

# Monitoring Electronic Trading Environments using Spark

Fergal Toomey and Pierre Lacave  
Corvil



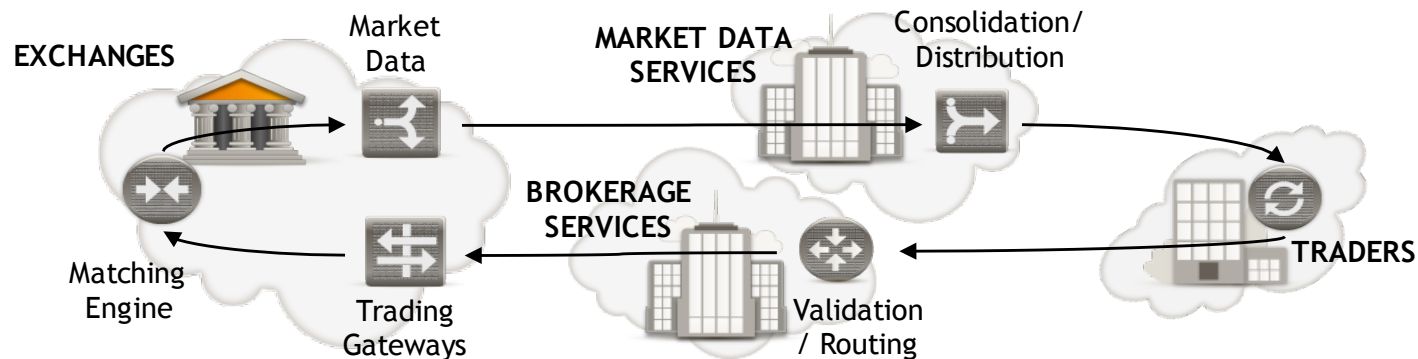
**SPARK SUMMIT EAST**  
DATA SCIENCE AND ENGINEERING AT SCALE  
FEBRUARY 16-18, 2016 NEW YORK CITY

# Monitoring Goals

- Performance and technical risk
- Client experience
- Compliance



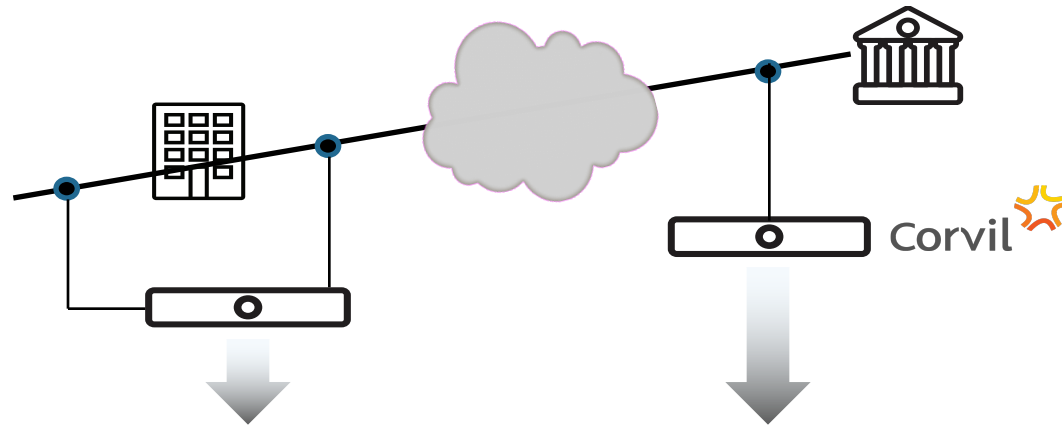
# Challenges



- Distributed across sites and firms
- Latency sensitive
- Multiple protocols and data formats

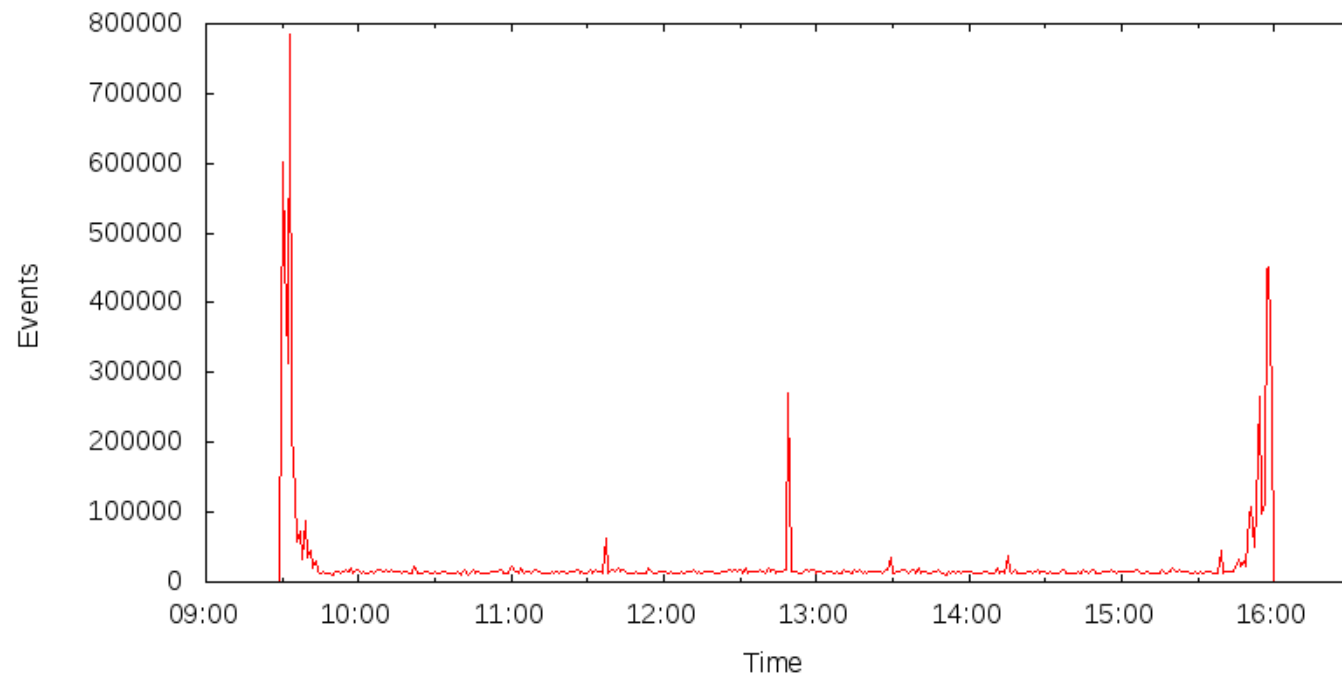
- Transformations:
    - Change of custody
    - Different symbologies
    - Parent/child orders
    - Variable routing
-

# Data Sources

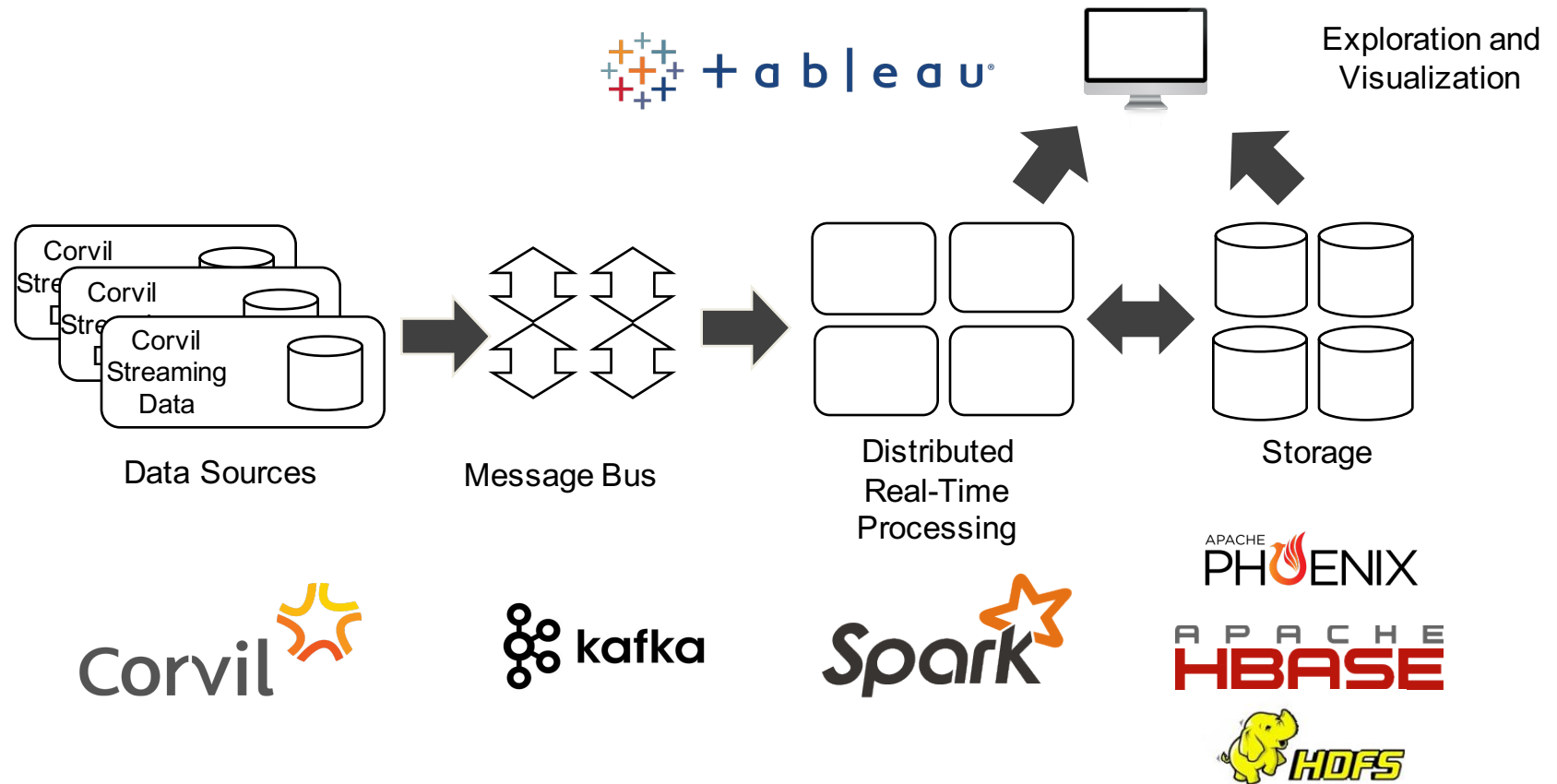


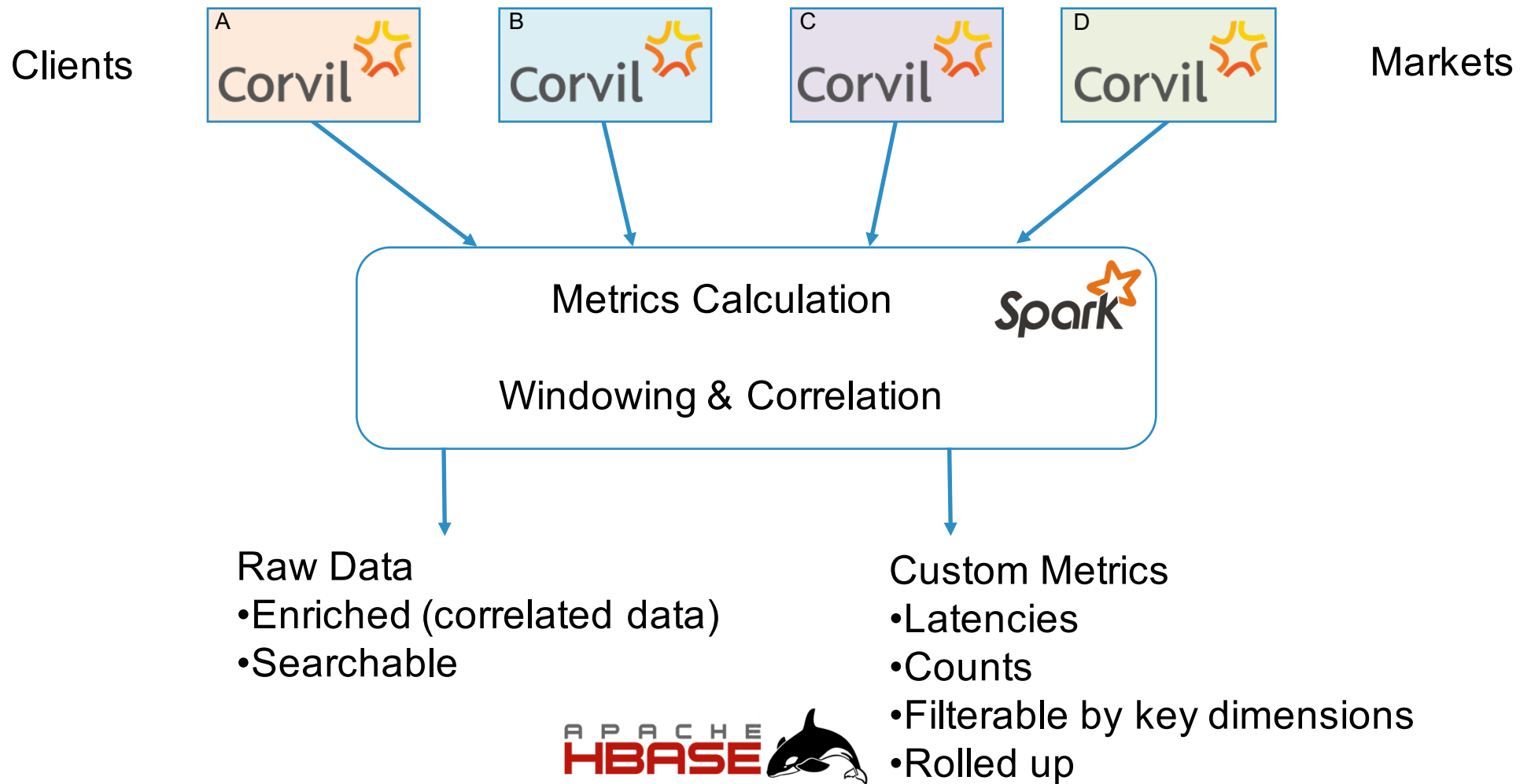
Timestamp	Event Name	protocol	symbol	firm_id	quantity	pri_token	alt_token
<u>2016/02/01 16:24:03.513,712,446</u>	Filled	Bus.parent	PRSNW		120	99376	49789
<u>2016/02/01 16:24:03.513,845,446</u>	Order	Bus.child	ARCI		5	411825	99388
<u>2016/02/01 16:24:03.514,222,446</u>	Filled	FIX	SALE	Silver Lake Group	60	SLVL-6.97	49787
<u>2016/02/01 16:24:03.514,524,446</u>	Accepted	Bus.child	PSCE			411808	49790
<u>2016/02/01 16:24:03.514,917,446</u>	Filled	FIX	SALE	Silver Lake Group	60	SLVL-6.97	49787
<u>2016/02/01 16:24:03.515,407,446</u>	Filled	Bus.parent	ASTE		120	99380	49790
<u>2016/02/01 16:24:03.515,678,446</u>	Accepted	FIX	ARCI	WFG Advisors		WFGA-35.89	

# Data Volumes

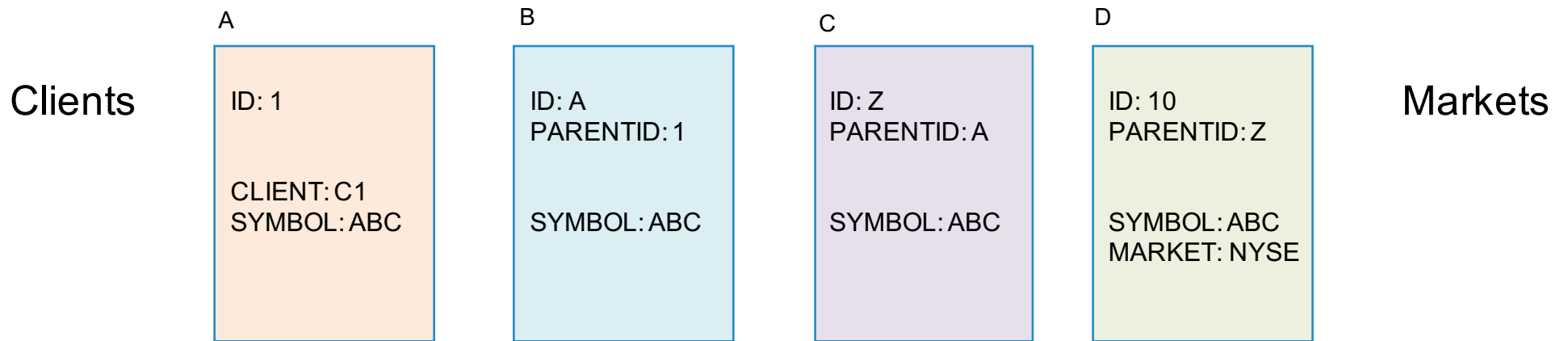


# Platform





# Events Correlation

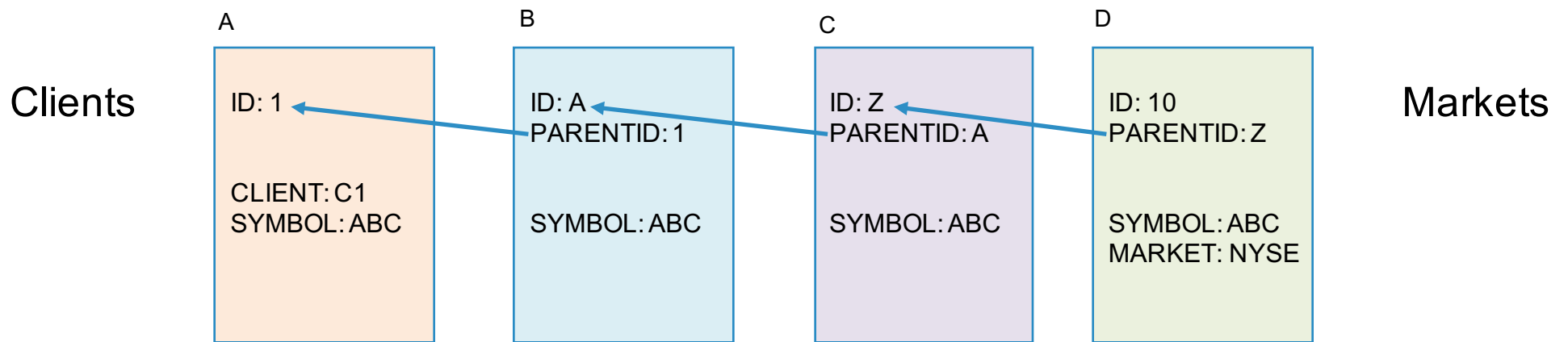


Filter by destination market in point A?

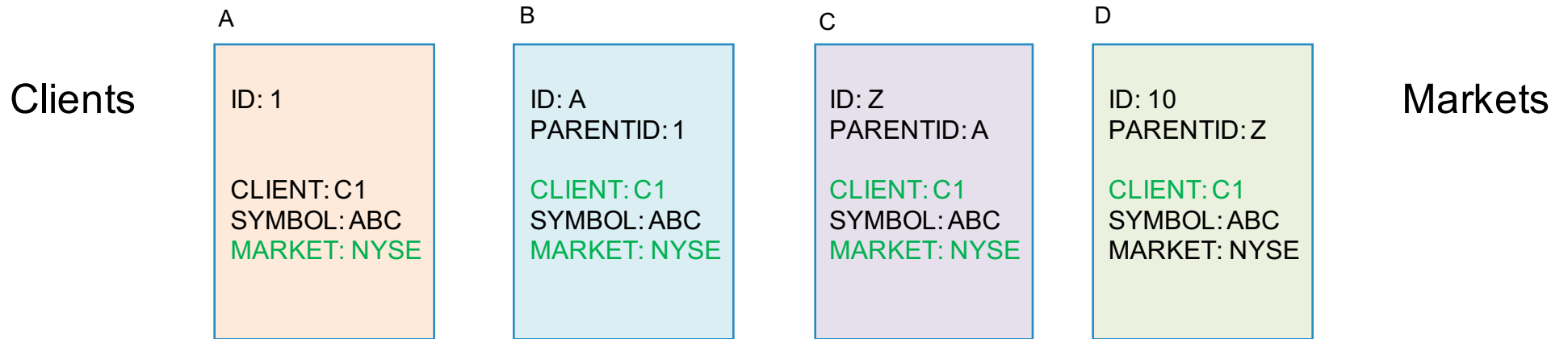
---



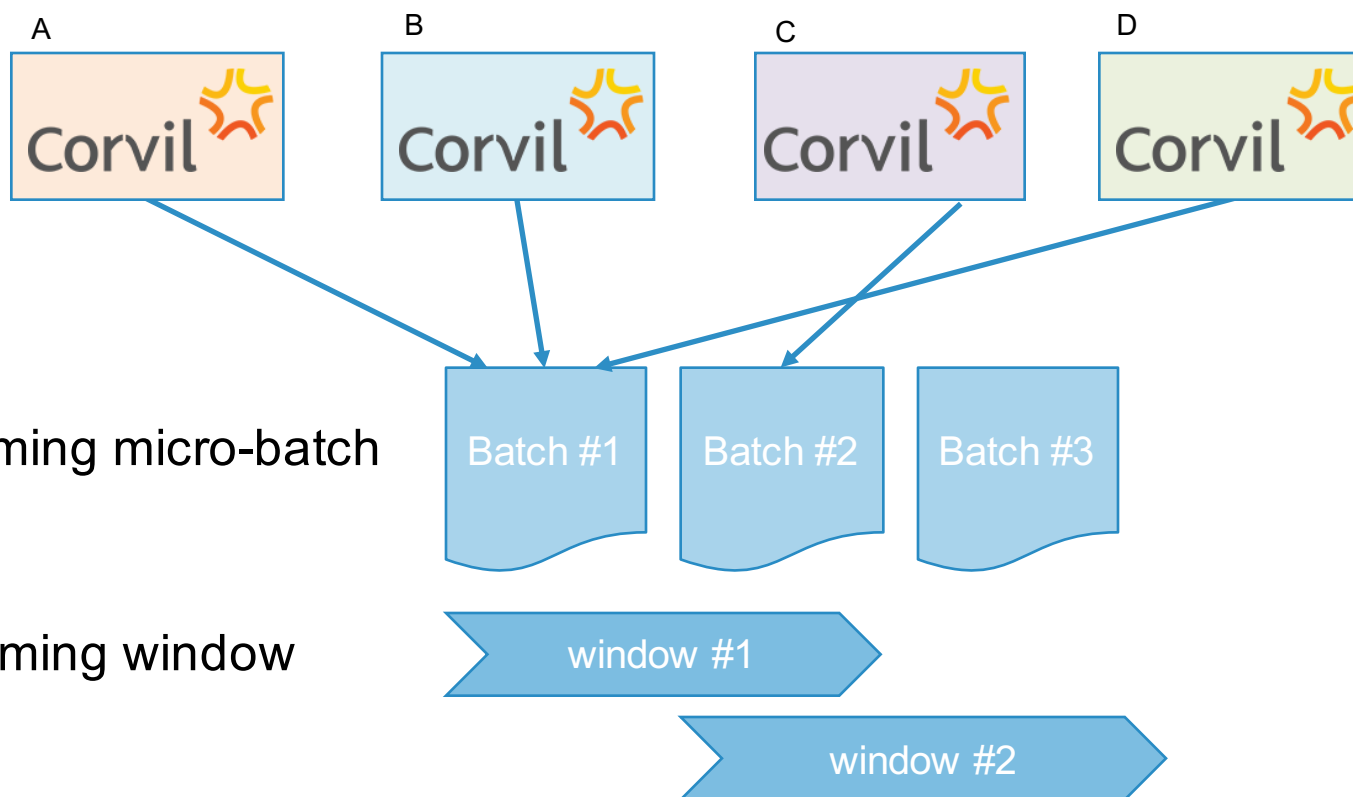
# Events Correlation



# Events Correlation



# Events Windowing



Spark Streaming micro-batch

Spark Streaming window

Reconstruct batch based on original time

# Direct Data Access



Access metrics via SQL / JDBC

BI tools, self-developed application, etc

Calculation done in HBASE

---

Order Response Latency  
for all Client Orders

DATE &amp; TIME:

PERIOD:

1 Hour

VIEW:

Current

Historic



## Top 5 Clients by MAX Latency (μs)

	ORDERS	99%	MAX	AVG
Arcadia Investment	220,062	3,109	5,493	724
Birchview Capital	149,397	1,557	3,049	65
Iron Financial	13,121	1,375	2,681	69
Aozora Bank	6,456	826	1,586	66
Banyan Partners	6,545	326	588	64

Order Response Latency  
for all Venue Orders

Time: 2/12/2016 12:30:00

	ORDERS	99%	MAX	AVG
MAHWAH				
NYSE	326,851	2,270	4,308	486
NYSE Arca	326,586	2,197	4,509	352
CARTERET				
Nasdaq	293,930	2,185	4,091	395
Nasdaq-BX	305,164	2,203	4,108	519
Nasdaq-PSX	315,990	2,201	4,570	359
SECAUCUS				
BATS-Y	324,610	2,391	4,345	751
BATS-Z	326,883	2,439	4,142	892
BATS-EDGE-A	327,331	2,387	4,097	914
BATS-EDGE-X	993	2,120	9,985	769
Add Location				

# Data Extracts

## CSV extraction to HDFS

```
corvil@dc-hadoop-master2:~/extract-0.2-SPARK$ ./bin/extract dc-hadoop-master2:2181 \
> /user/corvil/export/fills-count_symbol-1454047717863797622-1454076517863797622.csv \
> RAW_EVENT 1454047717863797622 1454076517863797622 \
> "SELECT count(SYMBOL) COUNT, SYMBOL
> FROM RAW_EVENT
> WHERE
>   CHANNEL_ID = 'Traders'
>   AND EVENT_NAME = 'Filled'
>   AND SYMBOL LIKE 'M%';
> GROUP BY SYMBOL"
```

- Filtering pushed down to HBASE
- Aggregation done In SPARK (SparkSQL)

```
1 SELECT SYMBOL, COUNT
2 FROM parse_all(from_hdfs(E'corvil', E'dc-hadoop-master1',
3 E'/user/corvil/export/fills-
count_symbol-1454047717863797622-1454076517863797622.csv'))
ORDER BY COUNT DESC LIMIT 10;
```

Query History		
symbol	count	
MLNX	9649	
MAGS	9598	
MYOS	9451	
MRVC	9448	
MNGA	9351	
MCBK	9341	
MMLP	9328	
MXWL	9308	
MDVXW	9299	
MSTR	9279	

# Client Dashboard



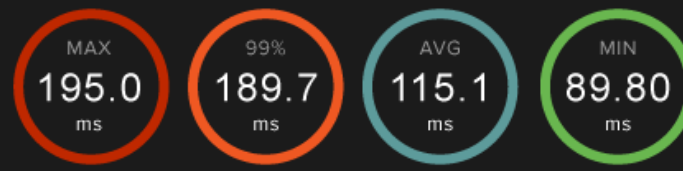
DATE & TIME: 27/10/15 00:00:00 SELECT PERIOD: 1 Month

## Order to Fill Rate



VENUES	
NYSE	93%
NYSE Arca	87%
Nasdaq	90%
Nasdaq-BB	94%

## Latency



VENUES	ORDERS	CANCELS	AVG
NYSE	112	8	92.6ms
Nasdaq	101	10	189.7ms
BATS-Y	94	11	177.1ms
BATS-EDGE-A	86	12	176.3ms

## TOP TRADES Bought



SYM	VOL	\$
GE	126k	79M
WTW	124k	75M
GE	102k	67M
INTC	116k	68M

## TOP TRADES Sold

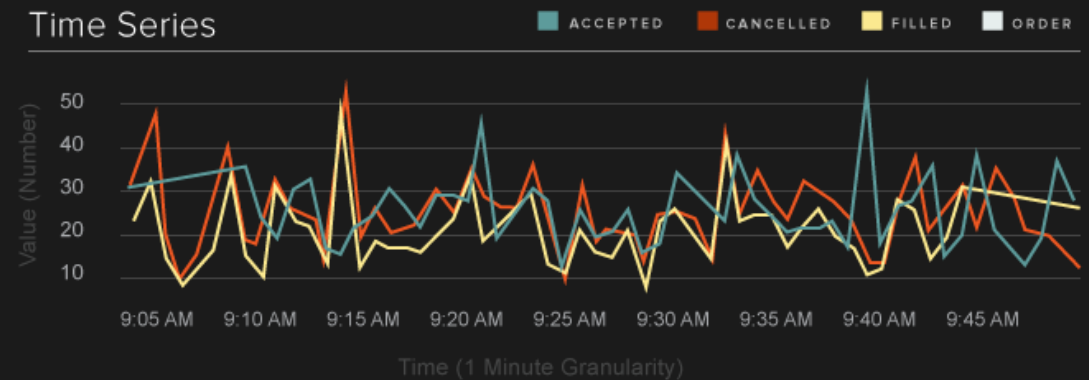


SYM	VOL	\$
AA	176k	132M
AAPL	176k	69M
AA	180k	67M
WTW	171k	64M

## Trader Activity

TOP POSITIONS	SYM	VOL	\$
Bob Spread	GE	19k	8M
Sarah Smith	WTW	18k	7M
Joe Bloggs	WMCS	17k	6.7M
David Clarke	INTC	17k	7.1M
Jane Murphy	AAPL	16k	5.2M

## Time Series



# THANK YOU.

[www.corvil.com](http://www.corvil.com)



**SPARK SUMMIT EAST**  
DATA SCIENCE AND ENGINEERING AT SCALE  
FEBRUARY 16-18, 2016 NEW YORK CITY