CLOUDERA | AMD

Innovating for Tomorrow by Solving Today's Analytics Challenges

Table of Contents

Introduction:
A Joint Approach to Modern Analytics Challenges

Today, we're operating in a dynamic data world. According to IBM, on average every person generated 1.7 megabytes of data per second in 2020. Internet users alone generate 2.5 quintillion bytes of data every day.[1]

Businesses realize that all of this data creates unprecedented opportunities. Since 2012, there has been a huge increase in the adoption of big data initiatives. In a 2020 survey of mainstream companies, 99% have made big data investments and 65% employ a Chief Data Officer (CDO).[2]

But they also realize that with unprecedented opportunities come unprecedented challenges. In that same decade, only 39.3% are effectively managing data as an asset; only 24% say they've created a data-driven organization; and only 24.4% have forged an effective data culture within their

organizations.[2] Many of these challenges stem from the inability to find the right data analytics architecture—on-premises, public or private cloud, or hybrid cloud. There are a number of factors that go into this decision:

Performance

Optimizing performance on-premises and in the cloud to analyze—and get the most out of—the zettabytes of data available today.

Security

Protecting data through stronger security and governance as more sensitive data and workloads are moved to the cloud.

Sustainability

Helping organizations avoid too many disparate/redundant solutions and that data centers are expanded in an environmentally-conscious way, as the organization grows.

Cloudera Data Platform (CDP) running on AMD EPYC™ CPU-based servers empowers organizations to implement big data analytics solutions that scale as businesses grow. The technical partnership between Cloudera and AMD delivers a big data solution that takes a proactive approach to addressing these concerns.

Chapter 1: Performance

The Right Platform + The Right Hardware

5000% **increase** in data interactions between 2010 and 2020[3]

79 **zettabytes** of data generated worldwide in 2021[4]

>150 **zettabytes** of big data will need analysis by 2025[5]

Analyzing petabytes of data in real-time—and having the ability to act on it quickly—has never been more critical. The technical partnership between Cloudera and AMD helps organizations achieve this goal by bringing together a world-class data platform and top-tier hardware in one place.

Cloudera's data and AI platform allows data practitioners to optimize their operations by analyzing vast amounts of data anywhere in its lifecycle. CDP is a hybrid data platform that allows organizations to analyze data wherever it lives—on premises, on public or private cloud, or in a hybrid cloud environment—managing and securing the big data lifecycle from edge to AI.

AMD EPYC™ processors are the highest-performing x86 server processors on the market[6], demonstrating consistently faster performance for cloud[7] and enterprise.[8]

CDP Hybrid Cloud Benefits

Extensible modern architecture with the flexibility to run any analytics on any private or public clouds and data, regardless of format with:

- Optimal performance, scale, and cost
- Single pane of glass visibility
- Universal security and governance

Optimize Data Analytics with Solutions Powered by AMD EPYC™

60%

Higher Performance
1x AMD EPYC™ 75F3 vs
2x Intel Xeon® 6262V

TPCx-HS@3TB

61%

Better Price/Performance
1x AMD EPYC™ 75F3 vs
2x Xeon® 6262V

TPCx-HS@3TB

Apache Hadoop Performance
Vs. Intel Processors[9]

Chapter 2: Security

Always-On Protection from Chip to Edge

In the world of big data analytics, data security is always a chief concern. As the amount of data increases, so do the risks of information theft, cyberattacks, ransomware, and other malicious activities. For one thing, this data explosion has led to less visibility over what data is being shared and transferred, where it's being transferred and how it's being transferred.

But visibility is just the tip of the security iceberg. A more connected, digitally-transformed organization has more endpoints that are potentially vulnerable to attack, creating an increased potential for threats across the entire network. Compliance regulations are ever-changing and geographically varied, which increases regulatory risk in complex hybrid IT ecosystems.

The bottom line is that more data created and living in more places means more vulnerabilities. With these challenges in mind, AMD and Cloudera deliver comprehensive security features and benefits at both the chip and software level.

AMD has a full suite of cutting edge security features built into the silicon to defend against internal and external threats. **AMD Infinity Guard**[10] goes beyond encrypting data while
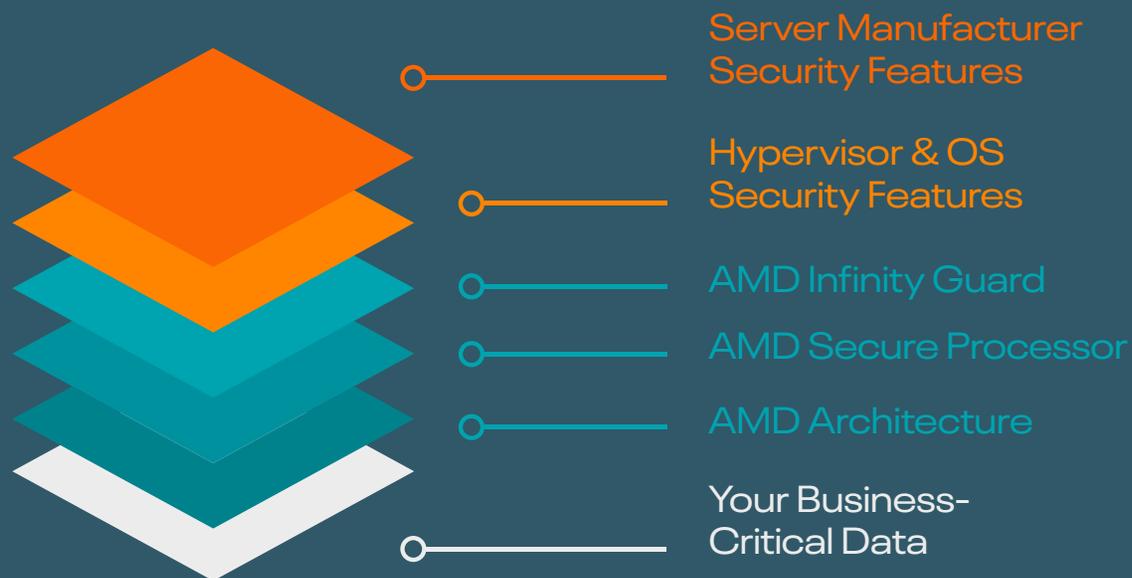
it is at rest and in transit. It encrypts in-use data, as it is being processed, with virtually no impact on system performance. And **Secure Encrypted Virtualization (SEV)** provides additional security to protect virtual machines with encryption keys known only to the processor itself.

Cloudera Shared Data Experience (SDX) provides both compliance and self-service data access for all users, with constant security and governance at the software level in public, private, and hybrid cloud environments.

The AMD Approach to Enterprise Security

AMD processor embedded security features help mitigate risk with a modern, multi-faceted approach to data center security.

Server Manufacturer
Security Features

Hypervisor & OS
Security Features

AMD Infinity Guard

AMD Secure Processor

AMD Architecture

Your Business-
Critical Data

Cloudera Shared Data Experience (SDX)

SDX is a fundamental part of Cloudera Cloud Platform architecture, making CDP secure by design with consistent policies everywhere.

- Robust Data Catalog

- Comprehensive security

- Enterprise-grade governance

- Full encryption

- Control across hybrid cloud

Chapter 3: Sustainability

Streamlined Solutions and a Cleaner Planet

Sustainability has two primary implications when it comes to big data analytics. For organizations, it means avoiding tech sprawl by consolidating disparate, redundant systems. For the planet, it means reducing the environmental footprint of data centers…which is quite significant.

Data centers account for **1% of annual global electricity demand** (200-250 TWh)[11]

Cryptocurrency mining **uses an additional 100 TWh** annually[12]

The Cloudera and AMD technical partnership takes on the challenges of sustainability and turns it into a core competitive and environmental advantage.

Cloudera Data Platform delivers everything data service organizations need in a single, secure platform. This reduces total cost of ownership (TCO) and reduces technical debt by eliminating the need for IT teams to deal with disparate point solutions. Open source software is the foundation of CDP, ensuring interoperability with a broad range of analytic and business intelligence applications. This helps minimize the expense and effort of connecting to organizations' existing IT infrastructures. And CDP's integrated data lifecycle eliminates the ongoing orchestration integration and development work to implement business use cases with individual data or analytics services from cloud providers.

AMD EPYC™ processors power the most energy-efficient x86 servers, delivering exceptional performance while lowering energy costs[13] through reduced heating and cooling requirements. AMD is focused on accelerating server energy efficiency, delivering high-performance computing (HPC) to help tackle the world's toughest challenges, and lowering data center TCO. AMD's goal is to deliver a **30x increase in energy efficiency** for AMD processors and accelerators powering servers for AI and HCP from 2020-2025, and a **97% reduction in energy use** per computation by 2025.[14]

Cloudera + AMD Reference Architecture

Cloudera Data Platform Private Cloud Base

Pod Network:
PowerSwith S5248F-ON series switch

Cluster Aggregation Network:
PowerSwitch Z9432F-ON series switch

Infrastructure Nodes:
AMD EPYC-based PowerEdge R6515

(3) Master nodes

(1) Utility node

(1) Edge node

(3+) Worker Nodes:
AMD EPYC-based PowerEdge R6515 (Config 1) or
AMD EPYC-based PowerEdge R7515 (Config 2)

GPU Accelerated Worker Node Option:
AMD EPYC-based PowerEdge R7525

HDFS:
Powerscale H5600 (Configuration 1)

or Additional Worker Nodes (Configuration 2)

Learn more about data management with
Cloudera Data Platform on AMD-powered
Dell Infrastructure.

CDP Data Center
Installable Software

Cloudera Manager

Bare Metal

CLOUDERA SDX

Physical Clusters

Data Centers

Storage

Cloudera Runtime

Configuration 1

Configuration 2

Independent Compute & Storage

Combined Compute & Storage

RECOMMENDED

AMD

DELL Technologies

Make Sense of Everything in Real-Time

The big data future is only getting bigger, and the pressure is real. But pressure can burst a pipe, or it can make a diamond. The ability to take on the challenges of performance, security, and sustainability simultaneously would be too much for any one solution—or provider. That's where the powerful technical partnership between Cloudera and AMD comes in.

AMD and Cloudera open a world of possibilities for organizations to accelerate the processing of real-time and high-volume data to stay agile in a dynamic data world. For cloud, IT, and data decision makers, AMD EPYC™ CPUs combined with Cloudera Data Platform deliver the right

hardware on the right data platform. We enable customers to create new insights from the wealth of data they have at their fingertips on a unified, high-performance platform with baked-in security from chip to edge. And with a commitment to sustainability that's not only good for business, but good for the planet.

Discover how Cloudera Data Platform and AMD EPYC™ CPUs can bring the sustainable, secure performance you need to implement a big data analytics solution that's ideal for today's needs...and scalable for tomorrow's.

Learn more at cloudera.com | US: +1 888 789 1488 | Outside the US: +1 650 362 0488

Sources

[1] Data Never Sleeps 5.0 https://www.domo.com/learn/infographic/data-never-sleeps-5

[2] New Vantage Partners Big Data and AI Executive Survey 2021

[3] Forbes, 54 Predictions About The State Of Data In 2021

[4] First Site Guide, Big Data Statistics 2022: How Much Data is in The World?

[5] Statistica, Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2025

[6] MLN-016B: Results as of 07/06/2021 using SPECrate®2017_int_base. The AMD EPYC 7763 scored 854, which is higher than all other 2P scores published on the SPEC® website. SPEC®, SPECrate® and SPEC CPU® are registered trademarks of the Standard Performance Evaluation Corporation.

[7] MLN-129: VMmark® 3.1 - vSAN™ comparisons based on best performing systems published by VMware. Configurations: 4-n, 2x AMD EPYC 7763 (39.01 score @ 40 tiles) versus 4-n, 2x Intel Xeon Platinum 8380 (24.26 @ 26 tiles). VMmark, vSAN, and Horizon are registered trademarks of VMware in the US or other countries.

[8] MLN-101: SAP® SD 2-tier comparison based on best performing systems published at www.sap.com/benchmarks. As of 5/5/2021, 2x AMD EPYC™ 7763 scored 75,000 benchmark users which supports 56% more benchmark users than the top "Ice Lake" 3rd Gen 2x Intel® Xeon® Platinum 8380 that scored 48,000 benchmark users. 2x AMD EPYC 7H12 scored 69,499 benchmark users. SAP and SAP logo are the trademarks or registered trademarks of SAP SE (or an SAP affiliate company) in Germany and in several other countries.

[9] MLN-069K: TPC Benchmark Express™ HSph@3TB comparison based on highest system results published as of 03/12/2021. Configurations: 17-node, 1x AMD EPYC 75F3 (34.52 HSph@3TB, $35,615.50/HSph@3TB, Framework 1 MapReduce) versus 17-node, 2x Intel Xeon Gold 6262V (21.52 HSph@3TB, $91,276.91/HSph@3TB, available 11/09/2020, Framework 2 Spark) for 60% higher [1.6x the] Hadoop Sort throughput performance at 61% better price/performance.

[10] GD-183: AMD Infinity Guard features vary by EPYC™ Processor generations. Infinity Guard security features must be enabled by server OEMs and/or Cloud Service Providers to operate. Check with your OEM or provider to confirm support of these features. Learn more about Infinity Guard.

[11] iea Data Centres and Data Transmission Networks Tracking Report, November 2021

[12] Capgemini Research Institute Sustainable IT Report

[13] EPYC-028: As of 2/2/22, of SPECpower_ssj® 2008 results published on SPEC's website, the 55 publications with the highest overall efficiency results were all powered by AMD EPYC processors. More information about SPEC® is available at http://www.spec.org. SPEC and SPECpower are registered trademarks of the Standard Performance Evaluation Corporation.

[14] Includes AMD high-performance CPU and GPU accelerators used for AI training and High-Performance Computing in a 4-Accelerator, CPU hosted configuration. Goal calculations are based on performance scores as measured by standard performance metrics (HPC: Linpack DGEMM kernel FLOPS with 4k matrix size. AI training: lower precision training-focused floating-point math GEMM kernels such as FP16 or BF16 FLOPS operating on 4k matrices) divided by the rated power consumption of a representative accelerated compute node including the CPU host + memory, and 4 GPU accelerators.

Privacy Policy | Terms of Service

CLOUDERA