1 Accessing TAQ data from D-CAF Servers at Aarhus University

As of 2021, students can only access the TAQ database at the Fuglsangs Allé campus using the computers located in C0. You will still need login credentials provided you contact us.

In order to access and download Trade And Quote (TAQ) data from the servers at Aarhus University you need:

1. A working copy of MATLAB

    (a) **Students:** Computers in room C0 (the one with Bloomberg Terminals installed) already have MATLAB installed.

2. Access to the networks at Aarhus University either by Ethernet cable (at Campus) or VPN tunnel (WiFi at Campus or access from outside Campus).

3. Login for the network drive:

    (a) Username: uni\svc_taquser

    (b) Password: Contact dbsupport@econ.au.dk for the password[1][2]

4. Template files available on the server

1.1 Access

The TAQ data is saved on a local server in HDF5[3] format. This highly compressed format enables us to store large amount of data in small footprint. The HDF5 format is widely supported by most major programming languages[4], and has libraries independently developed[5]. If you are located on AU campus and using Ethernet cables or WiFi you have access to the au network. Otherwise you must use a VPN tunnel https://remote.au.dk. If you are employed at AU you use your AU username. Others must ask for an AU username.

After ensuring that you are connected to Aarhus University's network, you have to add the network drive located at "\\cafmaster.econ.au.dk\taq\". On Windows this is done by right-clicking "Computer" and choosing "Map Network Drive". After this you are presented with a window as in Figure 1. Make sure that you to check the "Connect using different credentials" and write the above username and password.
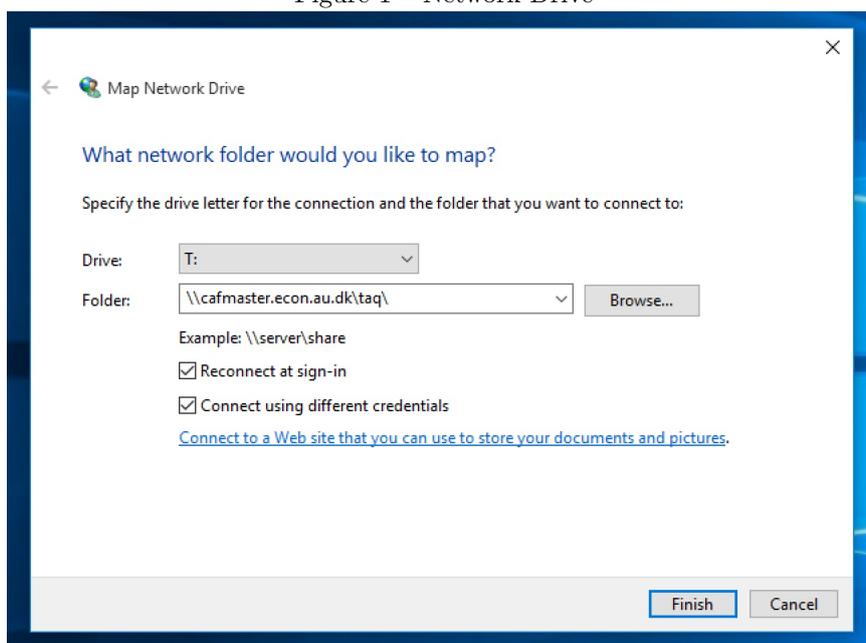
---

[1]The password changes every year around october.

[2]Please contact through your study mail otherwise your inquiry will be marked as spam.

[3]`http://www.hdfgroup.org/`

[4]`http://www.hdfgroup.org/products/hdf5_tools/index.html`

[5]`http://www.hdfgroup.org/products/hdf5_tools/SWSummarybyName.htm`

Figure 1 – Network Drive

If you cannot gain access to the network drive ensure and check if you have fullfilled Step 2 above and make sure that you check the troubleshooter in section 1.3. Otherwise, contact *dbsupport@econ.au.dk*

For Mac users the address is "smb://cafmaster.econ.au.dk/taq".

1.2   Downloading data

In the TAQ folder we have stored the raw TAQ files in HDF5 format, .h5 files, as well as several program files for MATLAB. If you need heavy computations you can apply for access to the cafux/creates cluster. If just need to retrieve data you can use your own computer. You will need a folder on your computer for storing MATLAB files. In these examples I have created the folder **C:\TAQ.** In "matlab\windows" you will find two folders named hdf5 and templates. Copy the files from the directory **C:\TAQ\matlab\windows\hdf5** to **C:\TAQ.** They must always be available in your MATLAB path. You can copy and modify the files in the template as needed.

In "hdf5" there are the three functions we need to get access to the TAQ data, each defined as an .m file. **convert_to_escaped.m** is a helper function that cleans up any issues with characters. **taq_h5_filename_daily.m** is helper function that tells the **taq_h5_read_daily.m** file where to look for the .h5 files.

Figure 2 – taq_h5_filename_daily.m

```matlab
function [shortFileName,fullFileName] = taq_h5_filename_daily(date,base)

if nargin==1
    base = [];
end

shortFileName = ['taq' num2str(date) '.h5'];

if isempty(base)
    base = '//cafmaster.econ.au.dk/TAQ/TAQHDF5/';
end
fullFileName = fullfile(base,shortFileName);
```

The function **taq_h5_filename_daily.m** (Shown in Figure 2) takes 2 arguments, date and base. The date argument is used to construct the filename to look for in our database, while base is the location of the actual data. You do not need to specify the database if you are using the standard database. If base is not specified, the function will by default assume that the base is "//cafmaster.econ.au.dk/TAQ/TAQHDF5/".
taq_h5_read_daily.m contains the function that does the actual reading of the data. If you are used to load in .txt and .csv files in MATLAB, then consider this as a similar function, where instead of .csv/.txt files we load in a special type of file, .h5, called HDF5. The taq_h5_read_daily function takes 5 arguments: type, ticker, date, base, metaData. These are explained in Table 1.

Table 1 – Arguments

Argument	Value	Mandatory
type	Character - "Quotes", "QuoteIndex", "Trades" or "TradeIndex"	Yes
ticker	Character - Ticker name (e.g. "IBM", "MSFT", or "SPY")	Yes
date	Date - Must be open trading day	Yes
base	file path - Location of the HDF5 files	No
metaData	Helper-data used when loading indeces	No

Finally, in the folder "templates" we have files called **many_days.m** and **daily.m**. Many days retrieve data for several days. These MATLAB files put together all the functions explained above, and shows you how to access the data. Open up the **daily.m** file in MATLAB and you will see Figure 3. Remember to modify "addpath" if you have copied the hdf5 files somewhere else.

Figure 3 – daily.m

```matlab
clear;
clc
close all;

% Matlab hdf5 utilities for TAQ
addpath('C:/Taq');

TAQTr = taq_h5_read_daily('Trades', 'SPY', 20020521);

% You can also use 'Quotes', and change 'SPY' to another ticker.

TAQQu = taq_h5_read_daily('Quotes', 'SPY', 20020521);

% To make it faster with repeated stocks on the same day

metadata = taq_h5_read_daily('TradeIndex', [], 20020521);
```

Next we consider the first example of reading in trades, to the object called TAQTr. Here we call the function **taq_h5_read_daily** using the three mandatory arguments as specified earlier. We want "Trades", we want it for the S&P500 tracker, which has the ticker name "SPY" and we want it for the 21th May, 2002 - 20120521. Running the first few lines of code, should now return a 1x1 struct in your workspace called TAQTr, which holds the data requested as shown in Figure 4. Note here that figure 4 only applies to trades from 1993 to 2013. When running the code for dates within 2014 and forward we obtain a different and longer 1x1 struct. In figure 5 we extract data for the 21th of May, 2014 (20140521). The different variables are explained in Table 2.
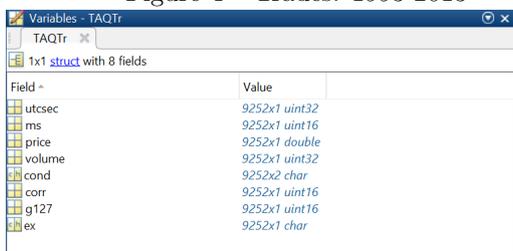
Figure 4 – Trades: 1993-2013
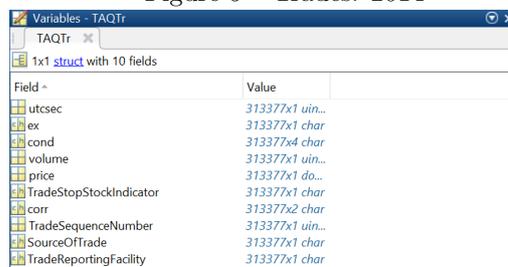

Figure 5 – Trades: 2014 –

Table 2 – TAQ Variables - Trades

utcsec	Coordinated Universal Time in seconds - minus 5 hours. Similar for 2014– but in nanoseconds
ms	Obsolete (always equal to zero)
price	Price
volume	Volume
cond	Condition of trade
corr	Correction
g127	?
ex	Exchange abbreviations, see Table (4)
TradeStopStockIndicator	?
TradeSequenceNumber	Message sequence number
SourceOfTrade	?
TradeReportingFacility	Exchange where trade was reported, if applicable.

Similarly, we can request quotes for the same two dates as we did for trades, this is done in the next example :

```matlab
clear;
clc
close all;

% Matlab hdf5 utilities for TAQ
addpath('C:/TAQ');

TAQTr = taq_h5_read_daily('Trades', 'SPY', 20020521);

% You can also use 'Quotes', and change 'SPY' to another ticker.

TAQQu = taq_h5_read_daily('Quotes', 'SPY', 20020521);

% To make it faster with repeated stocks on the same day

% metadata = taq_h5_read_daily('TradeIndex', [], 20020521);
```

In this case, we request the quotes for the same day and ticker as with the trades. Now we have a TAQQu 1x1 struct in our workspace with the data requested as shown in figure 7 and figure 8. The different variables are explained in Table 3.

Figure 7 – Quotes: 1993-2013

Figure 8 – Quotes: 2014 –

Table 3 – TAQ Variables - Quotes

utcsec	Coordinated Universal Time in seconds - minus 5 hours. Similar for 2014– but in nanoseconds
ms	Obsolete (always equal to zero)
bid	Bid price
ofr	Offer price
bidsize	Bid size in number of round lots (Counted in 100s)
ofrsize	Offer size in number of round lots (Counted in 100s)
mode	Quote condition
ex	Exchange on which the quote occurred (See Table (4))
MarketMaker	?
BidExchange	?
AskExchange	?
SequenceNumber	Message sequence number
NationalBBOInd	?
NaqdaqBBOind	?
QuoteCancelCorrection	Indicates that this record is a cancel or a correction of a previous quote
SourceQuote	?
RPI	?
ShortSaleRestrictionIndicator	Short Sale Restriction status
CQS	?
UTP	?
FinraAdfMpidIndicator	?
SIPGeneratedMessageIdent	?
NationalBBOLuldIndicator	?

Table 4 show the Exchanges and their corresponding abbreviations. Which are denoted as ex in table 2 and 3.

Table 4 – Exchanges and abbreviations

| Exchange name	Abbreviation
NYSE	N
Boston	B
Arca	P
INSX	C
NASDAQ	T
NASD ADF	D
Philadelphia	X
ISE	I
Chicago	M
CBOE	W
BATS	Z

For more information on Daily TAQ Client Specification go to https://www.nyse.com/publicdocs/nyse/data/Daily_TAQ_

1.3   Troubleshooting

1. Are you currently connected by Ethernet, WiFi or VPN to AU's network?

2. Have you properly copied the template files to your own computer?

3. Have you added the correct folders, with the MATLAB files containing functions, to your MATLAB Path?

4. Is the ticker you are using correct?

5. Is the date provided a trading day on the exchanges? `http://www.nyx.com/en/holidays-and-hours/nyse`

6. If you get the SMB1 protocol error (Windows 10) then use the following step:

The error message would look like:

Figure 9
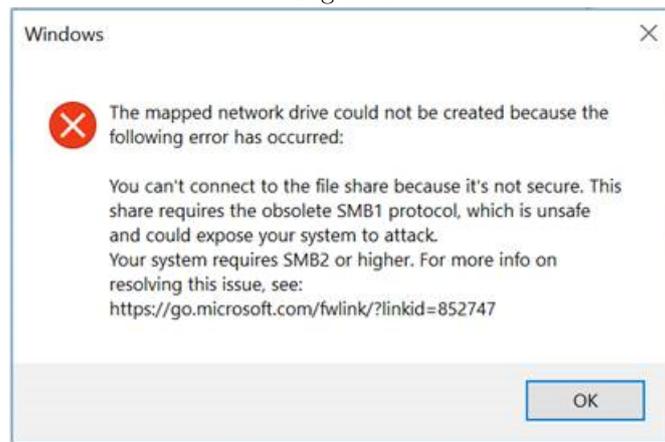
The solution is then:

- Connect the laptop to AU network by Ethernet cable, WiFi or VPN

- Press 'start' and type in 'Windows PowerShell'

- Right-click on 'Windows PowerShell' and choose 'Run as Administrator'

In the Windows PowerShell command window type in the following command: **Enable-WindowsOptionalFeature -Online -FeatureName smb1protocol**
Observe that there are three spaces marked by red in the figure below

Figure 10
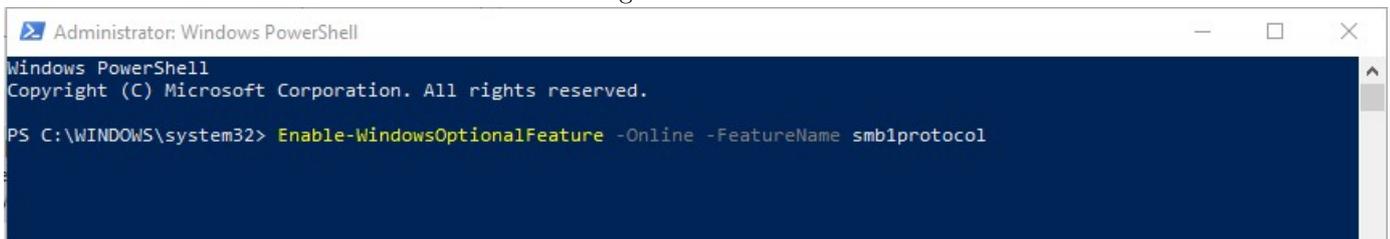
- Press Enter and wait (this can take a few seconds)

- Restart Windows and reconnect to TAQ.