



ASTER

Do more.

The Impact of Cloud Computing on Data Warehousing

Tasso Argyros , Co-Founder and CTO

April 15th, 2009

Topics

Aster Data Systems

Introduction to Data Warehousing

Impact of Cloud on Data Warehousing

Aster nCluster Cloud Edition

Introduction

Aster Data Systems

Who is Aster Data Systems?

Relational database for data warehousing

- ➔ software that runs on big clusters of cheap servers

Founded in 2005

- ➔ Mayank Bawa, CEO [Stanford InfoLab]
- ➔ Tasso Argyros, CTO [Stanford DSG]
- ➔ George Candea, Chief Scientist [Stanford ROC]

Roots



Investors



Recognition

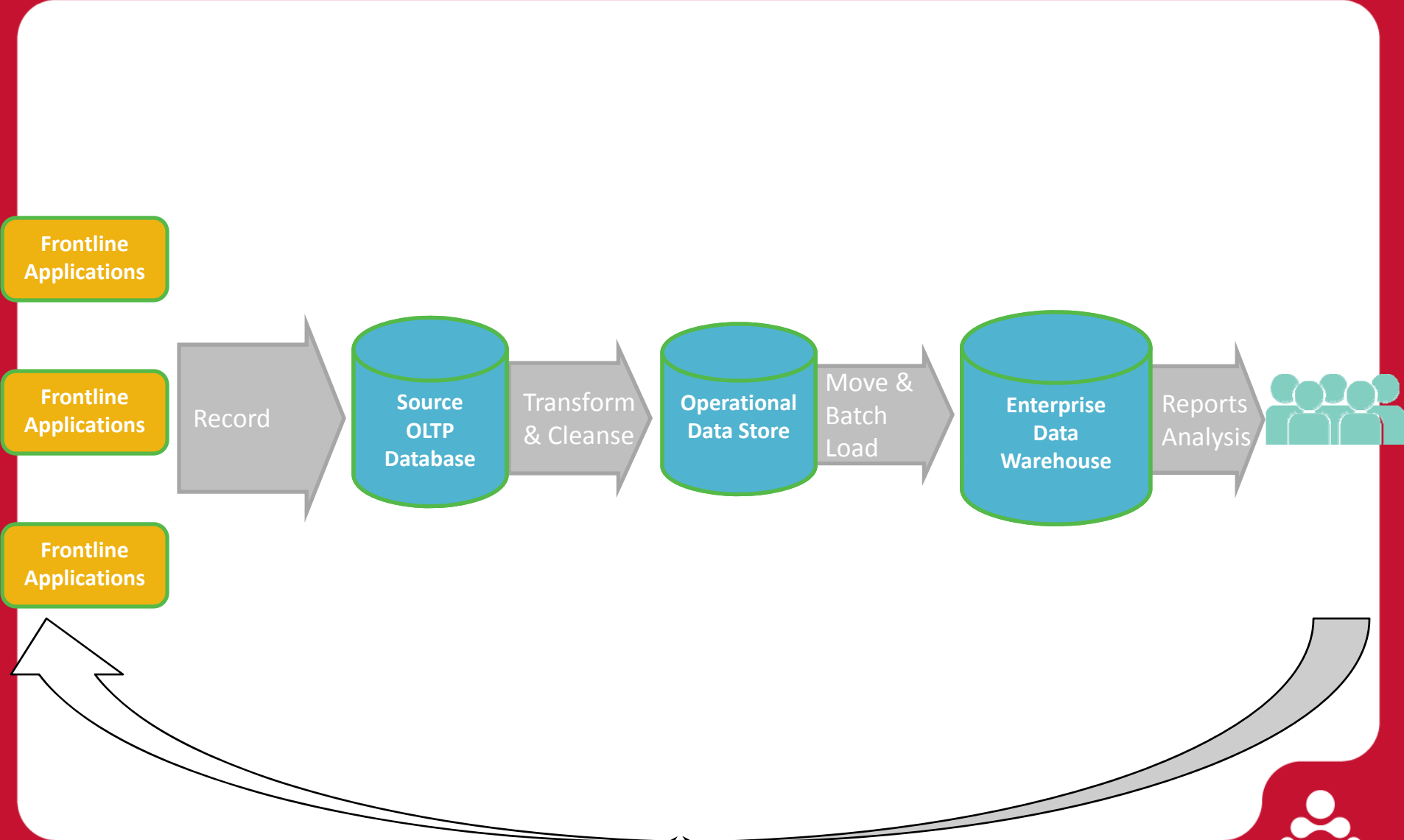


Gartner

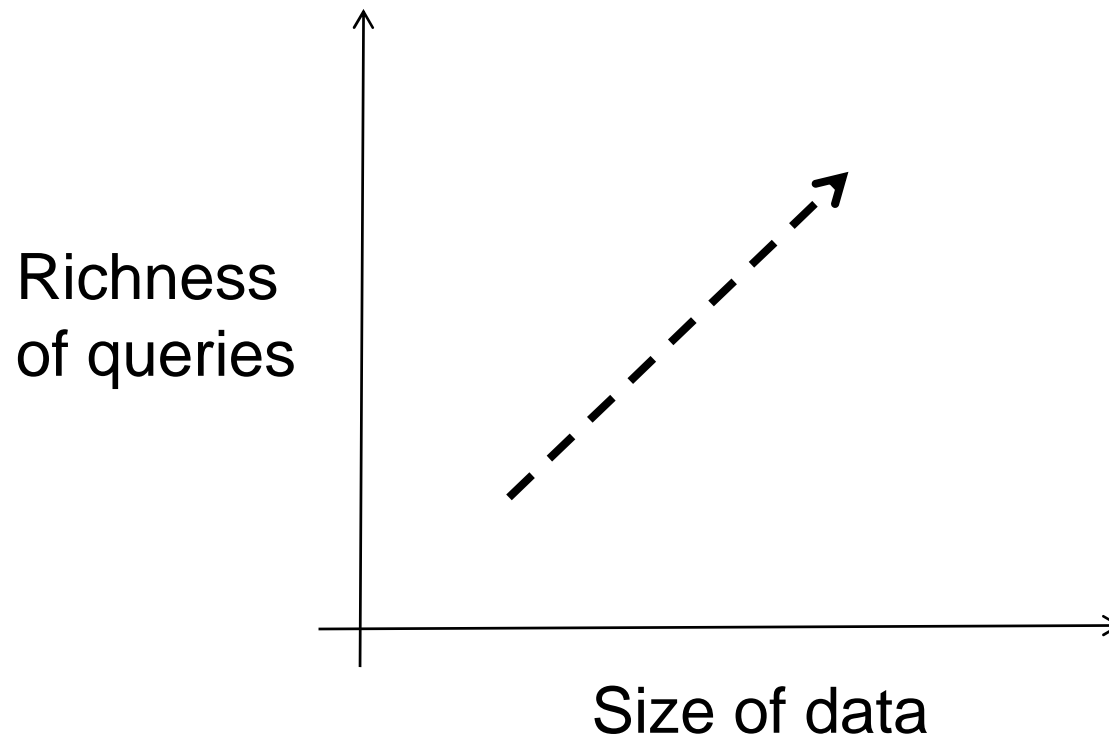
Introduction

Data Warehousing

Enterprise Data Warehousing



Trends in Data Warehousing



1. Mix of queries changes as more users are added
2. Usage has peaks and troughs within a day

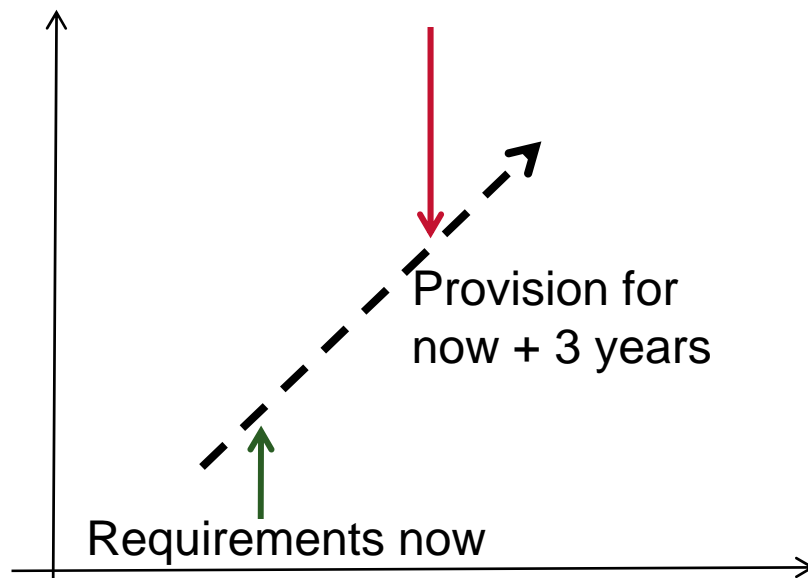
Implications on Infrastructure

➔ Compute and storage requirements are high & increasing

- Big SMP and SAN deployments

➔ Infrastructure footprint is large

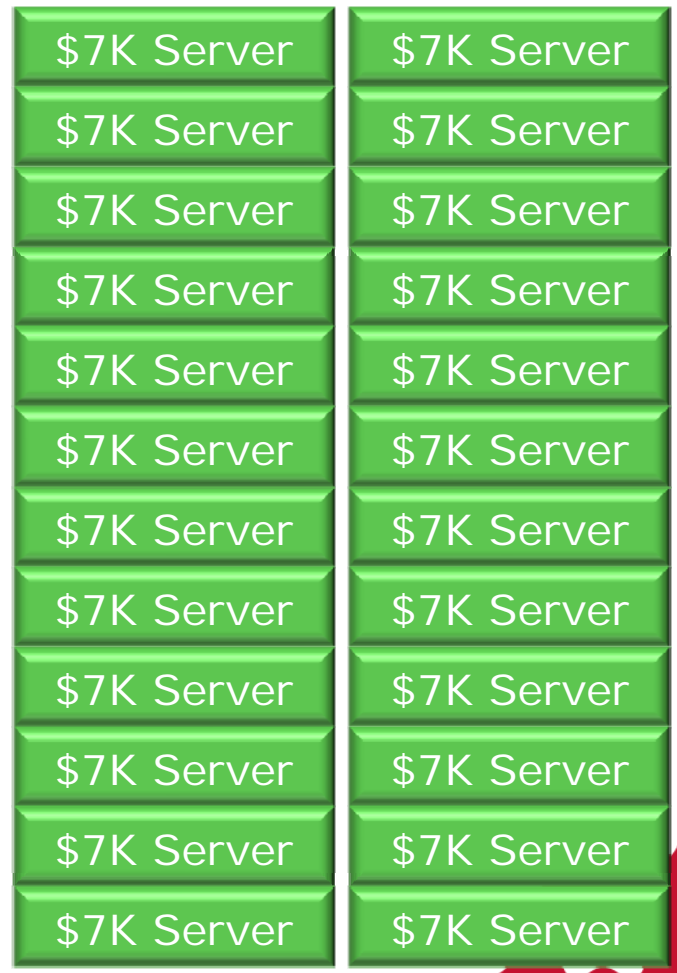
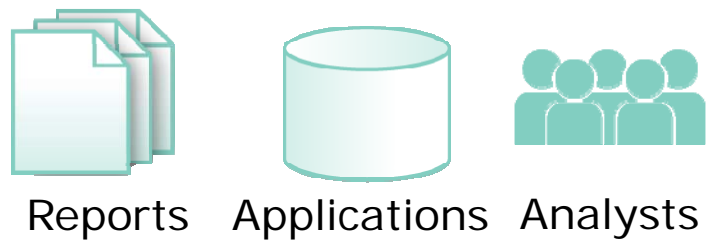
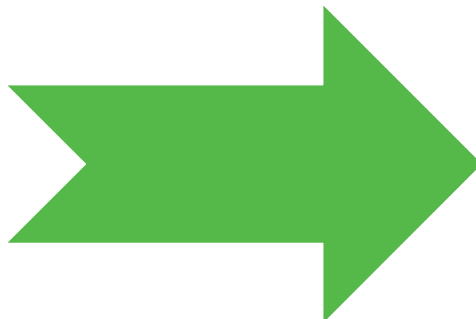
- Upgrades are expensive in time and effort



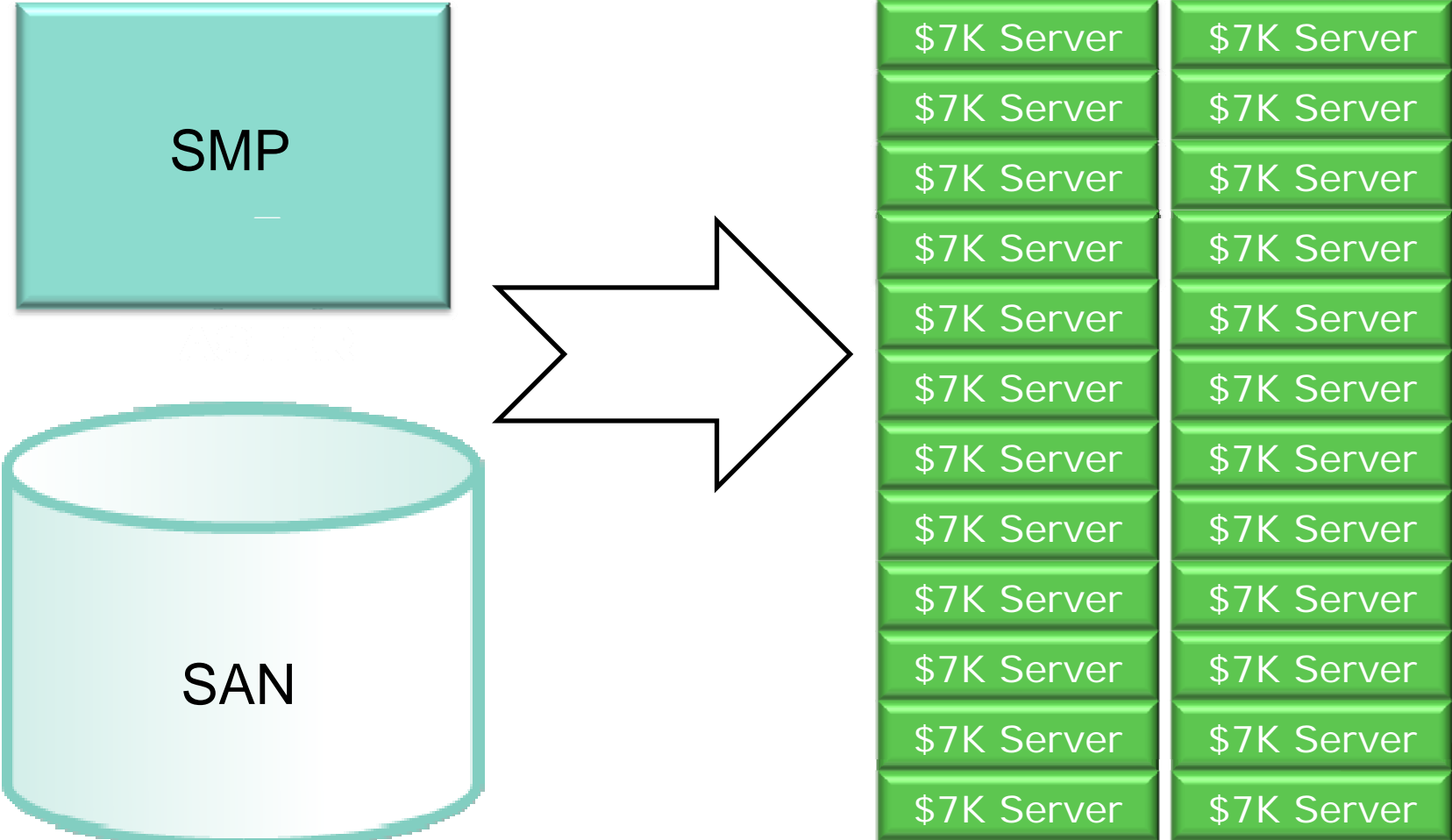
➔ 80% of initial cost is infrastructure cost

MySpace (2007-09): Actual Deployment

3500TB



Data Warehousing is now "Cloud-Friendly"



Cloud Computing Impact on Data Warehousing

Public and Private Clouds

BENEFITS

- Pay only for what you use
- Fast scale-up (or down)
- Reduce admin overhead

ENTERPRISE CONCERNS

- Privacy/security of data in a shared infrastructure
- Data transfer speeds over public Internet

Cloud Variants

PUBLIC CLOUD

- Example: Amazon EC2
- Typical users
 - Startups/developers
 - Enterprise experimenters

PRIVATE CLOUD

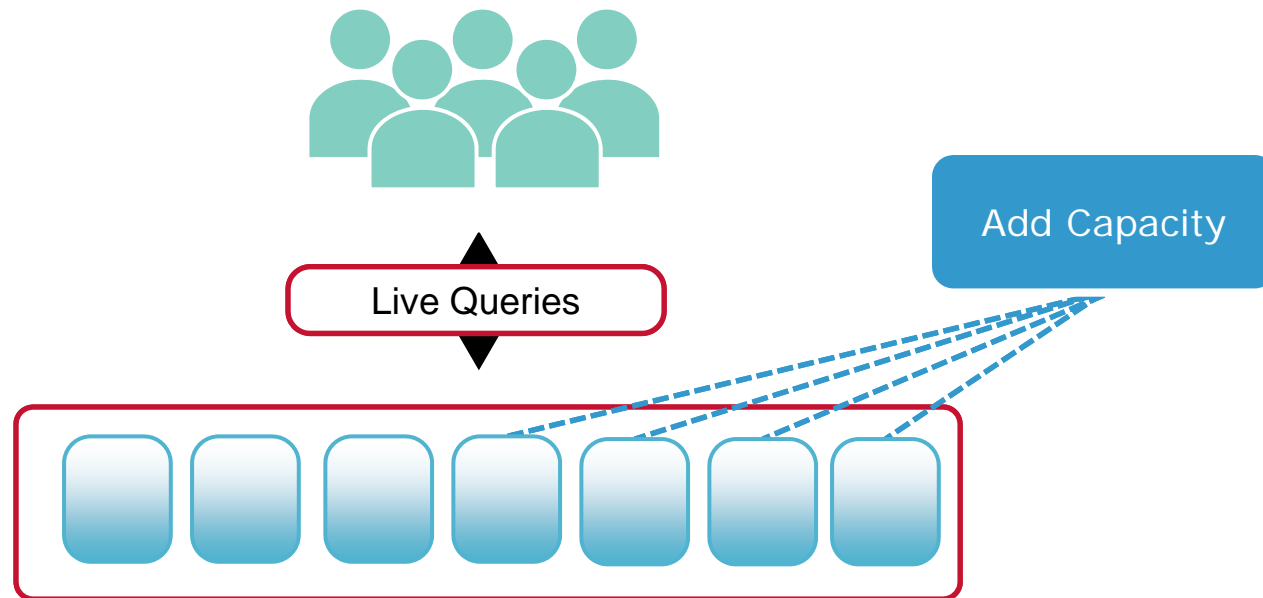
- Owned by large enterprise IT groups
- Centralized infrastructure for use across the company
- Address enterprise concerns of security and data transfer speeds

Aster *n*Cluster Cloud Edition

- ➔ Proven: ShareThis is largest cloud-based DW in world on AWS
 - (2.2TB, growing to 10-18TB by year-end)
- ➔ Easiest on-demand scaling in the market
- ➔ First host-vendor-neutral offering

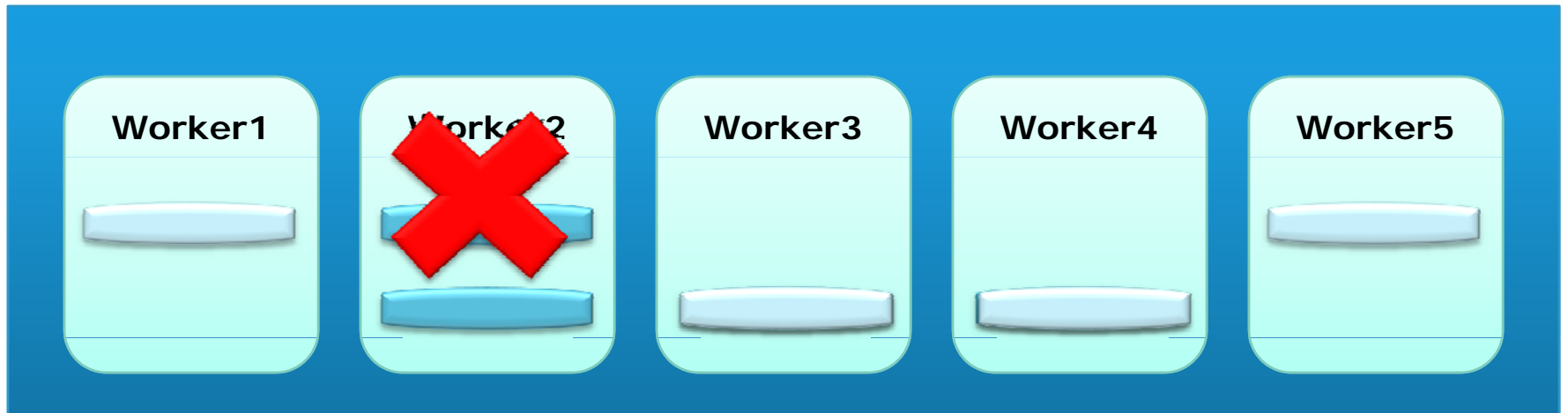


1. Elastic Scalability



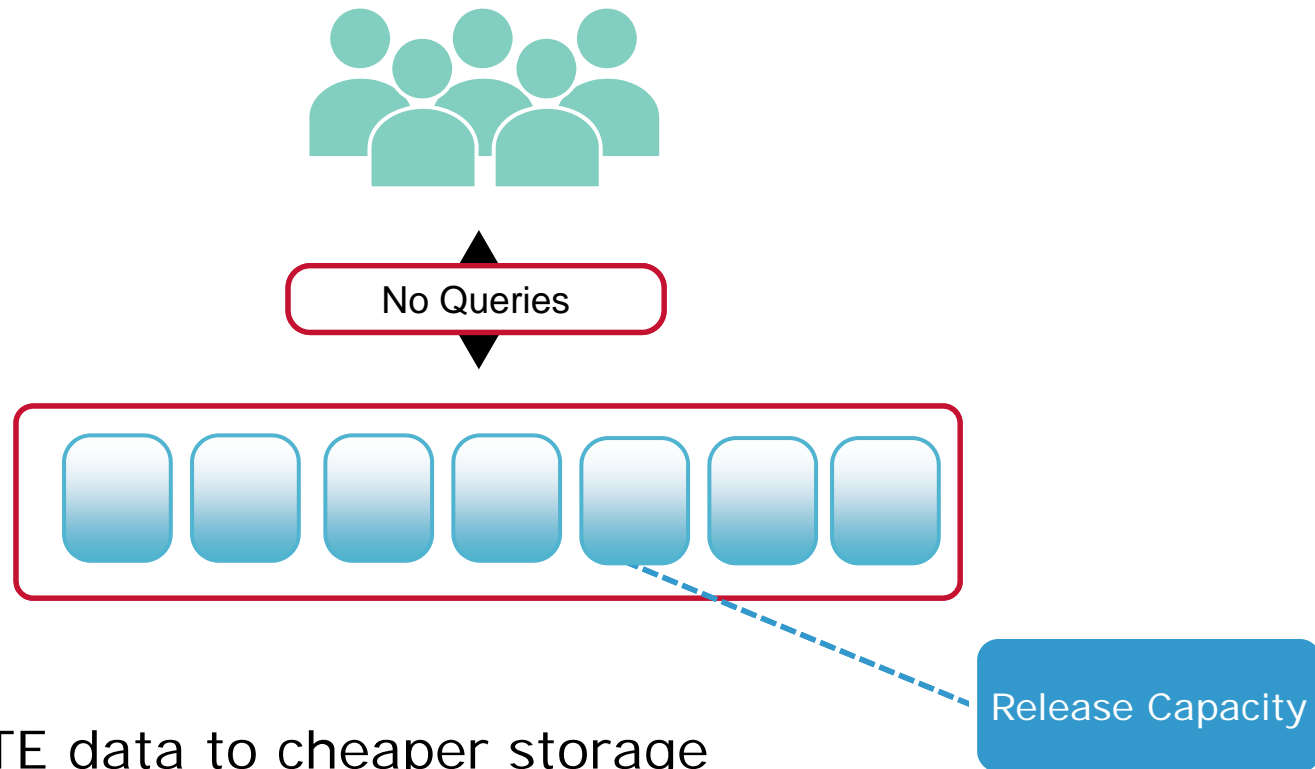
- Single-click scale-out and scale-down with no downtime
- Automated incorporation and load balancing in minutes
- Database available even while loading, backup, export, restore, scale-up, re-provision, fault recovery

2. "Always On" Availability



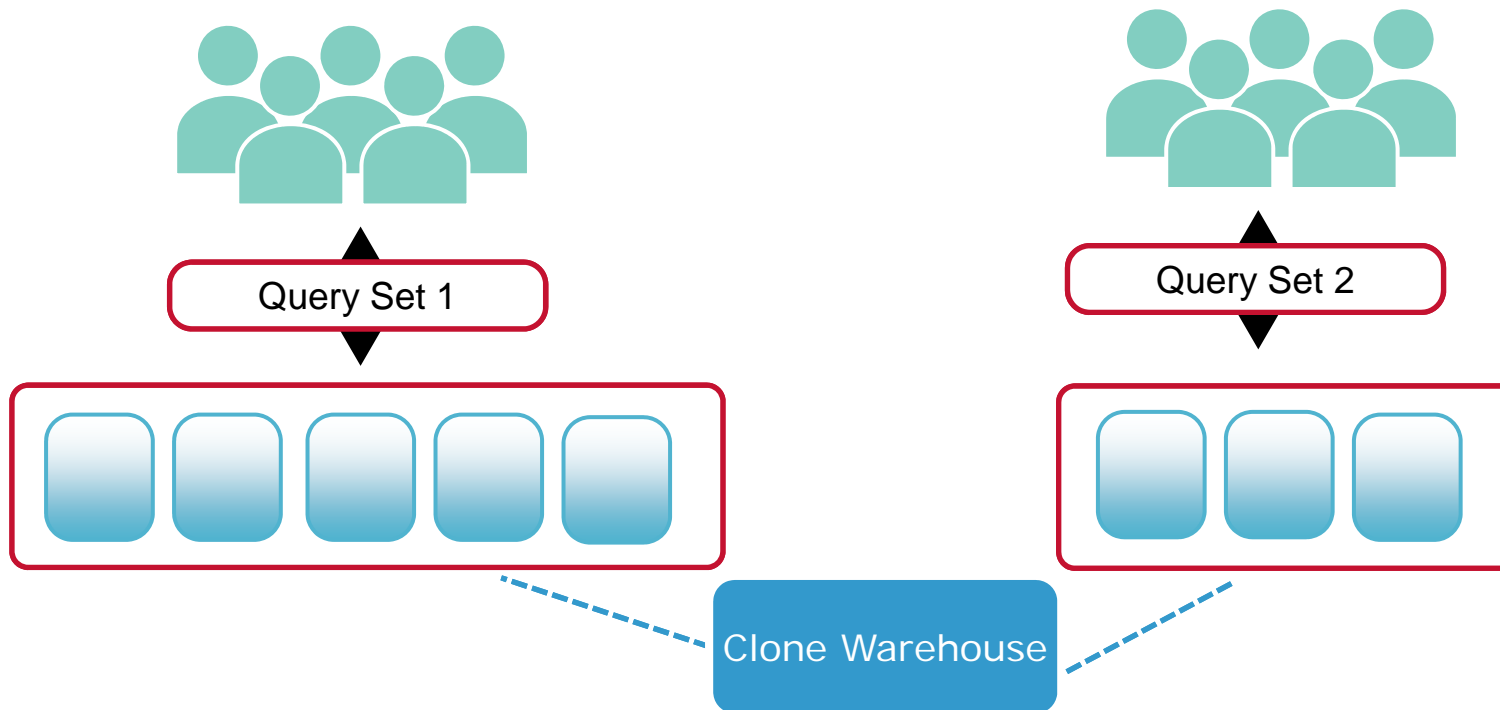
- Cloud units WILL FAIL
- Online backup and restore
- Online load and export

3. Hibernating Services



- HIBERNATE data to cheaper storage
- Release cloud units when no usage
- Revigorate on-demand

4. Managing Workloads



- CLONE service (data + compute) to a new pool
- Re-assemble pool when usage declines

SUMMARY: Data Warehousing in the Cloud

1. Port Product

- ➔ Ensure compatibility
- ➔ Ensure performance
- ➔ Ensure features

2. Innovate Product

- ➔ Leverage “infinity”
- ➔ Leverage service APIs
- ➔ Enable new features