Demystifying the Fog: Why Cloud Computing is Requiring us to Rethink Resiliency at the Edge

A greater dependence on cloud-based applications means businesses must rethink the level of redundancy of the physical infrastructure equipment remaining on-premise, at the "Edge".

Kevin Brown, SVP of Innovation and CTO, IT Division, Schneider Electric

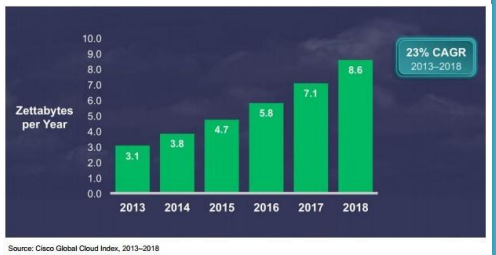
BICSI

Agenda

1. Trends: IoT, cloud, edge
2. What the “edge” looks like today
3. Common data center physical infrastructure practices
4. Re-thinking how we design at the edge
5. Re-defining a “failure”
6. Building reliability at the right levels

Digital traffic is expanding annually by 23%+

..by 2018



8.6 Zettebytes of IP Traffic

that's **8,600,000,000,000,000,000,000** bytes

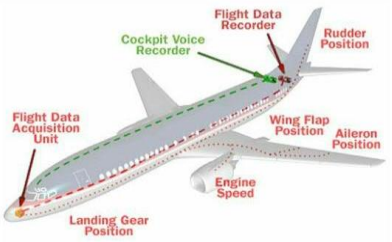
Source: Cloud Index Report 2014

Growth is driven by the number of connected users and the Internet of Things (IoT)

3.7 Billion Internet users

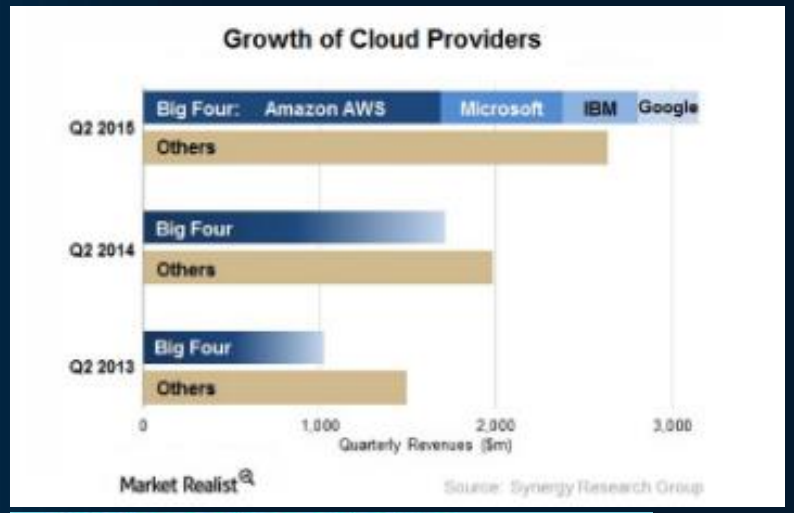
21 Billion network devices

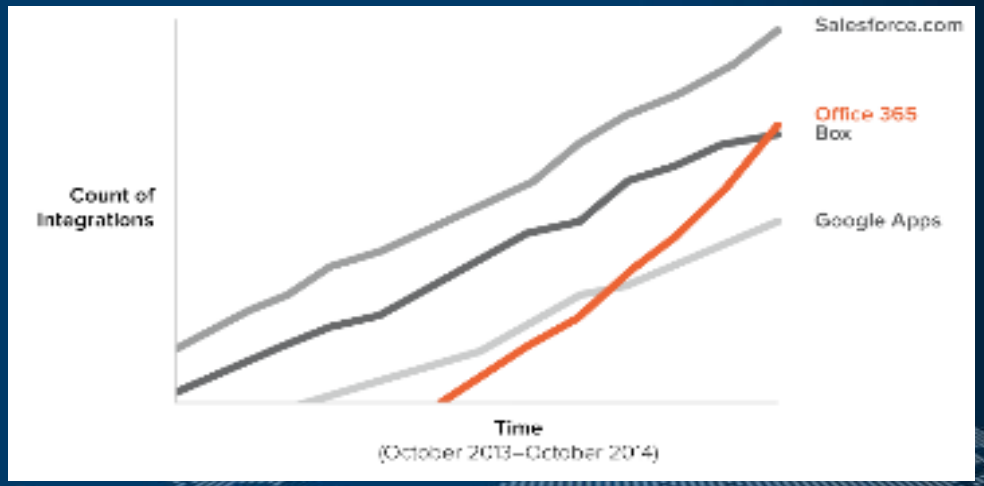
1.3 Million video views per min

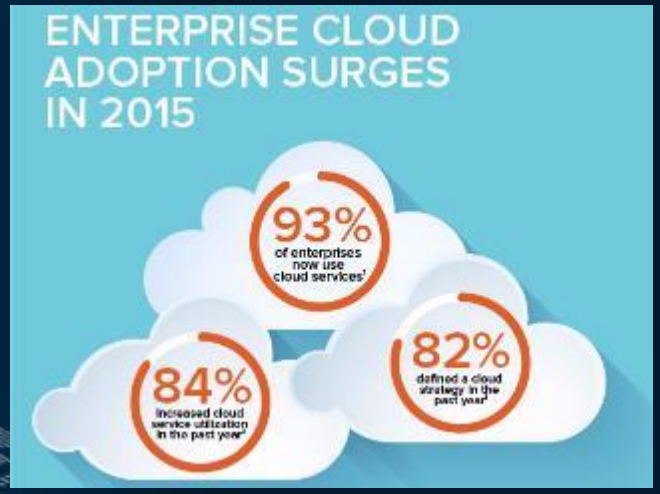


40 TB transmitted / flight hr

Enterprise cloud-based IT is growing rapidly

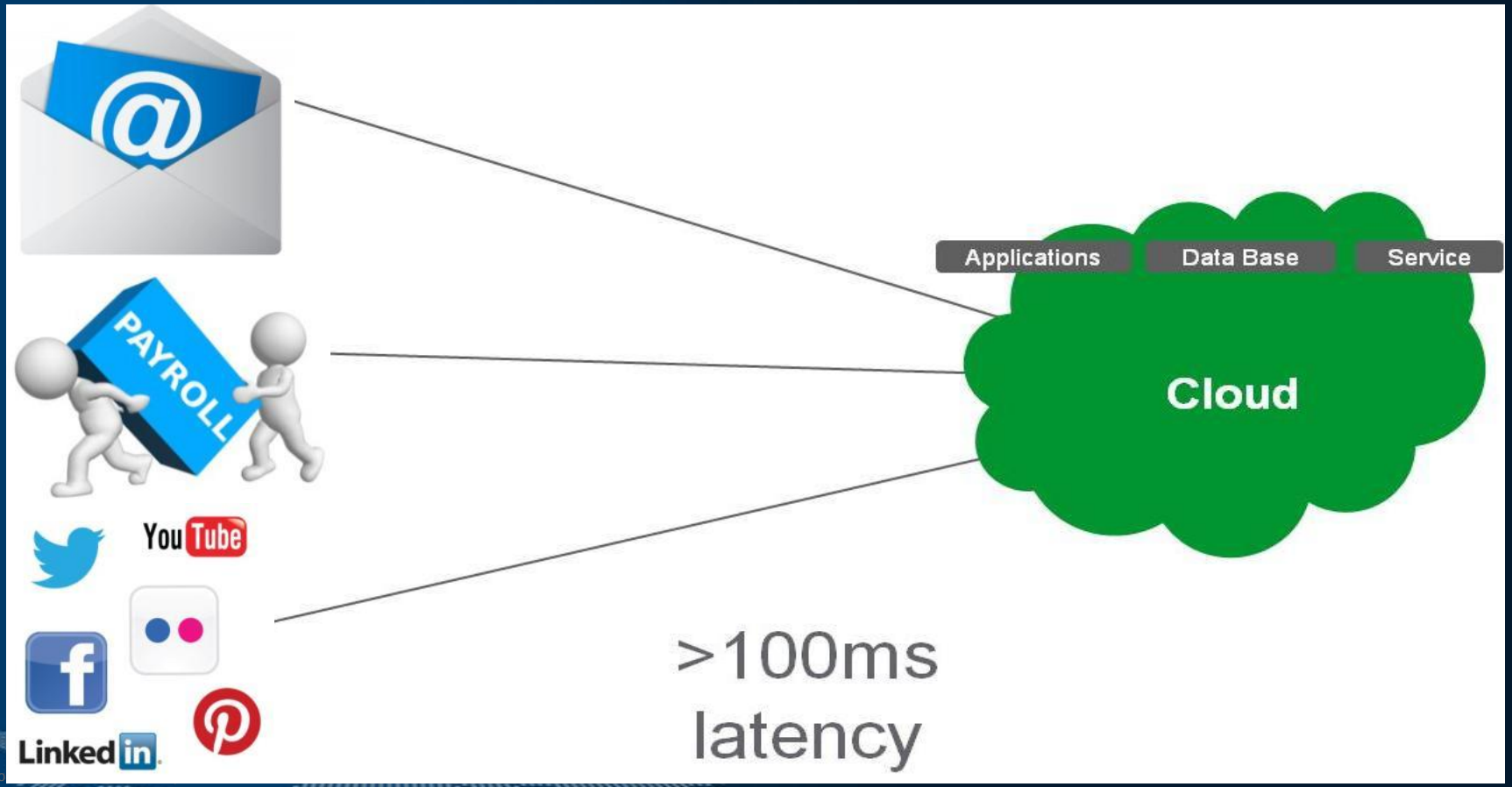




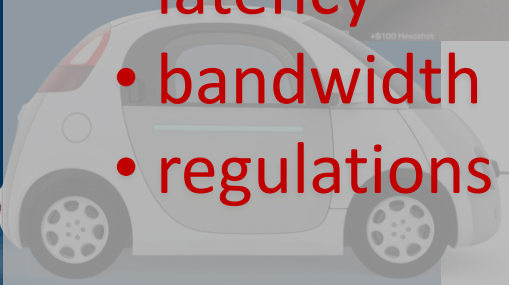


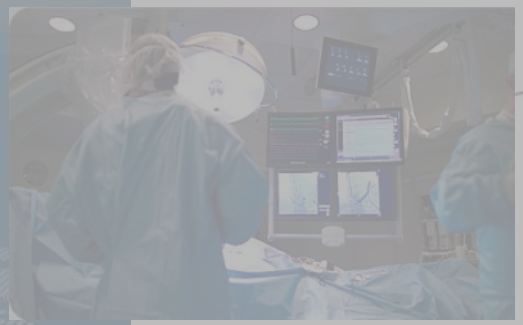
Bicsi

The ‘centralised cloud’ was conceived for certain applications



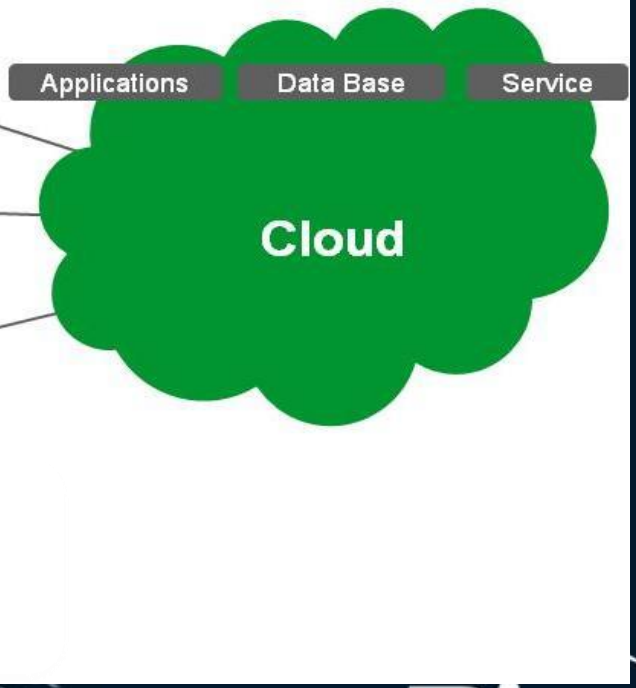
But it didn't anticipate…





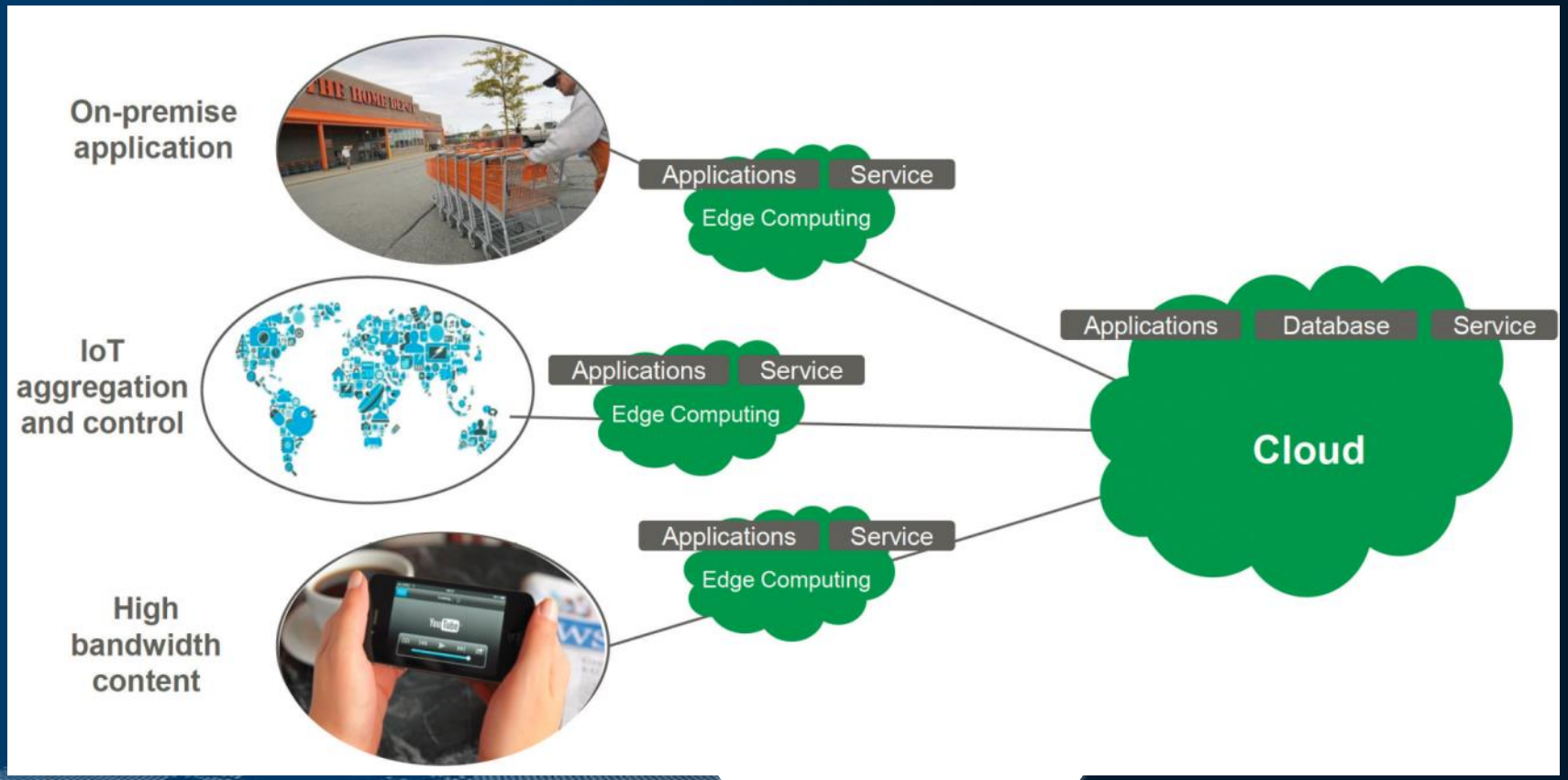
<u>limitations with:</u>

- latency
- bandwidth
- regulations



Bicsi

Edge/Fog computing provides a ‘high performance bridge’ to the centralized cloud

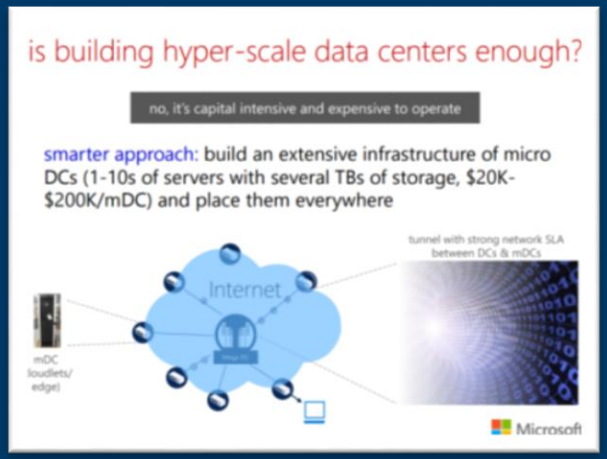


Convergence of telco and cloud services

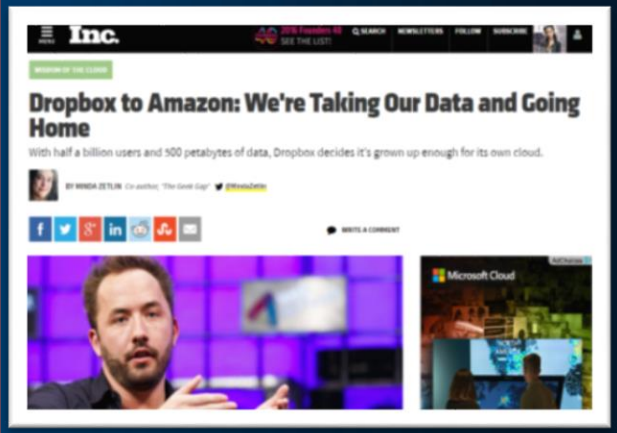
Mobile-Edge Computing (MEC)

Moving the gateway and application server closer to the radio can significantly reduce latency. Services are no longer tied to a single point-to-point IP connection, enabling the connectivity path to be freely chosen according to actual service demand.

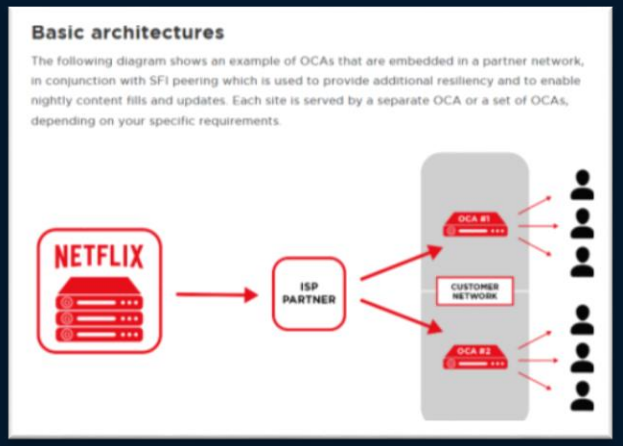
And even the big cloud providers are moving to a hybrid environment



May 15, 2015



March 15, 2016



Microsoft

Dropbox

NETFLIX
OPEN CONNECT

Bicsi

Which leads to three types of Data Centers all of which are mission critical

1

Centralized Cloud Data Center

2

Regional Data Centers

3

Localized or Micro Data Centers

Bicsi

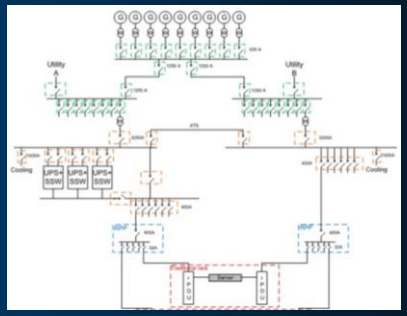
However, best practices seen in centralized and regional data centers…

Biometrics at doors

Man-traps

Security guards

Locked, organized racks



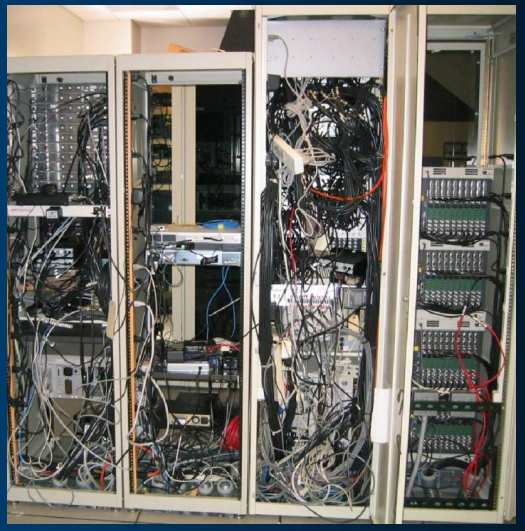
Redundancy of critical systems

Monitoring at all times

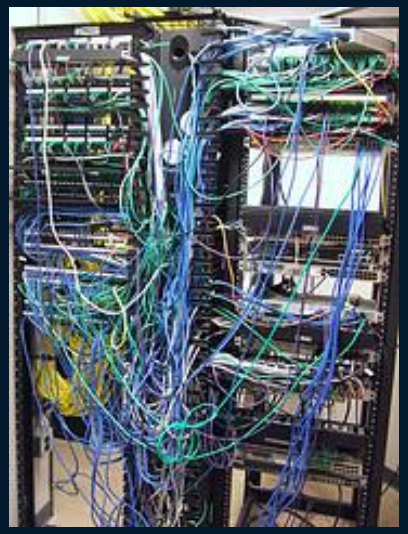
…are usually not at the localised edge…

There's not a shortage of server rooms, branch offices, and wiring closets that look like this…

Poor cable management



No monitoring



No redundancy

Unsecured racks

Cabling – nothing to do with me!!!

Moving apps to the cloud makes 'edge' sites and their connection to the cloud mission critical

The trends and changes we are faced with…

Millennials are coming and they have different expectations

Nature of computing is leading to a very complex hybrid environment

Or worse still…. What if my teenagers couldn't access…

We will evolve from 'access is available' to 'access must meet users expectations'….and expectations will have to meet Pokemon GO….

Bicsi

Hyperconverged cloud architectures also reduce what used to be a 1 MW on premise data center to potentially just a few racks

or even….

The resiliency and operation of "what's left" should be treated the same as the 1MW data center

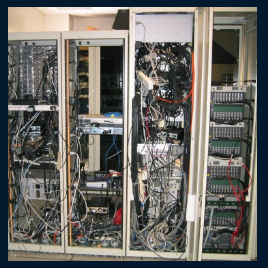
Bicsi

Availability levels typically seen in hybrid architecture today

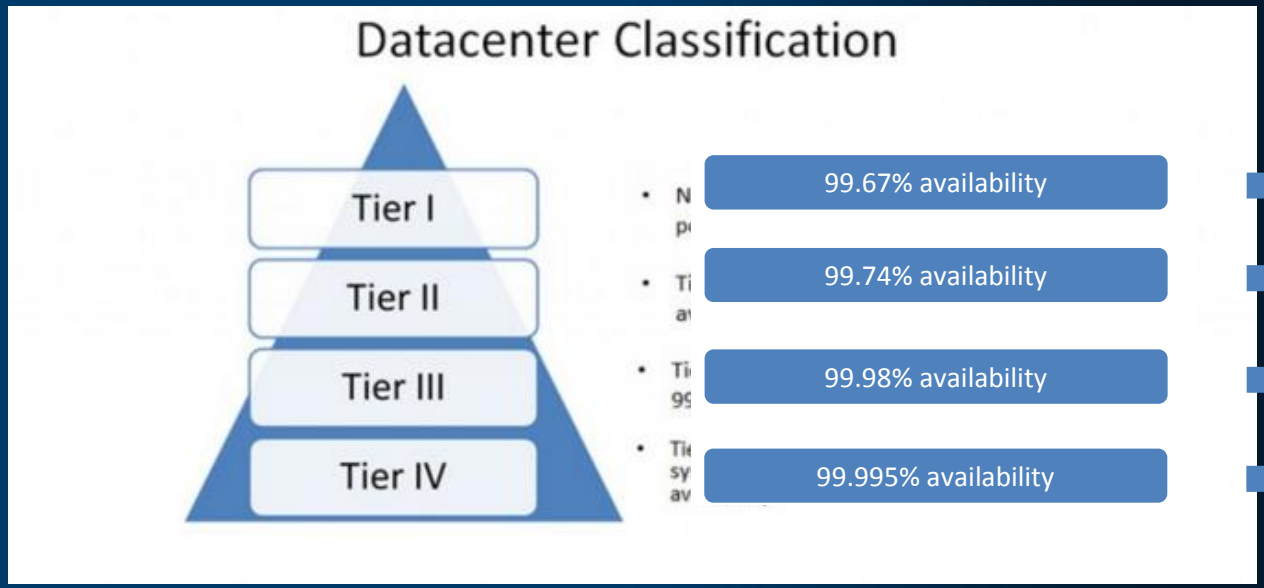
Tier 3+ Cloud data center

Tier 3 Regional data centers

Tier 1 Localized micro data centers



Current thinking on availability focuses on individual sites



When you look at downtime, the business impact becomes more apparent

- 99.67% availability → 28.8 hours of downtime
- 99.74% availability → 22.7 hours of downtime
- 99.98% availability → 1.6 hours of downtime
- 99.995% availability → 25 minutes of downtime

Source: http://www.nexdatacenter.com/blog/data-center/types-and-tiers-of-data-centers/

Our perception of “failure” is inadequate and needs to evolve

- **Current paradigm**

•Failure is a disruption to any IT equipment within a single data center

- Focused on the centralized data center
- Failure of IT rack meant a failure
- Doesn’t comprehend branch/remote sites

- **New paradigm**

•Failure comprehends user interruption, including loss of connectivity at localized / micro data centers

- Focuses on the system performance
- Considers employees at localised sites
- Considers functions at localised sites

Availability of dependent systems

If my focus is the availability of only the centralized Tier 3 data center...

Centralized data center

Availability = **99.98%**

Downtime = **1.6 hours/year**

But, if I take the viewpoint of the employees in an edge data center…

Centralized data center

Edge data center

$Availability_{system} = Availability_1 * Availability_2$

Tier 3 Cloud Data Center Availability = 99.98%

Tier 1 Edge Data Center Availability = 99.67%

Availability = 99.98% x 99.67% = **99.65%**

Downtime = **30.7 hours/year**

And add in the number of people impacted…

Data Center Availability						
Description	**Availability**	**Downtime (hrs)**	**# Sites**	**# people/site**	**Total people impacted**	**People-hours of downtime/yr**
Tier 1 edge data centers	99.67%	28.82	10	100	1,000	28,820
Tier 3 central datacenter	99.98%	1.58	1	0	1,000	1,580
					Total people-hours of downtime/yr	30,400
					Availability	99.65%

Availability of the edge dominates the equation

Business function also matters

- Some edge sites are business critical…

And some are not!

An effective metric for the hybrid cloud architecture includes weighting by employee count AND business function

A scorecard can give you the full picture…

Data Center Scorecard					
Site Name	Availability	Annual Downtime (hours)	Severity of Effects of Failure (1-5)*	Score (weighted for criticality)	Site impact on Score
1	99.98%	1.752	2	3.5	0.4%
2	99.20%	70.08	4	280.3	30.0%
3	99.60%	35.04	1	35.0	3.7%
4	98.60%	122.64	5	613.2	**65.5%**
5	99.98%	1.752	2	3.5	0.4%
			Overall criticality score:	*935.6*	

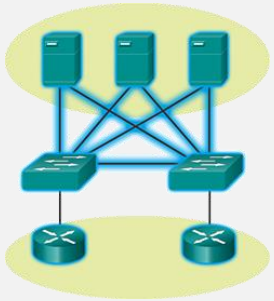
- Include all data centers in hybrid architecture
- Weight the effect of the failure of each site by number of employees AND importance of function performed at the site
- Focus improvements first on the sites with greatest impact on scorecard

Business critical edge sites should be designed to achieve high availability

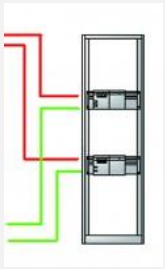
Bicsi

We need to rethink robust architectures for the localised data center – focus on security, redundancy, and management

Dual network connectivity



Redundancy in critical components of power/cooling



Secure, safe environment

Bicsi

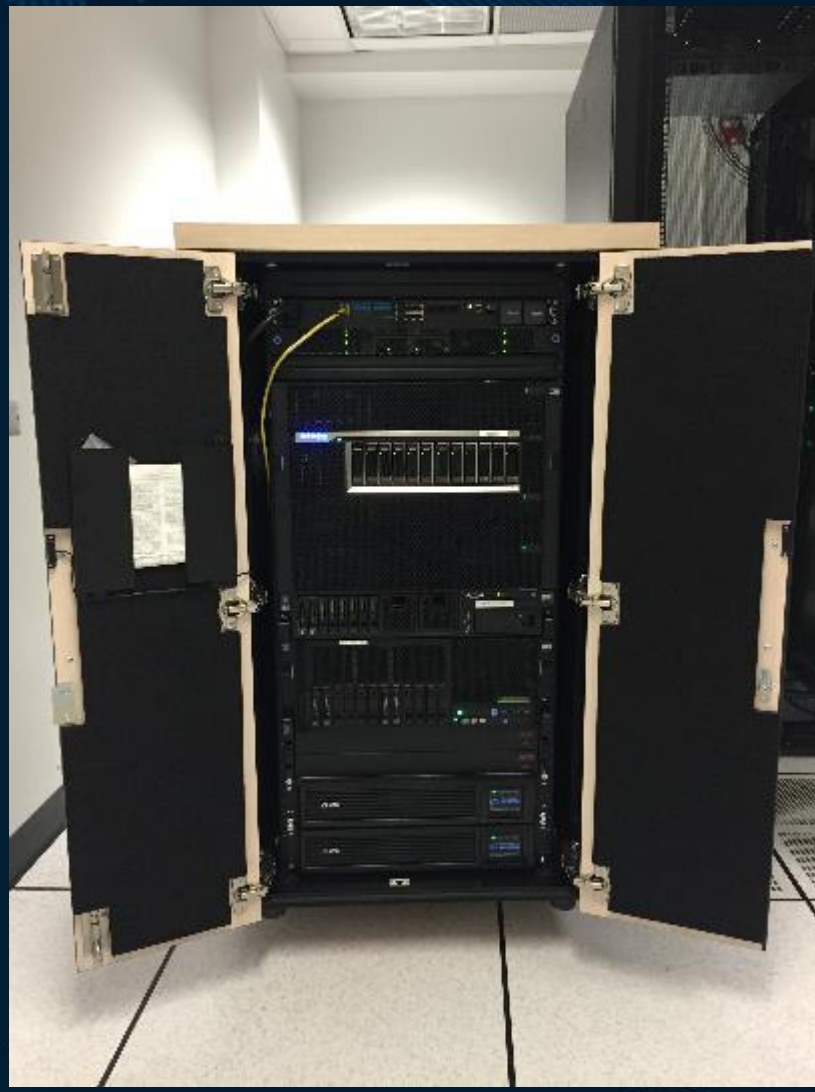
Schneider Electric's micro data center definition

A micro data center is a self contained, secure computing environment that includes all the storage, processing and networking required to run the necessary applications. It ships in single enclosure and includes all necessary power, cooling, security, and associated management tools (DCIM). Micro data centers can be assembled and tested in a factory environment.

Example of a micro data center: IBM Flash Storage equipment in the SmartBunker CX

The configuration consists of:

- IBM A9000
- Power8 Server
- X3650 server
- Network switch
- Brocade Switch
- UPS unit
- Netbotz
- RM PDUs
- DCE
- Wireless temperature sensors



Considerations to “harden” your edge infrastructure

- Physical Security
- Monitoring (DCIM), operational practices, remote monitoring
- Redundant power and cooling
- Concurrent maintainability
- Dual network connectivity

Bicsi

Physical Security

- Challenges
  - Micro data centers are often placed within a highly accessible room (i.e. shared office space)
  - No dedicated space, so open racks are unsecured
- Recommended steps
  - Move equipment to locked room or locked enclosure(s).
  - Ensure biometric or other access control
  - For harsh environments, secure equipment in enclosure that protects against fire, flood, humidity, vandalism & EMF effects
  - Deploy security & environmental monitoring 24x7, video surveillance

Data center management

- Challenges
  - No standard management & operations protocol from site to site
  - Many sites to manage can be costly
  - Availability depends on shared facility resources
- Recommended steps
  - Take inventory of existing management methods and systems
  - Consolidate to centralized monitoring platform of all assets across sites
  - Deploy remote monitoring when resources are constrained

Power & cooling

- Challenges
  - Single points of failure
  - Over-heated closets/rooms
  - Shared facility infrastructure

- Recommended steps
  - Measure temperature and humidity to understand level of cooling needed (i.e. passive airflow, active airflow or dedicated cooling)
  - Consider redundant power paths for concurrent maintainability
  - Ensure critical circuits are on emergency generator

Network / Internet connectivity

- Challenges
  - Single internet service provider represents single point of failure
  - Rats nest of cables breeds human error
- Recommended steps
  - Consider adding a second network provider
  - Organize network cables with network management cable devices (raceways, routing systems, ties, etc.)
  - IIMS
  - Label, and color-code network links to avoid human error

Key take-aways

1. Connectivity on the edge is more critical with cloud-based data center architectures
2. The resiliency and operation of remaining “edge” equipment in a hybrid architecture should be treated the same as the traditional enterprise data center
3. A more comprehensive availability metric is needed focused on measurement of connectivity to the cloud in this distributed environment