

Large Scale Cloud Computing: Opportunities and Challenges

Yousef A. Khalidi
Distinguished Engineer
Microsoft

Overview

- ④ Cloud computing promise
 - ④ Architectural principles
- ④ Key: Large scale sharing
- ④ Challenges in achieving the promise
- ④ How Windows Azure is addressing these challenges
- ④ Remaining challenges and research opportunities

What is the **cloud**?



Delivering Information Technology as a Standardized Service

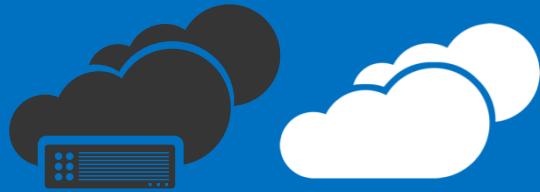
The illusion of infinite compute and storage

Customer datacenter

Partner datacenter

Public datacenter

Cloud Design Point



continuous services



connected devices



Architectural Principles

Virtualized Compute Fabric

Hypervisor-Based Isolation

Virtualized Network

Secure Connectivity & Isolation

Scale-Out Compute Model

Uniform Nodes,
VM as Unit of Capacity,
Optimize for MTTR

Each Node is a Cache

State must be externalized

Automation

App, OS, & HW Lifecycle Management

Rich Services

Distributed systems are hard to get right

High Scale and Sharing are Key

- ④ **Scale** is required to achieve cloud promise
 - ④ Economies of scale
 - ④ Elasticity
 - ④ Increased utilization
- ④ To afford the scale, cloud providers must **share** resources among many customers
- ④ **Uniform** systems are needed allow resource fungibility
- ④ **Virtualization** is necessary but not sufficient

(Some) Cloud Challenges

Targeting applications to the cloud

- Forklift approach does not work

Security

- Security of data and apps in a shared cloud

Compliance

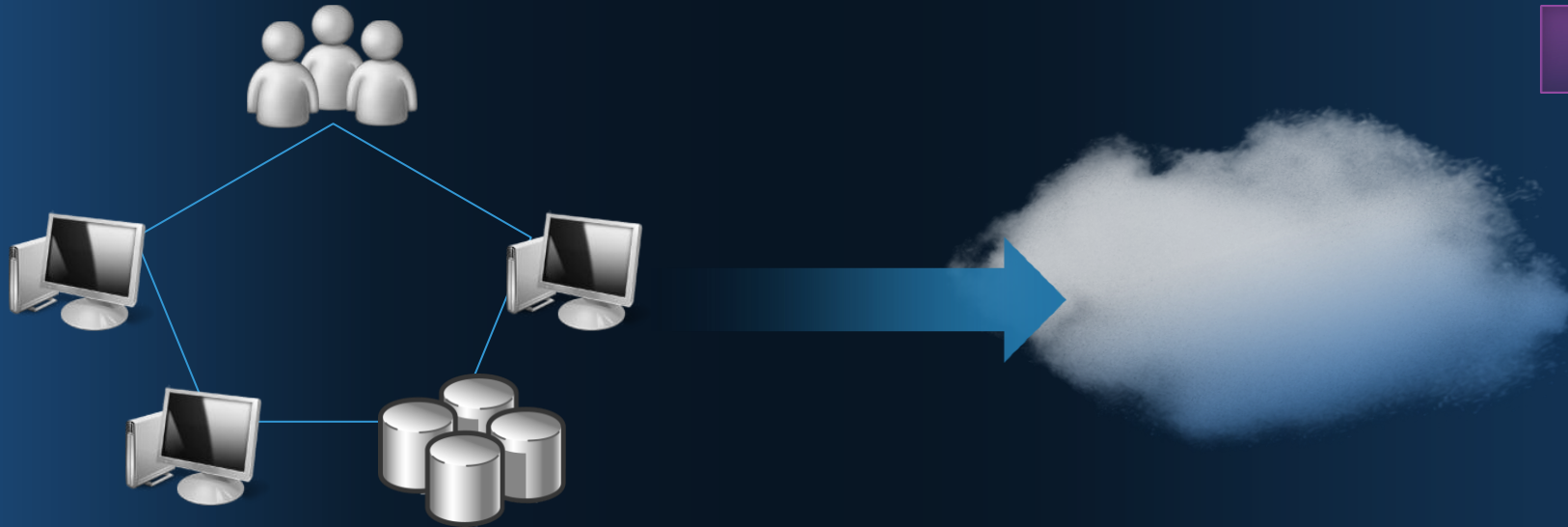
- Regulations, verification

Service level guarantees

- Availability SLAs, consistent performance SLAs

Not discussing today challenges of building and operating very large scale systems, business models, etc.

Targeting Apps to Cloud



Questions To Consider

Application State

Application Scale

Application Dependencies

Connectivity Needs

Performance Needs

Hardware Needs

Distributed Systems are hard

Some Easy Cases

e.g., web site sharing public data

Often, Forklift Approach
Will Not Work

Careful decomposition needed

Security, Trust, Regulations

- ④ Application isolation and data privacy
 - ④ Over shared network and compute fabrics
- ④ Identity and access management
- ④ Policy controls for data & applications
 - ④ Do you know where your data reside?
- ④ Independently verified compliance
 - ④ Audit certifications, e.g., ISO/IEC 27001:2005

Service Level Guarantees

⌵ Availability

- ⌵ Typically expressed as percentage of total operations to succeed
- ⌵ Calculated over some fixed period
- ⌵ Example
 - ⌵ “We guarantee that at least 99.9% of the time we will successfully process correctly formatted requests that we receive to add, update, read and delete data”

⌵ Performance

- ⌵ Typically expressed as minimum (or a range) of memory, network bandwidth, or VM computing power
- ⌵ Applications want end-to-end guarantees

How Windows Azure is addressing these challenges



Windows Azure



Developer Experience



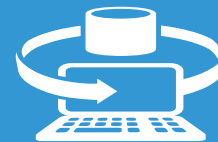
Use existing skills and tools



Compute



Storage



Database



CDN



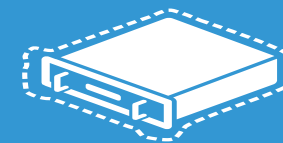
Service Bus



Access Control



Caching



VMs



Reporting

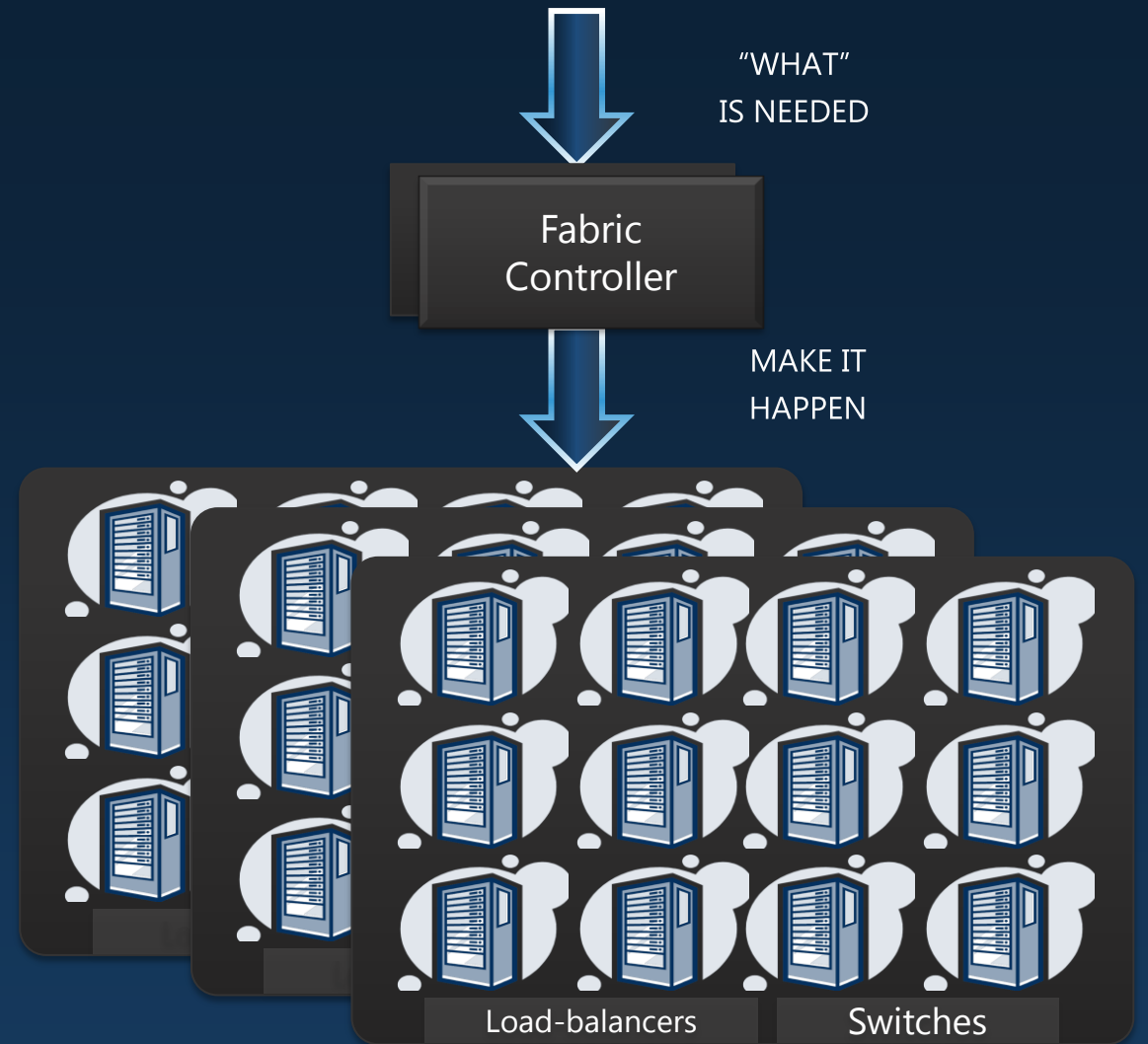
Computing At Scale: Windows Azure

- Redundancy at multiple levels
- Virtualized compute and network

- Model-driven automation
 - Logical application graphs mapped over physical topology

- Development and Management tools
 - REST APIs, multiple language and tool support
 - Web and enterprise management tools

- Rich services - built as Azure apps
 - Storage, SQL Azure, etc.
 - Secure cloud federation services



Windows Azure Global Presence

North America Region

Europe Region

Asia Pacific Region



Applying Architectural Principles in Azure

Virtualized Compute Fabric

HyperV based virtualization

Virtualized Network

Logical networks over physical network

Scale-Out Compute Model

Fixed set of VM/mem/bw sizes, all components optimized for MTTR

Each Node is a Cache

Durable network drives, local drives used as a cache

Automation

Model-driven automation of sw + hw, provisioning, configuration, and health

Rich Services

Blob, table, & database services, queues, caching, identity, ...

Cloud Challenges

Targeting applications to the cloud

- Forklift approach does not work

Security

- Isolation, data privacy, ...

Compliance

- Regulations, verification

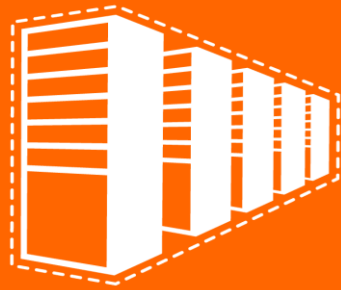
Service level guarantees

- Availability SLAs, consistent performance SLAs

Windows Azure – a platform for apps



Physical



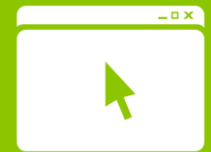
Virtual



IaaS



PaaS



SaaS

The Foundation for
Private Cloud

You can build apps that use both

Finished Services

Cloud Application Development

Rich Development Tools

Toolkits & SDKs

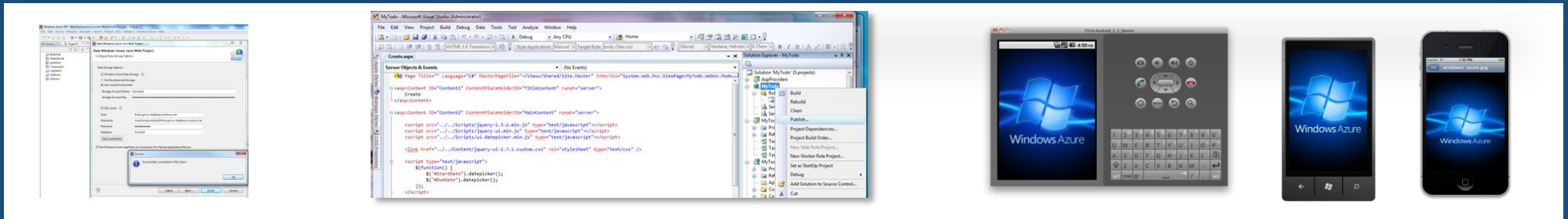
- ▶ SDK with platform emulator
- ▶ Phones – WP7, iOS, Android
- ▶ Languages – .NET, Java, PHP, node.js
- ▶ Social – Facebook, games
- ▶ Windows 8 Metro

Programming tools

- ▶ Microsoft Visual Studio
- ▶ Eclipse

Rich Services

- ▶ Storage, Database as a service, queues, Hadoop, ...





worker role



SQL Azure
database



blob storage



access control
service



content delivery
network



web role

building blocks



connect



caching



workflow



compute



content delivery
& storage



database



security
& identity



integration



networking



Service Bus
queues



Windows Azure
drive



traffic manager



Windows Azure
queues



SQL Azure
data sync

Managing Applications

Rich Management Tools

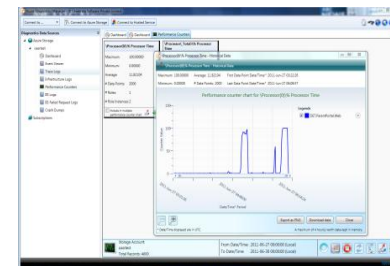
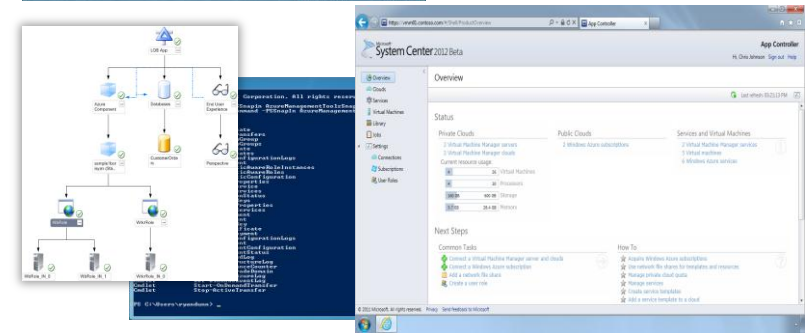
Web-based management

Ease of use
High visibility and control
REST public interfaces

Enterprise management

System Center
PowerShell
REST public interfaces

Growing ecosystem of 3rd party tools



Hybrid Cloud Connectivity

Cloud



Data Synchronization
SQL Azure Data Sync

Application-layer
Connectivity & Messaging
Service Bus

Security
*Federated Identity
and Access Control*

Secure Network Connectivity
Windows Azure Connect

On-premises



Cloud Challenges

Targeting applications to the cloud

- Forklift approach does not work

Security

- Isolation, data privacy, ...

Compliance

- Regulations, verification

Service level guarantees

- Availability SLAs, consistent performance SLAs

Security

⌵ Compute isolation

- ⌵ Hypervisor, Root OS, Guest VMs
- ⌵ Guest VMs are untrusted

⌵ Network isolation

- ⌵ Packet filtering at network edge

⌵ Control software

- ⌵ Fabric Controllers isolation
- ⌵ SSL mutual auth for internal traffic
- ⌵ Encrypted store for platform secrets

Authentication mechanisms

Subjects	Objects	Auth mechanism
Customers	Subscription (account)	Windows Live ID
Developers & Operators	Web portal REST APIs	Live ID (portal) Self-signed certificate
Applications	Storage	Storage account key
External applications	Azure apps	Customer-defined Platform support for single-sign on and federated identity scenarios

World Class Foundation



Physical facilities

Redundancy everywhere
State-of-the-art security and access control
Innovation in power efficiency



Geo-distribution

Diversity of location
Local and geo-replication
Redundant platform services and failover



Availability & security

Highly available platform services
Isolation and boundaries
Multiple lines of defense



Compliance & DR

Physical facility
Service specific
Preparedness, testing, refinement

Cloud Challenges

Targeting applications to the cloud

- Forklift approach does not work

Security

- Isolation, data privacy, ...

Compliance

- Regulations, verification

Service level guarantees

- Availability SLAs, consistent performance SLAs

Comprehensive Compliance Framework

Industry Standards and Regulations

Payment Card Industry Data Security Standard
Health Insurance Portability and Accountability Act

Media Ratings Council
Sarbanes-Oxley, GLBA, etc.

Controls Framework

Identify and integrate

Regulatory requirements
Customer requirements

Assess and remediate

Eliminate or mitigate gaps in control design



Predictable Audit Schedule

Test effectiveness and assess risk
Attain certifications and attestations
Improve and optimize

Examine root cause of non-compliance
Track until fully remediated

Certification and Attestations

ISO/IEC 27001:2005 certification

SSAE 16 attestations

Cloud Challenges

Targeting applications to the cloud

- Forklift approach does not work

Security

- Isolation, data privacy, ...

Compliance

- Regulations, verification

Service level guarantees

- Availability SLAs, consistent performance SLAs

Monthly Service Level Agreement

Compute connectivity

- Your service is connected and reachable via web
- Internet facing roles will have external connectivity

>99.95%

Instance monitoring and restart

- All running roles will be continuously monitored
- If role is not running, we will detect and initiate corrective state

>99.9%

Storage availability

- Storage service will be available/reachable (connectivity)
- Your storage requests will be processed successfully

>99.9%

Database availability

- Database is connected to the internet gateway
- All databases will be continuously monitored

>99.9%

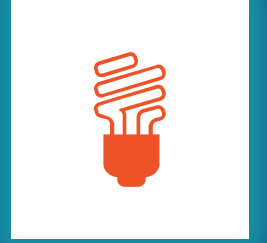
Service bus and access control availability

- Service bus and access control endpoints will have external connectivity
- Message operation requests processed successfully

>99.9%

Remaining Challenges & Opportunities

- ⌚ Network performance guarantees
 - ⌚ End-to-end latency and bandwidth guarantees
- ⌚ Optimizing large-scale computing over data
 - ⌚ Shipping computation to data (and vice versa?)
- ⌚ Intersection of economics and cloud computing
 - ⌚ Use of economic models for pricing and resource management
- ⌚ Machine learning and very large system management
 - ⌚ Error prediction, root cause analysis, trend analysis



Microsoft[®]

© 2012 Microsoft Corporation. All rights reserved. Microsoft, Windows, Windows Vista and other product names are or may be registered trademarks and/or trademarks in the U.S. and/or other countries. The information herein is for informational purposes only and represents the current view of Microsoft Corporation as of the date of this presentation. Because Microsoft must respond to changing market conditions, it should not be interpreted to be a commitment on the part of Microsoft, and Microsoft cannot guarantee the accuracy of any information provided after the date of this presentation. MICROSOFT MAKES NO WARRANTIES, EXPRESS, IMPLIED OR STATUTORY, AS TO THE INFORMATION IN THIS PRESENTATION.