

AIOps: Reactive to Proactive – Revolutionizing the IT Mindset

Live Webinar

April 9, 2024

10:00 am PT / 1:00 pm ET

Today's Presenters



Michael Hoard

Chair, SNIA Cloud Storage Technologies Initiative
Intel



Pratik Gupta

Chief Technology Office
IBM Automation

The SNIA Community



200
Corporations,
universities, startups,
and individuals



2,500
Active
contributing
members



50,000
Worldwide
IT end users and
professionals

What We Do



Educate vendors and users on cloud storage, data services and orchestration



Support & promote business models and architectures: OpenStack, Software Defined Storage, Kubernetes, Object Storage



Understand Hyperscaler requirements
Incorporate them into standards and programs



Collaborate with other industry associations

SNIA Legal Notice

- The material contained in this presentation is copyrighted by SNIA unless otherwise noted.
- Member companies and individual members may use this material in presentations and literature under the following conditions:
 - Any slide or slides used must be reproduced in their entirety without modification
 - SNIA must be acknowledged as the source of any material used in the body of any document containing material from these presentations.
- This presentation is a project of SNIA.
- Neither the author nor the presenter is an attorney and nothing in this presentation is intended to be, or should be construed as legal advice or an opinion of counsel. If you need legal advice or a legal opinion please contact your attorney.
- The information presented herein represents the author's personal opinion and current understanding of the relevant issues involved. The author, the presenter, and the SNIA do not assume any responsibility or liability for damages arising out of any reliance on or use of this information.

NO WARRANTIES, EXPRESS OR IMPLIED. USE AT YOUR OWN RISK.

Agenda

❖ Redefining AIOps

- Thinking holistically

❖ From Reactive to Proactive

- Shift Left and Future-proofing

❖ Beyond siloed solutions

- Holistic collaboration

❖ Automating the process

- Insight to Intelligent Action

❖ Conclusion



Redefining AIOps

Thinking holistically

The AIOps Outcome

To deliver IT services more reliably (match SLAs) with fewer service outages and to reduce MTTR (Mean Time to Recovery or Resolution) by using the power of AI.

Main Cause of Server and Application Downtime :

The Culprit

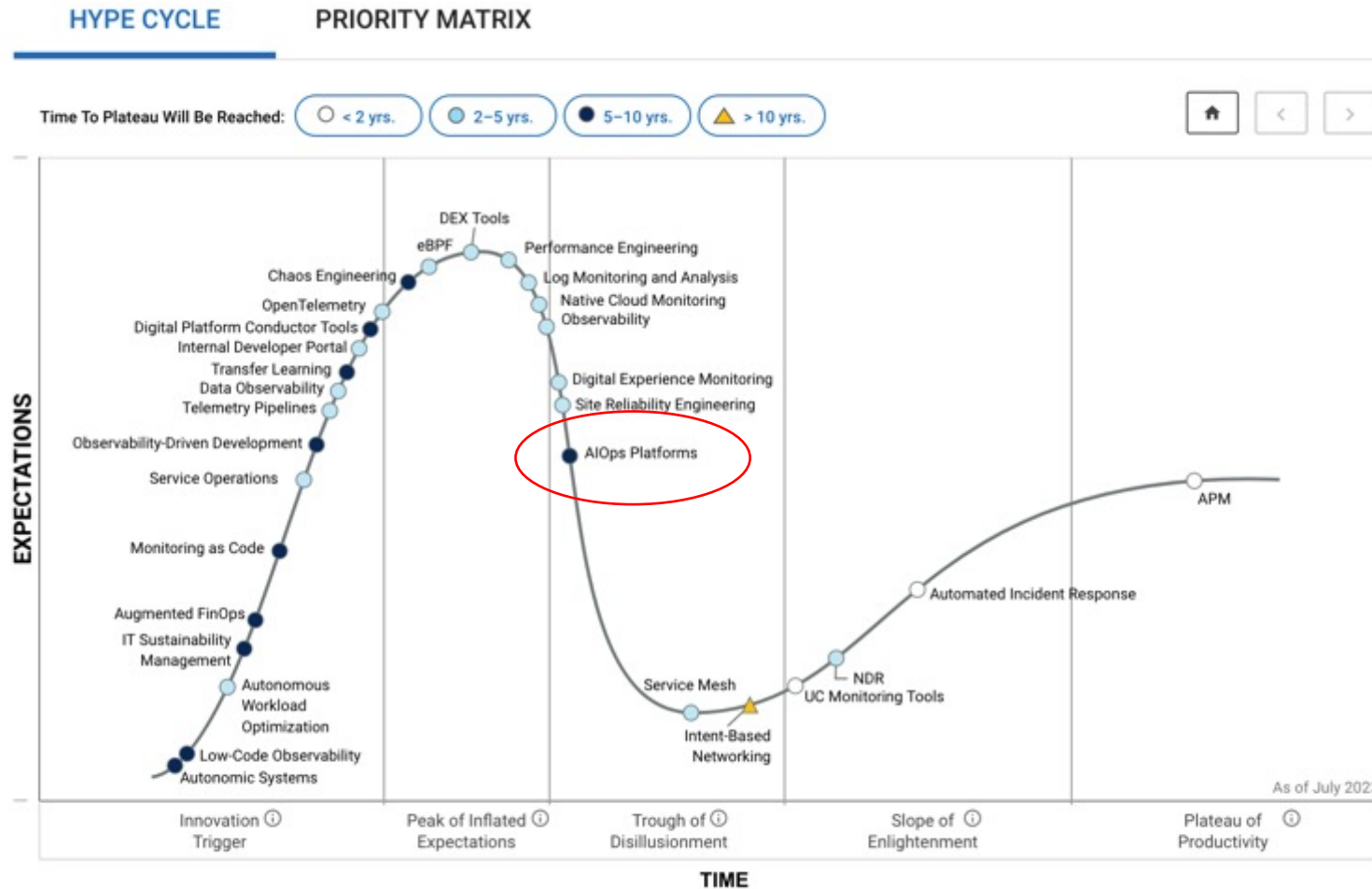
Security, Human Error, Software bugs

Exhibit 2. Security is the Top Cause of Server, Application Downtime



Source: ITIC 2022 Global Server Hardware, Server OS Security Survey

Disillusionment with AIOps?



Gartner Hype Cycle, 2023

- Considered part of ITOM (IT Operations Management)
- Considered a 'Tool'
- Reactive – focused on how to detect problems and not how to prevent problems.
- The technology is complex to deploy with high false positives.

Cost of Defect

Cost of Defects

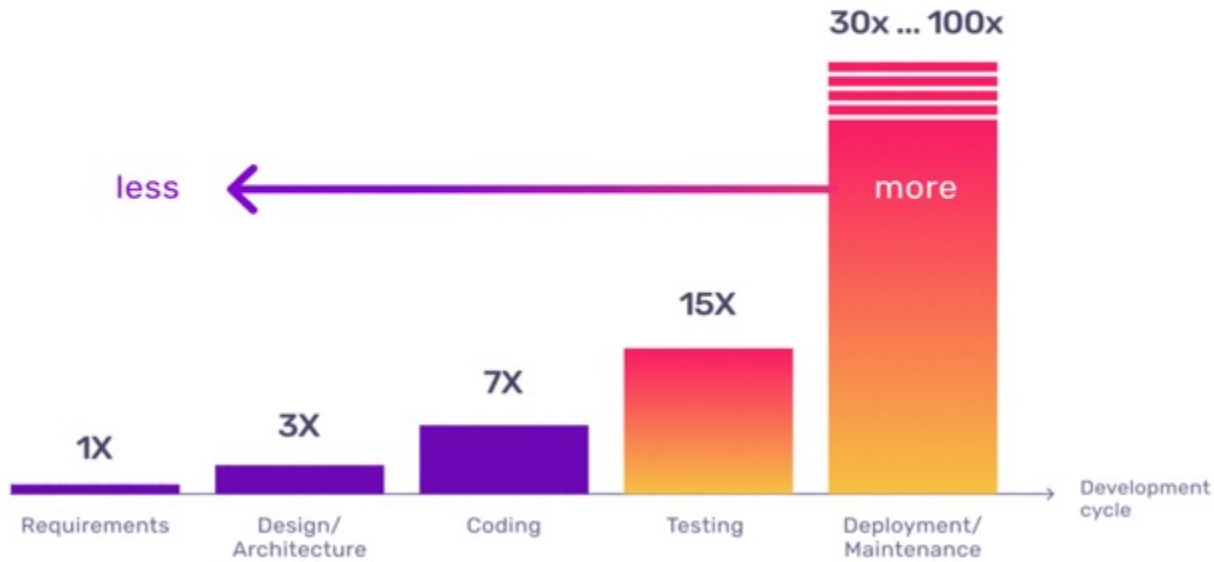
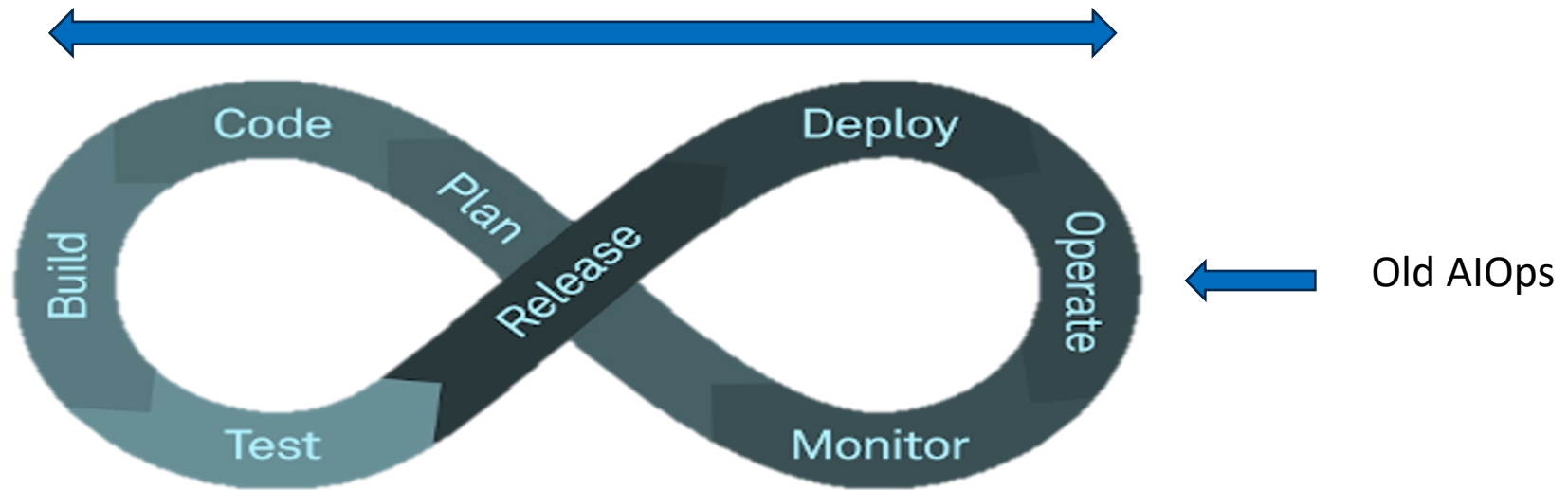


Image Source: "The Cost of Finding Bugs Later in the SDLC", Functionalize Blog (2023) – sourced from NIST

AIOps has to focus on finding defects earlier in the process.

The Technology of AIOps is “Shifting Left” to AI for IT

AI / ML across the Software Development Lifecycle (SDLC)



What is AIOps ? - The Practical Redefinition....

Old Definition: AIOps stands for "artificial intelligence for IT operations". It's a technology that uses machine learning (ML) and analytics to automate IT operations.

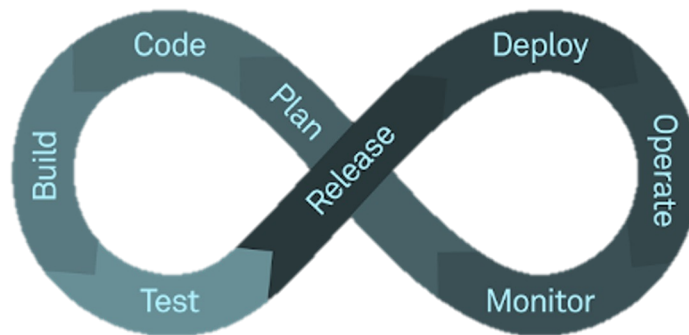
New Definition:

AI for IT is a set of **processes** that facilitate operational excellence and continuous improvement across Dev to IT, a set of **technologies, tools, and practices** that leverage AI and ML for delivering superior reliability outcomes across the complete software development and operations lifecycle.

From Reactive to Proactive

Shift Left and Future-proofing

Shift Left of AIOps



RCA -> AI Enabled Remediation

Event/Alert Reduction -> AI Enabled Root Cause Analysis

Dynamic tuning and optimization -> AI Enables Placement & Tuning

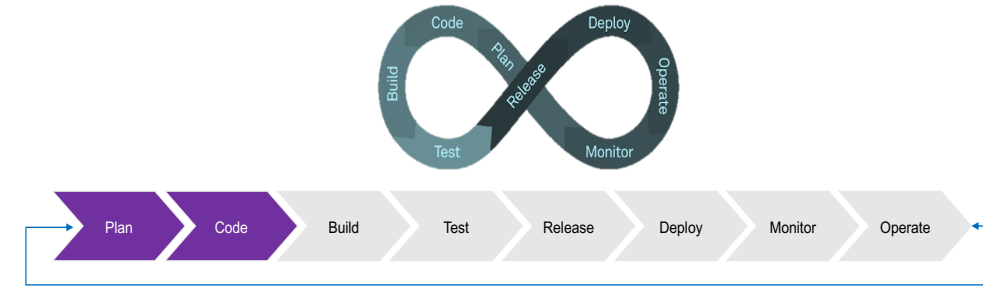
Shift-Left Observability -> AI Enabled Insights, Predictive Alerts

SAST, DAST, SCA -> AI Enabled SAST, DAST, SCA..

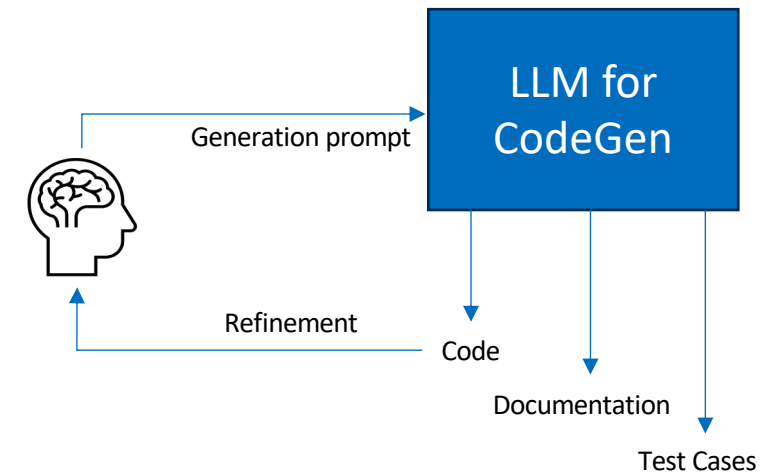
Change Risk Assessment -> AI for Change Risk Assessment

Reduce complexity
Modular services -> AI for CodeGen and Review

Reduce Complexity – AI for CodeGen and Review

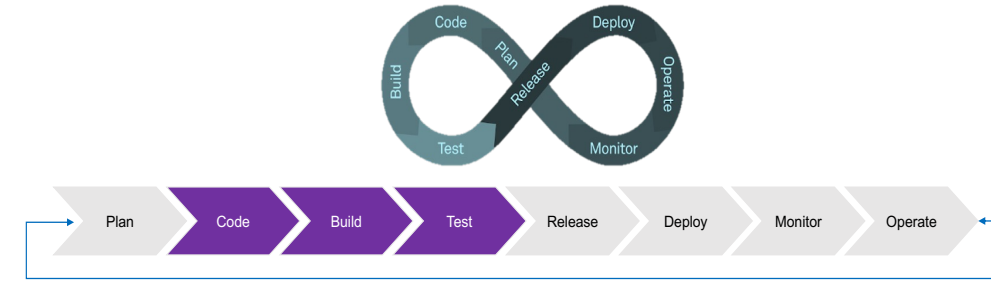


- ❖ LLMs being used for Code Analysis and Generation
 - ❖ Better explainability – explain existing code
 - ❖ Modular code generation
 - ❖ Code modernization
 - ❖ Automatic test case generation, including coverage
 - ❖ Automatic Documentation Generation
 - ❖ Almost any programming language

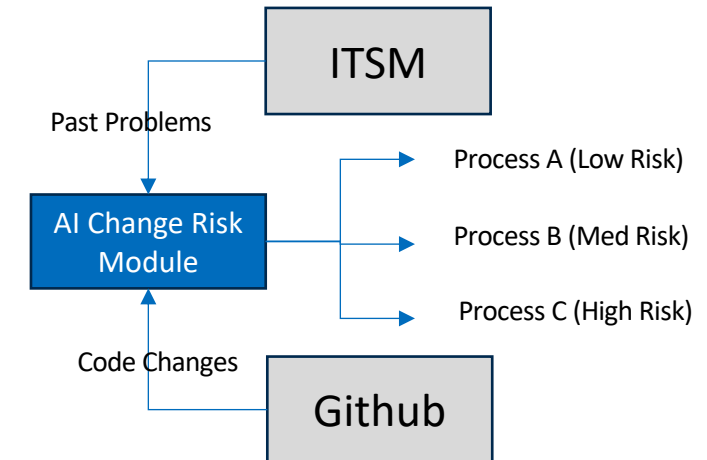


Early

AI for Change Risk Assessment

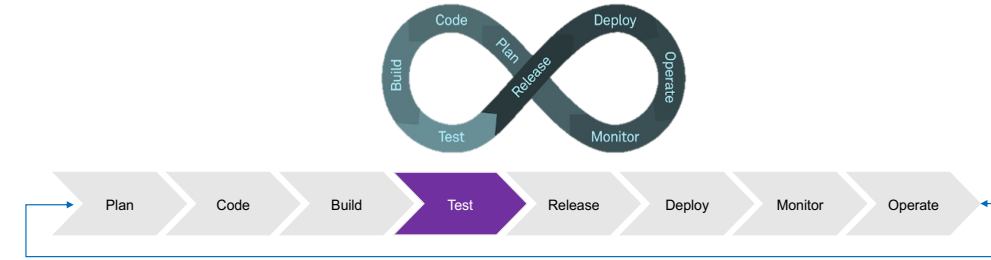


- ❖ Correlating modules changed with:
 - ❖ Past propensity for the introduction of problems
 - ❖ Criticality of module
 - ❖ Volume of changes
 - ❖ Other factors.....
- ❖ Route the request for different levels of review/testing.



Mature

AI-Enabled SAST, DAST, SCA

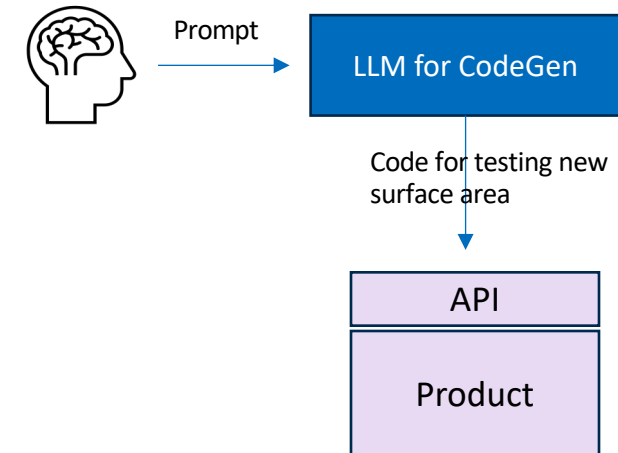


SAST: Static Application Security Testing (Code Scanning)

DSAT: Dynamic Application Security Testing

SCA: Software Composition Analysis

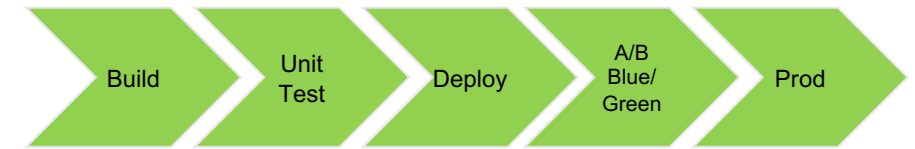
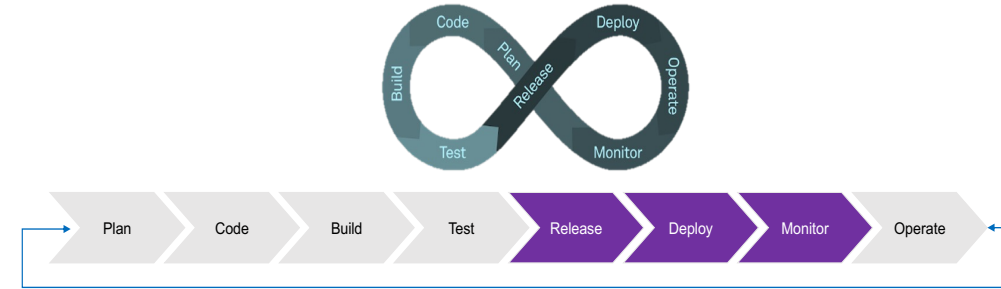
- ❖ Traditional SAST tools have a high rate of false positives and false negatives.
 - ❖ AI can help create custom query for testing specific vulnerabilities
 - ❖ Contextual code analysis by AI improves detection
- ❖ Dynamic (Black Box) testing – OWASP – Open Worldwide Application Security Project
 - ❖ AI to detect brute force attacks
 - ❖ AI-driven autonomous testing (learning-based API/response testing).



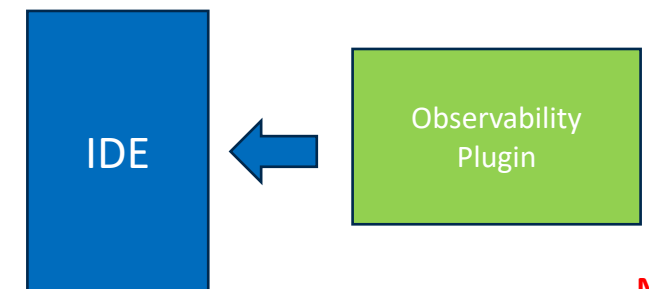
Early

Shift Left Observability – Insights and Predictive Alerts

- ❖ Observability tools are expanding to observe Dev Pipelines/CI/CD, test environments, security compliance
- ❖ Most modern Observability tools now have AI-based metric anomaly detection
 - ❖ Logs, Traces, Metrics collected across the microservices
 - ❖ ML used for predictive alerting (training still an issue)
- ❖ IDEs are integrating with observability tools for DevSecOps teams to see production performance (Developer Native Observability)

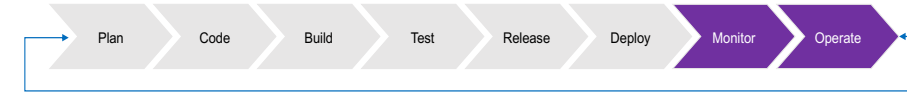
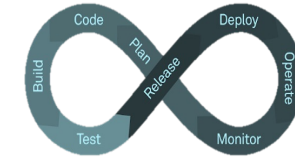


Single Observability Tool
Unified Anomaly Detection & RCA



Mature

Dynamic tuning and Optimization – Placement and Tuning



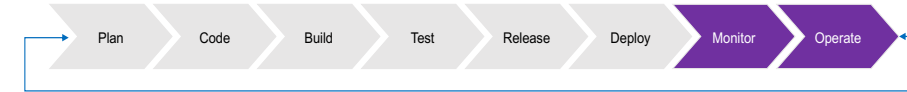
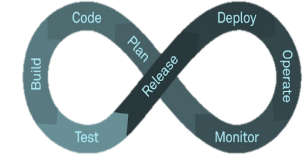
- ❖ Cloud Cost Optimization emerging as a bridge between FinOps and IT Operations
 - ❖ AI/analytics-based dynamic tuning of CPU, Memory, Storage and Cloud Services based on performance and cost objectives
 - ❖ Purchase of Reserved Instances(RI) calculated by advanced tools to optimize cost
- ❖ Reduces Errors and Alerts due to resource shortage and load spikes

- Measure end-to-end performance across the Application stack including intermediate points
- Understand end-user performance Latency
- Identify bottlenecks
 - CPU Scaling recommendations (CPU size purchase)
 - Memory size recommendations
 - Type of Storage (SSD, Spinning)
 - Placement (Which Cluster)
 - Predictive RI purchase
- Remediation Automation

- ☐ Resize down VMem for Virtual Machine mo1-035a-ixb-bp from 16 GB to 7 GB
Underutilized VMem in Virtual Machine mo1-035a-ixb-bp
- ☐ Resize down VCPU for Virtual Machine mo1-035a-ixb-bp from 4 to 2
Underutilized VCPU in Virtual Machine mo1-035a-ixb-bp
- ☐ Move Volume Vol-mo1-031a-ixb-bp-389_S_016 of Virtual Machine mo1-031a-ixb-bp from 389_S_016 to 389_S_028
Storage Amount Congestion
- ☐ Resize down Heap for Application Component mo1-031a-ixb-bp_csa02 [10.62.29.31] from 1 GB to 896 MB
Underutilized Heap in Application Component mo1-031a-ixb-bp_csa02 [10.62.29.31]
- ☐ Resize down Heap for Application Component mo1-031a-ixb-bp_csa01 [10.62.29.31] from 1 GB to 896 MB
Underutilized Heap in Application Component mo1-031a-ixb-bp_csa01 [10.62.29.31]

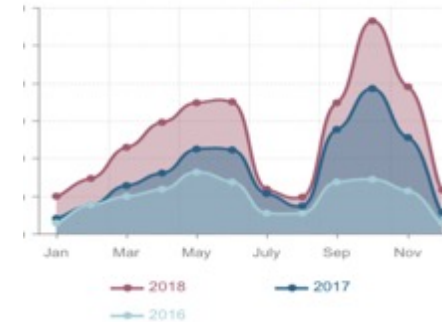
Mature

Event/Alert Reduction – AI based RCA

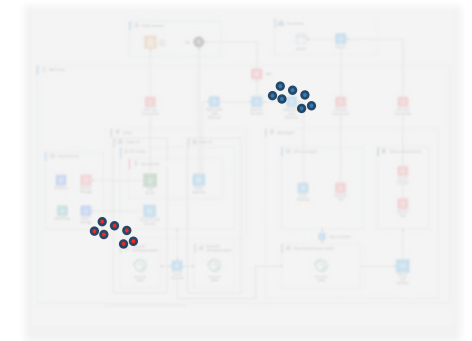


- ❖ Event analysis
 - ❖ Deduplication
 - ❖ Clustering
 - ❖ Seasonality
 - ❖ Event storm reduction
 - ❖ Root Cause detection (First Fault Detection)

Total Events By Month



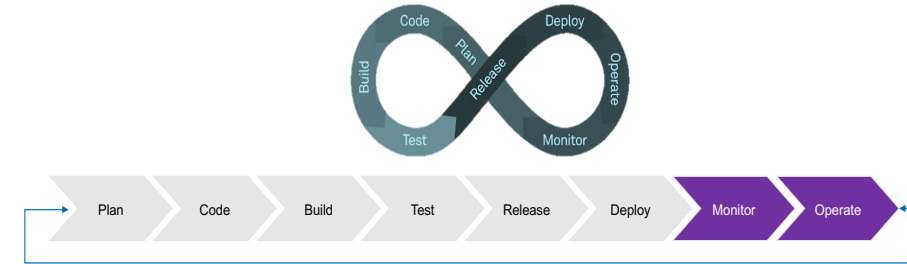
Seasonality



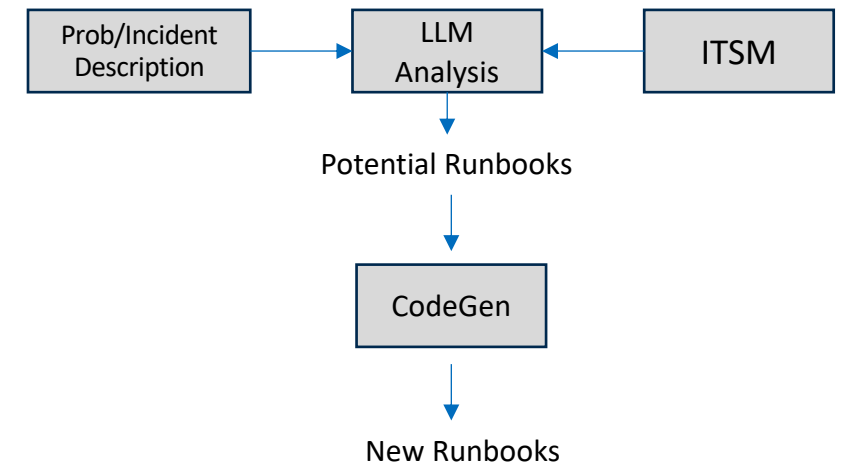
Clustering

Mature

RCA – AI-based Remediation



- ❖ Automated Run-book Identification
 - ❖ Use LLMs against Problem tickets to identify Potential Runbooks for use
 - ❖ Use AI CodeGen tools to update the runbooks to suit incidents quickly
 - ❖ Add to the library for automated execution



Early

Beyond Siloed Solutions

Holistic collaboration

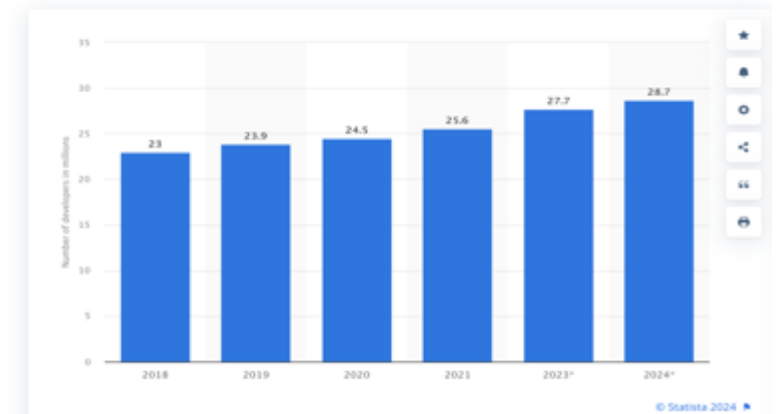
AI Ops is Enabling Collaboration Across Boundaries

Job Roles converging



- The emergence of AI-enabled, Developer Observability Platforms are blurring the lines between Developers, DevSecOps and SREs & IT Operations.
- Cybersecurity and AI revolutionizing DevSecOps.
- Convergence of tools taking place
- AI is breaking down information silos across Development, Security and Operations
- Creating a shared understanding of development and operations information and implications
- Enabling all roles to make better, more collaborative decisions

Number of software developers worldwide in 2018 to 2024
(in millions)



SRE Activity Breakdown



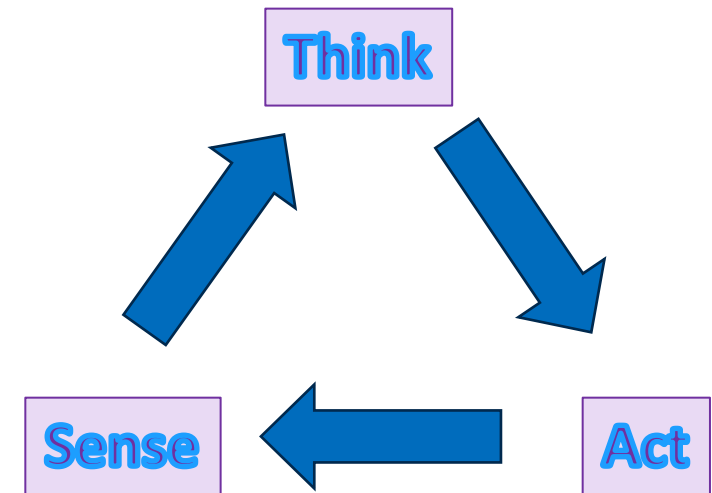
<https://pages.catchpoint.com/hubfs/Report/Catchpoint-2021-SRE-Report.pdf>

Automating everything

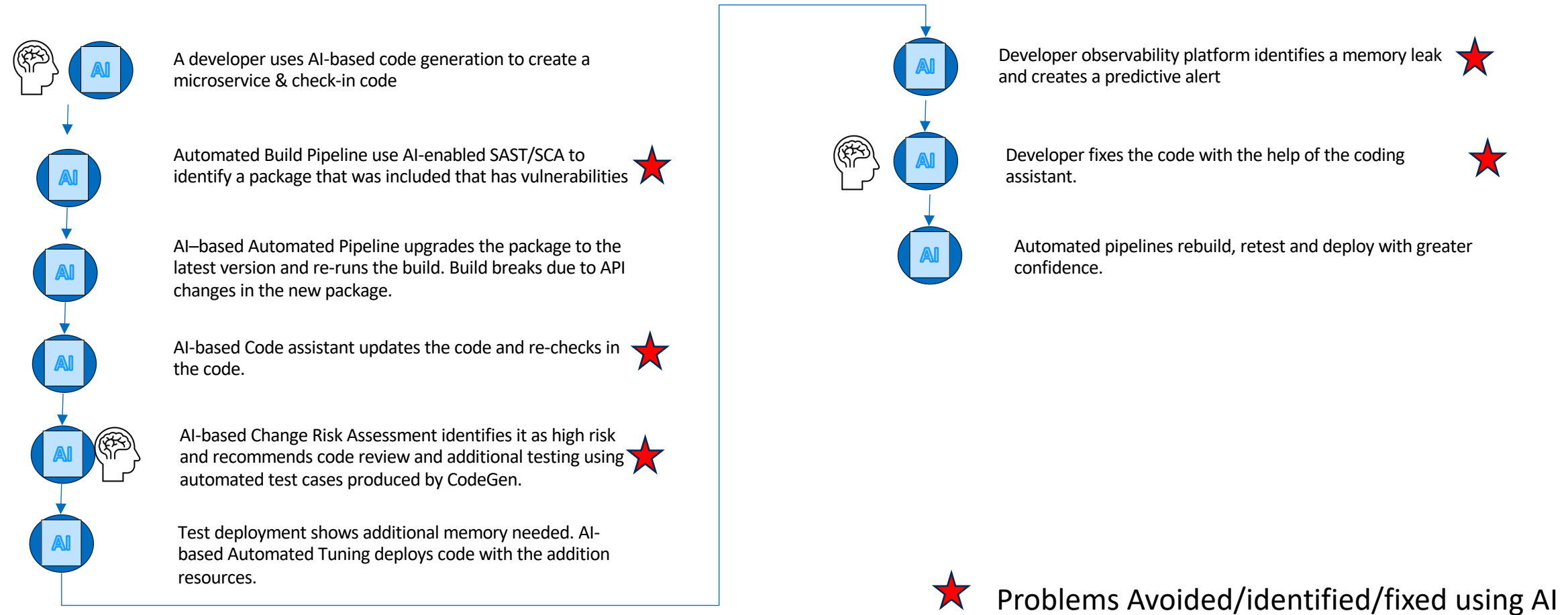
Insight to Intelligent Action

AI-Powered Automation

- Key Trends in Automation
 - AI-powered automation allows automation across
 - CodeGen – functional and security test – code review – deployment – Operations
 - Digital Assistants using Generative AI capable of reading documentation and past incident reports and providing support.
 - Digital Assistants using Generative AI to perform toil, fix bugs
 - Generative AI used to build more automation.



Imagine This is Possible Today !



Conclusion

- AIOps has evolved into AI Everywhere
- Job Roles are converging
- Continuous improvement of process and removal of silos are needed to take advantage of the advancements
- Problem avoidance and AI-based RCA are used early in the process to achieve improvements in Service Delivery (SLAs) and MTTR.





Questions?

References

Code generation

- <https://huggingface.co/blog/starcoder>
- <https://www.marktechpost.com/2024/03/14/top-artificial-intelligence-ai-tools-that-can-generate-code-to-help-programmers/>
- <https://www.ibm.com/products/watsonx-code-assistant>

AI based test case generation

- <https://read.engineerscodex.com/p/metasploit-llm-based-test-generator>

AI based change risk assessment

- <https://www.sogeti.com/ai-for-qe/section-3-1-inform-measure/chapter-5/>

SAST,DAST, SCA

- <https://www.mend.io/newsroom/mend-io-launches-mend-ai/>
- <https://www.kiuwan.com/blog/ai-how-it-is-changing-application-security-testing/>
- <https://apriori.dev/blog/dynamic-application-security-testing-the-advent-of-ai-driven-autonomous-testing>

Shift left observability

- <https://stackoverflow.blog/2022/07/18/how-observability-is-redefining-the-roles-of-developers/>
- <https://www.practical-devsecops.com/ai-in-devsecops>

Cloud Cost Optimization

- <https://www.ibm.com/products/turbonomic>
- <https://learn.g2.com/cloud-cost-optimization>

Automated remediation

- <https://thenewstack.io/aioas-done-right-automating-remediation-and-resiliency/>
- <https://www.redhat.com/en/technologies/management/ansible/ansible-lightspeed>

Thanks for Viewing this Webinar

- Please rate this presentation and provide us with feedback
- This webinar and a copy of the slides are available at the SNIA Educational Library <https://www.snia.org/educational-library>
- A Q&A from this webinar will be posted to the SNIA Cloud blog: www.sniacloud.com/
- Follow us [@SNIACloud](https://twitter.com/SNIACloud)



Thank You