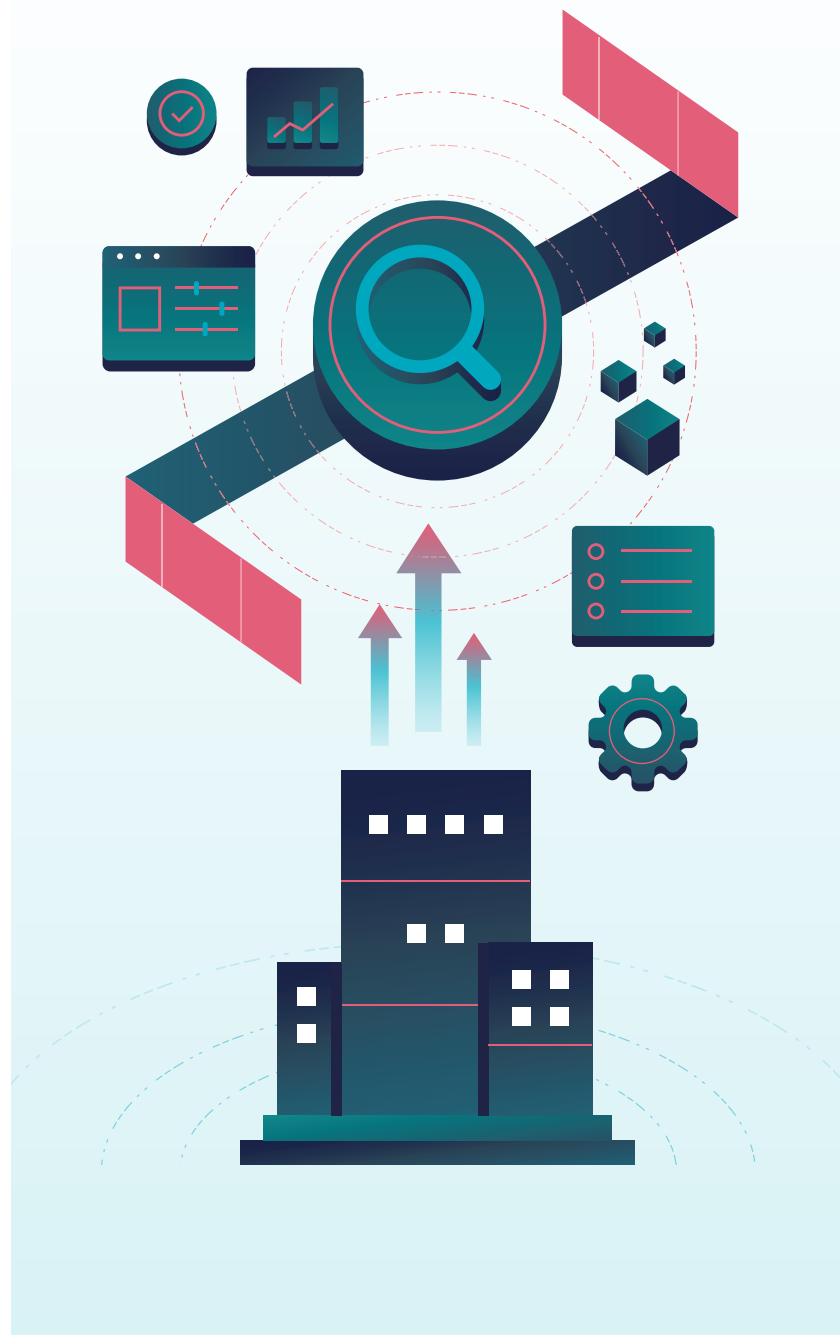


# Big Data Observability & Continuous Tuning at Scale

Modern enterprises are struggling to manage and understand unprecedented amounts of data. Leveraging this data is key to business success—but optimizing performance and keeping costs in line is a technical challenge. A key to meeting this challenge is effective observability.

Observability relates to how effectively an enterprise can understand their big data system by analyzing that system's outputs. Gartner defines observability as “the characteristic of software and systems that allows them to be ‘seen’ and allows questions about their behavior to be answered.” Effective observability is key to understanding the internal states of a big data system so that action can be taken.

To achieve proper observability, organizations need the ability to correlate application and infrastructure performance metrics. They should also be able to take action to optimize performance in real time, before problems occur. IT teams and app developers need observability and continuous tuning for the entire analytics stack—at scale.

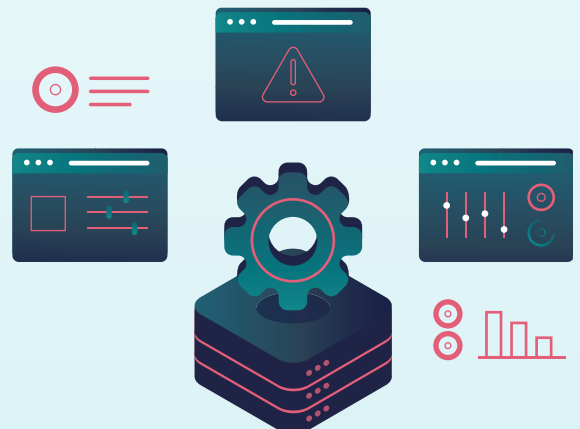
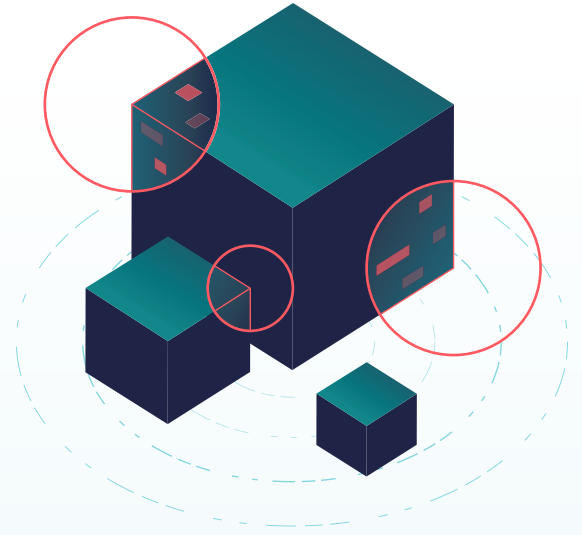


# The Difference Between Monitoring and Observability

Observability and monitoring are different, yet complementary, solutions. You need both, and one doesn't replace the other. Monitoring is critically important, and most enterprises use a combination of traditional Application Performance Management (APM) solutions, platform add-on tools, and homegrown solutions to perform monitoring. However, monitoring alone isn't enough to derive maximum value from today's big data ecosystem.

**Monitoring is the When** – Monitoring collects metrics and logs that provide information on whether the system is working, and it lets you know when something went wrong, so that you can react quickly.

**Observability is the Why** – Observability enables users to gather actionable data that provides not only the when of an error or issue, but—more importantly—the why. Observability typically shortens the duration, and reduces the impact, of incidents. Only with observability can you move from reactive to proactive to predictive IT.



## Why Traditional Monitoring Isn't Enough

Developers and IT Ops teams work with enormous amounts of data that increase exponentially every day. Workloads are increasingly resource-hungry, and meeting business-driven SLAs is increasingly more difficult.

Optimizing the performance of big data applications is often a challenge in and of itself. When it comes to operating big data environments at scale, performance relies on dependencies among the various layers within the architecture—the infrastructure, the virtualization layer, and the underlying hardware. On top of all this, the environment is continually changing by the second.

Faced with this challenge, many organizations find that their current legacy monitoring solutions are inadequate in today's modern IT world. In fact, monitoring on its own isn't enough. Maintaining and improving performance requires the why as well as the when. With modern software systems, observability enables teams to respond quickly and resolve emergencies. Only with observability can teams:

- Optimize the performance of cloud and modern investments.
- Understand the real-time performance of their digital infrastructure and make data-driven decisions.
- Be proactive and predictive.

## Where Monitoring Comes up Short of Expectations

Identifying the best big data monitoring solution and integrating it with your system to get actionable insights about your operational data is rife with complexities, and many solutions can fall short of expectations and value. Additionally, most monitoring solutions are behind the curve without the automation that is required in the modern data center.

**"Many monitoring tools advertise the ability to consume and analyze multiple types of machine data. It is more important to be concerned with having the right data in the right place at the right time than to prioritize consolidation."**

Gregg Siegfried, Gartner: *Monitoring and Observability for Modern Services and Infrastructure*.



## Homegrown Monitoring Solutions

Homegrown monitoring solutions don't scale and perform at enterprise levels, and they are typically time-consuming and costly to maintain. Additionally, it is difficult to support homegrown tools when the talent that created them leaves the company.

## Siloed Solutions

Some organizations rely on the free or low-cost monitoring solutions provided by public cloud vendors. These solutions only monitor and manage workloads running on that vendor's infrastructure. The problem is that if the IT Ops team is monitoring workloads for multiple vendors, they have to use several different solutions. These tools don't provide a consolidated view of the different environments, including workloads running in private clouds or traditional infrastructure in their own data centers.

## Cloud Support for Legacy Solutions is Lacking

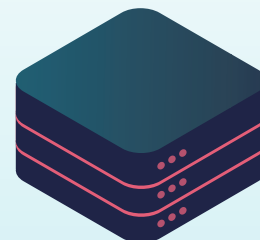
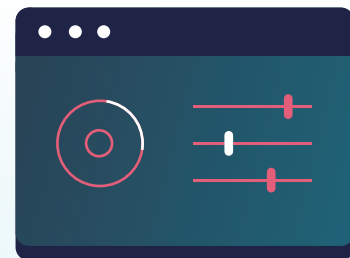
Many legacy monitoring tools were designed for in-house data centers, not cloud monitoring. Although many legacy on-prem monitoring tools have implemented cloud support, many still do not provide the level of detail and comprehensive capabilities that enterprises require to succeed.

## Selecting a Multi-Platform Monitoring Solution Requires Due Diligence

Many enterprises want to implement a hybrid- or multi-cloud monitoring and management strategy. These solutions are more complex and, often, more costly. Different and wide-ranging feature sets make comparing them difficult, and the research and evaluation typically require a significant level of effort to determine whether the solutions can perform as expected and get ROI.

## Lack of Automation

Many DevOps teams are still manually tuning, and they spend a lot of time doing that. As multi-tenant environments become more complex and dynamic, the burden to keep up the pace falls on the shoulders of the DevOps teams. The volume, velocity, and variety of the data being collected in the modern IT world demands automation and is fundamentally unmanageable by humans.



## Moving Beyond Legacy Monitoring

Moving beyond the limitations of legacy monitoring requires two critical components:

1. Unlocking the why that comes with observability
2. Implementing continuous tuning

Big data application performance requires deploying a solution that correlates application performance with infrastructure performance data to optimize the entire system, workloads included, within the context of cluster operations.

Observability enables you to move beyond just knowing you have an issue. With a real-time view of your cluster, a contextual why, and context-aware recommendations, you can speed up time to resolution as well as proactively optimize and automate applications and infrastructure to prevent issues.

In multi-tenant environments, operations metrics are of critical importance. When something is not working optimally or especially if issues impact revenue, knowing why is critical to resolving the issue. Properly resolving bottlenecks and failures requires rich contextual operational data that traverses infrastructure and application performance.

## How Continuous Tuning Enhances Observability

By instrumenting systems and applications, and collecting metrics and logs that enable the understanding of system behavior, the goals of observability are to:

- Create end-to-end system visibility.
- Examine the sequence of a problem through monitoring, correlating system data and automating with ML.
- Get accurate operational insights.
- Explain system behavior over time.

Continuous tuning allows an enterprise to move beyond manual tuning, which many DevOps teams are still doing. This is very time-consuming, and it doesn't scale. Worse, it's near impossible with the scale of data being collected—thousands of applications per day and a growth rate of dozens of nodes per year—for humans to manage. Even the most experienced IT Ops teams and capacity planners can't manually tune every application and workflow with the precision and speed that an automated solution does.

As new technologies and practices increase velocity and reduce the friction of getting software from code to production, they introduce greater complexities. Achieving system observability and reaping the benefits of continuous tuning might be the difference between success and failure for the modern IT team.

## Achieving End-To-End Visibility with Actionable Insights

True observability means understanding the why of big data events. Seeing the full picture, or having the full context to understand why these issues occur, is critically important. When holistically optimizing big data performance, there are several key areas that every solution must master.



## Classifying Performance Problems

Applications are often plagued with multiple performance problems. Classifying or categorizing them accelerates problem resolution. Bottlenecks and other external factors require a different approach than problems like configuration or inefficient use of resources. Observability and continuous tuning help IT to divide and conquer challenges more efficiently by providing actionable insights to both IT and developers with clear indications of which applications are causing problems. This triage serves to vastly reduce the time it takes to ultimately resolve problems and improve performance overall.



## Providing Definitive Analysis

Observability and continuous tuning remove the guesswork from performance analysis and deliver more precision. Without precise data, performance analysis may hinge on experience or even gut instinct. With real insights, data guides the process, revealing what is—and what is not—causing problems, enabling the team to move on without wasting time or effort.



## Solving Intermittent Problems

Intermittent performance problems tend to be the most challenging to diagnose, for several reasons:

- The conditions of the failure are often elusive.
- Recurrence is unpredictable, which lessens the effectiveness of pattern matching or pattern recognition as troubleshooting steps.
- The environment itself is changing throughout the duration of these problems, which lessens the effectiveness of pattern matching or pattern recognition as troubleshooting steps.

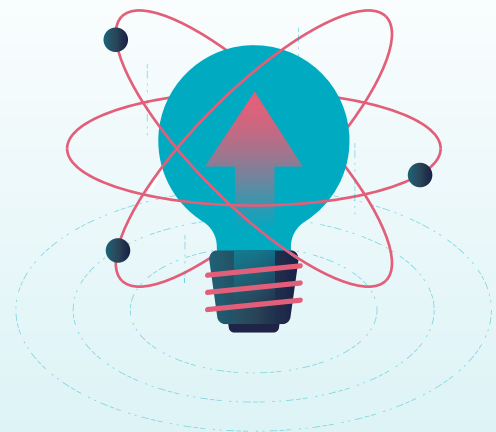
Observability and continuous tuning address these challenges and enable IT to speed up diagnosis and resolution time of intermittent problems. Observability, when fine-grained and properly scoped, can help address intermittent challenges by providing the data to support the trend analysis and anomaly detection required to flush out the source, or sources, of intermittent problems. Continuous tuning can help eliminate problems on the fly and report on the scale of inefficiencies found as an additional data point in the effort to diagnose intermittent problems.

## Gaining Insight into Highly Dynamic Environments

Observability and continuous tuning are extremely effective at quickly diagnosing problems in cloud-based, virtualized, or containerized environments. In these dynamic application environments, the infrastructure is constantly changing, and a sampled approach can miss the state changes that occur as components are spun up and down.

Continuous tuning eliminates guesswork and enables you to capture every state transition to accurately portray how a system is behaving.

A holistic approach empowers you to do more than just monitoring and diagnostics. It can be used to methodically audit and analyze application and infrastructure performance. The availability of deep performance data and big data analytics becomes the basis for continuous performance improvement, enabling developer and operations teams to better understand and optimize where their time is being spent.

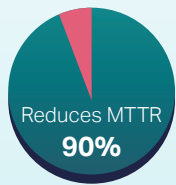
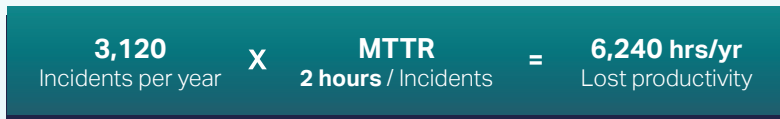


# Pepperdata: Observability and Continuous Tuning

Pepperdata provides observability and continuous tuning for the big data analytics stack. For optimal performance on premises or in the cloud, Pepperdata provides real-time visibility for troubleshooting, debugging, planning, and automated tuning.

## Observability ROI: Fortune 200 Company

### Existing Environment



= **5,616 hours saved/year**

Unlike data sampling or the use of triggers, Pepperdata-driven observability and continuous tuning leverages performance data in real time to provide insights and benefits to the platform and end users. The system effectively resolves performance problems faster and optimizes performance to reduce mean time to resolution by 90%, in many cases, and improve platform performance by up to 50%, even in dynamic application environments that rely on micro-services, virtualization, and cloud services. Pepperdata is able to do this within a small footprint—taking up just 1% of one core and 300 MB of memory per server instance.

The Pepperdata PepAgent collects hundreds of metrics every few seconds: application metrics, JMX metrics, cluster metrics, kernel metrics, and custom metrics of any kind. It records user activity data, event data, and performance data. Pepperdata also enables you to create customized real-time alerts on specific behaviors so you immediately know about critical conditions that could affect system performance. The UI empowers all users, from IT professionals to developers, to quickly diagnose bottlenecks and tune performance. Recommendations provide a level of observability into your applications and give you the guidance to correct potential issues before they become worse.

Increasingly, many organizations find that their current legacy monitoring solutions are no longer adequate in today's modern IT world. Big data application performance requires observability and continuous tuning to treat infrastructure and application performance as an integrated function, capturing and correlating the performance data from each domain to fully inform both development and operations teams.