

# CLOUDERA SDX

Built for multi-disciplinary analytics | Optimized for cloud

### Cloudera SDX Benefits

**INCREASED**

- \_ Business insights
- \_ Business agility
- \_ Governance capability
- \_ User self-service

**DESCREASED**

- \_ Operational costs
- \_ IT staff overhead
- \_ Security risks
- \_ Data redundancy

Spotting and preventing cyber-attacks in real-time. Delivering more effective treatment to patients. Keeping factories, fleets, and financial markets running. The highest-value use cases today rely on a variety of data analysis capabilities working in concert. The Cloudera Enterprise platform delivers them all:

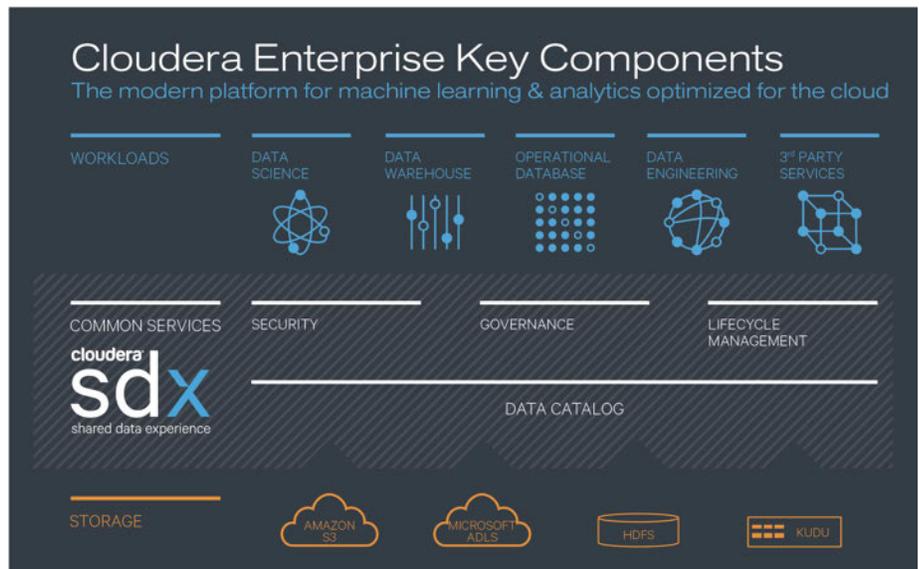
- \_ Data Engineering to speed ETL processes and train machine learning models
- \_ Data Science to enable exploratory workloads at big data scale
- \_ Data Warehouse for high-performance, high-concurrency SQL
- \_ Operational Database to power data-driven applications that deliver near real-time insights
- \_ Extensible services that integrate with third-party applications

Most business applications actually require a combination of two or more of the above functions to operate against the same data set. This is a challenge in traditional analytic environments where each of these functions exists within an application silo that comes with its own distinct security, governance, management, data lifecycle policies and separate data catalog. It's even tougher in transient cloud deployments where data context – data about the data – does not persist and may be lost when a cluster is terminated.

“SDX includes comprehensive, granular security to protect against cyber threats, and unified governance for the audit and search capabilities that the modern world demands, especially with standards like PCI-DSS and GDPR.” Mike Olson, Chief Strategy Officer and Co-founder, Cloudera

### Delivering a Shared Data Experience

Cloudera SDX (shared data experience) is a powerful software framework that makes multi-disciplinary analytics easier to develop, quicker to deploy, more cost-effective, and more secure. By applying stateful, centralized, consistent data context services that reside with the persistent object storage, not the transient compute nodes, SDX enables hundreds of different workloads to run against shared or overlapping sets of data.



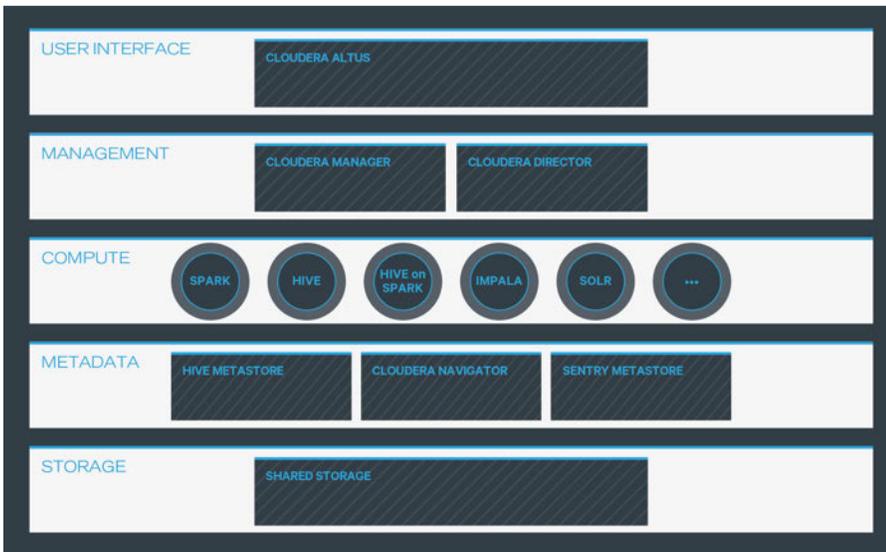
SDX is a core element of the Cloudera Enterprise platform today and enables diverse workloads to act as a singular unit. It is comprised of five discrete functions that together solve a really hard problem — providing a shared data experience for a platform that supports a diverse set of workloads and user interactions models.

FUNCTION	CUSTOMER BENEFIT
Shared catalog	Provide a common catalog of schema and lineage metadata to each workload and user accessing the platform for maximum efficiency and productivity
Shared security	Implement unified and consistent, granular authentication, authorization, encryption, and compliance controls
Shared governance	Easily discover new data, understand where that data originated, and track how it has been modified
Shared data lifecycle	Aggregate a single copy of all data, provide disaster recovery, and manage the full data lifecycle from ingestion to purge

## SDX in the cloud

Cloudera SDX makes it easier for customers to deploy multi-disciplinary analytics on cloud infrastructure without sacrificing the shared data experience of on-premises deployments. SDX is even more valuable in the cloud because workloads tend to run on isolated infrastructure and are often transient in nature. Entire data context should be automatically supplied and harvested for each workload.

Without SDX, each workload degenerates into a silo of isolated security policies and metadata context that becomes a challenge for the data and IT teams to manage. SDX makes it possible to create a single logical cluster that supports multi-disciplinary analytics and simultaneously allows each workload to take full advantage of cloud infrastructure.



The diagram above highlights the details of how SDX works in the cloud. Starting from the bottom upwards:

- \_ The storage layer is implemented via Shared Object Storage and requires only a single copy of raw data to implement SDX in order to maximize efficiency, security, and governance
- \_ The metadata layer is implemented through a shared metastore and related tools that maintain consistent data catalog and policies across the platform
- \_ The compute layer is implemented by running each workload in a dedicated cluster such that each workload can be fully optimized for cloud infrastructure as a service (IaaS)
- \_ The management layer is implemented via Cloudera Manager to make it easy to create and manage both transient and persistent workloads
- \_ The user interface layer is implemented in Cloudera Altus, making it easy for end users to create and troubleshoot their jobs in a shared environment managed by the data team

Cloudera SDX supports multiple public cloud, private cloud and bare metal configurations. The service is embedded into the following commercial offerings of Cloudera Enterprise:

- \_ Cloudera Data Science and Data Engineering
- \_ Cloudera Operational Database
- \_ Cloudera Data Warehouse
- \_ Cloudera Enterprise Data Hub
- \_ Cloudera Altus Data Engineering
- \_ Cloudera Altus Data Warehouse