



Towards Building Autonomous Data Services on Azure





Introduction

- **Challenge:**
 - Growing data services and configurations at scale.
 - Cloud users
 - Cloud providers
- **Solution Approach:**
 - Utilize advances in data science and machine learning, combined with cloud technology, to develop autonomous data services.



Viewpoints

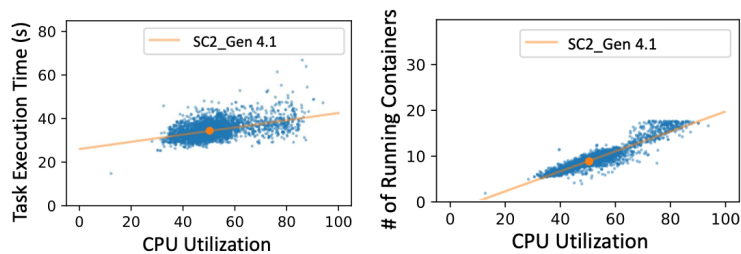
Perfect timing for Autonomous Data Services

- Economic scale of the cloud necessitates autonomous data services.
- Autonomy spans cloud infrastructure, query engine, and service layers.
- Objectives: Improve ease of use, optimize performance, reduce costs, maintain data privacy.

Challenges in automating in different layers

Cloud Infrastructure Layer

- Complex Resource Management
- Balance Quality of Service, such as low latency, with the operational costs of maintaining that service.



(a) CPU vs Task Exec Time (b) CPU vs Runinng Containers

Figure 1: Models to predict machine behavior [53]

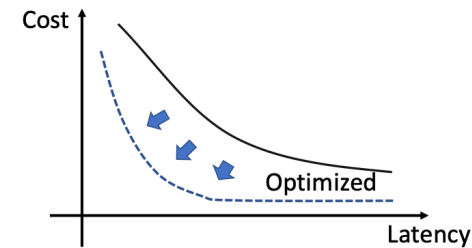


Figure 2: Pareto curve depicting the trade-offs between the QoS (x-axis) and the cost (y-axis)



Challenges in automating in different layers

Query Engine Layer

- Real production systems are often more intricate than the academic prototypes
- Learn from the past to improve the future
- Principals:
 - Minimal changes to the existing engine
 - Us ML only when it make difference



Challenges in automating in different layers

- **Service Layer**

- The need to make customer-facing decisions and choices regarding the system at the service level

Global Model
+ Enough Data
- Not accurate
- Hard to build

Segment Model
+ Transfer learning
among similar
customers

Individual Model
+ Accurate
- No enough Data
- Too many model





Lessons Learned

- Simplicity rules. Simplicity helps with:
 - Cost
 - Scalability
 - Manageability
 - Explainability
- One size does not fit all
- Feedback loop is indispensable



Future Direction

- Reuse!
- Standardization
- Optimization across components jointly.
- Responsible AI (RAI)



Question?

