**Notes on the CCASS schema**

**Copyright and disclaimer**

Please read the separate “Copyright and disclaimer” Notice.

**Background**

The Central Clearing And Settlement System (**CCASS**), launched in 1993, is operated by Hong Kong Securities Clearing Company Limited (**HKSCC**), which was then a not-for-profit company, but in 1999 was [subsumed](https://webb-site.com/articles/uturn.asp) into the merger of the Stock and Futures exchanges to become a wholly owned subsidiary of Hong Kong Exchanges and Clearing Limited (**HKEX**) which also owns the Stock Exchange of Hong Kong Limited (**SEHK**).

Shares must be deposited into CCASS in order to settle sales on SEHK, so nearly all the publicly held shares are held in CCASS (mostly via custodians, banks and brokers). CCASS appears on the issuer’s Register of Members (or share register) as a single shareholder, HKSCC Nominees Limited (**HKSCCN**).

That makes inspecting the share register (which can only be done in person) rather useless. In order to re-establish the level of transparency that existed before CCASS, while I was a shareholder-elected (not Government-appointed) Independent Non-Executive Director (**INED**) of HKEX, I lobbied to get the CCASS data published. The service was [launched](https://www.hkex.com.hk/News/News-Release/2008/0804232news?sc_lang=en) by HKEX on 28-Apr-2008, with data extending back for 1 year.

**The CCASS schema**

The CCASS schema contains most but not all of what you need to reproduce the Webb-site CCASS analysis. You will also need, from the enigma schema:

* The issue table, which contains details of each issue of securities
* The organisations table, which contains the details of each issuer of securities
* The secTypes table, which has details of the security type
* The stockListings table, if you want to find past and present stock codes of each issue. Note that HKEX recycles stock codes over time, and some companies have changed stock codes either because of a move from the GEM to Main Board; because they had a special arrangement; or because they delisted and relisted.
* The listings table, which has details of each board on each exchange (1=HK Main Board, 20=HK GEM, 23=HK REITs and so on).
* The log table, which includes various items such as CCASSdateDone and URLs for scraping.

However, we have also produced a CSV with extracts which contain details of all the issues in CCASS.holdings, the name and ID of the issuer.

**One schema, Two tables**

The ccass schema has 2 very large tables which contain the same data with different primary keys:

* holdings, which is arranged with primary key (issueID, partID, atDate). partID is our participant ID. This is used to analyse changes in the holders of a stock between 2 dates.
* parthold, which has primary key (partID,issueID,atDate) and is used to analyse changes in a participant’s holdings in all stocks between 2 dates or in one stock over time.

You might think that it would be easier to just put another Index on the holdings table instead of storing the same data in parthold with a different PK. We tried this in the past and found that the query speed was too slow because of the fragmentation of the data – but your mileage may vary, so if you think you can run it off one table, have at it. Remember we’ve been running the CCASS system for 18 years so technology may have improved by now.

To save space, we don’t record unchanged CCASS holdings; only about a quarter of the holdings change from one day to the next. This does make the computation of changes between 2 dates a little more challenging, but it also allows us to show the date on which the holding last changed.

Our ccass dataset stretches back to 26-Jun-2007, unlike HKEX, which deletes it from public view after a year and has never made any effort to add the functionality for investors that we have produced. The SFC could also do that – it is within their [statutory remit](https://www.elegislation.gov.hk/hk/cap571?xpid=ID_1438403465238_001). We also keep all the data for delisted stocks. On the delisting date, we set the final holdings to zero as the stock is removed from CCASS.

The ccass schema needs to be resident in RAM if you want it to run at any reasonable speed. Our collection code typically takes about 45 minutes to run on a Windows 10 PC with 32GB of RAM, so configure MySQL accordingly. The Webb-site.com server has 128GB of RAM, more than ample because it runs other tasks.

**Dailylog**

This table contains the summaries of each CCASS stock on each day, including our calculation of the concentration of CCASS holdings (top 5, top 10 etc).

**Quotes**

The schema also contains 2 tables with historic prices, volumes and turnover – these were collected from the official Daily Quotations Sheets (**DQS**) for the Main Board and GEM, at SEHK. The tables are “quotes” and “pquotes”, the latter for when a stock is in “parallel trading” typically during a stock split or consolidation.

Parallel trading is a unique quirk of HK – as an HKEX INED I campaigned for it to be abolished, and on 22-Apr-2008, HKEX [announced](https://www.hkex.com.hk/news/news-release/2008/080422news?sc_lang=en) that it would be, but on 23-Jul-2008 (after I had resigned) the move was [delayed](https://www.hkex.com.hk/news/news-release/2008/080723news?sc_lang=en) to “the second half of 2009” and 16 years later, still hasn’t proceeded.

Our quotes database stretches back to 3-Jan-1994, so we can also calculate “Webb-site Total Returns” on all stocks since that date. That requires a lot of adjustments for dividends, rights issues, bonus issues, share splits/consolidations, bonus warrant issues and so on; the details are in the enigma schema.

**Specialdays**

The “specialdays” table records days on which the market did not trade in either the morning, afternoon or both – including public holidays and bad-weather days, or when the morning session started late or either session was terminated early due to weather.

There are also some days which trade but have no settlement – for example, a number of eves of Christmas, New Year or Chinese New Year in which the market closes at lunch time.

Each year, we need to enter the public holidays in advance, and during each year, we need to enter any days where the market is disrupted by weather or other events, then regenerate the settlement calendar (see below) so that the CCASS data make sense.

**Calendar**

The “specialdays” table is used by our code to generate the entries of another table, ccass.calendar, which contains the settlement dates corresponding to each trade date (normally T+2).

As a consequence of special days, there are some trade dates which share a settlement date.

**Shortnames**

This table contains the short-names of stocks used by SEHK in the Daily Quotations Sheets. Each day, our code adds new-listed stocks to this or finds stocks which have changed their short-name (usually because of a company name change).

**Unquotes**

If a new stock has appeared in the DQS but is not yet in our enigma.issue table, then we temporarily store the quotes in the unquotes table.

**sehkmonthend**

I can’t remember why I had this table, which records the last trading date of each month from Nov-2006 to Dec-2010. Seemed like a good idea at the time. We might have been trying to record outstanding shares at month-end, but we now do that based on the Monthly Returns at calendar month-end. You can drop this table if you don’t need it.

**Maintaining the dataset**

To run our code on the collection machine you will need an ODBC connection with full read-write privileges on the ccass tables.

Obviously, you need to collect the data every day, so it’s important to get your system up and running soon after the date of this data dump, before the data disappear.

To make that easier, we intend to provide at least 2 data snapshots, so that you can set up your system with the first one, test your collection code, and then import the data from the second snapshot.

Beware that if you miss a day, and the stock delists, then HKEX deletes the CCASS records immediately, so you won’t be able to go back and collect the data for that stock.

For the quotes data, the DQS are on the HKEX web site for the current and previous month – so on the first day of a month, you can only look back one month.

For CCASS data, we are including our VB.net code, in ccass.VB, ScraperKit.vb and JSONkit.vb. ScraperKit is our module of commonly used subs and functions. JSONkit is our self-developed module for reading JSON packages.

For the quotes data, we are including quotes.vb, which collects the daily prices. That module also has code to calculate the adjustments used in “Webb-site Total Returns”, including collecting FX data from Yahoo, collecting the entitlements data from the HKEX quote pages, and converting dividends which are declared in a currency different to the quote currency (or taking the issuer’s converted rate if they disclose it).

The HKEX entitlements data appear are hand-coded, so can sometimes contain errors. They can also contain details (such as a distribution in specie of shares in another company) which are beyond the scope of our code. So, for a few entitlements each month, we have to process them manually. In the Access application (enigma.accdb) there is a form called “Undone Entitlements” where those show up.

**Rendering the data on a website**

Obviously you can present the data any way you want, but we are including the ASP files used in the CCASS directory on Webb-site.com as a starting point. Some of these include SQL code, while others call MySQL stored procedures in ccass.

These ASP files also include some files from the dbpub directory on Webb-site, namely the core “functions1.asp” library and the “navbars.asp” file for navigation menus.

If you are using MySQL then on the web server you will need an ODBC connector with a user which has full read access to the ccass tables.

DMW

2025-04-15