Demo Script – Microservices Pipelines to the Edge

***<Title Slide>***

In this HPE Ezmeral demonstration, we’ll see how the HPE Ezmeral Data Fabric can be used to create a global data footprint, moving data around the world across data centers, cloud providers, and all the way out to the Tactical Edge.

***<Use Case Diagram>***

The use case we’ll be developing looks like this. We have a regional fixed-site data center, which might have a large HPE Ezmeral installation containing 10s, hundreds, or even thousands of nodes. In this example, that the HPE Ezmeral Data Fabric cluster will be executing a data processing pipeline, pulling in new imagery data and moving it through various stages of data enrichment. From there, we want to send the relevant data out to the soldiers in the field who are running a deployable HPE Ezmeral Data Fabric Edge system.

***<Edge Diagram>***

That Edge system might look like this, where we’ve installed the Data Fabric on a single HPE Edge Line 8000 unit allowing us to deploy 3 to 4 nodes of the Ezmeral Data Fabric in a compact ruggedized space. That’s enough hardware to run the complete HPE Ezmeral Data Fabric footprint, the same software that powers enormous data centers running at Exabyte scale. Any capabilities built to run in the cloud or on premise will run identically out here at the Edge and can compactly scale with the HPE Edge line products in any edge environment.

***<Return to Use Case Diagram>***

Often in these edge environments we are limited on bandwidth, we need to be very selective about what information consumes the limited communication resources available. So initially, we are going to broadcast announcements of new data as it becomes available, and have the soldiers indicate which of those assets they want delivered. But these requests are going to be forwarded from the fixed site data center into a cloud-based HPE Ezmeral Data Fabric environment, which will be collecting information from this site as well as sites from all over the world. With this global view of data including who is requesting information from what location, data scientists will be able to build advanced analytics and recommendation engines, so that we can smartly and proactively deliver the right information to the right missions. This is all utilizing open APIs and tools in conjunction with multi-modal replication technology in a global namespace built into the HPE Ezmeral Data Fabric. You could easily extend this picture into more complex topologies using multiple clouds, data centers, and edge devices allowing for effortless scale and survivability in even the most trying of situations.

***<Demo Screen>***

Let’s start building this data fabric. On the left we have a dashboard showing the data processing pipeline that’s running at the fixed-site data center. Each tile represents a new image being processed by a specific stage in the pipeline. The entire pipeline is built on the HPE Ezmeral Data Fabric based on the MapR Data Platform using a Microservices-based architecture. The image files themselves are being stored in the Ezmeral file system. The image metadata is being stored in the Ezmeral Database. And the Ezmeral Event Store for Apache Kafka, is the pub/sub streaming system that is used to tie all the stages of the pipeline together, sending real-time updates to a dashboard anywhere in the world it needs to be.

On the right side of the screen is the field dashboard that the forward-deployed soldiers are seeing. That dashboard is connected to the local EDGE system, but the data hasn’t been transferred yet. It’s all still sitting at the fixed site data center…..

So let’s fix that, and get these broadcasts out into the field. In this service application, the mission requires two-way data replication, such that we can send announcements out to the edge, and users can send requests for information back.

***<left-side: MCS, right-side: edge dashboard>***

Setting this up in Ezmeral Data Fabric is pretty simple. Here’s the Data Fabric Control System. We simply tell Ezmeral which data stream we want to replicate from the fixed site, and where we want it to go (which in this case is to the Edge1 cluster). We’ve also selected multi-master, so that the Edge users can send their information requests back using the same channel. We complete the action by clicking “Add Replica”.

In a matter of seconds, the Data Fabric will create the downstream replica, configure each site to replicate to each other, and send the data currently sitting in the stream out over the wire to the Edge. This did not require new software stacks, new servers or new storage complications. Point. Click. Done.

Watching the dashboard on the right, we see that replication has been established, and the notifications have started coming in. Ok… So that was pretty easy…. You can send data from a single source to multiple sites, you can also chain sites together, so they propagate data from one area of the world to the next. If you replicate all the way back to the source, HPE’s Ezmeral Global Data Fabric has built in loop-detection, to make sure you’re not duplicating your data.

In this corner, I used another pretty neat feature called Streaming Audits to show that replication was established. The Ezmeral Data Fabric can audit system events a real time, publish those event, and you can subscribe to those feeds, integrating that system information into your own dashboards and applications.

I configured the replication to be active/active, or bidirectional. So when I click on any of these Request links out at the Edge, those requests are automatically sent back to the fixed site data center. Looking at the fixed site dashboard on the left, we see the various Asset Requests popping up on the grid and being serviced.

Ezmeral data replication is very resilient, and it’s designed to work with low bandwidth and intermittent connections. I’m now going to simulate a catastrophic network failure completely disconnecting my forward-deployed Edge unit.

When I do that, we see that the Edge dashboard stops receiving messages, even though we can see on the left side that new data is still coming in to our data center.

Now,…

Although they aren’t being replicated, they are continually being added to the outbound message queue. Similarly, at the Edge, Users can continue to make Requests, and they’ll just queue up at the Edge. When communications is reestablished, we’ll see that the messages once again start to flow across the fabric, as soon as the restored connection is detected.

You can also pause the replication intentionally by user interaction or through automation. ***<click>*** You may want to do this when you want to reserve all your available bandwidth for higher-priority data. With the HPE Ezmeral Data Fabric, it’s not all or nothing – you can set up different data streams with different message priorities, and you can control them independently. Here we see the same behavior as before – the messages are queueing up locally but are NOT being sent over the wire. When we’re ready or when conditions are optimal, we can simply resume replication. ***<click>***

Another aspect that we see here, is how one can access all functionality not only through the graphical control system and the command line, but also programmatically using standard REST APIs.

***<left-side: MCS, right-side: Edge>***

You may have noticed that the users at the Edge still haven’t seen any actual images. To demonstrate additional capabilities, we will replicate these files from the fixed-site data center to the Edge using another option, I’m going to set up a Volume *Mirror*. A mirror allows you to replicate specific directories to 1 or more remote sites. The process to create one is very similar to the one we used to set up Streaming replication. Here, back in the Ezmeral Data Fabric Control System, we create a new mirror volume, specifying that the source of this data is a folder of images at the fixed site data center. We can set additional controls on the volume, such as encryption and auditing, and when we are done, we click Create Volume. ***<click>*** Once created,we initiate the mirroring operation by selecting **Start Mirroring**.

Having done that, we’ll soon see the images start to appear in the Edge Dashboard. Unlike Streams and Database replication, where new data is propagated as soon as it arrives, volume mirroring take place on a set schedule. Here, incremental changes to any data has been set to be delivered every 15 minutes.

After the initial mirror operation completes, incremental updates will occur periodically. The UI allows for increments of 1 minute in adjusting your mirroring frequency.

This gives us the flexibility to establish what data pipelines need immediate propagation of information and what data can be relegated to times that don’t compete with mission critical communication flow.

Having established our mirror, we now see the requested images start to appear in the Edge dashboard.

***<Cloud>***

The next piece we’re going to tackle is sending all of these requests for information to a third location in our HPE Ezmeral Data Fabric that is running in the cloud. Data scientists will have immediate visibility into what data is being requested, all over the world. With that comprehensive data set, they’ll be able to build much more advanced analytical models that enable us to proactively push data out into the field, tailored for each soldier, each mission and each location. To do this, we’re going to go back to the Ezmeral Data Fabric Control System formerly MapR. Notice that the AWS cluster appears in our drop-down list of possible destinations. Now, we just have to replicate the appropriate data stream to that destination.

On the right, We’ve got a standard terminal window open, showing the contents of our new data set.

When we create the new replica, in this example from from the command line, we’ll see the new information requests being seamlessly sent into the cloud, ready to be analyzed.

At this point, we’ve shown how to deploy The HPE Ezmeral Data Fabric to successfully create a global data footprint that extends from the Edge, to a Fixed Core Site, and into the Cloud. We saw data represented in both files and streaming formats being transported with synchronous and asynchronous methods across a mutli-site footprint. Additionally, database records can also be replicated in an active-active fashion using the exact same process we saw earlier. This is because the Ezmeral Data Fabric is a fully integrated data scape, and *not* just a collection of disparate technologies. Connecting to a fixed site data center, a simple Linux command shows all of the connected locations in the global data namespace.

We can drill into any of those to see the data at that particular geographic location.

And within that location, We have an integrated view into all of the files, database records, and data streams, that make up the data pipeline used to impact the missions represented in this demonstration. This level of architectural integration on a global, edge to core to cloud data fabric is not available anywhere else, and it’s is a key reason HPE Ezmeral Data Fabric is able to scale these capabilities across the world

This is what a Global Data Fabric is supposed to be.

This is Ezmeral.

This is HPE.

Thanks for watching, let us know how we can help.