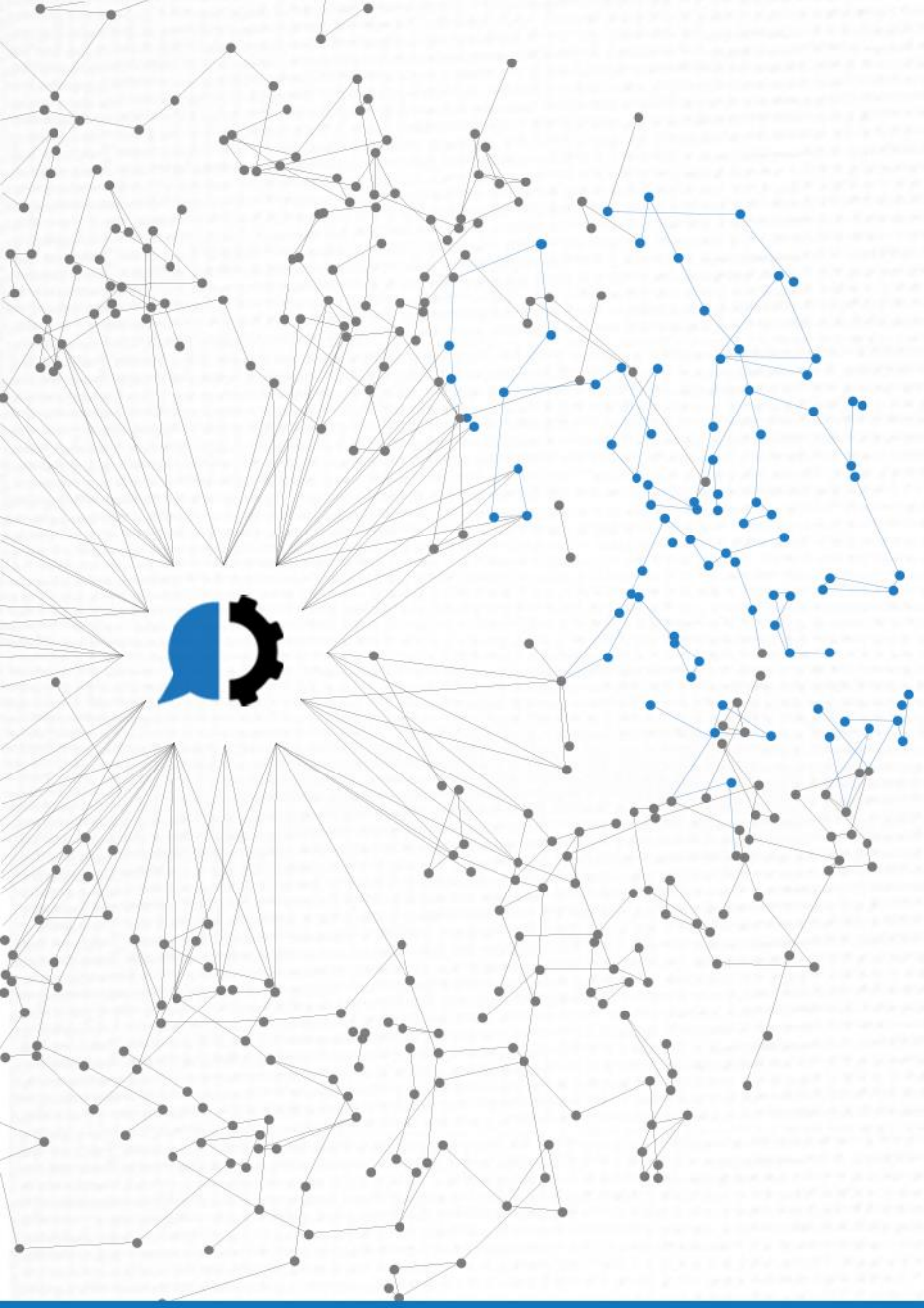


helpsystems

How To Optimize Capacity in the Public Cloud

Per Bauer

www.helpsystems.com

Your Presenter

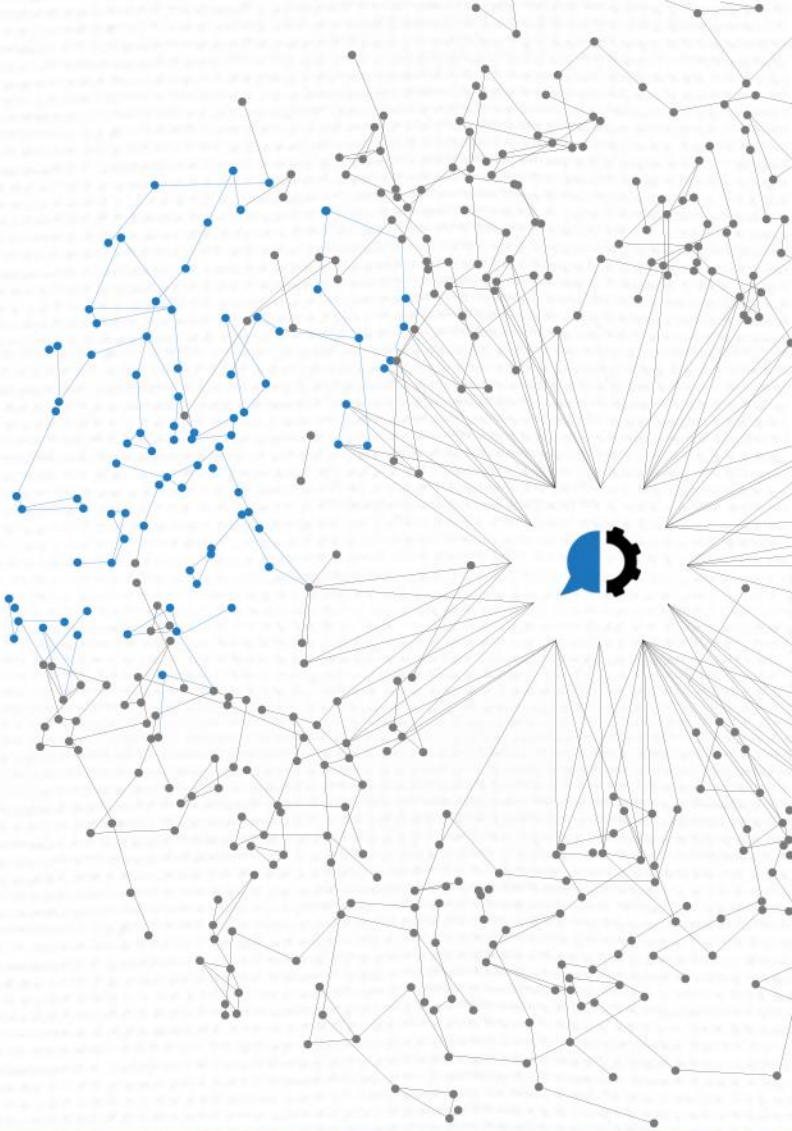
Per Bauer

Director of International Services

Agenda

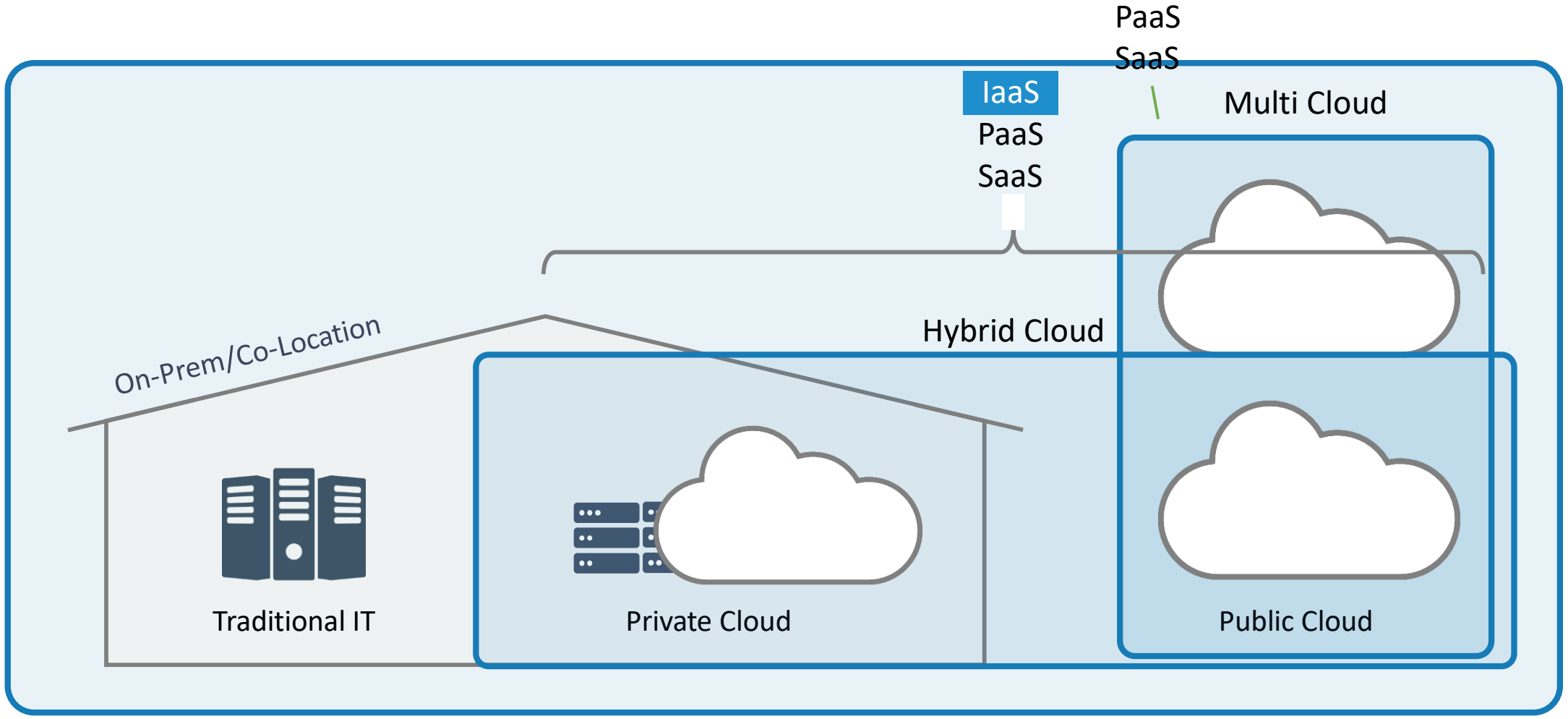
1. Cloud Facts
2. Impacts on Capacity Management
3. Common misconceptions
4. Optimizing the migration
5. Forecasting requirements and cost
6. Other considerations
7. Summary and Q&A

Cloud Facts



UP NEXT

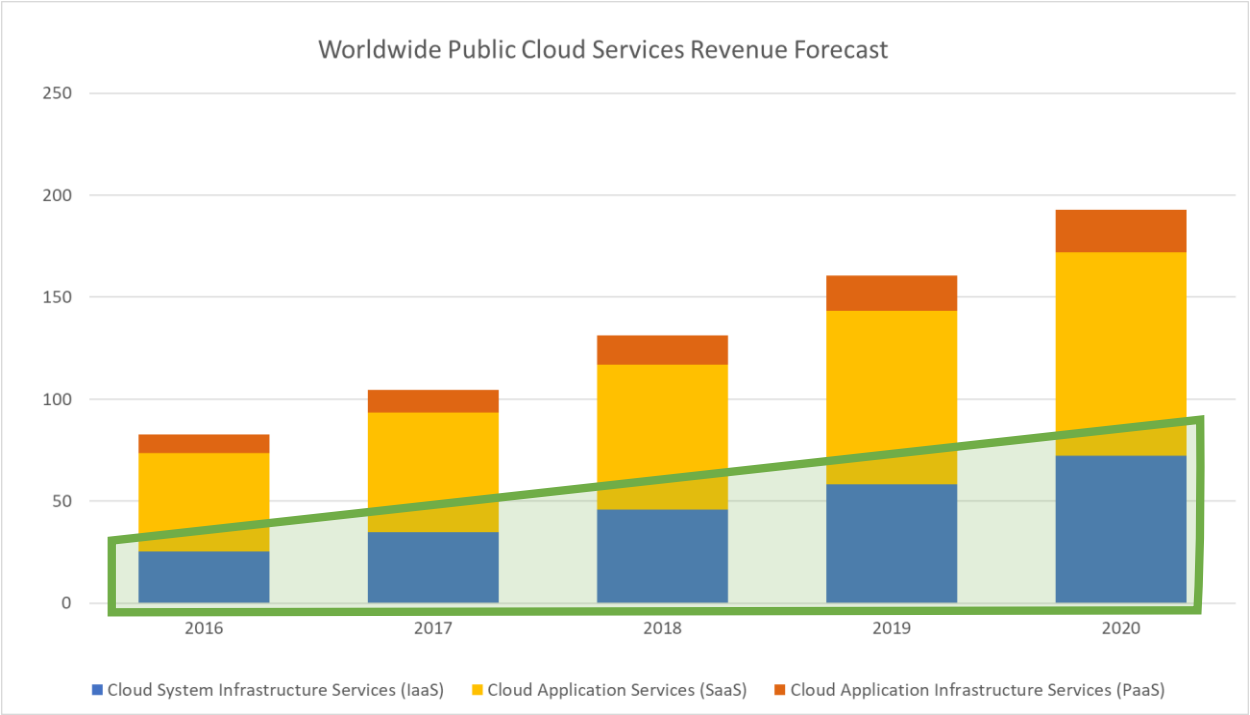
Terminology



Why Are Organizations Adopting IaaS?

1. Cost Reduction
2. Agility
3. Strategic Focus
4. Assurance
5. Scalability

Exodus to the Public Cloud



Source: Gartner (October 2017)

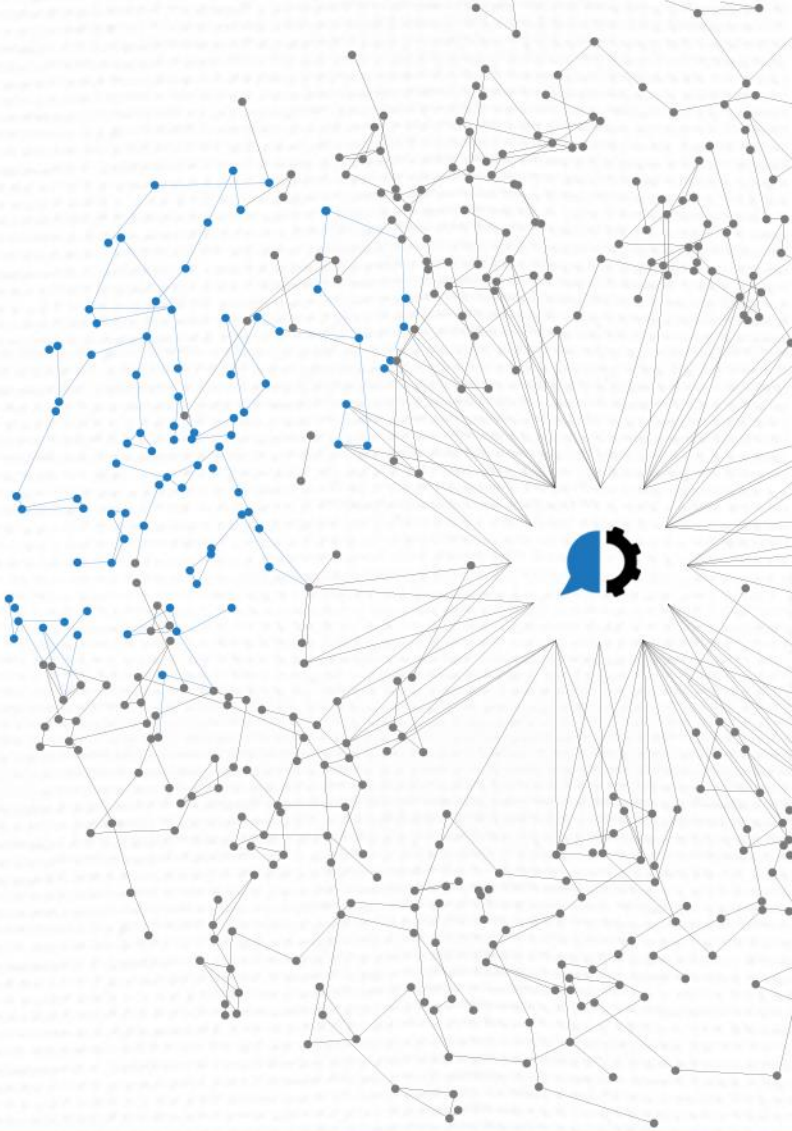
No-Cloud → Cloud-First → Cloud-Only

What Runs in the Cloud?

Cloud Native applications are software specifically designed to run in the cloud:

Micro-services	Built on small components talking through APIs
Scale-out	Architected to provide elasticity (add/retract nodes)
Stateless	Shared state is maintained and mirrored across the nodes
Automation	Designed for automation using management frameworks like Kubernetes, Swarm, Mesos, etc.

Impact on Capacity Management



UP NEXT

The Changing Role of Capacity Management?

Traditional On-premise IT	⟷	Hosted in the Public Cloud
Optimize the use of your limited data center resources	→	Optimize the use of your limited budget
Driving Factors for Capacity Management: • Business critical services • Seasonality and peaks in business activity • Business growth • Regulatory requirements • Provisioning lead times • Cost optimization • Agility	→	• Business critical services • Seasonality and peaks in business activity • Business growth • Regulatory requirements • ~~Provisioning lead times~~ • Cost optimization • Agility
Efficiency gains realized over time	→	Efficiency improvements has immediate impact

helpsystems

Capacity Management for the Cloud

Optimization

- CapEx to pure OpEx
- No “sunk cost”, optimization pays off immediately
- Deliver the “right” capacity at the lowest possible cost

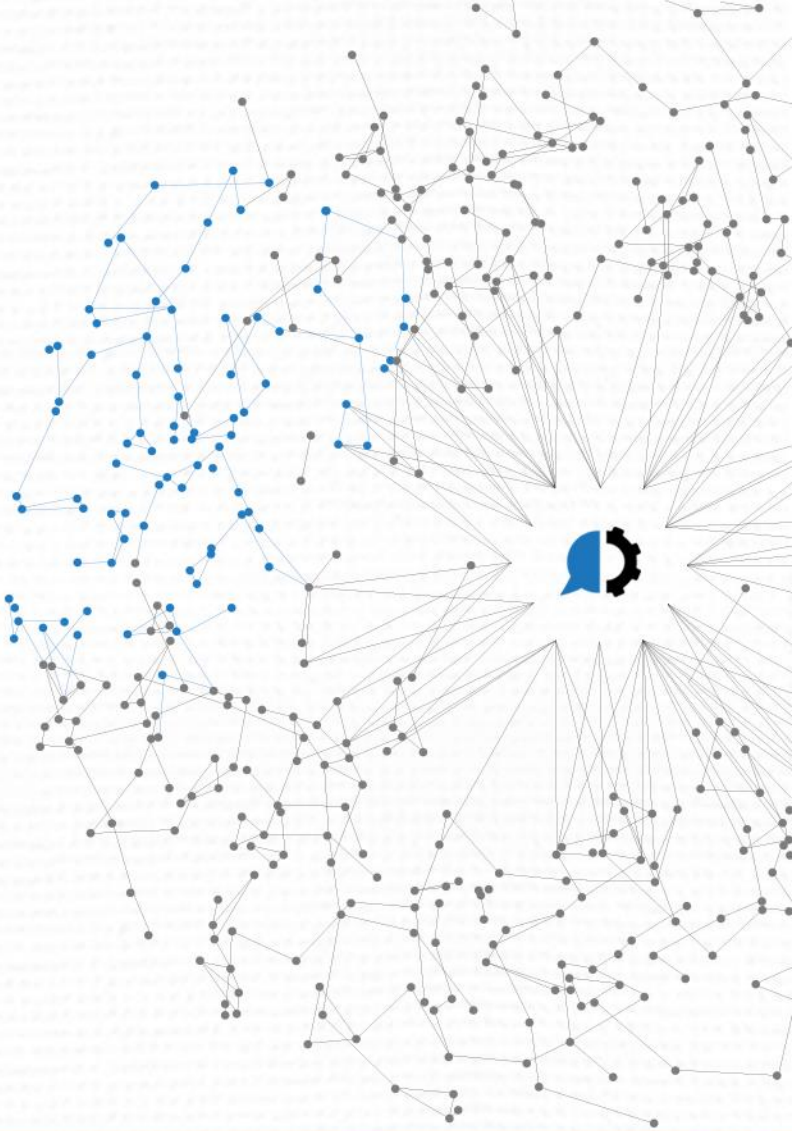
Predictability

- Capacity is infinite, but not budgets
- Qualified prediction of cost - organic growth and forecasted demand

Proactive to Reactive

- Proactivity used to be the axiom
- Harder with less control over provisioning
- Improve reactive clean-up and right-sizing procedures

Misconceptions about Public Cloud



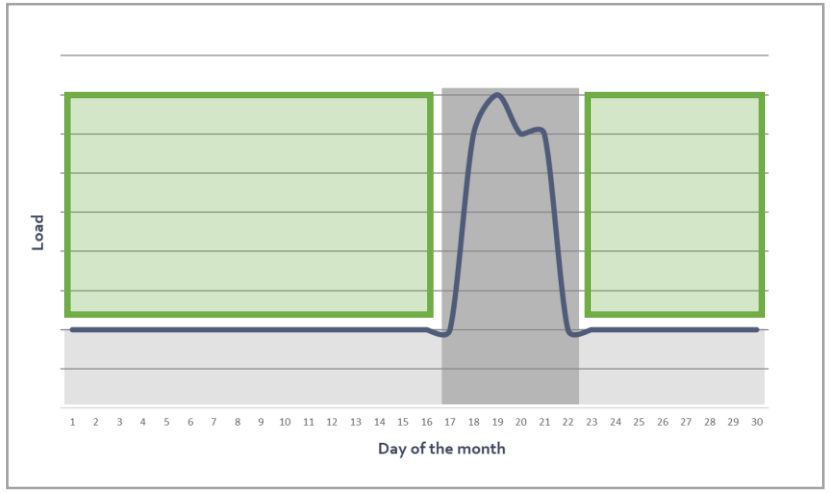
UP NEXT

“Capacity is Cheap”

It can be, but it depends on a number of parameters:

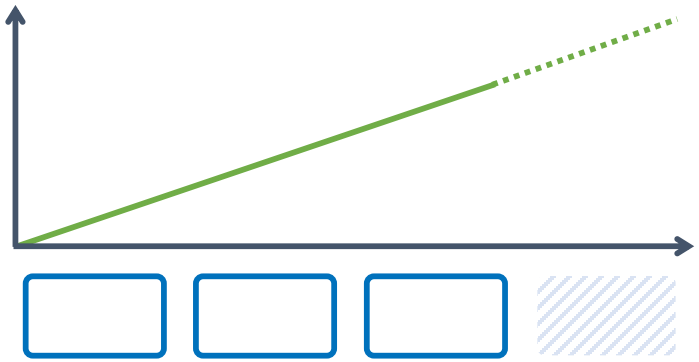
- Type of instance (on-demand, reserved, dedicated, etc.)
- Workload profile
- Sustained activity level
- Data volumes exchanged

75% savings



“Capacity Can Be Added Instantaneously”

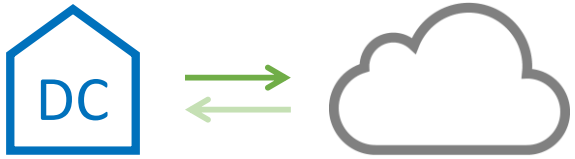
- Is the application requiring additional capacity designed to scale out?
- Data center location affinity?
- At what cost? Can it be justified? Has anyone budgeted for it?



“Burst During Peak Demand”

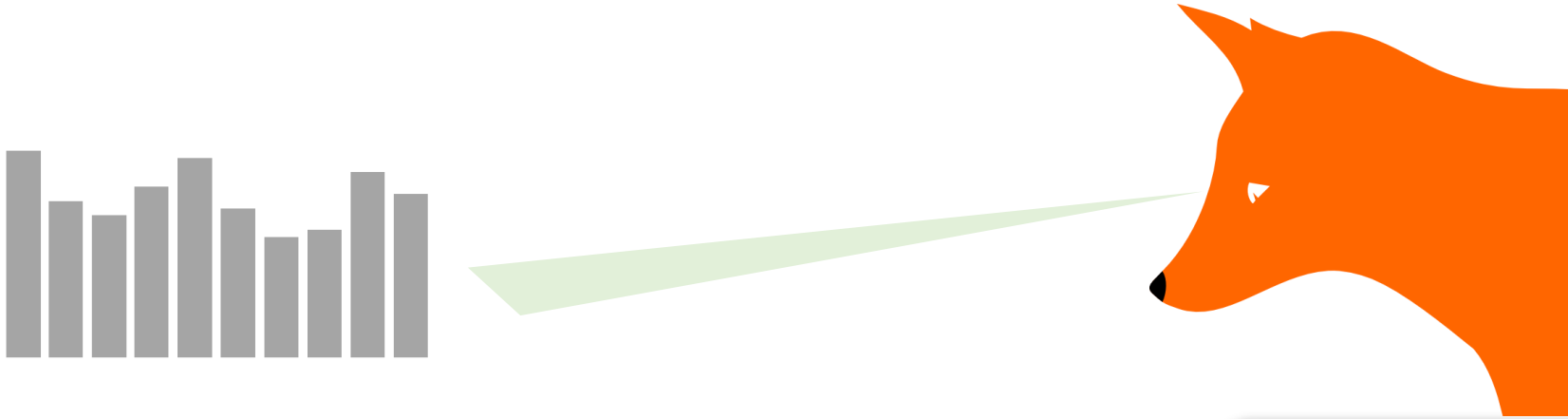
Moving workloads to accommodate changing demands

- Lack of compatibility
- Potential benefit lost having to customize, test, validate, and nurse the solution
- Compelling concept, but few successful real-life examples



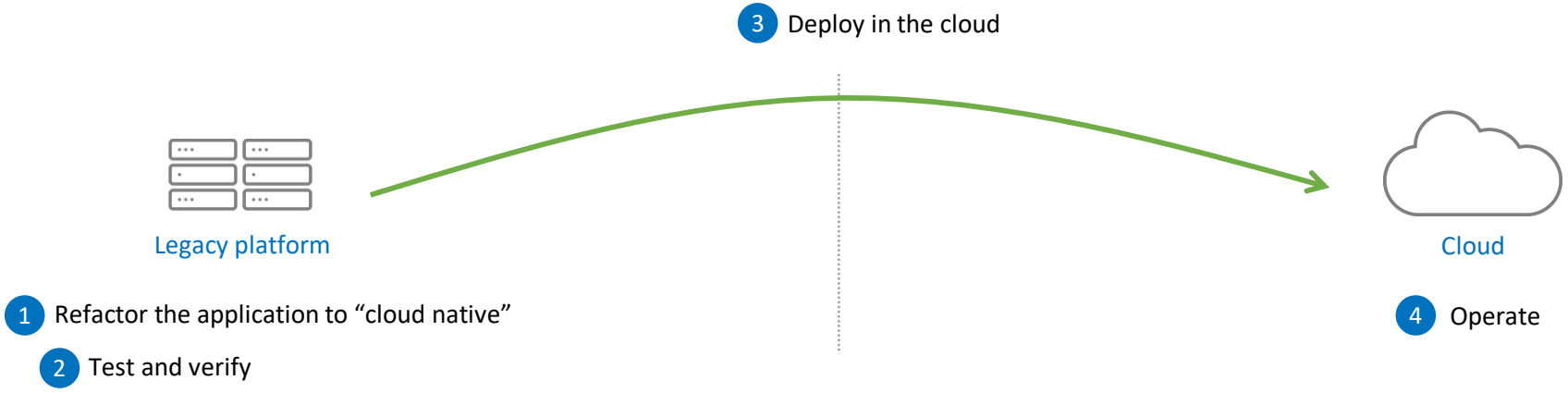
“Planning Can Be Delegated to the Provider”

- Demand forecasts and sanity checks requires understanding of the business
- Multiple providers?
- Lack of incentives

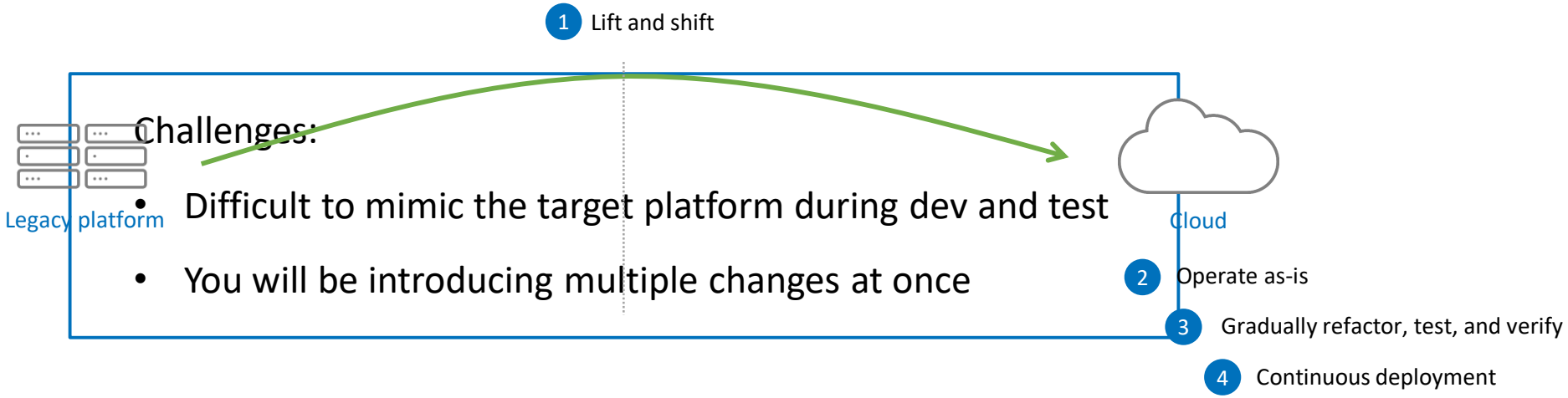


“Refactoring Happens Prior to Migration”

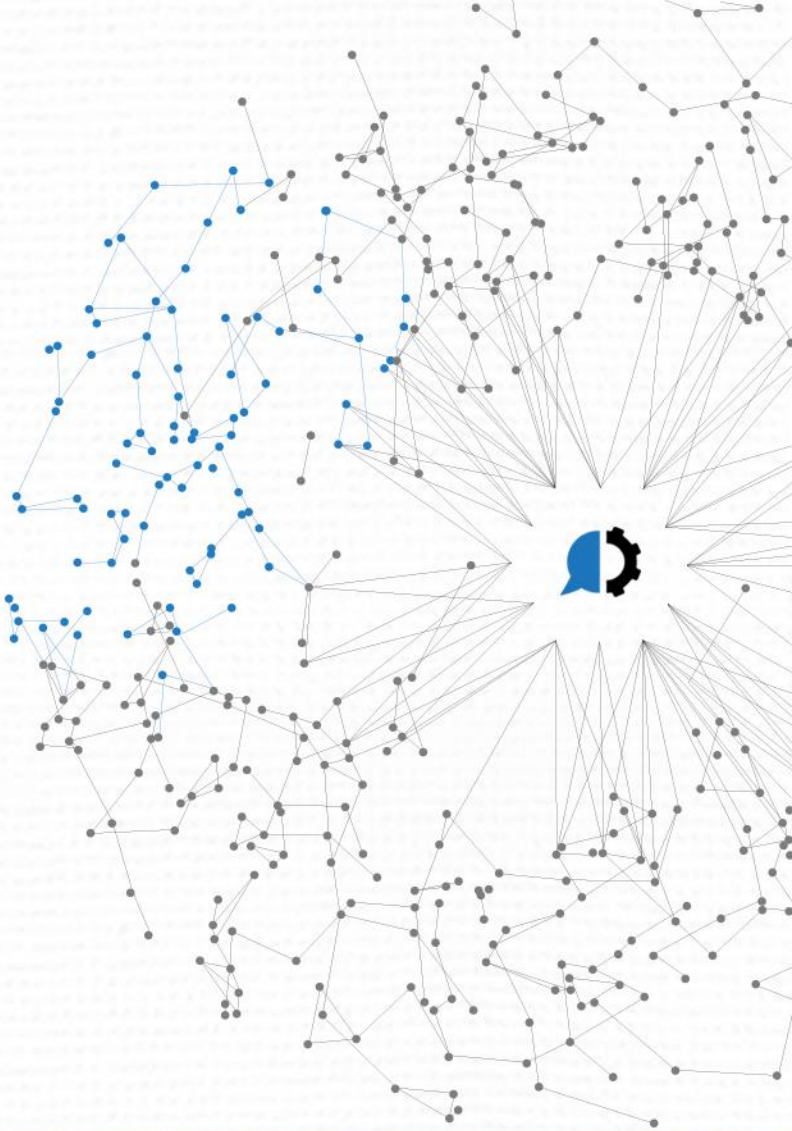
Textbook



Lift-and-shift



Optimizing the Migration



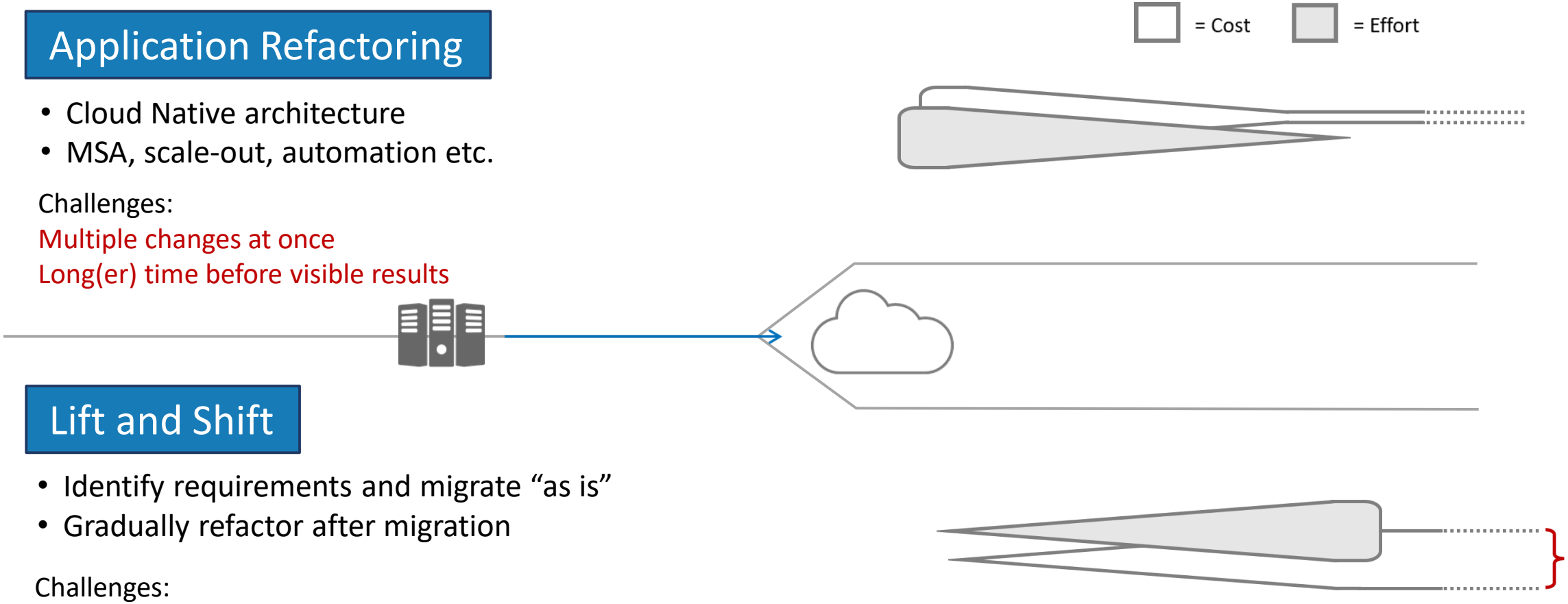
UP NEXT

Cloud Migration challenges

Application Refactoring

- Cloud Native architecture
- MSA, scale-out, automation etc.

Challenges:
Multiple changes at once
Long(er) time before visible results

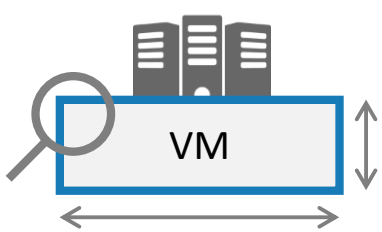


Lift and Shift

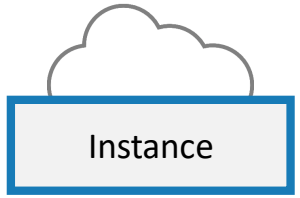
- Identify requirements and migrate “as is”
- Gradually refactor after migration

Challenges:
Newly migrated workload will cost more than “expected”

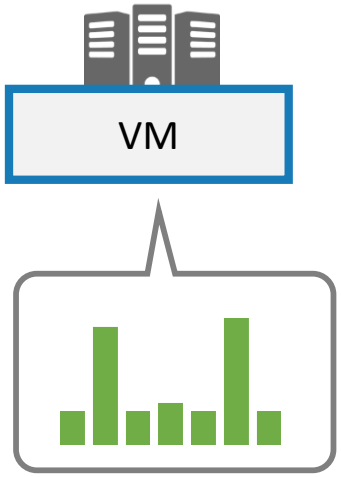
Optimizing the target instance



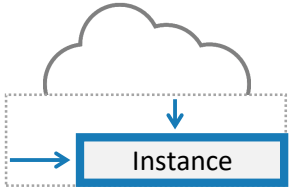
Lift and Shift



Instance types
- t2.micro
- t2.small
- t2.medium
- t2.large
- ...

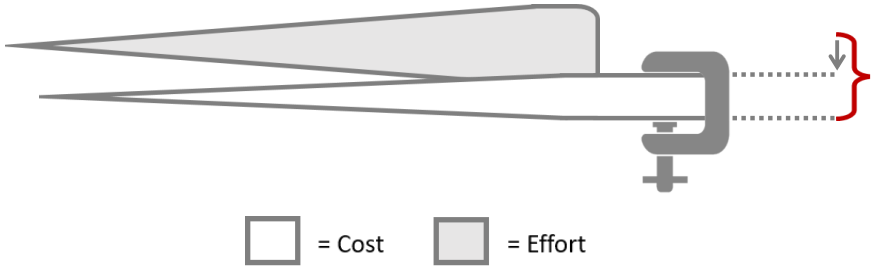


Lift, Trim and Shift



- Utilization of resources
- Business Activity Cycles
- On-Demand vs Reserved
- Growth Headroom

35%-85% savings*

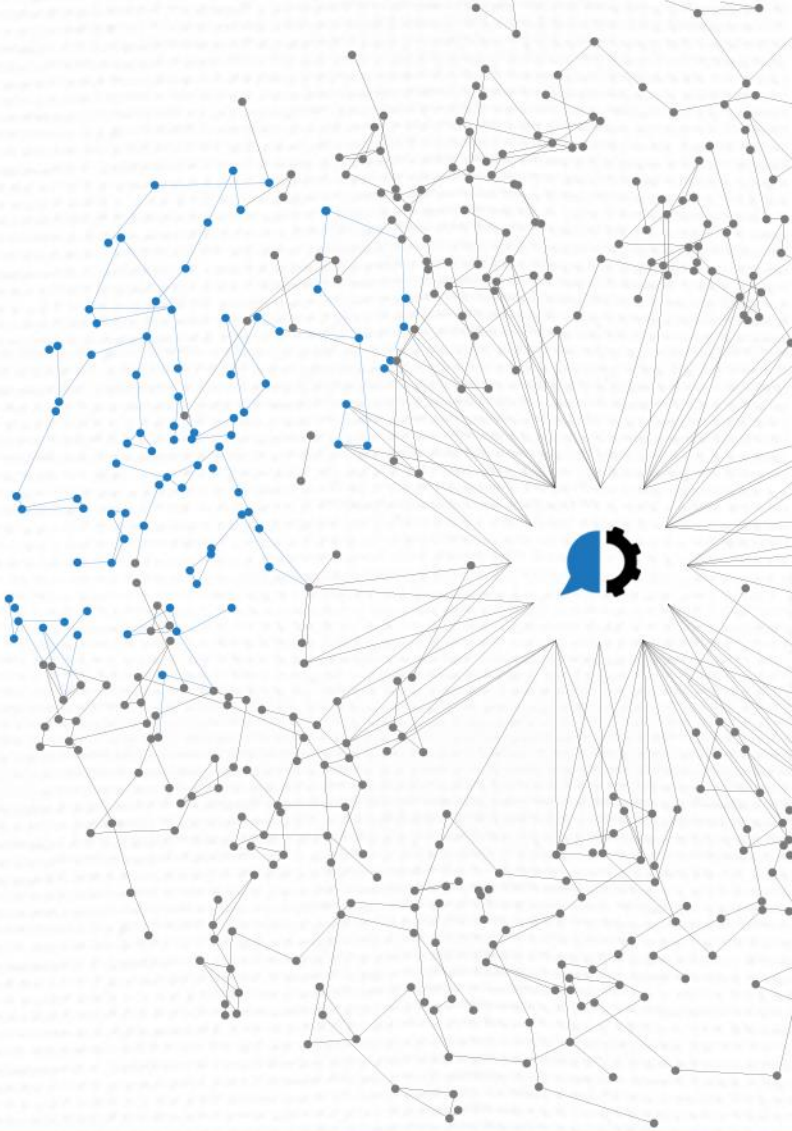


Where should I start?

	1	2	3	4	5
Architectural Fit	Monolith ⟷				Service Oriented
Current Cost of Ownership	Low ⟷				High
Business Process	Static ⟷				Dynamic
Demand	Stable ⟷				Variable

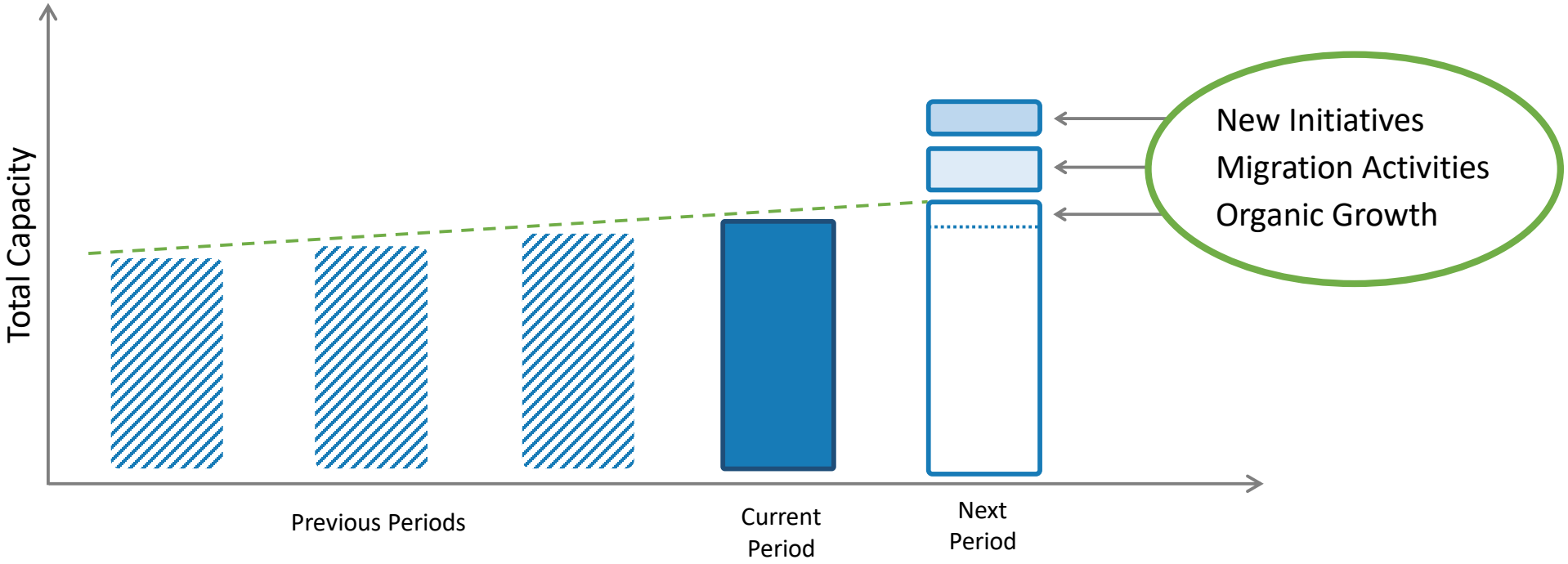
helpsystems

Forecasting Capacity Requirements and Cost



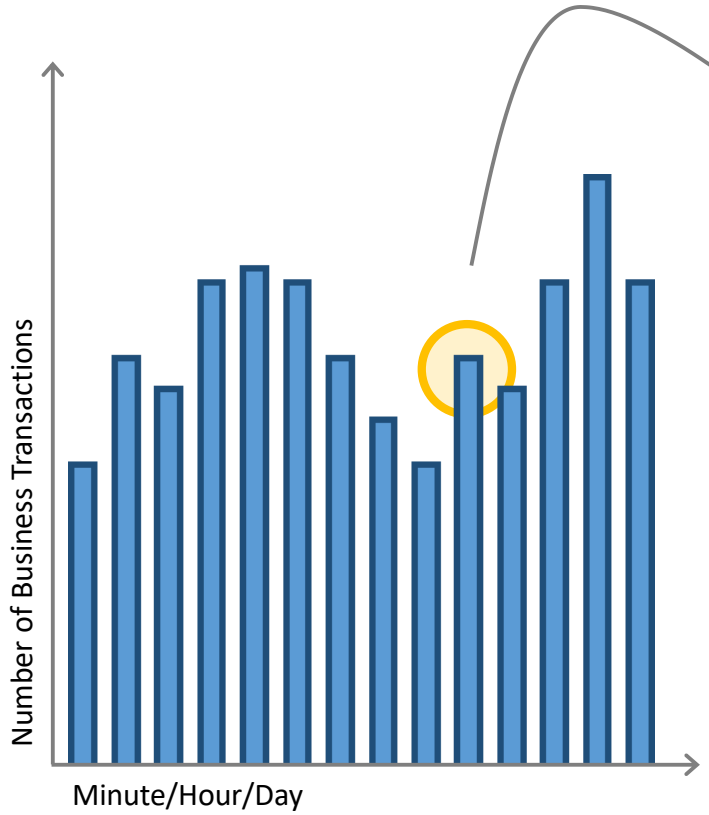
UP NEXT

Forecasting future Cloud spend



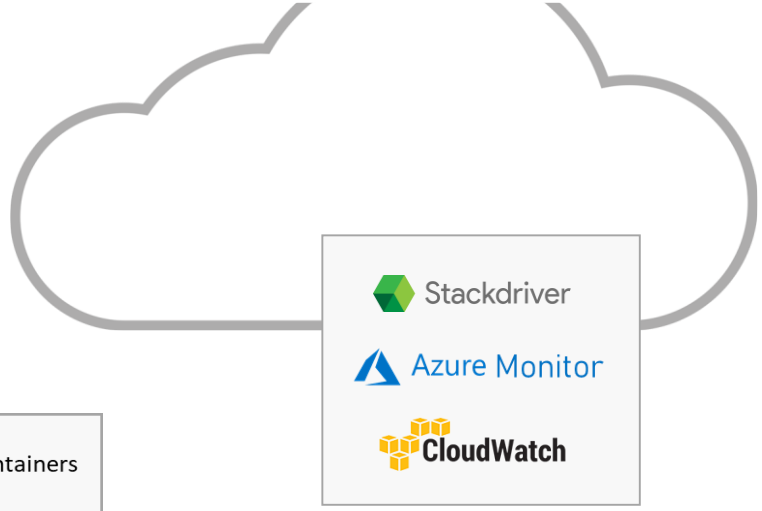
Understanding the dynamics of organic growth

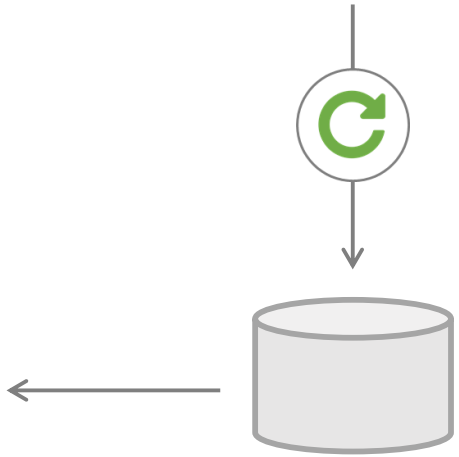
Correlate Business Activity with Provisioning



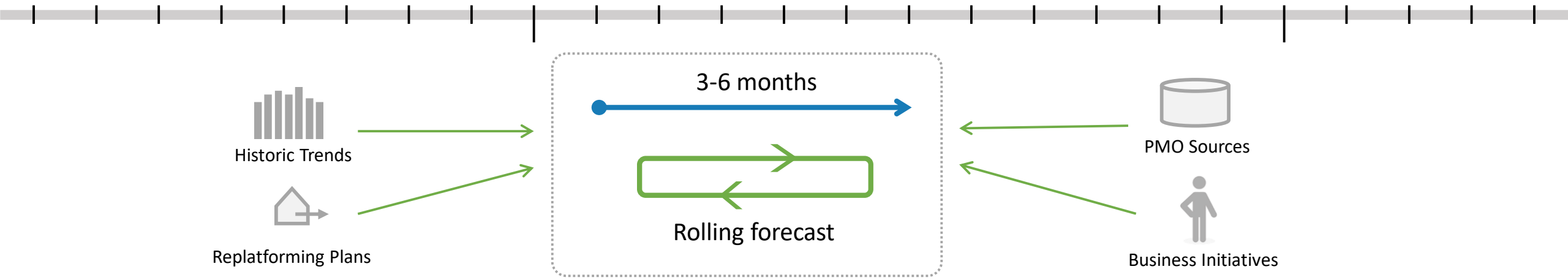
Snapshot

Service / Component	Instances & Resource Utilization	# Containers
respon		28
inact		18
libm		7
busf		13
claws		24
gic		8
…		…



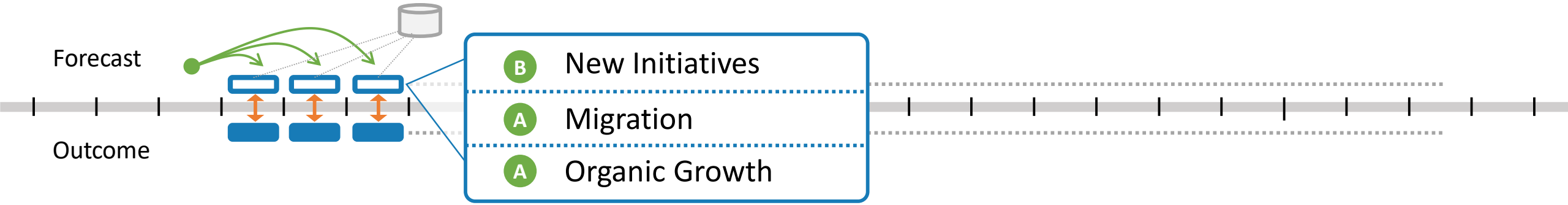


Build a Demand Calendar



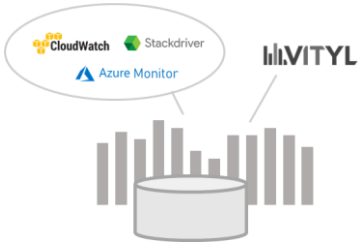
- Focus on the biggest drivers of variance
- Automate repetitive aspects
- Analyze input - Socialize sources
- Regular reviews of KPIs with stakeholders

Continuously Improving the Forecast



- Record the forecasts for correlation with actual outcome
- Analyze growth factors individually
  - A. Faulty statistical models, Inadequate data coverage
  - B. Incorrect assumptions, Reality vs. Aspirations
- Adjust future forecasts with probability scores based on time and correlation analysis

Predictable Costs

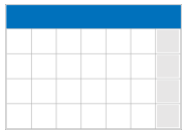


Usage Profiles

Trends and seasonality patterns

Rates and Charges

Current rates and accumulated charges

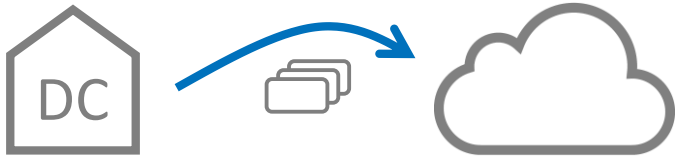


Demand Calendar

Forecasts about new initiatives

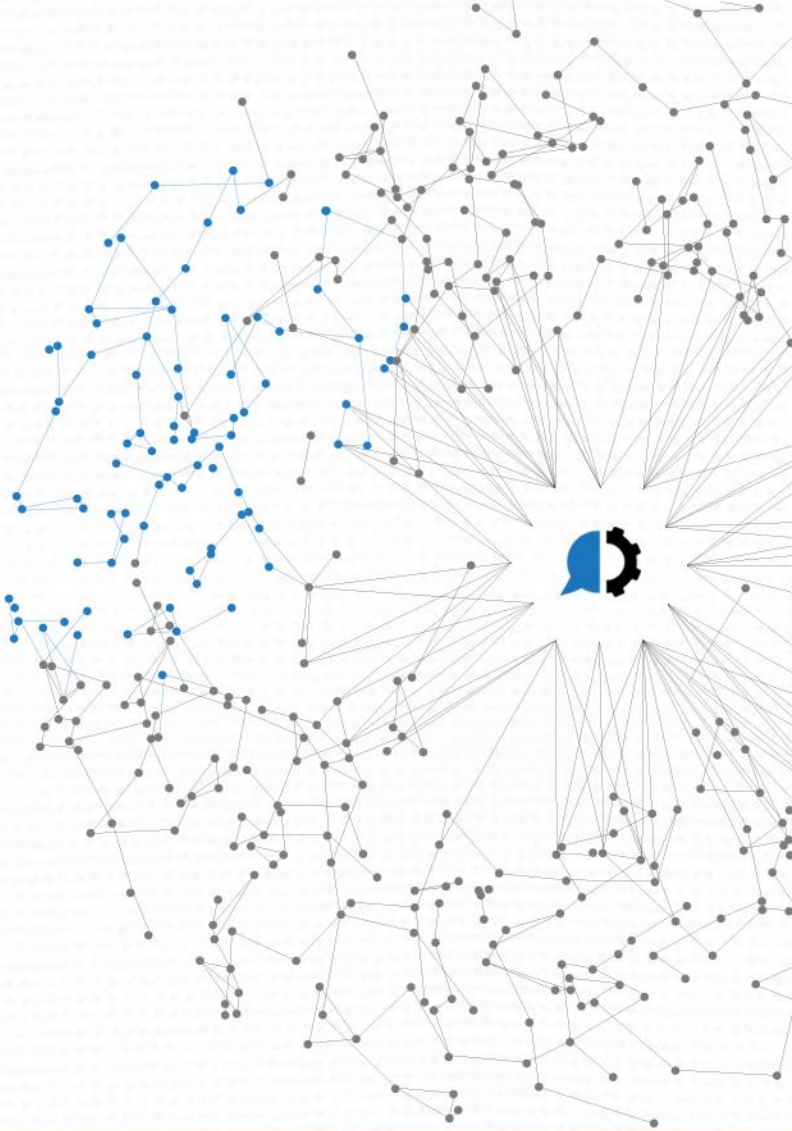
Rolling, automated prediction of future allocation and cost using trends, seasonal profiles and predictive analytics

Ongoing Management and Optimization



- Identify and address inefficiencies
- Track charges vs. budget
- Forecast future capacity requirements
- Predict cost

Other considerations and recommendations



UP NEXT

On-Demand vs Reserved Instances

- Discount Range depends on:
  - Commitment period (1-3 years)
  - Type (None/Partial/All Upfront)
  - Class (Standard/Convertible)
- Affinity – reservations can “float” across any accounts within a consolidated bill. “Capacity Reservations” can not float.

Example:

Reserved Instances - Recommendation

- Moderate gains for instances used less than 65% of the time
- "Per Instance" RIs quickly becomes unmanageable – group by application, function etc.
- Lifecycle Management
- R.o.T
  - 1-year term reservations typically break even after 6 months
  - 3-year term reservations typically break even after 9 months

Cost Management Challenges

Lack of overview:

Allocation and consumption scattered across the organization

No incentive to curb cloud spending:

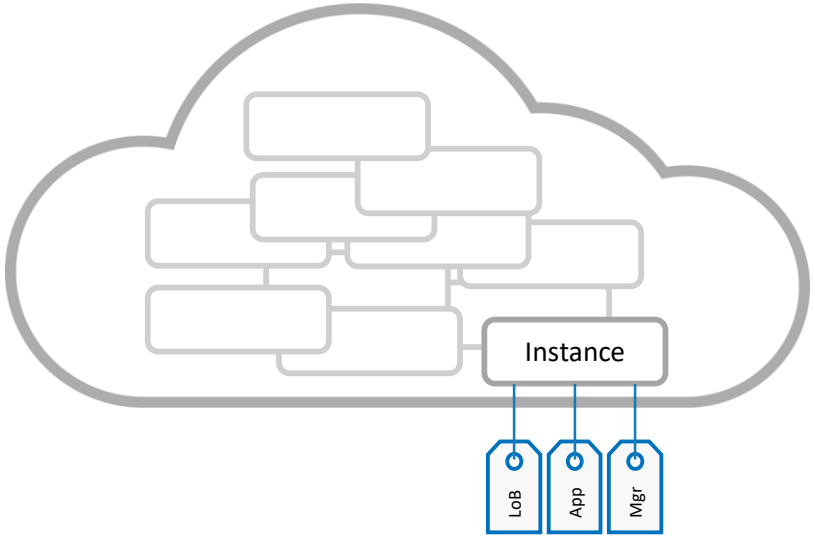
Standard billing insufficient for charge-out

Service valuation:

Lack of fully burdened costs and tracking by business unit

Cost Management Using Tags

- Automate tagging of cloud allocations:
  - Business Unit
  - Service or Application
  - Owner
- Global structure and naming conventions
- Track and report on consumption and cost by "peer" to enable charge out



Becoming Cloud Agnostic

Pros:

- Flexibility, more options
- Avoid lock-in
- Enable a multi-cloud strategy

Cons:

- Restricted to least common denominator
- Cloud Service Broker (CSB) or Cloud Mgmt Platform (CMP) dependency
- Cost vs. benefit?

Summary

Capacity Management in the Cloud:

Optimization

Delivering the “right” capacity at the lowest possible cost throughout the lifecycle

- Right-sized migration
- Continuous analysis
- Tagging and charge out policies

Predictability

Quantification of organic growth and forecasted demand to predict cost

- Modeling of growth factors
- Demand calendar
- Continuous improvement

Thank you for your attention.

Questions?