Load Testing Report of the Moscow Exchange Trading and Clearing Systems October 26, 2024

Contents

Testing objective

The objective of the load testing was to verify the performance of the trading and clearing systems under peak load and increased number of orders and trades. The trading systems of the following markets were tested:

a) the Securities market,

b) the FX market,

c) the Derivatives market.

Main results

The Securities market Trading/clearing system

Testing was performed on the production version and configuration of the trading/clearing system with independent trading and clearing engines.

The achieved data capacity and performance parameters, compared to the maximum number of orders and trades in production environment and load testing from 2023, are provided in the table below.

Definition of the 'accepted transactions': all incoming orders resulting in either registration of a new order or cancelation of a previously registered order.

	Transactions	Orders	Trades
Achieved values, units, 2024	193,335,233	143,369,861	13,258,761
Achieved values, units, 2023	165,131,183	111,259,021	13,072,341
Maximum values in production environment	170,000,000	101,500,000	12,000,000
Maximum processing rate for accepted transactions, units per second, 2024	62,215	41,803	4,223
Maximum processing rate for accepted transactions, units per second, 2023	59,978	40,609	9,180

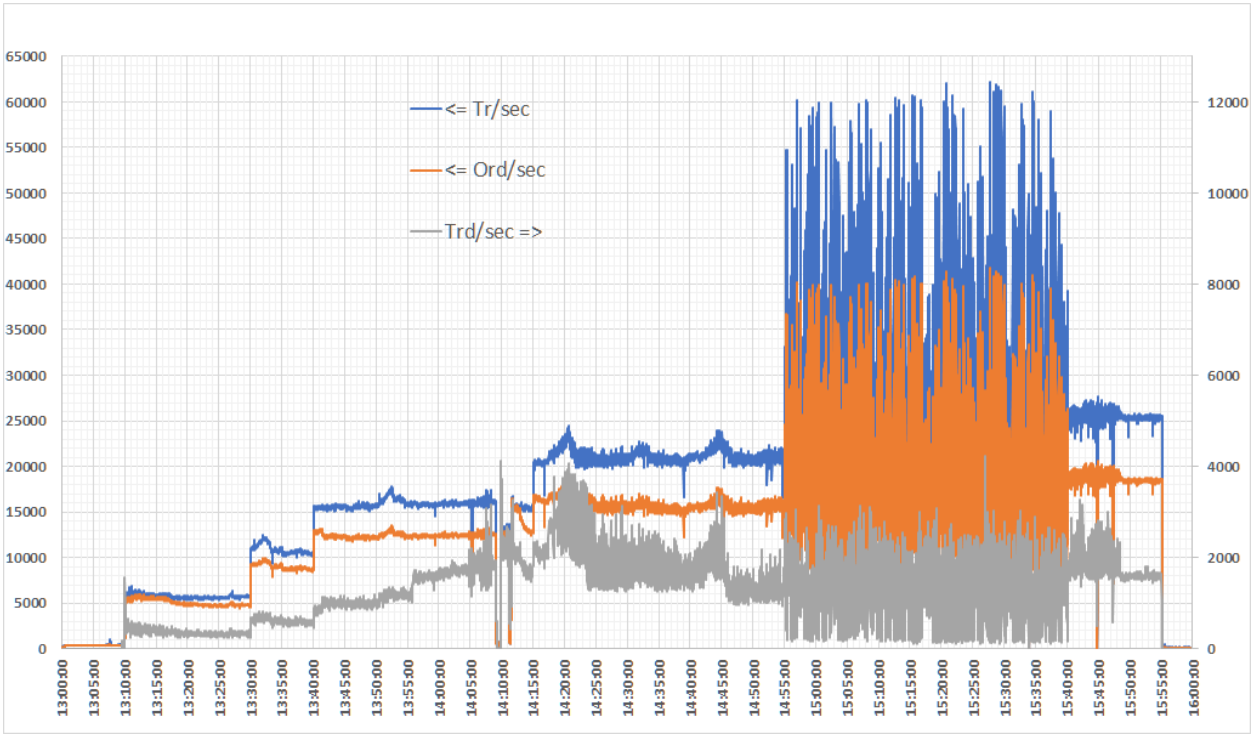
Peak numbers of orders and trades were not reached during testing. The peak frequency of trades was not reached during testing.

Peak frequencies of transactions and orders matched or exceeded 2023 values.

The average transaction latency for the fastest trading protocol FIFO TWIME ASTS was 220 microseconds, 27% lower than other protocols.

The Securities market access servers, index servers, FIX, FAST, SIMBA, and FIFO TWIME ASTS services operated without issues.

The graph below shows frequencies of transactions, orders, and trades.



Clients generated 4.7% of the transactions, nearly double the 2023 level.

The FX market Trading/clearing system

Testing was performed on the production version and configuration of the trading/clearing system with independent trading engine and several independent clearing engines.

The achieved data capacity and performance parameters, compared to the maximum number of orders and trades in production environment and load testing from 2022, are provided in the table below.

Definition of the 'accepted transactions': all incoming orders resulting in either registration of a new order or cancelation of a previously registered order.

	Transactions	Orders	Trades
Achieved values, units, 2024	169,827,222	124,728,799	2,944,171
Achieved values, units, 2023	170,000,062	112,544,472	3,527,920
Maximum values in production environment	120,000,000	85,000,000	1,500,000
Maximum processing rate for accepted transactions, units per second, 2024	59,628	44,537	4,468
Maximum processing rate for accepted transactions, units per second, 2023	60,421	40,027	3,041

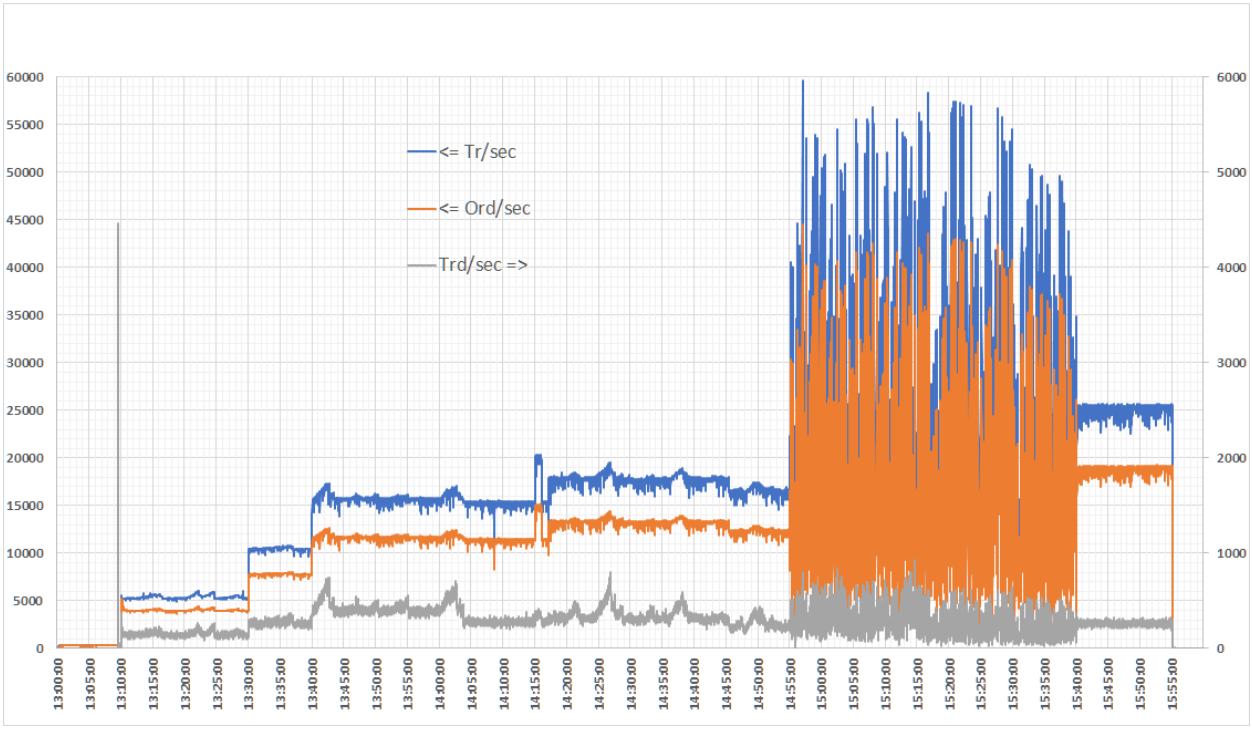
Peak numbers of orders and trades were not reached during testing. The peak frequency of trades was not reached during testing.

Peak frequencies of transactions and orders were consistent with 2023 levels.

The FX market access servers, FIX, FAST, SIMBA, and FIFO TWIME ASTS services operated without issues.

The average transaction latency for the fastest trading protocol FIFO TWIME ASTS was 220 microseconds, 27% lower than other protocols.

The graph below shows frequencies of transactions, orders, and trades.



Clients generated 1.8% of the transactions, 40% lower than in 2023.

The Derivatives market Trading/clearing system

Testing was performed on a prototype version of the system with hot backup enabled for the trading engine modules:

- message dispatcher,
- pre-trade risk and collateral control modules,
- trading system matching module.

Other system services and modules corresponded to SPECTRA 7.27 used in production since September 21, 2024 on the servers of the Data Space and M1 data centers.

The focus of the testing was to evaluate the impact of hot backup on latency and network bandwidth, as well as to measure the failover time of engine modules during simulated hardware failures in production infrastructure.

Failover testing results:

During simulated message dispatcher failure, failover occurred within 4 seconds.

During simulated pre-trade risk and collateral control server failure, failover occurred within 2 seconds.

During simulated matching server failure, the trading engine modules reported data desynchronization between the main and backup matching modules. The Kill Switch mechanism was triggered, halting the trading engine.

An automatic system shutdown was performed at 13:41:11.

Manual synchronization of data files on the matching servers and subsequent trading engine restart were completed by 14:20:18.

The test demonstrated that the hot backup prototype requires further improvements.

Impact of hot backup on trading/clearing system transaction latency:

Transaction latency increased from 70-75 to 120-130 microseconds. Refer to the next section for more details.

Impact of hot backup on trading/clearing system network bandwidth:

The order-to-trade ratio in the testing was near the production value. During the testing, 133 million transactions were sent, and 3,6 million trades were executed. **The peak transaction processing rate during the testing reached 72,000 transactions per second (120,000 transactions per second in 2023).**

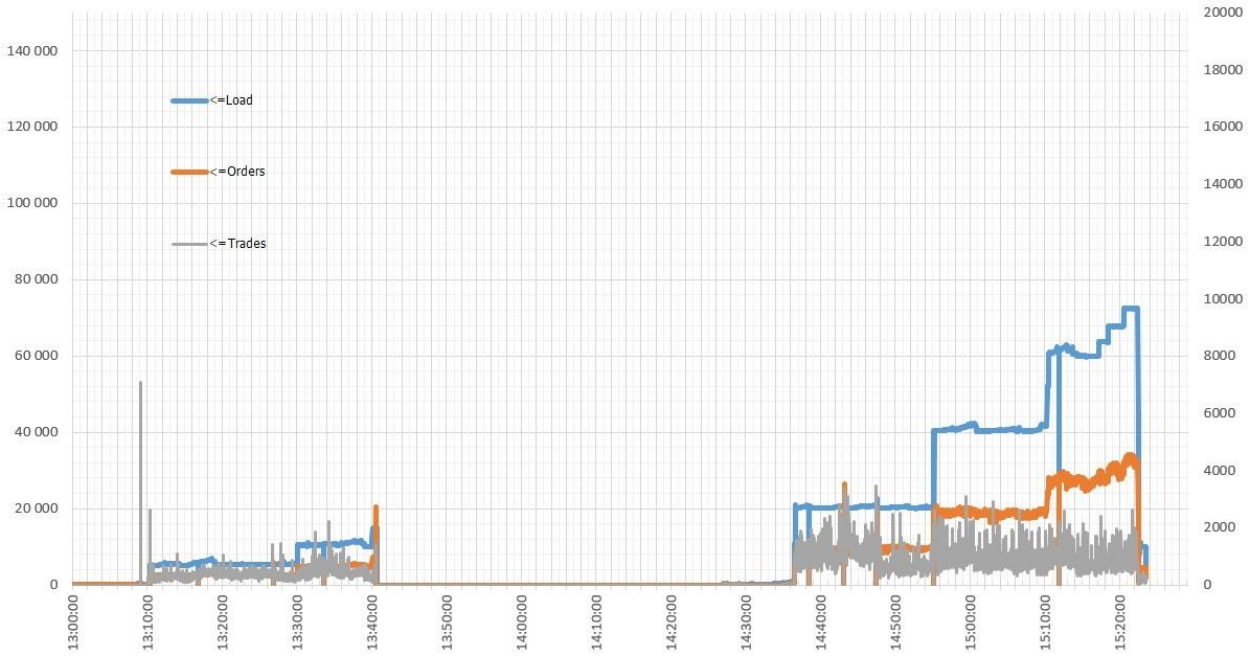
The significant reduction in the total number of transactions submitted during the testing compared to 2023 was due to the system shutdown during hot backup testing and the resulting 40-minute loss of testing time to resolve the issue.

	Transactions	Orders	Trades
Achieved values, units, 2024	133,949,925	65,608,305	3,643,536
Achieved values, units, 2023	282,801,252	127,872,026	9,560,134
The peak performance, transactions per second, 2024	72,000	—	—
The peak performance, transactions per second, 2023	120,000	—	—
Performance in production environment, transactions per second, 2024	25,000	—	—

Clients generated 2.11% of the transactions, 0.75% lower than in 2023.

No members with colocation connection or with third-party software using TWIME / SIMBA / FAST protocols took part in the testing.

The graph below shows transaction load on the Derivatives market trading/clearing system.

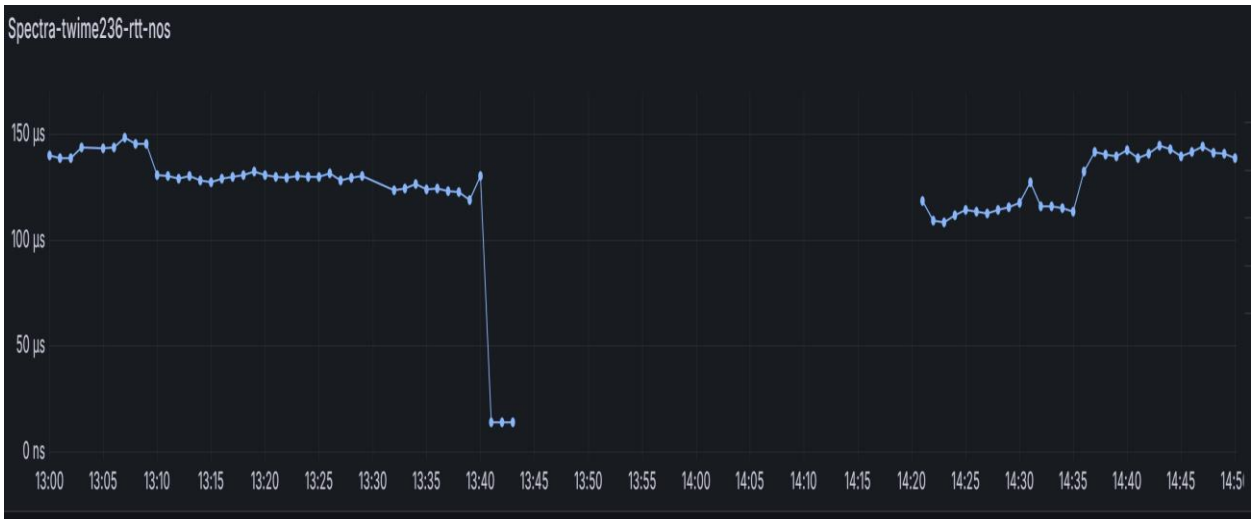


During the load testing, an intraday clearing session was conducted. It was successfully completed within 1 minute and 38 seconds.

At the end of the testing, an evening clearing session was conducted. The evening clearing session operations were completed in 15 minutes, which aligns with expectations based on data volumes.

Transaction latency. Spectra trading/clearing system, SIMBA servers, and FAST UDP multicast marketdata of the Derivatives Market

To measure the Derivatives market trading/clearing system latency, the Corvil system was used alongside a traffic analysis and visualization setup in Grafana, and logging of transaction flows from sell-side trading robots.



Within the transaction frequency range from 5,000 to 20,000 per second, the median RTT for TWIME gateway was between 120 and 130 microseconds. The peak observed on the graph between 13:00 and 13:10 was caused by the opening auction on the Derivatives market. The absence of data between 13:40 and 14:20 was due to the lack of transaction flows on the TWIME gateway (trading engine failure).

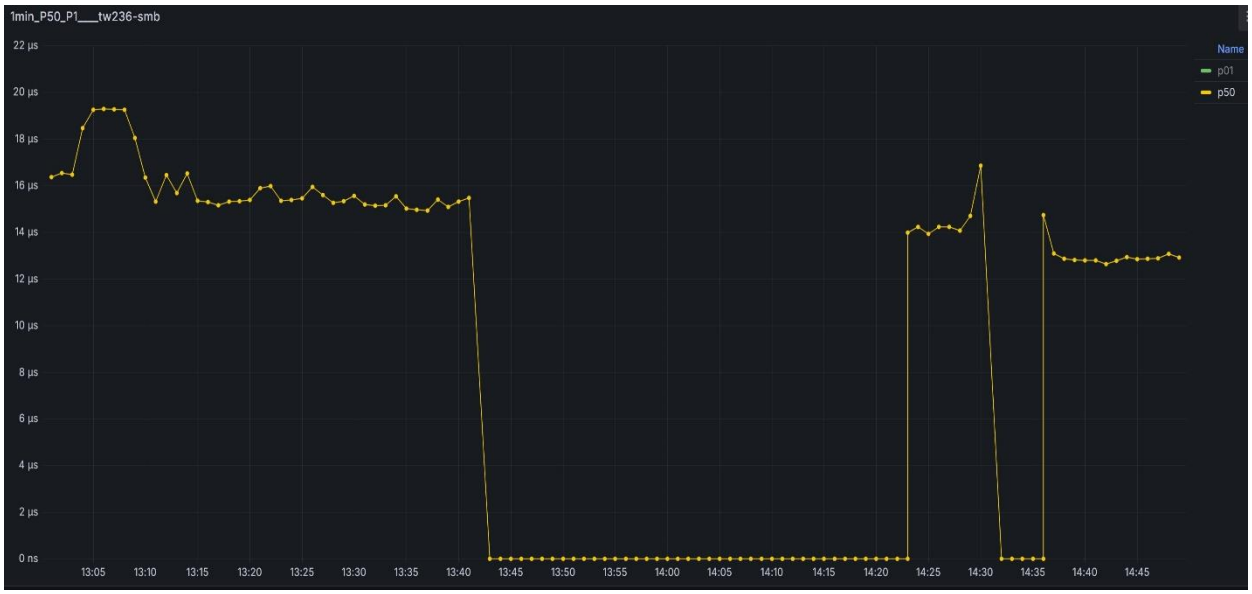
SIMBA and FAST servers' configuration for the Derivatives market was the same as in production. The servers have operated correctly across the entire load levels during the testing.

Traffic capture was performed using the Corvil equipment. Statistical data on TWIME protocol order accepted messages (NewOrderSingleResponse) relative to new order messages from the trading/clearing system (Order log feed) in SIMBA protocol was collected using the traffic analysis and visualization setup in Grafana. The graphs below show the median difference between the TWIME and SIMBA timestamps. The positive value indicates that, on average, SIMBA outpaced TWIME by 15 microseconds.

Statistical data of this type is continuously collected during normal trading. The median publication times for the Derivatives market are shown on the graphs below. Peaks on the graph was caused by the opening auction and intraday clearing session on the Derivatives market.

The absence of data between 13:40 and 14:20 was due to the lack of transaction flows on the TWIME gateway. The absence of data between 14:32 and 14:36 was due to the intraday clearing session on the Derivatives market.

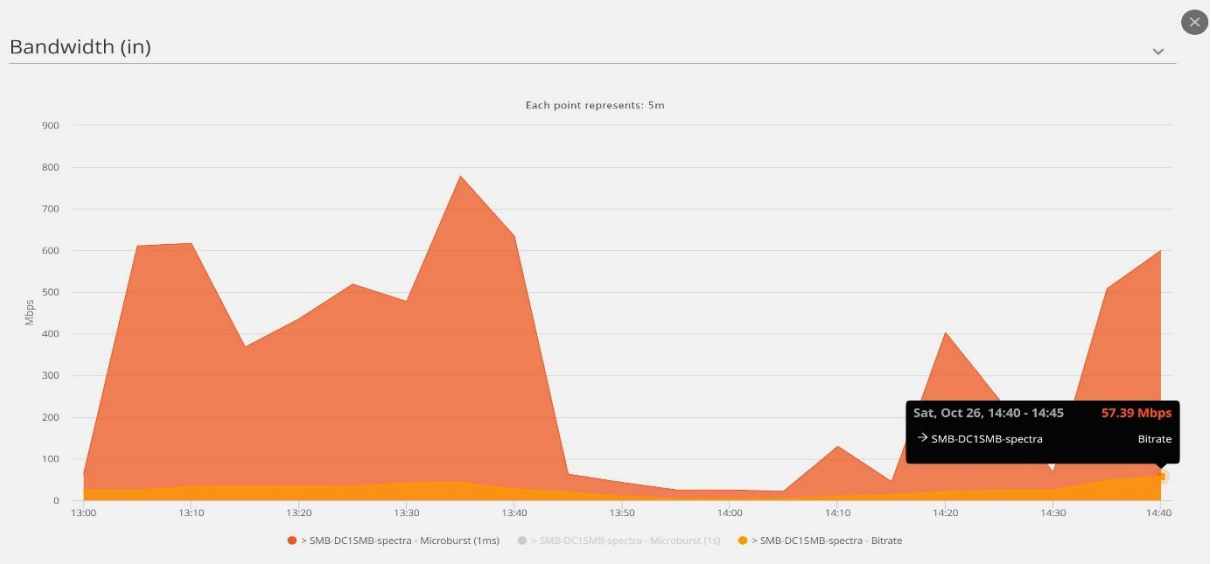
SIMBA versus TWIME comparison:



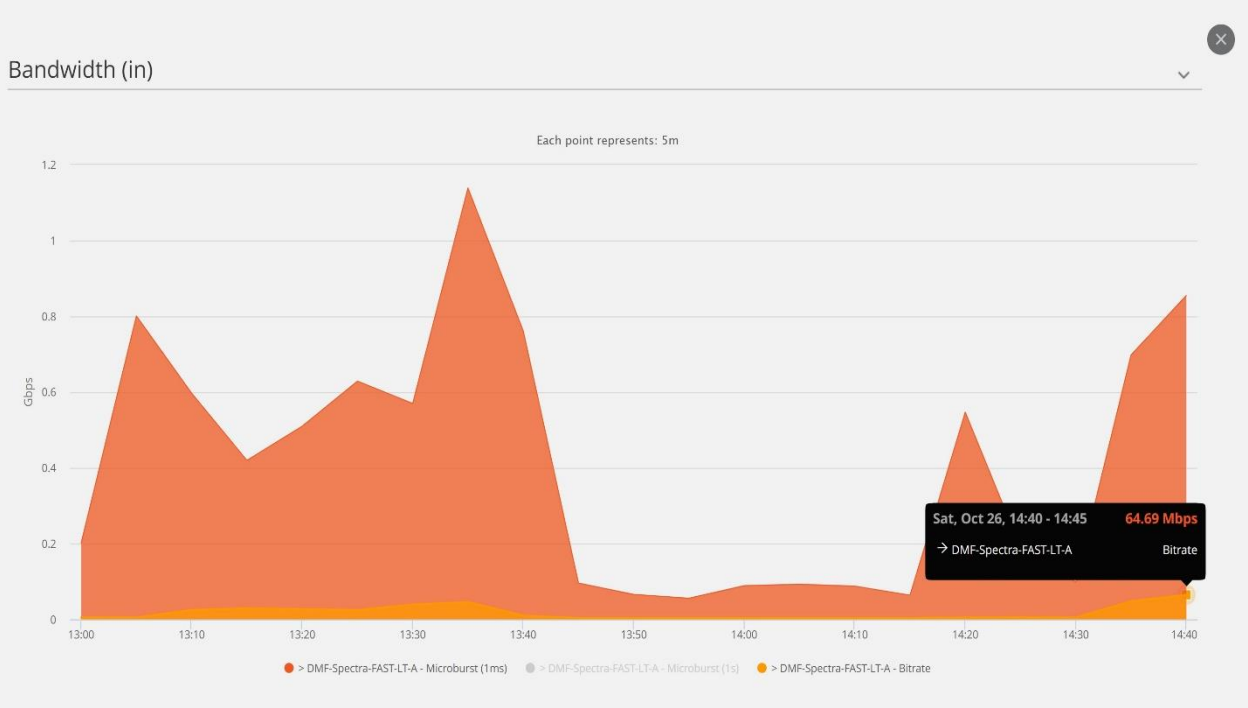
In 99% of cases, SIMBA market data for orders was sent at least 10 microseconds faster than TWIME transaction responses.

The median publication time for SIMBA during operations at 20,000 transactions per second was approximately 14 microseconds.

Graph of SIMBA UDP multicast traffic in copy A and millisecond interval peaks:



Graph of FAST UDP multicast in copy A and millisecond interval peaks:



FAST UDP multicast traffic reached 65 Mbps in each copy (A and B) with millisecond interval peaks of up to 1 Gbps.

SIMBA UDP multicast traffic reached 57 Mbps in each copy (A and B) with millisecond interval peaks of up to 0.75 Gbps.

The volume of transmitted market data directly depends on the system's network bandwidth, which in 2024 was lower than in 2023 due to reduced peak performance of the engine.

The required network bandwidth for clients using the FAST service to receive the ORDERS-LOG in real-time is 100 Mbps per feed. To receive two feeds, FEED A and FEED B, or data from multiple markets, the 1-10 Gbps bandwidth is recommended.

Exchange network and colocation network

Monitoring of the internal exchange network and colocation network, including the colocation core, revealed no anomalies. No packet losses, retransmissions, or delays were detected.

We remind you that network bandwidth recommendations and requirements are published on MOEX website at https://www.moex.com/a1873.

Conclusions

The Securities and FX markets

1. Load testing confirmed that the Trading/clearing system of the Securities and FX markets have significant capacity margin for maximum orders and trades compared to production environment.

2. The maximum network bandwidth of the Trading/clearing system remained unchanged compared to 2023 values.

3. A scheduled upgrade of the Trading/clearing system servers in 2025 may result in a 1.5x increase in peak traffic for the FAST UDP multicast marketdata.

Derivatives market

1. Testing revealed that enabling hot backup adversely affects the performance of the Spectra trading/clearing system both in terms of latency (increasing from 70-75 to 120-130 microseconds) and network bandwidth (decreasing from 120,000 to 72,000 transactions per second).
   Despite this, the Trading/clearing system performance remains sufficient to meet member demands under peak load. Performance during testing (72,000 transactions per second) was nearly three times higher than the recorded maximums in production environment (25,000 transactions per second).
2. The engine modules failover time ranged from 2 to 4 seconds, aligning with the initial expectations. An incident during the failover of one of the modules is under investigation; its root cause will be addressed in the production version of the hot backup system.

Comparison of load testing parameters and peak load values in production environment

This section provides a table comparing the load testing parameters to the maximum values of similar parameters in production environment.

Due to the automatic scaling of data table sizes, the concept of maximum number of orders and trades is not tied to initial table size settings. For main servers of the Trading/clearing system, memory limitations are at least twice the measured values.

Parameter	FX market	Derivatives market	Securities market
Peak number of trades per day	1,500,000	2,039,051	12,000,000
The number of trades in load testing	2,944,171	3,643,536	13,258,761
Peak number of orders per day	85,000,000	67,180,418	101,500,000
The number of orders in load testing	124,728,799	65,608,305	143,369,861
Peak numbers of orders, trades, and transactions, production environment	N/A		
Second interval peak number of transactions, production environment	12,000	25,000	21,000
Peak transaction frequency, load testing	59,628	72,000	62,215