Thirty-Fourth AAAI Conference on Artificial Intelligence

Microsoft

Workshop

Cloud Intelligence: AI/ML for Efficient and Manageable Cloud Services

February 7th, 2020, New York, New York - USA

AI For Cloud - Toward Intelligent Cloud Platforms and AIOps

Microsoft Azure

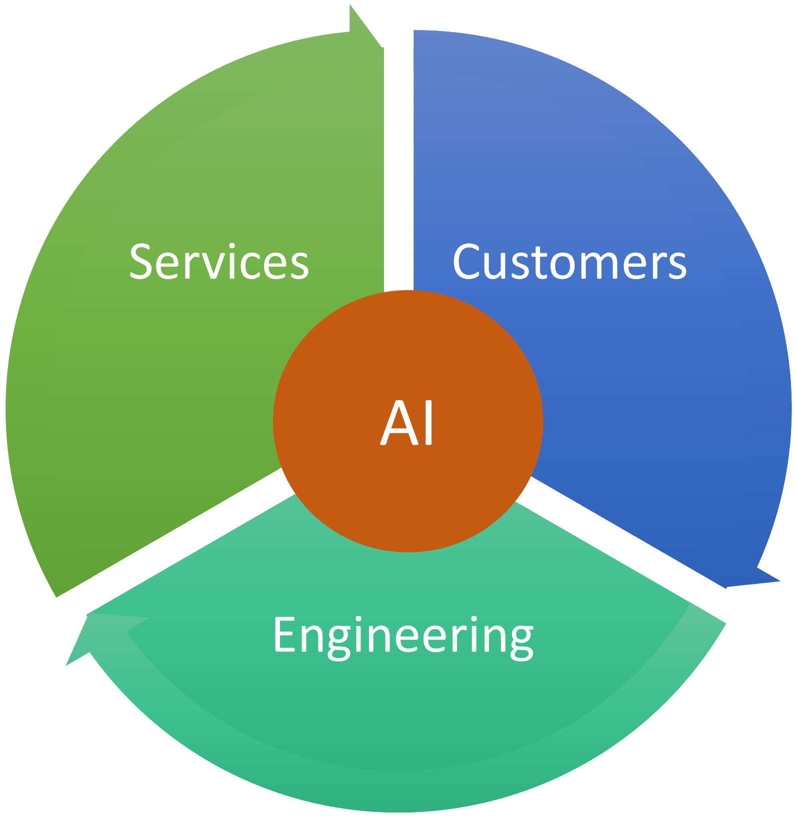
Marcus Fontoura, Technical Fellow

Murali Chintalapati, Partner SWE Manager

Yingnong Dang, Principal Data Scientist Manager

AI for Cloud

- Ongoing digital transformation across all industries
- Scale and complexity as the biggest challenge
- AI/ML is a key technology in addressing this challenge



Infusing AI into Systems & Operations

Infusing AI into Systems & Operations: What Do We Need?

Microsoft

AI For Cloud

AIOps

AI for Platform
- Self-awareness: prevent regressions
- Self-healing: predict failures & protect customers' workloads
- Self-adaptation: predict workload behavior & adjust for efficiency
- …

AI for DevOps
- Programming intelligence
- Regression prevention
- Triage and diagnosis support
- Mitigation intelligence
- …

AI for Customers
- Customer behavior intelligence
- Intelligent customer engagement
- Customer support intelligence

AI-Serving Platform

Online Serving | Data & Model Management | Resource Management

AI For Cloud: AI-Serving Platform for Azure

AI-Serving Platform for Azure

Microsoft

Online Serving
- Resource Central (foundational)
- Azure ML (higher levels)

Data & Model Management
- Azure Data Explorer
- Azure Data Lake
- Resource Central
- Azure ML

Resource Management:
- Impact-free server defragmentation
- Safe core oversubscription
- Etc.

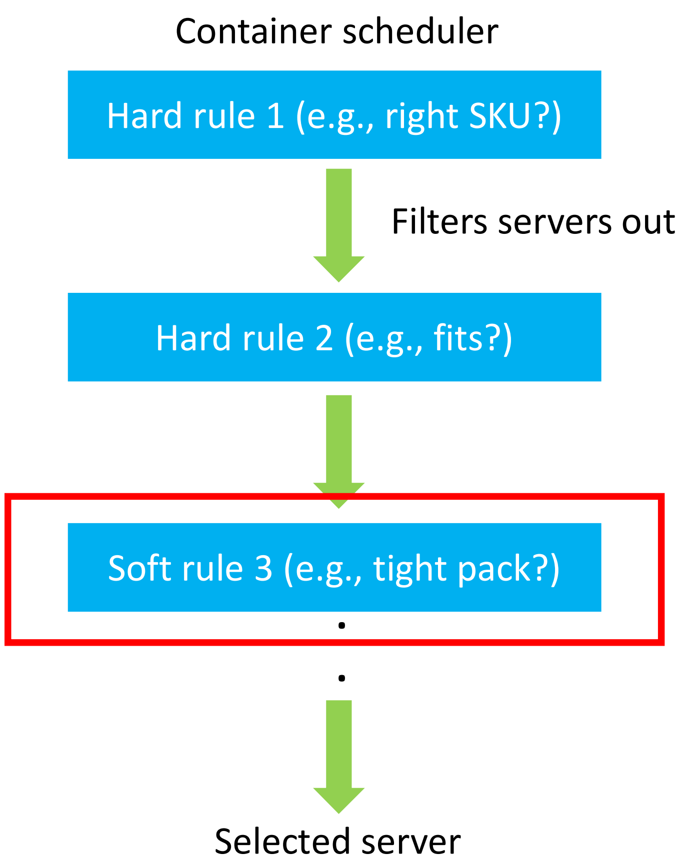
Resource Central

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

Case study: Smart CPU oversubscription



Goals:

- Be conservative! Stick with P95, 1st-party loads
- Don't oversubscribe servers running prod VMs
- Oversubscribe other servers up to a percentage over capacity and a max predicted (P95) utilization

New rule checking the sum of the P95 utilizations

Mispredictions: only issue is consistent under-prediction

RC-informed CPU oversubscription

Simulation results

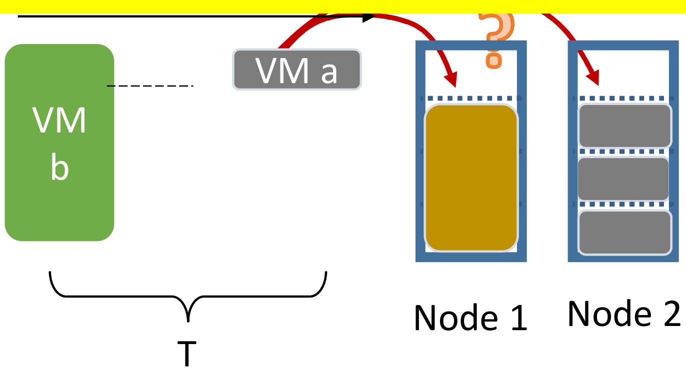
Version	Description	Behavior
Baseline	No oversubscription	Low capacity; many VM allocation failures
Naive	25% oversub without predictions	No failures; 6x resource exhaustion
RC-informed	25% oversub with RC predictions	No failures; rare exhaustion
RC-right	25% oversub with oracle predictions	No failures; same exhaustion

Multi-Dimension Optimization

- Container scheduling should achieve high utilization across all resource dimensions
  1. Multi-dimensional resource packing
  2. Take into account online nature of service allocation

- Simple example: **Assume every VM has**

Lifetime prediction is important for container scheduling



$$\left(\frac{1}{2}\right)^{2} + \left(\frac{1}{2}\right)^{3} = \frac{6}{16}$$

- If new VM is placed on Node 2:

$$\left(\frac{1}{2}\right) + \left(\frac{1}{2}\right)^{4} = \frac{9}{16}$$

→ Placing new VM on Node 2 is better!

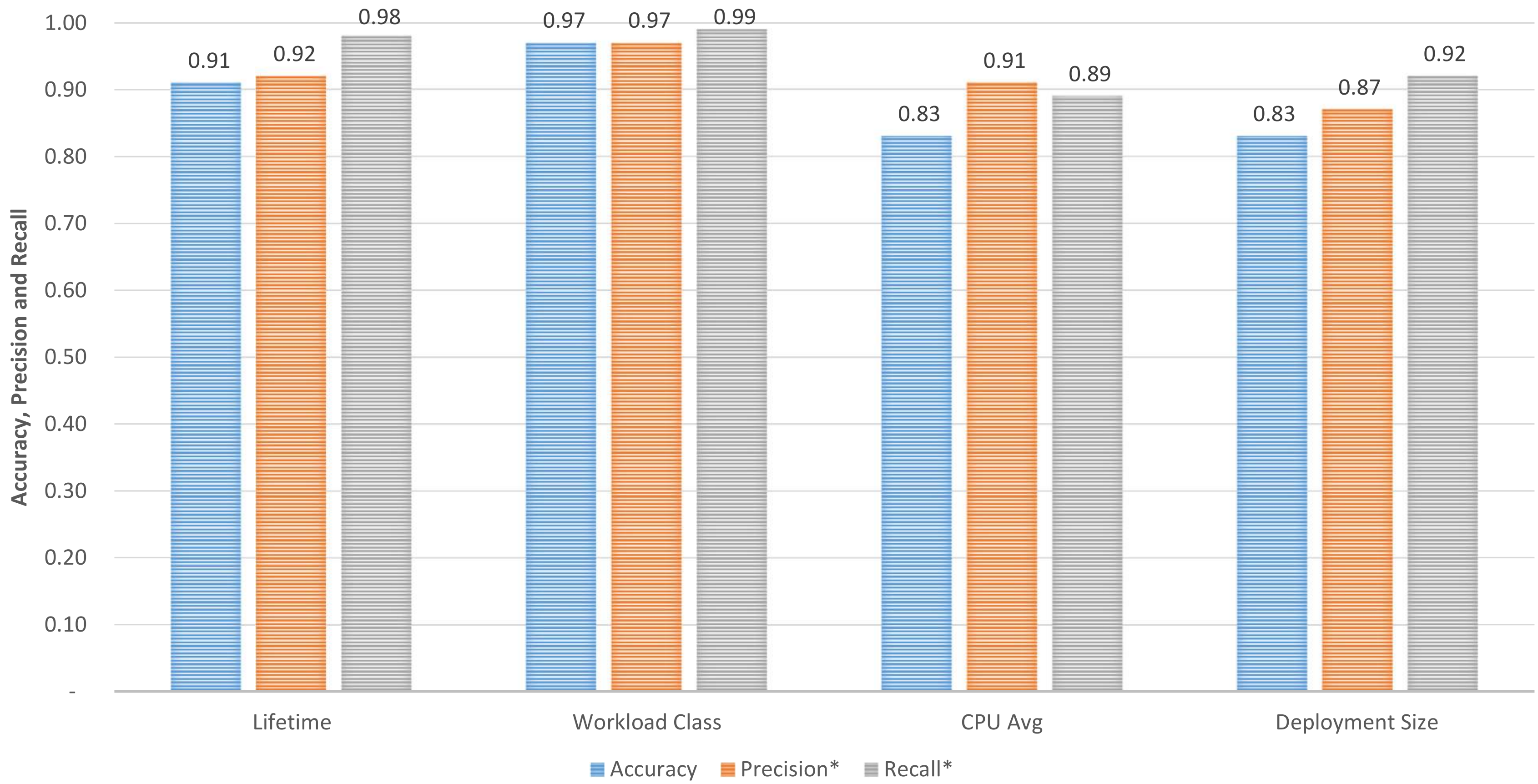
Resource utilization in Azure

- Each 1% of utilization gain results in huge savings

Container scheduling algorithms are crucial for operating the cloud effectively!

Prediction Quality

Accuracy ≥ 83%
$Precision^{\theta}$ ≥ 87%
$Recall^{\theta}$ ≥ 89%



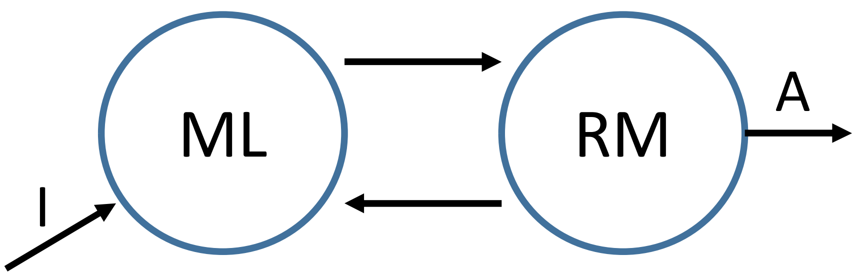
Approaches to adding ML

Passive, external to managers:

Predict load intensity, utilization

Cluster workloads, resources

ML as an insight provider



I = Inputs; A = Actions
RM = Resource Manager

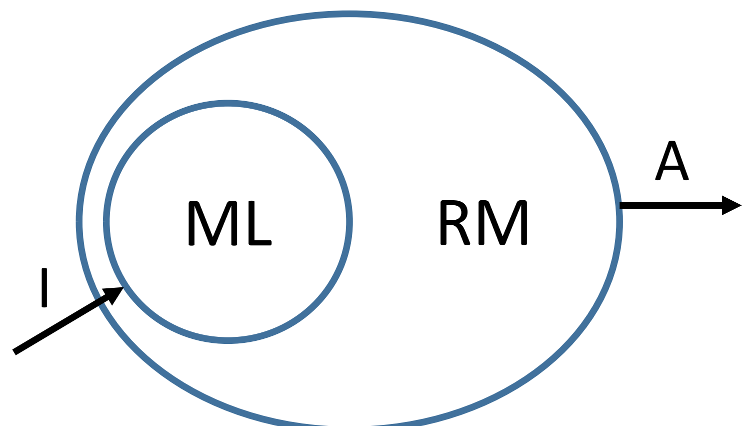
Debuggable; simpler RMs

Active, built into managers:

Adjust parameters of policies

Select actions to be performed

ML has deep knowledge of policies

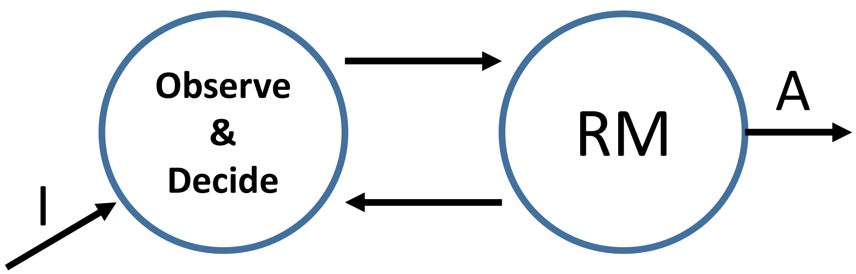


Along a different dimension

Iterative observe and decide:

After each action, observe & decide

Management as a control problem

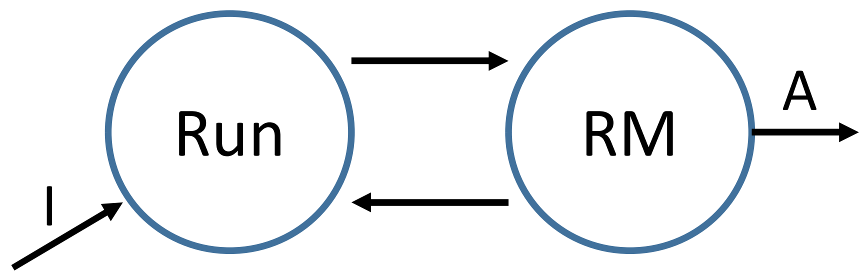


I = Inputs; A = Actions
RM = Resource Manager

Delayed observation:

Generate model offline, run it online

Re-generate model periodically



Simpler

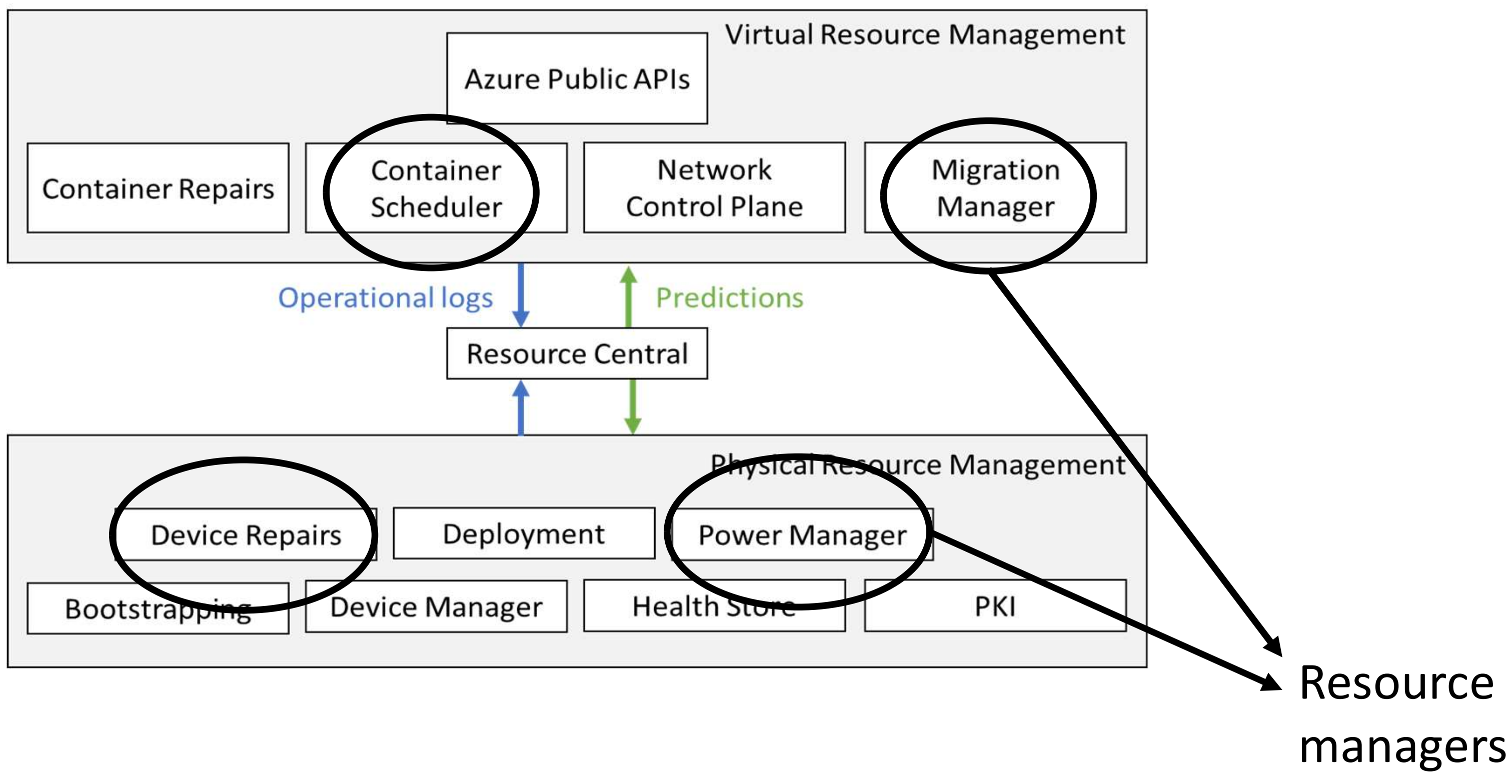
Summary of our approach

A general, passive and delayed-observation framework for all ML tasks

Management	Useful predictions
Container scheduling	VM resource utilization
Scavenging	VM workload class

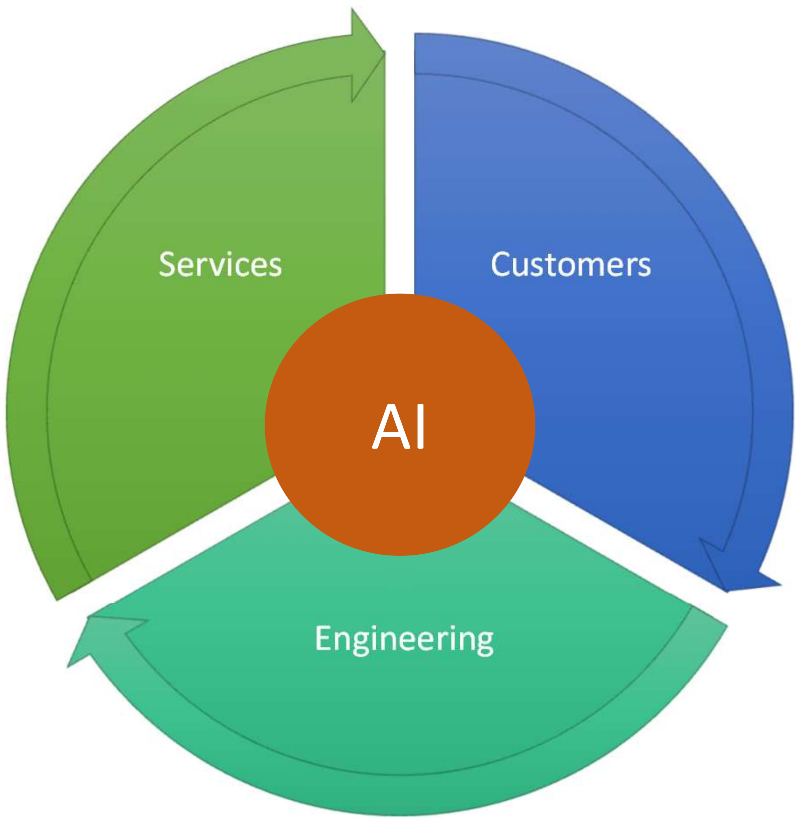
We are building Resource Central and modifying resource managers to use its predictions in Azure Compute

RC at the center of Azure Compute

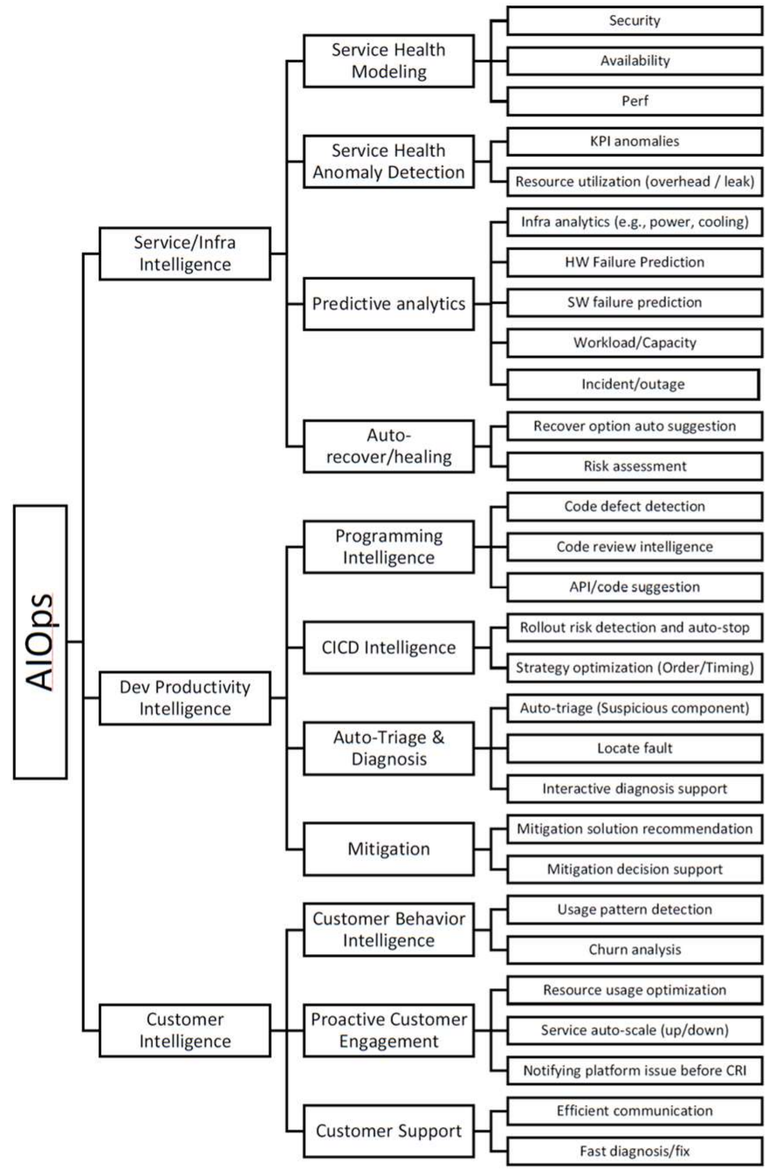


AI For Cloud: AIOps Solutions for Azure

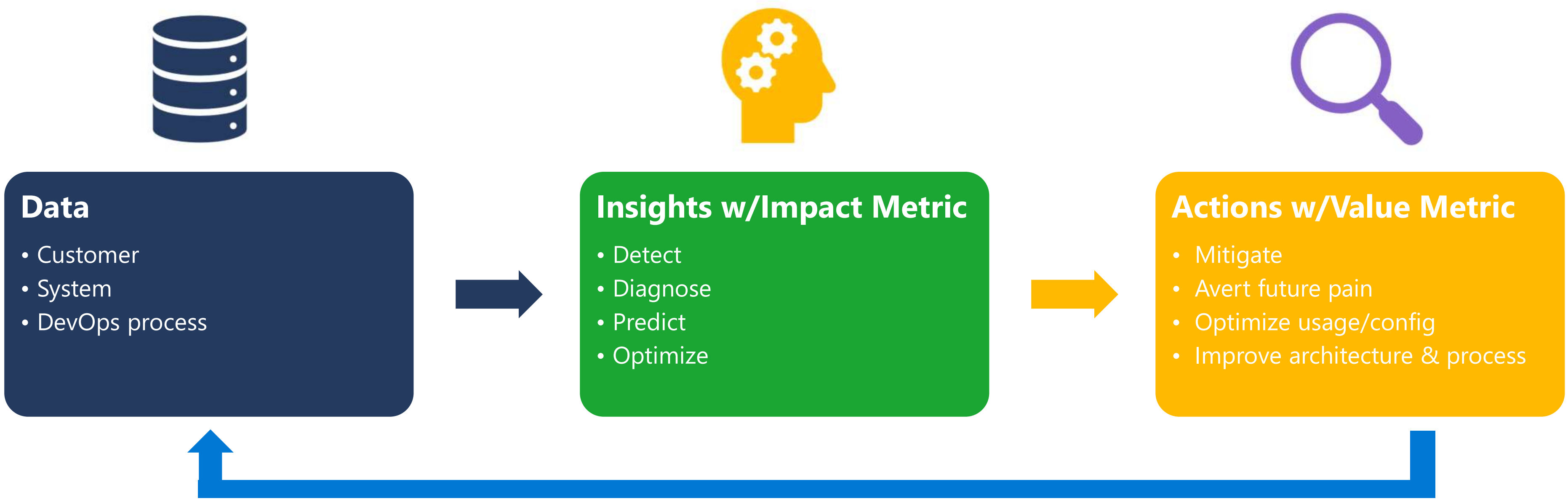
Problem Space



Infusing AI into Systems & Operations



Methodologies: From Data to Actions



- Measuring Customer and COGS impact of both Insights and Actions
- Improving intelligence through continuous feedback loop
- Driving architectural improvements for scalability, availability and reliability

Example: Dealing with Mem Leaks in Cloud

Data

Memory usage per Process for many instances

Training data: past several weeks, numerous time-series, large number of pivots

Volume: TBs of process data

Insights

Process **Foo** has memory leak

Mem consumption increase to '2n' MB on average (previous baseline: 'n' MB) in past **x** days

Geo scope: **y** machines in **z** clusters

Customer impact: creating new VMs in these nodes has 50% probability to fail

Repair action

Mitigation: restart process Foo

Repair:

- Collect memory dump
- Identify root cause
- Bug fix
- Testing in Stage
- Rollout to production
- Validation in production

Case Studies

1. Self-adapting platform through smart thresholds

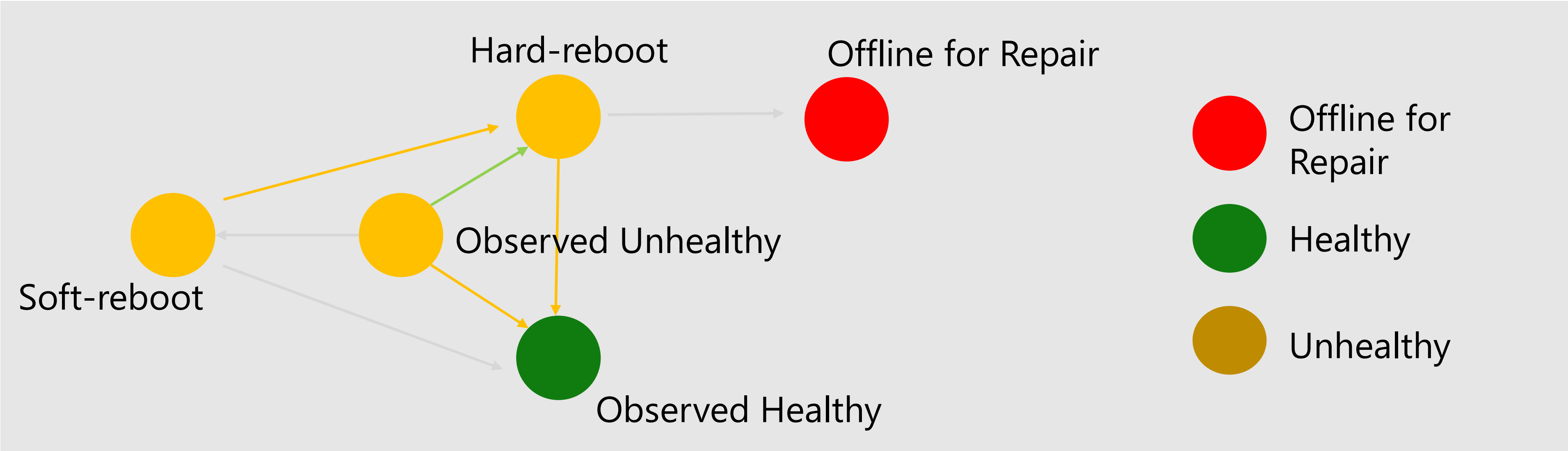
2. Resilient platform through failure prediction

3. Preventing platform regressions through Safe deployment

A Typical Problem: When to Timeout?

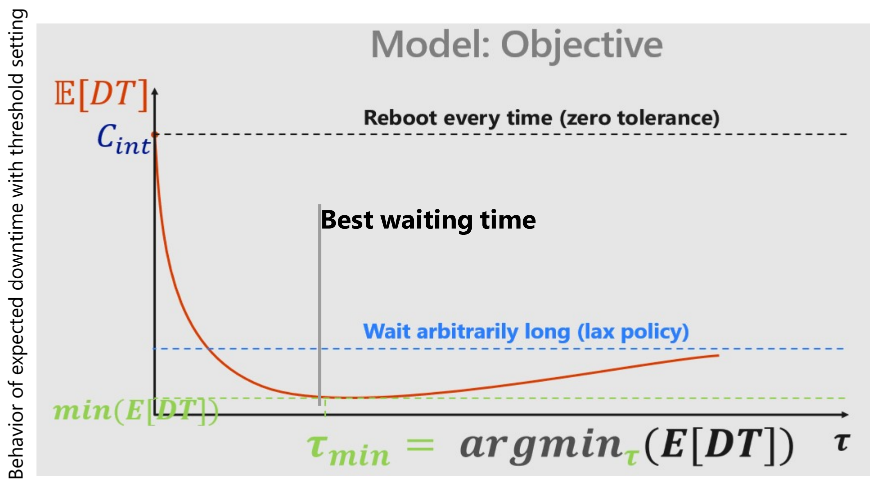
- Hard-coded thresholds in the platform leading to suboptimal decisions
- Thresholds can't be optimized in isolation

Example: node journey between online and offline



Self-Adapting Platform: Optimizing Timeout Thresholds

- Objective: minimize customer downtime caused by unhealthy host
- Unhealthy host: reboot or wait for auto-recovery?
  - Waiting too long will lead to long downtime duration.
  - Waiting too little will lead to more VM reboots.



Prediction Helps Improve Customer Experience

Microsoft

Transform Customer Experience through Prediction

- Avoid Customer Interruption such as VM Reboot and VM Downtime
- Optimize the data placement
- Proactive maintenance
- Allow high availability with low cost

Single Instance IaaS

Multi-Instance IaaS

PaaS Service

Storage System

Disk Configuration

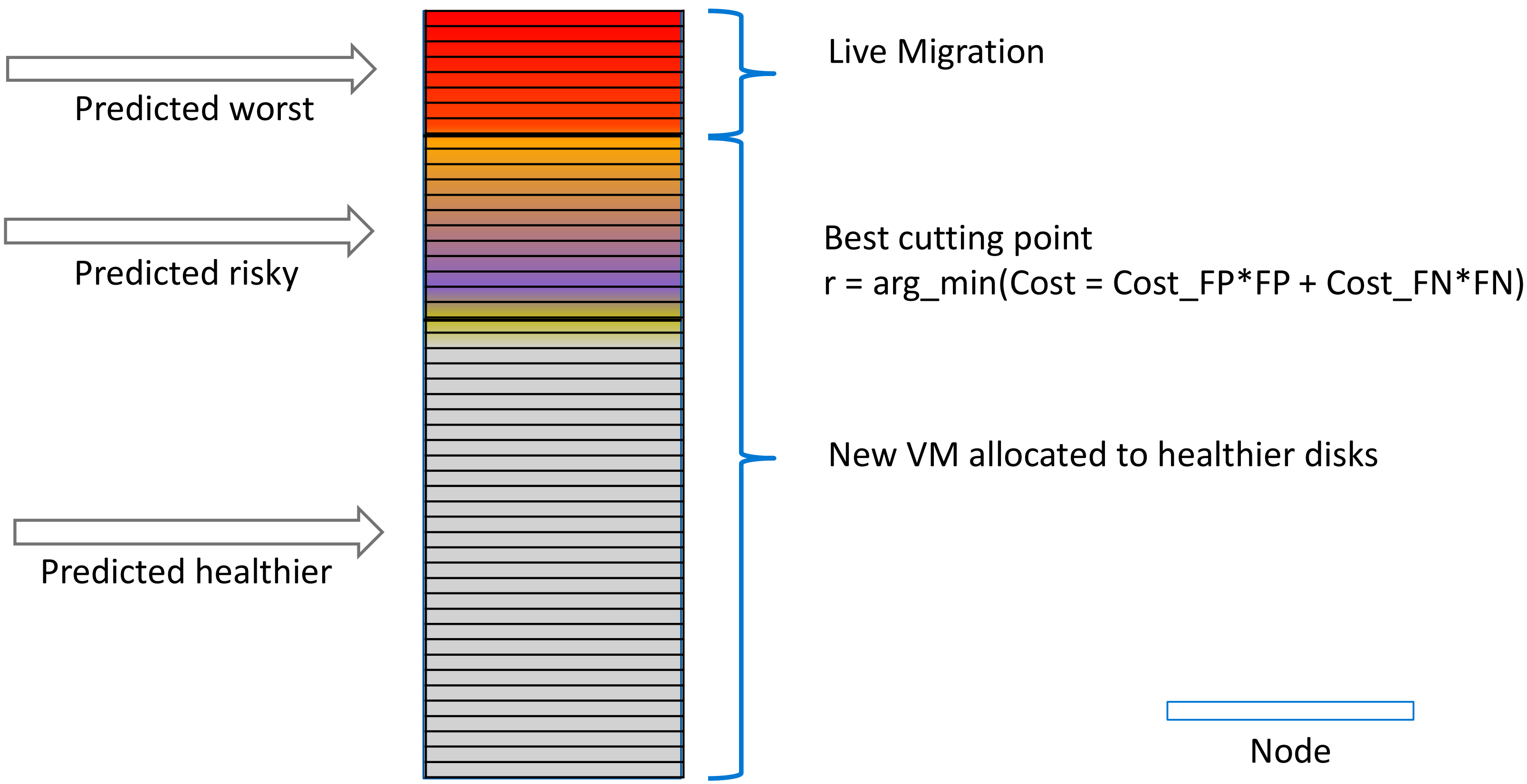
Low cost

Redundancy

Low throughput High cost Disk AFR	High throughput High cost Low client failure
Low throughput Low cost Disk AFR	High throughput Low cost Exaggerated failure

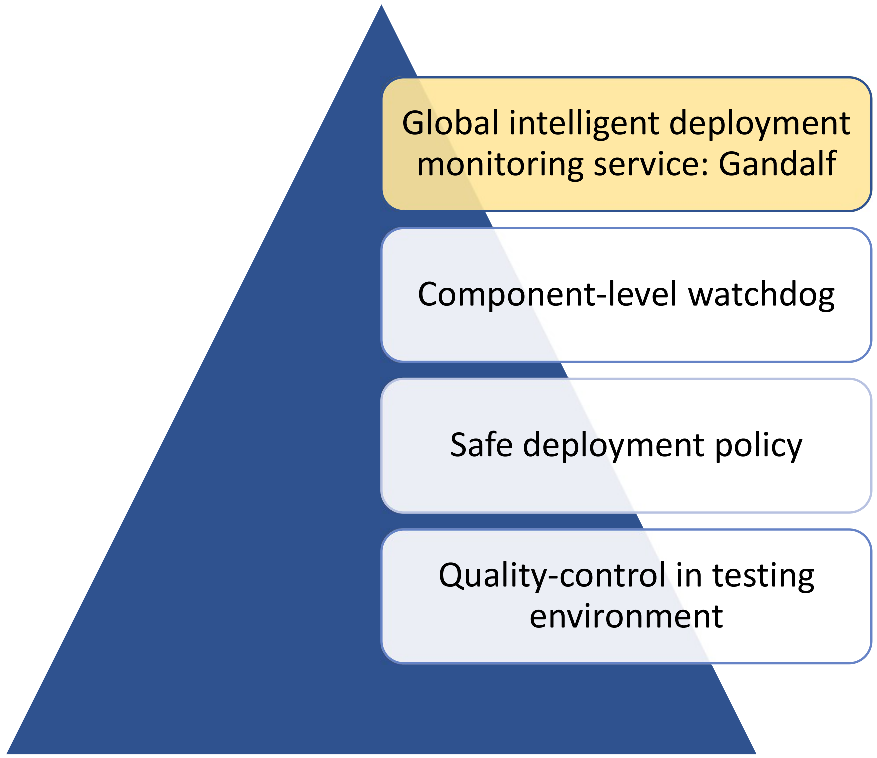
Disk Striping

Approach: Ranking Instead of Classification



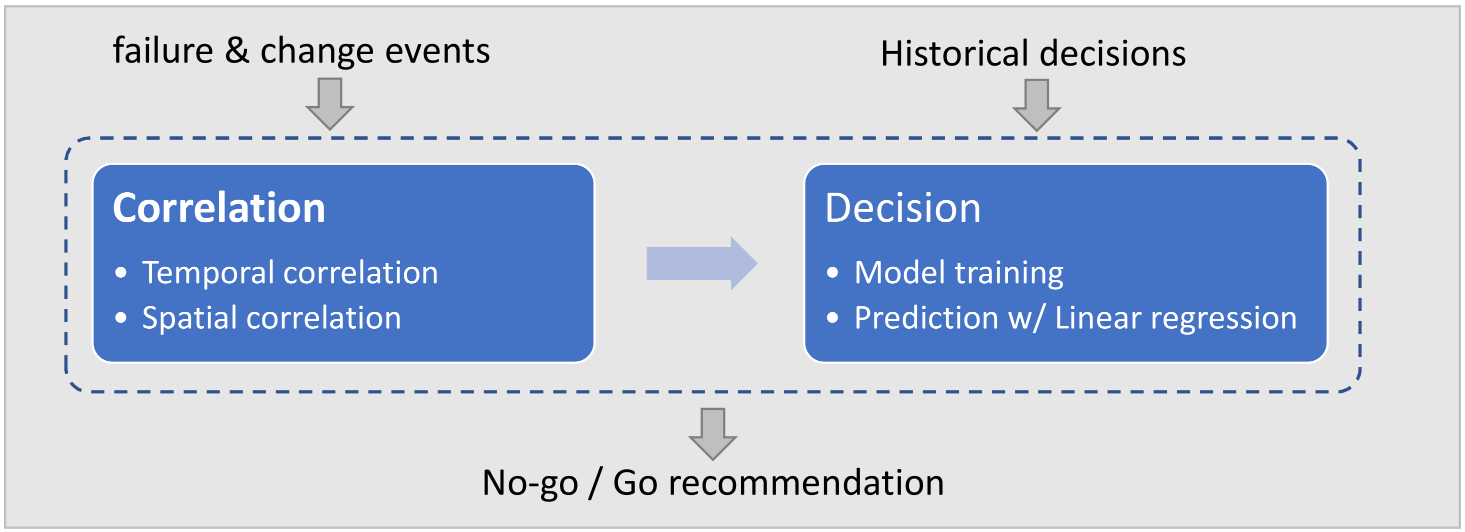
Improving Service Availability of Cloud Systems by Predicting Disk Error, USENIX ATC 2018

Azure's Four-layer Mechanism to Ensure Safe Deployment



Gandalf: Intelligent Global Watchdog

Model



System

- Lamda architecture for supporting both batch and stream-processing based decisions
- REST API to notify rollout orchestrator and Web frontend for supporting evidence

Gandalf: An Intelligent, End-To-End Analytics Service for Safe Deployment in Large-Scale Cloud, NSDI’20

AIOps in Azure: Summary

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

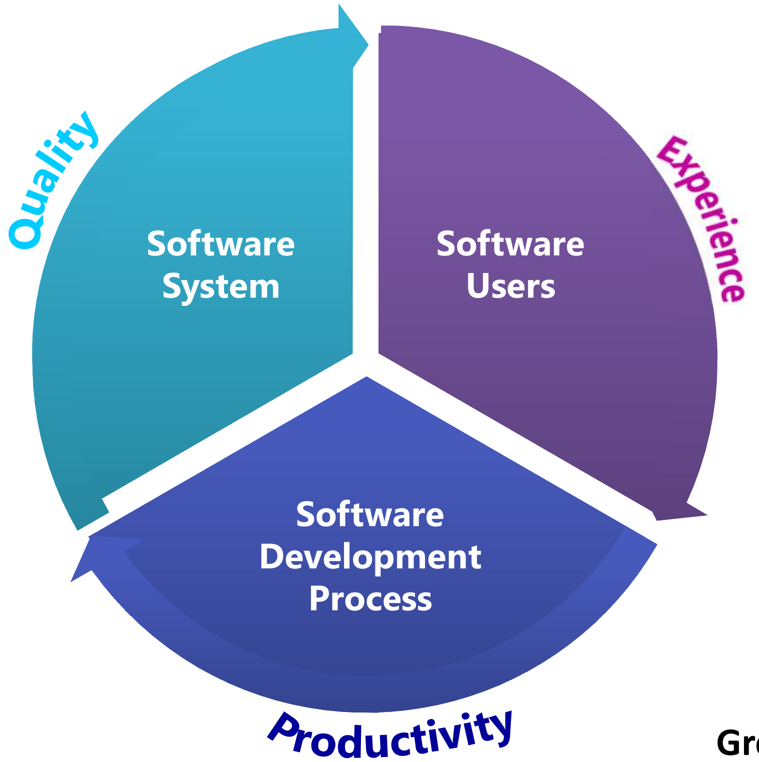
Microsoft

AI for Cloud: Related Research Areas

- **Software Analytics**
  - "Software analytics aims to obtain insightful and actionable information from software artifacts that help practitioners accomplish tasks related to software development, systems, and users.“ – *Dongmei Zhang, Microsoft Research*
- **Machine learning for systems**
  - “Traditional low-level systems code (operating systems, compilers, storage systems) does not make extensive use of machine learning today ” – *Jeff Dean, Google Brain*
  - MLCS 2018: First workshop on Machine Learning for Computing Systems

And more…

Software Analytics Research: 10+ Years from Microsoft Research Asia



Vertical

Horizontal

Information Visualization

Data Analysis Algorithms

Large-scale Computing

Great synergy between Software Analytics and AIOps

Making Industrial and Academic Impact

Contributing to broad Microsoft products

Azure
Microsoft Visual Studio
Windows
Power BI
Office 365
bing

Microsoft

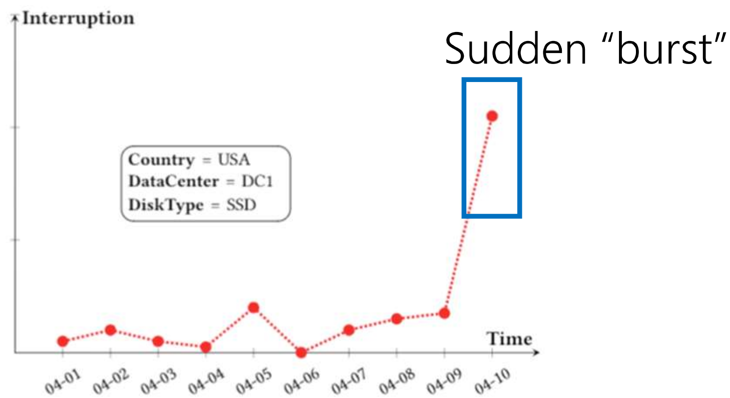
Contributing to multiple research communities: Software Engineering, Systems, Data Mining, and ML

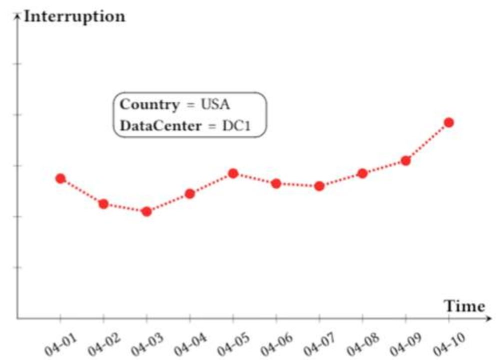
- An Intelligent, End-To-End Analytics Service for Safe Deployment in Large-Scale Cloud, NSDI'20
- Robust Log-based Anomaly Detection on Unstable Log Data, FSE'19
- Towards More Efficient Meta-heuristic Algorithms for Combinatorial Test Generation, FSE'19
- Local Search with Efficient Automatic Configuration for Minimum Vertex Cover, IJCAI'19
- Cross-dataset Time Series Anomaly Detection for Cloud Systems, USENIX ATC'19
- AIOps: Real-World Challenges and Research Innovations, Tech briefing, ICSE'19
- An Empirical Investigation of Incident Triage for Online Service Systems, SEIP, ICSE'19
- Outage Prediction and Diagnosis for Cloud Service Systems, short, WWW'19
- Identifying Impactful Service System Problems via Log Analysis, FSE'18
- Predicting Node Failure in Cloud Service Systems, FSE'18
- BigIN4: Instant, Interactive Insight Identification for Multi-Dimensional Big Data, SigKDD'18
- Improving Service Availability of Cloud Systems by Predicting Disk Error, USENIX ATC'18
- iDice: Problem Identification for Emerging Issues, ICSE 2016
- Log Clustering based Problem Identification for Online Service Systems, SEIP, ICSE 2016
- An Empirical Study on Quality Issues of Production Big Data Platform, SEIP, ICSE 2015
- YADING: Fast Clustering of Large-Scale Time Series Data, VLDB 2015
- Log2: A Cost-Aware Logging Mechanism for Performance Diagnosis, USENIX ATC 2015
- Correlating Events with Time Series for Incident Diagnosis, SigKDD'14
- Identifying Recurrent and Unknown Performance Issues, ICDM, 2014
- Mining Historical Issue Repositories to Heal Large-Scale Online Service Systems, ICDSN, 2014
- Where Do Developers Log? An Empirical Study on Logging Practices in Industry, ICSE 2014
- Contextual Analysis of Program Logs for Understanding System Behaviors, MSR 2013
- Software Analytics for Incident Management of Online Services: An Experience Report, ASE 2013
- Healing Online Service Systems via Mining Historical Issue Repositories, ASE 2012
- Performance Issue Diagnosis for Online Service Systems, SRDS 2012
- Mining Invariants from Console Logs for System Problem Detection, USENIX ATC 2010
- Mining Program Workflow from Interleaved Traces, SigKDD 2010
- Execution Anomaly Detection in Distributed Systems through Unstructured Log Analysis, ICDM, 2009
- Mining Dependency in Distributed Systems through unstructured log analysis, SIGOPS OS review 2009
- …

iDice – Identifying Emerging Issues From High Dimensional Data

Daily aggregation of Service Interruptions

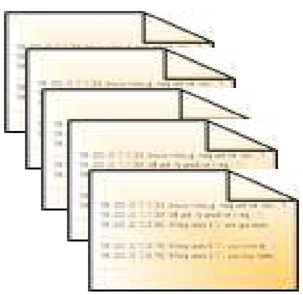
Time	Country	Datacenter	Disk Type	Interruption
2019-04-01	USA	DM1	SSD	1
2019-04-01	Australia	MEL21	SSD	1
2019-04-01	USA	DC1	HDD	4
2019-04-01	India	BL1	SSD	10
2019-04-01	UK	SN6	Hybrid	3
2019-04-01	USA	DM1	HDD	0
………	………	………	………	………



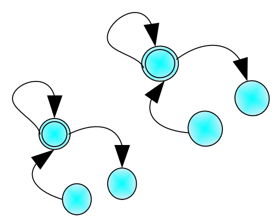


iDice: Problem Identification for Emerging Issues, ICSE 2016

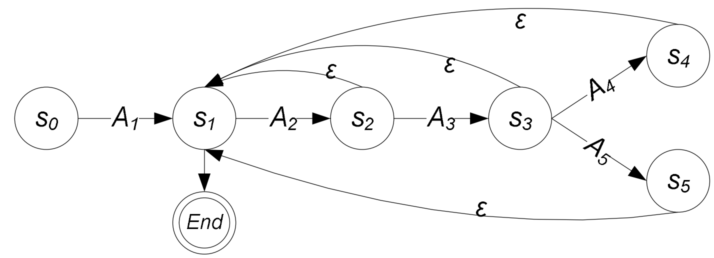
Interpreting System Behavior Semantics through Recovering Program Workflow from Logs



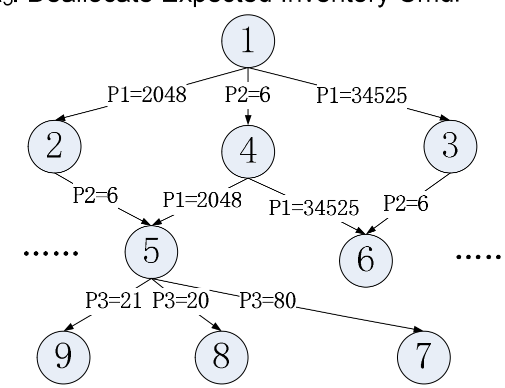
Log Sequences



Workflow Models



A_1: Find State Order Items; A_2: Verify State Order Items; A_3: Is Using ATP; A_4: Deallocate Existing Inventory Cmd; A_5: Deallocate Expected Inventory Cmd.



Learn Contextual Factors

*Mining Program Workflow from Interleaved Traces, SigKDD 2010

*Contextual Analysis of Program Logs for Understanding System Behaviors, MSR 2013

Conclusion

- AI for Cloud: an important vertical that AI can generate great value
- Our vision is infusing AI into platform and DEVOps process
- Azure experience and learnings on AI for Cloud
- Contributions to multiple research communities

You are welcome to visit Microsoft booth during main conference Feb 8-11!

Microsoft booth is #211 the 2nd floor - Rhinelander

Thank You!