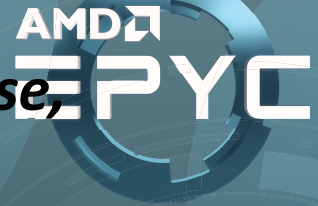


AMD and Cloudera®: Big Data Analytics for On-Premise, Cloud and Hybrid Deployments



August, 2018

Standards Based

AMD is committed to industry standards, offering you a choice in x86 architecture with design innovations that target the evolving needs of modern datacenters.

High Density, Low Cost

Compute requirements are increasing, datacenter space is not. AMD's EPYC processor offers high core density with full access to all features. Innovative architecture means outstanding performance at a low cost.

Partner Ecosystem

AMD's broad partner ecosystem and collaborative engineering provide tested and validated solutions that help lower your risk and your total cost of ownership

Cloudera Enterprise Data Hub: A Complete Multi-Discipline Big Data Analytics Solution

Cloudera Enterprise Data Hub answers an organization's most valuable questions by integrating data warehousing, data science, data engineering, search, streaming and real-time analytics into a unified platform with a single source of truth.

Cloudera SDX: Shared Data Experience for On-Premise, Cloud and Hybrid Deployments

Cloudera Shared Data Experience (SDX) creates a seamless integration of all analytic disciplines with full security, governance and administration across any set of deployments: on-premise, cloud and hybrid

AMD EPYC is Cloudera Certified Technology

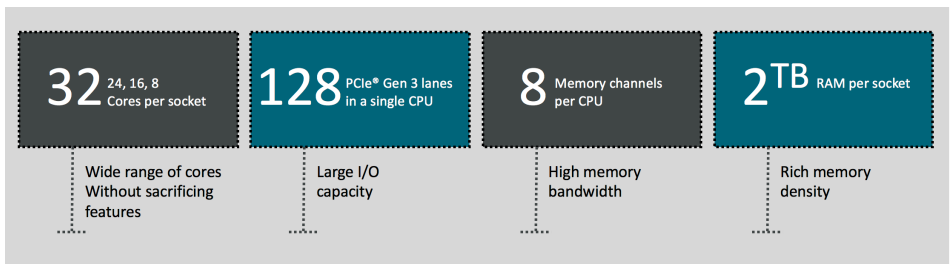
The [Cloudera Certified Technology](#) program is designed to make choosing the right technology easier. It means the product has been tested and validated to work with CDH – an enterprise-ready distribution of Apache Hadoop and related projects



AMD EPYC™ Processor Meets the Needs of Today's Datacenter Workloads

Designed from the ground up for a new generation of solutions, AMD EPYC implements a philosophy of choice without restriction. Choose the number of cores and sockets that meet your needs without sacrificing key features like memory and I/O.

Each EPYC system-on-chip (SoC) can have from 8 to 32 cores with access to incredible amounts of I/O and memory regardless of the number of cores in use, including 128 PCIe lanes, and access to 2 TB of high speed memory per socket.



EPYC's innovative architecture translates to terrific performance at a low cost. More importantly, the performance you're paying for is appropriate to the performance you need.

Storage intensive workloads can utilize the plentiful I/O bandwidth with the right number of cores (avoiding overpaying for unneeded power), while compute-intensive workloads can make use of fully loaded core counts, dual sockets and plenty of memory.

AMD EPYC processors help enable more performance, flexibility, and security

PERFORMANCE. The AMD EPYC SoC brings a new balance to the datacenter. Utilizing an x86-architecture, the AMD EPYC processor, brings together high core counts, large memory capacity, ample memory bandwidth and massive I/O with the right ratios to help performance reach new heights.

FLEXIBILITY. Match core count with application needs without compromising processor features. EPYC's balanced set of resources means more freedom to right-size the server configuration to the workload.

SECURITY. AMD EPYC features the industry's first dedicated security processor embedded in an x86-architecture server SoC. The processor manages secure boot, memory encryption, and secure virtualization on the SoC itself. Encryption keys never leave the processor where they can be exposed to intruders.

SCALABILITY. Scale-up or scale-out, AMD and its ecosystem partners offer high-performance network connectivity options for applications at massive scale.

AMD EPYC for Big Data Analytics

The advent of big data revolutionized analytics and data science by allowing enterprises to store, access and analyze massive amounts of data of almost any type from any source. This created the need for a new generation of data processing architecture; one where all the data is stored in a vast, ever-expanding data lake, then accessed by various applications using new technologies appropriate to the need.

The EPYC Advantage: Outstanding performance and scalability – both scale-up and scale-out. AMD and its ecosystem partners offer high-performance network connectivity options for massive scale-out systems such as Apache™ Hadoop®.

These needs included traditional historical analysis, now done with the distributed processing methods of big data technologies, as well as integration with existing enterprise systems now able to take advantage of this new source of business insight.

In addition, new capabilities came online in the form of real-time streaming analytics, and ad hoc, interactive analysis on enormous datasets. Click stream analysis, log analytics, recommendation engines are just a few examples of the new generation of applications now in wide use. Some of these applications need massive data injections and high-speed I/O, while others need large amounts of memory - all need the capabilities enabled in the underlying hardware and the software.

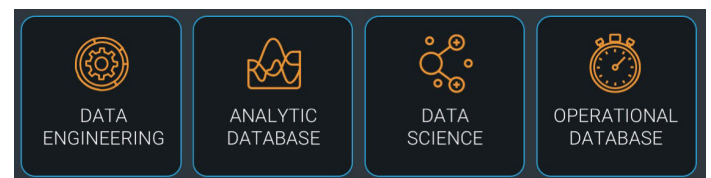
The EPYC Advantage: Performance - The AMD EPYC SoC brings new balance to the datacenter. The highest core count yet in an AMD x86-architecture server processor, large memory capacity, memory bandwidth and I/O density are all brought together with the right ratios to help performance reach new heights.

AMD has partnered with leading software and server vendors to create certified reference architectures for customers to demonstrate strong performance, lower risk and to help reduce implementation cost.

Cloudera Enterprise

Cloudera is a thought leader in the creation and application of enterprise-wide big data analytics solutions enabling organizations to answer their most valuable questions. Cloudera Enterprise Data Hub brings together data warehousing, data science, data engineering, search, streaming and real-time analytics into a unified platform with a single source of truth.

Organizations now run diverse, multidisciplinary big data workloads that span analytic databases, operational databases, data engineering applications, and data science applications. Many of these workloads operate on the same underlying data.



Recognizing that most business applications require a combination of core functions to solve real-world problems, Cloudera identified then integrated these functions enabling them to operate on the same data set.

Data Engineering enables the business to run batch or stream processes that speed ETL and train machine learning models.

Data Science enables the business to do exploratory data science at big data scale with full data security and governance.

Analytic Database delivers fast time-to-insight with the flexibility and agility to run in any environment and against any type of data.

Operational Database enables the business to build data-driven applications that deliver near real-time insights.

This integration solves a thorny problem, but it exists within the context of a larger issue. Compute has moved to the data and is stateless, either transient or long-running, but the data may be on-premise or in the cloud, or both.

Further, the meta-data that describes the data: table definitions, governance artifacts, metadata classifications, access permissions, and business definitions, often called the ‘data context,’ is difficult to keep consistent across the increasing number of workloads and data locations.

Cloudera SDX: Shared Data Experience

Cloudera SDX offers a modular software framework that ensures a shared data experience across all deployment types, including multiple public clouds, private cloud, hybrid cloud, and bare metal configurations. SDX applies stateful, centralized, and consistent data context services making it possible for hundreds of different workloads to run against shared or overlapping sets of data. SDX makes multi-disciplinary data applications easier to develop, less expensive to deploy and increasingly important in today’s environment, more consistently secure.

SDX 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 interaction models.

Shared Governance provides the ability to govern the data in a unified manner so users can easily discover new data, understand where that data came from, and track how it has been modified.

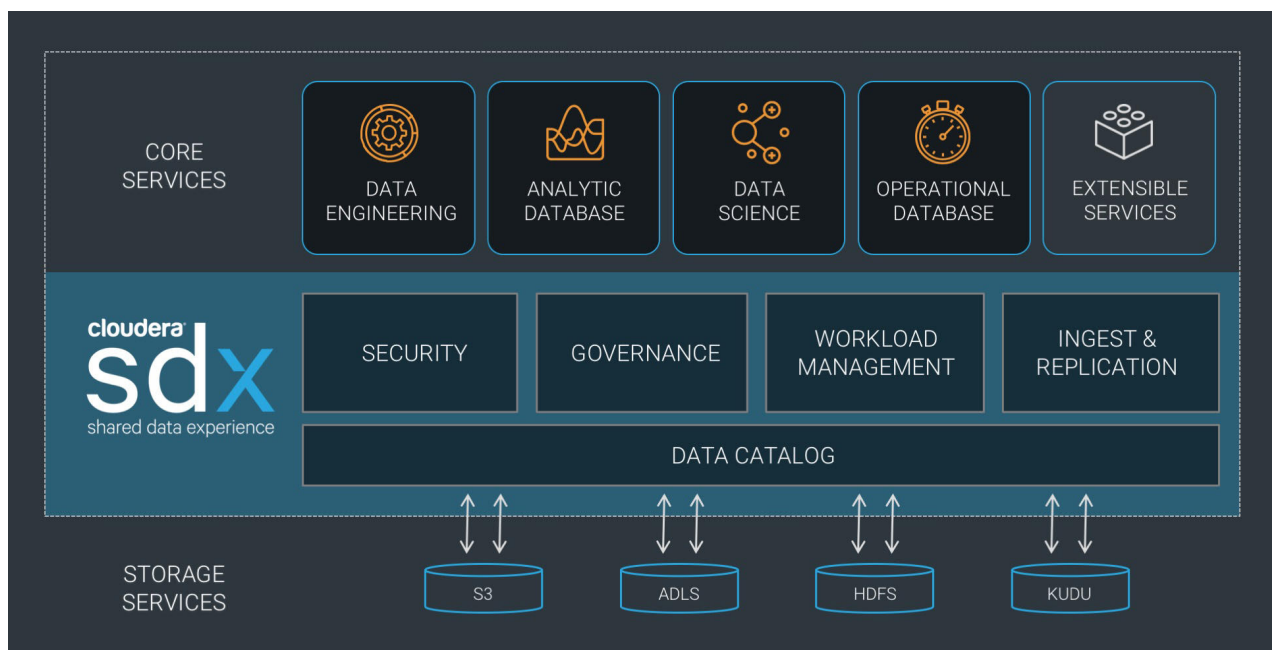
Shared Security implements consistent, granular authentication, authorization, encryption, and compliance controls in a unified manner across the entire platform.

The EPYC Advantage: Security - AMD EPYC features the industry’s first dedicated security processor embedded in an x86-architecture server SoC. The processor manages secure boot, memory encryption, and secure virtualization on the SoC itself. Encryption keys never leave the system where they can be exposed to intruders.

Shared Workload Management enables administrators to create, manage and optimize workloads individually or as a collection and to allocate resources and assign workload priority based on business requirements.

Shared Ingest & Replication provide the ability to ingest data once and make it available to all applications and users without additional ingest pipelines or copies of data, and to replicate data on demand to remote locations or directly to the cloud.

Shared Data Catalog provides a common catalog of schema and lineage metadata to each workload and user accessing the platform for maximum efficiency and productivity.



Reference Architecture for AMD EPYC and Cloudera Enterprise

The reference architectures for AMD EPYC processors and Cloudera Enterprise provide options for the performance and scalability requirements needed to maximize the investment in Big Data Analytics.

AMD EPYC's ability to provide a no compromise single-socket solution ensures you are only paying for the processing power the application needs. A single-socket server has all of the I/O and memory bandwidth available in a dual-socket server without the extra cost

For high-performance applications, AMD EPYC-based dual-socket servers offer phenomenal core density.

The EPYC Advantage: Flexibility - Match core count with application needs without compromising processor features. EPYC's balanced set of resources means more freedom to right-size the server configuration to the workload.

These reference architectures offer the foundation for creating custom configurations that meet unique application demands with the various storage and connectivity options to choose from our leading server OEM partners.

No Compromise Single-Socket		Versatile Dual-Socket	
Management Nodes: 3		Management Nodes: 3	
CPU	1 x AMD EPYC 7601, 7551, or 7501*	CPU	2 x AMD EPYC 7451 or 7401*
Cores/Threads	32/64	Cores/Threads	24/48 per CPU, 48/96 per node
Memory	256GB	Memory	256GB
Network	10/25/40/100 Gigabit Ethernet	Network	10/25/40/100 Gigabit Ethernet
Storage: OS	2 x SSDs or 2 HDDs	Storage: OS	2 x SSDs or 2 HDDs
Storage: Data	2 x NVMe drives, or 4 x SAS/SATA SSDs, or 8 x SAS/SATA HDDs	Storage: Data	2 x NVMe drives, or 4 x SAS/SATA SSDs, or 8 x SAS/SATA HDDs
Data Nodes: Minimum 3, scalable to hundreds		Data Nodes: Minimum 3, scalable to hundreds	
CPU	1 x AMD EPYC 7601, 7551, or 7501*	CPU	2 x AMD EPYC 7451 or 7401*
Cores/Threads	32/64	Cores/Threads	24/48 per CPU, 48/96 total
Memory	128/256 GB (supports up to 2 TB)	Memory	128/256 GB (supports up to 4 TB)
Network	10/25/40/100 Gigabit Ethernet	Network	10/25/40/100 Gigabit Ethernet
Storage: OS	2 x SSDs or 2 x HDDs	Storage: OS	2 x SSDs or 2 x HDDs
Storage: Data High Performance Performance Capacity	4-8 x NVMe drives or 8-12 x SAS/SATA SSDs 24 x SAS/SATA HDDs (SFF) 12-16 x SAS/SATA HDDs (LFF)	Storage: Data High Performance Performance Capacity	4-8xNVMe drives or 8-12 SAS/SATA SSDs 24 x SAS/SATA HDDs (SFF) 12-16 x SAS/SATA HDDs (LFF)

*other AMD EPYC 7000 series processors are also supported

Conclusion

Versatility and agility are among the most important requirements in modern datacenters. The AMD EPYC system-on-a-chip (SoC) enables organizations to deploy systems that precisely meet today's needs while positioning themselves for tomorrow's requirements.

Cloudera delivers on the promise of complete integration of different analytic disciplines for big data: operations, data science, data engineering and machine learning, all optimized for on-premise, cloud and hybrid deployments.

Together, AMD and Cloudera empower the development of modern data applications that implement solutions to a diverse set of business problems with high-performance processing and cost-effective solutions that are perfectly sized for current needs and easily scalable as your business grows.

For more information about AMD's EPYC line of processors visit: <http://www.amd.com/epyc>

For more information about Cloudera visit: <http://www.cloudera.com>.

For more information about Cloudera Certified Technology program visit: <https://www.cloudera.com/partners/partners-listing.html>



©2018 Advanced Micro Devices, Inc. All rights reserved. AMD, the AMD Arrow logo, EPYC and combinations thereof are trademarks of Advanced Micro Devices, Inc. Cloudera and the Cloudera logo are registered trademarks or registered trademarks of Cloudera, Inc. in the United States and other countries. Apache and Apache Hadoop are either registered trademarks or trademarks of the Apache Software Foundation in the United States and other countries. Other product names used in this publication are for identification purposes only and may be trademarks of their respective companies.

The information presented in this document is for informational purposes only and may contain technical inaccuracies, omissions and typographical errors. AMD reserves the right to revise this information and to make changes from time to time to the content hereof without obligation of AMD to notify any person of such revisions or changes.