DATERA

POWERING DATA LAKES WITH DATERA ELASTIC DATA FABRIC.

DATA LAKES

Within the last decade, there has been an explosion of tools and technologies used to manage data and the analytics process used to transform "big data" into business value. The term Data Lakes is used to describe the next state for big data analytic environments.
A Data Lake is a concept in which infrastructure has the flexibility to manage data sets and tools in a singular environment. As more companies proceed to grow their big data infrastructure and adopt Data Lakes, they face a number of new challenges: multi-tenancy, resource utilization and diverse performance needs, all while trying to control cost.

Datera Elastic Data Fabric (EDF) is the most advanced block storage technology that can replace the use of traditional, locally attached, or centralized storage offerings. n much the same way Amazon offers the EBS platform for all types of data storage in the AWS cloud, Datera EDF can bring block storage to your on premises cloud or Data Lake implementation at a more cost effective and manageable manner.

Datera offers the only unique data fabric that can be used to power large scale operations. By leveraging scale out technology, Datera scales horizontally, yet linearly in performance, supporting a wide breadth of cloud native technologies. As the needs of the environments grow, the Datera EDF can expand effortlessly, adding more performance and capacity in a linear, inexpensive fashion. Datera also offers multi-tenant features needed for cloud scale, which provides the ability to logically and physically separate workloads and tenants from each other. This feature set is extremely valuable for analytic environments, especially when coupled with Datera's heterogeneous nodes types. Each node type has varying capabilities: large capacity, low cost per GB hybrid, or high performance all flash nodes. This ability
to blend different types of nodes into a single control plane allows the Datera EDF to underpin the entire analytics environment.

DATERA BENEFITS:

True Scale for Scale Model: Start small and be able to scale fast. Traditional storage is limited in scale by the proprietary hardware frame size and performance, while Datera delivers heterogeneous scale-out based on commodity off-the-shelf technology. Software-based systems will organically evolve as they tend to grow with new hardware, and they need to decommission obsolete old hardware without any disruption. With Datera EDF, data gets rebalanced and access optimized for the runtime workloads. No data migration. No forklift upgrades.

Wide Price/Performance aperture: Given the wide variety of cloud applications, Datera provides the elastic price bands that fit the economic value for particular data. Each volume's requirements are composable on-demand and that articulates storage placement. The Datera system allows for storage to be placed intelligently maintaining the intended service level objectives for any application workload, even in multi-tenant clouds.

Workload Isolation: With multiple tenants and varying types of applications in the cloud, Datera provides the the ability to manage performance at application or volume-basis through an array of configuration options that enable segmented and deterministic performance, such as QoS controls to addressing the noisy neighbor problem. The system is deeply multi-tenant aware with resource segmentation and workload isolation natively across the architecture.

Transparent DevOps to IT Ops Transition: Datera is designed for application related storage provisioning to be decoupled from the physical infrastructure management. Moreover, customers can seamlessly transition from development to test to IT operations, without application configuration changes, but morphing the infrastructure constraints through policy overrides based on application, user, or tenant.

Automated operations Model: Datera provides application-driven real-time resource consumption, evolving data centers from slow infrastructure-centric IT to agile DevOps-centric IT. Application requirements are captured in context-aware policies ("intents") that deeply shape the fabric, accommodating failure domains, network topology, multi-tenancy, workload isolation, etc. An intent-based REST API makes the infrastructure programmable and composable, and allows instantiating entire storage clouds with single API calls (datacenter-as-code). Datera transforms datacenter speed, agility, and experience.

Datera EDF provides the flexibility to create a Data Lake. Building a Data Lake requires a number of different components all working together. The Datera technical solution guide for Kafka can be used as a technical reference for deploying this important component of the analytics pipeline. For Hadoop and Spark deployments see the Datera technical solution guide for Hadoop.