Aeron Cluster

A framework for high-performance, in-memory, fault-tolerant services

Aeron Cluster enables resilient, 24/7 workloads in the cloud, on-prem or in a hybrid deployment, simplifying overall development efforts.

Organisations using Aeron Cluster operate some of the world's highest throughput, and lowest latency applications with a reliable and predictable latency of 18 microseconds on physical hardware and less than 100 microseconds in Cloud environments, with throughput of over 1 million messages per second for 32 byte messages.

Aeron Cluster implements the Raft Consensus Algorithm to provide log replication and allow multiple nodes to maintain the same state as well as automated leader election to ensure that there is a single leader within the cluster. Automatic failover with Aeron Cluster takes less than a second, minimising the impact on the system and ensuring that critical data is not lost.

Aeron Cluster Replication and Recovery

Within Aeron Cluster, replication (ensuring that a follower node has the same state as the leader) and recovery (restoring a stopped node to its previous state) are similar challenges, solved with the same mechanism.

That solution relies on three things:

1. A snapshot of either the node's initial state when first provisioned, or a state from another point in time.
2. An ordered log of all input messages, handled by Cluster's Raft implementation.
3. Deterministic application logic (ie, logic that, when presented with the same starting point and the same inputs, will always arrive at the same result state.)

As it is impractical to replay all data from the "beginning of time", the system can take snapshots periodically, e.g. daily or hourly. The frequency of snapshots should be determined by the volume of data into the system, the throughput of the business logic, and the desired mean time to recovery.

Use cases

- Host one or more central limit order books
- Host RFQ, IOI or similar negotiation logic
- Message sequencer
- Any other scenario that requires fault-tolerant, high-performance state

Cluster Capabilities

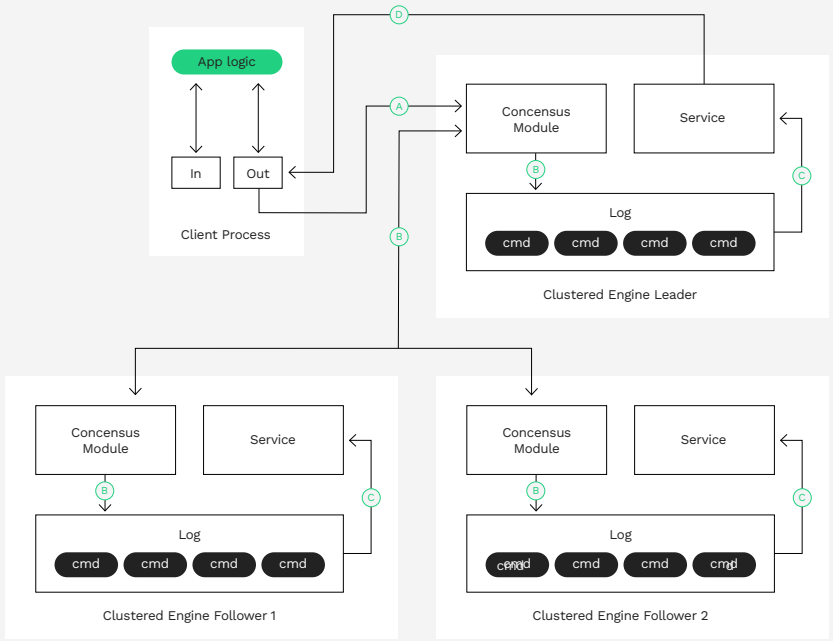
+ Reacts to partial system failure, including failure of the leader node, as long as a quorum of (n/2) + 1 nodes remain available
+ Sequences multiple client connections into a single, replicated log
+ Provides efficient fault tolerance with 3 or more nodes
+ Supports snapshotting, which reduces recovery time by limiting the number of messages that need to be replayed
+ Allows 1 or more clustered services to be run, with support for inter-service sequenced messaging
+ Enables reliable, sequenced timers to trigger business or technical events
+ Offers high levels of performance, in particular with Aeron Premium features for Aeron Transport such as Solarflare

Raft Consensus and Replication

The Raft Consensus algorithm has growing usage in distributed systems used for infrastructure as well as business logic services such as matching engines and trading platforms in financial trading systems.

Raft is an architecture for replicating logs from a leader to followers. Raft clients send in commands and receive back events, as in the following diagram:

Aeron Cluster Message Flow:



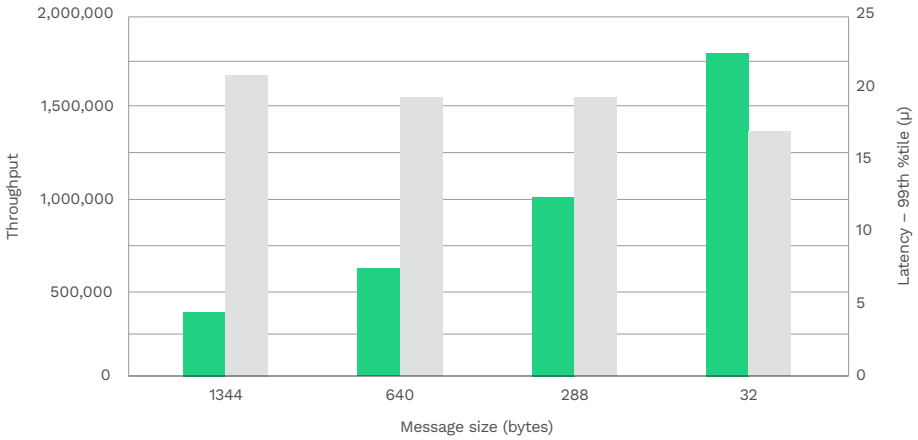
A The client process sends the command into the leader server consensus module.

B The leader consensus module appends the command to the local log. In parallel, the consensus module replicates the commands to the follower nodes. Once a quorum of nodes accepts the new command, the command is committed and the committed command is then handed off to the service for processing.

C All three service instances accept and process the command.

D Application logic produces 1 or more events, leader sends them to the client.

Aeron Cluster Throughput and Latency for Different Message Sizes*



*Aeron Cluster running on physical hardware, leveraging Aeron Transport Kernel Bypass. Aeron Cluster is receiving a message, persisting, replicating across nodes and responding to the message.

By Adaptive

Adaptive builds & operates bespoke trading technology solutions across asset classes for financial services firms wanting to own their tech stack to differentiate and compete in the long-term. Central to Adaptive's offering is Aeron, the global standard for high-throughput, low-latency and fault-tolerant trading systems - the open-source technology supported and sponsored by Adaptive.

AERON GITHUB COMMUNITY

AERON COOKBOOK

aeron.io / info@aeron.io / weareadaptive.com