# 7 WAYS BMC HELIXGPT REDUCES MANUAL TOIL TO ACHIEVE ZERO-TOUCH OPERATIONS

Modern enterprise applications are deployed in hybrid multi-cloud environments. The system of engagement for end users is supported by modern, cloud-scale architectures deployed as containerized microservices. The system of record requires hybrid architectures spanning cloud to mainframe for seamless integration between modern and legacy systems.

For end users—global customers, partners, or employees who interface with mission-critical business services for financial transactions or administrative tasks—business services are synonymous with the brand. Bad service quality or outages can have a negative impact, resulting in financial penalties and brand damage. The attention span of a mobile end user is measured in seconds; a bad mobile experience in any industry can result in subscriber churn. App stores place the power of switching loyalties in the hands of the end user, who can download, install, and delete applications at will in seconds. This is why it is critical to measure end user experience in the context of end-to-end business service performance and availability.

Organizations are constantly looking to improve operational efficiencies, reduce errors, and optimize productivity; however, these same organizations are also challenged with the burden of manual toil. The goal is to achieve zero-touch operations, where processes and operations require little to no manual intervention during unexpected disruptions to business services that span a multi-layered public and private IT landscape.

For zero-touch operations, the solution—in addition to ingesting observability artifacts and integrating with service management solutions for incident and change—will apply correlation, predictive, causal, and generative artificial intelligence, and machine learning (AI/ML) algorithms to recommend and automate actions to remove manual steps. Additionally, a conversational AI-based experience enriches and personalizes the user experience.

In this blog post, I would like to highlight seven core AIOps capabilities that are required by organizations to successfully achieve zero-touch operations (see Figure 1):

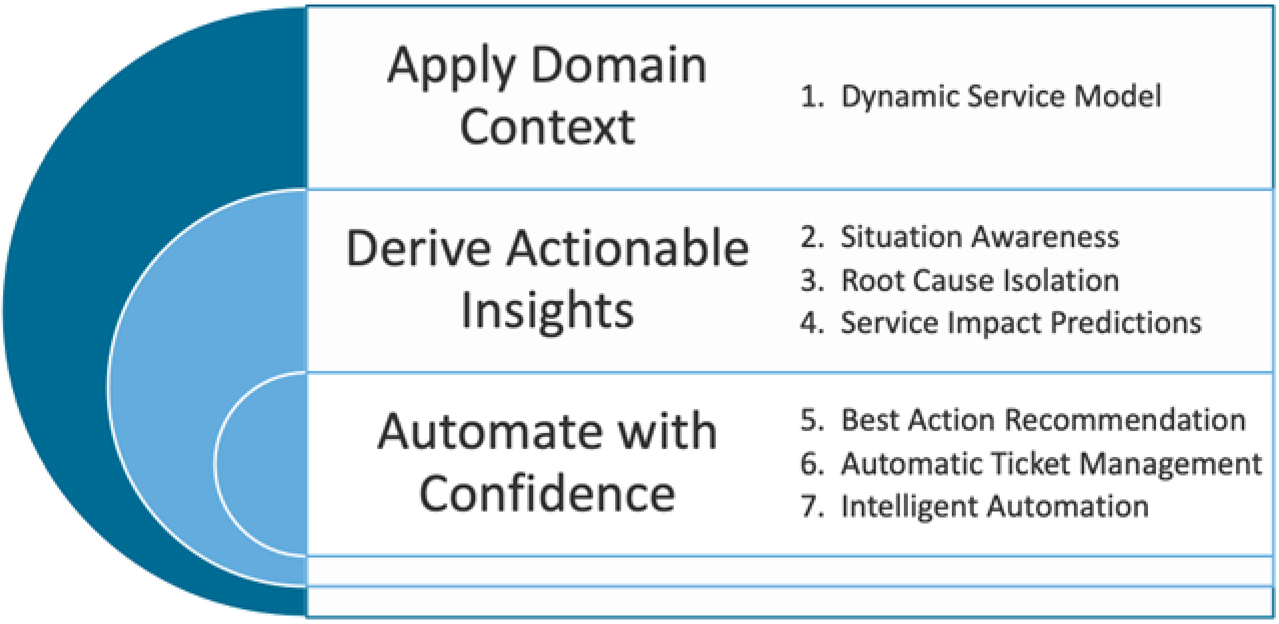

Apply Domain Context
1. Dynamic Service Model
Derive Actionable Insights
2. Situation Awareness
3. Root Cause Isolation
4. Service Impact Predictions
Automate with Confidence
5. Best Action Recommendation
6. Automatic Ticket Management
7. Intelligent Automation

Figure 1. Seven core AIOps capabilities for managing complex and constantly changing hybrid cloud environments.

These seven core AIOps capabilities are used to apply domain context, derive actionable insights from data, and, finally, automate the best action with confidence.

## Apply domain context and derive actionable insights

The criticality for an impacted business service is measured by the service impact. Figure 2 below overlays the hybrid multi-cloud deployment with the first four AIOps capabilities that model the service and build Situation awareness to identify root cause and assess service impact.

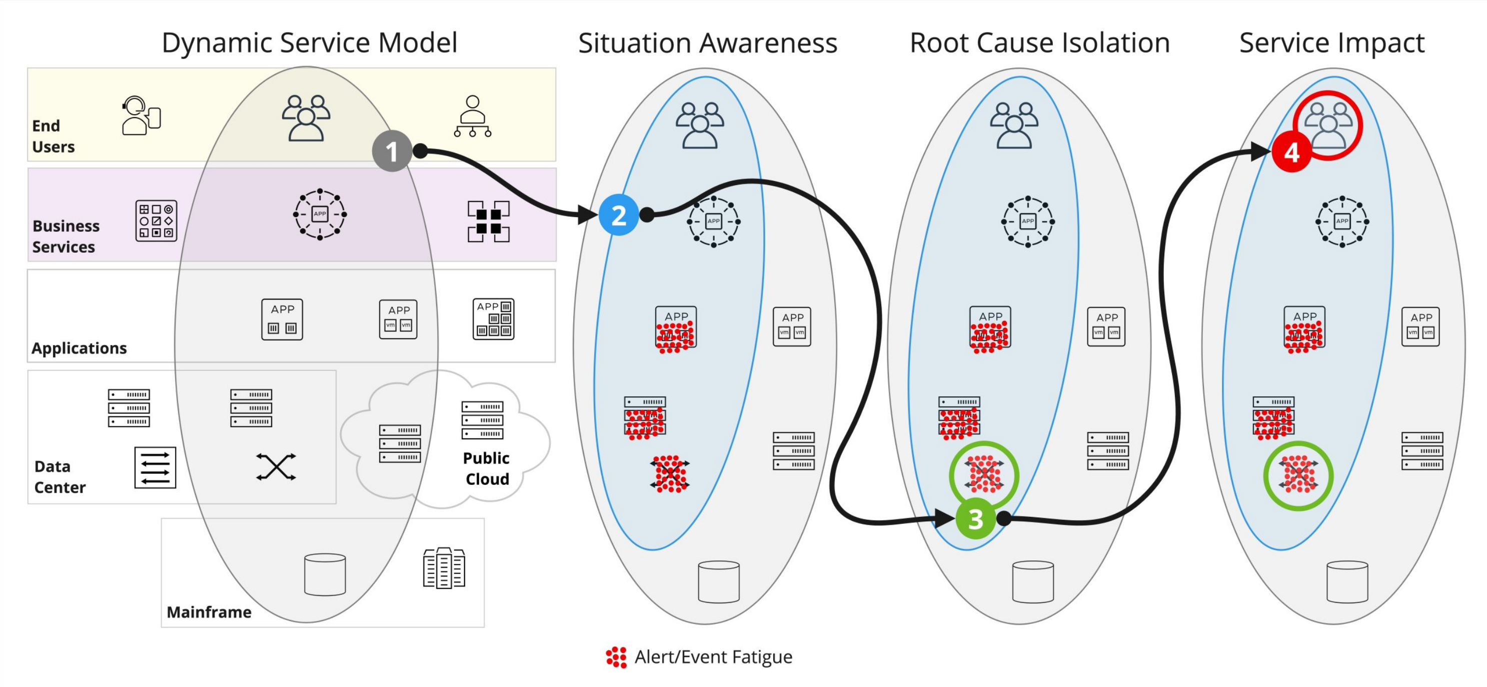


Figure 2. Model a dynamic service (1), build Situation awareness based on alertevent fatigue (2), perform root cause isolation (3), and assess service impact (4).

## 1. Dynamic service model

Dynamic service models are used to represent a business service (e.g., mobile banking, voicemail, etc.). The business service adds domain-specific knowledge, which in turn adds context and improves decision-making.

A business service is modeled from topology ingested by discovery and monitoring tools. The goal is to eliminate any manual tasks and automatically reconcile and dynamically update the end-to-end topology that spans application to network and cloud to mainframe. In Figure 2 above, the grey oval shape (1) encapsulates all the configuration items (CIs) and their relationships that represent a business service.

The dynamic service model provides the underlying connected topology as a foundation to apply AI/ML algorithms. The model represents domain-specific knowledge of the business service. Putting a boundary around the impactful CIs helps make informed predictions, root cause determination, and recommendations.

Business services are built automatically starting with a CI that represents a service (e.g., application name, database cluster name, VMware cluster name, Kubernetes name space, or network switch). Based on the starting point CI, related CIs are pattern-matched to automatically build and update the service model. This removes the need to manually build or maintain service models. You can refer to my previous blog post for additional details on service modeling and blueprints.

## 2. Situation awareness

A Situation represents awareness around an active issue that is impacting or has the potential to impact a business service (e.g., network switch issues could potentially impact mobile banking users if not triaged and fixed in a timely manner).

Alert/event fatigue is a common problem for the service desk that is compounded by the complex and distributed nature of modern applications. For example, a network problem can trigger an alert/event storm that impacts every layer of the business service, from application to network and everything in between.

Deciphering the signal from the noise is beyond human cognitive abilities and requires an AI/ML-based approach to build Situation awareness of the problem. Correlation AI/ML algorithms are used to group alerts/events across the dimensions of time and text using clustering and natural language processing (NLP) algorithms. The resulting noise reduction draws focus to the problematic cluster(s).

The Situation clusters alerts/events in relation to the impacted CIs and is represented as a slice of the dynamic service model, as shown above in the blue oval shape (2) in Figure 2. While this Situation puts a lens on the brewing problem, at this stage, there is no root cause isolation or indication of current or future service impact. Noise reduction based on event/alert clustering is not enough to determine root cause.

## 3. Root cause isolation

Root cause isolation is required to identify the culprit in an active Situation, and then engage and/or automate best actions to restore an impacted business service.

Correlation AI/ML algorithms do a good job of noise reduction; however, they lack causation, which is required for root cause isolation. By the same token, algorithmic root cause isolation is required to eliminate the blame game that takes place in a typical war room. The ultimate goal is to minimize the impact of an outage and restore service.

To be deterministic and explainable, root cause isolation requires domain-specific knowledge. To achieve this, dynamic service models provide a third dimension of topology. Causal AI/ML algorithms are used to build a causal graph for the active Situation, perform graph traversal in the context of the alerts/events, and identify the root cause CI(s). In our example, the root cause CI is the network switch, represented by a green circle (3) in Figure 2 above.

The active Situation is automatically updated with the root cause CI. However, at this stage, the impact and urgency of the Situation is unknown. This can be determined by predicting the service impact.

## 4. Service impact predictions

Service impact analysis is required during an active Situation to assess the current or potential future impact to a business service. This helps identify the criticality to determine the best action needed to engage with ticketing and automation systems.

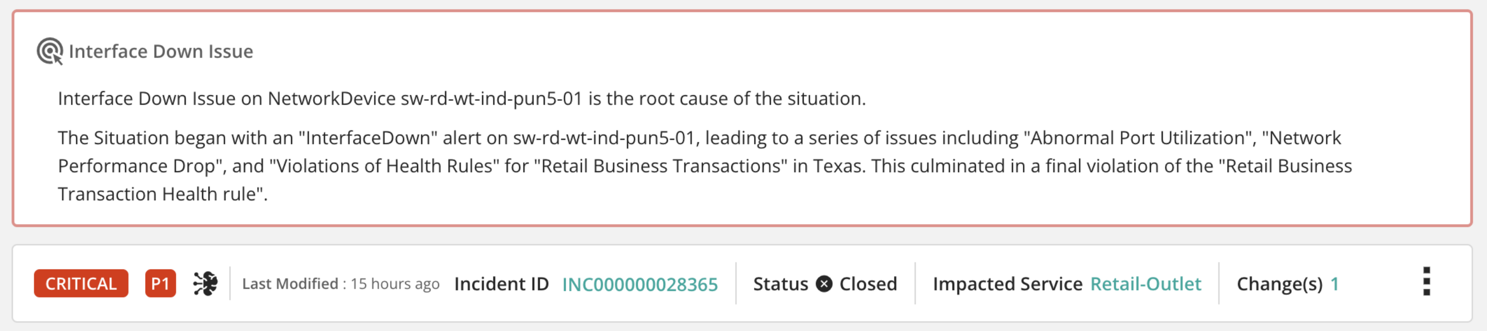
Service impact to a mission-critical business service can result in brand damage and financial penalties for an organization. Proactively assessing service impact is critical to prioritize the criticality of the active Situation, as shown in the red circle (4) in Figure 2 above.

Service impact is linked to the key performance indicators (KPIs) that are used to measure and assess the health and performance of a business service. KPI examples are service- and industry-specific (e.g., end user response time for a mobile banking application, voice quality for a voicemail application, user sentiment from social media feeds, number of transactions processed to measure revenue, or saturation events for infrastructure resources).

Predictive AI/ML algorithms are used to assess current or future KPI impact to the business service. They are applied against historical and current data to identify patterns and predict future outcomes. Common proactive approaches include:

- Predictions are used to report near-future deviations from threshold and/or normal behavior. This helps fix issues before users are impacted.
- Forecasting is used to assess and report on resource saturation. Proactively fixing resource saturation helps with planning and avoids potential outages.
- Univariate anomaly detection is used to observe metrics over time to find and report on outliers. This helps identify the needle(s) in the haystack, especially when looking across thousands of metrics.
- Multivariate anomaly detection is used to compare multiple metrics to find and report deviations from normal metric patterns. This helps find abnormal trends in data patterns across multiple dimensions.

With the Situation in place, the AIOps solution has successfully modeled the business service, reduced noise, identified the root cause, and assessed the service impact to establish criticality. The Situation also provides contextual input (prompts) for generative AI algorithms to accurately build a human-readable summary as shown in Figure 3 below. Note how context is captured based on the root cause, service impact, and causal chain of alerts/events to write an accurate problem summary based on event/alert data.



Interface Down Issue

Interface Down Issue on NetworkDevice sw-rd-wt-ind-pun5-01 is the root cause of the situation.

The Situation began with an "InterfaceDown" alert on sw-rd-wt-ind-pun5-01, leading to a series of issues including "Abnormal Port Utilization", "Network Performance Drop", and "Violations of Health Rules" for "Retail Business Transactions" in Texas. This culminated in a final violation of the "Retail Business Transaction Health rule".

CRITICAL P1 Last Modified : 15 hours ago Incident ID INC000000028365 Status ⊗ Closed Impacted Service Retail-Outlet Change(s) 1

Figure 3. Human-readable generative AI problem summary identifying root cause CI and its impact.

## Automate with confidence

Once the Situation has matured, the next step is recommending and automating the best action to take based on past Situations and historical ticket resolutions. The last three AIOps capabilities (5-7 below) help you make an informed decision and automate with confidence.

## 5. Best action recommendation

Best action recommendation is based on the analysis of past Situations represented in a knowledge graph and the processing of historical ticket data using generative AI.

### Knowledge graph and past Situations

The knowledge graph is a graph-based reasoning framework used to represent past Situations. The Situation captures the causal chain of alerts/events across the different layers (cloud to mainframe), isolates the root cause CI, and identifies the end user impact.

The Situation enriches the knowledge graph with a semantic meaning that represents real-world knowledge, allowing for intelligent reasoning and inference. A good example is the generative AI Situation summary described in the previous section, which summarizes the problem across text, time, and topology for event/alert data. The knowledge graph also pattern-matches similar Situations from the past, as shown in Figure 4 below.

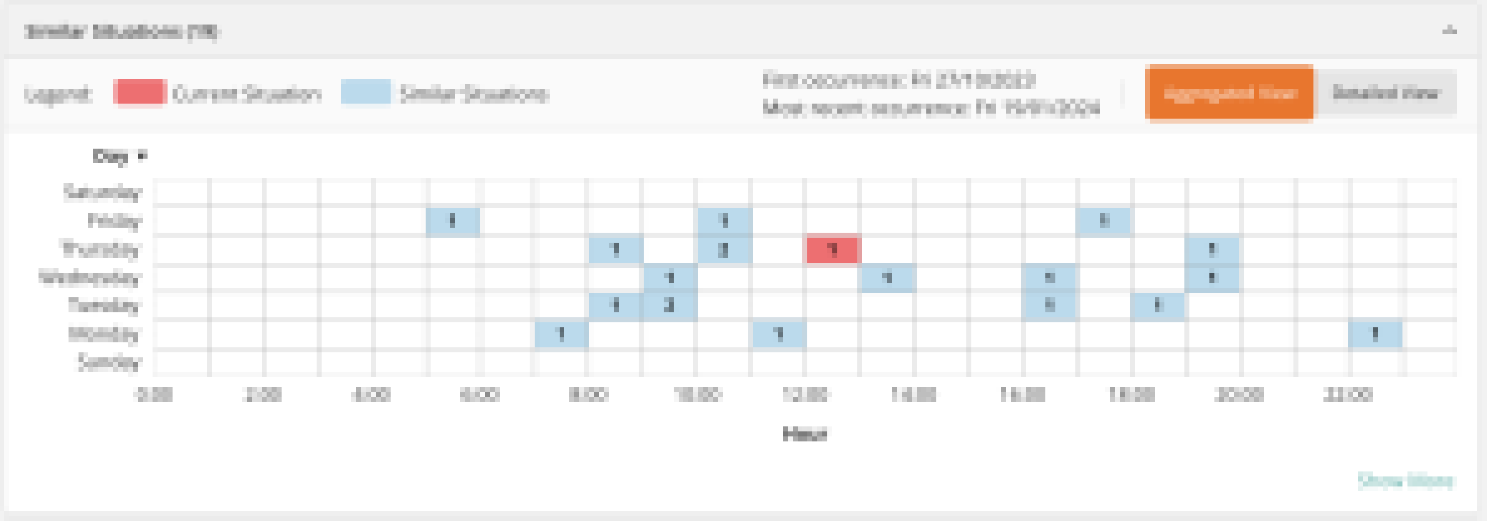
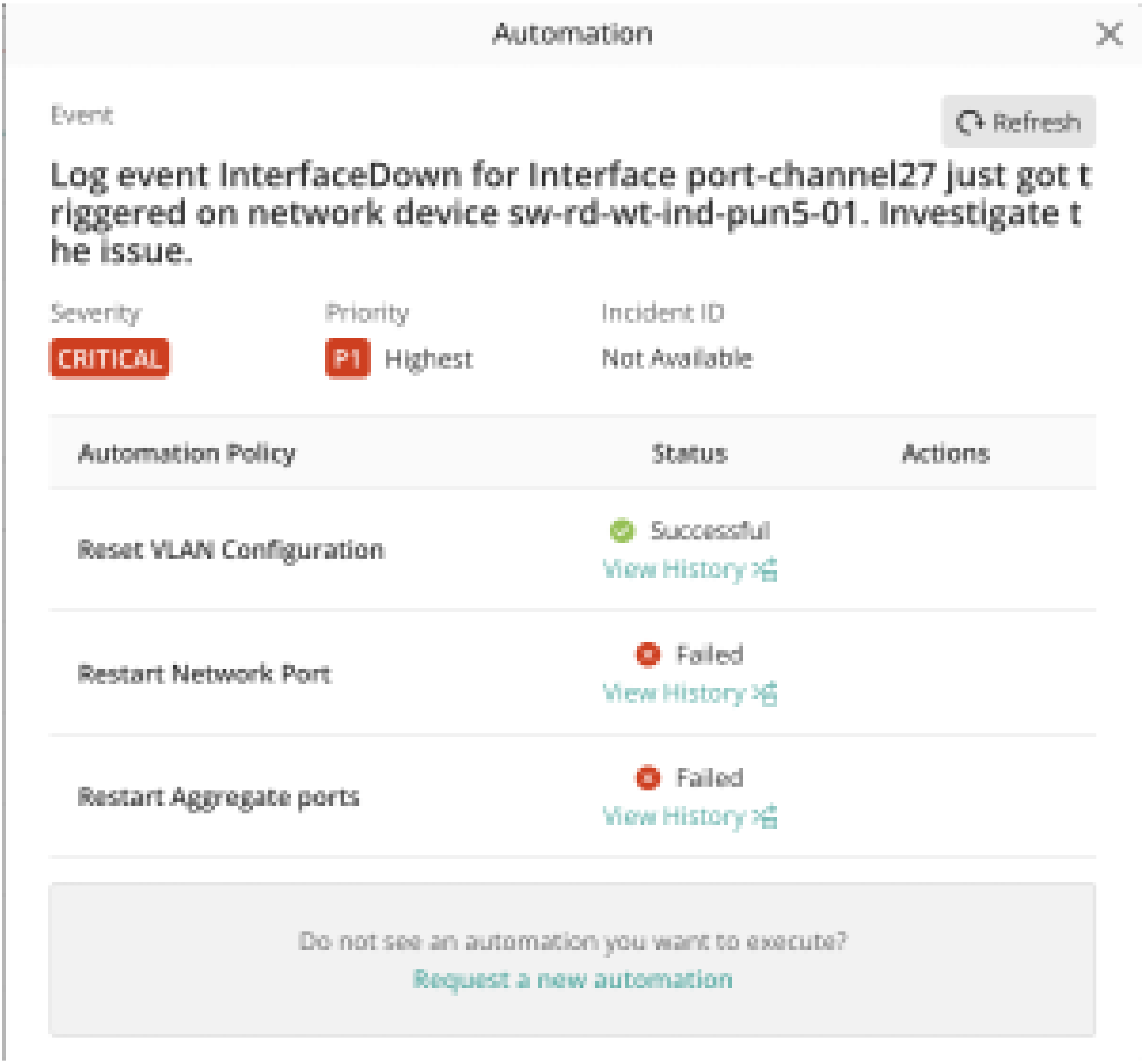


Figure 4. Similar Situation aggregated view for the past four months.

The knowledge graph learns and grows over time, recommending the best action based on past behavior. Past similar Situations in the knowledge graph are clustered and pattern-matched to provide automation recommendations based on the success or failure of past actions, as shown in Figure 5 below.



Event

Log event InterfaceDown for Interface port-channel27 just got triggered on network device sw-rd-wt-ind-pun5-01. Investigate the issue.

| Severity | Priority | Incident ID |
|---|---|---|
| CRITICAL | P1 Highest | Not Available |

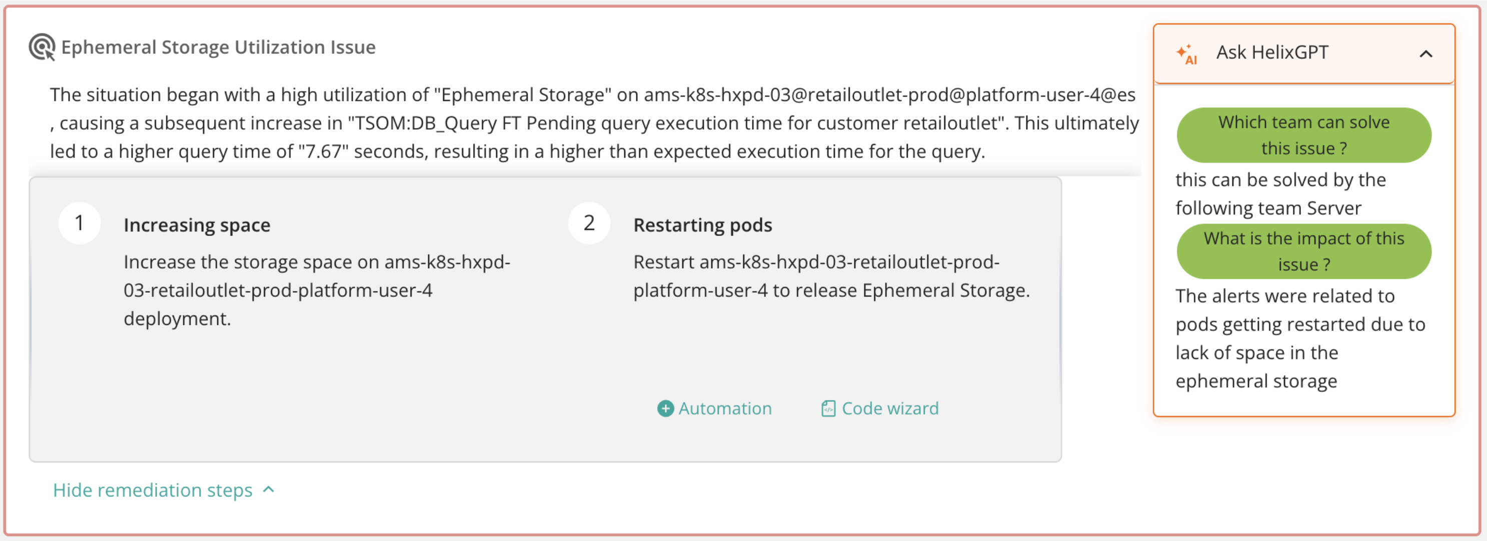
| Automation Policy | Status | Actions |
|---|---|---|
| Reset VLAN Configuration | ✓ Successful<br>View History | |
| Restart Network Port | ● Failed<br>View History | |
| Restart Aggregate ports | ● Failed<br>View History | |

Do not see an automation you want to execute?
Request a new automation

Figure 5. Automation recommendation based on past Situations.

## Historical ticket and large language models

Additionally, historical ticket data (incident, change, defect) is used to train a large language model (LLM) that provides a best action recommendation based on historical resolutions. Figure 6 below shows a generative AI-based Situation event summary, along with two recommended actions based on the processing of historical ticket data. Using generative AI, the trained model is asked, for example, how to fix, a storage issue. Based on the results, the user can automate or manually run a recommended automation, ask the LLM to generate an automation script using the code wizard, and/or chat with the model to get more information. In the example below, we click on “Ask BMC HelixGPT” to ask questions and better understand the issue’s impact and which team has solved the issue in the past.



Ephemeral Storage Utilization Issue

The situation began with a high utilization of "Ephemeral Storage" on ams-k8s-hxpd-03@retailoutlet-prod@platform-user-4@es , causing a subsequent increase in "TSOM:DB_Query FT Pending query execution time for customer retailoutlet". This ultimately led to a higher query time of "7.67" seconds, resulting in a higher than expected execution time for the query.

1 Increasing space

Increase the storage space on ams-k8s-hxpd-03-retailoutlet-prod-platform-user-4 deployment.

2 Restarting pods

Restart ams-k8s-hxpd-03-retailoutlet-prod-platform-user-4 to release Ephemeral Storage.

Which team can solve this issue ?

this can be solved by the following team Server

What is the impact of this issue ?

The alerts were related to pods getting restarted due to lack of space in the ephemeral storage

Figure 6. Generative AI problem summary and best action recommendation with actionable insight and conversational UI.

The Situation at this stage has enough context to confidently engage with other solutions to create incident(s) and change request(s), and then act by running automation tasks using automation tools. Based on root cause isolation, service impact prediction, and recommended best action, the system can automate with confidence.

## 6. Automatic ticket management

With the root cause and service impact identified, we can now automate the creation, prioritization, and routing of a ticket.

Create a single incident ticket for the Situation (not per each individual event) and target the right support group based on the CI(s) identified as the root cause. This eliminates the need for the first level of support to triage the Situation, and a second level of noise reduction is applied to reduce help desk incident fatigue. This also bypasses the need for a war room and eliminates the back and forth between different monitoring teams to establish root cause and ownership to fix the issue.

Assign ticket severity based on the current or predicted service impact. This helps prioritize the Situation so that support staff can focus and work on the most critical business-impacting issues.

## 7. Intelligent automation

Run the recommended automation based on resolution insights from similar Situations and past tickets. The recommended action(s) is based on the success of past actions and ticket resolutions, eliminating the need for support teams to manually process historical Situations and tickets. Change request approvals can also be automated depending on the change request risk assessment (e.g., restarting pods or scaling out virtual machines may be considered low risk, allowing for automated approvals).

To summarize, zero-touch operations can help revolutionize how organizations minimize customer outages and improve the end user experience. Adoption of AI/ML and automation with the seven core AIOps capabilities discussed in this blog provide the foundation to apply domain-centric service context, derive actionable insights from monitoring data that spans a complex multi-cloud IT landscape, and, finally, automate a best action recommendation based on historical success with confidence.

The AIOps capabilities in our BMC Helix Operations Management and BMC Helix Discovery solutions provide the foundation, and BMC HelixGPT brings the domain knowledge—typically held by subject matter experts—to apply correlation, predictive, causal, and generative AI/ML algorithms to solve complex IT operational issues.