

Onstream's Device-First Intelligence Framework Enabled by Docker and Automated with Amazon



Executive Summary

Onstream provides a software framework for OEMs to turn connected devices into smart devices. In a world where "smart devices" are simply connected devices streaming data to some distant database, Onstream activates the brain of each device, driving intelligence, responsive behavior, and new product features to networks of connected smart devices.

Onstream's framework shortens the time from connectivity to action by transforming raw data streams into useful information as soon as it leaves the device. For companies wanting to do more than just product monitoring and data analysis, Onstream provides simple visual tools to quickly and cost-effectively develop intelligent products; products that change their behavior in response to things happening near the device, far from the device, or wherever. And OEMs can extend Onstream's flexibility to their customers, allowing end users to tailor responsive product behavior to their specific needs. With Onstream, introducing and distributing new product features and capabilities to your connected devices is easily enabled within the framework.

The Challenge

Onstream had developed a new technology framework that enables connected devices to become smart devices without having to hardcode intelligence at the device. Their smart device framework enables early transformation of raw streams of device data, and can run on-premise or in the cloud. Portability was a key requirement to ensure the app could run in a cloud or on-premise environment.

Onstream was in stealth mode, preparing for launch, and focusing on building their application. They needed a partner who could design and build out the cloud infrastructure to support Onstream Mesosphere, the solution that runs in cloud environments.

The ClearScale Solution

Onstream wanted to follow emerging DevOps best practices with a focus on automation and decided to partner with ClearScale to build the infrastructure to power Onstream Mesosphere.

ClearScale designed an architecture that could support the application at scale and included cutting-edge container technology from Docker and the open-source Samza data stream processing framework from the Apache Software Foundation. For automation, ClearScale also recommended that the team use Amazon EC2 Container Service (ECS), a Docker orchestration service, which had just been released to general availability.

Following the architecture design, build, and test phases, ClearScale had built the new cloud infrastructure and rolled out the brand new cloud infrastructure on AWS in time to support Onstream's launch.

Docker Orchestration with Amazon ECS

From the project outset, the team decided that Docker would be used to containerize everything in Onstream's infrastructure including the application APIs, web apps, and data processing clusters.

[Docker](#) containers allow users to package an application with all of its dependencies into a standardized unit for software development. Individual services are compartmentalized into single components that can be mixed and matched depending on needs. By design, Docker enables application portability, which was a key project requirement.

[Amazon EC2 Container Service](#) (ECS) is used to automate launching Docker containers on EC2 instances. ECS provides all functionality to enable automated deployment and configuration of components within a distributed system.

Together, Docker and ECS give the ultimate in infrastructure flexibility and automation to build, ship, and run distributed applications.

ClearScale created Dockerfiles for each individual piece of Onstream's infrastructure – the Onstream Mesosphere APIs, the web app that end users interact with, and each Kafka, MongoDB, YARN, and Zookeeper node. A Dockerfile is a text document that contains all of the commands you would normally execute manually in order to build a Docker image. From the command line, Docker will build the image step by step according to the Dockerfile instructions. Onstream's images are saved in a local registry, which Amazon ECS then uses to launch containers on EC2 instances in Onstream's clusters.

ClearScale used [Tiller](#), an open-source tool that dynamically generates configuration files from templates inside Docker containers. Environment variables are used as input parameters to define values for configuration files. Each ECS cluster has an individual set of input parameters.

ClearScale set up Onstream's ECS deployment to run in Amazon VPC for increased security and control. The deployment takes advantage of Amazon ELB for load balancing and IAM for identity and access management.

Open Source Big Data Tools

Onstream's framework performs early transformation of raw streams of device data to enable real-time action and is designed to capture massive volumes of data. The open-source Apache Samza stream processing framework from the Apache Software Foundation for its ability to process data in near real-time with low latency.

Apache Samza is a lightweight distributed stream processing framework that allows for continuous data processing. It uses Apache Kafka for messaging and Apache Hadoop YARN (MapReduce 2.0) to provide fault tolerance, processor isolation, security, and resource management. The Samza framework is optimized for low-latency processing and messaging and allows for data to be processed in near real-time. This type of data processing framework is mandatory for an application like Onstream Mesosphere which was built to capture a deluge of data from the connected devices.

Samza continuously computes results as data arrives, which makes sub-second response times possible.

The Benefits

ClearScale was able to meet Onstream's deadline and deployed the new cloud infrastructure in time for launch. With this new product backbone, Onstream is able to easily onboard new Onstream Mesosphere customers. The framework is fast and efficient and allows for low-latency data processing, which is critical in a high volume environment. Partnering with ClearScale to build the cloud infrastructure enabled Onstream to focus on their core mission of developing and delivering the Onstream framework.