

Akridata and Dell technologies Frequently Asked Questions January 2021

What is Akridata?

Akridata provides the only Edge Data Platform for the Autonomous World.

It solves the exascale-class data problem that is common when AI go beyond automation and towards autonomy data at the edge is applied to continuous learning. The Akridata solution can reduce time to access the right data by 10X and improve data scientist productivity by 2X as it improves infrastructure efficiency by 4X.

What is Data Ops? Is it just another word for data management?

Data Ops is the discipline at the intersection of data management, data processing, and data communication. It combines data, storage, and workflow management with a focus on complex data, data volume, and use by AI/ML algorithms.

We often hear Edge-Core-Cloud but in many cases core is really the cloud. Could you elaborate? Is there a typical distribution of data centers locally, regionally, centrally, or in the clouds? How do you see Dell within this model?

The answer to percentage between core data centers (DCs) and cloud depends on the customer. We usually find a mix, with the core DCs preferred for reasons of control, cost and performance, and cloud DCs preferred for reasons of shared access and more familiar dev environments for data scientists. Akridata supports all major clouds (AWS, Azure, GCP) and is unique in supporting configurations that are all core, or a hybrid of core and cloud.

Is there a model to find the tipping point where the cost of collecting and storing more data exceeds the value of the latent information?

You are bringing up a key underlying point but note that the notion of "value of latent info" is use-case specific and evolves as the models mature. Akridata provides the flexibility for customers to specify what different teams consider valuable (at any given time) and efficiently extract/prioritize that subset of the data. We do have statistical algorithms that identify and emphasize the most diverse examples in the data.

“

We are seeing excellent customer reception to the combined Akridata-Dell solution since it allows customers to ideally configure their edge-to-core-to-cloud AI environment based on Akridata's edge data platform and Dell's comprehensive infrastructure offerings at many price points, all wrapped into a seamless solution by various Dell software and professional services.

”

One of your slides mentions \$100M storage cost. Where was that in the end-to-end system configuration?

AI data sets continue to grow and are more and more decentralized. This is why we see the data problem as an exascale-class problem. The \$100M reference was actual to in-core DC storage at one of the customers.

For Edge opportunities, do you start with the data or with the workload?

Our customers usually think of data and workflows in an integrated fashion. Workflows are typically broken up into “standard” workflows (something akin to data pipelines, which define common prep/tagging that happens on all incoming sensor streams) and “use-specific workflows”, which are triggered by specific data scientist groups depending on the models they are training/refining.

How do you position Akridata and the cloud data platform Snowflake?

Akridata is designed for the world of edge-to-core-to-cloud. What does this mean? It means an open ecosystem that brings together multiple tools to achieve the end-to-end solution. It leads to a blend of data + workflow + orchestration that integrates with existing execution services, storage, and databases. And because a lot of that is motivated by AI, it also leads to the seamless handling of complex unstructured data (images, videos, objects).

Snowflake is billed as a proprietary cloud data warehouse, so it addresses different needs. It's apparent success shows that the data problem is significant and the AI data problem is even more so. It shows that the market opportunity for the Dell-Akridata solution is even bigger.

Which K8s engines are you using in cloud? Running own clusters, Anthos, other?

We are fairly flexible regarding the K8s engines. On the clouds, we can use the native ones (e.g., AWS EKS) or deploy our own K8s cluster. On prem, we deploy our own.

Where do the "Operators" belong? Is it a customer responsibility or is it a service from Akridata/Dell?

Akridata's existing deployments have customer-provided operators, but we are in the middle of a bid (together with Dell) where the ask is for offering the edge-related functionality as a managed service.

Is VMware used in any of the solution?

Not explicitly, but if customers have a VMware environment in their core DCs, that would be a good platform on which to deploy the Akridata functionality. We have a notion of a “soft edge cluster”, which has the ability to scale up and down based on dynamic demand. This elastic scaling would be a good fit to the VMware virtualized environment. On the edge, most of our deployments today are either on bare metal or container environments.

VMware is a very common part of customers' environments in general, and especially those who have multiple data centers in multiple continents. Typically, AI sensor data is very large to move and very important, so engineers connect remotely to run simulations where the data is. There is an active Dell-Akridata opportunity in EMEA where VMware would help with urgent data access.

Are containers/K8s the "secret" sauce to portability?

Containers (Docker images) is one aspect of portability. But we also support modules written in standard languages/frameworks (Java, Python, TensorFlow, PyTorch) and handle the dependency management ourselves. Basically, our philosophy is to support whatever customers already use rather than force them to rewrite their processing.

If several users require the same data from the same edge for their different ML work, wouldn't it end in this case with copying multiple of the same data in the core, or does the core have some intelligence using some identification of same data?

This is one of the key value propositions of the Akridata platform. Each data object created in the system has a globally unique ID and we manage the efficient transfer/storage and access of data using these IDs. In the scenario you bring up, the system uses reference counts and other mechanisms to avoid redundant transfers.

The Akridata platform also tracks lineage and provenance. Each data object and meta data record is uniquely identifiable, version controlled, and logged/audited with regards to operations against it. This is of course implemented in a very efficient and scalable fashion such that it can handle 10s to 100s of PBs of data and billions of metadata records.

Have you encountered AWS Outposts with any of your customers?

Not yet. Today, these deployments tend to be in non-traditional DC environments (e.g., garages, depots, etc.) where there is value from ruggedized infrastructure. In the core DCs, our customers tend to gravitate to more traditional compute and storage solutions.

Besides the automotive industry, what other use cases are you pursuing? For example, what about Healthcare Digital Twin in management of chronic disease?

We are starting discussions in the manufacturing and smart retail verticals, which have similar complex data + AI models. We have also had discussions with medical imaging use cases, which is related to the point you bring up. In general, healthcare seems like a suitable use case. We also see some interest from supercomputing sites with large scale data problems.

Where can I go to get more help if my customer could use the Akridata solution?

Akridata is working very closely with Dell around the world and is ready to get engaged. Just let us know! Please also check the special [Akridata.com/dell](https://akridata.com/dell) page for additional info.