Web Services on Ampere® Altra® Processors

Real world solutions using popular Cloud Native applications on

Ampere® Altra®

Ampere® Empowering What's Next

Web services are commonly deployed services comprised of multiple Cloud Native applications working together to deliver content over the Internet. These are increasingly built using a microservice-based architecture with individual microservices using different programming languages and frameworks, but loosely coupled to communicate with each other. They can be easily deployed, managed, and scaled using a containerized environment such as Kubernetes.

Cloud Native Advantage

Cloud Native is a modern approach to building and running software applications that make use of the flexibility, scalability, and resilience of cloud computing. More and more developers are embracing Cloud Native microservices-based architecture to develop and deploy applications such as web services to the cloud.

The web services used here simulate real-world services using many of the popular Cloud Native applications including NGINX, Apache HTTP server, Redis, Memcached, MySQL and MongoDB. These applications are run as microservices using a Kubernetes cluster with Docker containers.

Key Benefits

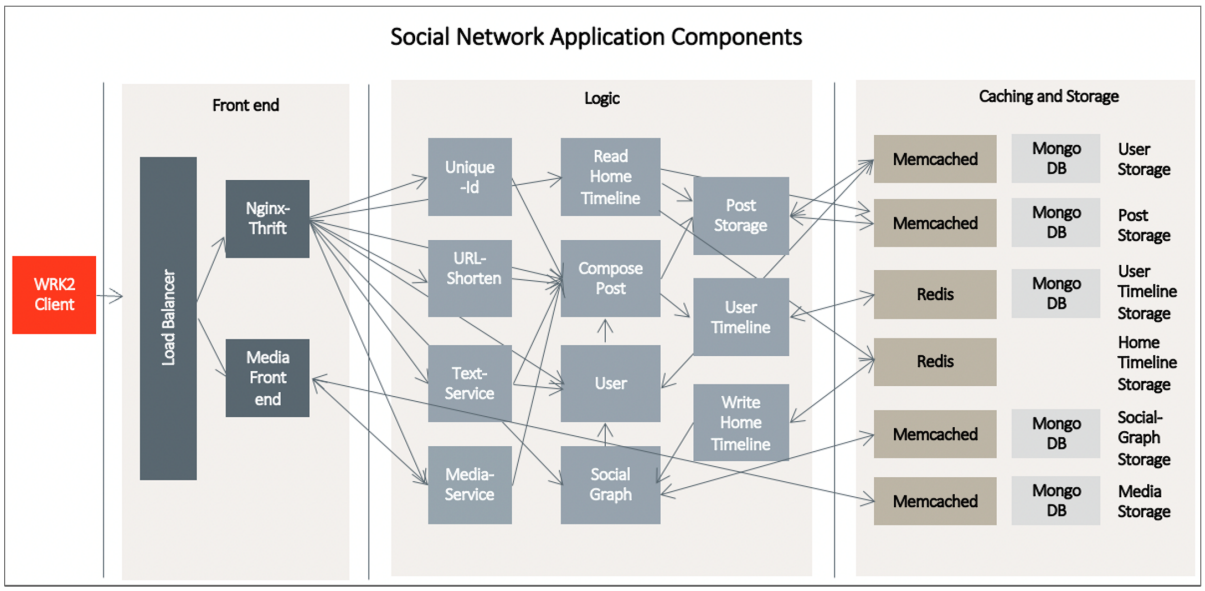
- **Cloud Native**: Ideal for web services that use common cloud-native applications and are deployed in a containerized environment.
- **Scalable**: Predictable performance for web services even under high utilization. Real world web services simulated on Ampere systems shows better throughput and lower latencies as compared to legacy x86 platforms.
- **Developer Friendly**: Robust ecosystem of applications supported on Ampere processors and ease of application portability to AArch64.
- **Power Efficient**: Competitive levels of raw performance while consuming less than half the power compared to the competition.

The Demo – Replicating Real-World Web Services

Web services showcased here are deployed on Ampere Altra servers or Virtual Machines running Ampere Altra instances in the cloud. The load generator is a client workload that simulates multiple simultaneous connections. The demo workloads use a Kubernetes deployment with the applications running as a collection of multiple Docker containers. The number of replicas of each pod, CPU and memory allocation for the pods is tuned to achieve the lowest P99 latency and highest throughput. At the end of the test, the load generator results indicate the average and 99th percentile of latency or response times for all client connections as well as the throughput measured in terms for requests per second.

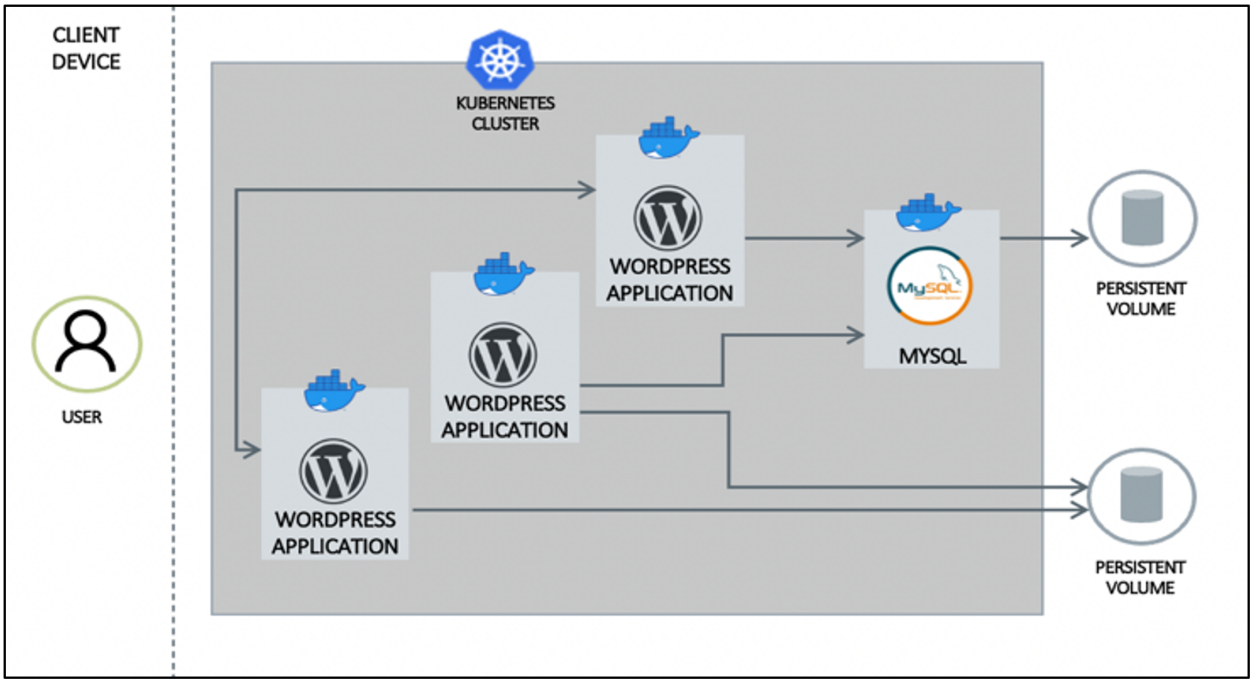
DeathStarBench Social Network Application

The DeathStarBench social network application running Ampere Altra servers simulates a twitter-like application running at scale with thousands of users connecting to the frontend using http, composing posts, tagging other users, adding media or URLs to the posts, and saving to the user and home timelines.



Scalable WordPress Deployment Using Kubernetes

The Wordpress solution is an example of a LAMP stack deployment built on PHP and MySQL. It is a popular website development framework used to create websites, blog, or apps. The goal is to benchmark the performance of a WordPress hosting website and determine the peak performance. Peak performance is measured using a load testing service such as K6 which emulates large numbers of users visiting a website and watching how well the hosting responds under these stressful conditions.



For additional information, visit the Ampere Solutions Portal.

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