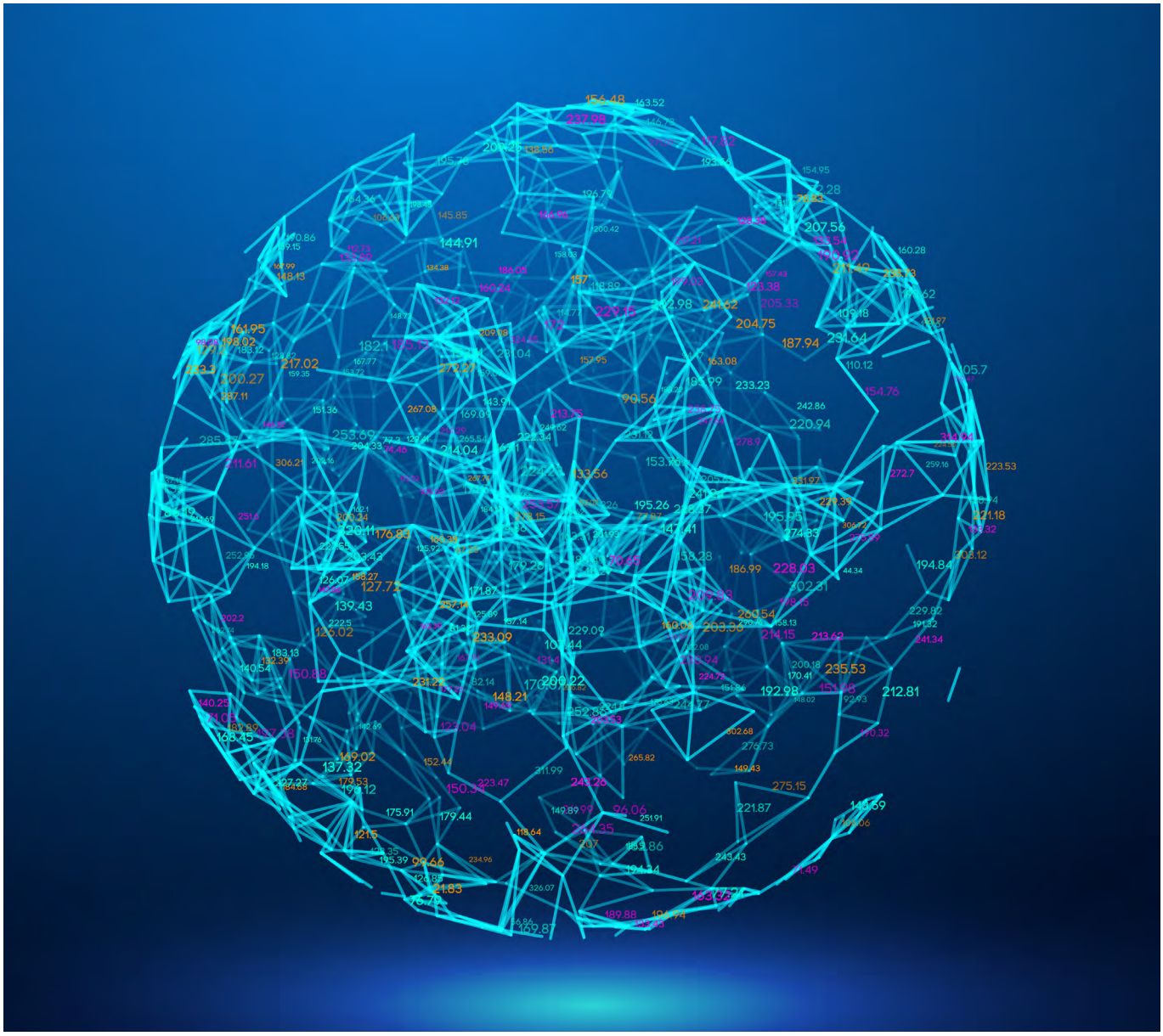
CIO FROM IDG | Red Hat CLOUDERA

Enterprise analytics: Ideal vs. Reality

In any iterative process, speed matters.

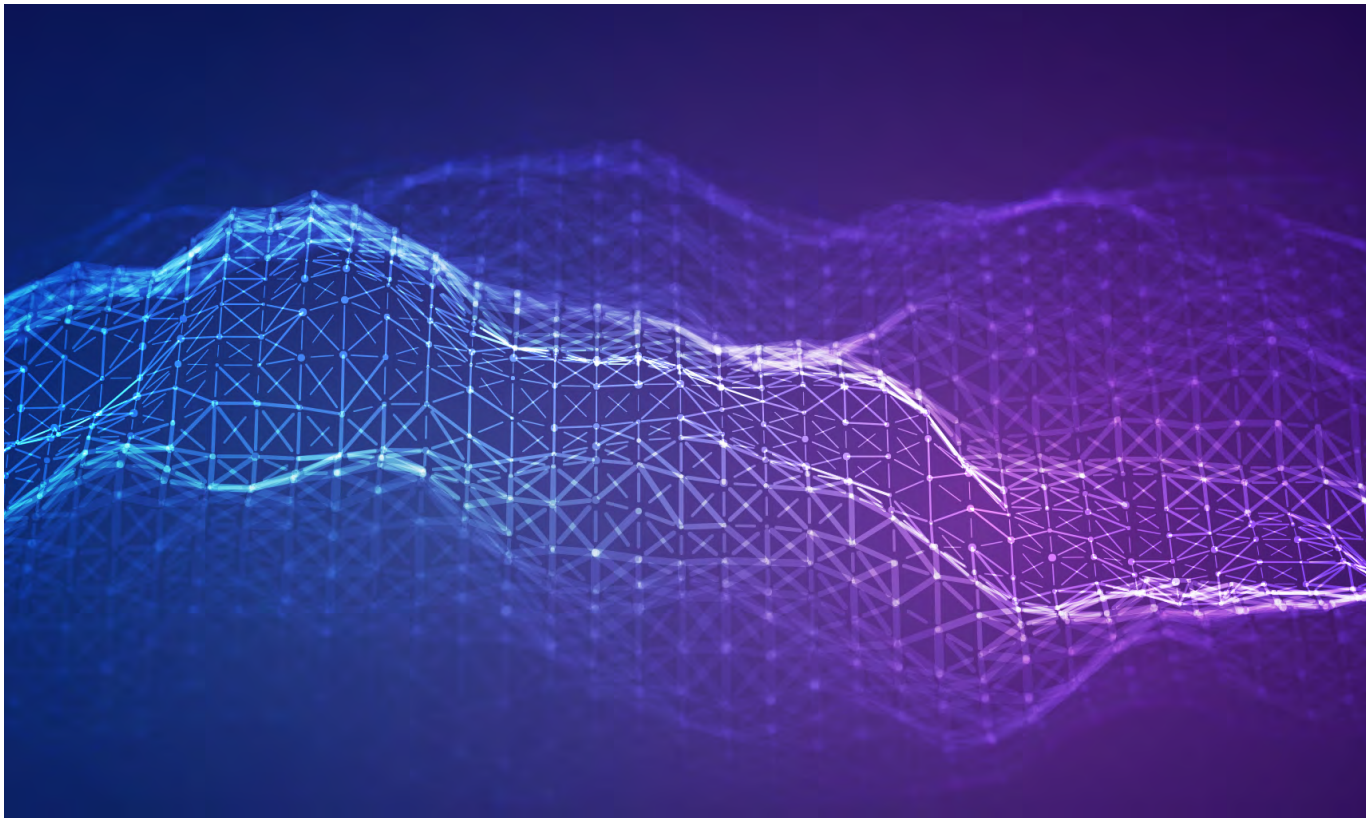


In analytics, the ideal is a steady cadence of query, response, and business decision-making. You need the results of the query that sits in front of you to inform tactics and strategy as soon as possible. When that job is complete, space for the next query opens up. The result should be a cycle of continuous improvement in products, services, and business processes.

All too often, the reality is far less impressive.

Traditional on-premise architectures impose delays. From the user's perspective, delay might involve a dashboard frozen by excessive demand on a shared cluster. Or it might take the form of multi-week waits for new hardware provisioning or the completion of an upgrade. The cumulative effect is the same: a brake on innovation.

However you look at it, the rigidity of traditional on-premise architectures extracts a high price all round, for analytics teams, IT organizations and the business decision makers who rely upon them.



On-premise analytics: traditionally slow, painful, and expensive

We can identify five sources of delay in analytics that runs on traditionally-configured hardware and storage. Too often, the only way of avoiding these delays has been to spend more money and create more challenging trade-offs for the platform team to manage.

1) Slow onboarding

Why wait weeks or months to provision the resources for a new compute cluster in the data center? If they have sufficient seniority, many executives won't wait. Instead, they bankroll point solutions. The results nearly always involve additional costs, and often include the challenges inherent in shadow IT.

2) Noisy neighbours

In traditional data centers, where rigidly-defined compute clusters may be shared between different analytics teams, it's a challenge to estimate peak demand. Unplanned spikes in usage on one team can disrupt other teams' ability to access the compute resource, leading to SLA breaches.

3) Data silos

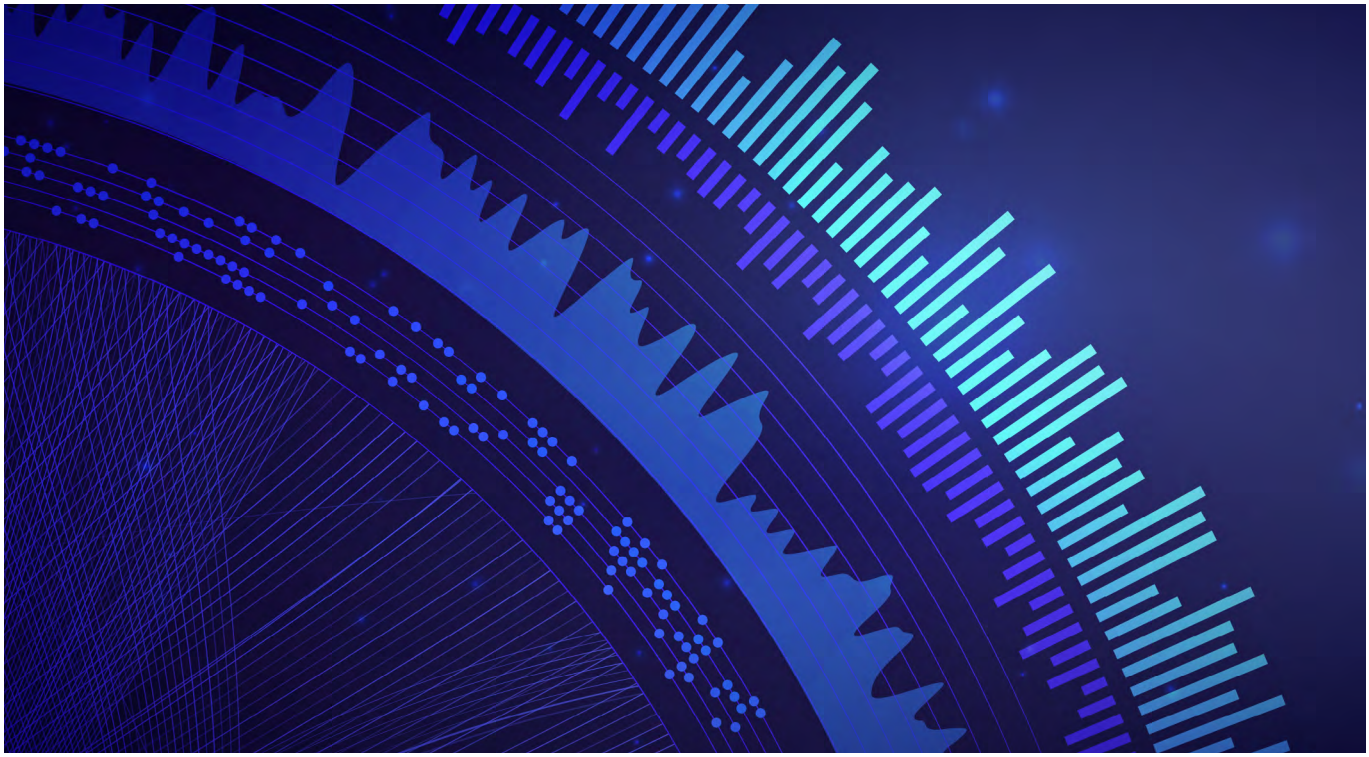
When new applications arrive in-house, they require a new cluster. And with each cluster comes the potential for more siloed data. These silos join the other silos, including those created by adventures with shadow IT. Pretty soon, the sprawl starts to look ugly. More to the point, it's increasingly hard to manage.

4) Lack of elasticity

In traditional on-premise architectures, there's often only one answer to the noisy neighbour problem: more resource. However, over-provisioning breeds poor utilization. Once this was simply the price of doing business. Today, the economics of public cloud allow us to see utilization rates of 30%-40% for what they are: wasteful.

5) Complex upgrades

If you have a finance team with predictable yet mission-critical workloads sitting on the same cluster as an analytics team dealing with an unpredictable volume of cutting-edge engineering queries, it's going to be hard to reach agreement on the frequency with which upgrades should occur. Building a plan might take months, depriving some users of tools they could successfully exploit.



The ideal platform for on-premise analytics: what does it look like?

What does the ideal platform for on-premise analytics look like? How would this platform minimize the delay, cost and the stuttering cadence of so much on-premise analytics?

Here are the features that Cloudera and Red Hat believe both data scientists and platform teams deserve:

Cloud-like onboarding

The ideal on-premise analytics architecture would eliminate lengthy provisioning schedules and offer data scientists what developers have enjoyed for the past decade: the autonomy of self-service. The analyst who needs 20 nodes to run a series of complex queries would simply request the capacity on a screen. In an ideal world, we could reduce the lead-time for provisioning that capacity from five days to perhaps five minutes.

Eliminating noisy neighbours

In an ideal world, each application would be encapsulated, with guaranteed access to the resources it requires from a shared environment. Result: much quieter neighbours, no frozen screens and much more attractive utilization rates across the entire compute resource.

Minimizing data silos

In traditional architectures, we create new clusters to underwrite the performance of specific applications. Along the way, we replicate a lot of data, and carry it to the new cluster, creating silos that are challenging to manage. The weak link is the need to create new clusters. In a better world, we could avoid much of this effort by offering our applications a true multi-tenancy guarantee of performance levels.

Disaggregate for elasticity

The logic of the cluster is broken because it ties compute and data together so completely. A far better approach borrows from virtualization: separate compute from data and allow your applications to access both as required.

Upgrade autonomy

Disaggregation will also allow the finance team and the engineering team (in our example above) to run applications with their own libraries – upgraded regularly or not, as each team sees fit.

If all of this reminds you of the public cloud, you're not wrong. In fact, our ideal on-premise analytics architecture replicates much that's familiar from that world, including disaggregation, multi-tenancy, isolation, elasticity, and self-service provisioning.

A best of breed solution for traditional data center environments

Too many enterprises are living with a crisis of under-performance in traditional data centers running analytics workloads. For these organizations, analytics has become slow and painful, a pale imitation of the rapid, iterative processes that characterize data-driven enterprises.

To address this crisis, Red Hat and Cloudera, two of the world's leading open source distributors, have developed CDP Private Cloud, an integrated combination of infrastructure orchestration, application management and next-generation analytics. Available as an upgrade to existing Cloudera software or as a migration from third-party solutions, CDP Private Cloud brings cloud-native speed and flexibility to on-premise analytics. It embodies the solutions that on-premise analytics so clearly needs.

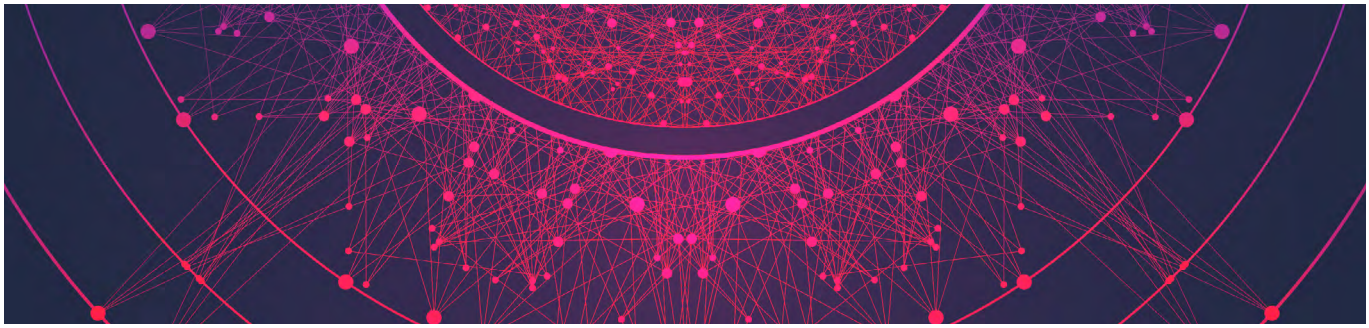
The cloud-native foundation: Red Hat OpenShift

Red Hat OpenShift

Red Hat's contribution to CDP Private Cloud is Red Hat® OpenShift®, the world's most flexible and widely-deployed multi-cloud container platform. Driven by Kubernetes, Red Hat OpenShift automates the provisioning, management and scaling of compute, storage, and applications across multiple, heterogeneous environments.

You've probably encountered Red Hat OpenShift being discussed in the context of multi-cloud and hybrid cloud operations, where its extreme flexibility plays a major role in realizing cloud strategy. But Red Hat OpenShift has been engineered to work in traditional data centers, too.

It performs with ease across bare-metal servers, virtual environments and cloud platforms, isolating workloads on disaggregated compute and storage. On premise, Red Hat OpenShift allows you to run your choice of applications on a private cloud, with greatly reduced provisioning times and far higher levels of hardware utilization.



The data management platform: Cloudera Data Platform

CLOUDERA

Cloudera Data Platform (CDP Private Cloud) accelerates the management and analysis, as well as the securing and governance, of all enterprise data sets. Sitting on top of Red Hat OpenShift, CDP Private Cloud addresses the full spectrum of requirements for analytics, including real-time data ingestion, data engineering, data warehousing, data science and machine learning. Built on open source code like Red Hat OpenShift, CDP Private Cloud supports the widest possible range of analytics workloads, tools, and applications.

CDP Private Cloud allows different users to work on the same data with the tools of their choice, avoiding the need for replicated data sets and the silos that result. CDP Private Cloud includes Cloudera SDX, an advanced data fabric that organizes data, enforces access policies and enables self-service.

So say goodbye to frozen screens. Say goodbye to lengthy waits for provisioning. Whether you are working with a complex legacy environment that's not ready for migration to the public cloud, or with data sets that simply belong on-premise, CDP Private Cloud is a uniquely powerful solution. Above all, it gives your traditional architecture the opportunity to become a high-performance engine room at the heart of the data-driven enterprise.

Cloudera

At Cloudera, we believe that data can make what is impossible today, possible tomorrow. We empower people to transform complex data into clear and actionable insights. Cloudera delivers an enterprise data cloud for any data, anywhere, from the Edge to AI. Powered by the relentless innovation of the open source community, Cloudera advances digital transformation for the world's largest enterprises.

Cloudera CDP Private Cloud

Red Hat

Red Hat OpenShift is a Kubernetes container platform with full-stack automated operations to manage hybrid cloud, multicloud, and edge deployments. Red Hat OpenShift is optimized to improve developer productivity and promote innovation.

Red Hat OpenShift

CIO FROM IDG | Red Hat CLOUDERA