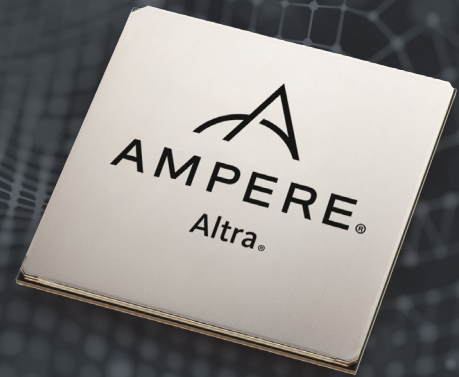




Workload Brief

Canonical Anbox Cloud on Ampere® Altra®



Ampere®—Empowering What's Next

The Ampere® Altra® and Ampere® Altra® Max processors are complete system-on-chip (SOC) solutions built for Arm native applications. Ampere Altra supports up to 80 AArch64 cores and Ampere Altra Max supports up to 128 cores. In addition to incorporating a large number of high-performance cores, the innovative architecture delivers predictable high performance, linear scaling, and high energy efficiency.

Anbox Cloud supports cloud phone and cloud gaming use cases and can be deployed as a service on public cloud or on bare-metal Ampere platforms. It is Linux container based and uses the popular Ubuntu Linux operating systems and Canonical services such as MaaS and Juju for deployment and orchestration. Anbox Cloud is a very scalable solution, providing deployment on a single node or in multi-node clusters.

In this workload brief, we focus on cloud gaming workload running 3D games on Canonical Anbox Cloud software stack on Ampere Altra platform – Gigabyte G242-P33 which is a 2U single socket Altra Q80-30 – populated with a single Nvidia A16 GPU. We will measure numbers of streaming instances, FPS and CPU utilization for this workload.

Canonical Anbox Cloud on Ampere Altra

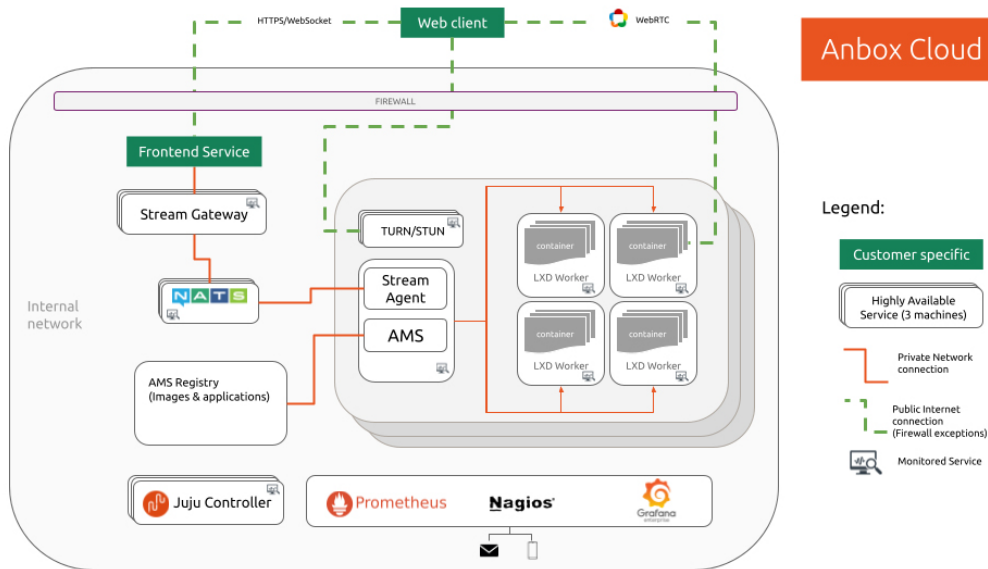
Ampere Altra processors are designed to deliver exceptional performance for Arm native applications like Anbox Cloud. This is accomplished through an innovative architectural design, operating at consistent frequencies, and using single-threaded cores that make applications more resistant to noisy neighbor issues. This allows workloads to run in a predictable manner with minimal variance under increasing loads. The processors are also designed to deliver exceptional energy efficiency. This translates to industry leading performance/watt capabilities and a lower carbon footprint.



Benefits of running Canonical Anbox Cloud on Ampere Altra

- **Cloud Gaming:** Designed from the ground up with high core density, provide high instance density for cloud gaming.
- **Consistency & predictability:** Single-threaded cores running at fixed maximum frequencies ensure linear scaling.
- **Energy efficiency:** With energy-efficient Arm cores, Ampere Altra can consume lower power while maintaining competitive levels of performance.
- **Data center class platform:** Broad platform ecosystem with high GPU density.
- **Best performance per watt:** Outstanding TCO (Total Cost of Ownership) – best performance per dollar for a solution of equivalent capacity.

Canonical Anbox Cloud Architecture Diagram



Anbox Cloud comes with additional components to build a control plane which simplifies and abstracts the use of the underlying container platform. The Anbox Management Service (AMS) provides an abstraction layer on top of LXD which adds resource management, container orchestration, application lifecycle management and various other things which are needed to run Android at scale and high density on a single machine. To facilitate Anbox Cloud handling scale out across multiple regions, the Anbox Stream Agent connects a single region powered by AMS and LXD via a NATS message queue to a centralized management service called Anbox Stream Gateway. The Anbox Stream Gateway allows the creation of user specific streaming sessions and routes users to the nearest region.

Benchmarking Configuration

Anbox Cloud enables us to run automated tests for Android applications at scale. A benchmark for streaming requires more automation than just starting containers. Therefore, Anbox Cloud provides a dedicated benchmark tool for this purpose. The tool automates the following tasks:

- Create a streaming session
- Receive the video/audio stream
- Collect various statistics
- Optional: Dump the received stream to a local file

Ampere Altra

- 80 64-bit cores at 3.30GHz
- 64KB i-Cache, 64KB d-Cache per core
- 1MB L2 Cache
- 32MB System Level Cache
- 2x full-width (128b) SIMD
- Coherent mesh-based interconnect

Memory

- 8x72 bit DDR4-3200 channels
- ECC and DDR4 RAS
- Up to 16 DIMMs and 4 TB addressable memory

Connectivity

- 128 lanes of PCIe Gen4
- Coherent multi-socket support
- 4x16 CCIX lanes

System

- Armv8.2+, SBSA Level 4
- Advanced Power Management

Performance

- SPECrate®2017 Integer Estimated: 300

The benchmark tool is provided through the *anbox-cloud-tests* snap. Use the following command to install it:

```
snap install anbox-cloud-tests
```

To run the benchmark, we must provide an authentication token for the Anbox Stream Gateway. The results are printed out as text. Alternatively, we can change the output format to JSON with *--format=json* and save the results to a file with the *--report-path=/path/to/report.json* option.

The following command runs the benchmark against an existing Anbox Cloud deployment:

```
anbox-cloud-tests.benchmark \  
--screen-width=1280 \  
--screen-height=720 \  
--screen-fps=30 \  
--stream-dump-path=/path/to/stream/dump/output \  
--application=my-application \  
--url=<https:// URL of the Anbox Stream Gateway> \  
--auth-token=<valid auth token for the Anbox Stream Gateway> \  

```

Fig 1: FPS at 80 instances

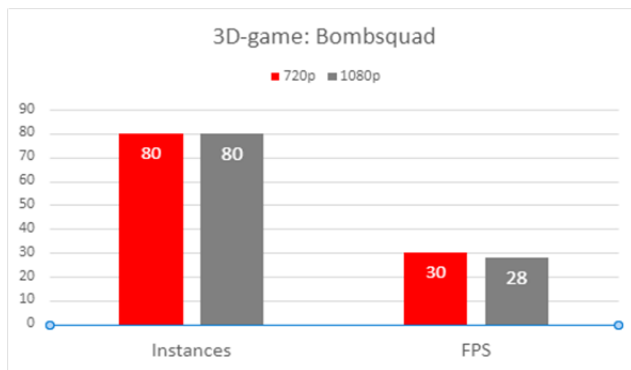
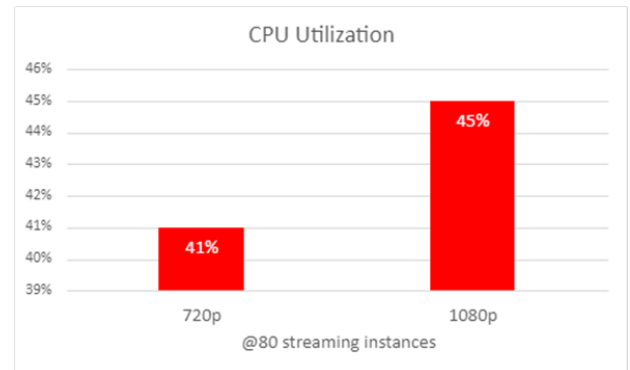


Fig 2: CPU Utilization



Benchmarking Results and Conclusions

Results in Fig 1 from Altra 1P server with 1x Nvidia A16 GPU running Canonical Anbox Cloud. Instances of 3D Android games are rendered and encoded at 1280x720@30fps and 1920x1080@30fps. We can achieve high instance density of 80 instances and consistent 28-30fps as shown in figure-1, with roughly 40 percent of CPU usages as shown in figure 2.

Ampere Computing reserves the right to make changes to its products, its datasheets, or related documentation, without notice and warrants its products solely pursuant to its terms and conditions of sale, only to substantially comply with the latest available datasheet. Ampere, Ampere Computing, the Ampere Computing and 'A' logos, and Altra Max are registered trademarks of Ampere Computing. Arm is a registered trademark of Arm Limited (or its subsidiaries) in the US and/or elsewhere. All other trademarks are the property of their respective holders.

©2022 Ampere Computing. All Rights Reserved.

Ampere Computing® / 4655 Great America Parkway, Suite 601 / Santa Clara, CA 95054 / amperecomputing.com

