Using ROOT to analyse High-Frequency Finance Data

20th International Workshop on Advanced Computing and Analysis Techniques in Physics Research

01-12-2021, Philippe Debie, on behalf of Project HighLO

Debie, P.M.M., Naumann, A., Pennings, J.M.E., Tekinerdogan, B., Catal, C., Rembser, J., Verhulst, M.E., Leeuwen, P., Moneta, L., & Alskaif, T.

WAGENINGEN UNIVERSITY & RESEARCH

CORMEC COMMODITY RISK MANAGEMENT EXPERTISE CENTRE

CERN

Project HighLO & Research goal

Collaboration between

1. Wageningen University & Research (WUR)
2. CERN
3. Commodity Risk Management Expertise Centre (CORMEC)

Research goal

1. Describe and detect manipulation of financial markets
2. Help regulators and lawmakers

WAGENINGEN UNIVERSITY & RESEARCH

CORMEC COMMODITY RISK MANAGEMENT EXPERTISE CENTRE

CERN

Background info

Limit orders (LO)

- Bid LO = buy for maximum price
- Ask LO = sell for minimum price

Limit order book (LOB)

- Summary of all demand and supply
- Price of any ask LO is greater than price of any bid LO

	Level	Price	Volume
Ask (sell)	5	$55.00	2
	4	$54.60	10
	3	$53.50	2
	2	$53.00	4
	1	$50.00	9
		Spread	
Bid (buy)	1	$49.00	12
	2	$48.00	4
	3	$47.50	38
	4	$46.00	3
	5	$40.00	1

WAGENINGEN UNIVERSITY & RESEARCH

CORMEC COMMODITY RISK MANAGEMENT EXPERTISE CENTRE

CERN

Background info

Data

- Commodity futures from the Chicago Mercantile Exchange (CME)
- 300TB of messages
    - For each order, for each transaction, etc.
    - Nanosecond timestamp
    - Irregularly spaced

WAGENINGEN UNIVERSITY & RESEARCH

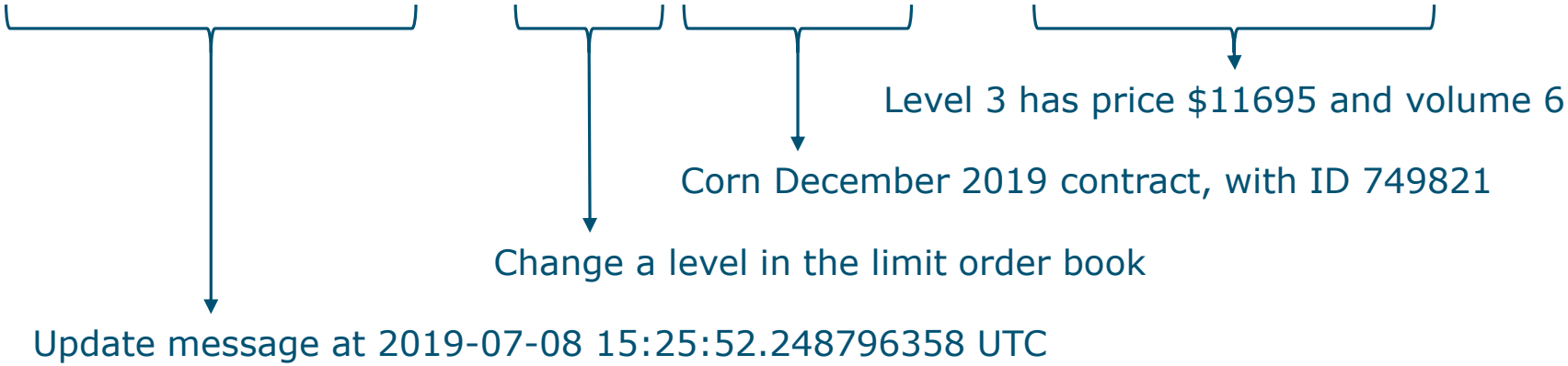
CORMEC COMMODITY RISK MANAGEMENT EXPERTISE CENTRE

CERN

Background info

Message example (FIX protocol)

```
35=X;52=20190708152552248796358;268=1;279=1;269=1;48=749821;55=ZCZ9;83=1847;270=11695.0;271=6;346=2;1023=3;10=140
```



→ 15TB in ROOT

WAGENINGEN UNIVERSITY & RESEARCH

CORMEC COMMODITY RISK MANAGEMENT EXPERTISE CENTRE

CERN

Why ROOT

Finance research

- Data is noisy, irregular in shape, and large in size
- Current storage tools are basic (e.g., csv files)

The power of ROOT

- TFile and TTree are perfect for LOB data
- Transform timeseries into events → Apply HEP statistical methods

WAGENINGEN UNIVERSITY & RESEARCH

CORMEC COMMODITY RISK MANAGEMENT EXPERTISE CENTRE

CERN

Data iteration using a TimeFrame

Overview

- Counterpart of RDataframe for timeseries
- Templated (LimitOrderBook, TimeNS, Message)
- Keeps track of a state (LimitOrderBook)
- Based on lambda functions

Data iteration using a TimeFrame

Create a TimeFrame object

```
TimeFrame timeFrame;

timeFrame.add(chainSoybean);
timeFrame.add(chainCorn);
```

Data iteration using a TimeFrame

Keep track of the internal state

```
timeFrame.setStateInitializer([&](int id)
{
    return LimitOrderBook(metaData.at(id).Name, id);
});

timeFrame.setStateUpdater([](int id, TimeNS time, LimitOrderBook& lob, const Message& message)
{
    lob.update(time, message);
});
```

WAGENINGEN UNIVERSITY & RESEARCH

CORMEC COMMODITY RISK MANAGEMENT EXPERTISE CENTRE

CERN

Data iteration using a TimeFrame

Simple iteration

```
timeFrame.setForEachRow([&](int id, TimeNS time, const Message& message, const LimitOrderBook& lob)
{
    std::cout << lob.getName() << " has " << lob.getTradeVolume() << " transactions so far\n";
});
```

Making snapshots

```
timeFrame.setForEachSnapshot(T_Second * 10, [](TimeNS time, const map<int, LimitOrderBook>& lobs)
{
    std::cout << lobs.size() << " internal states tracked at " << nsToTimestamp(time) << '\n';
});
```

WAGENINGEN UNIVERSITY & RESEARCH

CORMEC COMMODITY RISK MANAGEMENT EXPERTISE CENTRE

CERN

Data iteration using a TimeFrame

Start iteration

```
timeFrame.run();
```

What happens?

1. Synchronize the 2 chains
2. Build the state for each message
3. Call the lambda functions

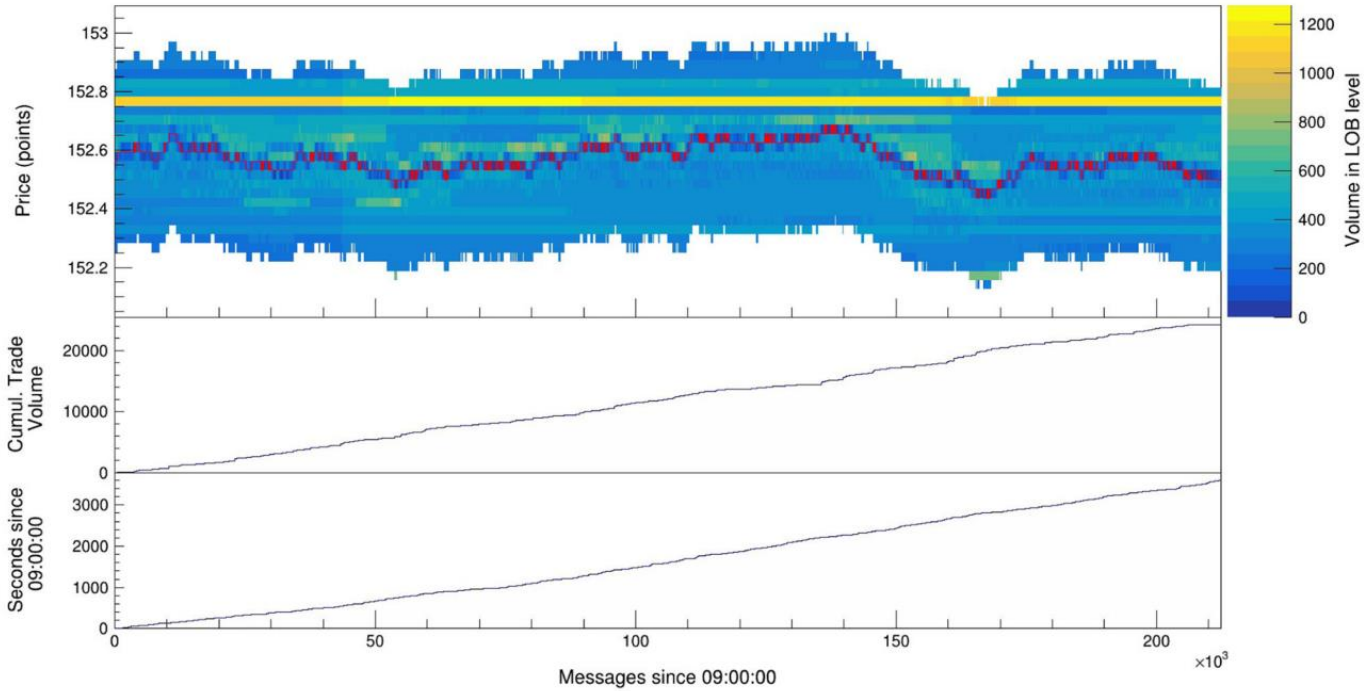
→ Constant memory usage

WAGENINGEN UNIVERSITY & RESEARCH

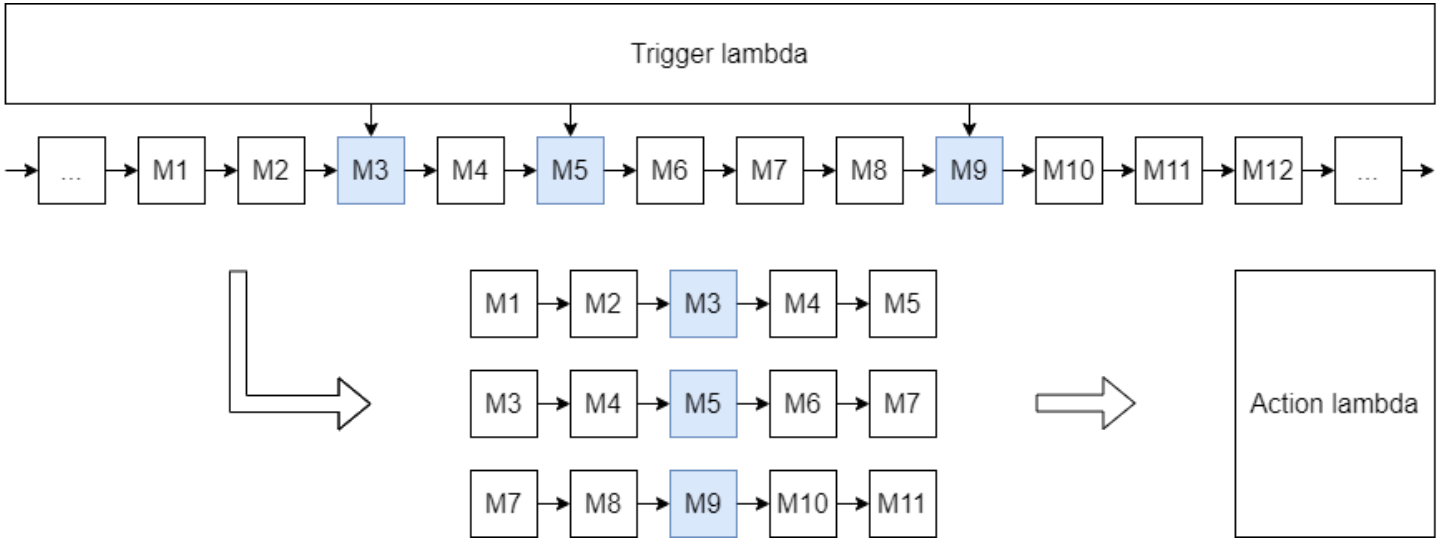
CORMEC COMMODITY RISK MANAGEMENT EXPERTISE CENTRE

CERN

Results – Visualization of limit order book



Trigger-filter-action system



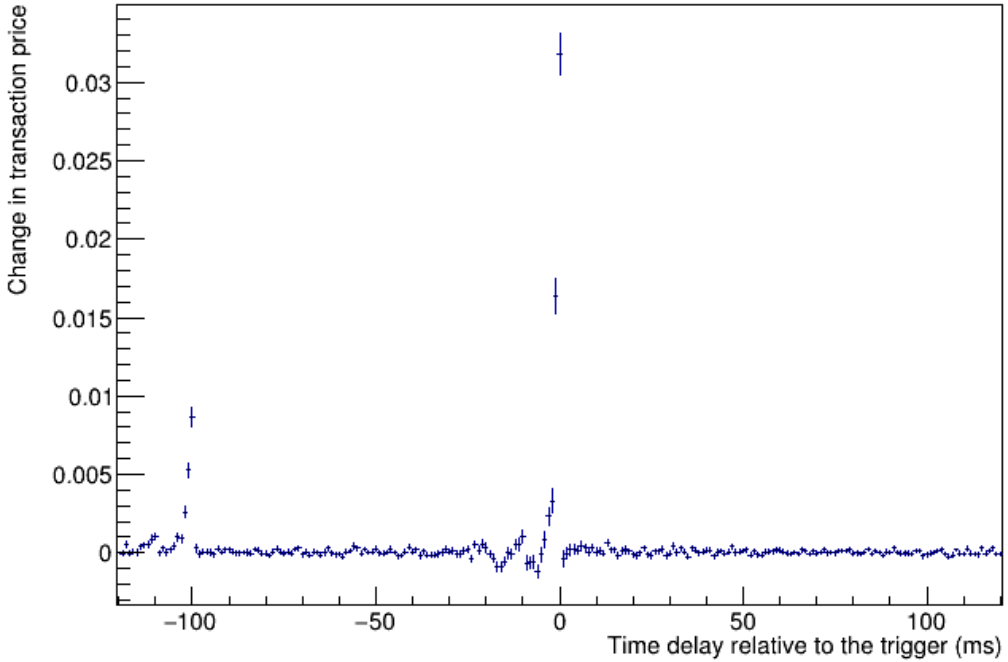
Trigger-filter-action system

```
timeFrame.setTrigger([&](int id, TimeNS time, const Message& message)
{
    return isLOBUpdate(message);
});

timeFrame.setFilter([&](int id, TimeNS time, const Message& message)
{
    return true;
});

timeFrame.setAction(-T_Minute, T_Minute, [&](int id, TimeNS time, const Message& triggerMessage,
                                             const std::list<std::pair<TimeNS, Message>>& msgs)
{
    std::cout << msgs.size() << " msgs in the 2 minutes before and after the trigger at " << nsToTimestamp(time) << '\n';
});
```

Results – Impact plot



WAGENINGEN UNIVERSITY & RESEARCH

CORMEC COMMODITY RISK MANAGEMENT EXPERTISE CENTRE

CERN

Conclusion

TimeFrame

- Use ROOT with timeseries
- Introduce HEP tools into Finance

References

- Verhulst, Marjolein E., Philippe Debie, Stephan Hageboeck, Joost ME Pennings, Cornelis Gardebroek, Axel Naumann, Paul van Leeuwen, Andres A. Trujillo-Barrera, and Lorenzo Moneta. "When two worlds collide: Using particle physics tools to visualize the limit order book." Journal of Futures Markets 41, no. 11 (2021): 1715-1734.

- https://github.com/HighLO/TimeFrame

Questions