AQUAQ Analytics

AquaQuarantine:

Level 2 and 3 Data Storage Optimisation

Experts in fast data solutions

for demanding environments

AQUAQ Analytics

Intro

- What is market depth and its various levels
- 3 common methods of storage
- Comparisons between them
- Optimisations that can be implemented

AQUAQ Analytics

Market Depth

- Records orders to buy and sell an instrument
- Orders can consist of different prices and quantities
- The orders shape how the price changes through trades
- The collection of all orders is called the order book

AQUAQ Analytics

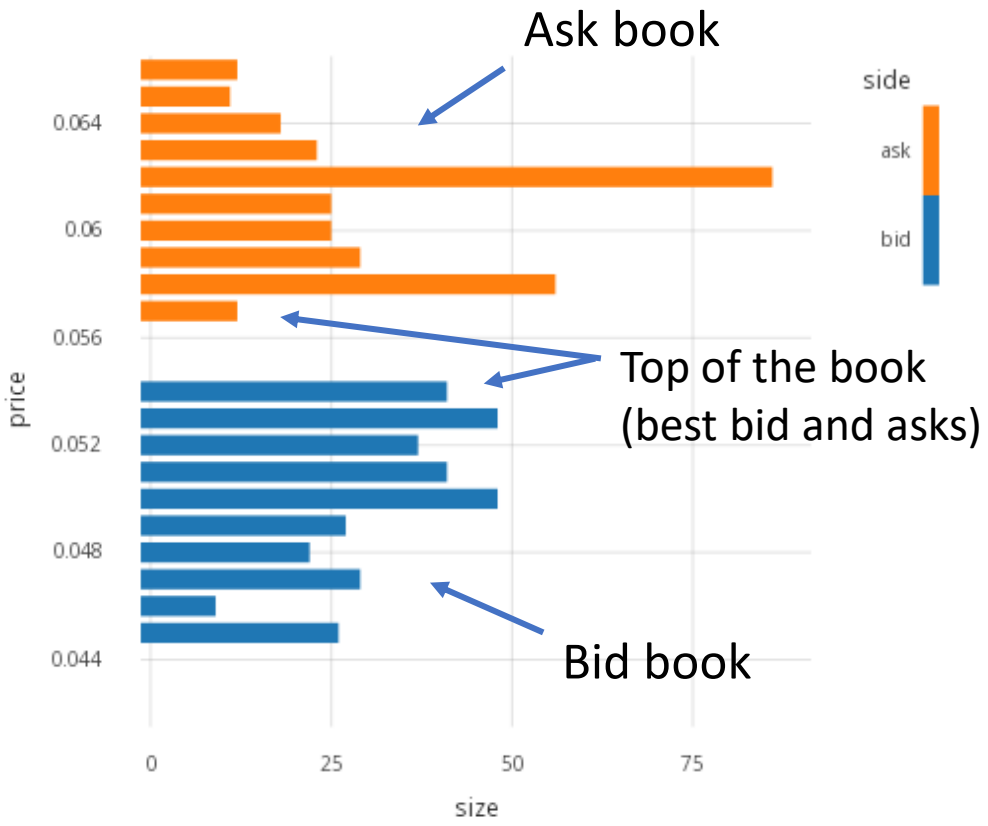
Example: dogecoin

- Cryptocurrency introduced in 2013

Last traded price →

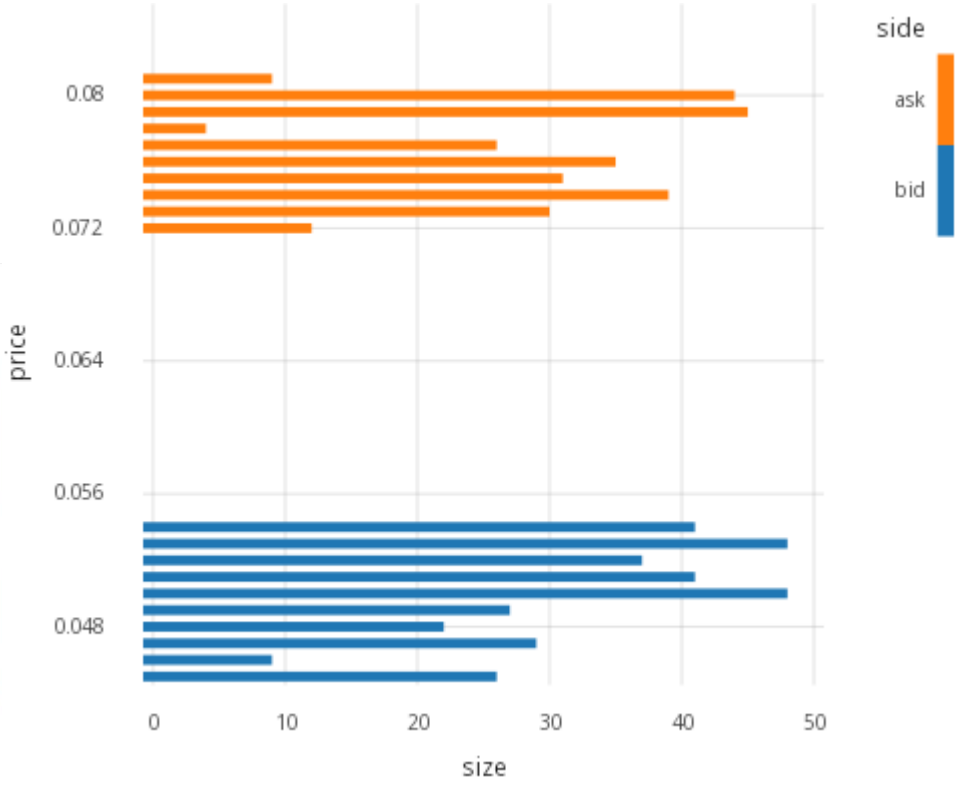
- Price determined by the order book

Dogecoin order book



10:00

10:05



10:15

AQUAQ Analytics

Considerations when storing

- Lots of changing prices and quantities at any point in time
- Number of different prices may continually change
- Potentially unlimited number of prices in bid and ask books
- Different levels of granularity

AQUAQ Analytics

AQUAQ Analytics

Case study

- Used TorQ-CME to process CME MDP FIX 3.0 files
- Looked at 10 different symbols for Soybean Futures
- Maximum depth of 10 price levels
- Compared 3 methods of storage

AQUAQ Analytics

Method 1: rawquote

- Stores orders as they come in throughout the day
- Each row represents an update for a price level (Level 2 data)

```
q)rawquote
time                          sym  action side  price  size  numberoforders
-------------------------------------------------------------------------
2016.03.15D15:59:01.362460418 ZSF7 NEW    OFFER 914.75 5     1
2016.03.15D15:59:01.362460418 ZSF7 NEW    OFFER 912    25    1
2016.03.15D15:59:01.463521766 ZSF7 NEW    BID   911    7     1
2016.03.15D15:59:01.385523763 ZSF7 CHANGE OFFER 912    40    2
2016.03.15D15:59:01.924056282 ZSF7 DELETE BID   911    7     1

q)count rawquote
1179899

q)sum select bookchanges:count distinct time by sym from rawquote
bookchanges| 845673
```

Method 1: rawquote

Pros:

- Easy to store in real-time
- Light on resources

Cons:

- Difficult to see the full order book and how it changes

View rawquote order book

```
1. orders:select by side, price from rawquote where sym=`ZSN6, time<16:00

2. openorders:delete from orders where action=`DELETE

3. select[10;> price] from openorders where side=`BID          /  bid book
   select[10;< price] from openorders where side=`OFFER        /  ask book
```

Bid book

Level 1 →

desc ↓

©¬ side	©¬ price	time	sym	action	size
BID	895.25	2016.03.15D15:59:58.240370947	ZSN6	CHANGE	6
BID	895	2016.03.15D15:59:58.239020998	ZSN6	CHANGE	51
BID	894.75	2016.03.15D15:59:58.239055066	ZSN6	CHANGE	43
BID	894.5	2016.03.15D15:59:50.964555374	ZSN6	CHANGE	87
BID	894.25	2016.03.15D15:59:57.059684072	ZSN6	CHANGE	22
BID	894	2016.03.15D15:59:46.545694250	ZSN6	CHANGE	66
BID	893.75	2016.03.15D15:59:46.545694250	ZSN6	CHANGE	11
BID	893.5	2016.03.15D15:59:05.317977689	ZSN6	CHANGE	13
BID	893.25	2016.03.15D15:59:57.864313077	ZSN6	CHANGE	11
BID	893	2016.03.15D15:59:46.748850370	ZSN6	CHANGE	8

Ask book

→

asc ↓

©¬ side	©¬ price	time	sym	action	size
OFFER	895.5	2016.03.15D15:59:57.862328950	ZSN6	CHANGE	41
OFFER	895.75	2016.03.15D15:59:57.862328950	ZSN6	CHANGE	42
OFFER	896	2016.03.15D15:59:57.860387201	ZSN6	CHANGE	61
OFFER	896.25	2016.03.15D15:59:57.858361572	ZSN6	CHANGE	29
OFFER	896.5	2016.03.15D15:59:50.961888903	ZSN6	CHANGE	24
OFFER	896.75	2016.03.15D15:59:46.499826036	ZSN6	CHANGE	27
OFFER	897	2016.03.15D15:59:47.195614877	ZSN6	CHANGE	36
OFFER	897.25	2016.03.15D15:59:51.677606974	ZSN6	CHANGE	12
OFFER	897.5	2016.03.15D15:59:50.964515195	ZSN6	CHANGE	7
OFFER	897.75	2016.03.15D15:59:39.204285986	ZSN6	NEW	6

AQUAQ Analytics

View order book transitions

1. Scan through quotes to build book snapshots for each sym
2. Group by sym, time

time	sym	bksnap
2016.03.15D00:00:01.457837343	ZSH7	(flip (`side`price)!(`OFFER`OFFER;915 917f))!flip (enlist `size)!enlist 45 5f
2016.03.15D00:00:01.662006984	ZSH7	(flip (`side`price)!(`OFFER`OFFER;915 917f))!flip (enlist `size)!enlist 46 5f
2016.03.15D00:00:01.662754484	ZSH7	(flip (`side`price)!(`BID`OFFER`OFFER;911.25 915 917))!flip (enlist `size)!enlist 1 46 5f

©¬ side	©¬ price	size
BID	911.25	1
OFFER	915	46
OFFER	917	5

- Slow and resource intensive

AQUAQ Analytics

Method 2: widebook

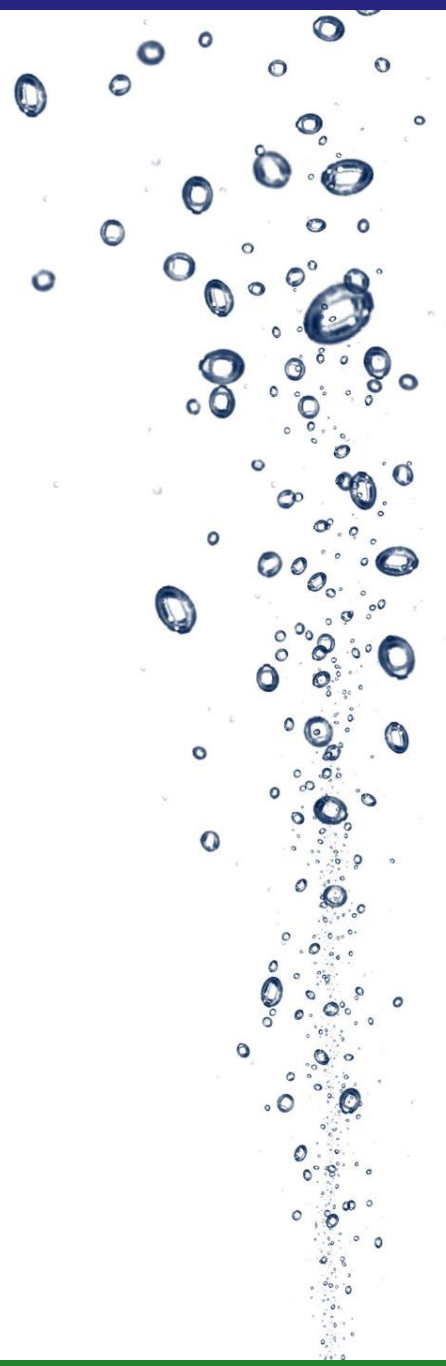
- Store book snapshots using nested lists

time	sym	bprice	bsize	aprice	asize
2016.03.15D13:45:10.832831941	ZSU6	901.25 901 900.75 900.5 900.25 898 896.25 895 894 893.75	7 7 4 3 1 3 3 3 14 1f	901.75 902 902.25 902.5 902.75 903 906 906.75 907.5 908	6 5 4 2 1 3 4 2 1 3f
2016.03.15D13:45:10.833000290	ZSU6	901.25 901 900.75 900.5 900.25 898 896.25 895 894 893.75	7 7 4 3 1 3 3 3 14 1f	901.75 902 902.25 902.5 902.75 903 906 906.75 907.5 908	8 5 4 2 1 3 4 2 1 3f
2016.03.15D13:45:10.833119483	ZSU6	901.25 901 900.75 900.5 900.25 898 896.25 895 894 893.75	6 7 4 3 2 3 3 3 14 1f	901.75 902 902.25 902.5 902.75 903 906 906.75 907.5 908	8 5 4 2 1 3 4 2 1 3f
2016.03.15D13:45:11.999552130	ZSU6	901.25 901 900.75 900.5 900.25 898 896.25 895 894 893.75	6 7 4 3 2 3 3 3 14 1f	901.5 901.75 902 902.25 902.5 902.75 903 906 906.75 907.5	1 8 5 4 2 1 3 4 2 1f
2016.03.15D13:45:11.999631958	ZSU6	901.25 901 900.75 900.5 900.25 898 896.25 895 894 893.75	6 7 4 3 2 3 3 3 14 1f	901.5 901.75 902 902.25 902.5 902.75 903 906 906.75 907.5	1 9 5 4 2 1 3 4 2 1f
2016.03.15D13:45:14.188065655	ZSU6	901.25 901 900.75 900.5 900.25 898 896.25 895 894 893.75	6 7 4 3 2 3 3 3 14 1f	901.75 902 902.25 902.5 902.75 903 906 906.75 907.5 908	9 5 4 2 1 3 4 2 1 3f
2016.03.15D13:45:14.192813529	ZSU6	901.25 901 900.75 900.5 900.25 898 896.25 895 894 893.75	6 7 4 3 2 3 3 3 14 1f	901.75 902 902.25 902.5 902.75 903 906 906.75 907.5 908	8 5 4 2 1 3 4 2 1 3f

Level 1

AQUAQ Analytics

Method 2: widebook



Pros:

- Easy to see book transitions and any number of levels
- Can store unlimited number of price levels

Cons:

- Lots of repeated values
- Heavy on memory and storage
- Difficult to vectorise calculations due to nested lists

AQUAQ Analytics

View top of widebook

```
1. tbk:select time, sym, bestbid:bprice[;0], bestbsize:bsize[;0],
                bestask:aprice[;0], bestasize:asize[;0] from widebook

2. tbk:select from tbk where any differ each (bestbid;bestbsize;bestask;bestasize)
```

time	sym	bestbid	bestbsize	bestask	bestasize
2016.03.15D13:32:48.622111815	ZSH7	910	2	911	2
2016.03.15D13:32:50.711499047	ZSH7	910.25	1	911	2
2016.03.15D13:32:50.714650161	ZSH7	910	2	911	2
2016.03.15D13:32:50.717274882	ZSH7	910.25	2	911	2
2016.03.15D13:32:50.723049225	ZSH7	910.25	3	911	2
2016.03.15D13:32:51.208545587	ZSH7	910.25	3	911	3
2016.03.15D13:32:51.290134851	ZSH7	910.5	1	911	3

AQUAQ Analytics

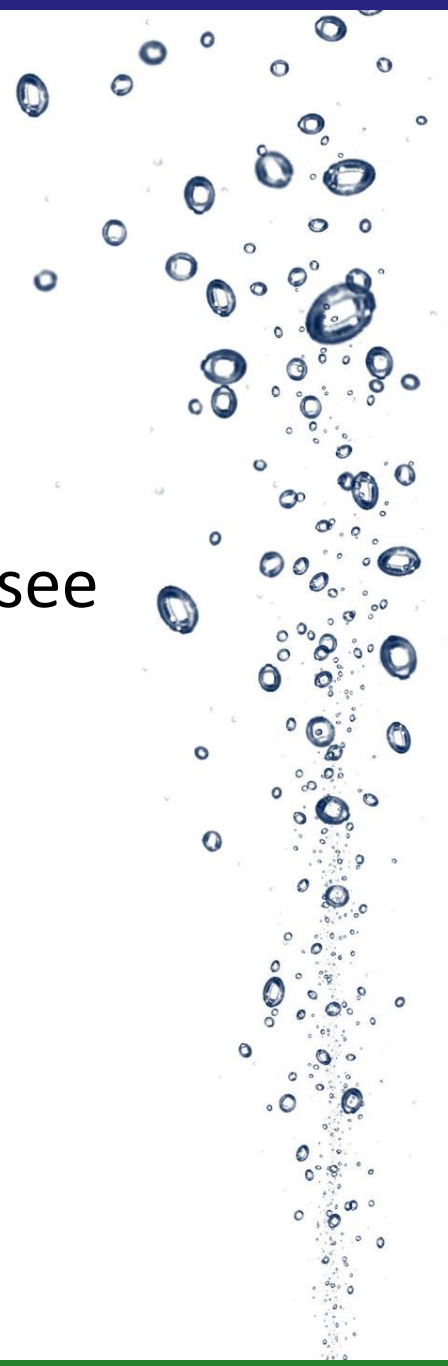
Method 3: fixedbook

- Obtained from widebook
- Price levels are separated into columns for each book snapshot

time	sym	bid1	bid1size	bid2	bid2size	bid3	bid3size	bid4	bid4size	bid5	bid5size	bid6	bid6size	bid7	bid7size	bid8	bid8size	bid9	bid9size	bid10	bid10size	ask1	ask1size	ask2	ask2size
2016.03.15D13:45:00.001...	ZSU6	901.25	6	901	7	900.75	4	900.5	3	900.25	2	898	3	896.25	3	895	3	894	14	893.75	1	901.5	1	901.75	9
2016.03.15D13:45:02.532...	ZSU6	901.25	5	901	7	900.75	5	900.5	3	900.25	2	898	3	896.25	3	895	3	894	14	893.75	1	901.5	1	901.75	9
2016.03.15D13:45:06.808...	ZSU6	901.25	6	901	7	900.75	4	900.5	3	900.25	2	898	3	896.25	3	895	3	894	14	893.75	1	901.5	1	901.75	9
2016.03.15D13:45:06.810...	ZSU6	901.25	7	901	7	900.75	3	900.5	3	900.25	2	898	3	896.25	3	895	3	894	14	893.75	1	901.5	1	901.75	9
2016.03.15D13:45:06.811...	ZSU6	901.25	7	901	7	900.75	3	900.5	3	900.25	2	898	3	896.25	3	895	3	894	14	893.75	1	901.75	9	902	5
2016.03.15D13:45:06.812...	ZSU6	901.25	8	901	7	900.75	3	900.5	3	900.25	1	898	3	896.25	3	895	3	894	14	893.75	1	901.75	9	902	5
2016.03.15D13:45:06.812...	ZSU6	901.25	9	901	7	900.75	3	900.5	3	898	3	896.25	3	895	3	894	14	893.75	1	892.5	1	901.75	9	902	5
2016.03.15D13:45:06.813...	ZSU6	901.25	9	901	7	900.75	3	900.5	3	898	3	896.25	3	895	3	894	14	893.75	1	892.5	1	901.75	8	902	5
2016.03.15D13:45:06.813...	ZSU6	901.25	9	901	7	900.75	3	900.5	3	898	3	896.25	3	895	3	894	14	893.75	1	892.5	1	901.75	7	902	5
2016.03.15D13:45:06.815...	ZSU6	901.25	9	901	7	900.75	3	900.5	3	900.25	1	898	3	896.25	3	895	3	894	14	893.75	1	901.75	7	902	5

AQUAQ Analytics

Method 3: fixedbook



Pros:

- Easy to see book transitions
- More efficient than widebook for viewing levels and analytics (see vwap in blog)

Cons:

- Heavy on storage and memory
- Limits the number of price levels

AQUAQ Analytics

Performance comparisons

- Ran queries to return time series of best bids and asks

table	query time (ms)	memory used (MB)
rawquote	6418	634
widebook	301	60
fixedbook	26	28

AQUAQ Analytics

Memory and space comparisons

table	in-memory (MB)	space on disk (MB)
rawquote	80	55
widebook	430	283
fixedbook	336	272

- Storing book snapshots significantly increases memory and space
- widebook and fixedbook are similar on disk but not in-memory
- All tables are larger in-memory than on disk

AQUAQ Analytics

Widebook vs fixedbook (in-memory)

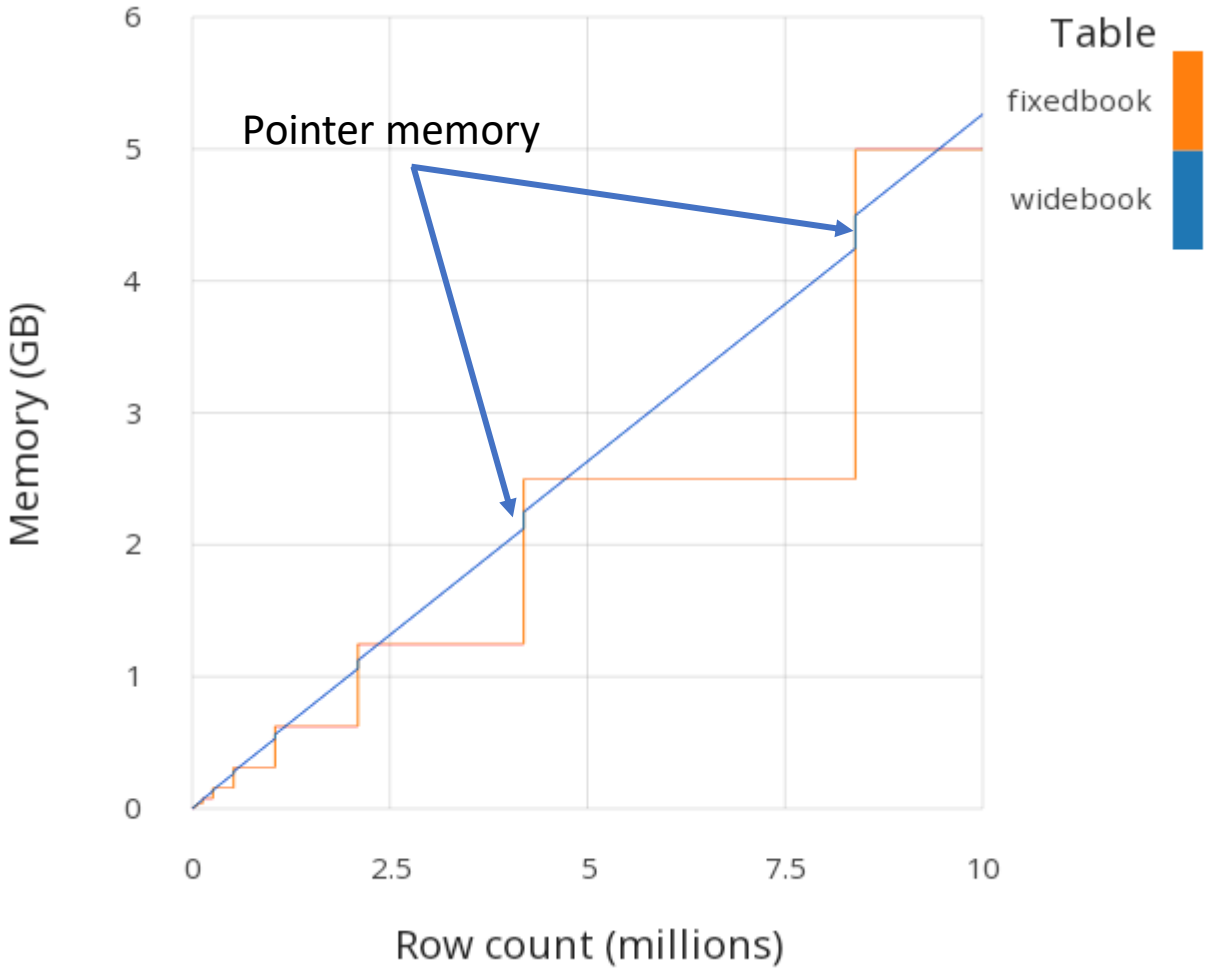
- Kdb+ uses a buddy system when allocating in-memory
- in-memory simple lists sizes are rounded up to nearest power of 2

- wide order book consists of lots of small simple lists
- fixed order book consist of several large simple lists
- widebook is larger 87.5% of the time *

* Assuming all nested lists contain 7 or more elements

AQUAQ Analytics

Widebook vs fixedbook (in-memory)



AQUAQ Analytics

Compression

- .z.zd : 16 2 4 / zip defaults

table	in-memory (MB)	disk space (MB)	compressed (MB)
rawquote	80	59	6.7
widebook	430	283	15
fixedbook	336	272	7.7

- Overall compression ratio of close to 21 : 1

AQUAQ Analytics

Compression

- Compression generally impacts query performance
- Ran queries for top of book time series

Table	Uncompressed time (ms)	Compressed time (ms)
widebook	301	926
fixedbook	26	125

AQUAQ Analytics

Compression strategies

1. Compress less used columns in fixedbook (eg levels 6 to 10)
2. Join uncompressed top of fixedbook to compressed widebook

time	sym	bprice	bsize	aprice	asize	bid1	bid1size	ask1	ask1size
2016.03.15D02:33:56.865253747	ZSF7	907.75 848	1 1f	908.75 911 911.75 912 914.75 915 917	1 1 1 5 1 11 2f	907.75	1	908.75	1
2016.03.15D02:33:56.867676532	ZSF7	907.75 907.5	1 1f	908.75 911 911.75 912 914.75 915 917	1 1 1 5 1 11 2f	907.75	1	908.75	1
2016.03.15D02:33:56.867838219	ZSF7	907.75 907.5	1 1f	908.75 909 911 911.75 912 914.75 915	1 1 1 1 5 1 11f	907.75	1	908.75	1
2016.03.15D02:33:56.870722160	ZSF7	907.5 848	1 1f	908.75 909 911 911.75 912 914.75 915	1 1 1 1 5 1 11f	907.5	1	908.75	1
2016.03.15D02:33:56.871078736	ZSF7	907.5 848	1 1f	909 911 911.75 912 914.75 915 917	1 1 1 5 1 11 2f	907.5	1	909	1
2016.03.15D02:33:56.875990377	ZSF7	848 760.75	1 1f	909 911 911.75 912 914.75 915 917	1 1 1 5 1 11 2f	848	1	909	1

Compressed

Uncompressed

Table	Space on disk (MB)	Top of book time (ms)
widebook	283	301
Compressed widebook	15	926
Hybrid widebook	41	69

AQUAQ Analytics

Less price levels (widebook)

Max depth	On disk %	In-memory %	Row count %
1	8.1	5.1	57.1
2	27.1	29.2	79.3
3	37.8	51.6	88.1
4	47.6	53.6	92.2
5	56.8	54.8	94.6
6	65.8	55.5	96.1
7	74.4	97.2	97.2
8	83.1	98.4	98.3
9	91.9	99.6	99.5

Price level	Changes in widebook %
1	57.1
2	40.0
3	27.7
4	19.8
5	16.3
6	12.0
7	8.7
8	7.8
9	6.9
10	5.1

- On disk decreases linearly
- In-memory decreases in steps
- Significant reductions from 2 to 1 (remove enlistings)
- Lot of activity in first few levels

AQUAQ Analytics

Conclusion

- rawquote the leanest of the 3 but difficult to work with
- Storing book snapshots much easier to work with but resource intensive
- fixedbook the best for analytics but expensive to store and limits levels
- Saving memory/space is a trade off with query time or maximum depth
- Every data set is different

AQUAQ Analytics

- Questions?

Links:

- TorQ-CME: https://github.com/AquaQAnalytics/TorQ-CME
- Blog post: https://www.aquaq.co.uk/data/level-2-storage-formats/

Upcoming talks:

1st April – Interfacing kdb+ with Refinitiv's TickHistory in Google Big Query

15th April – Introduction to Shakti