

# SERRA LABS CPO

Cost-Performance Optimization for Cloud Applications

Q3/2020

Private and Confidential

# Cloud Cost-Performance Problem



Cloud Application Usage <u>Dynamically</u> changes while Cloud Application Capacity remains <u>Inelastic</u> resulting in Waste or Slowdown or Both

Waste due to Over-Capacity relative to Usage

Excess Costs



35% Waste (Flexera Cloud Surveys 2018-20) Slowdown due to Under-Capacity relative to Usage

User Dissatisfaction

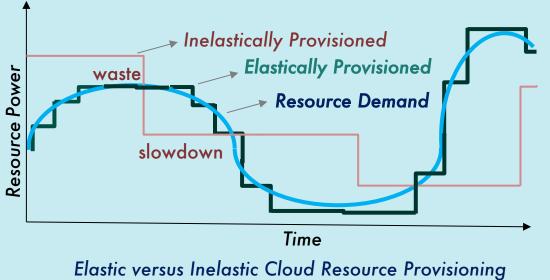


48% Slowdown (HIMSS Cloud Survey 2018)

# Addressing the Cloud Cost-Performance Problem

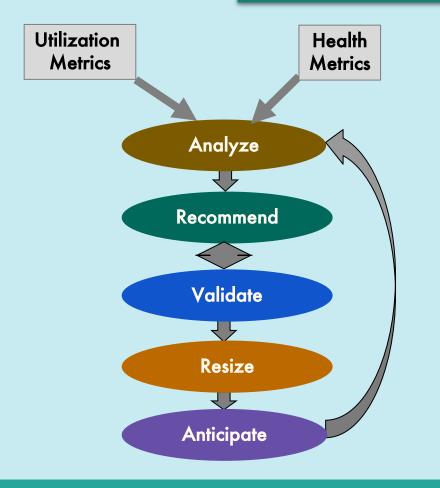


Cloud Capacity Management needs to match the Elasticity of Cloud Application Resource Demand to Minimize Waste and Avoid Slowdown



to Meet Dynamic Resource Demand

# **CPO** Approach

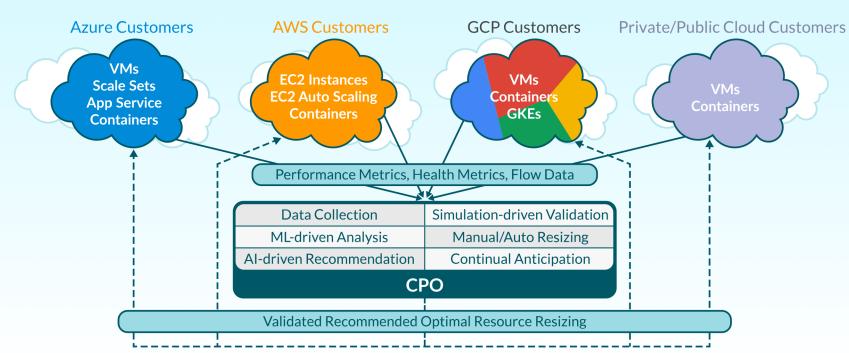


- ML-driven Cloud Application Resources for Utilization and Health Levels
- AI-based Recommendations for Multiple
  Optimization Goals that Differently
  Trade-Off Cost and Performance
- Simulation-based Validation
  Optimization Benefits of Selected
  Recommendation Before Resizing
- ML-driven Continual Anticipation of Suboptimality



# **CPO Deployment Architecture**

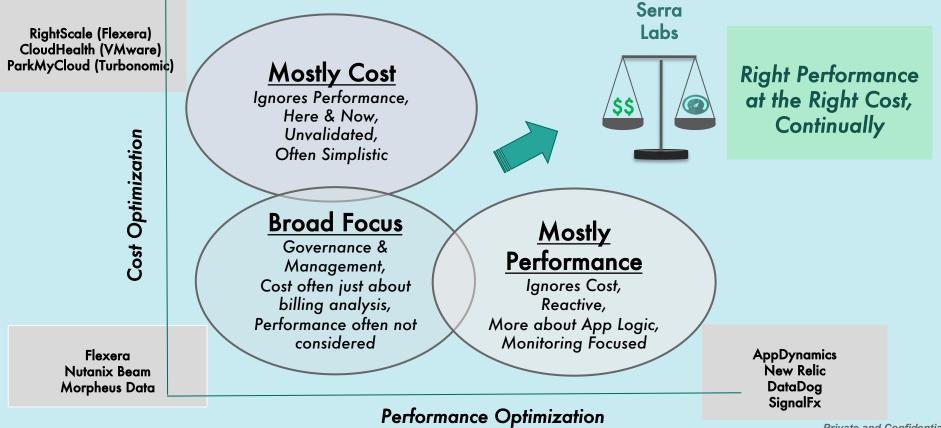




#### **CPO SaaS Architecture**

# **Comparison with Current Solutions**

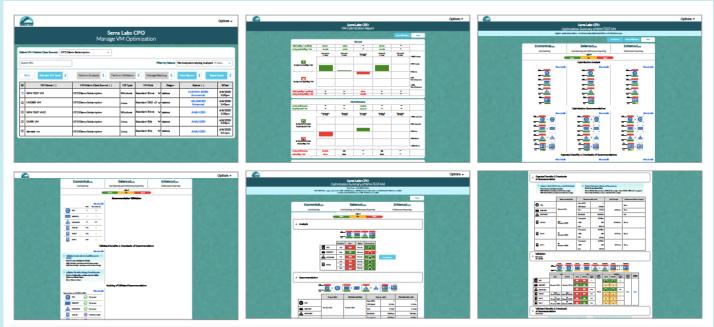




### **CPO** Solution

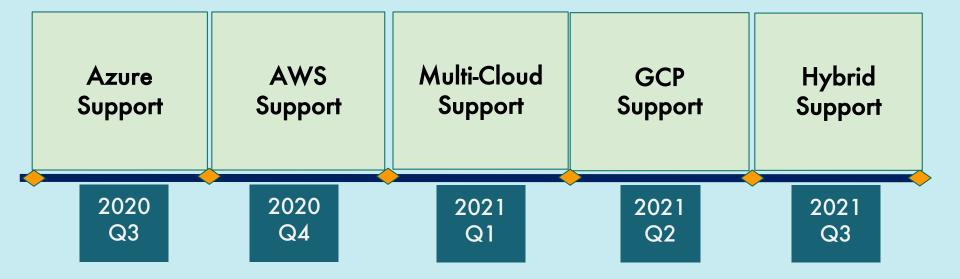


- Delivered as SaaS
- Pay-Per-Use & Subscription Plans
- Browser-based UI, No Installation Needed



## **CPO** Roadmap





# Summary



- Excess Costs and Poor Performance of Cloud Applications Incurred due to Inelastic Resource Sizing
- CPO is the Only Solution that Optimizes Resource Sizing for Cost and Performance Together
- Each Cloud Application can be Optimized for Its Own Cost-Performance Objective
- CPO available as SaaS for Microsoft Azure Cloud
- AWS Support in Progress, GCP and Multi/Hybrid Cloud Support Planned