

# Intel, Cloudera, and SAP Collaborate to Improve Business Intelligence and Accelerate Predictive Analytics

A joint SAP-Cloudera solution running on Intel® hardware can improve decision-making, empower data analysts, and produce deep insights.



Data-driven organizations reap the value of analytics extracted from the blend of business data (CRM, ERP) and Big Data (sensor data, log files, images, free text, video, and social media). Technologies such as SAP HANA and Cloudera Enterprise, operating on Intel® Architecture enable these organizations to deliver new insights that drive action. For example:

- **Predictive maintenance.** Combining production, operations, maintenance, and service data with sensor and IoT data from devices in the field to help manufacturing companies obtain better visibility of machine health and avoid costly downtime.
- **Next-best offer.** Blending transaction and web browsing activities to help retailers gain a more complete view of a particular customer's buying habits in order to drive promotions straight to a customer's mobile device at the time of sale.
- **Risk modeling.** Working with various operational and analytical datasets—whether historical or real-time/streaming—to help financial institutions make better decisions on product, currency, or investments to avoid potential catastrophic losses.

The Cloudera and SAP platforms dovetail to provide seamless Hadoop\*-based operations and advanced predictive analytics. Cloudera Enterprise includes open source tools on top of advanced system management, data manage-

ment, and comprehensive security and governance. It delivers a powerful, cost-effective solution for storage, processing, discovery, and serving of any type of data. The SAP HANA platform melds transactional and analytical processing, dramatically reducing the data footprint with sophisticated data modeling and compression techniques. SAP HANA includes an in-memory columnar database, text, graph, spatial, predictive analytics, business rules, engines, and libraries.

SAP Hana's innovative architecture enables a completely new way of designing and developing business applications and results in unprecedented operational speed and a simplified IT landscape, among other benefits. When paired with Cloudera Enterprise, it helps generate insights and puts analytics into action at extraordinary scale.

Acting as a landing zone, processing engine, and data discovery platform, Cloudera Enterprise feeds data into SAP HANA for analytics, via certified connector to Hive or Impala (coming soon). SAP BI users can run queries against data in Cloudera Enterprise using tools like SAP Business Objects, SAP Predictive Analytics, and SAP Lumira, pulling data directly from Impala.

The combination of SAP HANA and Cloudera Enterprise, running on Intel® processors optimized for big data, provides customers the ability to drive deeper analytic insight from more complete and varied datasets.

### Spotlight on SAP

As the market leader in enterprise application software, SAP is at the center of today's business and technology revolution. Our innovations enable more than 296,000 customers worldwide to work together more efficiently and use business insight more effectively. SAP helps organizations of all sizes and industries overcome the complexities that plague our businesses, our jobs, and our lives. With "Run Simple" as our operating principle, SAP's nearly 75,600 employees focus on a singular purpose that inspires us every day: To help the world run better and improve people's lives.

For more information, visit [www.sap.com](http://www.sap.com).



### SAP and Cloudera

An environment using both SAP HANA and Cloudera Enterprise (Figure 1) offers several benefits:

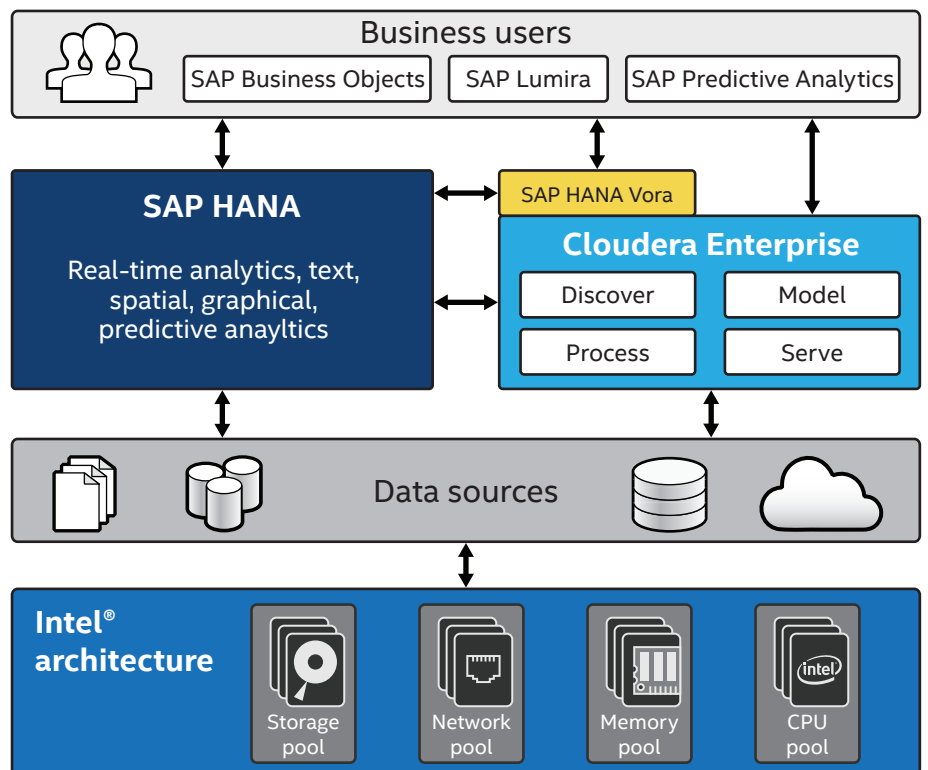
- Ability to capture, store, and process heterogeneous types of data (structured/unstructured, internal/external, SAP/non-SAP).
- Support for a variety of analytical workloads, including agile data exploration, pattern detection, visualization, real-time event detection, interactive SQL, full-text search, and machine learning.
- Real-time integration with business processes and operational systems.
- Continuous innovation through the open source community and leading software engineers, developers, and support teams.
- High-performance computing, optimized TCO, and scalability.

### Apache Spark\*

Cloudera Enterprise includes Apache Spark, an open source, parallel data processing framework in the Apache Hadoop ecosystem that makes it easy to develop fast, end-to-end Big Data applications that combine batch, streaming, and interactive analytics on all of your data. The key component of real-time interactive analytics, Spark runs the types of compute jobs that enable organizations to quickly crunch massive datasets and make smarter decisions out of the data. For analysts and data scientists who rely on iterative algorithms, Spark can deliver results 10 to 100 times faster than MapReduce.

Because it is open source, Spark is fully compatible with existing data-center environments and tools, including SAP's new HANA Vora query engine.

Figure 1 SAP HANA and Cloudera system architecture. With SAP and Cloudera, forward-thinking enterprises can integrate next-generation analytical systems like SAP HANA and business intelligence applications like SAP Lumira and SAP Predictive with their Cloudera-powered Big Data sources.



### SAP HANA Vora\*

SAP HANA Vora is an in-memory query engine that runs on the Spark framework supported by Cloudera and extends the Apache Spark execution framework to provide enriched interactive analytics. It accelerates queries of unstructured data in Hadoop environments, as well as structured information in common enterprise data sources, including SAP HANA.

HANA Vora leverages the Spark data processing engine to mine business insights from vast stores of data produced by machines, business transactions, and sensors. It puts big data analytics in context with an understanding of business processes to pull business insights from the data.

Many organizations that will likely be early adopters of HANA Vora operate in tightly regulated industries that limit who can see and act upon their data. The data should also be encrypted and secure-in transit and at rest. Those requirements make Cloudera Enterprise an easy choice for Spark, with its shared data management and governance model that closely tracks how data is brought into Hadoop, accessed, and transformed, as well as where it goes outside of Hadoop. Only Cloudera offers such comprehensive security and governance.

With two complementary platforms optimized for specific purposes—SAP HANA, designed for in-memory transactional and analytical processing, and Cloudera Enterprise, for secure, scalable processing, discovery, modeling, and serving of Big Data—the last piece an organization needs to make this a viable solution is a hardware platform that can quickly execute the operations these software solutions require.

### Intel® processors

The more powerful your hardware platform is, the better it can leverage the capabilities of SAP HANA and Cloudera Enterprise, and the more flexibility it can offer for right-sizing your IT environment and reducing TCO. Intel's latest generation of processors have embedded features that optimize performance in both SAP and Cloudera environments.

The recently released Intel® Xeon® Processor E5-2600 v3 product family, for example, adds 50 percent more cores and cache<sup>2</sup> over the previous generation and includes other hardware enhancements, such as Intel® Advanced Vector Extensions 2 (Intel® AVX2) and Intel® Quick Path Interconnect (Intel® QPI), which significantly boost output across a broad set of workloads. SAP HANA scan performance, which determines how many items in a database can be scanned per second, has been enhanced to benefit from Intel® AVX2.

As production big data environments have grown and an increasing quantity of sensitive data (personal, financial, IP) has found its home in Hadoop clusters, encrypting data at rest and in transit has become a business-critical need. Cloudera Enterprise is the only Hadoop distribution that includes comprehensive data encryption and key management, a requirement for data protection regulations like HIPAA and PCI-DSS.

Intel optimizes each new generation of processor to improve encryption/decryption performance and reduce processing times to the point of negligibility for Hadoop environments. The Intel® Xeon® Processor E5-2600 v3 product family, with built-in enhanced Intel®

### Spotlight on Cloudera

Cloudera is revolutionizing enterprise data management by offering a unified platform for Big Data, an enterprise data hub built on Apache\* Hadoop\*. Cloudera offers enterprises one place to store, access, process, secure, and analyze all their data, empowering them to extend the value of existing investments while enabling fundamental new ways to derive value from their data.

Cloudera's open source Big Data platform is the most widely adopted in the world, and Cloudera is the most prolific contributor to the open source Hadoop ecosystem. As the leading educator of Hadoop professionals, Cloudera has trained over 40,000 individuals worldwide. Over 1,900 partners and a seasoned professional services team help deliver greater time to value. Finally, only Cloudera provides proactive and predictive support to run an enterprise data hub with confidence. Leading organizations in every industry plus top public sector organizations globally run Cloudera in production.

For more information, visit [www.cloudera.com](http://www.cloudera.com).

**cloudera**®

Advanced Encryption Standard New Instructions (Intel® AES-NI), reduces performance latency for encryption/decryption dramatically. Tests prove organizations can encrypt an entire Hadoop database with no significant performance penalty.<sup>1</sup>

Combined with Intel® Data Protection Technology (Intel® DPT) and Intel® Secure Key random number generation, the Intel® Xeon® Processor E5-2600 v3 family provides stronger data protection than previous Intel® processors.

For online transaction processing (OLTP), Intel® Transactional Synchronization Extensions (Intel® TSX) capability, which provides hardware-supported lock elision for improved transactional data processing, boosts the performance of in-memory transactional data processing on systems with high core counts by increasing the scalability of thread synchronization. SAP HANA leverages Intel® TSX capability to improve its existing lock-based programming model, resulting in faster system performance and extended scalability.

Intel is committed to enhancing machine learning and graph analytics frameworks on the open source Apache\* Hadoop data platform, in addition to optimizing Intel® Xeon® processors for encryption, data protection, and other technologies.

1. Intel® Xeon® Processor E5-2600 v3 Accelerates Hadoop HDFS Encryption, Doc# 332307. Intel Corporation. 2015.

## Summary

Organizations of all types and sizes maintain increasingly large data stores to support business activities and decision making. By blending operational data with big data and speeding up analytical processing and application response time, they can use this information to better understand customer behavior, predict demand, uncover new revenue opportunities, and increase their overall agility.

- **Real-time answers.** Organizations can provide businesspeople across the enterprise with better decision support, and empower employees to get answers and respond to queries in real-time.
- **Better decisions.** The combination of SAP HANA and Cloudera running on Intel hardware can improve decision-making and empower data analysts and lines of business to gain deeper insights on more diverse data.
- **Integration.** The solutions are tightly integrated, limiting data movement and delivering a more holistic experience to end-users.

If your company is looking for a scalable, secure, Big Data-driven analytics solution that can process massive amounts of data and is fast enough to provide the timely analytical insights you need while maintaining data security, look no more.

## Contact us

For information on Intel Big Data Professional Services, visit [Intel.com/bigdata/services](http://Intel.com/bigdata/services).

For details on the latest Intel Xeon processor, visit <http://ark.intel.com> and select *Intel® Xeon® Processor E5 v3 Family*.

For more information on Cloudera products and services, visit <http://www.cloudera.com/content/cloudera/en/products-and-services.html>.

For more information on SAP HANA, visit [www.saphana.com](http://www.saphana.com).



Intel does not control or audit the design or implementation of third-party benchmark data or websites referenced in this document. Intel encourages all of its customers to visit the referenced websites or others where similar performance benchmark data is reported and confirm whether the referenced benchmark data is accurate and reflects performance of systems available for purchase.

Copyright © 2015 Intel Corporation. All rights reserved.  
Intel and the Intel logo are trademarks or registered trademarks of Intel Corporation in the United States and other countries.