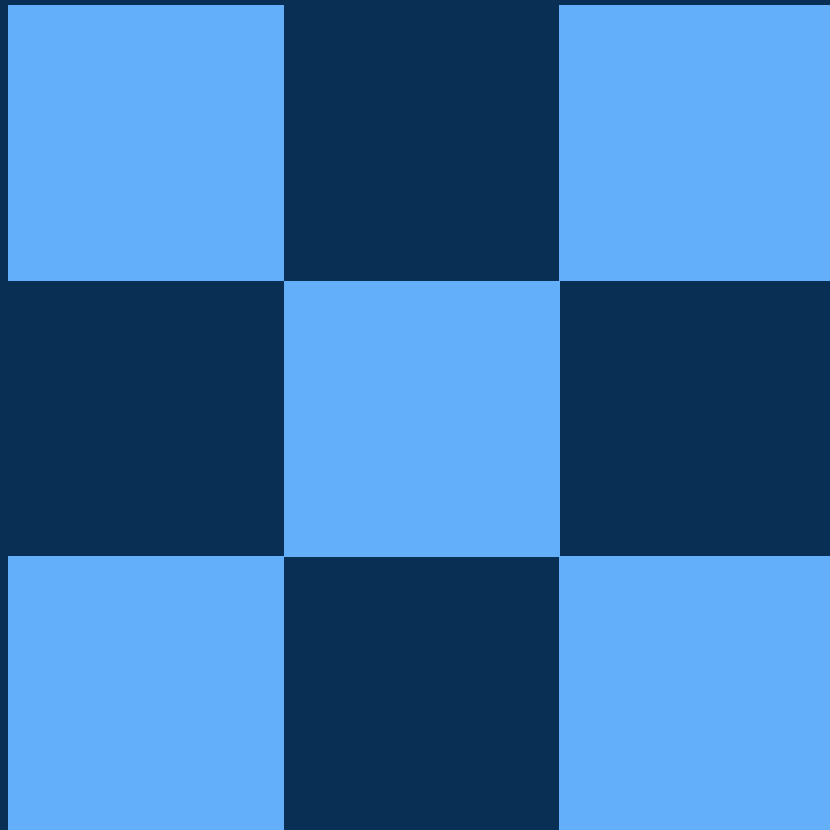


Vault Core

Peak Performance, Proven the Right Way



Background

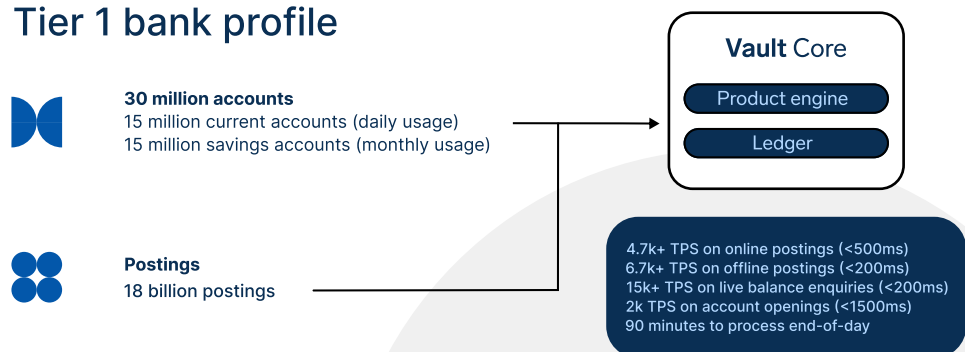
A typical bank processes billions of transactions daily, from retail payments to payroll, and poor performance in its core banking system can lead to slowness, outages, and operational risk. If the system struggles to process a bigger-than-expected volume of daily transactions, it can cause backlogs and even complete outages. Potentially halting all banking operations and causing significant financial and reputational damage, such as financial losses and regulatory non-compliance with fines.

Any bank needs confidence that its core will scale to handle spikes in demand with predictable latency and resilience, especially when unexpected. This confidence can be provided by a rigorous testing framework that delivers transparent, reliable results that reflect the complexity of real-world banking.

The usual approach is often unclear and lacks reliability. You'll see headline numbers like Transactions per Second (TPS) without any details on what they actually mean, how they were calculated, or if they're even reproducible.

Thought Machine's Vault Core is a high-performance banking platform built for modern financial services. Its performance testing proves that a real-time, cloud-based system can meet the demands of speed and reliability and support the bank at any scale.

Tier 1 bank profile



In this paper, we will detail our performance testing methodology and the results we've achieved with the Vault platform. We'll also discuss our plans to continue pushing the limits and aim to serve the world's largest banks.

Our high-integrity performance testing methodology

Our approach is built on a foundation of realism and reproducibility. We pride ourselves on being leaders in transparent, meaningful performance validation. To achieve this, we've developed a comprehensive testing framework that showcases Vault Core's capabilities in real-world production environments, built on the following key pillars:

- Realistic financial products
- Real-world production environments and pre-loaded data
- Clear real-world journeys
- End-to-end reproducibility
- Continuous performance gate

Realistic financial products

Many performance tests are artificial and don't reflect real banking. We solve this by employing three distinct financial product portfolios that cover a wide range of features and use cases found globally. This ensures our tests are grounded in authentic banking scenarios, providing you with insights that are directly applicable to your real-world operations.



Daily Use Accounts

- Such as current accounts and credit cards
- High-priority transactions: 6 peak / 4 median
- Low-priority transactions: 3 peak / 2 median



Monthly Use Accounts

- Such as savings accounts, loans, and mortgages
- High-priority transactions: 1 peak / 1 median
- Low-priority transactions: 1 peak / 2 median



Standard Retail Blend

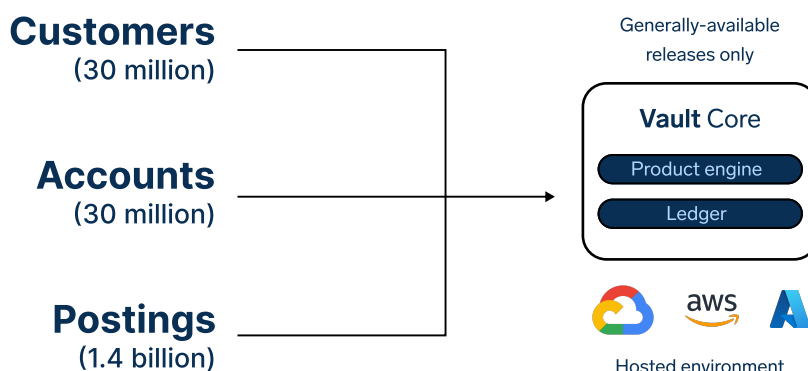
- 50/50 blend of daily and monthly use accounts

Real-world production environments and pre-loaded data

How can you be confident in performance results without a true production environment? Our performance tests are conducted in environments that meticulously replicate real-world production setups. We utilise documented infrastructure configurations on leading cloud providers, including AWS, Azure, and GCP. This ensures our results are directly translatable to your potential deployment, giving you a direct preview of Vault Core's performance in your operational landscape. In fact, you could achieve even better performance by leveraging more advanced infrastructure setups.

The test environment is also pre-loaded with a significant, production-scale dataset, including millions of customers, accounts, and a comprehensive history of transactions. This is crucial for mimicking the data density and complexity of a real-world banking system.

Only general availability releases of Vault Core are deployed onto the environments, and we strictly avoid any experimental code or infrastructure configuration.



Clear, real-world journeys

Vault Core's performance is measured across seven meaningful test journeys designed to cover a full range of banking operations, from single actions like opening an account to the intense, simultaneous demands of peak high-priority transactions and complex end-of-day batch runs. These test journeys are measured in four different test types, providing a holistic view of the platform's efficiency, scalability, and stability under various loads.

Clear, real-world journeys

This comprehensive approach empowers a clear assessment of the platform's ability to handle transaction volumes and understand its behaviour under peak conditions. Ultimately, this helps ensure the platform consistently meets your business requirements and provides a reliable user experience, preventing issues like slow response times that could impact customer satisfaction and business operations.

Test journeys

- High-priority transactions
- Low-priority transactions
- Account opening
- Live balance enquiry
- End-of-day processing
- End-of-day balance reconciliation
- Smart contract upgrade

X

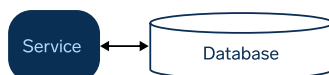
Test types

- Max throughput within SLO
- Throughput
- Round-trip time
- Fixed duration

End-to-end reproducibility

Unlike legacy systems that often rely on low-level, hardwired tests which bypass critical API layers and services, our testing framework ensures every test request passes through the same public APIs, internal services, event streams, and database tables as a live, production transaction. This guarantees that test results are not only realistic but also fully reproducible, giving you a clear and honest representation of how many production accounts can be supported.

Legacy



- Typically low-level tests and hardwired paths to specific use cases
- Blended journeys (transactions and balance inquiry) with API layers stripped away and services written directly to database
- Black box with mixed read and writers, not producing realistic production results

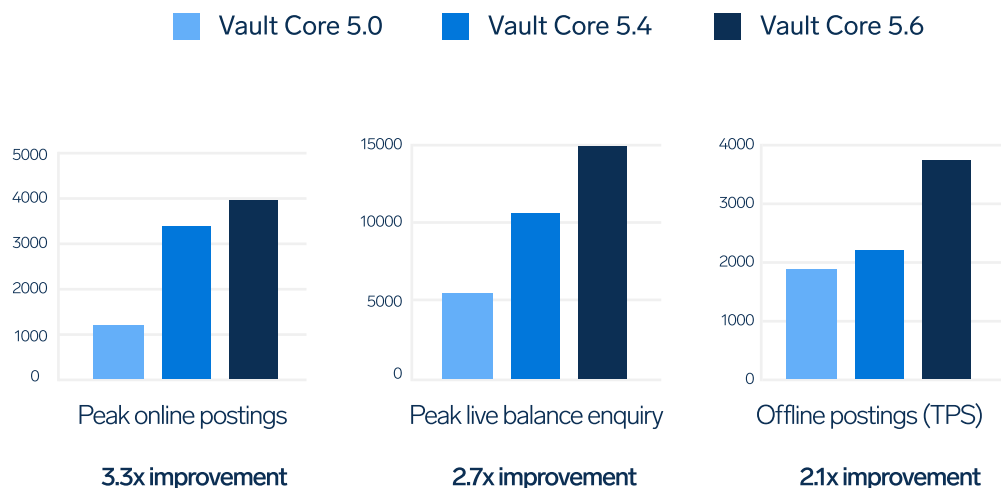
Vault Core



- Realistic and reproducible tests with full Product Library functionality maintained
- Split journeys for full transparency
- Replicating real-world scenarios, clearly identifying how many production accounts can be supported.

A continuous performance gate

To ensure consistent quality and improvement, every new Vault Core release must pass the target set before it is approved for deployment. This rigorous process guarantees a stable and continuously improving performance baseline with each update, giving you peace of mind.



Our performance results demonstrated

Certified results (baseline)

Building upon our rigorous testing methodology, we produce a certification performance report with every release. Among other things, the report shows the performance results of Vault Core within a Tier 1 banking environment, simulating 30 million live accounts and over a billion transactions—a blend of 15 million current accounts with daily usage and 15 million savings accounts with monthly usage. Our report measures six distinct test journeys against four test types, assessing performance both in isolation and in blended scenarios. This approach simulates live operational conditions, providing an authentic depiction of Vault Core's power and efficiency in real-world banking.

Below are the results for the key journeys. See the appendix for the full report:

- **High-priority transactions (Online Postings):** Achieving 4,769 TPS with 95% of responses delivered in under 500ms, ensuring rapid and seamless customer interactions.
- **Low-priority transactions (Offline Postings):** Demonstrating robust processing with 6,768 TPS, where 95% of operations complete in less than 200ms, underscoring our efficiency in batch processing.

Our performance results demonstrated (cont'd)

- **Account opening:** Facilitating expansion with 2,000 TPS for account openings, maintaining a 95th percentile response time of under 1,500ms, and empowering swift client onboarding.
- **Live balance enquiry:** Providing immediate financial insights with 15,035 TPS for live balance enquiries, with 95% of responses delivered in under 200ms, enhancing real-time financial visibility.
- **End-of-day processing:** Consolidating daily operations in just 90 minutes, processing at a rate of up to 20,676 TPS, ensuring timely and accurate financial reconciliation.

Journey	Performance results (p95)
High priority transactions	4,769 TPS (<500ms)
Low-priority transactions	6,768 TPS (<200ms)
Account opening	2,000 TPS (<1,500ms)
Live balance enquiry	15,035 TPS (<200ms)
End-of-day processing	<90 minutes at 20,676 TPS

Lab studies (beyond certification)

Beyond standard certification, Thought Machine is committed to continuously pushing the boundaries of Vault Core's capabilities. In a **Tier 1 banking environment with 70 million live accounts and 28 billion transactions**, we have successfully developed and implemented the following showcases for two of our Tier 1 clients. This underscores our dedication to innovation and our pursuit of unparalleled performance in the financial technology sector.

Pushing throughput to the limit (TPS)

In this lab study, we demonstrated the maximum throughput achievable with Vault Core for a bank with a very high volume of accounts (70 million). Blended tests showed a sustained capacity of **8,000 TPS for high-priority transactions under 151ms, and 8,000 TPS for live balance enquiries under 42ms** (95% percentile).

Environment	70 million accounts (50% current accounts, 50% savings accounts)
Goal	Demonstrate how much throughput Vault Core can sustain
Result	High-priority transactions: 8,000 TPS (<151ms) Live balance enquiries: 8,000 TPS (<42ms)

Lab studies (beyond certification)

Supporting Very High Volume Accounts

We define a high-volume account (HVA) as an account with more than 50,000 daily transactions. This lab study demonstrated that Vault Core can sustain performance levels comparable to certified ones in an environment with 70 million accounts, including 32 HVAs with 50,000 daily transactions each and three more with 140,000 daily transactions each. This demonstrates the platform's ability to support enterprise banks.

Environment	70 million accounts, including • 32 HVAs (50,000 daily transactions each) • 3 HVAs (140,000 daily transactions each)
Goal	Demonstrate Vault Core's ability to support enterprise banks
Result	Similar performance to the certification performance report

Conclusion

Vault Core's meticulously engineered architecture and rigorous testing methodology solidify its position as an industry-leading core banking solution. The detailed test results and real-world showcases demonstrate its capacity to handle the immense scale and complexity of modern banking operations with industry-leading throughput, low latency, and reliability.

By choosing Vault Core, financial institutions gain a platform that not only meets today's performance demands but is also engineered to scale for the future. Global banks such as JPMorgan Chase, Standard Chartered, Intesa Sanpaolo, and Lloyds have already chosen Vault Core as their strategic core banking platform, investing in a long-term partnership with Thought Machine. This partnership will continue to uphold the highest engineering standards and ensure their performance capabilities evolve with their ambitions.

Appendix

Preface

What this report contains

This report details the results of performance tests conducted on this Vault release in a standardised environment configuration.

This report will assist clients in understanding Vault's performance capabilities and improvements in performance across Vault releases.

- The overview describes the types of performance tests undertaken, the environment setup, and how to interpret the results.
- Realistic Bank Performance summarises the results for tests with production-like contracts, data loads and configurations.

Who should read this report?

This report is intended for both existing Thought Machine clients and prospective clients interested in evaluating Vault's capabilities.

Typical readers would be:

- Senior Project Stakeholders
- System Architects
- Developers
- QA Engineers

Overview

About the performance tests

Performance tests use the public endpoints (HTTP or Kafka) that Vault exposes to replicate behaviour as closely as possible to real-world behaviour.

In all cases, the reported results are end-to-end measurements as observed by our load test frameworks and therefore representative of Vault API usage.

Many parameters adjust the trade-offs in Vault between cost, availability, and performance. These tests have been executed using realistic cloud hardware for banks of particular sizes.

Overview (cont'd)

Journey types

A journey is a typical use case for Vault, and we are interested in measuring its performance. Each journey is classified as one of these four types:

Journey Type	Units	Description	Example
Throughput	TPS / RPS	Run a process for a fixed period of time, then measure the median and peak TPS.	Offline postings have no round-trip time SLO but must have high throughput.
Round Trip Time	Seconds	Run a process at fixed TPS for a fixed period of time, and measure the 95th percentile round-trip request time	Median online postings is a journey that tells us how quickly Vault can process high-priority postings at our expected median rate
Max Throughput Within SLO	TPS / RPS	Run a process at a low TPS initially, and ramp it up until we fail the round-trip time SLO. Measure the maximum TPS achieved before the SLO breach	Peak online postings is a journey that tells us the maximum TPS Vault can sustain before it breaches SLO
Fixed Duration	Seconds	Measures the total time taken by a process. The target is the same for all deployment sizes and profiles.	End of Day should take 30 minutes or less.

Environment configuration

All Vault performance tests are carried out on infrastructure capable of handling the provided test load.

Infrastructure Config: Please refer to the Infrastructure Config sections for a given Bank size or profile for more details, as the configuration may vary by account volume.

Realistic bank performance

Bank profiles

To reduce the combinations required for testing (and to provide confidence in unseen bank workloads), we categorise a bank's product performance into two basic profiles and one blended profile. We aim to test these over several different deployment sizes:

Realistic bank performance (cont'd)

- Small: up to 1 million accounts
- Medium: between 1 million and 10 million accounts
- Medium-Large: between 10 million and 30 million accounts
- Large: between 30 million and 100 million accounts

Certain results for specific deployment sizes will be omitted while the test framework is continuing to be developed.

Daily use

This profile captures customer accounts that have multiple transactions each day, and have a daily schedule of interest accrual:

- Example products: Current Account, Credit Card, Wallets, Charge Cards
- Product Breakdown Used: 80% Current Accounts, 20% Credit Cards
- Performance Profile: This profile models a median of 4 online transactions per account per day and a peak of 6. A median of 1 offline transaction a day, and a peak of 3. It contains one interest schedule per account at 9pm
- Smart Contract Implementation: This contract uses a simple latest balance check for authorisations (pre-posting), and accrues interest on the current balance at 9pm
- Important metrics: peak online TPS under SLO, round-trip time (RTT) profile of median daily online TPS, Balance enquiry QPS and RTT, and Offline TPS

Monthly use

This profile represents customer accounts that generally do not see daily use, but operate on monthly cycles:

- Example Products: Instant Access Savings, Personal Loans, Mortgages, BNPL
- Product Breakdown Used: 40% Savings, 30% Personal Loans, 30% Mortgages
- Performance Profile: This profile models a median of 1 and a peak of 2 offline transactions per account per day. It has a single online transaction per account per day and a single monthly interest schedule at 9pm on the 28th of each month.
- Smart Contract Implementation: Latest balance check on online postings, interest is paid on the balance held in the account on the 28th of the month
- Important metrics: Offline TPS

Realistic bank performance (cont'd)

Standard retail blend (SRB)

Based on an analysis of the account volumes provided during RFPs, one additional blended performance profile is to be used for the testing and certification of Vault Core. This is a standard retail bank that includes a mix of daily use products (current accounts, etc) and some longer-term savings and lending products

- 50% Daily Use: 40% Current Accounts, 10% Credit cards
- 50% Monthly Use: 20% Savings, 15% Personal Loans and 15% Mortgages

Headline results for each performance test are reproduced below. For the full results for each scenario, see the sections below.

Interpretation of results

Each reported performance metric is a test of a specific scenario, carried out on one of the performance profiles detailed above; these constitute our current best understanding of a production-like bank setup and aim to be as representative as possible of a production environment with BAU traffic.

For each user journey, Thought Machine runs tests at different volumes. The reported results that have been selected:

- Are representative of Vault API usage
- Are reproduced multiple times to account for variance in test results,
- Set at a throughput for which the system response times remain within our internal target SLOs

There are many parameters, such as business, platform, and cloud infrastructure, that can be adjusted to tune variants for operational costs, availability, and performance.

Where an error rate has not been explicitly recorded, it can be assumed to be 0%.

Realistic bank performance (cont'd)

Metric calculations

Our load test framework runs within the System Under Test's cluster, and as such, we have no network latency in the Round Trip Times reported in this report.

Round Trip Times (RTT)

For Kafka-based load tests:

- We store `time.Now()` right before we publish to the request Kafka topic
- We store the time when the responses hit the response Kafka topic, taken from the timestamp of the Kafka messages themselves
- We subtract the request time from the response time to get our individual RTT and calculate percentiles on the full dataset of that

For HTTP-based load tests:

- We store the time when the HTTP request is sent
- We store the time when the HTTP response is received
- We subtract the request time from the response time to get our individual RTT and calculate percentiles on the full dataset of that

Performance profile results

30 million bank

Infrastructure config

The current set of Vault performance tests is carried out on the following environment setup.

Component	Description
Cloud Provider	• GCP
Kubernetes nodes	• e2-custom-8-24576 • 8 vCPUs + 24 GB RAM
Kafka	• 6 brokers • 3 Ti storage per broker • 7-day retention period • Replication factor: 3
Database	• PostgreSQL-16 compatible • AlloyDB Primary cluster • 1 primary instance, no read pool • Contains all logical databases except Warm Storage • 64 vCPUs and 512 GB RAM • Scaling storage
Database flags	• alloydb.iam_authentication to on • autovacuum_analyze_scale_factor to 0.05 • autovacuum_analyze_threshold to 500 • autovacuum_freeze_max_age to 800000000 • autovacuum_max_workers to 6 • autovacuum_multixact_freeze_max_age to 800000000 • autovacuum_vacuum_cost_delay to -1 • autovacuum_vacuum_scale_factor to 0.1 • autovacuum_vacuum_threshold to 500 • log_autovacuum_min_duration to 0 • log_min_duration_statement to 4000 • log_statement to ddl • maintenance_work_mem to 2097152 • max_connections to 5000 • max_locks_per_transaction to 2048 • max_pred_locks_per_page to 64 • max_pred_locks_per_transaction to 2048 • max_wal_size to 8192 • password_encryption to scram-sha-256 • timezone to UTC • vacuum_freeze_min_age to 5000000 • vacuum_multixact_freeze_min_age to 5000000 • work_mem to 65536
Dependency versions	• Kafka: 3.9.0 • Istio: 1.26 • Kubernetes: 1.32 • HashiCorp Vault: 1.16
Vault configuration	• common.deployment_size: medium-large
Warm Storage Database	• PostgreSQL-16 compatible • AlloyDB Primary cluster • 1 primary instance, no read pool • Only contains the Warm Storage logical database • 64 vCPUs and 512 GB RAM • Scaling storage

This configuration will ensure the environment will not hit out-of-shared-memory errors caused by the limitation in the number of locks it can give on exclusive grants on the tables.

Performance profile results

Standard retail

Journey Name	Journey Type	Units	Target	Result	Progress
Peak Online Postings	Max Throughput Within SLO	TPS	1,042	4769.05	457.68%
Peak Account Opening	Max Throughput Within SLO	TPS	1,170	2900	247.86%
Peak Live Balance Enquiry	Max Throughput Within SLO	Effective QPS	8,334	13787.46	165.44%
Peak Live Balance V2 Enquiry	Max Throughput Within SLO	Effective QPS	8,334	15035.84	180.42%
Offline Postings	Throughput	TPS	869	4666.56	537.0%
Median Online Postings	Round-trip time	Milliseconds	500	161	310.56%
Median Online Postings Sync	Round-trip time	Milliseconds	500	158	316.46%
Median Account Opening	Round-trip time	Milliseconds	1,500	125	1,200%
Median Live Balance Enquiry	Round-trip time	Milliseconds	200	33	606.06%
Median Live Balance V2 Enquiry	Round-trip time	Milliseconds	200	28	714.29%
End Of Day	Round-trip time	Seconds	7,200	5,700	126.32%
End Of Day Balance Reconciliation	Round-trip time	Seconds	-	-	-

Performance profile
results

Standard retail

Blended Name	Journey Name	Journey Type	Units	Target	Result	Progress
Online Postings And End Of Day	Median Online Postings	Round Trip Time	Milliseconds	500	420	119.05%
	End Of Day	Fixed Duration	Seconds	7,200	5,912	121.79%
Online Postings Sync And End Of Day	Median Online Postings Sync	Round Trip Time	Milliseconds	500	448	111.61%
	End Of Day	Fixed Duration	Seconds	7,200	5,461	131.84%
Online Postings And Live Balance Enquiries	Median Online Postings	Round Trip Time	Milliseconds	500	159	314.47%
	Median Live Balance V2 Enquiry	Round Trip Time	Milliseconds	200	43	465.12%
Online And Offline Postings	Median Online Postings	Round Trip Time	Milliseconds	500	141	354.61%
	Offline Postings	Throughput	TPS	869	1984.56	228.37%

Performance profile results

Standard retail (cont'd)

Blended Name	Journey Name	Journey Type	Units	Target	Result	Progress
Online Postings With HVA And VHVA Postings	Median Online Postings Sync	Round Trip Time	Milliseconds	500	177	282.49%
	HVA Offline Postings	Throughput	TPS	1,042	1521.46	146.01%
	VHVA Offline Postings	Throughput	TPS	125	233.32	186.66%
Smart Contract Upgrade And Account Opening	Smart Contract Upgrade	Fixed Duration	Seconds	14,400	14,308	100.64%
	Median Account Opening	Round Trip Time	Milliseconds	1,500	101	1485.15%
Online Postings, Account Opening And Live Balance Enquiries	Median Online Postings	Round Trip Time	Milliseconds	500	195	256.41%
	Median Account Opening	Round Trip Time	Milliseconds	1,500	144	1041.67%
	Median Live Balance V2 Enquiry	Round Trip Time	Milliseconds	200	53	377.36%



www.thoughtmachine.net

