

AIOps: AI for IT Operators

Jian Zhang

Microsoft Azure
May 19th, 2021

Stanford University - EE392B





Agenda

- What's AIOps
- AIOps at Azure
- Annoucement



What's AIOps



AIOps definition (Gartner)

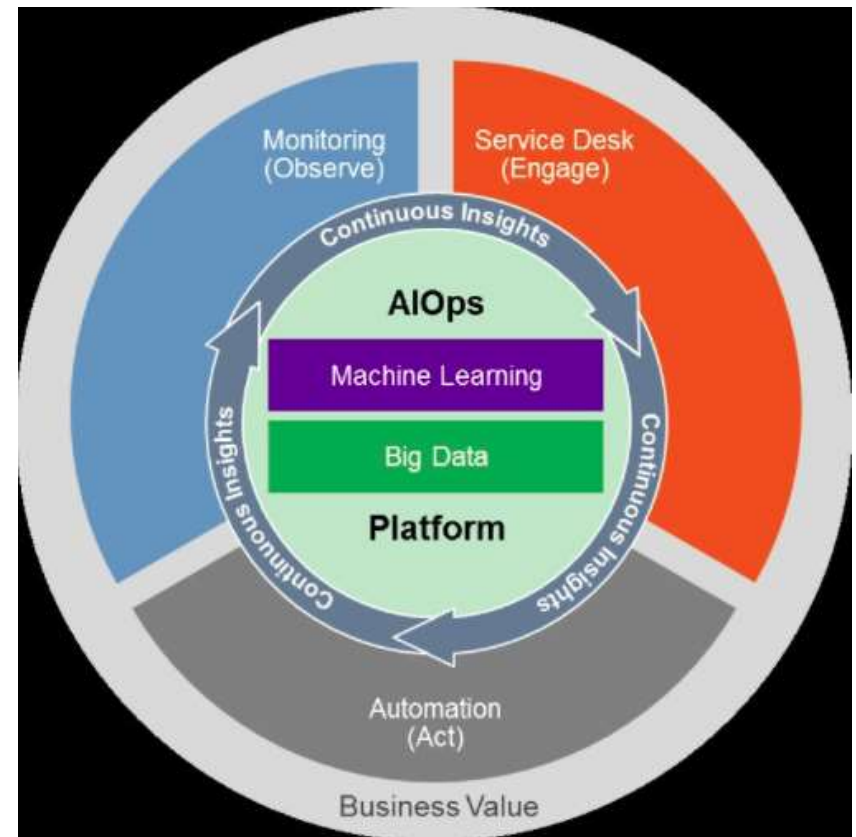


Adoption has increased with the uptick of digital transformation



Business value

Improve engineering efficiency and service quality; Reduce COGS



Source: Gartner



AI Ops Problem Space



Detection



Diagnosis



Prediction



Optimization



AIOps Products

Embedded AIOps features from cloud service providers

- Amazon
- Google
- Microsoft

AIOps product offering from more IT service providers

- IBM
- Splunk
- Cisco (AppDynamics)
- New Relic
- DataDog

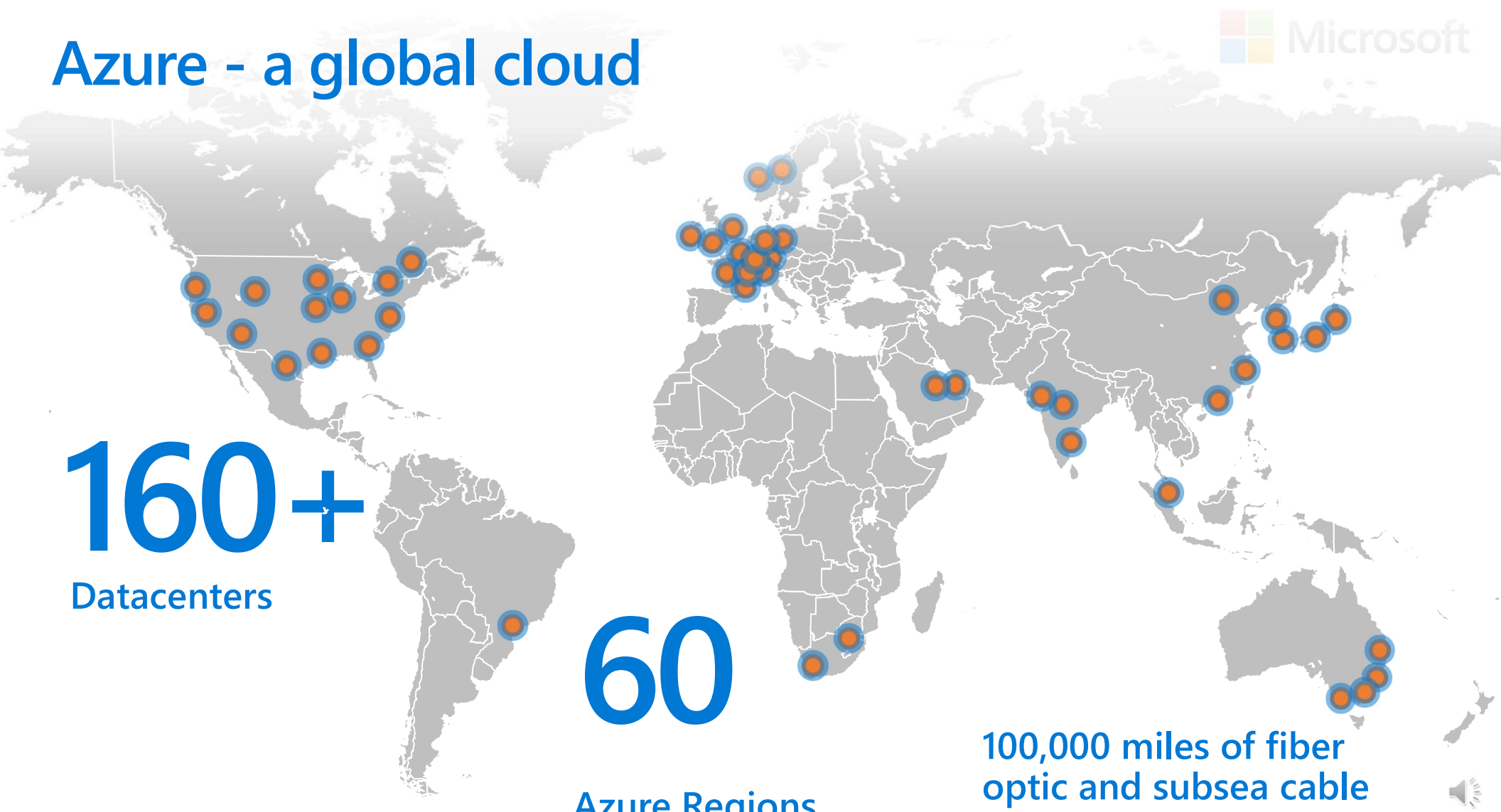


Azure - a global cloud

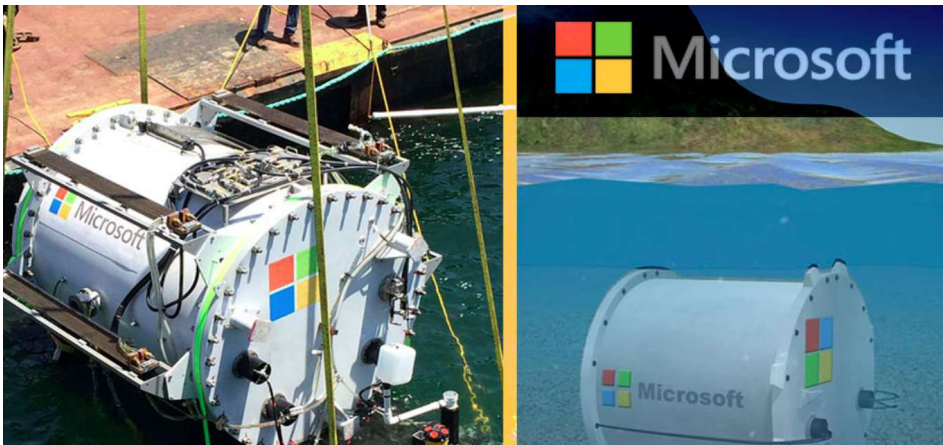
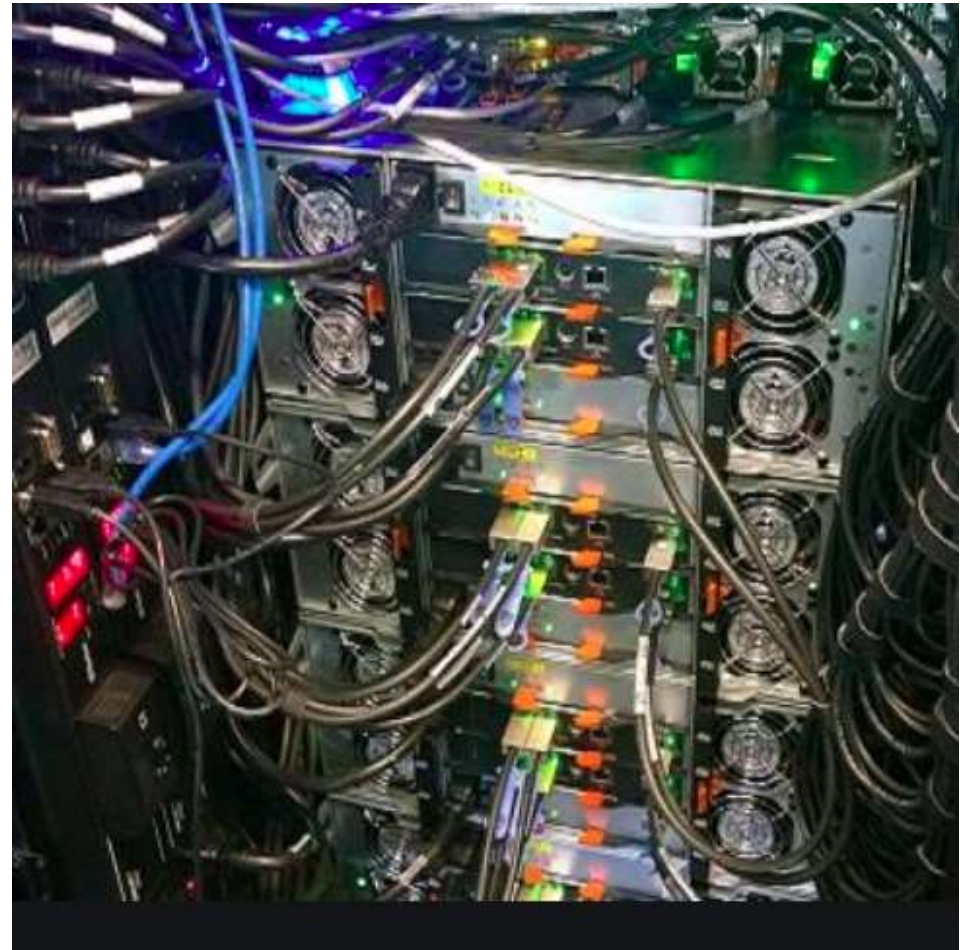
160+
Datacenters

60
Azure Regions

100,000 miles of fiber
optic and subsea cable



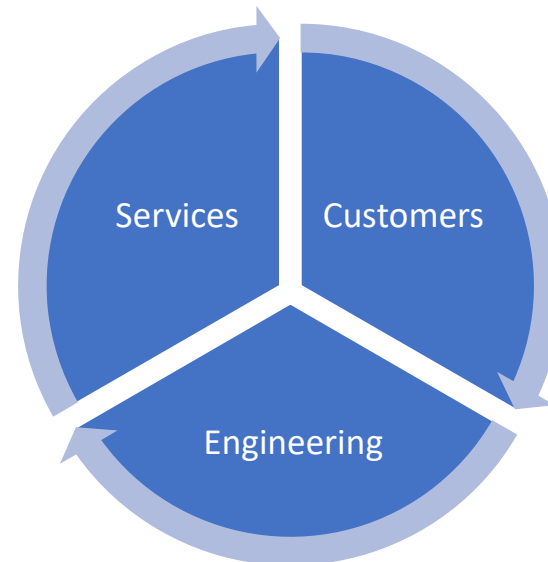
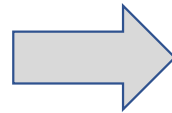




AI Ops at Azure

Motivation

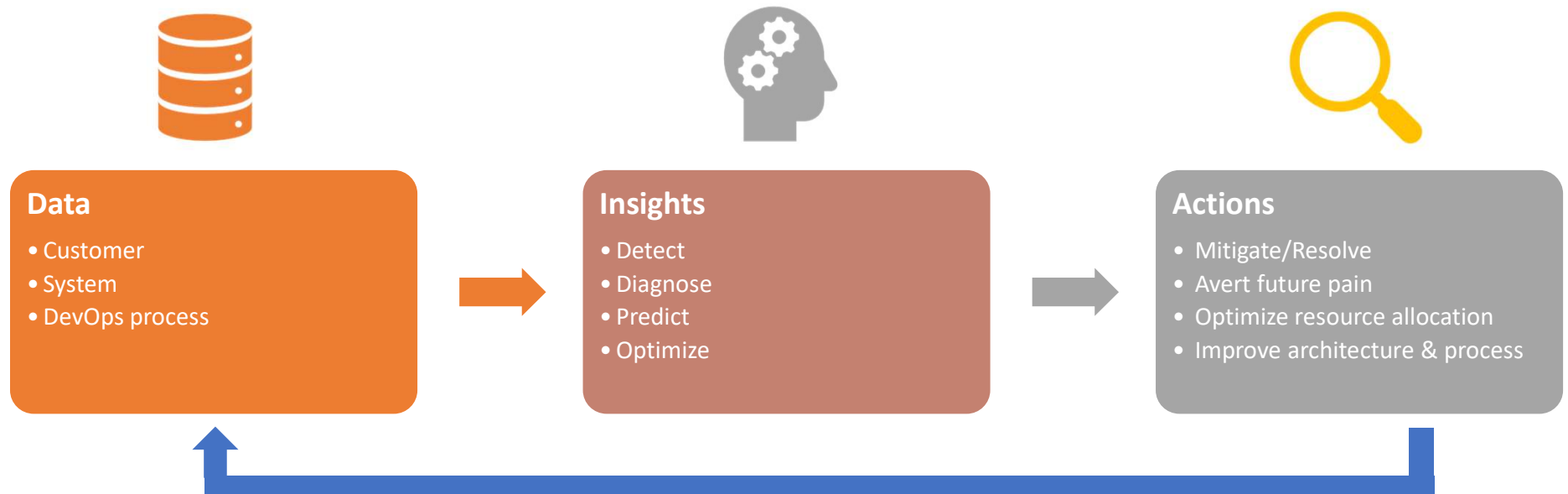
- Scale
- Complexity
- Quality



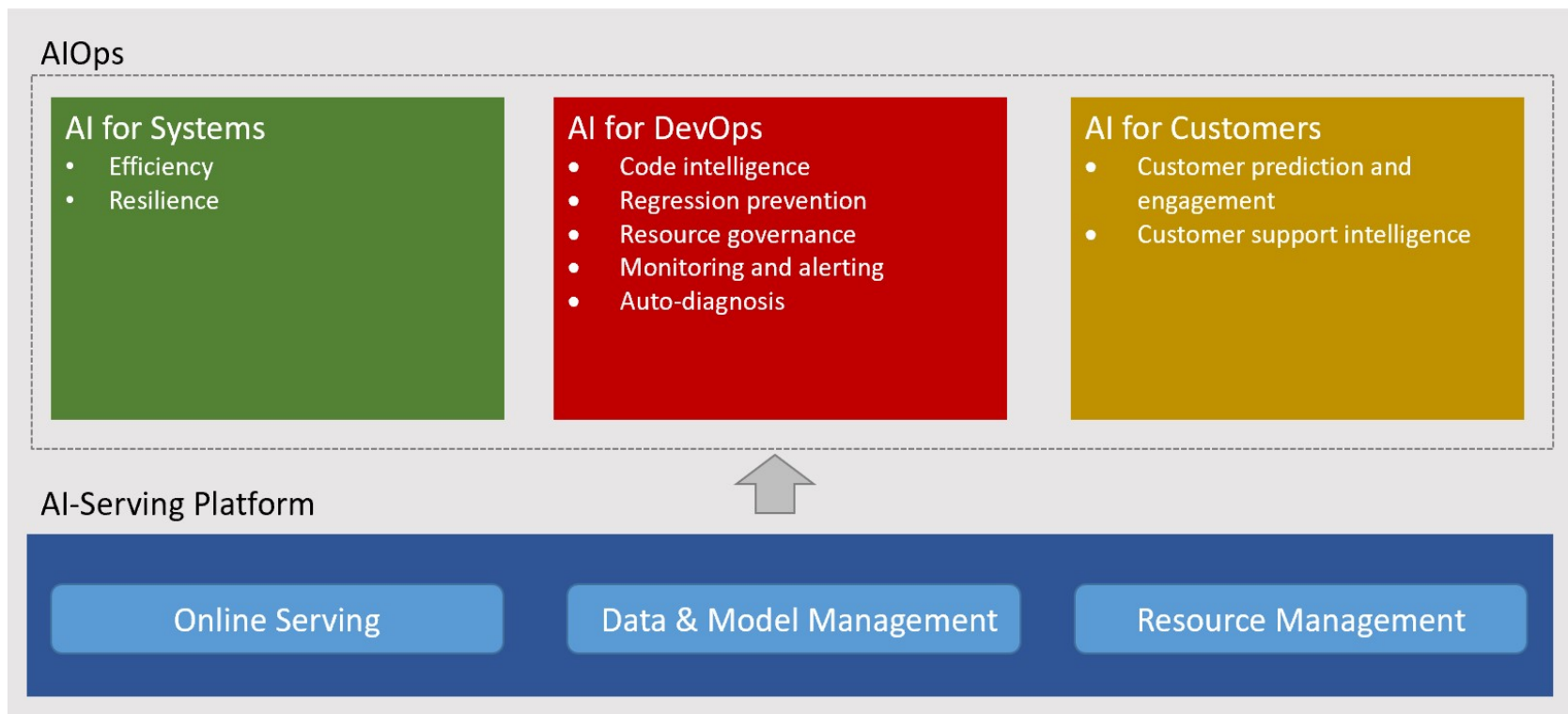
Empower Building & Operating
Services at Scale



AI Ops Methodologies: From Data to Actions



Infusing AI into Systems & Operations



AI for System Resilience



Project Tardigrade

➤ Scenario – Platform resilience

- Reduce VM service interruption caused by physical host reboot

➤ Approach

- Hardware failure prediction
- Memory leak detection
- Physical host recovery only pauses VM execution
- VM live migration

➤ Result & Learnings

- VMs: can survive host reboot and continue running
- Improved VM service availability



AI for DevOps (Sample Solutions)

Intelligent Development



Sankie

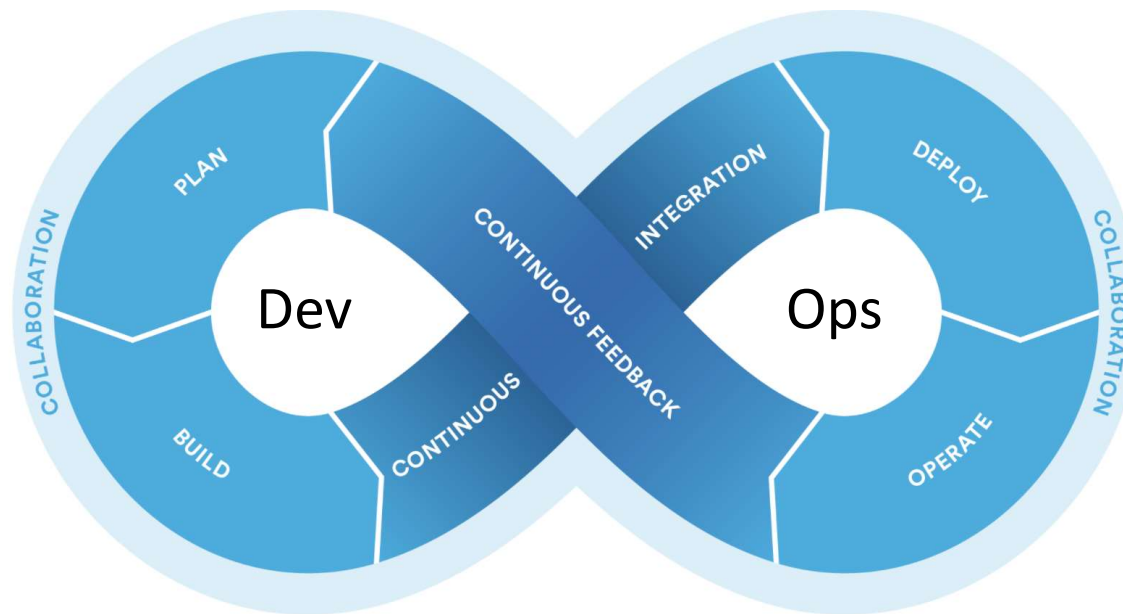
Effective localization

Safe Deployment




Gandalf

Automatic stop



Picture courtesy of addteq.com

Intelligent Anomaly Detection



AiDice

Monitoring & Diagnosis

Intelligent Live Site Management

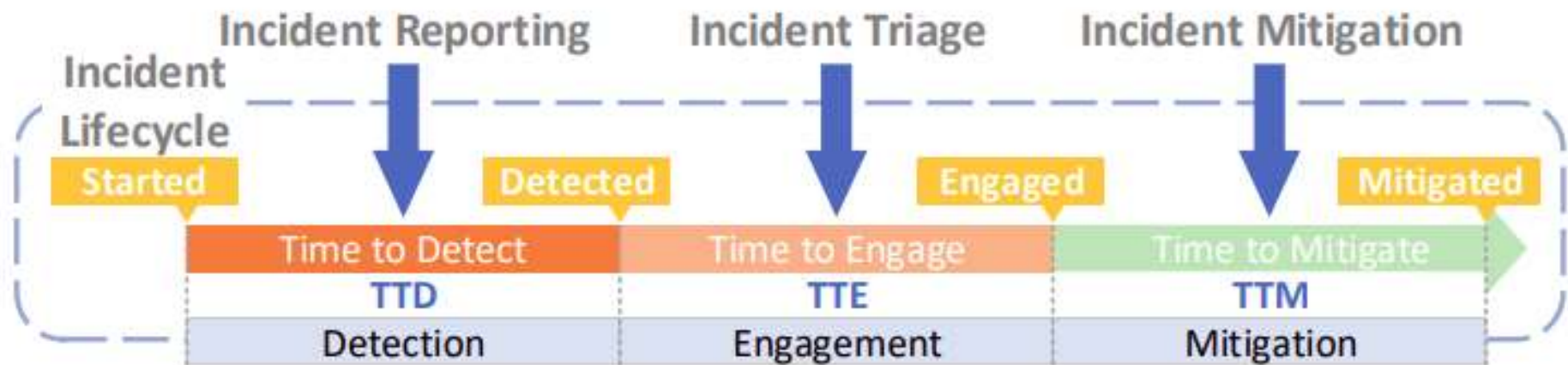


BRAIN

Fast mitigation/fix



Incident Management Procedure



Z. Chen et al, "AIOps Innovations of Incident Management for Cloud Services", AAAI-20

Gandalf Safe Deployment

Problem

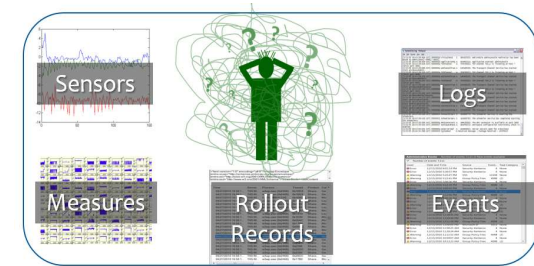
Manual assessment: *labor intensive, time consuming, error prone*

Goal

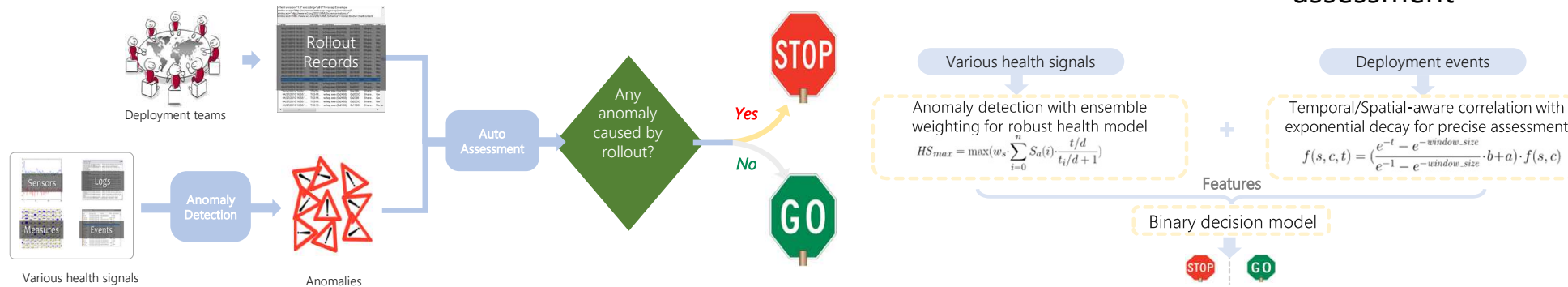
Systematic, data-driven solution for deployment management

KEY TECH

- Robust anomaly detection with ensemble weighting
- Temporal/spatial-aware correlation with exponential decay



Manual deployment assessment



Möbius: Online Anomaly Detection

Problem

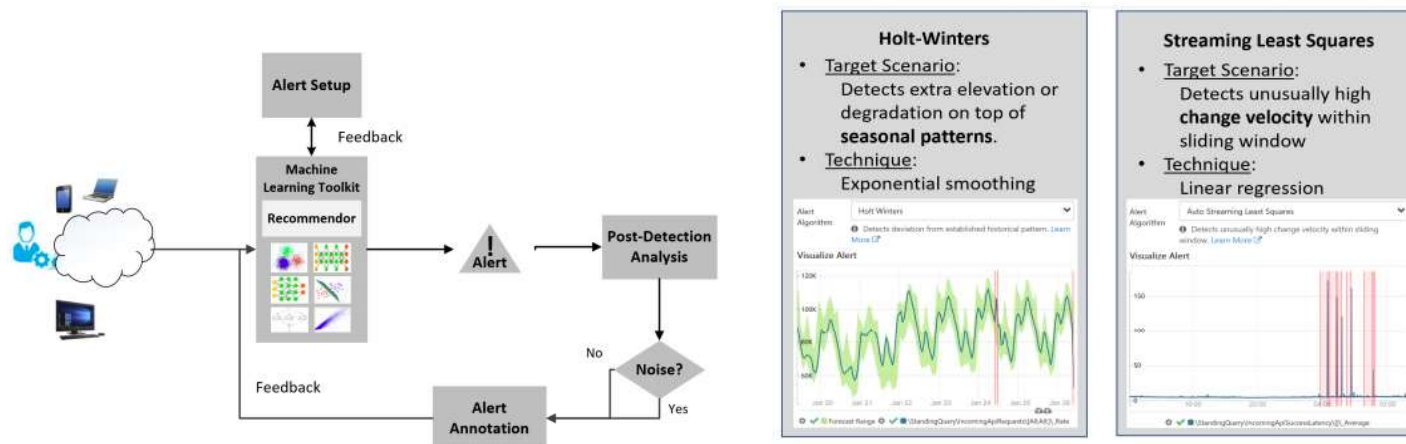
Traditional thresholder alerts generate noise in detection

Goal

Improve signal to noise ratio

KEY TECH

- Model recommendation from a model pool based on workload characteristics
- Feedback loop for continuous improvement and adaptation to dynamic workload



AiDice

– Efficient Incident Identification

Problem

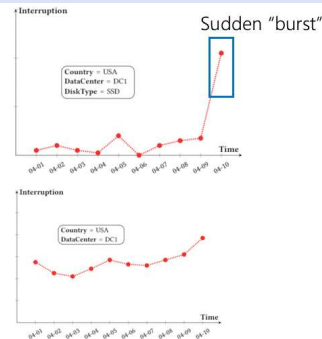
Needle in a haystack: Hard to detect & locate emerging issues from an enormous attribute combinations

Goal

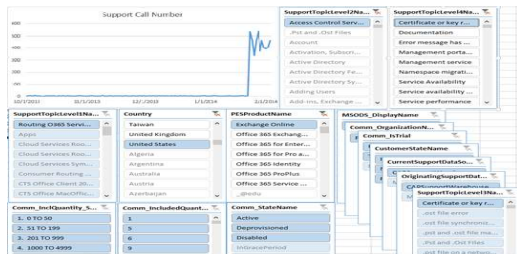
Actionable alert with precise Issue location

Daily aggregation of Service Interruptions

Time	Country	Datacenter	Disk Type	Interruption
xxxx-xx-xx	USA	DM1	SSD	1
xxxx-xx-xx	Australia	MEL21	SSD	1
xxxx-xx-xx	USA	DC1	HDD	4
xxxx-xx-xx	India	BL1	SSD	10
xxxx-xx-xx	UK	SN6	Hybrid	3
xxxx-xx-xx	USA	DM1	HDD	0

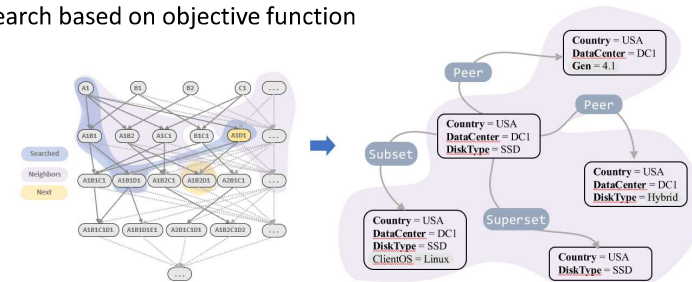


iDice: Problem Identification for Emerging Issues, ICSE 2016



Problem Formulation

Select attributes combinations from neighbors with heuristics search based on objective function



Object Function

$$R = p_{B+A} \ln \frac{p_A}{p_B}$$

where

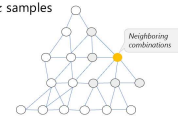
$$p_x = \frac{V_{Xt}}{V_t}$$

V_{Xt} : # of interruptions for current combination during a time period t
 V_t : # of total interruptions for during a time period t

Best from Multiple Selection

- 1) Randomly sample k combinations from neighbors
- 2) Select the one with largest obj among k samples

Theoretical Guidance for BMS:
 $k \geq \frac{\ln[1 - Pr_c]}{\ln p_c}$
 where p_c is the percentile among all results and Pr_c is the corresponding confidence.



Efficiency with informative guide

- 3) Maintain a forbidden list consisting of recently visited solutions



AI/ML Platform - Resource Central

ML and prediction-serving system for improving resource management



RC clients: Platform resource managers

VM scheduling

Cluster selection

Power
oversubscription

Server
maintenance

VM rightsizing
recommendation

R. Bianchini, M. Fontoura et al. "Toward ML-Centric Cloud Platforms", Communications ACM, Feb 2020



CONTRIBUTED ARTICLES

Toward ML-Centric Cloud Platforms

By Ricardo Bianchini, Marcus Fontoura, Eli Cortez, Anand Bonde, Alexandre Muzio, Ana-Maria Constantin, Thomas Moscibroda, Gabriel Magalhaes, Girish Bablani, Mark Russinovich
Communications of the ACM, February 2020, Vol. 63 No. 2, Pages 50-59
10.1145/3364684
[Comments](#)

VIEW AS: SHARE:



Credit: Marcel Clemens

Cloud platforms, such as Microsoft Azure, Amazon Web Services (AWS), and Google Cloud Platform, are tremendously complex. For example, the Azure Compute fabric governs all the physical and virtualized resources running in Microsoft's datacenters. Its main resource management systems include virtual machine (VM) and container (hereafter we refer to VMs and containers simply as "containers") scheduling, server and container health monitoring and repairs, power and energy management, and other management functions.

[Back to Top](#)

Key Insights

- There are many potential uses of ML in cloud computing platforms. The challenge is in defining exactly how and where ML should be infused in these platforms.
- Leveraging ML-derived predictions has shown promise for many resource managers in Azure Compute. Having a general and independent ML framework/system has been key to increasing adoption quickly.
- Many research challenges remain open, including how to make action-prescribing ML general enough for wide applicability in cloud platforms, how to manage (potentially partial) feedback at scale, and how to debug misbehaviors (especially when the ML is tightly integrated with resource managers).



Cloud platforms are also extremely expensive to



AI Ops in Azure: Summary

- AI Ops is critical for **digital transformation** and an emerging innovation area
- AI Ops is a **multi-discipline** research area involving software engineering, systems, machine learning, big data and visualization
- AI Ops is **comprehensive**: from making the system smart and resilient to enhancing developer efficiency and improving customer experience
- AI Ops is what makes modern clouds **scale** and **efficient** to support the next generation of Computing
- AI Ops calls for close **collaboration** between the industry and academia



Announcement





ICSE21 Workshop on Cloud Intelligence

In conjunction with the 43rd International Conference on Software Engineering

May 29th, 2021 Virtual (Originally Madrid, Spain)

[Home](#)

[Call For Papers](#)

[Accepted Papers](#)

[Organizers](#)

[Program](#)

2021 ▾

Cloud Intelligence

AI/ML for Efficient and Manageable Cloud Service

May 29th, 2021 Virtual (Originally Madrid, Spain)

<https://cloudintelligenceworkshop.org>

Registration deadline: May 23rd, 2021





The 4th International Workshop on Artificial Intelligence of Things

In conjunction with the 27th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2021)

Virtual conference, Aug. 15-16th, 2021

[Home](#)

[Call For Papers](#)

[Organizers](#)

[KDD 2021](#) ▾

[AAAI 2020](#)

[Submit](#)

Artificial Intelligence of Things

Learn and share enabling technologies of AI powered IoT.

Connect with experts in the field.

Virtual conference, Aug.14-18th, 2021

[Submit](#)

<https://aiotworkshop.github.io>

Submission deadline: May 24th, 2021



Reference

- Rex: Preventing Bugs and Misconfiguration in Large Services Using Correlated Change Analysis, Sonu Mehta, Ranjita Bhagwan et al., NSDI'20
- “An Intelligent, End-To-End Analytics Service for Safe Deployment in Large-Scale Cloud Infrastructure”, in NSDI'20
- Tardigrade: Leveraging Lightweight Virtual Machines to Easily and Efficiently Construct Fault-Tolerant Services, Jacob R. Lorch et al., NSDI'15
- Efficient Incident Identification from Multi-dimensional Issue Reports via Meta-heuristic Search, Jiazhen Gu, Chuan Luo et al., ESEC/FSE'20
- Toward ML-Centric Cloud Platforms, Ricardo Bianchini, Marcus Fontoura et al., Communications ACM, 2020