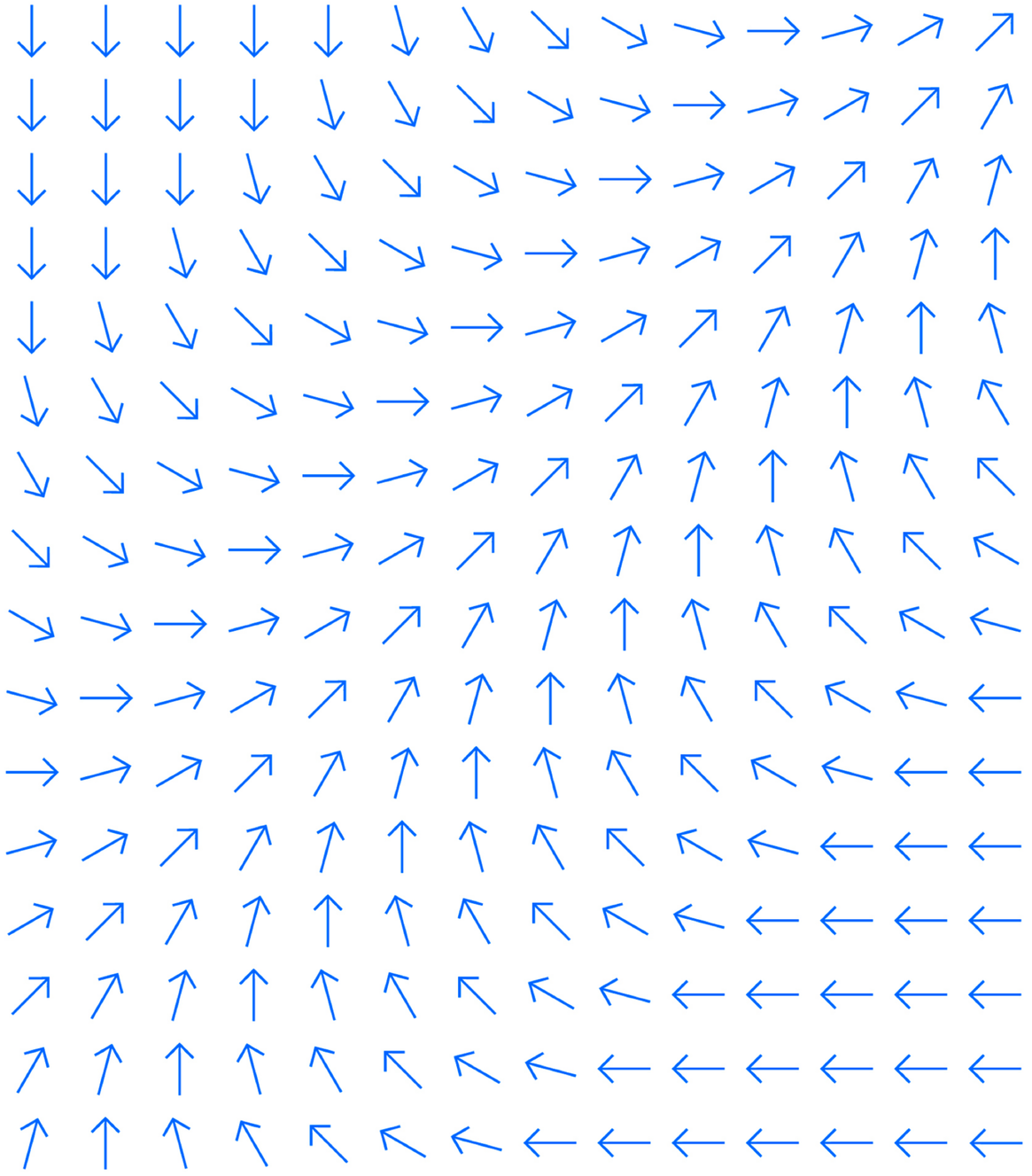
Continuous, Cost-Effective Application Performance

IBM Turbonomic

Dynamic IT resource management for continuous hybrid cloud optimization



IBM

As multicloud adoption accelerates... complexity and cost are increasing

More clouds

93% of organizations have a multicloud strategy.*

More workloads

In 2024, world-wide spending on public cloud services is forecast to grow 20.4% to $678 billion.**

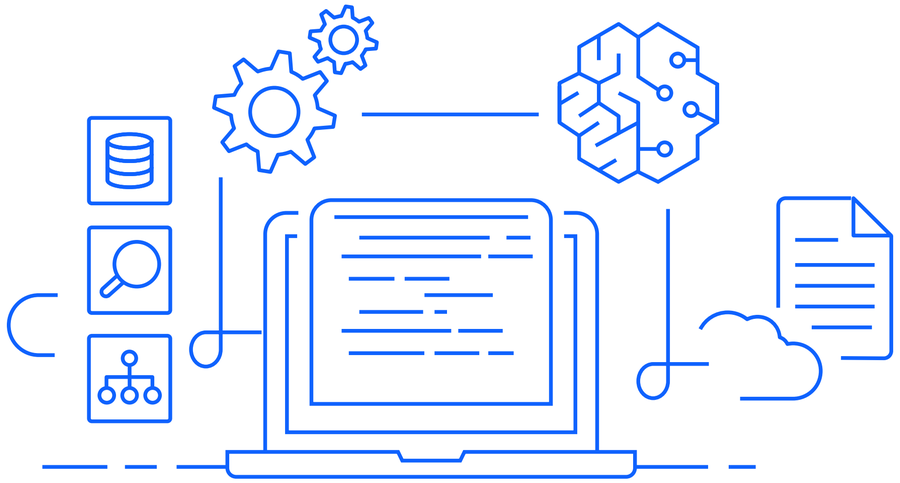
More cloud waste

Executives estimate that at least 30% of their cloud spending is wasted.*

More uncertainty

7 in 10 organizations have experienced higher-than expected cloud costs, demonstrating the need for effective cloud planning.***

This complexity brings new challenges



HOW do I?

- Assure application performance to meet customer demand?
- Optimize my environment to improve efficiency without risking performance?
- Prevent bottlenecks and correct issues before they happen and end the firefighting?
- Enable elasticity to meet the ever-changing demands of the business.

#1 Job for IT is delivering and running applications that help the business meet business objectives

To enable successful business outcomes and customer experiences, IT teams must consider:

Defensible IT Spending

Always know exactly what is needed to spend to deliver on IT business objectives with proof that IT is not spending more than the need.

Automation to protect the business

Spending less is meaningless if it puts the business at risk. Organizations need automated elasticity to continuously assures application performance at the lowest cost.

Future Proof Strategy

Future proof Kubernetes and cloud application resourcing to evolve and grow with the changing needs of the business. Ability to match supply and demand across hybrid cloud, multicloud, edge and beyond.

IBM Turbonomic

247% ROI and <6 months payback

Hybrid Cloud Optimization

Manage, automate and continuously optimize your hybrid cloud

1 Understand Performance Needs

2 Identify Waste

3 Automate Optimization

Hybrid cloud

Multicloud

PaaS/containers

Get 30% of your engineering time back.

Assure application performance.

Improve cloud investments by 35%

Reduce data center refresh costs by 75%

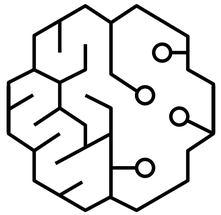
Use less carbon by spending less on IT

IBM Turbonomic Key Features

Full-stack visualization

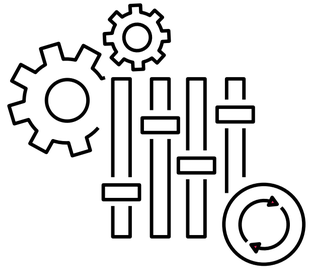
Complete visualization of your applications and infrastructure to stop guessing and start knowing the health of your environment

Analytical insights



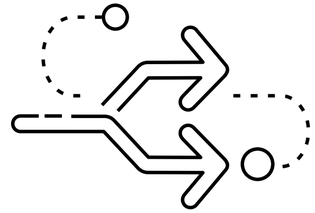
Patented and predictive analytics provide insights that empower teams to make informed decisions to enhance efficiency without sacrificing performance

Operationalize elasticity



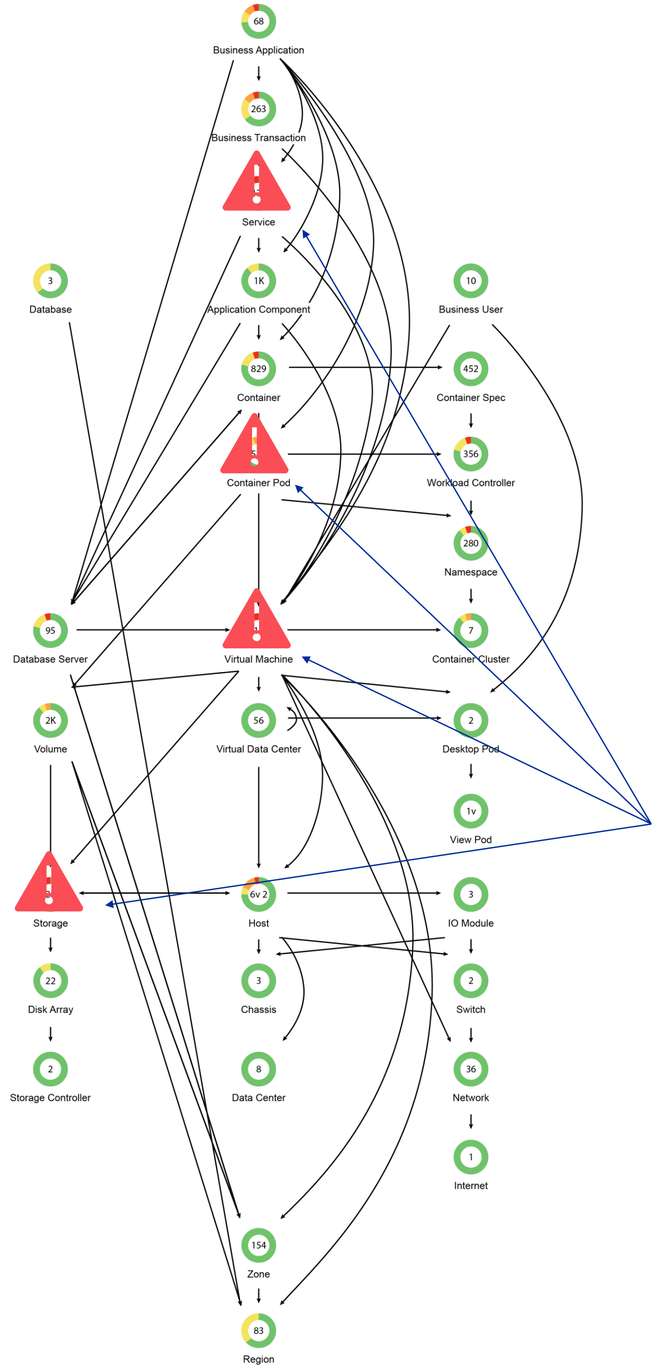
Operationalize elasticity to size up or down on demand to assure application performance at the lowest cost

Comprehensive integrations

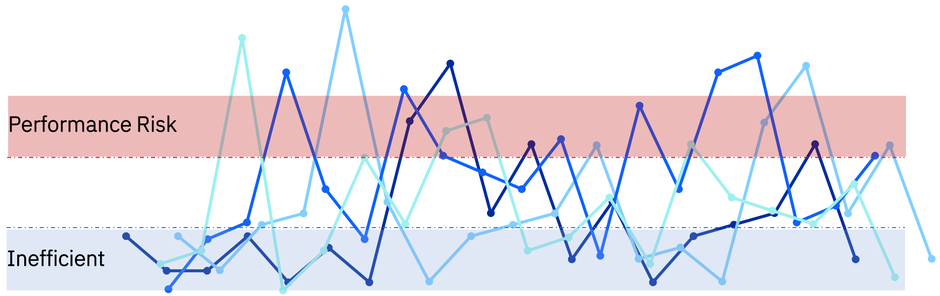


Seamlessly integrate with your organizations existing tools, enabling rapid time to value

The application stack



Driven by ever changing demand



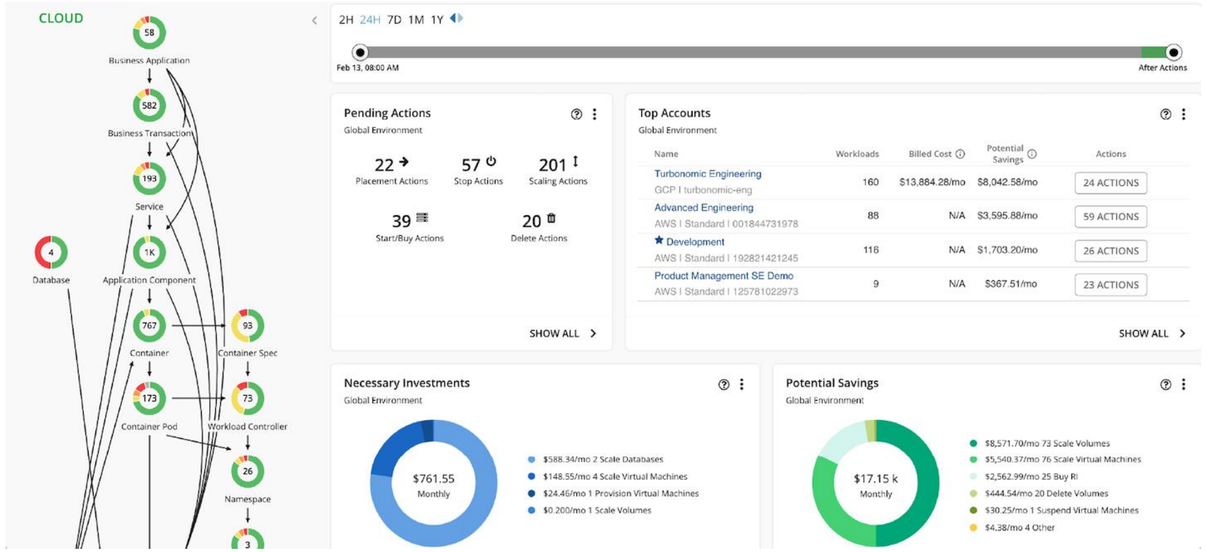
Transactions | Response Time | Memory | CPU | ...

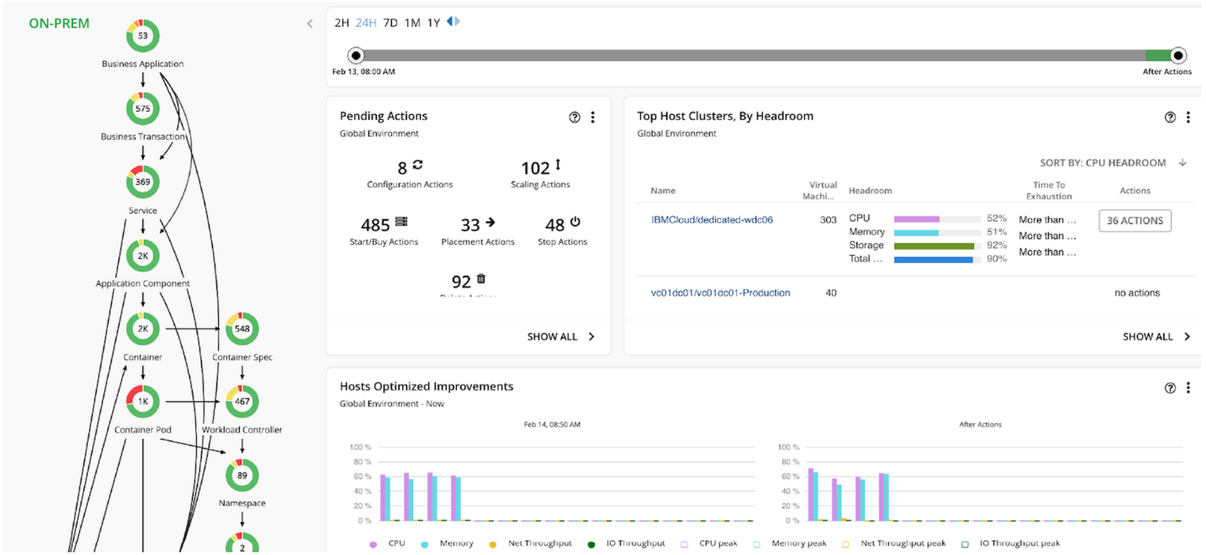
Continuous performance and the lowest cost, automatically

Generate actions that prevent problems, not just alerts…

With Turbonomic, applications get exactly the resources they need when they need them, continuously and automatically.

- App-first, demand-based analysis ensures actions can be safely automated.
- Automated actions prevent resource congestion (and cost overruns) across the hybrid cloud environment.





IBM Turbonomic Use Cases

- Cloud Cost Optimization
- Kubernetes Resource Optimization
- GPU Optimization
- Data Center Modernization
- Cloud Migration Planning
- VMware Optimization

Contain cloud spend while preserving application performance with IBM Turbonomic

IBM Turbonomic's AI-powered automation and cloud optimization solutions can continuously help ensure application performance and optimize costs.

With Turbonomic your cloud investments are smarter, automation is easier, and you can finally realize the promise of agility, elasticity and speed to market.

Make informed cloud management decisions

Turbonomic identifies cost-saving opportunities by analyzing resource usage, identifying underutilized resources and generating right-sizing actions continuously.

Enable a self-optimizing hybrid cloud

Optimize across public, private, hybrid and multicloud – from your own data centers to AWS, Azure and Google Cloud within a single instance.

Continuous optimization with Automation

Automate processes and eliminate guess work in resourcing applications, freeing up IT teams from repetitive tasks and enabling them to focus on innovation.

Enhance performance while controlling cost

Understand the interdependencies between applications and infrastructure while proactively identifying bottlenecks. Turbonomic can take optimization actions to enhance performance and reduce cost.

Effortlessly integrate with your cloud tools

Turbonomic supports a wide range of software tools and platforms to extend your ecosystem. Connect to public cloud tooling to optimize your environment, including reserved instances inventory.

Use Kubernetes resource optimization to align metrics and mission

IBM Turbonomic automatically determines the right resource allocation actions—and when to make them—to help ensure your Kubernetes and Red Hat OpenShift environments and mission-critical apps get exactly what they need to meet your SLOs.

Assure Performance for Kubernetes applications

Turbonomic proactively identifies and resolves bottlenecks, ensuring high performance and availability for your Kubernetes applications.

Real-time resource optimization

Continuously analyze application demand and automate resource allocation to ensure optimal performance at the lowest cost.

Reduce capacity spend

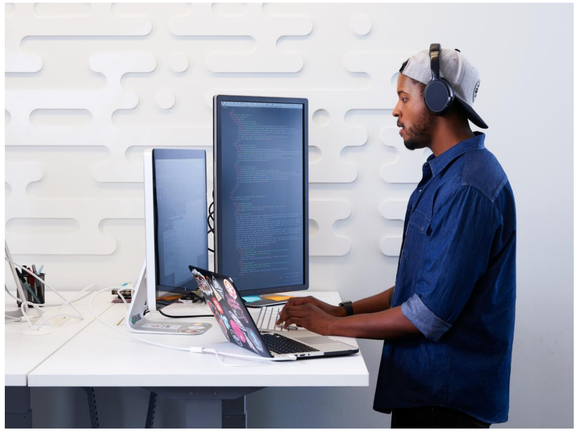
Turbonomic automatically sizes containers, moves pods and scales clusters based on application demand helping to ensure cost efficiency.

Automation for continued performance at the lowest cost

Turbonomic can automate resource allocation to meet demand as it happens, ensuring performance and cost efficiency.

Effortlessly integrate with your current tools

Turbonomic supports a wide range of software tools and platforms to extend your ecosystem. Turbonomic seamlessly and agentlessly integrates with your legacy and next-generation solutions to improve time to value.



GPU optimization as GPUs are expensive

IBM Turbonomic automatically determines the right GPU optimization actions—and when to make them—to help ensure your applications are getting the most out of the GPUs in your environments (Cloud, On-Prem and Kubernetes) as they are expensive, why pay for un-used GPU capacity when you don't have to.

Assure Performance for Gen AI LLM Inference workloads in Kubernetes

Turbonomic identifies GPU Queuing, Service and Transactions to reduce GPU Response Time bottlenecks, ensuring high performance and availability for your Gen AI LLL Inference applications.

Real-time GPU resource optimization

Continuously analyze GPU metrics from NVIDIA Data Center GPU Manager (DCGM) in IaaS VMs in the Cloud and in Kubernetes, Red Hat OpenShift and OpenShift AI to generate GPU specific actions.

Reduce capacity spend

Turbonomic automatically horizontally scales container pods, resizes IaaS VMs and moves VMware VMs to use only the required GPUs needed to based on application demand helping to ensure cost efficiency.

Automation for continued performance at the lowest cost

Turbonomic can automate actions specific to the GPU metrics discovered in different environments to meet demand as it happens, ensuring performance and cost efficiency.

Effortlessly integrate with your current tools

Turbonomic supports a wide range of software tools and platforms to extend your ecosystem. Turbonomic seamlessly integrates with your legacy and next-generation solutions to improve time to value.

Get ahead of the growing digital landscape with a modern data center

IBM Turbonomic can help you reimagine the data center as the next-gen hero of your ecosystem.

Turbonomic software automatically optimizes your applications' resourcing levels while dynamically scaling with business needs in real time.

With a modernization strategy you get a more efficient data center and better business outcomes.

Consolidate with confidence with refresh planning

Turbonomic planning helps organizations understand what hardware they can or should keep.

Put data center uptime on autopilot

Proactively identify bottlenecks and automate corrective actions to enhance performance and availability

Automate real-time IT resource optimization

Turbonomic eliminates over-provisioning and automate rightsizing operations aligned to your business objectives.

Full-Stack Visualization

Turbonomic provides a complete understanding of the application stack gives you the confidence to trust Turbonomic's recommendations and automate actions to improve performance without disruption..

Effortlessly integrate with your current tools

Turbonomic supports a wide range of software tools and platforms to extend your ecosystem. Turbonomic seamlessly and agentlessly integrates with your legacy and next-generation data center technologies

Cloud migration planning with IBM Turbonomic

IBM Turbonomic provides easy-to-use cloud migration planning that allows you to optimize your cloud consumption from the start and simplify your cloud migration process.

Whether your organization is pursuing a cloud-first, hybrid cloud or multicloud strategy, Turbonomic can deliver convenient cloud migration planning tools that can accelerate your digital transformation.

Migration plan optimization

Turbonomic analyzes current application performance and can deliver potential app migration optimization plans to avoid expensive lift and shift migrations.

Discount aware actions

Examines billing and price adjustments negotiated with your cloud provider and creates migration plans based that account for discounted pricing.

Guided workflows

Turbonomic provides a guided workflow to help you configure different scenarios for your team's dynamic cloud use case, then creates a migration plan in minutes.

“What if” planning scenarios

Turbonomic enables users to experiment with “what if” scenarios based on your entire environment and evaluate migrations strategies across AWS, Azure and GCP.

Streamline cloud migration costs through planning

Identify cost-saving opportunities or required investments during the migration process and contain unexpected costs through continuous optimization.

Optimize VMware Environments for Performance and Efficiency

If your VMware infrastructure costs are up, utilization is down, and overprovisioning is sideways, it's time to rethink your data center strategy.

Reduce by Increasing Utilization

IBM Turbonomic can help customers safely increase utilization across their environment through optimization and reduce the need to spend on VMware licensing and hardware. **Here's how.**

VM Rightsizing

Turbonomic takes rightsizing actions to safely increase density and avoid resource contention from oversized allocation. When appropriately sizing virtual machines (VMs), Turbonomic software considers VM virtual memory, CPU, and as demand across the stack.

Cross Cluster Optimization

Turbonomic enables you to create "super clusters," or virtual pools that unlock the total cumulative resource amount by allowing workloads to move between clusters and vCenters when demand increases. This delivers more elasticity, better performance and improved cloud economics.

Continuous compute placement

Turbonomic continuously analyzes app demand and resource supply levels. It generates automatable actions that reduce the potential for resource congestion, reduce the impact of latency, and safely use all hardware without putting performance at risk.

Capacity Management

Turbonomic uses an accurate view of the real-time environment to simulate changes you define, such as how much physical infrastructure is needed to support growth, migrate workloads, and shut down what isn't needed.

Gain Flexibility with Turbonomic

IBM Turbonomic helps you gain flexibility from reliance on VMware and enables your organization to reduce spend and better negotiate future VMware services. Turbonomic increases performance, improves efficiency, and frees you to move to other hypervisors or public clouds.

Benefits of IBM Turbonomic

Assure application performance

Intelligent automation proactively identifies bottlenecks and takes corrective actions to enhance performance and availability.

Optimize IT resources in real time

GPU and Application-aware analytics dynamically optimize resources ensuring that workloads receive the right amount of workload resources without over or under-provisioning.

Optimize costs

Automate processes and eliminate guess work in resourcing applications, freeing up IT teams to focus on strategic priorities.

Align infrastructure and application teams

Unified data model aligns infrastructure, application, and DevOps teams around a single view of application and infrastructure health in real time and over time.

Manage hybrid and multicloud

Integrations powered by a patented abstraction enable self-optimizing resource management from your own private data centers to AWS, Azure, and Google Cloud

Integrate seamlessly

Turbonomic seamlessly integrates with your organization’s existing tools enabling rapid time to value.

© 2024 IBM Corporation

The IBM Turbonomic Difference

Efficiency that never risks performance.

Application-aware, full-stack analytics ensure automation drives efficiency but never risks performance.

Automation that Engineering, Ops, and App teams can trust.

Engineering, Ops, and App teams always get the data they need to confidently execute actions, then adopt automation at scale.

Integrations that ensure operationalization.

Automation can be integrated with pipelines, processes, and workflows. Customizable policies ensure business context mitigates risk.

One platform, many ways to optimize for cost and performance.

Supports cloud cost optimization across compute, volumes, database, Kubernetes, including RIs/Savings Plans/discounts.

A platform that is future-proof.

Analytics can be extended to future tech because it is grounded in the fundamentals of IT, matching demand to supply across multicloud and beyond.

It’s time you Turbo

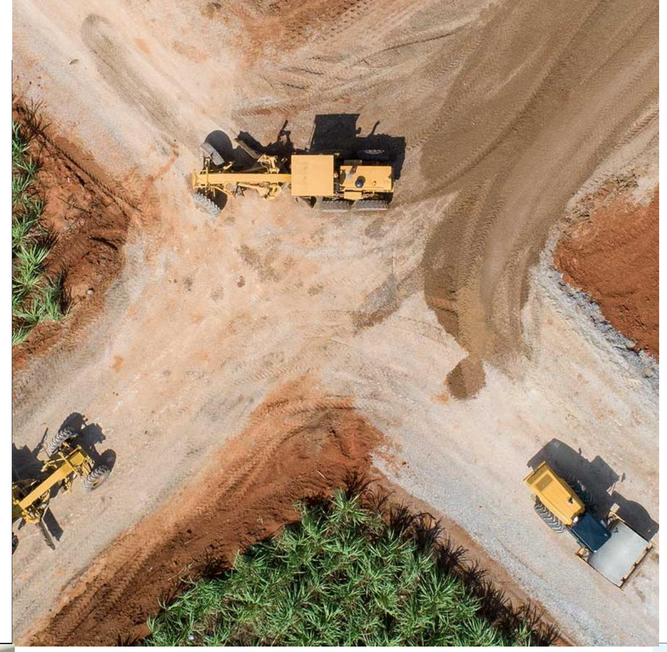
Start your journey to continuous application performance while safely reducing cost. Try Turbonomic today.

© 2024 IBM Corporation

KOMATSU

How does this multinational manufacturing firm use intelligent automation help assure application performance?

Read the full story



“In our organization, optimizing application performance is a continuous process that is beyond human scale. We see tremendous value in IBM Turbonomic to help us close that gap as much as possible within our organization.”

Matthew Koozer
Cloud Engineer, Komatsu Ltd.

Like many other IT organizations, the infrastructure team at Komatsu Ltd. historically relied on disparate monitoring tools, user complaints and manual intervention to address performance issues. Once they implemented the IBM Turbonomic hybrid cloud cost optimization solution, the team was finally able to break this cycle of fighting fires. Armed with Turbonomic’s full-stack visibility, the team at Komatsu is now able to quickly assess potential performance risk and identify the optimal way to reallocate resources to eliminate that risk without compromising performance in another layer of their technology stack.

- Reduced cloud spend by over USD 650,000
- Achieved 33% savings on server run rates
- Reduced user complaints to 10 tickets per year

Solution components:

IBM® Turbonomic®

DENVER
THE MILE HIGH CITY

How does Denver assure the performance of the various applications its citizens and employees rely on every day?

Read the full story

IBM Turbonomic gave us the data we needed to show the broader team that overprovisioning would not only fail to assure performance, but actually put performance at risk.

Nick Steensland
Service Delivery Manager,
City and County of Denver, CO

Historically, the City of Denver's IT Operations team used disparate monitoring tools and manual intervention to address performance issue. There was a misconception that overprovisioning could assure performance. Once they installed the IBM Turbonomic solution, the IT Operations team finally had a single source of truth to understand how their environment was performing. They had concrete evidence to show app owners where their VMs were overprovisioned. This allowed them to reallocate resources to improve performance and efficiency.

- 33% reduction in CPU and RAM usage
- 24x7 performance assurance of 1,500 VMs

"We chose IBM Turbonomic because we were looking for a common language that everyone in the organization could rely on as a source of truth for what we need to right-size our environments."

Nick Steensland
Service Delivery Manager, City and County of Denver, CO

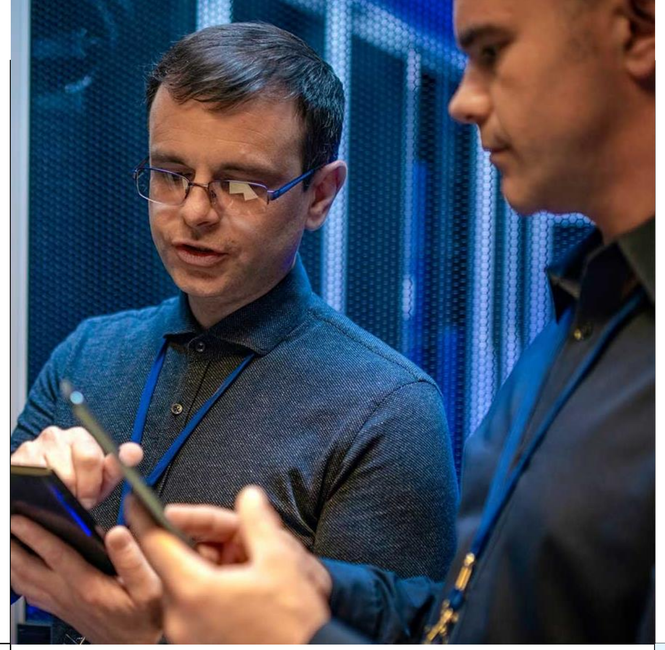
Solution components:

IBM® Turbonomic®

BlueIT

How does BlueIT use Turbonomic and Instana to assure application performance *and* cut carbon emissions?

Read the full story



"The sustainability paradigm in IT rests on an organization's ability to achieve the right balance between resource allocation and end-user experience. In my opinion, implementing Turbonomic and Instana is without a doubt the best way to right-size your infrastructure and reduce your carbon footprint without sacrificing performance."

Francesco Sartini
Chief Innovation Officer, BlueIT

Given the heterogeneity of their clients' environments, it is critical that BlueIT Chief Innovation Officer Francesco Sartini and his team have tools to efficiently onboard each new client and continuously optimize application performance moving forward. Ease of deployment and time to value are key.

"We have found that the initial setup of Turbonomic and Instana is simple and immediate"

Francesco Sartini,
Chief Innovation Officer, Blue IT

In a single day, both tools can be deployed and start proactively identifying resizing opportunities. With the help of Turbonomic and Instana Observability, BlueIT has achieved a ***50% reduction in MTTR*** and a ***60% reduction in the time required to implement a resourcing action***, while also reducing waste across their clients' environments.

For example, one BlueIT client achieved a ***10% reduction in memory and CPU over-allocation*** after executing Turbonomic's AI-powered resourcing recommendations.

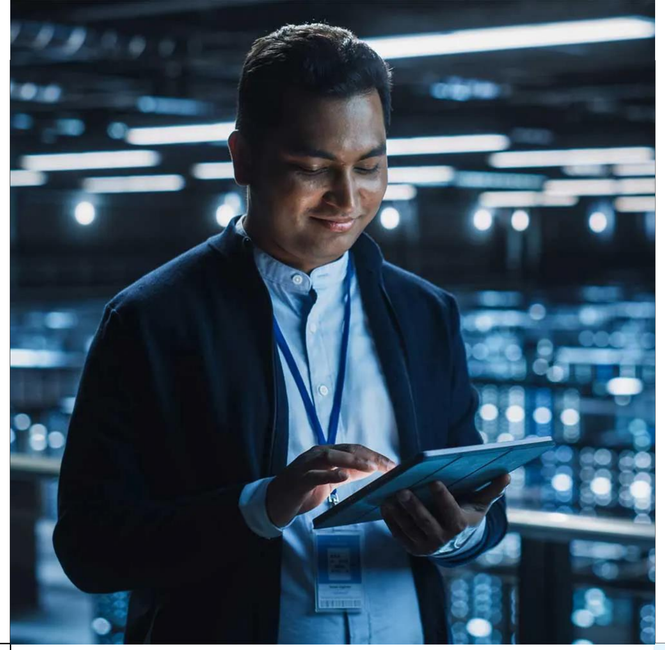
- With AIOps, the team is executing resourcing decisions 60% faster
- Achieved 10% reduction in memory and CPU over-allocation for one client
- Reduced MTTR by 50% across the organization

Solution components:

IBM® Instana® Observability
IBM Turbonomic®

How does an IT team supporting over 280,000 users all over the world continuously deliver performant IT services?

Read the full story



“When you look at the complexity of our environment and the impact Turbonomic has had, you can’t count the amount of time that we’ve saved by implementing Turbonomic automated resourcing actions because we could not execute that volume of actions without Turbonomic automation.”

Matt Lyteson
VP, CIO Hybrid Cloud Platforms, IBM

The IBM CIO Organization began adopting Red Hat OpenShift in 2019, as part of an overarching enterprise platform-oriented approach across IBM Z, IBM Power, x86 and container architectures. This meant adopting standard platforms at enterprise scale to increase velocity. With OpenShift, the migration to a large multi-tenant environment also made it increasingly difficult to rely on their existing tools and processes to manage capacity and maintain optimal application density.

Observability with IBM Instana

- 22 tenants and growing
- 1300+ hosts being monitored multiplatform with IBM Power, IBM Z, and Intel
- 370 critical applications observed in context
- 80+ websites observed with RUM
- OpenTelemetry enabled.

Dynamic resourcing with IBM Turbonomic

- 3.8 TB decrease in cumulative memory limits
- 64% decrease in CPU requests
- 45,000 automated resourcing actions per month, delivering 24x7 performance assurance

Solution components:

IBM® Instana® Observability
IBM Turbonomic®

How does IBM optimize GPUs for efficiency and performance for Gen AI LLM Inference workloads?

“Enabling Turbonomic to scale up and down our LLM inference servers has allowed me to spend less time monitoring performance.

Tom Morris
Infra and Ops Lead for AI Platform Research, IBM

[Read the full story]

The IBM® Big AI Models (BAM) team, that supports the primary research and development environment for engineering teams to test and refine their gen AI projects, saw an opportunity for improvement. As more projects went through the testing stage, the team recognized the importance of optimally using each instance to avoid resource wasting.

To optimize their GPU resources and manage their LLM Kubernetes instances, the IBM BAM team deployed IBM Turbonomic®, an advanced application resource management software.

As an IBM software explicitly tailored for optimizing hybrid cloud, including containerized applications, virtual machines and public clouds, IBM Turbonomic provided seamless integration within the existing infrastructure.

Dynamic GPU resourcing in Kubernetes with IBM Turbonomic

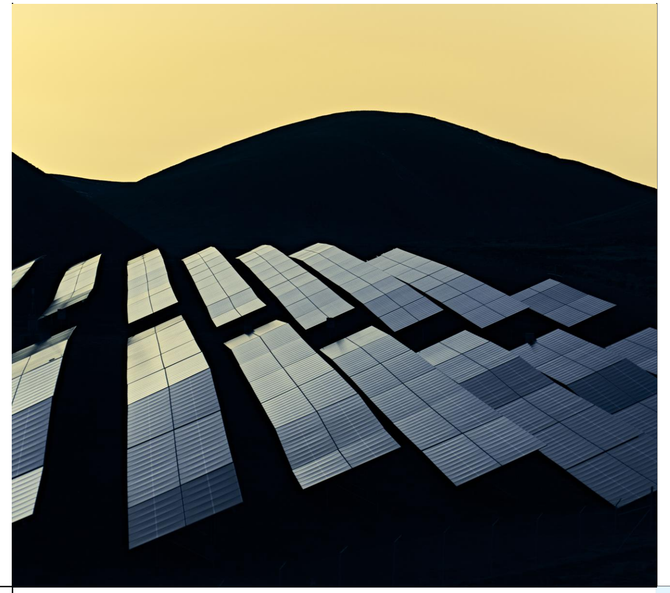
- 5.3x increase in idle GPU resources – making them available for new demand
- 2x GPU throughput without degrading the latency performance
- 960+ automated resourcing actions taken on the AI Inferencing workloads

Solution components:

IBM Turbonomic®

How does IBM Hursley implement automation in its journey to carbon neutrality?

Read the full story



IBM Turbonomic can help you make sure you are fully utilizing and optimizing your infrastructure to simultaneously assure performance and minimize cost (both financially and environmentally).

Colin Holyoake
Certified Datacenter Design and Sustainability Manager, IBM Hursley

IBM's goal is to achieve net zero greenhouse gas emissions by the year 2030. The IBM Turbonomic solution helps the IBM Hursley team identify opportunities to rebalance resources across the IBM Hursley data center and maximize their utilization of their existing infrastructure without sacrificing performance. By extending the use of their existing infrastructure, the IBM Hursley team avoids unnecessary expenditures on new hardware as well as the operational expense to power and cool it.

Dynamic resourcing with IBM Turbonomic

- Reduces performance risk across the data center
- Extends the lifespan of existing infrastructure and reduces environmental impact

With the help of Turbonomic, optimized
6,000
virtual systems

Those who rely on IBM Hursley's data center include
11,000
developers

Solution components:

IBM Turbonomic®
IBM® Instana® Observability