AIOps

For Cluster Orchestration

White Paper

By Eric Muccino

Abstract

In the fields of information technology and systems management, IT operations (ITOps) is an approach or method to retrieve, analyze, and report data for IT operations. AIOps is the class of methods and procedures associated with the application of Artificial Intelligence and Machine Learning for ITOps. Mindboard is seeking to apply AIOps to improve the operations of container orchestration. In the cloud computing environment, AIOps can be used in conjunction with container orchestration to perform capacity management, event monitoring, and alerting/remediation for micro-services within a service network.

Contents

1. Introduction

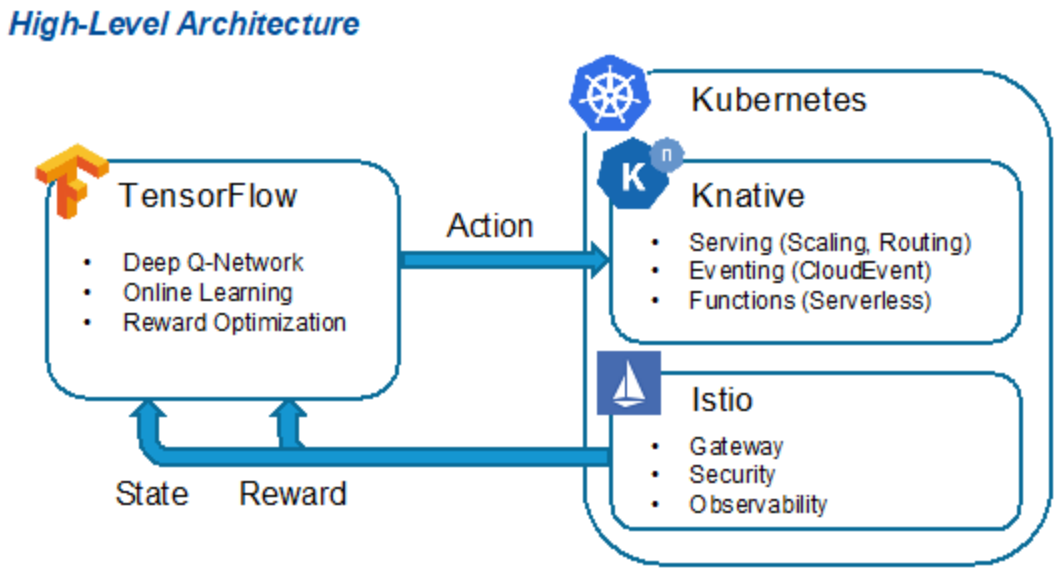
In the fields of information technology and systems management, IT operations (ITOps) is an approach or method to retrieve, analyze, and report data for IT operations. AIOps is the class of methods and procedures associated with the application of Artificial Intelligence and Machine Learning for ITOps. Mindboard is seeking to apply AIOps to improve the operations of container orchestration. In the cloud computing environment, AIOps can be used in conjunction with container orchestration to perform capacity management, event monitoring, and alerting/remediation for micro-services within a service network.

2. Use Cases

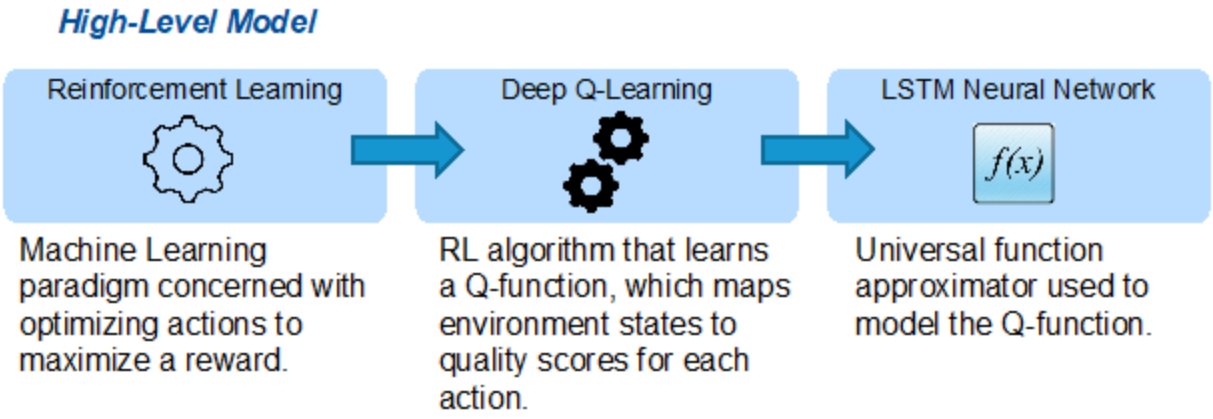
1. Capacity Management – Machine learning can provide resource autoscaling for optimized cluster performance and capacity. Specific implementations include:
    a. Vertical Scaling: Node instance type and node count scaling for decreased cluster resource cost.
    b. Horizontal Scaling: Replica set scaling for improved service response.
    c. Pod De-scheduling: Replica set placement for improved service resilience.
2. Event Monitoring – AIOps can be used for cluster data monitoring and synthesis for identification and prediction of network events and activities. Specific implementations include:
    a. Anomaly Detection: Classification of events, incidents, and traffic patterns to predict anomalies, distinguishing normal (routine/cyclical) from abnormal (threat potential) instances: involves association and ranking of the potential indicators around the core culprits, hence ensuring mission critical measures are taken before the faults develop in full.
    b. Event Correlation: Grouping of disparate metric, log, and telemetry data based upon hidden event cause and effect. Filtration and aggregation of data for more efficient event recognition, classification, and alerting.
3. Alerting and Remediation – Artificial Intelligence can enable proactive operation tools, covering:
    a. Alerting: Trigger incident escalation workflows, which involve the provisioning of alerts and response recommendations for relevant teams and subject matter experts, while reducing alarm fatigue and remediation bottlenecks.
    b. Incident Remediation: Automate remediation procedures by extracting and inferring incident root causes from alerts and event logs, and subsequently launching appropriate ITSM processes.

3. Implementation

AIOps is implemented by deploying an AIaaS (AI as a Service) for each task or use case. In the AIaaS architecture for AIOps, the AI agent acts as an external consultant to the service network cluster. The container orchestrator (e.g. Kubernetes) requests action recommendations from the AI agent by providing relevant cluster metrics. The AI agent uses the provided network data to update models and infer optimal action. The AI agent responds to the cluster with an action recommendation or prediction (e.g. scale resources, respond to security threat). AIOps leverages service meshes and serverless architectures (e.g. Istio, Knative) by making use of out-of-the-box cluster observability, security, and eventing, efficient and consistent implementation.



Almost all AIOps use cases require AI agents that can learn to produce actions that optimize a defined goal or reward function. Reinforcement Learning is a branch of Machine Learning that is designed for this type of task. There are many families of Reinforcement Learning algorithms, each with subtle advantages for different use cases. Deep Q-Learning, and its variants, provides a general algorithm framework that is suitable for learning to perform most tasks in the AIOps space, including autoscaling. Deep Learning models are implemented with open-source neural network libraries (e.g. TensorFlow, PyTorch).



4. Research

Mindboard's research goals are to research innovative AIOps design principles and develop commercial tools for canonical container orchestration platforms. Mindboard is interested in developing products that serve the use cases of capacity management, event monitoring, and alerting and remediation. Research and development efforts continue to build upon existing machine learning methods to create novel model architectures, data structures, and workflows, to be conducive for the container orchestration space.