# Application Capacity Planning with Agentic AI

## Introduction

Application Capacity Planning is a critical function for organizations managing IT infrastructure, ensuring that resources such as Virtual Machines (VMs), Physical Machines (PMs), Tanzu Application Service (TAS), Tanzu Kubernetes Grid Integrated Edition (TKGI), and OpenShift Container Platform (OCP) are optimally utilized. Leveraging Agentic AI, businesses can automate resource analysis, identify headroom issues, and proactively plan for scaling or reallocation.  
  
Agentic AI, characterized by its ability to autonomously interact with multiple systems, perform on-demand analysis, and drive decision-making, is well-suited for capacity planning. Below are some well-documented use cases demonstrating its application in this domain.

## Use Cases of Agentic AI in Application Capacity Planning

### 1. Intelligent Capacity Analysis and Optimization

Problem Statement:  
Organizations manage vast infrastructure with multiple computing resources, each generating performance and utilization data. These data points are siloed across different tables and databases, making real-time analysis and capacity forecasting difficult.

Solution Using Agentic AI:  
- A chatbot is developed to interact with the capacity planning system.  
- Upon a user query (e.g., 'For Application ID XYZ, show me resources with current headroom issues'), the AI agent autonomously queries multiple tables and aggregates the required information.  
- The AI model classifies resources based on over-utilization, under-utilization, and future headroom issues.  
- The chatbot suggests actions such as scaling up, scaling down, or reallocation based on predicted trends.

Business Benefits:  
- Eliminates manual analysis efforts, reducing human errors.  
- Provides real-time insights, enabling proactive decision-making.  
- Optimizes cost by ensuring efficient resource allocation.

### 2. Automated Resource Forecasting and Demand Prediction

Problem Statement:  
Organizations struggle with predicting resource needs for the future, often over-provisioning or under-provisioning resources, leading to inefficiencies and cost overruns.

Solution Using Agentic AI:  
The AI chatbot integrates historical usage trends, seasonal fluctuations, and machine learning models to predict future demand for resources.

Business Benefits:  
Avoids unexpected capacity shortages, helps in strategic budgeting, and reduces infrastructure costs through proactive scaling.

### 3. Dynamic Auto-Scaling Recommendations

Problem Statement:  
Current auto-scaling mechanisms are either manual or based on static thresholds, leading to inefficient scaling.

Solution Using Agentic AI:  
The AI agent monitors real-time usage data and triggers dynamic auto-scaling recommendations.

Business Benefits:  
Reduces downtime, optimizes performance, and automates scaling without human intervention.

### 4. Intelligent Incident Response for Capacity-Related Failures

Problem Statement:  
IT operations teams face challenges in identifying the root cause of performance issues related to capacity constraints.

Solution Using Agentic AI:  
The AI bot proactively analyzes system logs, infrastructure alerts, and telemetry data to detect issues.

Business Benefits:  
Reduces Mean Time to Resolution (MTTR), provides actionable insights, and improves system reliability.

### 5. Policy-Based Compliance & Cost Governance for Cloud Resources

Problem Statement:  
Many organizations experience resource sprawl due to uncontrolled provisioning, leading to excessive cloud costs.

Solution Using Agentic AI:  
The AI agent continuously monitors resource utilization vs. cost-effectiveness and enforces policy-based governance.

Business Benefits:  
Optimizes cloud costs, ensures compliance, and reduces financial waste while maintaining performance.

## Conclusion

Agentic AI presents a transformative approach to Application Capacity Planning, enabling organizations to automate resource analysis, predict future needs, and optimize infrastructure efficiently. By integrating AI-driven decision-making into IT operations, businesses can improve scalability, cost-effectiveness, and performance stability.  
  
The above use cases showcase how Agentic AI can be applied across different aspects of capacity planning, making operations more autonomous, intelligent, and efficient. Future developments in AI models and automation pipelines will further enhance these capabilities, leading to self-optimizing IT infrastructure.