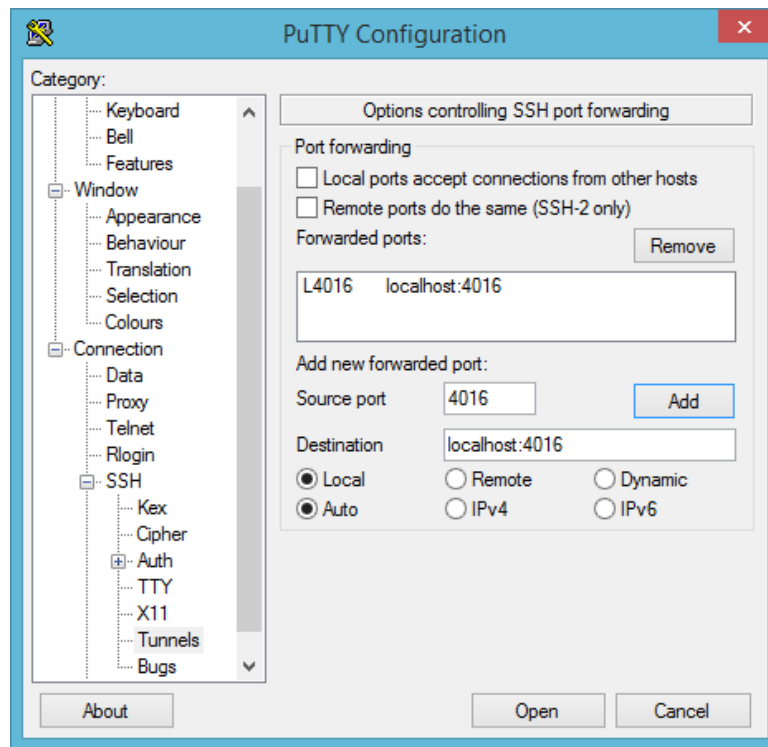# Appendix 1: How to Configure and Run Putty to Connect to WRDS

1. After you have downloaded and installed **PuTTY**, then:
   a. Launch **PuTTY**
   b. Type in the Host Name, **wrds-cloud.wharton.upenn.edu**
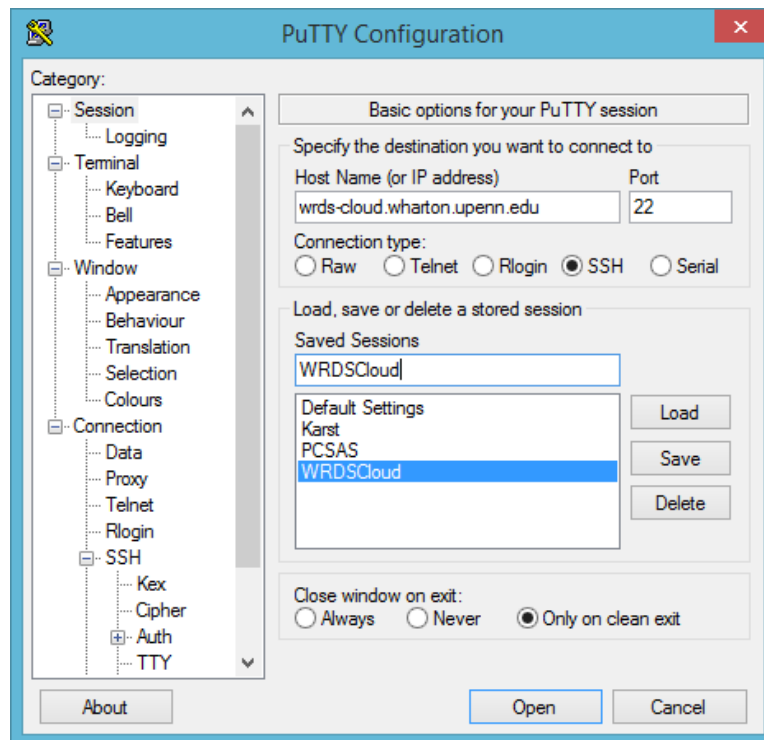   Default port should be **22**
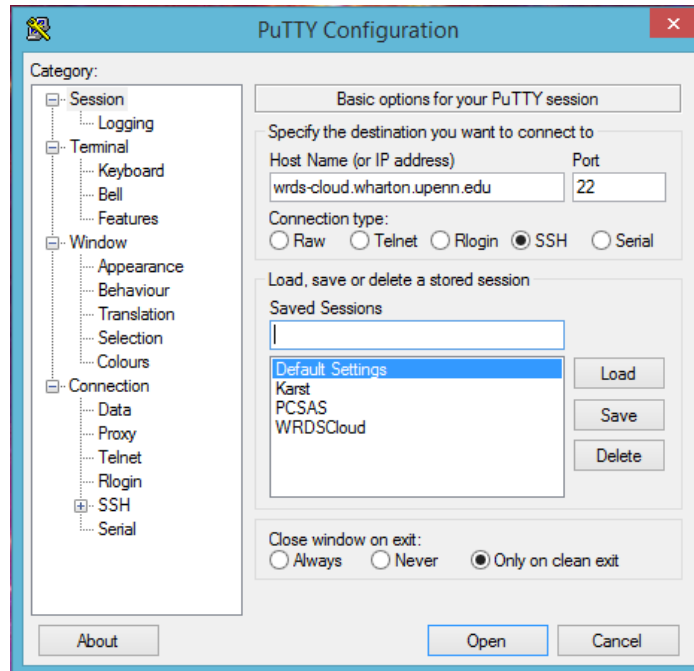
   c. Select **Connection->SSH->Tunnels**
   d. Enter Source Port **4016** and Destination **localhost:4016**
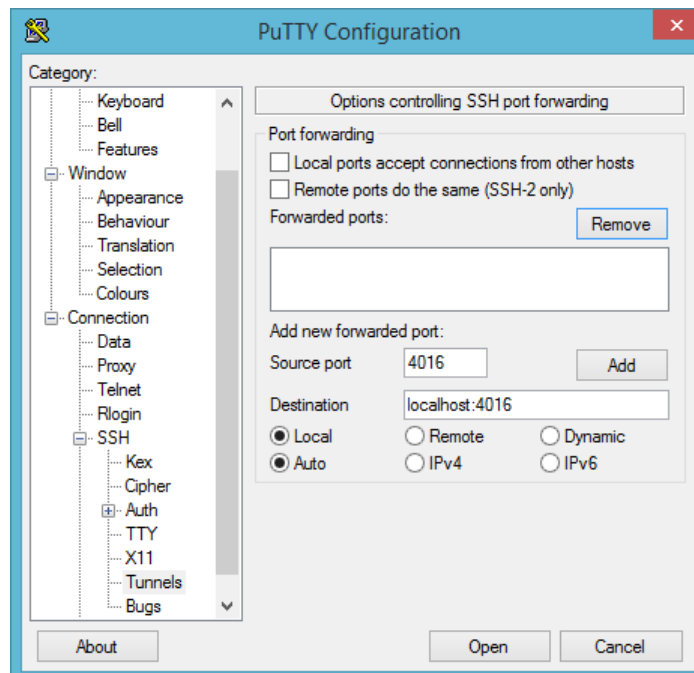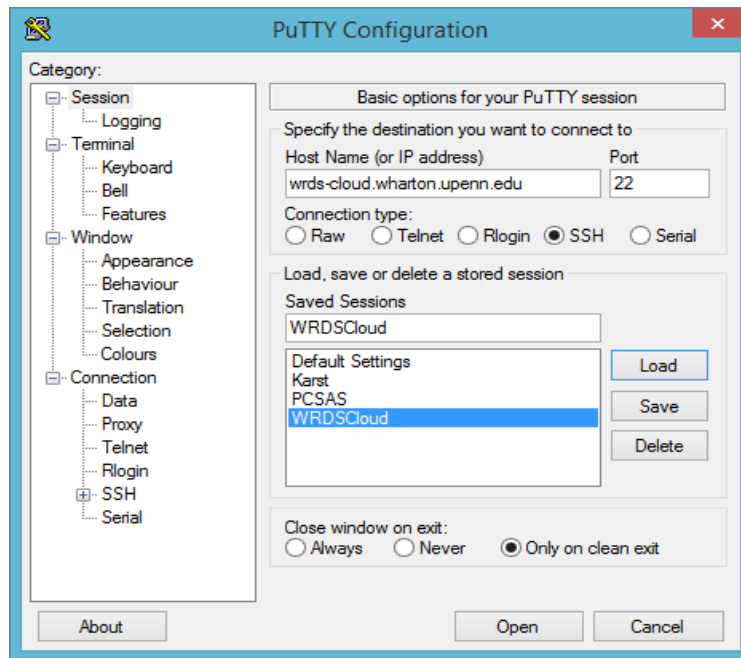
e. Click "**Add**"; It should now look like this:

f. Go back to **Session**. Enter **WRDSCloud** (or whatever you want) under **Saved Sessions**

Click "**Save**". You can now select a **WRDSCloud** session from **Saved Sessions** in the future.
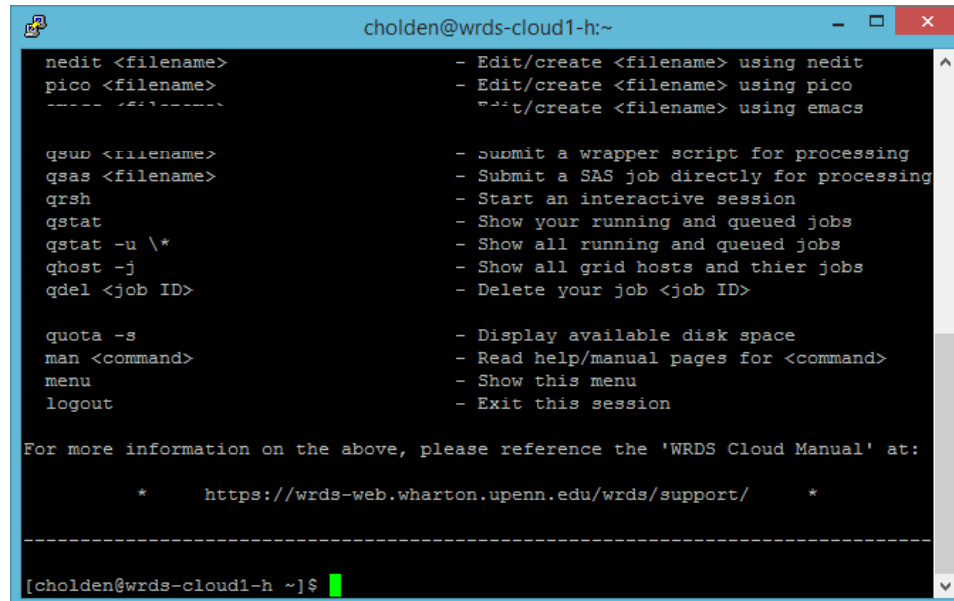
2. Launch **PuTTY**,
   a. click "**WRDSCloud**" then "**Load**" and then "**Open**" (or else just double- click "**WRDSCloud**") to open a new session.

   b. When you open a session, it will say: "login as:". Type your WRDS username and press enter. Then it will ask for your password. Type your WRDS password and press enter.
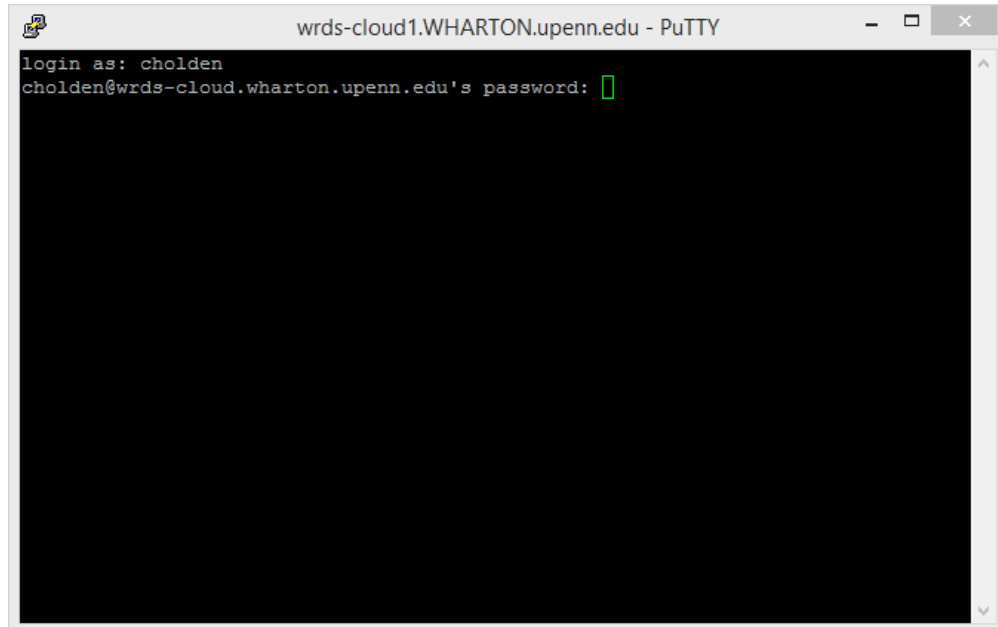
An established connection generates a "wrds-cloud" prompt that looks like this:



```
nedit <filename>                    - Edit/create <filename> using nedit
pico <filename>                     - Edit/create <filename> using pico
                                     E ''t/create <filename> using emacs

qsub <filename>                     - Submit a wrapper script for processing
qsas <filename>                     - Submit a SAS job directly for processing
qrsh                                - Start an interactive session
qstat                               - Show your running and queued jobs
qstat -u \*                         - Show all running and queued jobs
qhost -j                            - Show all grid hosts and thier jobs
qdel <job ID>                       - Delete your job <job ID>

quota -s                            - Display available disk space
man <command>                       - Read help/manual pages for <command>
menu                                - Show this menu
logout                              - Exit this session

For more information on the above, please reference the 'WRDS Cloud Manual' at:

        *       https://wrds-web.wharton.upenn.edu/wrds/support/       *


-----------------------------------------------------------------------------

[cholden@wrds-cloud1-h ~]$
```

Appendix 2: How to Configure and Run WinSCP to Connect to WRDS

1. After you have downloaded and installed WinSCP, then:
   a. Launch **WinSCP**
   b. Click on the **New Site** button in the upper left column
   c. Type in the Host name, **wrds-cloud.wharton.upenn.edu**
   d. Enter your WRDS user name
   e. Enter your WRDS password
   f. Click on **Save** and accept the default name or enter new name for it
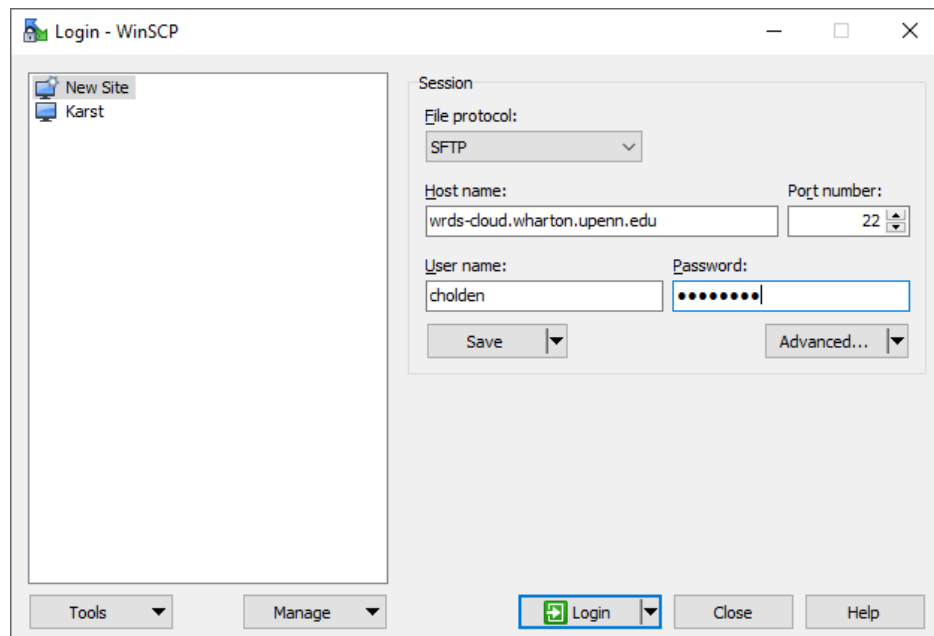
2. Launch **WinSCP**,
   a. Double-click the "**username@wrds-cloud.wharton.upenn.edu**" button (or whatever you named it) in the left column to login
   b. Confirm that this is what you want to do
   c. Enter your WRDS password
   d. A WinSCP session will open with two windows: (1) the local folder in one window and (2) the host folder in the other window – thus making it very easy to transfer files in either direction