* **Introduction of the New Data**

The Raw Data from Vertex includes 6 different types of data sets:

1. **ALL TRADES** (All trades and Volume): HELM\_OREC\_TRD
2. **ALL EVENTS** (all quotes + trades + admin messages): HELM\_OREC\_ALL
3. **PROCESSED TRADES** (includes Summary, Detail and Volume): HELM\_BOOK\_TRD
4. **BOOKS W/DEPTHS** (After each book event get a summary of the book): HELM\_BOOK\_QUO
5. **BOOKS ALL** (Define book depth and trades): HELM\_BOOK\_ALL

In both BOOKS W/DEPTHS and BOOKS ALL there is an additional field at the end of the URL that can be changed: Helm\_depth=

It can be set between 1 – 10 and signifies how many levels of the book you want the summary to have after each book event

1. **DAILY DATA** (get a range of daily data statistics like OHLC): HELM\_DEOS\_ALL

“Helm\_date=” is changed to “Helm\_date\_from=” and “helm\_date\_to=” which allows a range of data to be pulled i.e. not just a single day

* **Expected Scraping Process**

From the URL it offers, we are able to scrape all the 6 forms of data by using the downloader url with corresponding string of Date, Futures Name and the Name of Data Type above.

For all possible combinations:

If returned fExtnDesc=READY, download the url of zip file after fLeadpath string;

Else if fExtnDesc=TRANSFERING, rerun this combination;

Else break.

Unzip the zip file to csv.

* **The Types of Data:**

Raw data can be retrieved from as far back as 1/1/2011. However, during this period of time the CME changed the message protocol they used. This switch means that the messages returned from a raw data pull on dates **prior to 6/30/2015** will be in the **Fix/Fast** format and messages from dates **post 6/30/2015** will be in the **MDP 3.0** format.

Example:

**FIX/FAST**



In Fix/FAST, there are following keys:

1. QU = QUOTE: Message that contains information on the event that occurred in the book
2. TR = TRADE: Message that contains information about a trade that occurred
3. AD = ADMINISTRATION: Message that contains different market statistics. The type of Administration message is found in the TranType field OP, CL, ST, HI, LO, OI, etc.
4. SC = SUMMARY COUNTS MESSAGE: Message that provides the number of matches from a trade event
5. T = TRADE: Message that contains information about a trade that occurred (\*)
6. V = VOLUME: Message that updates volume data after a trade event (\*)
7. DH = DEPTH HEADER: Message that provides what changed in this depth update
8. Q = QUOTE: Message that contains information on the event that occurred in the book (\*)
9. (And Daily Summary)

(\*) --- Different from the Keys with same names in MDP 3.0

**MDP 3.0**



1. MBO: Market By Order (MBO) files provide all market data messages required to recreate the book. The consumer can see every level of every order in the book, as well as trade data for all CME Globex- traded products.
2. Q = QUOTE: Message that contains information on the event that occurred in the book
3. T = TRADE: Message that contains information about a trade that occurred
4. M = TRADE MATCH ELEMENT: Message that contains information on a match of a trade event
5. V = VOLUME: Message that updates volume data after a trade event
6. ST = DAILY STATISTIC: Message that provides information about a complete session
   1. Cleared Volume
   2. Fixing Price
   3. Open Interest
   4. Settlement Price
7. LM = LIMITS DAILY: Message that provides the daily limits for the current session
8. DS = SESSION STATISTICS: Message that provides information about the session during the session
   1. High bid and Low ask
   2. Indicative opening
   3. Opening price
   4. Session High and Low
9. SS = SECURITY STATUS MESSAGE: Message that provides the security group market state change
10. SC = SUMMARY COUNTS MESSAGE: Message that provides the number of matches from a trade event
11. DH = DEPTH HEADER: Message that provides what changed in this depth update
12. (And Daily Summary)

* **Variable Dictionary of the New Data Map with MDP 3.0 and Format**

1. **key**

Key (For all Keys): Message type (String)

1. **Time**

Time (For all Keys): Time, in nanoseconds, that the message occurred (Timeformat(hh:mm:ss.sss sss sss))

1. **GSEQ**

Gseq (Not for M/m): **Packet** level sequence number per channel (String)

\*\***Packet**: Each packet sent on all feeds (Incremental, Market Recovery, Instrument Definition, and TCP Replay) contains a packet sequence number and a timestamp. The packet number is unique to each packet sent and each channel has its own separate set of sequence numbers that increment sequentially with each packet and are reset weekly.

Packet = Packet Header + Message;

Packet Header = Packet Sequence + Time

Message = Message Header + FIX Message

Message Header = Message Size + Simple Binary Header (Block Length, TemplateID, SchemaID, Version)

<https://www.cmegroup.com/confluence/display/EPICSANDBOX/MDP+3.0+-+Packet+and+Message+Headers>

Q: Gseq

1. **Template**

TemplateID (Not for M/m and SS): Identifies what data is in this message (Int)

1. **SecurityID**

SecurityID (For all Keys): Unique instrument ID (Int)

1. **UAction**

MDUpdateAction (In T, Q and ST): MD Update Action 0=New, 1=Change, 2=Delete, 3=Delete Thru and 4=Delete From (Char)

OrderUpdateAction (In MBO): Order book update action to be applied to the order referenced by OrderID. 0=New, 1=Update, 2=Delete (uInt8)

1. **Side**

Updatetype (For MBO): Side of the book. 0 = Bid And 1 = Offer (Int)

SideOfLimit (For LM): Side of the limit. Can be high or low (String)

1. **Price**

MDEntryPx (For T, Q, MBO, LM, ST (SETTLEMENT, FIXING PRICE) and DS (OPENING PRICE)): Price of the MD Entry (Float, $)

1. **Quantity**

LastQty (For T & M/m): Quantity matched for this order (Int)

MDEntrySize (For V, Q & ST (CLEARED VOLUME)): Quantity of the MD Entry. In a Book Entry– order quantity In a Trade Entry – traded quantity In Electronic Volume - represents cumulative traded volume of the Daily Trade session. In an Indicative Opening - the indicative opening quantity (Int Qty)

MDDisplayQty (For MBO): Visible quantity of an order to the market (Int)

1. **OrderID**

OrderID (For M/m & MBO): Unique ID assigned by CME Globex to identify orders. (Int)

1. **Priority**

MDOrderPriority (For MBO): Order priority for execution on the order book. A lower value is a higher priority. Value is unique for all orders within a market segment and assigned for all orders. (Int)

1. **Seq**

Seq (Not for M/m, MBO & SS): Instrument level sequence number (reset weekly to 1 on Sunday) (String)

1. **bitstr**

Bitmask String (Not for M/m, MBO & SS): Used to determine if this is the last event for a given message. (String)

MatchEventIndicator (For SS): Used to determine if this is the last event for a given message. (String)

(CME Descriptive: "Bitmap field of eight Boolean type indicators reflecting the end of updates for a given CME Globex Event:

Bit 0: (least significant bit) Last Trade Summary message for a given event

Bit 1: Last electronic volume message for a given event

Bit 2: Last real quote message for a given event

Bit 3: Last statistic message for a given event

Bit 4: Last implied quote message for a given event

Bit 5: Message resent during recovery

Bit 6: Reserved for future use

Bit 7: (most significant bit) Last message for a given event")

1. **TxtSide**

MDEntryType (For T & Q):

Is it a Bid or an Offer. EOE (end-of-Event) appears when more than one book update comes on the same message and is used to announce the end of book updates for the message.

(Char)

1. **Plevel**

MDPriceLevel (For Q): Aggregate book level, any number from 1 to 10. (Int)

1. **orders**

NumOrds (For T & Q): Number of orders at this book level (Int)

1. **Aggressor**

AggressorSide (For T): "Indicates which side is aggressor of the trade. If there is a zero value present, then there is no aggressor. 1=Buy, 2=Sell and 0=Implied

Note: Trades without aggressors occur at Market Open, after a Pre-Open or after a Pause, and also when the event includes real order participation in a trade with a CME Globex-generated implied bid or offer." (Int)

1. **matchNum**

MatchNumber (For M/m): Order of this match (Int)

1. **etype**

MDEntryType (For V, LM, ST & DS): Is it a Bid or an Offer. EOE (end-of-Event) appears when more than one book update comes on the same message and is used to announce the end of book updates for the message.

(Char)

1. **settype**

SettlPriceType (For ST (SETTLEMENT)):

"Which type of settle price. Bitmap field of eight Boolean type indicators representing settlement price type:

Bit 0: (least significant bit):

1=Final

0=Preliminary

Bit 1:

1=Actual

0=Theoretical

Bit 2:

1=Settlement at Trading Tick

0=Settlement at Clearing Tick

Bit 3: (under development)

1=Intraday

0=Undefined

Bit 4-6: Reserved for future use, set to 0

Bit 7:

0=not NULL

1=entire set is a NULL"

(String)

1. **OIYN**

IsOpenInt (For ST (SETTLEMENT)): 0 for a settle message and filled for an OI message (Int)

1. **date**

Date (For ST): Date of trade session corresponding to the statistic entry (Timeformat (mm/dd/yyyy))

TradeDate (For SS): Indicates the date of the trade session. (Timeformat (mm/dd/yyyy))

1. **OpenInt**

MDEntrySize (For ST (OPEN INTEREST)): Quantity of the MD Entry. In a Book Entry– order quantity In a Trade Entry – traded quantity In Electronic Volume - represents cumulative traded volume of the Daily Trade session. In an Indicative Opening - the indicative opening quantity (Int Qty)

1. **ERROR**

ERROR = 1 If ST etype = NA; ERROR = 2 If Key = NA

1. **opnflg**

OpenPriceSettFlag (For DS (OPENING PRICE)): Is this the opening price. 0=Daily Open Price, 5=Indicative Opening (Int)

1. **status**

SecurityTradingStatus (For SS):

"Identifies the trading status applicable to the instrument or instrument group.

- 2 = Trading Halt

- 4 = Close

- 15 = New Price Indication

- 17 = Ready to trade (Start of Session)

- 18 = Not available for trading

- 20 = Unknown or Invalid

- 21 = Pre-Open

- 24 = Pre-Cross

- 25 = Cross

- 26 = Post Cross

- 103 = No Change"

(Int)

1. **Halt**

HaltReason (For SS):

"Identifies the reason for the status change- 0 = Group schedule

- 1 = Surveillance intervention

- 2 = Market event

- 3 = Instrument activation

- 4 = Instrument expiration

- 5 = Unknown

- 6 = Recovery in Process"

(Int)

1. **Tradeevent**

SecurityTradingEvent (For SS):

"Identifies an additional event or a rule related to the SecurityTradingStatus

- 0 = No Event (default)

- 1 = No Cancel

- 4 = Change of trading session (reset statistics)

- 5 = Implied matching ON

- 6 = Implied matching OFF"

(Int)

1. **Rootsym**

RootSymbol (For SS): Root symbol of the product (String)

1. **Rootasset**

RootOfAssetClass (For SS): Root symbol of the asset. EX: If pulling an option this will have the root of the option (can be blank as shown above) (String)