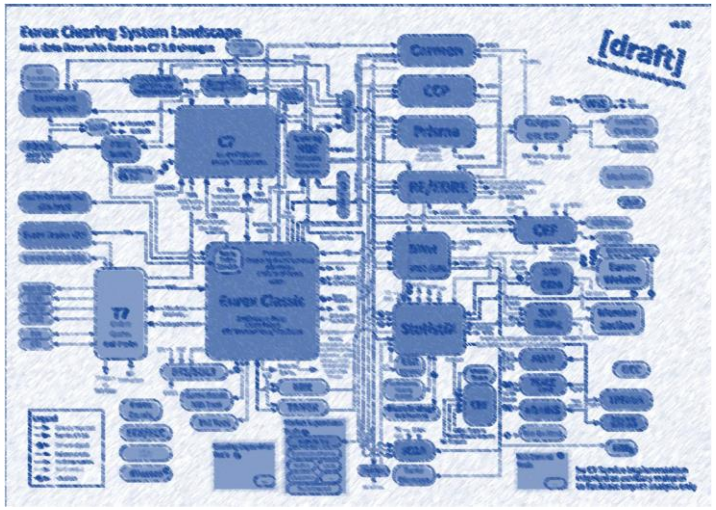
DEUTSCHE BÖRSE GROUP

Cloud Resource Optimzation using AI OPS

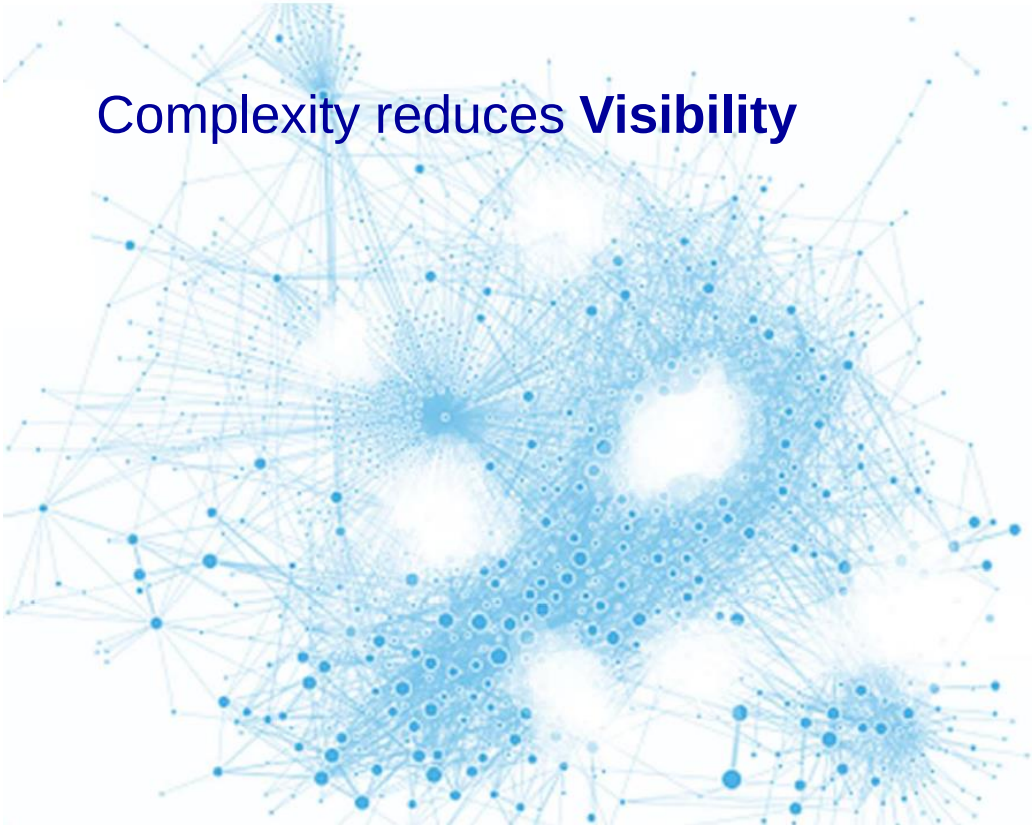
Cloud resource observability and optimization with smart Machine Learning AI

9/27/2024

Applications run well for the following 2 reasons



Microservices increase **Complexity**



Complexity reduces **Visibility**

Applications run well for the following 2 reasons

The developer writes good code.

INSTANA

Applications get the resources they need, when they need them.

IBM Turbonomic

Resource congenstion as a cause of poor application performance

The VM may not be receiving sufficient network throughput from the host.

Memory latency may occur with the VM.

The container POD suffers from CPU throtteling due to too low CPU limits.

The cloud instance type may not provide enough IOPS.

The Java heap might be exhausted.

The datastore may not be getting enough IOPS from the storage array.

A resource congestion anywhere in the stack can affect any application that depends on it.

The application may not be getting enough CPU from the VM.

The database connection pool may be exhausted.

How to keep IT continuously in the desired state:

What is the problem to be solved?

Ensure the application performance

while

IT infrastructure is used as efficiently as possible.

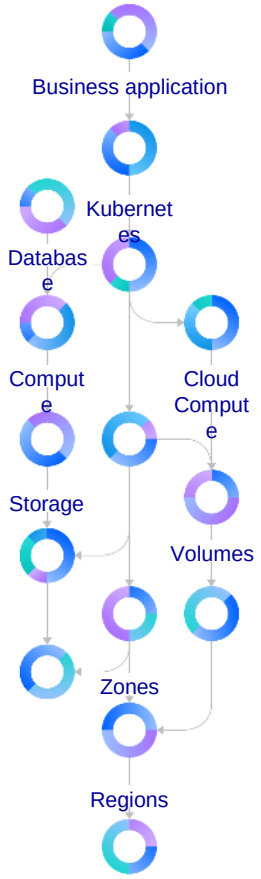
Data ↷

API-driven discovery

- Applications
- Databases
- Kubernetes
- Virtualizations
- Public Cloud
- Hyperconverged
- Compute
- Storage

Optimization ↷

Map applications to infrastructure…



…to generate trustworthy actions

- Intelligent sizing
- Continuous placement
- Dynamic scaling
- Start/stop
- What-If-Scenario Planning

Automation ↷

Execute & Integrate Actions

- Manually (with a click)
- Scheduled
- Pipelines & Workflows
- Real-time

Initiatives & Goals

Addressing Current DBG Initiatives

- Data Center Efficiency
  (Performance, Capacity Planning, cross Stack)
- Application Modernization
  (K8, OCP, Cloud Native)
- Multi-Cloud Transformation
  (Dyn. Cap.Mgmt., xTeam Visibility)
- FinOps
  (Cost Mgmt., License Compliance)
- Sustainable IT
  (e.g. reducing power consumption)
- Other (tbd)

Improved **capacity** management and planning for future hardware, software and platform changes, including multi-cloud transformation planning

Improve **Performance** Engineering, Stability and Resilience with performance assurance capabilities

Scalability through automation (built-in automatable actions with sufficient policy control features for DBG)

Eliminate complexity and manual processes where possible in the current operational process

Enable and future proof DBG's strategic decisions through a consistent and agnostic management platform (on-prem, public cloud, containerization)

Enable DBG to **optimize inefficiency & reclaim over-allocation** of resources

in various staging environments

Data ↷

API-driven discovery

DEUTSCHE BÖRSE GROUP

Applications

Databases

INSTANA

Kubernetes

RED HAT OPENSHIFT

Virtualizations

vmware

Public Cloud

aws

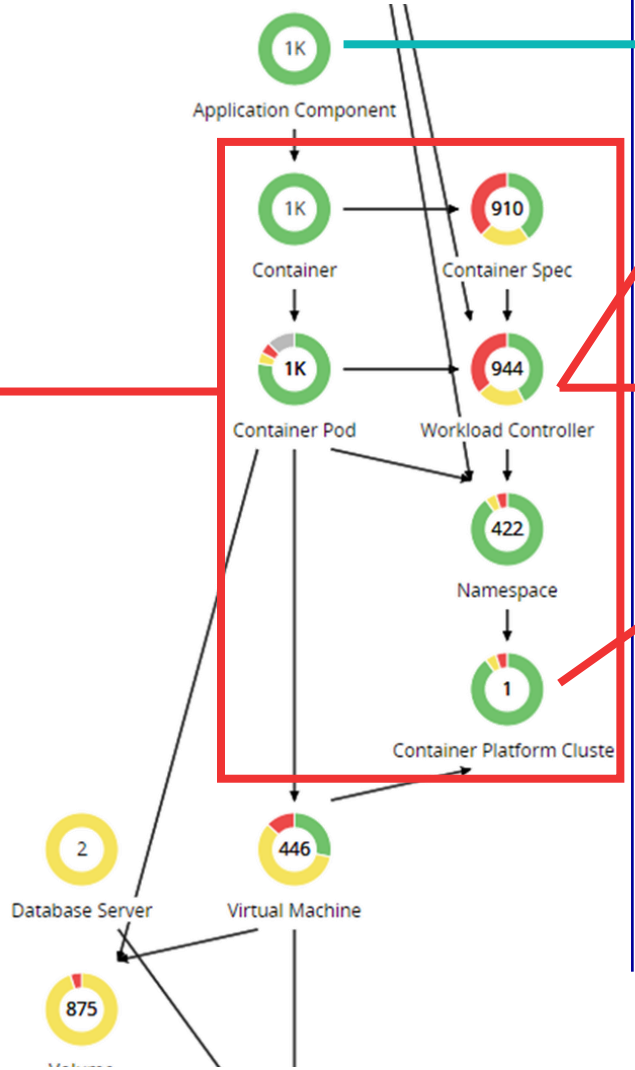
Hyperconverged

Compute

Storage

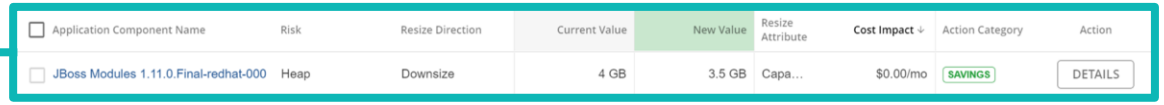
Optimization ↷

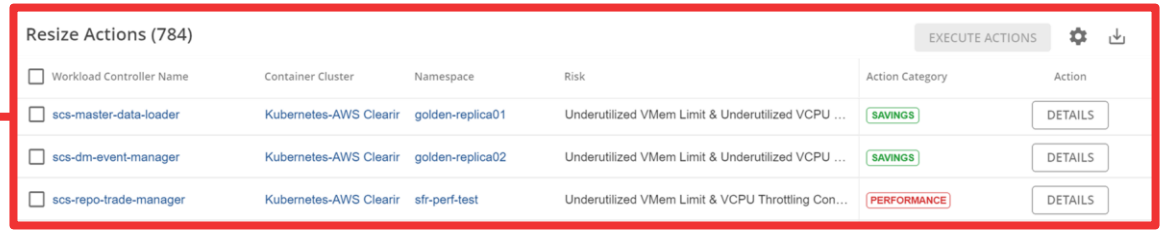
Map applications to infrastructure…

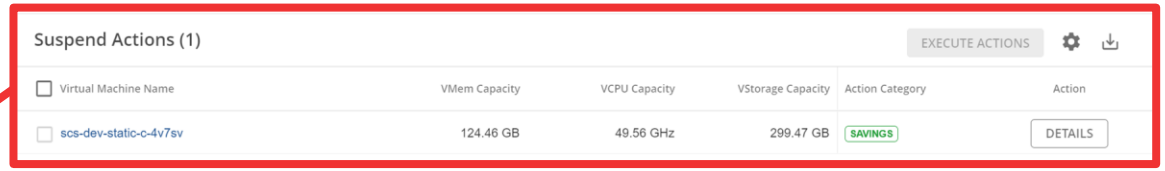


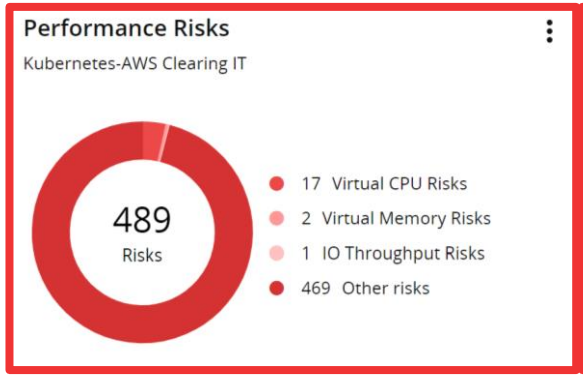
POV Extracts ↷

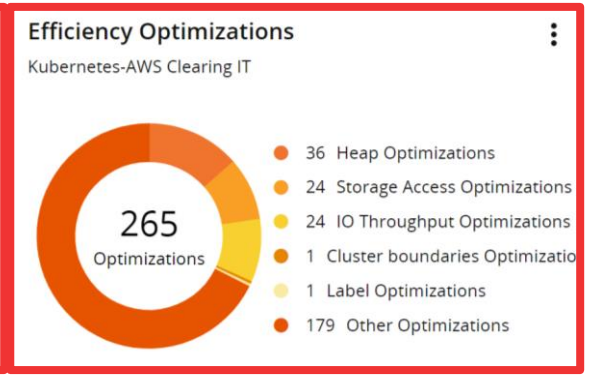
…to generate actions at all layers of the stack to optimize Kubernetes.











Data ↷

API-driven discovery

DEUTSCHE BÖRSE GROUP

Applications

Databases

INSTANA

Kubernetes

RED HAT OPENSHIFT

Virtualizations

vmware

Public Cloud

aws

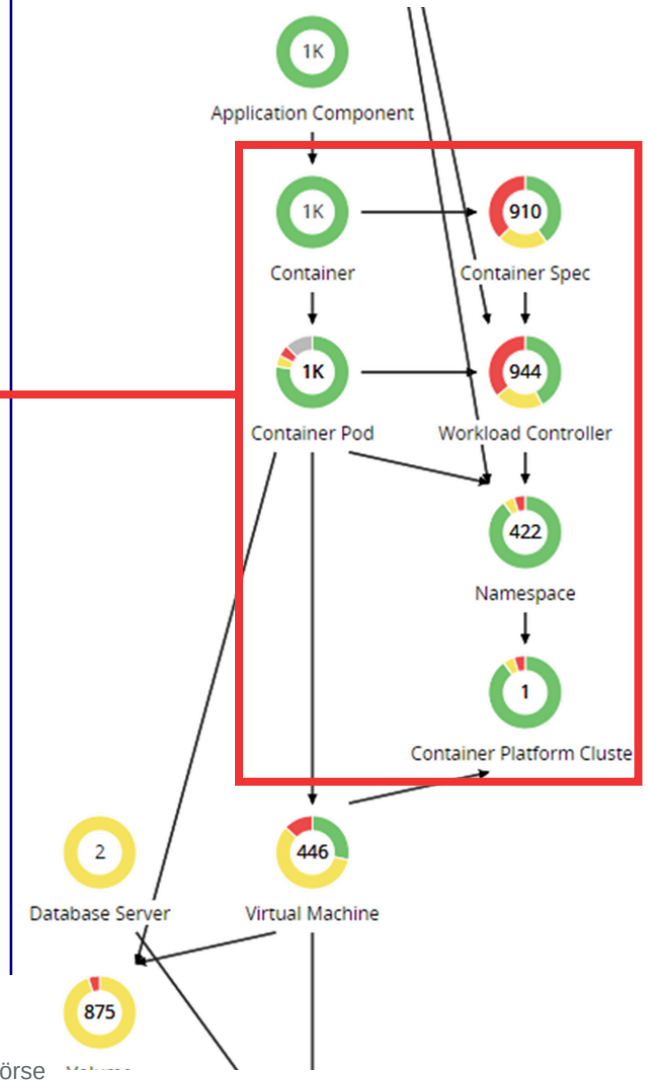
Hyperconverged

Compute

Storage

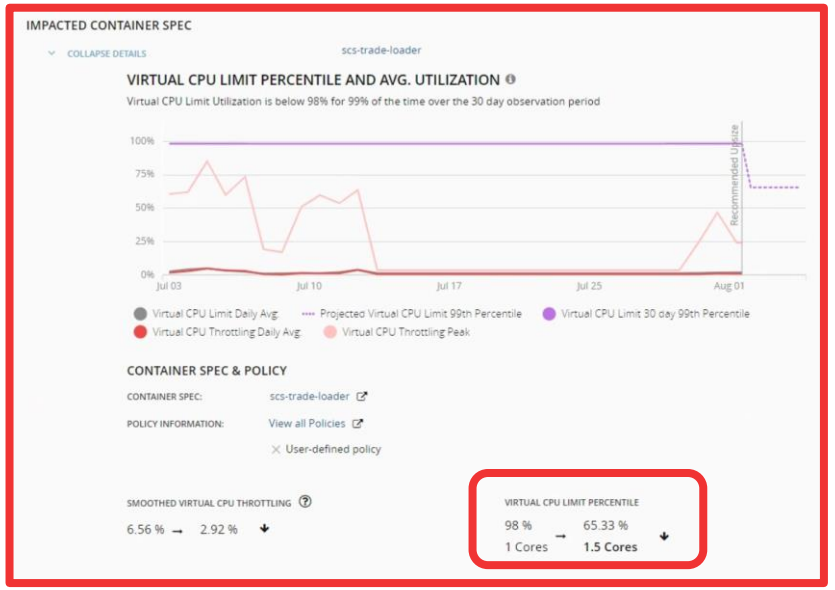
Optimization ↷

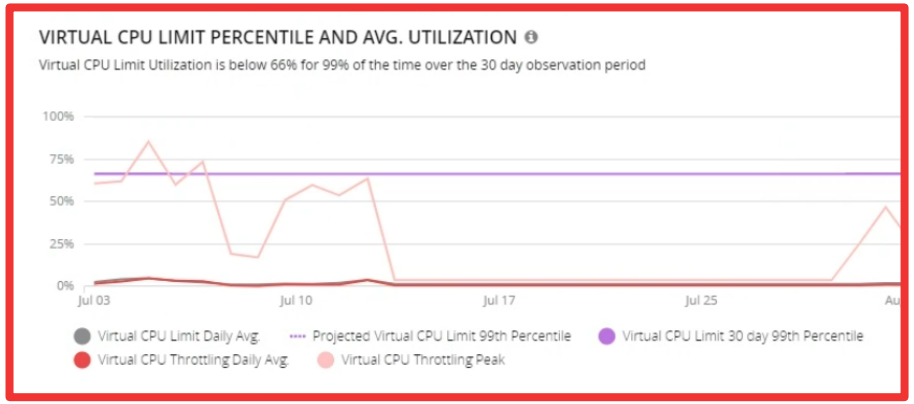
Map applications to infrastructure…



POV Extracts ↷

…to generate trustworthy actions.





Data ↷

API-driven discovery

DEUTSCHE BÖRSE GROUP

Applications

Databases

INSTANA

Kubernetes

RED HAT OPENSHIFT

Virtualizations

vmware

Public Cloud

aws

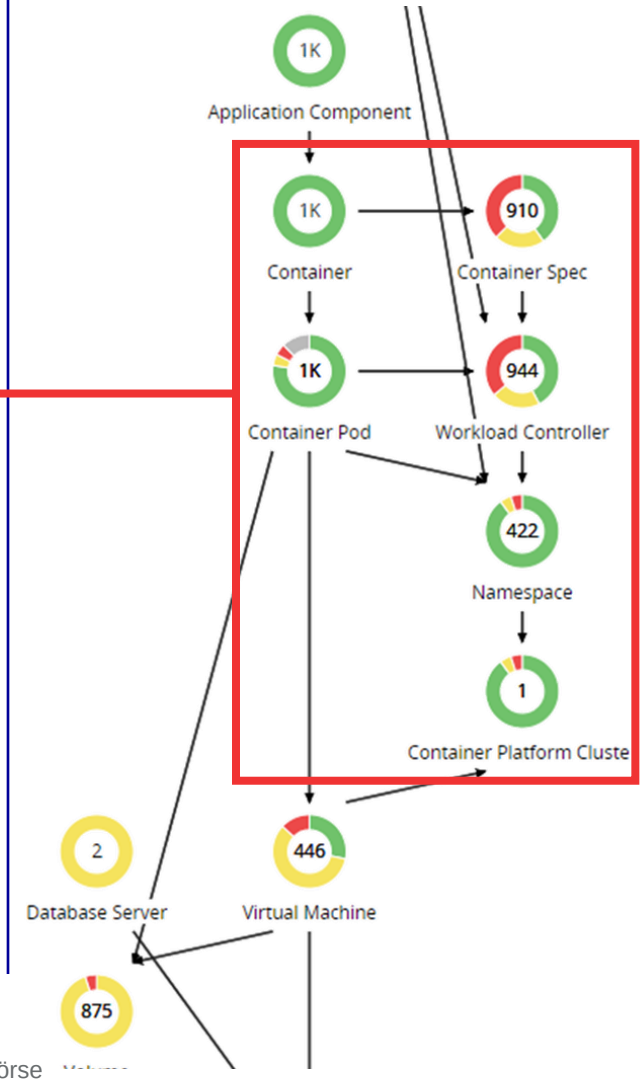
Hyperconverged

Compute

Storage

Optimization ↷

Map applications to infrastructure…



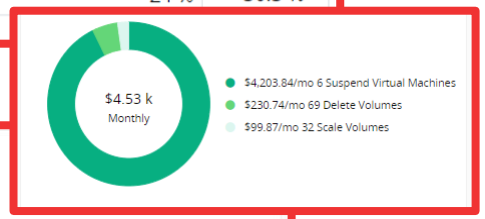
POV Extracts ↷

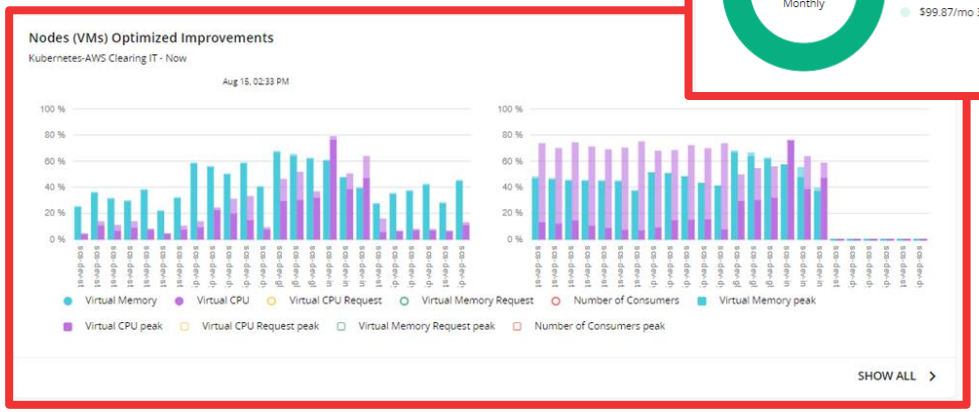
…to PLAN Kubernetes capacity.

Optimize Container Cluster Summary

Kubernetes-AWS Clearing IT

	Current	After Plan	Difference	%
Container Pods	1387	1297	90	▼6.5 %
Virtual Machines	24	18	6	▼25 %
Pod Density	57.8 : 1	72.1 : 1	14 : 1	▲24.7 %
Cluster CPU Capacity	324 Cores	228 Cores	96 Cores	▼29.6 %
Cluster CPU Allocatable	321.51 Cores	226.17 Cores	95.3 Cores	▼29.7 %
Cluster CPU Overcommitment	290.2 %	430.3 %	140.1 %	▲48.3 %
Cluster Memory Capacity	1.82 TB	1.33 TB	496 GB	▼26.7 %
Cluster Memory Allocatable	1.67 TB	1.23 TB	457.5 GB	▼26.7 %
Cluster Memory Overcommitment	79.3 %	55.3 %	24 %	▼30.3 %





Data ↷

API-driven discovery

DEUTSCHE BÖRSE GROUP

Applications

Databases

INSTANA

Kubernetes

RED HAT OPENSHIFT

Virtualizations

vmware

Public Cloud

aws

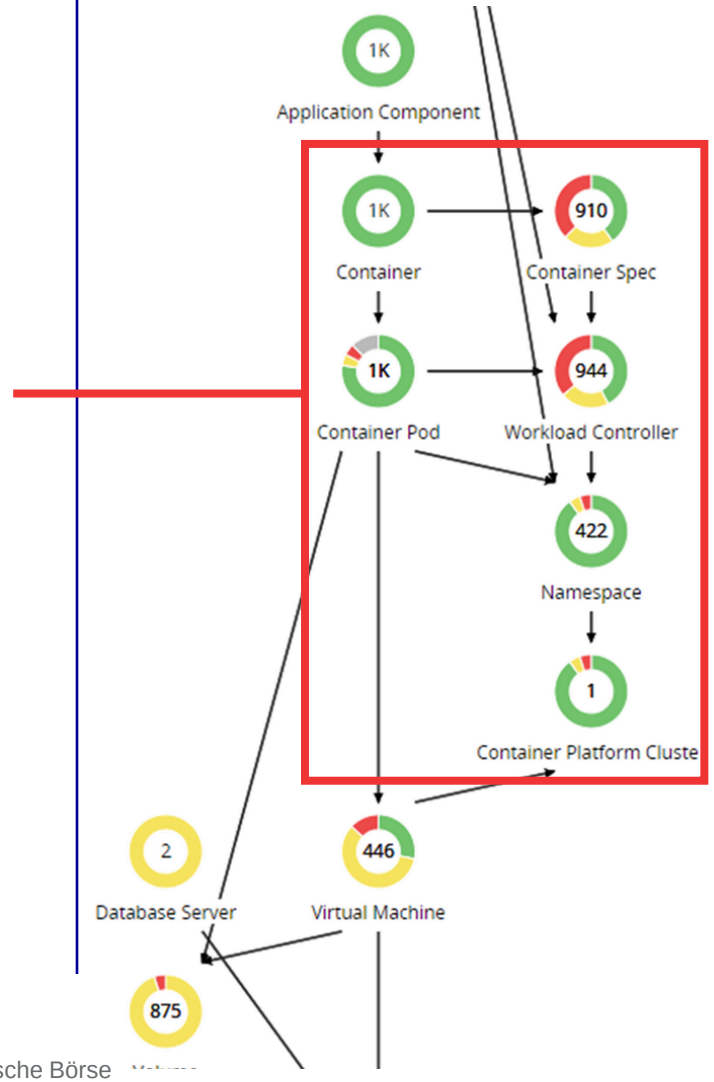
Hyperconverged

Compute

Storage

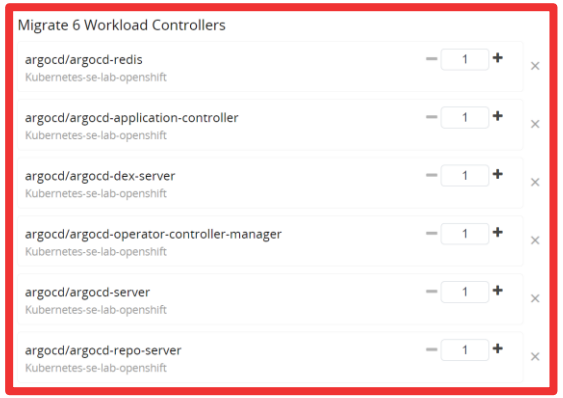
Optimization ↷

Map applications to infrastructure…



POV Extracts ↷

…to PLAN workload migrations.



DEMO DATA

Migrate Container Workloads Summary

Kubernetes-se-lab-aks

	BEFORE MIGRATION	LIFT & SHIFT	OPTIMIZED
Workload Controllers	34	40	40
Container Pods	90	103	75
Virtual Machines	7	8	4
Pod Density	12.9 : 1	12.9 : 1	18.8 : 1
Cluster CPU Capacity	20 Cores	24 Cores	16 Cores
Cluster CPU Allocatable	19.18 Cores	23.04 Cores	15.44 Cores
Cluster CPU Request	13.59 Cores	15.78 Cores	11.09 Cores
Cluster CPU Limit	39.4 Cores	43.9 Cores	39.01 Cores
Cluster CPU Overcommitment	197 %	182.9 %	243.8 %
Cluster Memory Capacity	62.26 GB	77.9 GB	62.54 GB
Cluster Memory Allocatable	45.34 GB	57.64 GB	49.21 GB
Cluster Memory Request	11.8 GB	14.74 GB	9.01 GB
Cluster Memory Limit	45.39 GB	48.61 GB	27.02 GB
Cluster Memory Overcommitment	72.9 %	62.4 %	43.2 %

Data ↷

API-driven discovery

DEUTSCHE BÖRSE GROUP

Applications

Databases

INSTANA

Kubernetes

RED HAT OPENSHIFT

Virtualizations

vmware

Public Cloud

aws

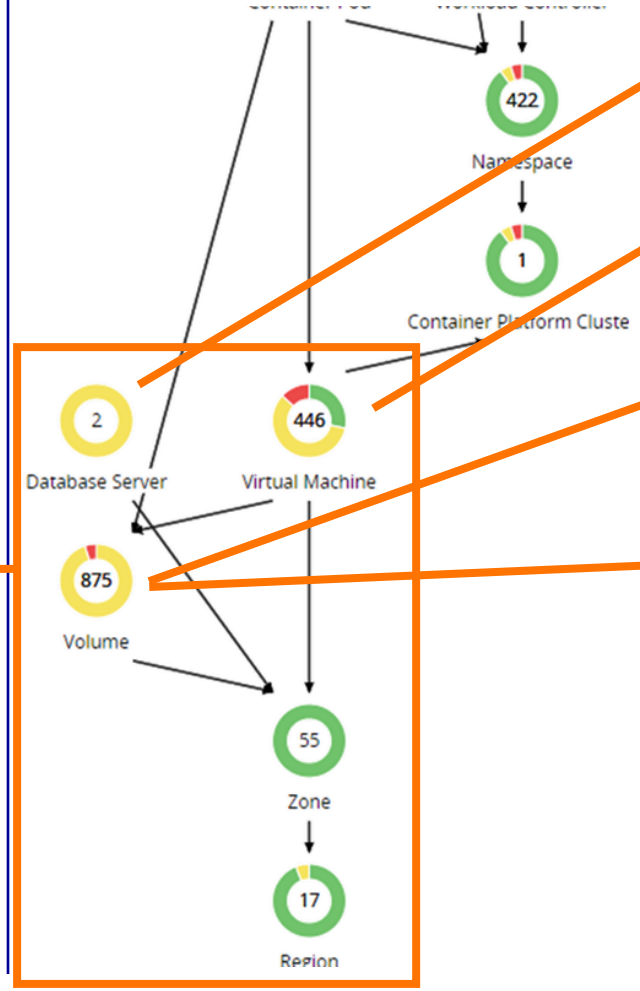
Hyperconverged

Compute

Storage

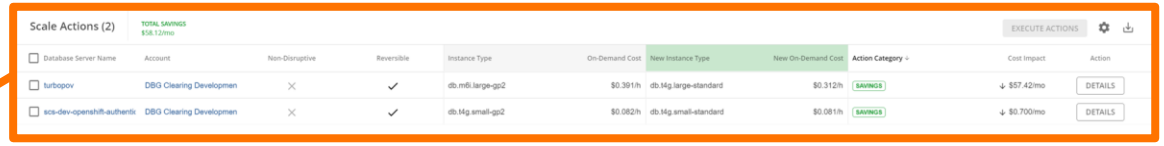
Optimization ↷

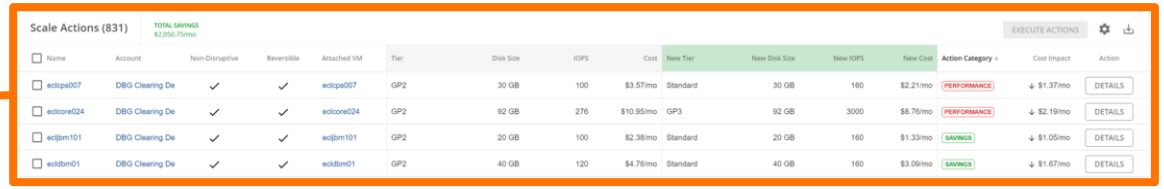
Map applications to infrastructure…

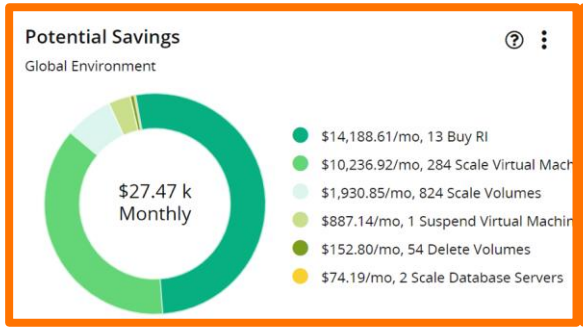


POV Extracts ↷

…to generate actions at all layers of the stack to optimize Public Cloud.







Data ↷

API-driven discovery

DEUTSCHE BÖRSE GROUP

Applications

Databases

INSTANA

Kubernetes

RED HAT OPENSHIFT

Virtualizations

vmware

Public Cloud

aws

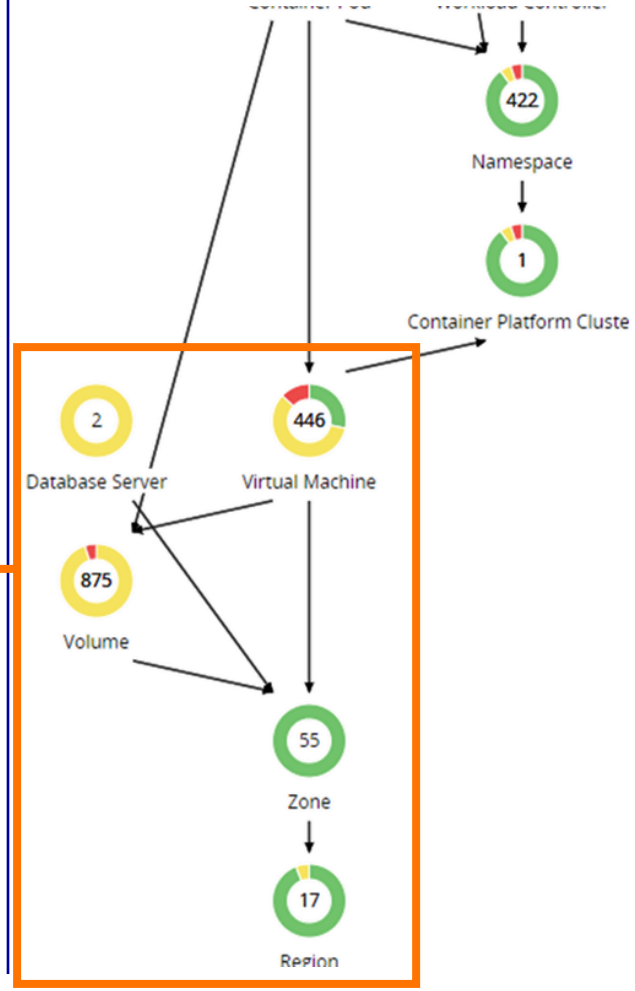
Hyperconverged

Compute

Storage

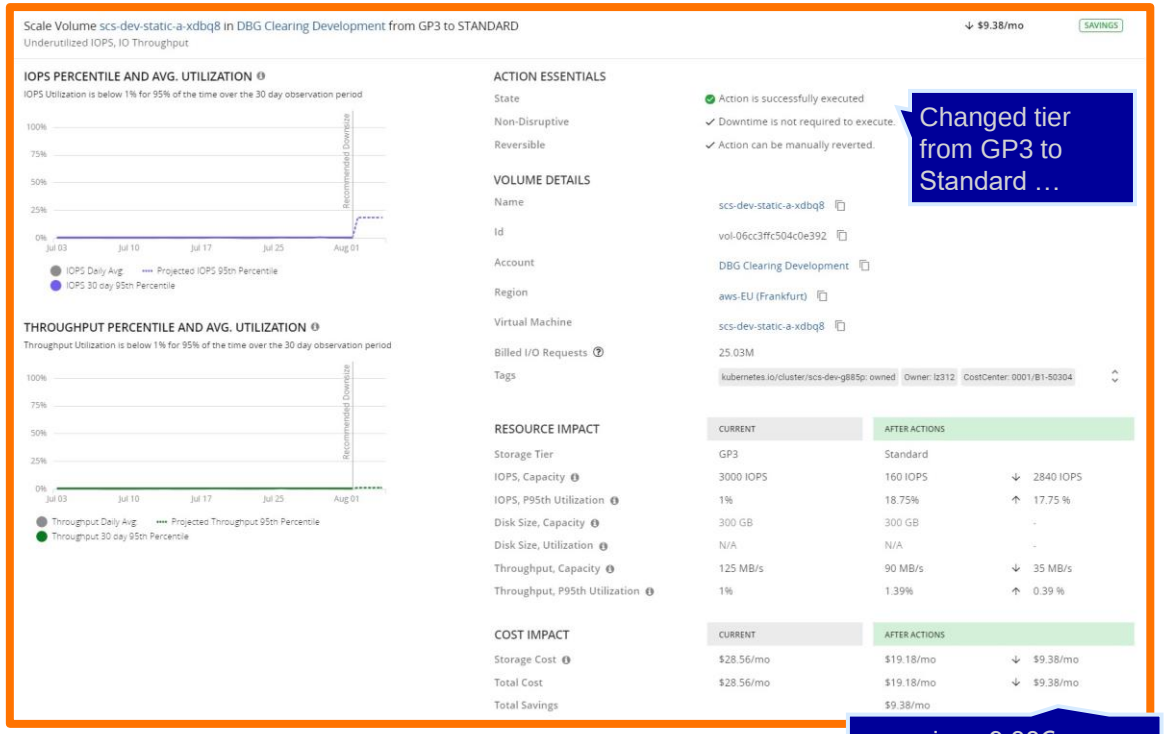
Optimization ↷

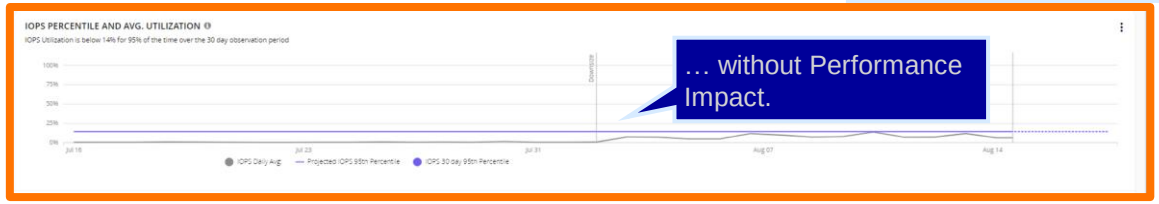
Map applications to infrastructure…



POV Extracts ↷

…to generate trustworthy actions.





Data ↷

API-driven discovery

DEUTSCHE BÖRSE GROUP

Applications

Databases

INSTANA

Kubernetes

RED HAT OPENSHIFT

Virtualizations

vmware

Public Cloud

aws

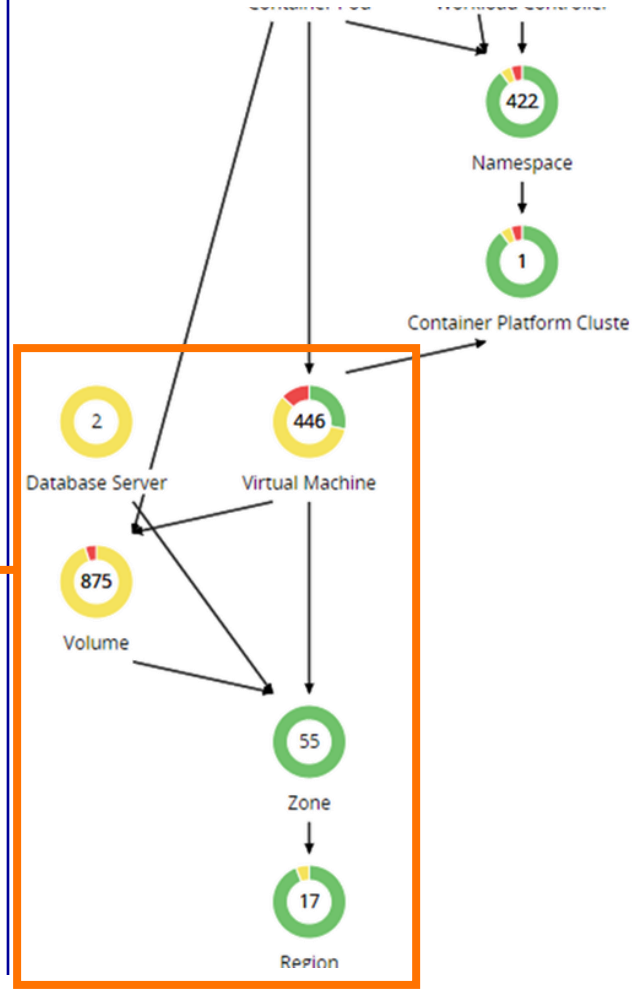
Hyperconverged

Compute

Storage

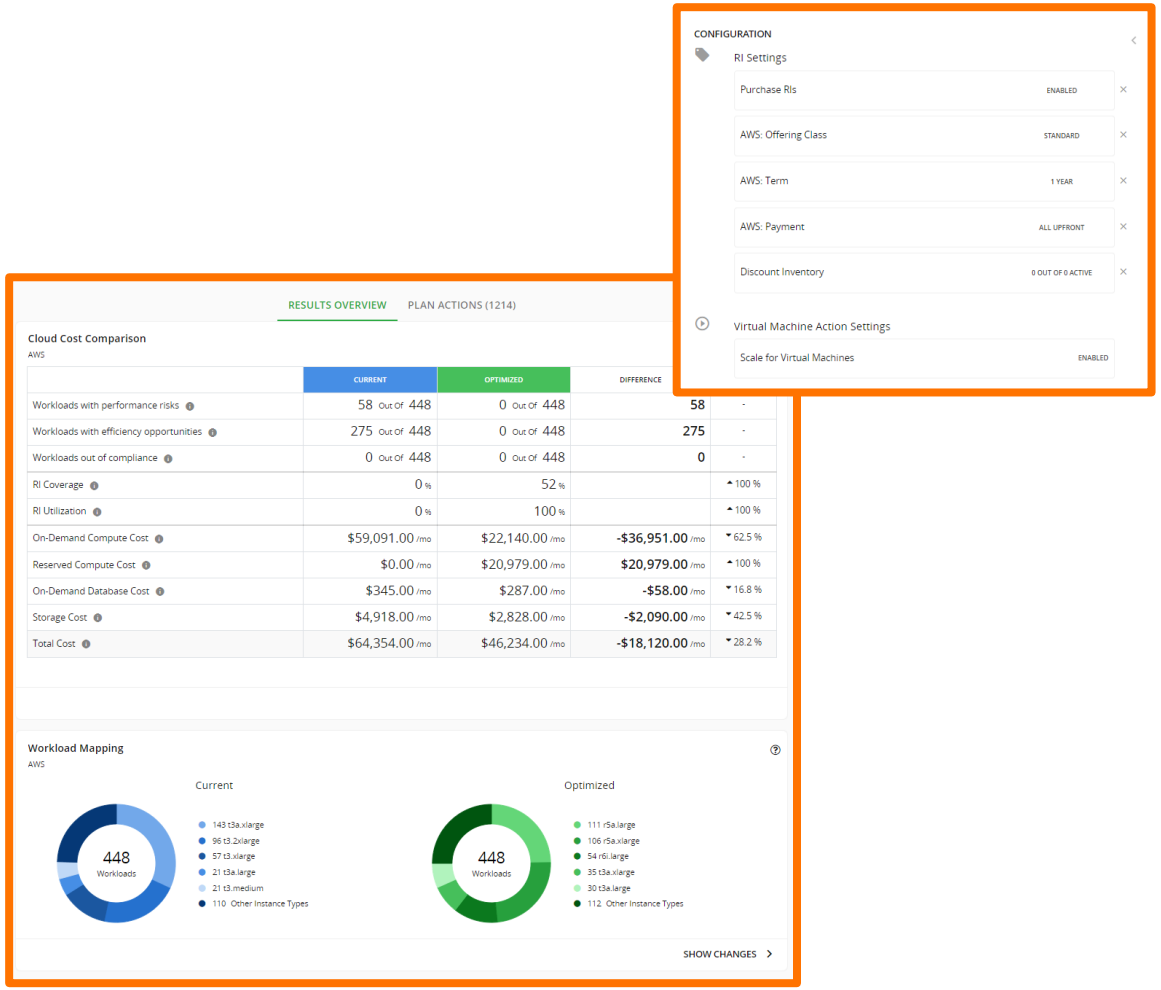
Optimization ↷

Map applications to infrastructure…



POV Extracts ↷

…to PLAN (simulate) Cloud optimization incl. Reservations/Discounts.



Data ↷

API-driven discovery

DEUTSCHE BÖRSE GROUP

Applications

Databases

INSTANA

Kubernetes

RED HAT OPENSHIFT

Virtualizations

vmware

Public Cloud

aws

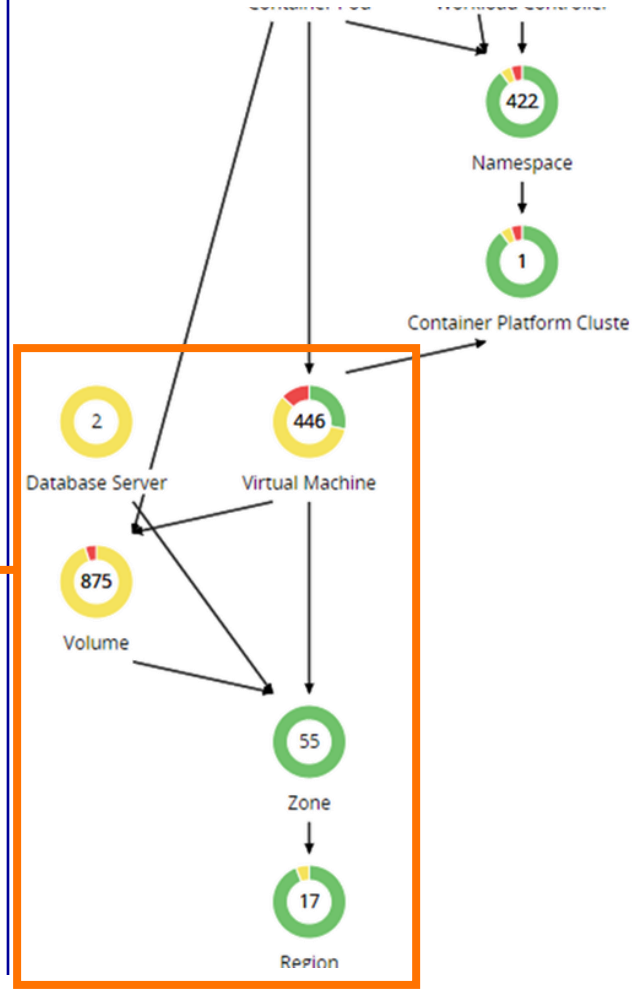
Hyperconverged

Compute

Storage

Optimization ↷

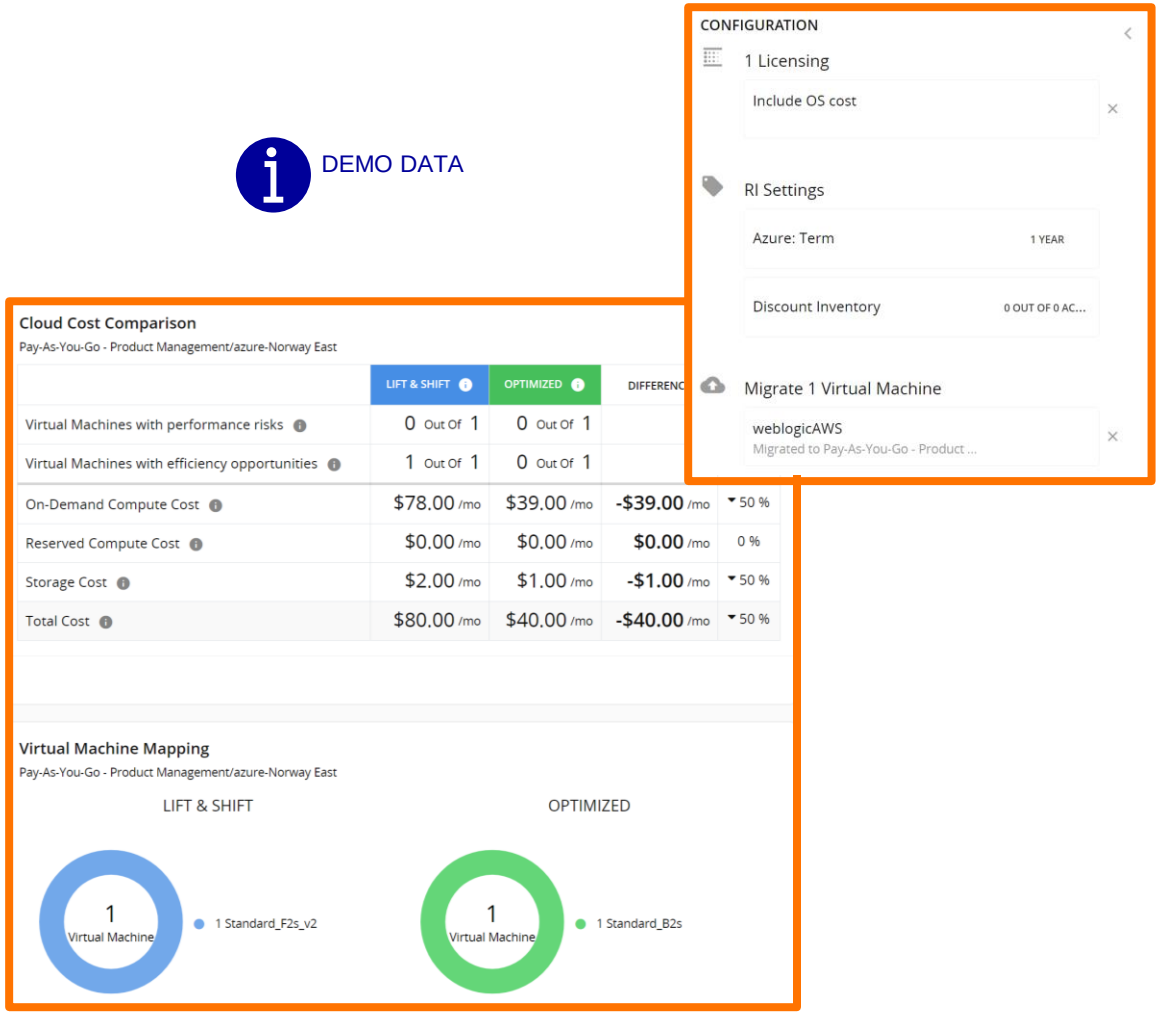
Map applications to infrastructure…



POV Extracts ↷

…to PLAN (simulate) Lift&Shift migrations to or between Cloud Providers.

DEMO DATA



Data ↷

API-driven discovery

DEUTSCHE BÖRSE GROUP

Applications

Databases

INSTANA

Kubernetes

RED HAT OPENSHIFT

Virtualizations

vmware

Public Cloud

aws

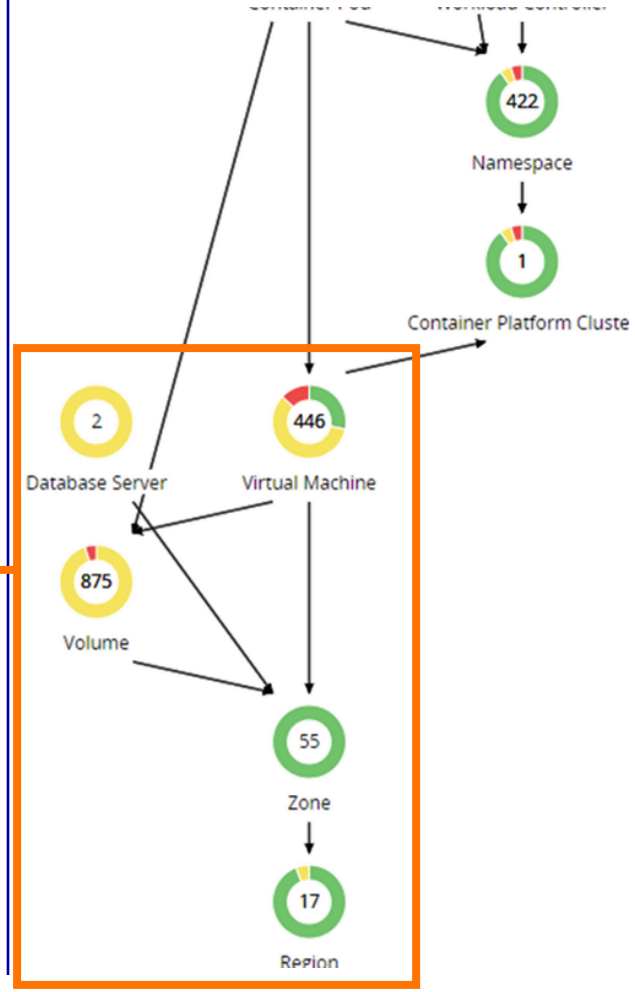
Hyperconverged

Compute

Storage

Optimization ↷

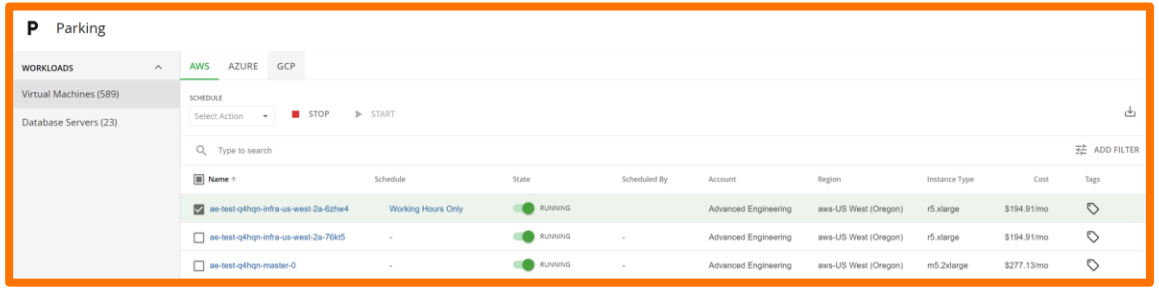
Map applications to infrastructure…

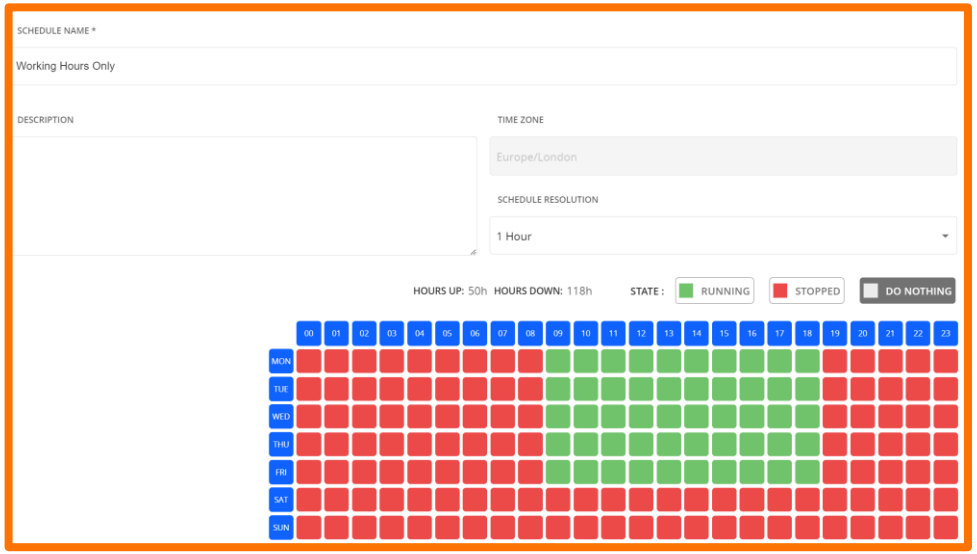


POV Extracts ↷

+ option for scheduled parking of Cloud Resources.

DEMO DATA





Summary

Visibility

Single Platform for Hybrid-/Multi-Cloud and Kubernetes

- API-based
- Application centric
- Contextual views + option for scoped & user defined views

Efficiency

Smart analytics of current and history data to spot trends and trigger alerts

- Rightsizing
- Cost efficency
- Optimized incident handling
- Reduced Time-2-Market

Automation

Actionable: changes can be managed and automated honoring ownership and responsibilities by implementing SOD with roles and permissions.

- Automatable resource decisions
- Dynamization of IT resource usage

Openess

Connects to existing INSTANA and CLOUDABILITY information

- Easy setup
- using industry standards (Webhooks)

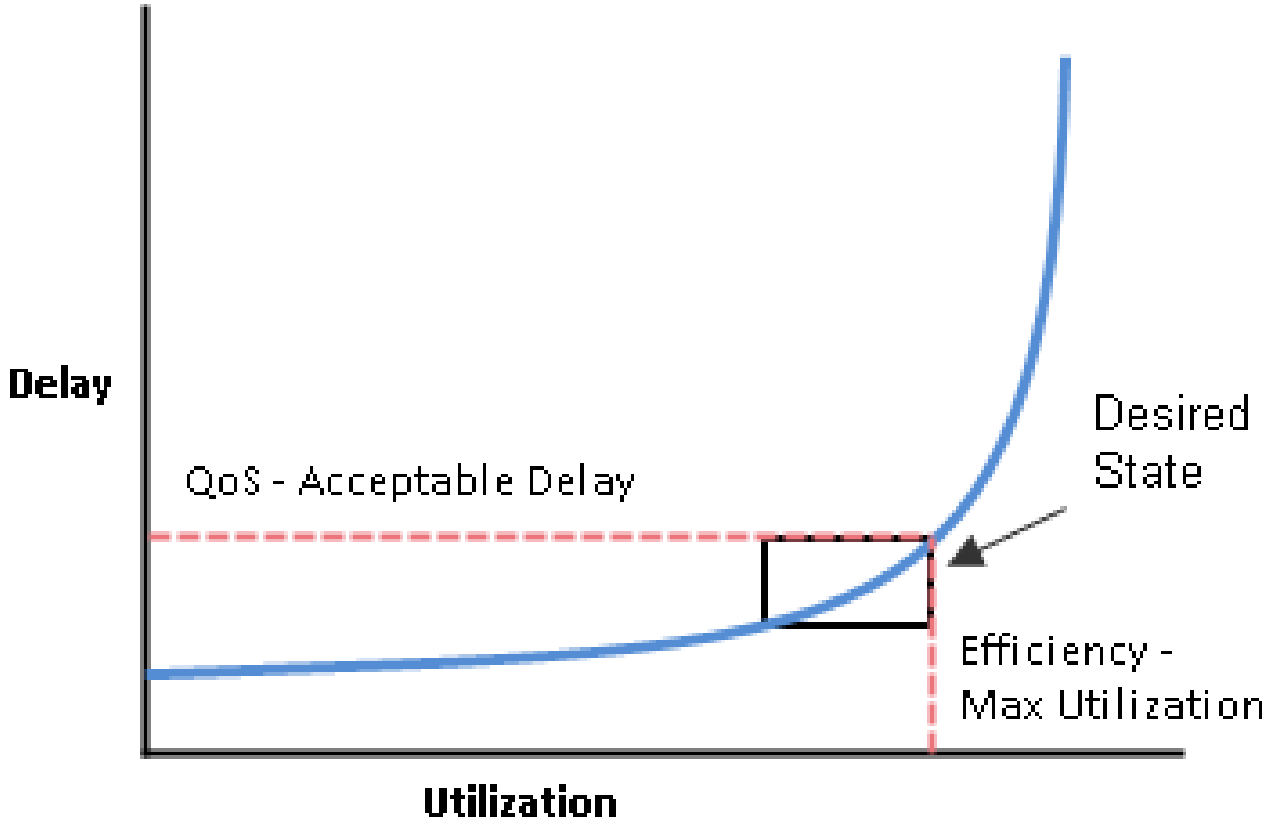
Initiatives & Goals

Addressing Current DBG Initiatives

- Data Center Efficiency *(Performance, Capacity Planning, cross Stack)*
- Application Modernization *(K8, OCP, Cloud Native)*
- Multi-Cloud Transformation *(Dyn. Cap.Mgmt., xTeam Visibility)*
- FinOps *(Cost Mgmt., License Compliance)* — Cloudability, IBM Turbonomic
- Sustainable IT *(e.g. reducing power consumption)*
- Other (tbd)

Improved capacity management and planning for future hardware, software and platform changes, including multi-cloud transformation planning	Improve Performance Engineering, Stability and Resilience with performance assurance capabilities	Scalability through automation (built-in automatable actions with sufficient policy control features for DBG)
↷ Change from static to dynamic capacity management.	↷ Ensure application performance while IT infrastructure is used as efficiently as possible.	↷ Take advantage of the variable cost model of the cloud.
Eliminate complexity and manual processes where possible in the current operational process	Enable and future proof DBG's strategic decisions through a consistent and agnostic management platform (on-prem, public cloud, containerization)	Enable DBG to optimize inefficiency & reclaim over-allocation of resources (e.g., in various staging environments)
↷ Trustworthy resource decision you can automate	↷ Support the full application stack in Hybrid and multi cloud incl. Kubernetes	↷ Usually cut cloud costs by 33% and achieve ROI in first year

The Desired State



Reference: *https://www.ibm.com/docs/en/tarm/8.13.5?topic=works-desired-state*

The Market and Virtual Currency

Turbonomic uses two sets of abstraction to model the environment:

- Modeling the physical and virtual IT stack as a service supply chain

- The supply chain models your environment as a set of managed entities. These include applications, VMs, hosts, storage, containers, availability zones (cloud), and data centers. Every entity is a buyer, a seller, or both. A host machine buys physical space, power, and cooling from a data center. The host sells resources such as CPU cycles and memory to VMs. In turn, VMs buy host services, and then sell their resources (VMem and VCPU) to containers, which then sell resources to applications.

- Using virtual currency to represent delay or QoS degradation, and to manage the supply and demand of services along the modeled supply chain

- The system uses virtual currency to value these buy/sell transactions. Each managed entity has a running budget — the entity adds to its budget by providing resources to consumers, and the entity draws from its budget to pay for the resources it consumes. The price of a resource is driven by its utilization — the more demand for a resource, the higher its price.

Reference: *https://www.ibm.com/docs/en/tarm/8.13.5?topic=works-market-virtual-currency*

Disclaimer

© Deutsche Börse Group 2024

This content is for informational purposes only. None of the information constitutes investment advice or an offer to sell or the solicitation of an offer to buy any contract, share or other financial instrument. This content is only for use as general information and all descriptions, examples and calculations contained are for illustrative purposes only.

Deutsche Börse AG, Frankfurter Wertpapierbörse (FWB®, the Frankfurt Stock Exchange), Eurex Frankfurt AG, Eurex Deutschland and Eurex Clearing AG and Deutsche Börse Digital Exchange do not represent that the information in this publication is comprehensive, complete or accurate and exclude liability for any consequence resulting from acting upon the contents of this or another publication, in so far as no willful violation of obligations took place or, as the case may be, no injury to life, health or body arises or claims resulting from the Product Liability Act are affected.

Securities traded on the Frankfurt Stock Exchange and Eurex derivatives (other than EURO STOXX 50® Index Futures contracts, EURO STOXX® Select Dividend 30 Index Futures contracts, STOXX® Europe 50 Index Futures contracts, STOXX® Europe 600 Index Futures contracts, STOXX® Europe Large/Mid/Small 200 Index Futures contracts, EURO STOXX® Banks Sector Futures contracts, STOXX® Europe 600 Banks/Industrial Goods & Services/Insurance/Media/Personal & Household Goods/Travel & Leisure/Utilities Futures contracts, Dow Jones Global Titans 50 IndexSM Futures contracts, DAX® Futures contracts, MDAX® Futures contracts, TecDAX® Futures contracts, SMIM® Futures contracts, SLI Swiss Leader Index® Futures contracts, RDXxt® USD - RDX Extended Index Futures contracts, Eurex inflation/commodity/weather/property and interest rate derivatives) are currently not available for offer, sale or trading in the United States nor may they be offered, sold or traded by persons to whom US tax laws apply.

The fund shares listed in XTF Exchange Traded Funds® are admitted for trading on the Frankfurt Stock Exchange. Users of this information service who legally reside outside Germany are herewith advised that sale of the fund shares listed in XTF Exchange Traded Funds may not be permitted in their country of residence. The user makes use of the information at their own risk.

Legal validity of this disclaimer

In the event that individual parts of or formulations contained in this text are not, or are no longer, legally valid (either in whole or in part), the content and validity of the remaining parts of the document are not affected.

Trademarks

The following names and designations are registered trademarks of Deutsche Börse AG or an affiliate of Deutsche Börse Group:

1585®; A7®; Buxl®; C7®; CDAX®; CEF®; CEF alpha®; CEF ultra®; CFF®; Classic All Share®; Clearstream®; CX®; D7®; DAX®; DAXglobal®; DAXplus®; DB1 Ventures®; DBIX Deutsche Börse India Index®, Deutsche Börse®; Deutsche Börse Capital Markets Partner®; Deutsche Börse Commodities®; Deutsche Börse Venture Network®; Deutsches Eigenkapitalforum®; DivDAX®; eb.rexx®; eb.rexX Jumbo Pfandbriefe®; ERS®; eTriParty®; Eurex®; Eurex Bonds®; Eurex Clearing Prisma®; Eurex Improve®; Eurex Repo®; Euro GC®; ExServes®; EXTF®; F7®; FDAX®; FWB®; GC Pooling®; GCPI®; GEX®; Global Emission Markets Access – GEMA®; HDAX®; iNAV®; L-DAX®; L-MDAX®; L-SDAX®; L-TecDAX®; M7®; MDAX®; N7®; ODAX®; ÖkoDAX®;PROPRIS®; REX®; RX REIT Index®; SCHATZ-FUTURE®; SDAX®; ShortDAX®; StatistiX®; T7®; TecDAX®; Technology All Share®; TRICE®; USD GC Pooling®; VDAX®; VDAX-NEW®; Vestima®; Xcreen®, Xemac®; Xentric®, Xetra®; Xetra-Gold®; Xpect®; Xpider®; XTF®; XTF Exchange Traded Funds®; We make markets work®

The names and trademarks listed above do not represent a complete list and, as well as all other trademarks and protected rights mentioned in this publication, are subject unreservedly to the applicable trademark law in each case and are not permitted to be used without the express permission of the registered owner. The simple fact that this publication mentions them does not imply that trademarks are not protected by the rights of third parties.

The STOXX® indices, the data included therein and the trademarks used in the index names are the intellectual property of STOXX Ltd., Zug, Switzerland and/or its licensors. Eurex' derivatives based on the STOXX indices are in no way sponsored, endorsed, sold or promoted by STOXX and its licensors and neither STOXX nor its licensors shall have any liability with respect thereto.

STOXX iSTUDIO® is a registered trademark of STOXX Ltd., Zug, Switzerland.

EEX® is a registered trademark of European Energy Exchange AG.

TRADEGATE® is a registered trademark of Tradegate AG Wertpapierhandelsbank.

Flexible is better.® is a registered trademark of Axioma, Inc.

Thank you!

Alexander Buschmann
VP Clearing & Risk IT Architects Ofiice

Deutsche Börse AG
Mergenthalerallee 61
65760 Eschborn

Phone +49 69 21118984
Fax +49 69 21118984
E-mail alexander.buschmann@deutsche-boerse.com

OpenDay 2024

DEUTSCHE BÖRSE GROUP