

REALIZE THE BUSINESS ADVANTAGES OF BIG DATA WITH CLOUDERA ENTERPRISE ON THE CENTURYLINK[®] CLOUD



There's little debate left about the need for enterprises to move quickly to embrace big data and generate business value from it by accelerating insights, improving decision-making, and getting products and services to market faster. Enterprises are increasingly turning to Apache Hadoop to store and process their big data, but a key challenge remains in deciding where and how to run Hadoop for the best performance and return on investment.

Challenge: *Storing, processing, and analyzing big data in the cloud offers great promise and flexibility—but only if the benefits can be combined with enterprise-grade performance, security, and manageability.*

At stake: *On-premises big data solutions can be expensive to run and labor-intensive to manage. Moving data management to the cloud first can give enterprises a significant, and lasting, competitive advantage.*

Solution: *Enterprises can now run Cloudera Enterprise on the CenturyLink® Cloud powered by Intel® Cloud Technology for fast deployment and time to insight, as well as the flexibility to move between cloud and on-premises solutions.*

On-premises Hadoop deployments are the most common, but they can be costly, complex, and unwieldy to manage. Running big data in the public cloud offers a promising alternative, with the potential to combine the business benefits of big data with the advantages of the cloud. That can mean fast deployments, leading to fast time to market; fast and flexible scaling of resources in the cloud, for maximum cost efficiency with varying workloads; and simplified management, which frees your IT staff to focus on innovation and strategic initiatives, instead of day-to-day tasks.

Despite the significant benefits, many enterprises are hesitant to run big data in the cloud because of traditional fears around public cloud deployments. Specifically, enterprises need to be confident they can run Hadoop in the public cloud while also maintaining the requisite levels of enterprise security, reliability, performance, and manageability. Getting the answer

now is critical, because if the competition gets there first, they'll reap the hefty benefits of running big data in the cloud and gain a significant—and potentially lasting—competitive advantage.

This paper examines the challenges of running Hadoop on-premises. It then details how Cloudera Enterprise on the CenturyLink® Cloud overcomes those challenges, providing a high-performing, secure, reliable, and easily-managed solution that enables enterprises to lower costs and drive excellent business outcomes.

Why Move Big Data to the Cloud?

Enterprises are increasingly turning to Hadoop to help store, process, and analyze the ever-increasing volume and variety of big data that's pouring in from social media, clickstreams, videos, sensors, and more. Currently, most of those enterprise Hadoop deployments are on-premises, but there are significant concerns and limitations in three main areas: people, processes, and technology.

- **People:** Many enterprises find that running Hadoop is not particularly easy or intuitive. It can be complex to deploy, and most enterprises don't have the in-house expertise to stand up the appropriate infrastructure, manage it on a daily basis, and stay on top of new trends and technologies as they emerge.
- **Processes:** Deploying Hadoop with enterprise-grade security and 24/7 reliability requires proven processes. In-house IT departments often lack the reference models and templates they need to ensure the level of performance enterprises require.
- **Technology:** Many enterprise data centers are already nearing capacity, so adding a complex infrastructure to support Hadoop may not be feasible or cost-effective. Legacy infrastructure platforms usually need to be upgraded to support the new workloads, which is a cost-intensive endeavor. And after all that work, the on-premises infrastructure may be more than is needed for most short-term workloads.

The Right Cloud Partners

By running Cloudera Enterprise on the CenturyLink Cloud powered by Intel® Cloud Technology, enterprises can quickly realize the full potential of Hadoop-powered big data deployments in the public cloud, while also overcoming the common on-premises challenges related to people, processes, and technology.

The most significant advantage of this joint solution is the unique combination of Cloudera's unparalleled expertise and experience with Hadoop; the flexibility, performance, and ease of use of the CenturyLink Cloud; and the enterprise-grade security, performance, and manageability of Intel Cloud Technology.

Let's take a closer look at how these advantages play out in the three key areas of people, processes, and technology.

People

With many public cloud providers, you're really on your own. You can run a small dev/test workload or proof of concept in the cloud, but if you decide to expand to an enterprise-wide deployment that spans

physical and cloud infrastructures, you're out of luck. You'll likely need to hire internally to manage the expanding deployment, and you'll face related delays and costs.

CenturyLink Cloud is different. Along with offering a complete enterprise cloud offering that includes one of the most rigorous SLAs in the industry, CenturyLink Cloud provides the people and the infrastructure necessary to build, host, and manage hybrid cloud and other solutions. So, you can move quickly to store, process, and analyze your big data in the public cloud, with the level of support you desire—from basic support through professional services. And when your needs expand, a highly-skilled and experienced team is available to help you take your next step quickly and cost-effectively.

Processes

Founded in 2008, Cloudera was the first company to bring Apache Hadoop to the enterprise market, smoothing the process for enterprises to take advantage of the benefits of Hadoop in their big data projects. Today, Cloudera remains the leading provider and supporter of Hadoop for the enterprise with its Cloudera Enterprise offering. With Hadoop at the core, Cloudera Enterprise delivers a single integrated system that brings diverse users and app workloads to one pool of data, with enterprise-grade security and governance. That means enterprises can put their data at the center of their operations to increase visibility and reduce costs while also managing risk and meeting compliance requirements.

The Cloudera enterprise data hub (EDH) (see Figure 1), a reference architecture realized through Cloudera Enterprise, is the key to enabling enterprises to realize the promise of big data. It allows enterprises to acquire and combine any amount or type of data in its original fidelity, in one place, for as long as needed—delivering insights to all kinds of users, as fast as possible. Enterprises can rely on Cloudera to streamline their processes and tackle a variety of critical business challenges.

- Automatically archive the complete set of enterprise data to meet compliance requirements
- Complement existing enterprise data warehouses to offload data and workloads, and improve performance while managing costs
- Support self-service business intelligence (BI) through familiar tools, on more data and more kinds of data than ever before
- Enable and consolidate enterprise search on data and documents in place within a single document
- Accelerate advanced analytics solutions such as recommendation engines, fraud detection, and image processing

Cloudera Enterprise offers several key advantages. One is an active archive, which means you can store all your data—in any format, at any volume—as long as you like. You can address your compliance

requirements and deliver data on demand to satisfy regulatory demands. And because it's secure, you control who sees what, tracking access over time. Another advantage is that you can run, extract, transfer, and load (ETL) workloads on the EDH, where they can run in parallel and at a low cost.

Cloudera Enterprise also supports self-service exploratory BI and advanced analytics. On the BI front, it allows users to explore data with full security, using the interactive BI tools they're used to. Advanced analytics are necessary to get the most out of your big data, and with Cloudera Enterprise, multiple computing frameworks enable analytics, search, machine learning, and more. That means you can examine all your historical data, in full fidelity, and combine it with comprehensive analyses. Simple tabular data can be mixed with complex, multi-structured data in ways you've likely never imagined before.

What all this adds up to are several benefits for your enterprise that are unique to Cloudera, including:

- The same full-fidelity data hub experience as an on-premises environment, from technology capabilities to system and data management tools—coupled with mission-critical support
- Enterprise-grade data security and data governance
- Support for the latest innovations in the Hadoop platform such as Apache Sentry and Cloudera Search
- An expanded partner program, with multiple pricing and support models that offer maximum flexibility in deployment and consumption

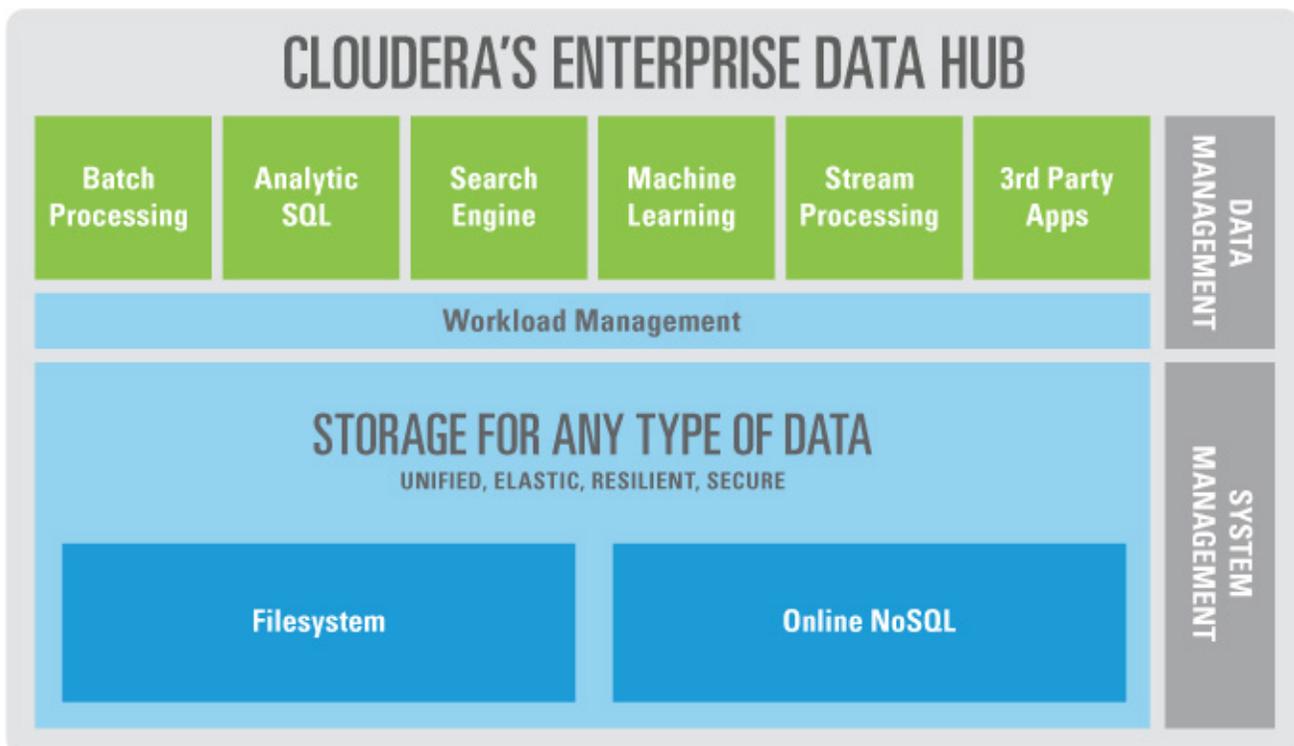


Figure 1. Cloudera's enterprise data hub combines unified storage with advanced data management tools.

Technologies

Cloudera works closely with Intel to optimize its distribution on Intel® architecture. In addition, CenturyLink Cloud is powered by Intel Cloud Technology, so enterprises that run Cloudera Enterprise on the CenturyLink Cloud can benefit from the technology leadership that Intel brings to the table, including in the areas of security and performance.

- **Security:** Enterprise deployments of Hadoop have long been held back by security and privacy concerns, and the difficulty of adding enterprise-grade security without hindering the performance, usability, and manageability of Hadoop.

To address these challenges, CenturyLink Cloud takes advantage of several advanced Intel® technologies, including Intel® Data Protection Technology with Advanced Encryption Standard New Instructions (AES-NI),¹ which reduces performance penalties and makes it feasible for enterprises to process data in Hadoop—even highly-sensitive data that must be encrypted at all times.

- **Performance:** CenturyLink Cloud Hyperscale servers achieve high levels of data management performance thanks to industry-leading Intel® Xeon® processors, high-end flash storage provided by Intel® Solid-State Drives (Intel® SSDs), and the added-performance-when-you-need-it of Intel® Turbo Boost Technology.

The Cloudera CDH Platform

At the heart of Cloudera Enterprise is Cloudera's open-source platform distribution, CDH. CDH is the world's most complete, tested, and popular distribution of Apache Hadoop and related projects. CDH is 100% Apache-licensed open-source and is the only Hadoop solution to offer unified batch processing, interactive SQL, and interactive search, and role-based access controls. More enterprises have downloaded CDH than all other similar distributions combined.

CDH includes the core elements of Hadoop plus several additional open-source projects, including:

- **Online NoSQL—HBase:** HBase is a distributed key-value store that helps you build real-time apps on massive tables with rapid access.
- **Analytics SQL—Impala:** Impala is the industry's leading massively-parallel processing (MPP) SQL engine built for Hadoop.
- **Search—Cloudera Search:** Cloudera Search lets users query and browse data in Hadoop just as they would search Google or similar sites.
- **In-Memory Machine Learning and Stream Processing—Apache Spark:** Spark delivers fast, in-memory analytics and real-time stream processing for Hadoop.

¹ No computer system can provide absolute security. Requires an enabled Intel® processor and software optimized for use of the technology. Consult your system manufacturer and/or software vendor for more information.

Due in large part to these advanced Intel technologies, CenturyLink Cloud Hyperscale servers deliver industry-leading IOPS performance and the highest CPU speed, as shown in independent research (see Table 1).²

CenturyLink Hyperscale vs. Competitors	2 Cores	4 Cores	8 Cores	16 Cores
CPU Performance: CenturyLink Performance Advantage	211%	167%	136%	136%
Disk Performance, Random 16K Blocks: CenturyLink Performance Advantage	319%	343%	329%	332%
Disk Performance, Random 64K Blocks: CenturyLink Performance Advantage	427%	430%	423%	409%

Table 1 – Compared to the competition, CenturyLink Hyperscale servers offer a CPU performance advantage of up to 211%, a disk read advantage of up to 343%, and a 64K disk read advantage of up to 430%.³

What Does This Mean for Your Business?

In the real world of business, running Cloudera Enterprise on the CenturyLink Cloud powered by Intel Cloud Technology can accelerate, deepen, and broaden insights—which leads to competitive advantage and differentiation.

Say, for instance, you have customer databases in multiple silos, and your CMO wants to combine all those databases across products, and add in a stream of Twitter feeds for real-time analysis. Setting up the

About CenturyLink Cloud Hyperscale Servers

CenturyLink's Hyperscale servers deliver an optimal combination of performance, security, and manageability for enterprises to run open-source NoSQL and big data workloads in the public cloud.

Performance

Thanks to 100% flash storage from Intel® Solid-State Drives (Intel® SSDs), customers can expect at least 15,000 IOPS from Hyperscale instances, and may experience even higher levels of performance.⁴

Security

Hyperscale servers are SSAE 16 SOC2 Type 2 certified and HIPAA compliant. Plus, the servers have added data protection from Intel, including advanced data encryption, protection against malware, and trusted compute pools that increase transparency, visibility, and control.

Manageability

Hyperscale instances can be built with any combination of compute, storage, and operating system—all as part of CenturyLink Cloud. Automated, self-service management tools allow enterprises to test new data management ideas quickly and at low risk and cost. Plus, flexible, pay-as-you-go pricing means enterprises pay only for what they use.

² See reports from Cloud Harmony (www.centurylinkcloud.com/lp/resources/Cloud-Harmony_CenturyLink-Hyperscale_Report.pdf) and Cloud Spectator (www.cloudspectator.com/reports/CenturyLink_Amazon_Value_Comparison_Processors.pdf). Tests were performed on cloud-based servers with 2, 4, 8, and 16 cores. The CenturyLink Cloud standard server was powered by an Intel® Xeon® E5-2680 2.70GHz. The CenturyLink Hyperscale server was powered by an Intel® Xeon® E5-2650 v2 2.60GHz. Competitor servers ran Intel® Xeon® E5-2680 v2 2.80GHz and Intel® Xeon® E5-2670 2.60GHz processors. The reference system used in the research was a bare-metal (non-virtualized) Dell M610 PowerEdge server. This server had two Intel® X5650 2.66 GHz CPUs (12 cores total), 48 GB DDR3-10166 memory, and a Seagate SAS 2.0 10k RPM drive dedicated for testing.

necessary big data processing in-house would typically take months, if not a year or more—time that you cannot afford to spend if you want to remain competitive and capture immediate business opportunities. Plus, the on-premises solution would be expensive and difficult to manage, and it might require hiring additional IT staff.

With the joint solution from Cloudera and CenturyLink, you can stand up that infrastructure in the cloud in far less time, with less cost, and without the management headaches. Hyperscale servers powered by Intel Cloud Technology will deliver analyses in minutes instead of hours. That means you can capture customer behaviors and buying trends to rapidly improve your marketing decision-making and ROI. CenturyLink Cloud will also host the deployment, with SLAs to ensure you get the performance you need. From there, you can ramp up and down quickly to maximize cost efficiency, based on the needs of specific workloads.

Of course, what the CMO will care most about is the result: a solution that's ready far faster, produces near-real-time responses to queries, enables deep insights, and ultimately supports better business decisions and faster time to market with new products and services.

In sum, the joint solution makes it possible for enterprises to:

- Accelerate deployment of production-ready big data solutions
- Capture and store enterprise data for all business functions
- Analyze and explore data in new ways to discover new patterns and insights
- Provide access to data for everyone in the organization to drive competitive business outcomes
- Have a platform to continuously accommodate more data volume and integrate big data solutions.

Conclusion: Capture the Full Potential of Big Data in the Public Cloud

The potential benefits of efficiently storing, processing, and analyzing big data are compelling to most enterprises, but so are the challenges. Hosting Hadoop deployments on-premises can be too time-intensive, complex, and expensive, but hosting those deployments in the public cloud has seemed too difficult or risky—at least until now.

By running Cloudera Enterprise on the CenturyLink Cloud powered by Intel Cloud Technology, enterprises can achieve the enterprise-grade security, reliability, and manageability they need to move to the cloud with confidence. The joint solution combines Cloudera's unrivaled Hadoop expertise with the flexibility and manageability of the CenturyLink Cloud, and the security and performance of Intel technologies.

³ Figures based on relative performance of CenturyLink Hyperscale server versus competitor average. <http://go.centurylinkcloud.com/CloudHarmonyReport> Tests were performed on cloud-based servers with 2, 4, 8, and 16 cores. The CenturyLink Cloud standard server was powered by an Intel® Xeon® E5-2680 2.70GHz. The CenturyLink Hyperscale server was powered by an Intel® Xeon® E5-2650 v2 2.60GHz. Competitor servers ran Intel® Xeon® E5-2680 v2 2.80GHz and Intel® Xeon® E5-2670 2.60GHz processors. The reference system used in the research was a bare-metal (non-virtualized) Dell M610 PowerEdge server. This server had two Intel® X5650 2.66 GHz CPUs (12 cores total), 48 GB DDR3-10166 memory, and a Seagate SAS 2.0 10k RPM drive dedicated for testing.

The solution makes it possible for enterprises to realize the full potential of big data in the public cloud—and to do so in far less time than with on-premises alternatives. With Cloudera Enterprise on the CenturyLink Cloud, enterprises can speed time to market by quickly scaling use cases, reducing the burdens of managing in-house deployments, implementing new data-driven business strategies, and building competitive advantage by running more comprehensive analyses faster to drive market-leading innovations.

About CenturyLink Cloud

CenturyLink Cloud, from CenturyLink Technology Solutions, is the complete platform to easily manage your entire business application portfolio, from development to business-critical workloads. CenturyLink Cloud offers high-performance, scalable, self-service virtual machines across our global network of data centers, including Hyperscale servers powered by Intel Cloud Technology for distributed workloads that require maximum performance.

And CenturyLink Cloud provides built-in automation, orchestration, and management tools for an IT-ready and developer-friendly platform that is flexible, scalable, cost-effective, and highly-manageable.

Further details can be found at www.centurylinktechnology.com/financial-services.

About CenturyLink Technology Solutions

CenturyLink Technology Solutions delivers innovative managed services for global businesses on virtual, dedicated, and colocation platforms. It is a global leader in cloud infrastructure and hosted IT solutions for enterprise customers. Parent company CenturyLink, Inc. is the third-largest telecommunications company in the United States, and empowers CenturyLink Technology Solutions with its high-quality advanced fiber optic network. Headquartered in Monroe, La., CenturyLink is an S&P 500 company and is included among the Fortune 500 list of America's largest corporations.

For more information, visit www.centurylink.com/technology.

About Cloudera

Cloudera is revolutionizing enterprise data management by offering the first 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

⁴ Tests were performed on cloud-based servers with 2, 4, 8, and 16 cores. The CenturyLink Cloud standard server was powered by an Intel® Xeon® E5-2680 2.70GHz. The CenturyLink Hyperscale server was powered by an Intel® Xeon® E5-2650 v2 2.60GHz. Competitor servers ran Intel® Xeon® E5-2680 v2 2.80GHz and Intel® Xeon® E5-2670 2.60GHz processors. The reference system used in the research was a bare-metal (non-virtualized) Dell M610 PowerEdge server. This server had two Intel® X5650 2.66 GHz CPUs (12 cores total), 48 GB DDR3-10166 memory, and a Seagate SAS 2.0 10k RPM drive dedicated for testing.

22,000 individuals worldwide. Over 1,200 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. www.cloudera.com

© 2014 CenturyLink. All Rights Reserved. The CenturyLink mark, pathways logo and certain CenturyLink product names are the property of CenturyLink. All other marks are the property of their respective owners.

Copyright © 2014 Intel Corporation. All rights reserved.

Intel, the Intel logo, and Xeon are trademarks of Intel Corporation in the U.S. and/or other countries.

*Other names and brands may be claimed as the property of others.

NOTE: Software and workloads used in performance tests related to Intel products may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.

Global Headquarters
St. Louis, MO
(800) 728-8471

EMEA Headquarters
United Kingdom
+44 (0)118 322 6000

Asia Pacific Headquarters
Singapore
+65 6591 8824