

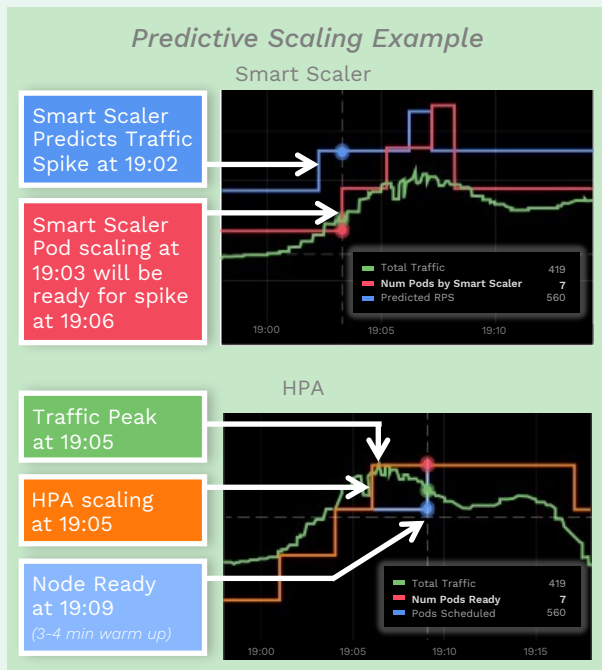


## Introduction

Smart Scaler is the world's first Gen AI based, predictive and fully-automated autoscaler for Kubernetes workloads. It is the only application aware autoscaler that optimizes for business logic outcomes. Smart Scaler has been shown to achieve up to 70% savings over HPA scaling and greatly improves DevOps productivity.

## How It Works

*Deployed as a SaaS, Smart Scaler analyzes application behaviors, service graphs to predict the pod count needed to maintain desired SLO. It uses Gen AI for pod scaling, adapting to traffic patterns and microservice behaviors, and integrates with an event calendar for scaling resources automatically.*



### Predictive Scaling

Note how in HPA (in bottom figure), the scaling is reactive and suffers from node startup delays. In Smart Scaler (in top figure), the scaling is predictive and hence able to mask node startup delays.

## Key Features

- Traffic-Based Scaling:** Utilizes algorithms to predict demand and adjust pod capacities in real-time.
- Calendar-Based Scaling:** Proactively scales pods ahead of scheduled events for seamless operations.
- Application-Metrics Based Scaling:** This is the only scaling automation that uses application behaviors such as service graphs, service-to-service latencies, error rates, RPS and CPU for accurately predicting scaling ratios for microservices.
- Lights-On/Lights-Off Use Case:** Automates pod startup and shutdown, aligning with working hours and schedule variations.
- Weekend and Weeknight Optimization:** Reduces resources during low-traffic periods to cut costs without affecting readiness.
- Karpenter Compatibility:** Enhances node scaling performance by offering predictability, reducing node latency.

## Benefits

- Resource Efficiency:** Addresses underutilization, with significant CPU savings in cloud-native applications.
- Cost Reduction:** Dynamically adjusts resources to avoid over-provisioning and idle costs.
- Reduces DevOps Burden:** To accurately scale based on application behaviors would take a DevOps person 2 weeks to set up for a set of microservices. Smart Scaler automates all that.
- Adaptability and Flexibility:** Easily adapts to schedule changes, maintaining operational efficiency.
- Sustainability:** Reduces resource wastage, aligning with environmental sustainability practices.

## Conclusion

Smart Scaler represents a shift in cloud-native pod scaling and resource management, leveraging application behaviors, predictive analytics and automation to enhance efficiency, reduce costs, and support sustainability.