Essential insight for leaders of the digital economy.

The Languages of Capacity Planning: Business, Infrastructure & Facilities

Amy Spellmann; St. Louis CMG 4/19/16

The Languages of Capacity Planning: Business, Infrastructure & Facilities

Amy Spellmann - amy.spellmann@451research.com

Richard Gimarc - richard.gimarc@ca.com

November 5, 2015

CMG 2015

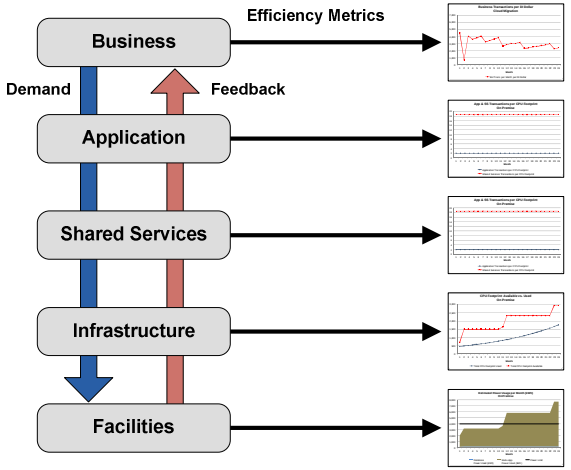
Session 511

The Languages of Capacity Planning

The Challenge

Capacity Planning Stack

- Multi-level hierarchy
  - Demand (↓)
  - Feedback (↑)
  - Efficiency Metrics (→)
- Supports all elements of today's Digital Infrastructure
- Implementation is straightforward & transparent



Challenge: ***Formulate & communicate demand***

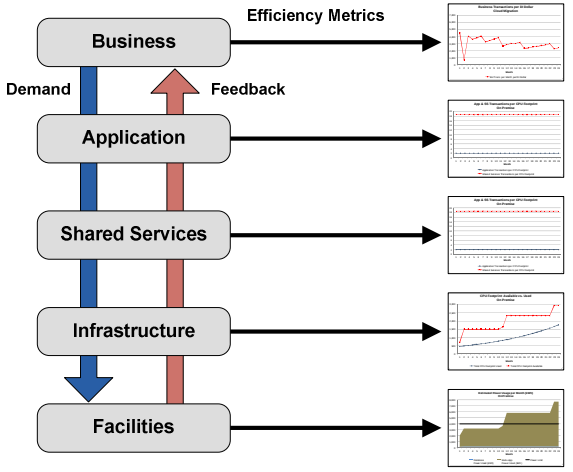
- How does Application talk to Business?
- How do Infrastructure planners talk to Facilities planners?

451 Research

The Languages of Capacity Planning

Our Approach

- Understand demand
- Revive the notion of Natural Forecasting Units (NFU)
- Use NFUs to express demand
- Apply this approach to a variety of execution environments

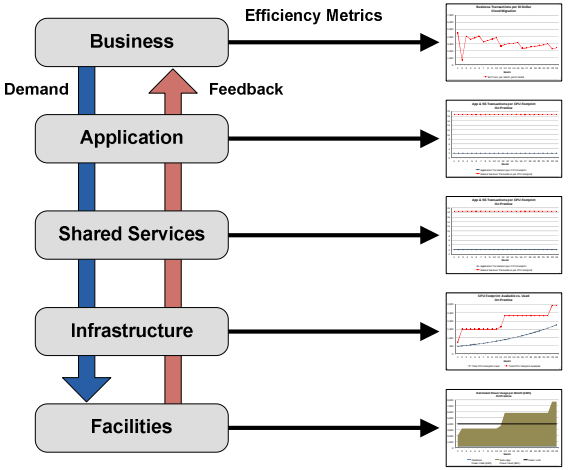


A Bit of History

2013 We presented the Capacity Planning Stack as a new way to view, analyze & communicate Digital Infrastructure capacity

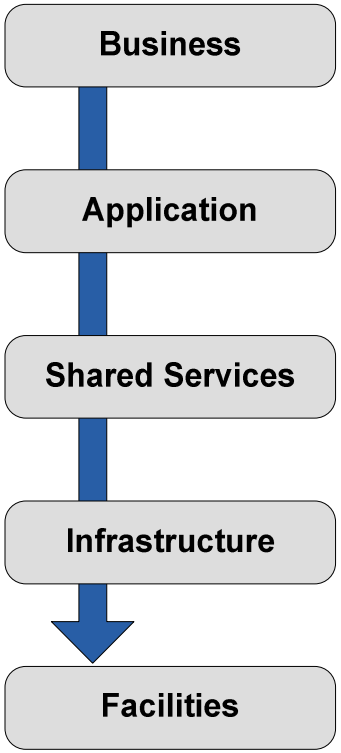
2014 We presented a taxonomy that organizes the metrics supporting the Stack

Today The Language of Capacity planning; utilizing the Stack to describe communication between the levels of the Stack for any service delivery model



451 Research

The Capacity Planning Stack



- Users of applications, infrastructure, facilities
- Focus on delivering business functions to end users

- Application owners
- Focus on the efficient delivery of applications to the business

- Provide shared services (e.g., message brokers, databases)
- Focus on efficient delivery of shared services to applications

- Hardware (e.g., servers, storage, network)
- Focus on optimized environment

- Data center space, power, cooling
- Focus on efficient design, maintenance, operations

451 Research

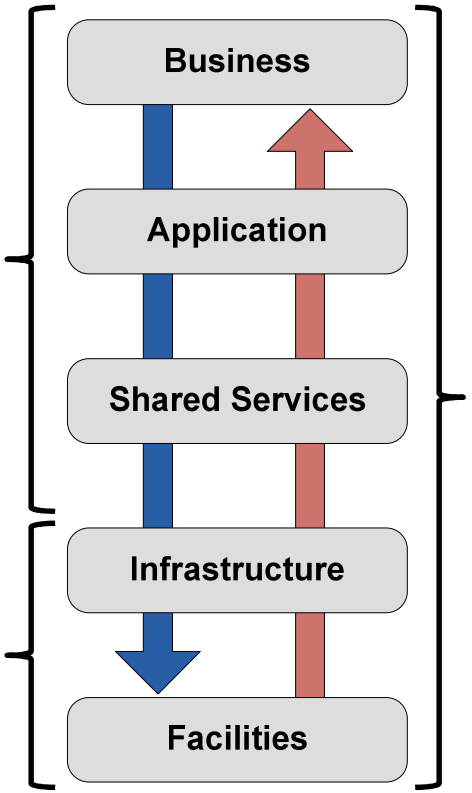
Demand & Feedback

Demand (↓)

- Volume & priorities
- Logical resource requirements
- Performance requirements & SLAs
- Budget

Demand (↓)

- Physical resource footprint & instances
- Performance requirements & SLAs
- Budget



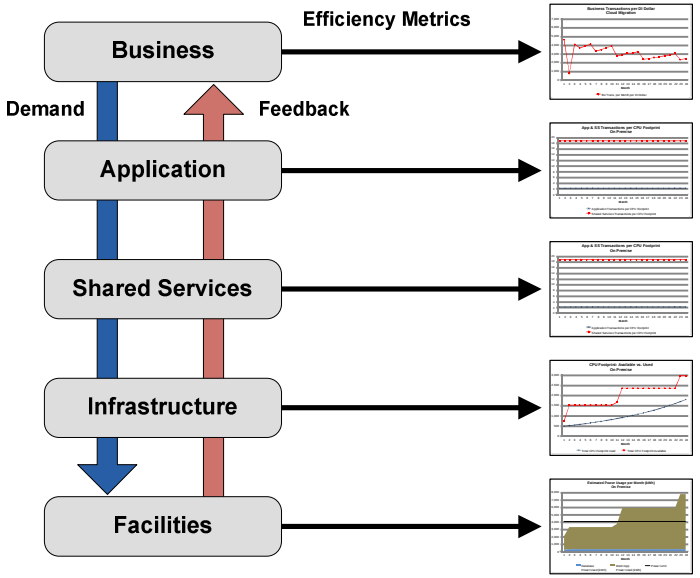
Feedback (↑)

- Cost
- Total time to satisfy
- Expected performance

451 Research

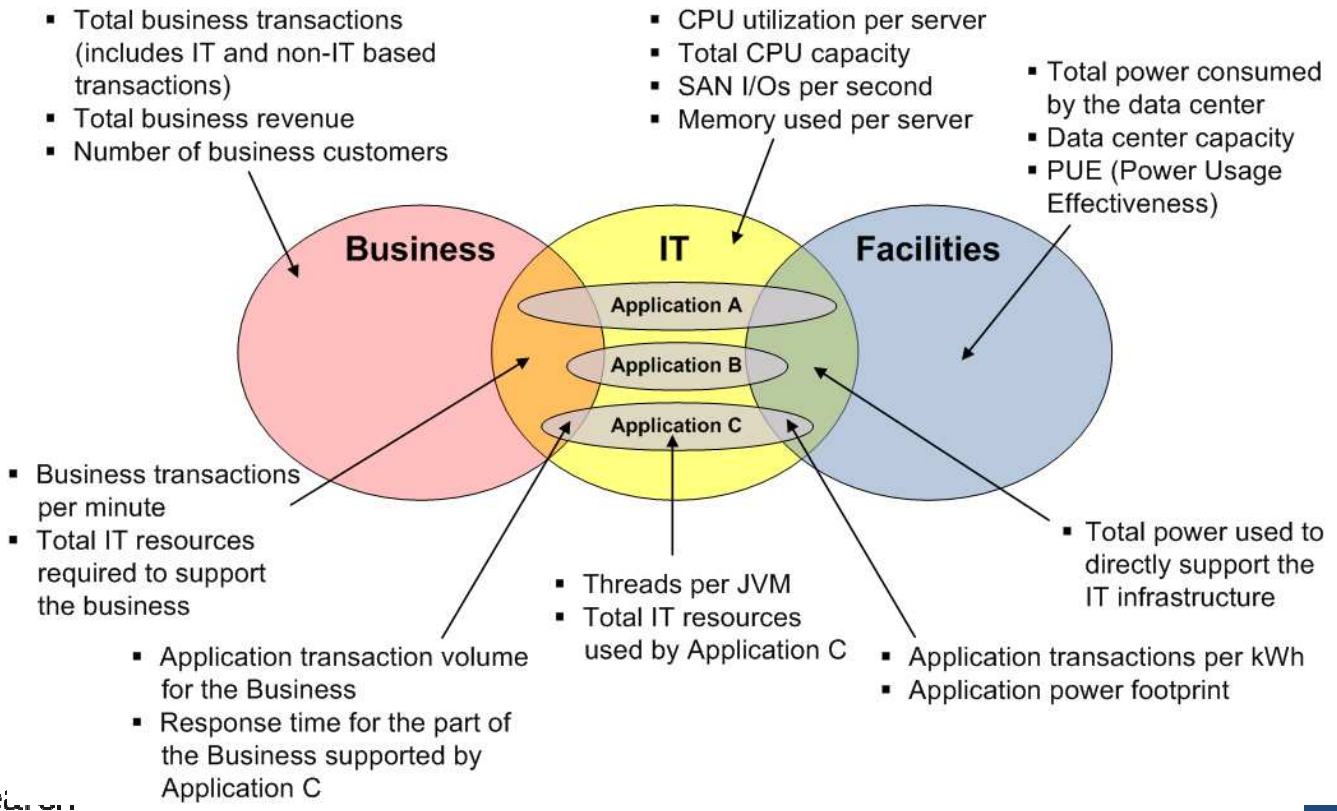
A Revolutionary Approach to Capacity Planning

The Capacity Planning Stack - a revolutionary approach to capacity planning that simplifies, structures and focuses the practice.



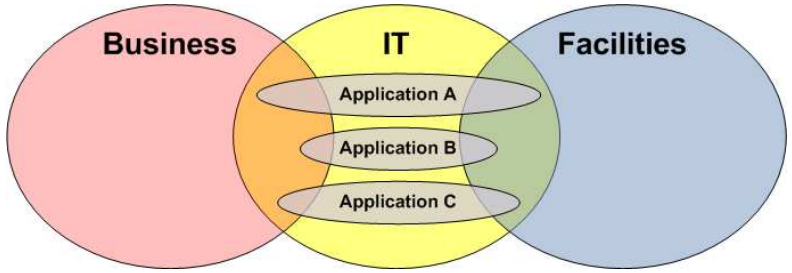
451 Research

The Capacity Planning Stack Taxonomy



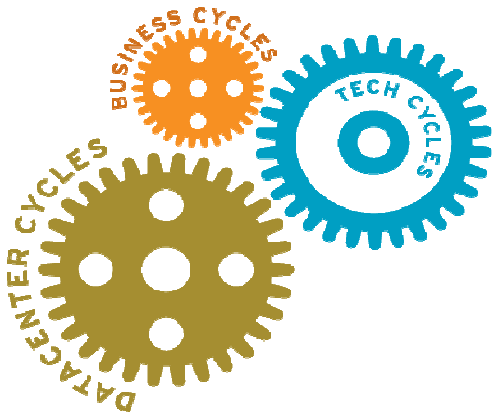
Benefits of the Stack Taxonomy

- Provides a structured approach to evaluating today's Digital Infrastructure
- Simplifies & guides the selection of metrics required to support decision making
- Equips the capacity planner with a structured way to think about, organize, communicate and collect the metrics that apply to the decision support process
- Paves the way for federated capacity planning across today's Digital Infrastructure



451 Research

Planning Horizons

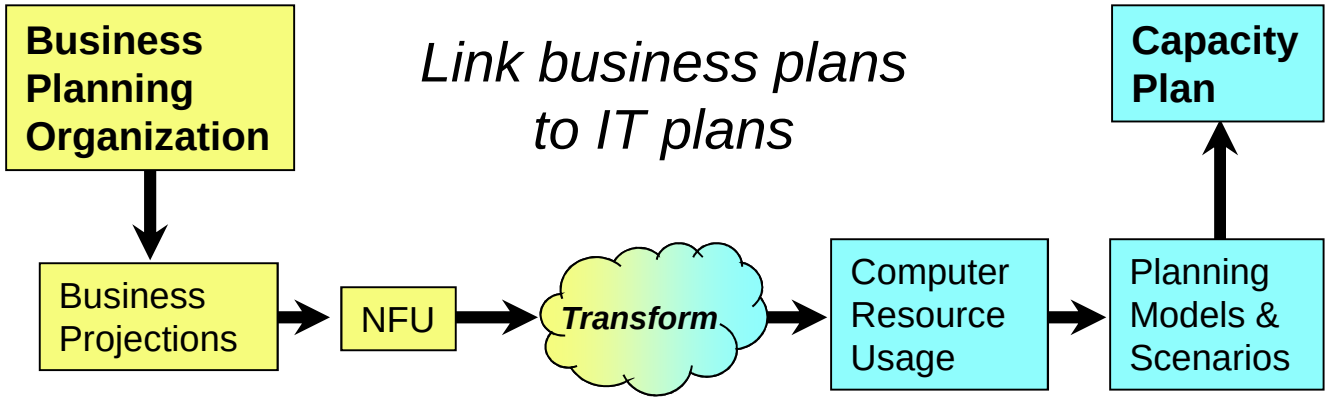


Business	Business planning horizons are generally in the range of 6 months to 1 year. Factors that influence their planning horizon include: ▪ New application deployment ▪ Seasonal fluctuations (e.g., Black Friday and Cyber Monday) ▪ Acquisitions and mergers ▪ Organic workload growth
Application	The Application planning horizon is generally in the 3 to 6 month range. Factors that influence their horizon include: ▪ New application rollout ▪ DevOps ▪ Organic workload growth for existing applications
Infrastructure	Infrastructure planning horizon is also in the 3 to 6 month range. Although the Infrastructure and Application levels have similar planning horizons, the factors that drive them are different. The factors that drive the Infrastructure planning horizon include the following: ▪ Technology refresh ▪ Application support ▪ Capacity demand ▪ Procurement timelines
Facilities	Facilities planning horizons are in the range of 5 to 10 years. Since they are concerned with the hosting data center they want to plan for as few changes as possible because change is expensive and time consuming. Changes to support additional power, space and cooling are not small or incremental; that is the primary reason for their long-term view. A consequence is that Facilities generally builds in more room for growth than the other Stack levels.

451 Research

Natural Forecasting Unit (NFU)

- Business metric that can be related to the use of computer system resources (e.g., CPU, I/O, memory, network traffic)
- Examples: Number of accounts, customers, loans, hotel beds, cars manufactured, product orders, insurance contracts, ...

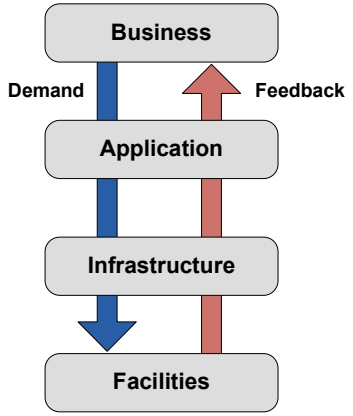


451 Research

NFUs & the Stack

- Each Stack level has its own notion of a Natural Forecasting Unit
- NFU demand is passed down the Stack from level to level.

	NFU Demand Factors (↓)	
Business	NFU:	Business volumetrics (e.g., number of loans)
Application	Task:	Translate Business NFU to application architecture & metrics
	NFU:	Application resource footprint and instance count (logical resources such as VMs, JVMs and threads)
Infrastructure	Task:	Translate Application NFU into physical Infrastructure requirements
	NFU:	Physical hardware requirements (e.g., servers, storage and network)
Facilities	Task:	Translate Infrastructure NFU into Facilities space, power & cooling
	NFU:	Power draw, space & cooling requirements



451 Research

NFU - Examples

Mortgage company

- Number of existing loans
- Number of new loans

Bank

- Number of accounts
- Number of customers using electronic bill payment

Insurance company

- Number of current insurance contracts
- Number of claims per month

Hotel chain

- Number of hotels & beds
- Number of registered guests

Challenge

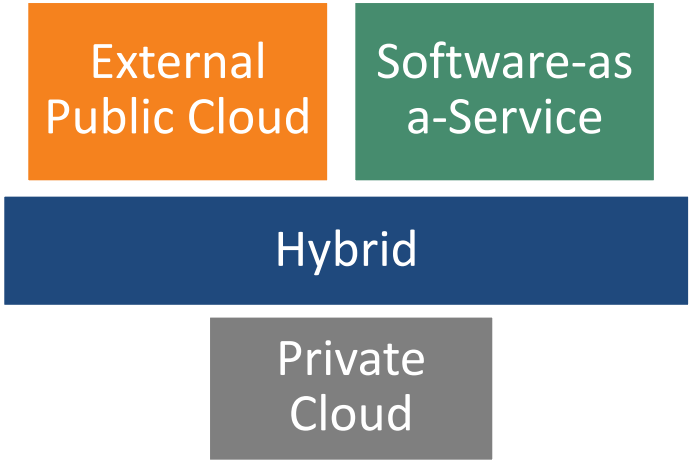
– How do you transform an NFU into computer system resource usage?

How much CPU, I/O, etc. are required to:

- Service an existing home loan
- Create a new loan
- Generate monthly statement
- Process a customer claim
- Register a new guest
- Generate bill for checkout

Best Execution Venue

Where is the best place to run a workload?



External Public Cloud:
Servers and storage on demand
(AWS, Rackspace)

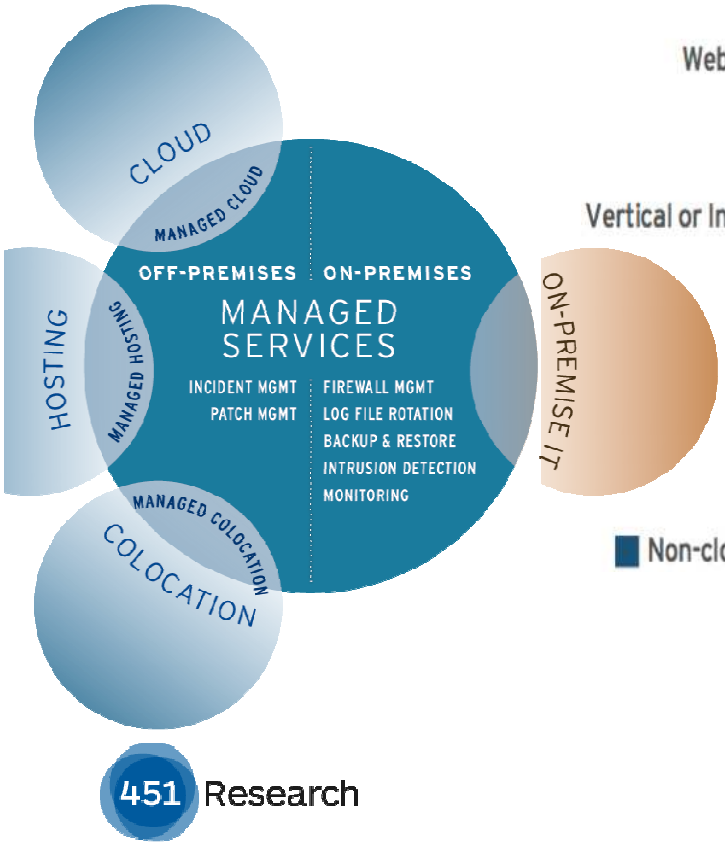
Private Cloud:
We own(or pretend we own) the servers and storage (Cloudstack, OpenStack, hosted private cloud)

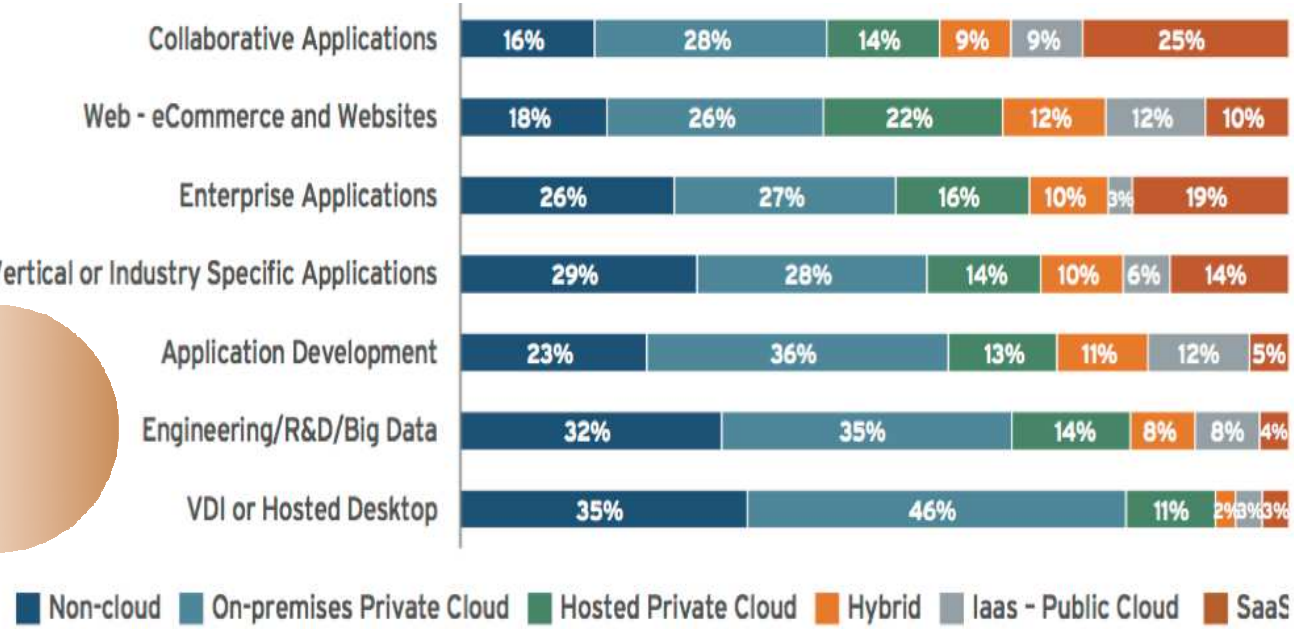
Hybrid:
Private and public resources managed together as needed (RightScale, Dell)

SaaS:
Applications on demand
(Salesforce, Google Docs)

451 Research

451 Research Best Execution Venue Thought Leadership

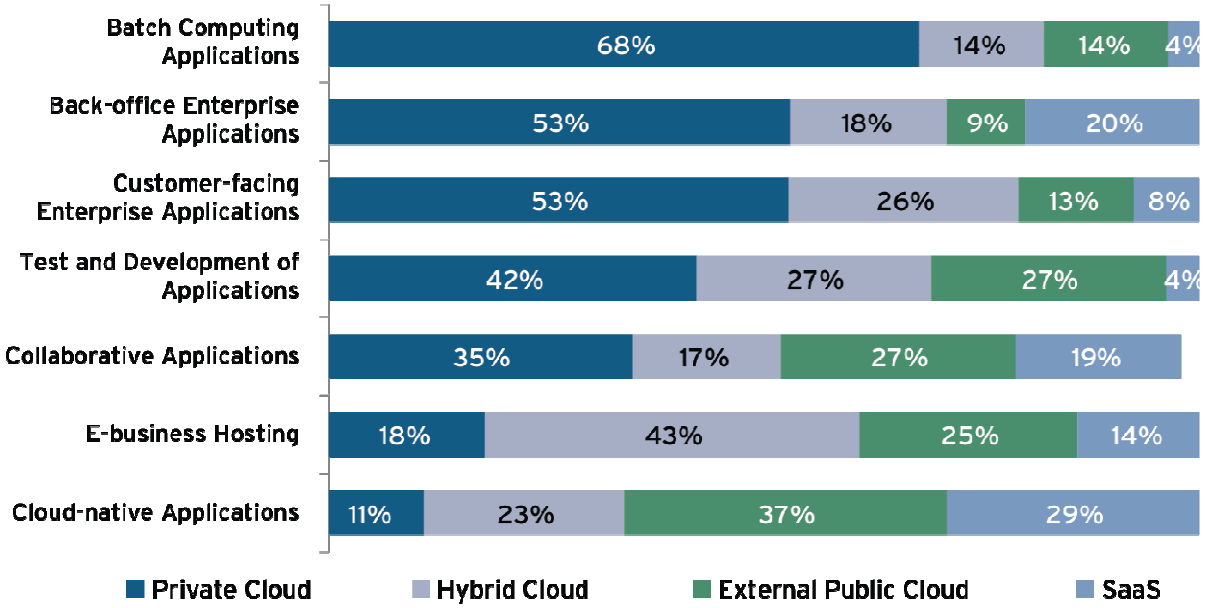




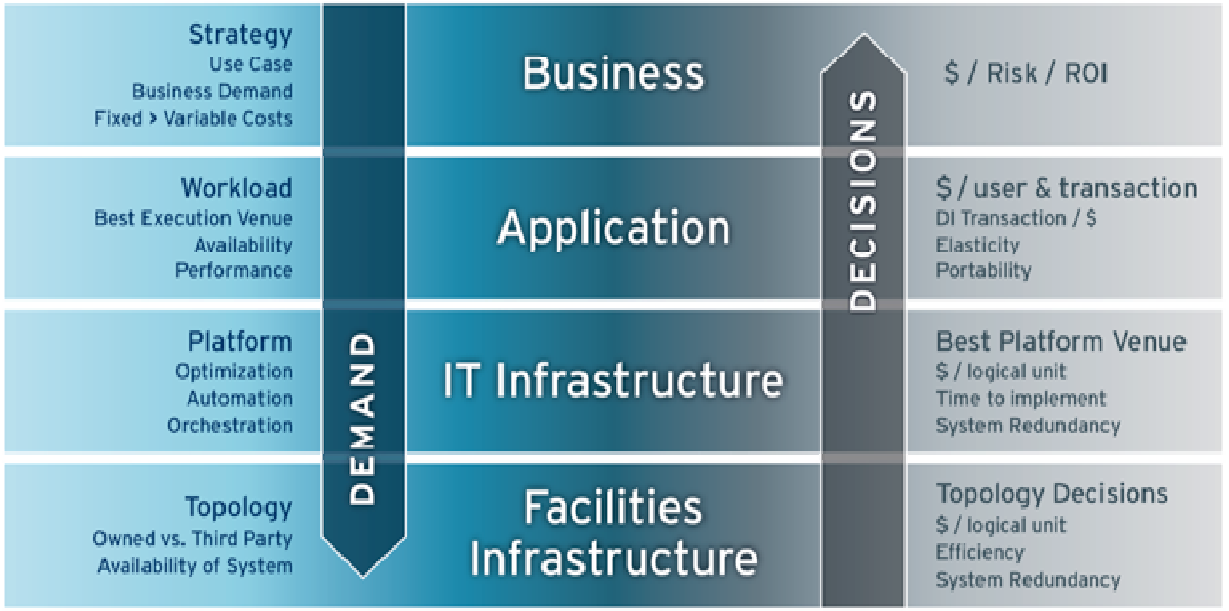
451 Research

Best Workload Execution Venue: Cloud

For each of the following categories of workload/business functions, what is your primary deployment method likely to be in the next two years (internal private cloud, external public cloud, hybrid cloud, or SaaS)?

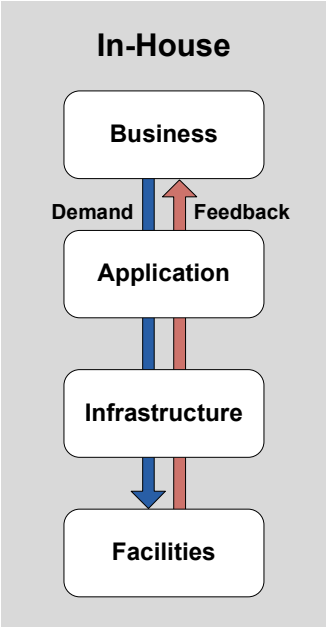


Business Demand Aligned to Best Execution Venue



451 Research

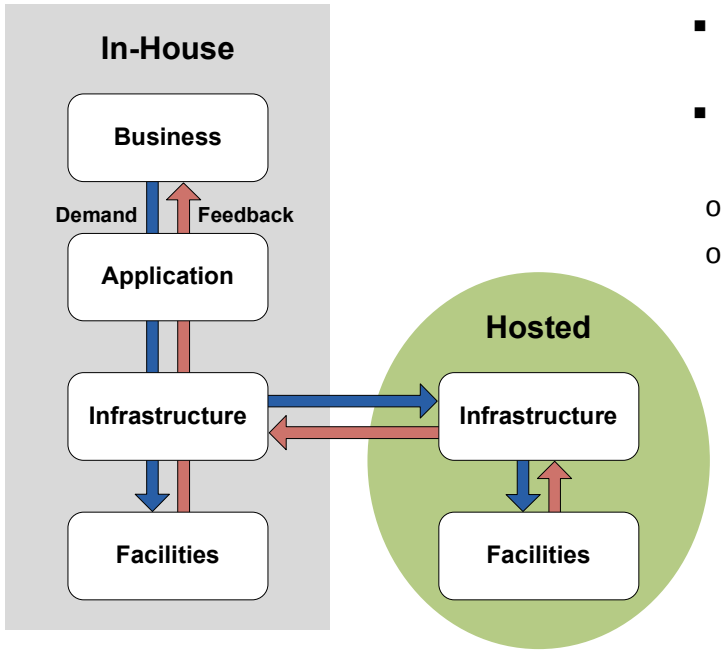
Traditional



- NFU Demand (↓)
  - Business to Application
    - Business volumetrics
  - Application to Infrastructure
    - Resource footprint & instance count
    - Logical resources
  - Infrastructure to Facilities
    - Physical hardware requirements
- Feedback
  - Describe what will be implemented to satisfy the demand at each level
- Budget (↓) & Cost (↑)

On-Premise Digital Infrastructure | NFUs & Communication/Language

Hosted

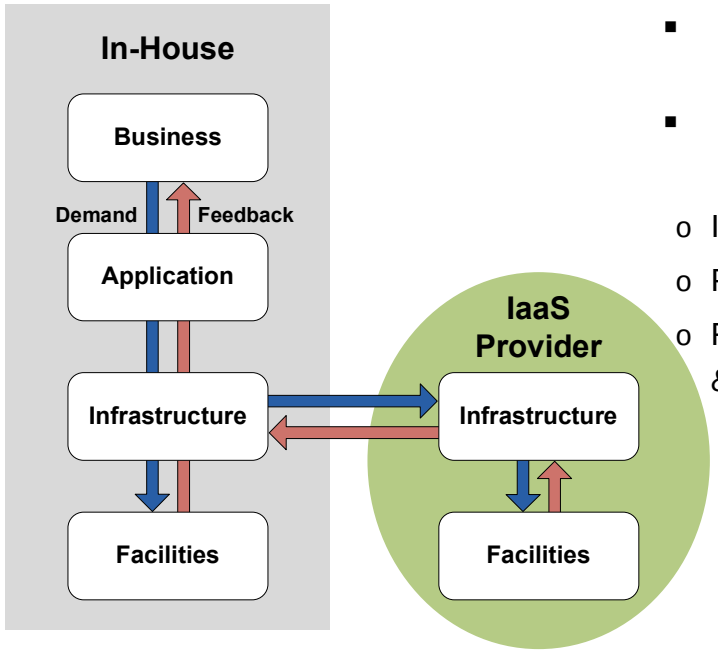


- Internal is identical to the traditional On-Premise model
- External communication of demand to hosting provider
  - Hardware requirements
  - Feedback is cost based on space, power and cooling at the provider's site

Add External Infrastructure & Facilities

NFUs & Communication/Language

IaaS

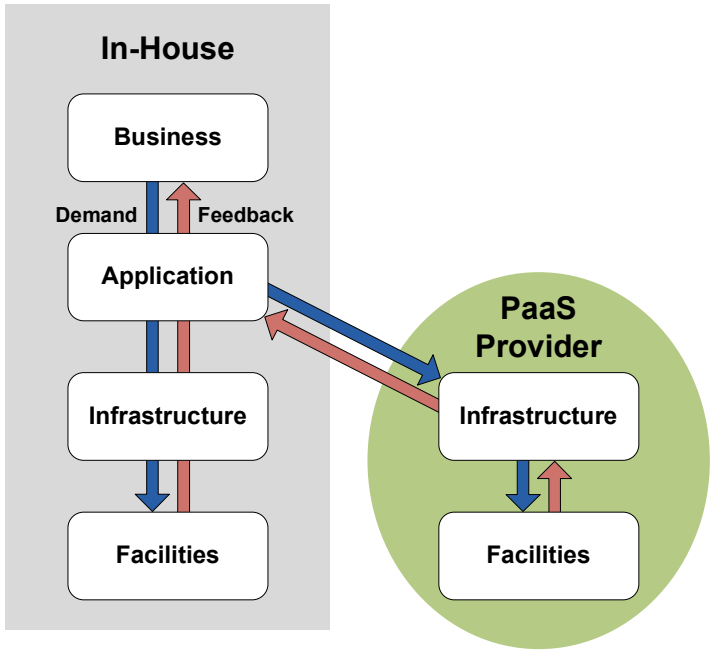


- Internal is identical to the traditional model
- External communication of demand to IasS provider
  - In-house Infrastructure to Cloud Provider
  - Physical hardware requirements
  - Feedback is cost based on instance usage & type, storage and network actuals

IaaS can be Public or Private Cloud

NFUs & Communication/Language

PaaS

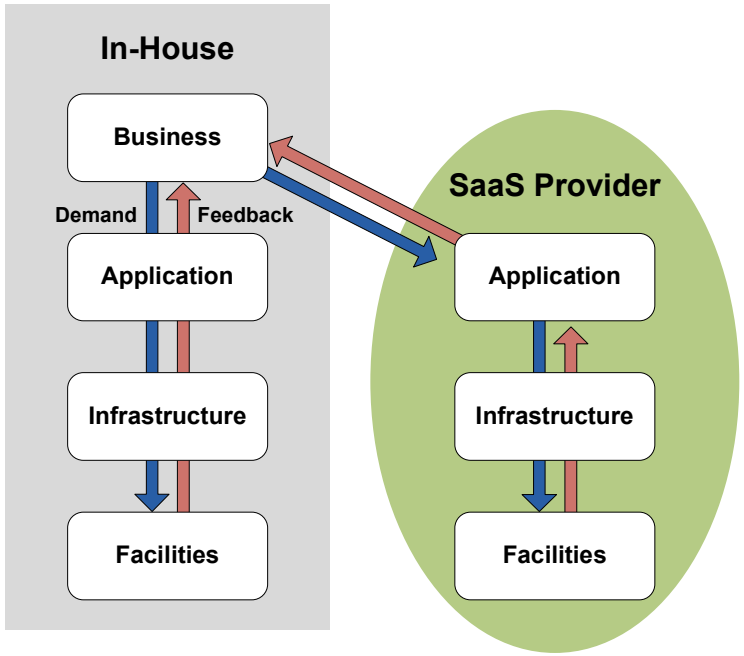


- Internal is identical to the traditional model
- External communication of demand to PaaS provider
  - Application to PaaS Provider
  - Resource footprint & instance count
  - Logical resources
  - Primary use is development
  - Feedback is cost based on instance usage & type, storage and network actuals

PaaS can be Public or Private Cloud

NFUs & Communication/Language

SaaS



- Internal is identical to the traditional model
- In-house LOB to SaaS Provider
  - Number of users, size of data stores, network traffic requirements, any specific SaaS customization or add-on features
  - Feedback is cost based on actual users, special features, storage and network

SaaS will be external Cloud

NFUs & Communication/Language

IT-AS-A-SERVICE

Managing Capacity Across
Best Execution Venues

451 Research

What is ITaaS?

- **ITaaS is a business and operating model**

  Recognizes that lines of business (LOBs) have options for IT resources and services and the IT organization must compete for their business

- **ITaaS is enabled by a software-defined architecture**

  For private, public and hybrid clouds where infrastructure (e.g., servers, storage and networking) is virtualized, automated by software and delivered as a service

- **Uses self-service catalogues**

  Expose services (applications, tools, resources) to LOBs and users

- **IT organization acts as an advisor and broker**

  To recommend and/or curate additional resources and services, on-demand, to maintain resiliency and accommodate changing business needs

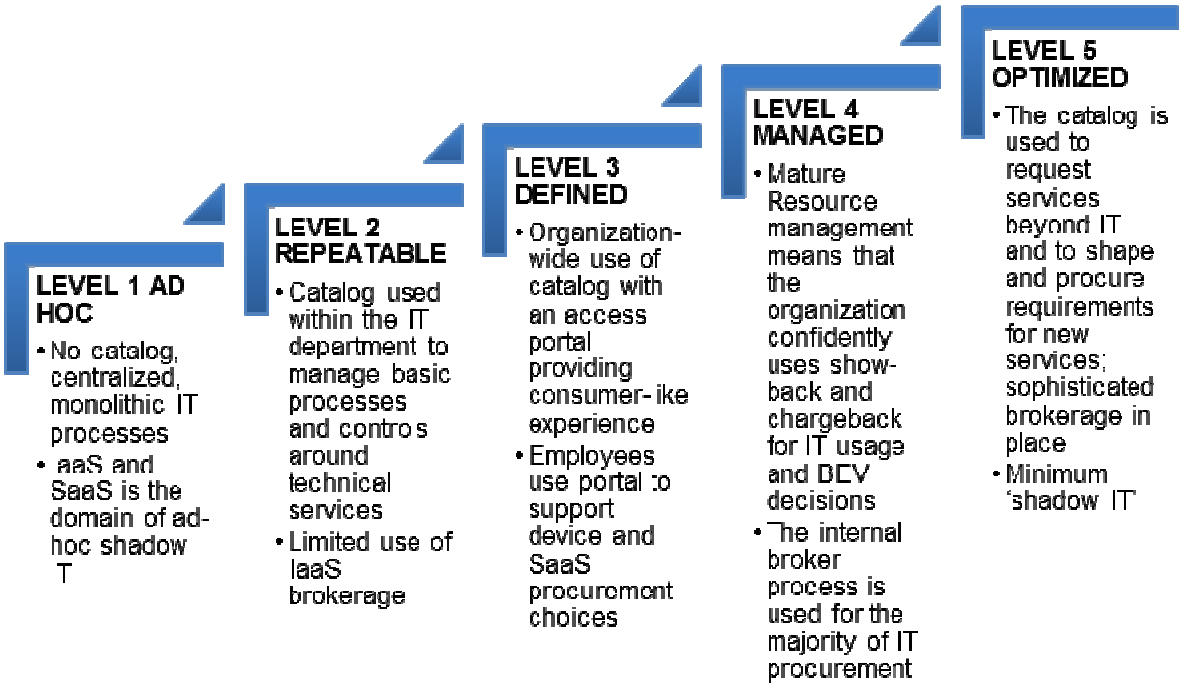
- **Performance metrics**

  Used to measure customer satisfaction and competitive positioning

ITaaS - Best Execution Venues

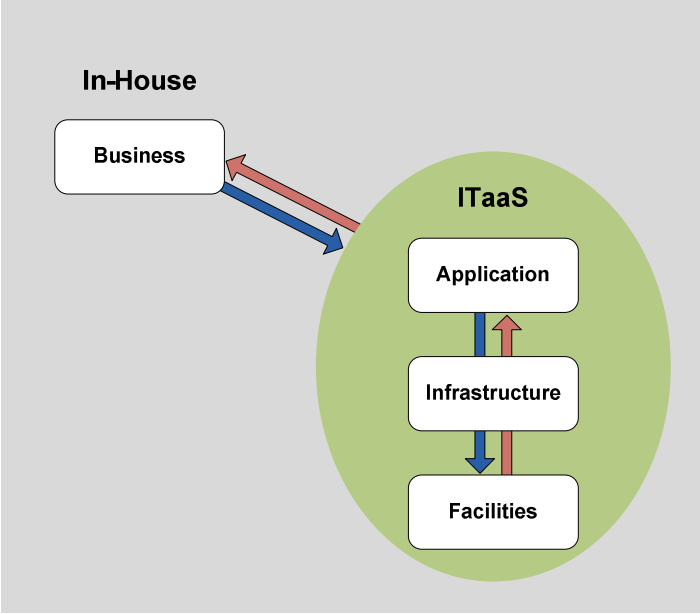
- **Every application has a best execution venue**
  - Some are mature, others are evolving
  - All are headed toward "the cloud"
- **Cloud computing is mainstream**
  - This means choice, access and diversity for IT
  - Benefits are material - business drives choice of venue
- **There are ways to systematically make the best choice of venue for each workload**
  - Enterprises are making these choices today
  - Providers are/should be targeting these workloads and creating the venues
  - There is a snowball effect across the IT landscape as cloud begets automation and automation begets growth

ITaaS - Maturity Model



451 Research

ITaaS - Basic

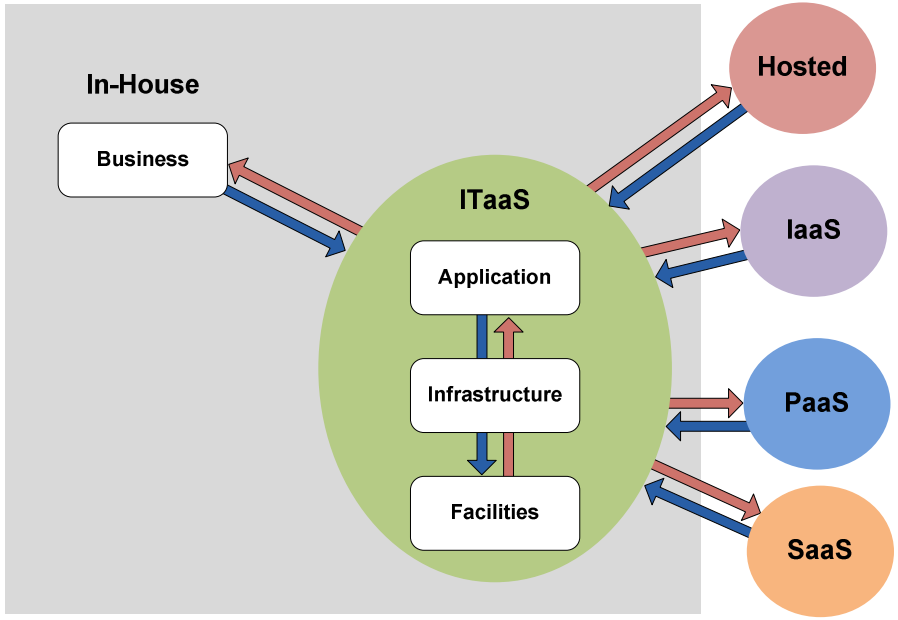


- Business communicates with ITaaS
- ITaaS makes its own internal decisions based on
  - Business demand
  - ITaaS internal operation
  - ITaaS will choose the best execution venue
- Feedback is cost based on what is required to provide the satisfy the Business demand

ITaaS communication

NFUs & Communication/Language

ITaaS - Evolved



- ITaaS will choose the best execution venue

ITaaS communication

NFUs & Communication/Language

Case Study: Global Fortune 500 Consumer Product Co.

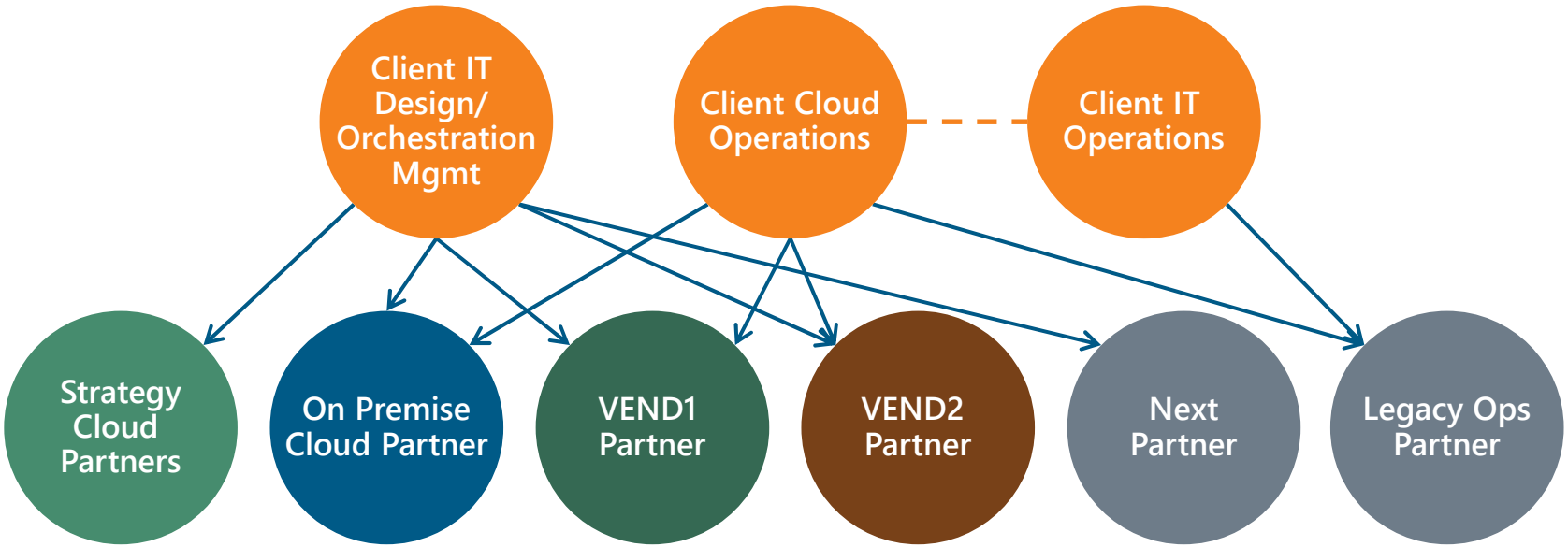
Pre transformation state:

- Owned/operated multiple data centers
- Utilized third party managed services outsource provider
- Typical infrastructure design/develop/deploy times of 6+ months
- Traditional siloes of technology, development, operations and management
- All APP development efforts were bespoke standalone projects with no sharing of infrastructure

Cloud-First post transformation state:

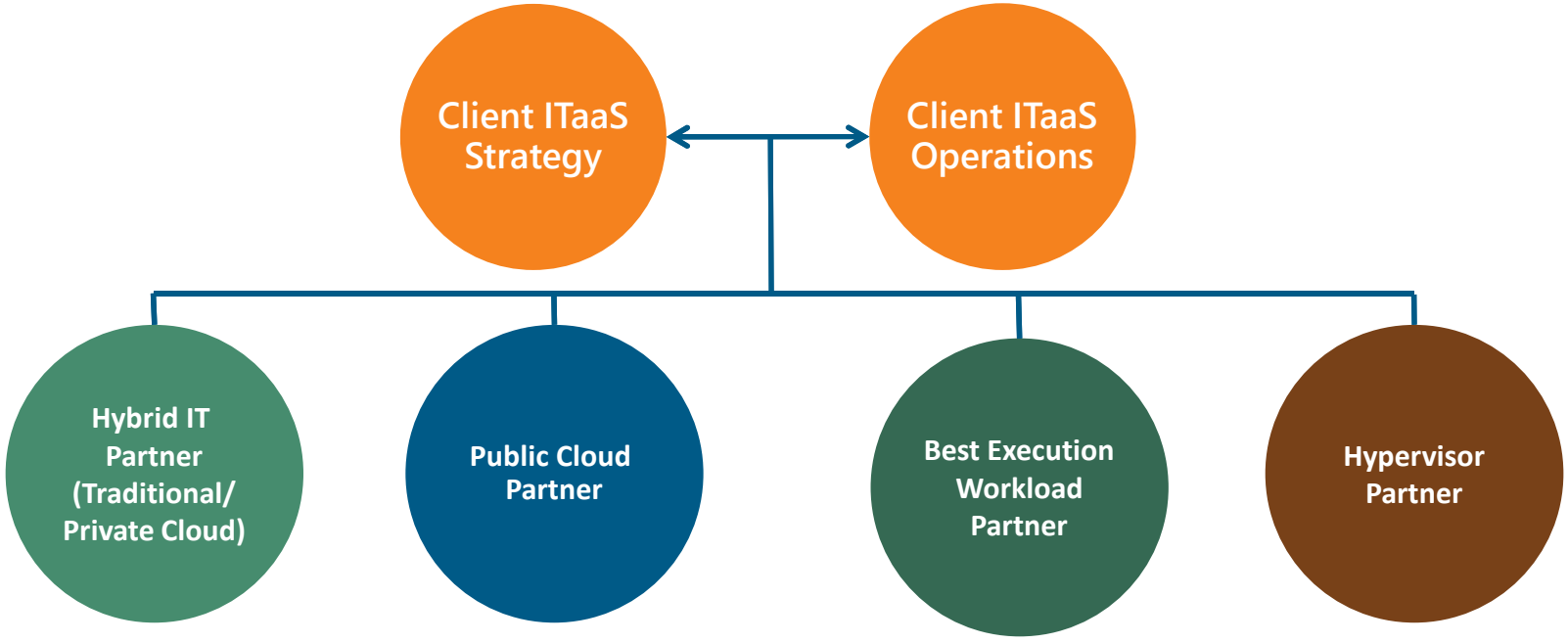
- Started with consumer facing APPS deployed in private cloud.
- Experienced sticker shock of 10% rise in APP environment cost increase
- Version 2 deployed to AWS cloud technologies. On demand, elastic and turned OFF when not in use
- 20% reduction in TCO vs. pre transformed (non cloud) state
- Design/build times of less than 2 weeks start to finish

Previous Client Organizational Structure: Complex/Siloes



451 Research

451 Research Recommended IT Organization



- Outsourcing?
- Legacy Owned
- Private Cloud
- DRaaS

- SaaS
- IaaS
- PaaS
- DRaaS

- Consumer Facing
- Enterprise Services
- Orchestration Execution
- Data portability

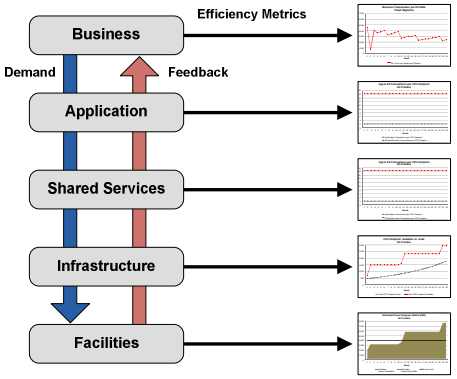
- Core Tools
- Automation
- Compute

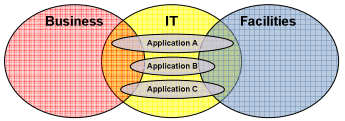
451 Research

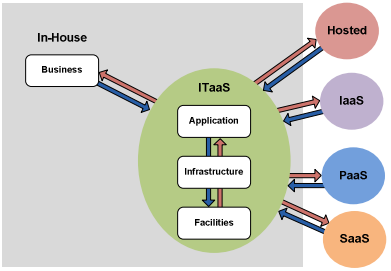
The Languages of Capacity Planning

Summary

- ✓Started with today's Digital Infrastructure
- ✓Created the Capacity Planning Stack to describe the capacity planning process across the breadth and depth of the Digital Infrastructure
- ✓Developed a taxonomy to organize our universe of metrics
- ✓Revived the notion of an NFU - used as touch points between Stack levels
- ✓Illustrated how the Stack can be applied to a variety of execution venues
- ✓Final step was to show how using the Stack is critical for ITaaS





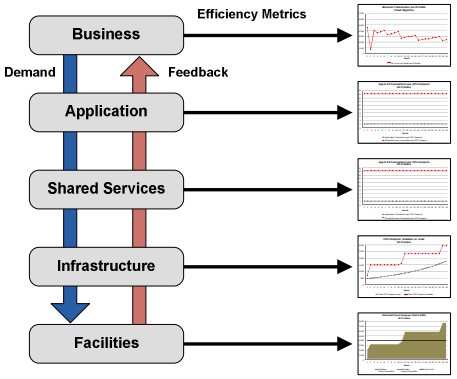


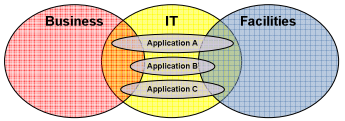
451 Research

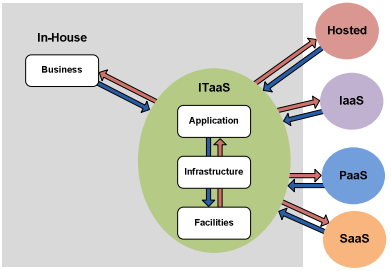
The Languages of Capacity Planning

What does this mean to you?

- ✓Use the Stack to think about, discuss & approach capacity planning – all execution venues
- ✓Leverage the Stack taxonomy to organize and describe your metric requirements
- ✓Think NFU
- ✓Language "challenges" can often be traced back to NFU usage & assumed understanding
- ✓Planning for the cloud fits into the Stack's framework
- ✓*Navigating the evolving landscape will require capacity planners to emerge as leaders in the transformation from traditional IT to ITaaS*







451 Research

Make an IMPACT in La Jolla