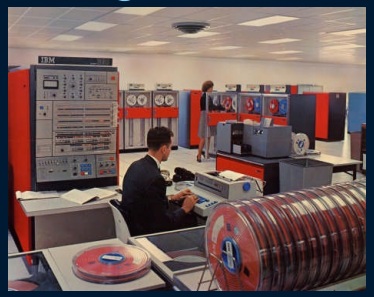
May 20, 2015

At Scale Enterprise Computing

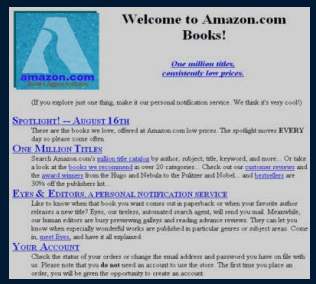
Dr. Chung-Sheng Li
IEEE Fellow &
IBM Academy of Technology Leadership Team
Director, Commercial Systems
IBM Research Division

Digital Transformation: The journey coincides with the evolution of computing



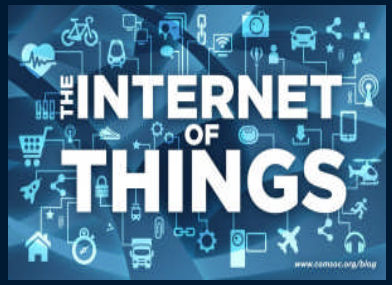
Digitization of transaction

Digitization of enterprise



Digitization of shopping

Digitization of interactions

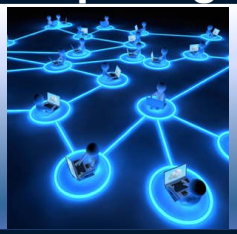


Digitization of environment & world

Mainframe Computing

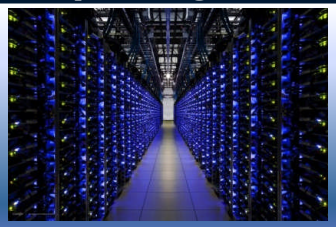
- Infrastructure: SMP,, channel, ESCON, Block Storage
- Middleware: TPF, IMS, CICS
- Application: Saber, SAP

Distributed Computing



- Infrastructure: client –server, scale out, TCP/IP, File storage
- Middleware: App Server, RDBMS, MQ Broker, SOA, ESB, BPM
- Application: 3-tier App

At Scale Computing



- Infrastructure: Warehouse Scale Computing, Flat network, Object Store
- Middleware: MapReduce, NoSQL, NewSQL, Micro Services, ZMQ,
- Application: FB, Google Search, Dropbox

Successful computing paradigms emerged from at scale industry transformation, and differentiated through full stack optimization that includes applications, middleware, compute, storage, networking and programming models.

Digitization: the Primary Catalyst for Industry Transformation and Refactoring

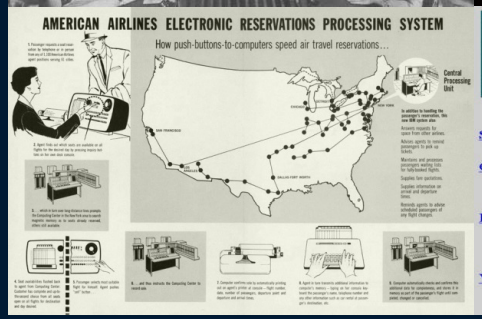
Manual airline reservation (Pan Am)

Traditional Bookstore

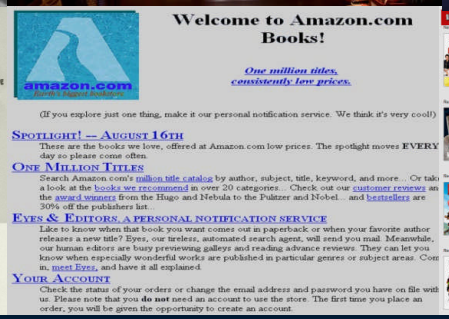
Video Rental

Texting

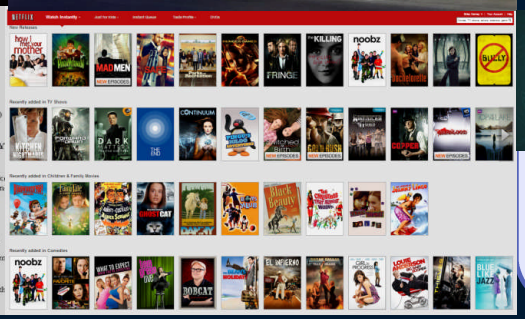
2010:
All major cellular carrier
$.20/SMS message



Fully automated airline reservation (Sabre)



E-commerce Bookstore



Video Streaming

2015: 30B message/day
700M monthly active user
$0.99/year/user

Case Study - Digital Transformation in Travel Industry: Then and Now

1950's


8 PAGES THIS WEEK

Travel Trade GAZETTE

THE NATIONAL WEEKLY OF THE TRAVEL & HOLIDAY SERVICES

THIS WILL BE BRITAIN'S GREATEST TOURIST YEAR

First bookings show trend to air travel

DEATH OF KLM PRESIDENT

HUNTING-CLAN EXTEND NORTHERN NETWORK

'Below-tourist' fares to Scandinavia

LONDON-MADRID AIR LINK

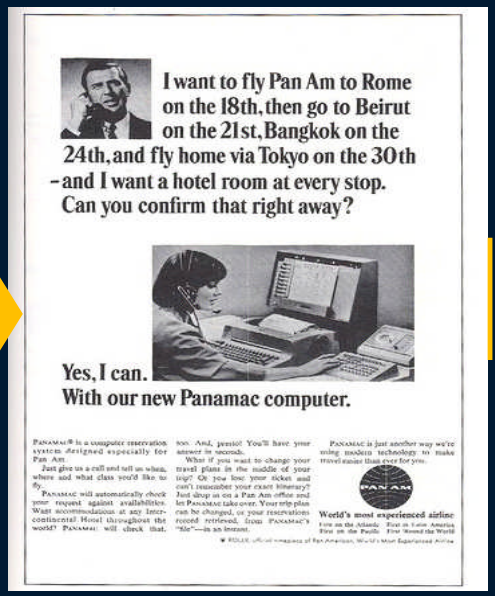
QUEEN HONOURS TRAVEL PERSONALITIES

A NEW ERA IN TOURIST TRAVEL

Luxury at Tourist Rates

GREEK LINE

1960's


I want to fly Pan Am to Rome on the 18th, then go to Beirut on the 21st, Bangkok on the 24th, and fly home via Tokyo on the 30th – and I want a hotel room at every stop. Can you confirm that right away?

Yes, I can. With our new Panamac computer.

PANAMAC® is a computer reservation system designed especially for Pan Am.

Just give us a call and tell us when, where and what class you'd like to fly.

PANAMAC will automatically check your request against availabilities. Want accommodations at any Intercontinental Hotel throughout the world? PANAMAC will check that, too. And, presto! You'll have your answer in seconds.

What if you want to change your travel plans in the middle of your trip? Or you lose your ticket and can't remember your exact itinerary? Just drop in on a Pan Am office and let PANAMAC take over. Your trip plan can be changed, or your reservations record retrieved, from PANAMAC's "file"—in an instant.

PANAMAC is just another way we're using modern technology to make travel easier than ever for you.

PAN AM

World's most experienced airline

IBM dominated the 1st phase of digital transformation of travel industry during 1960's with substantial productivity gain introduced by Sabre, Panamatic and Deltamatic

1990's

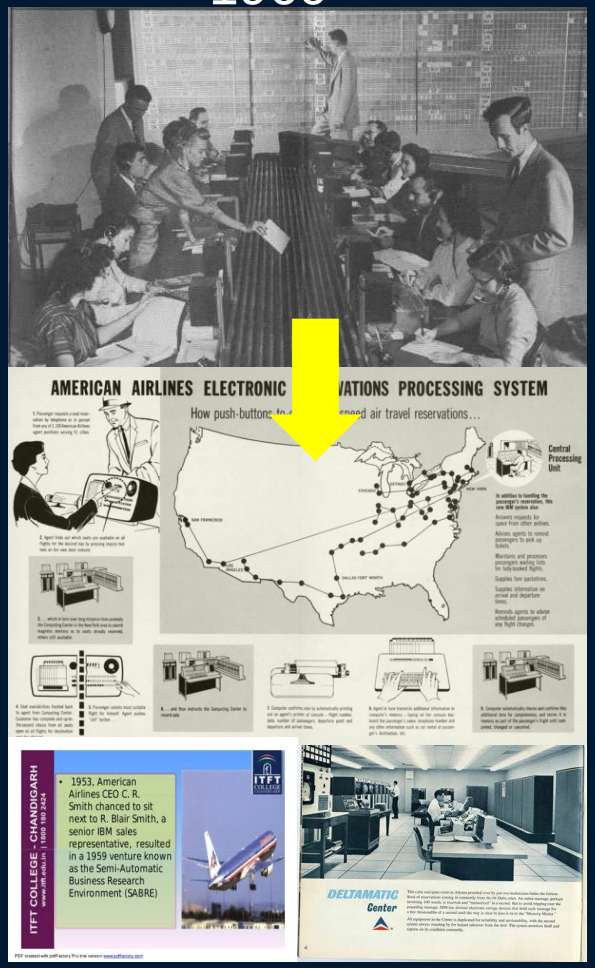
Internet democratized the shopping of travel packages and fares

2010's

Travel industry continued to be refactored by new players. IBM has little foot print left in the travel industry IT as of 2015

Case Study: Digital Transformation in Travel Industry – Key Events

1960



1969

1979

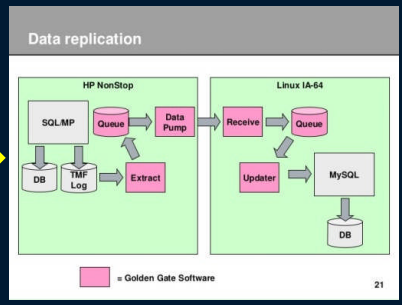
TPF

2000

Sabre Holdings

MySQL

2004



1996

ita Software™

2001

ORBITZ

2010

Google ita Software™

2015

What is at scale computing
IT@scale implies unprecedented scale in some of these dimensions:

- lines of code
- amount of data & metadata stored, accessed, manipulated, curated, and refined
- number of connections and interdependencies
- number of hardware elements
- number of computational elements
- number of system purposes and user perception of these purposes
- number of routine processes, interactions, and “emergent behaviors”
- number of (overlapping) policy domains and enforceable mechanisms
- number of people involved in some way

Born on the Web services

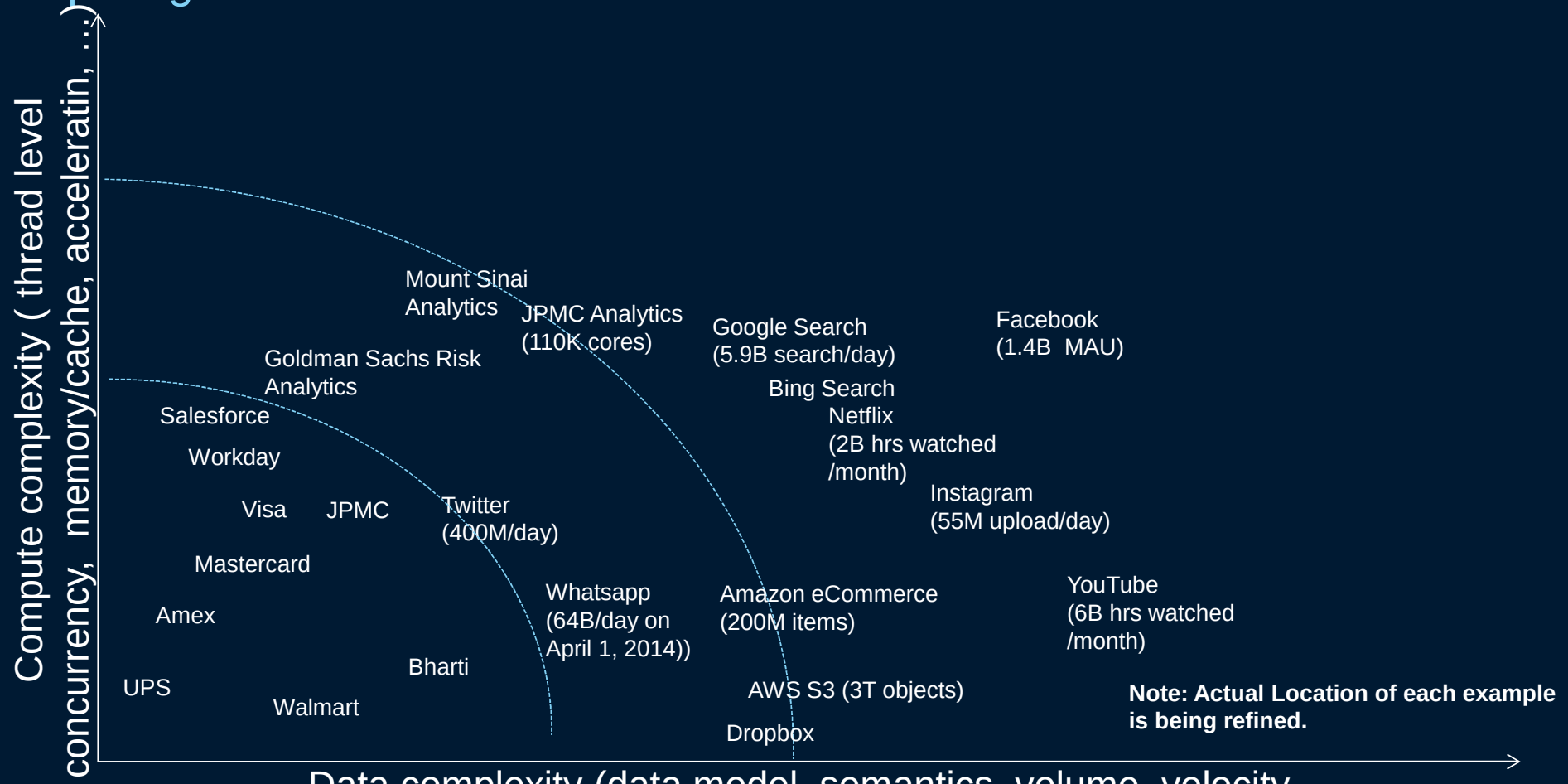
- AWS S3: 2T objects as of 2013
- AWS datacenter: ~2.5M nodes as of 2013
- Amazon: 36.M items sold on Cyber Monday 2013
- Facebook: 1.4B MAU (monthly active users) as of end of 2013
- Google: 5.9B search/day, 2.2T search/year (2013)
- Netflix: contributed to 30% of US Internet traffic during peak hrs, 2B hrs video watched
- Twitter: 400M tweets/day (Sept. 2013)
- Instagram: 55M photos/day uploaded
- MPRPG (many are hosted on SL): World of Warcraft (8.3M users/246 servers), MapleStory (5M users/96 server), Star Wars (1M users/214 servers);

Traditional enterprise

- Walmart: 100M customer/week
- UPS: 15.8M shipment/day (March, 2014)
- VISA+MC: 18B transactions/year

http://www.statisticbrain.com/google-searches/

Data and compute complexity of grand challenges define a unique category of at scale computing stack

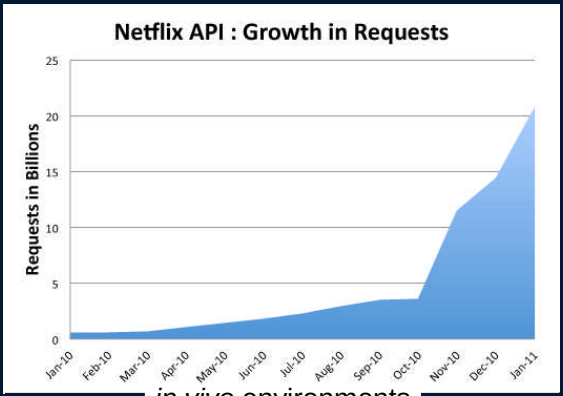


Current phase of digital transformation drove the adoption of at scale applications and services, resulting in new breed of middleware & infrastructure and deep stack optimization.

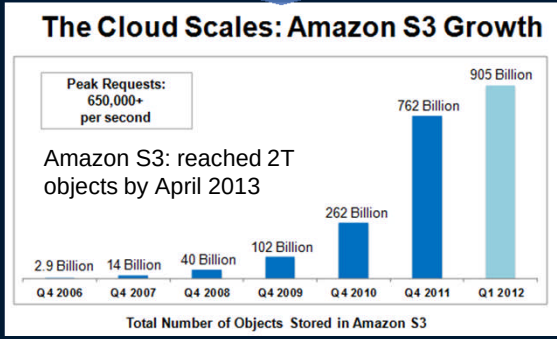
		Mainframe era	Distributed Computing era	At Scale era
why	Application	SAP, Saber	Client/Server, 3 tier architecture	Salesforce, Amazon commerce, Google search, Google map, Facebook, Netflix
why	Data	transactions	Web content + transactions	Social (FB), Streaming (Netflix, YouTube)
How	Integration	CICS	SOA, ESB, BPM, Workflow engine	Micro Service Architecture, Node.js, Apogee
How	OLTP/OLAP	IMS on Parallel Sysplex	Data partitioning and sharing (DB2 eee, Oracle RAC)	NewSQL (e.g. Google Spanner)
How	Content Store & NoSQL		Enterprise content management	NoSQL (document store, key value store, graph store, wide column store)
How	Big Data & Analytics	Batch	SAS, SPSS	MapReduce, Stream, Graph analytics
How	Messaging	MQ	MQ Broker	RabbitMQ, ZMQ, ActiveMQ
How	Virtualization/Isolation	VM, PR/SM	Xen, kvm, VMware	Software Defined Environments, Container (e.g. Docker)
How	Backup & Disaster Recovery	Rare failure & often handled locally	Rare failure & often handled locally	Fail in place, handled globally
How	Compute	SMP	Client/Server, peer-to-peer, Clustering	Warehouse scale computing
How	Network	Channel/ESCON	Ethernet/SAN	Flat Network (Spine-Leaf, Spline)
How	Storage	Block (VSAM)	File	Object Store

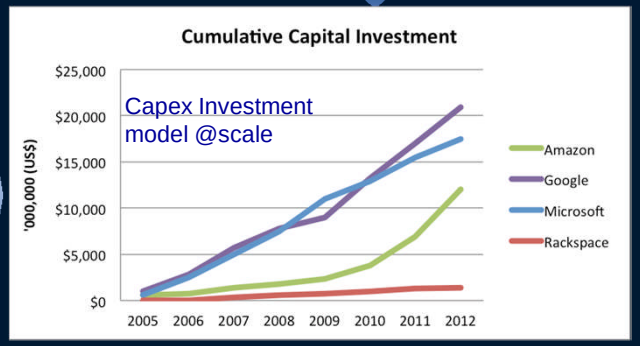
Scale is becoming the table stake & currency for CSPs for delivering computing as a service

- **Winner takes all**: when apps @scale reached, higher barrier to entry is established even for adjacent space (e.g. PayPal)
- CAPEX investment model @scale allow much more head room for pricing power, as evident by the recent pricing war among GOOG, AMZN, and MSFT



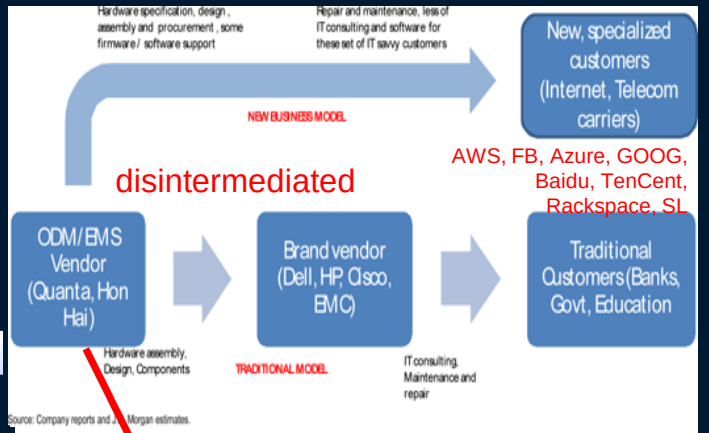
in vivo environments @scale drive fast infrastructure growth





Infrastructure @scale drive fast organic investment growth

Scale becomes the most important currency for @scale CSP to disintermediate traditional brand vendors



Vendor	4Q13 Revenue	4Q13 Market Share	4Q12 Revenue	4Q12 Market Share	4Q13/4Q12 Revenue Growth
1. HP*	$3,832	26.9%	$3,626	24.3%	5.7%
1. IBM*	$3,810	26.8%	$5,344	35.9%	-28.5%
3. Dell	$2,058	14.5%	$2,109	14.2%	-2.4%
4. Cisco*	$646	4.5%	$480	3.2%	34.5%
4. Oracle*	$589	4.1%	$602	4.0%	-2.2%
ODM Direct	$907	6.4%	$616	4.1%	47.3%
Others	$2,389	16.8%	$2,116	14.2%	12.9%
All Vendors	**$14,242**	**100%**	**$14,893**	**100%**	**-4.4%**

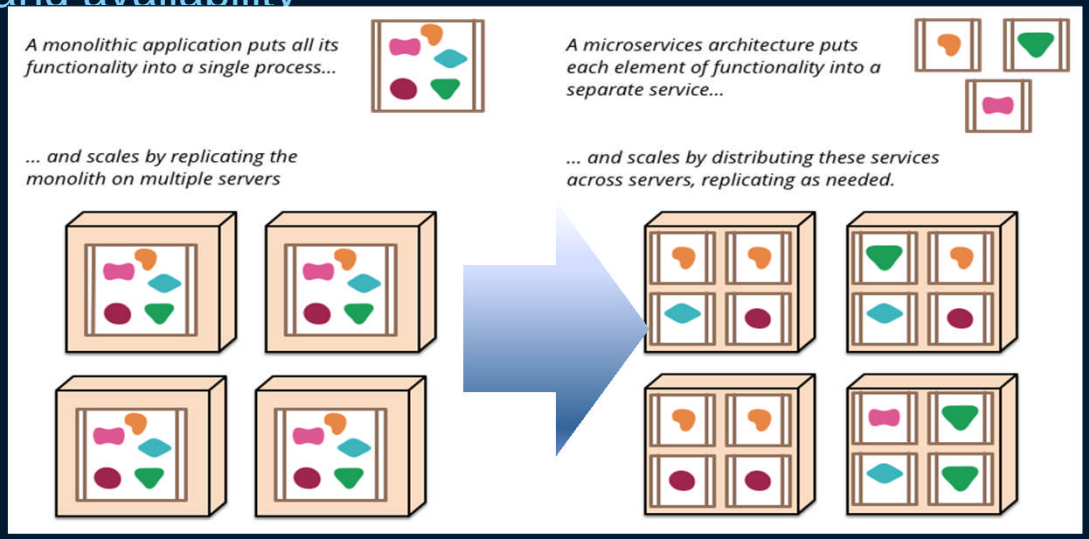
Source: IDC

Current phase of digital transformation drove the adoption of at scale applications and services, resulting in new breed of middleware & infrastructure and deep stack optimization.

	Mainframe era	Distributed Computing era	At Scale era
why: Application	SAP, Saber	Client/Server, 3 tier architecture	Salesforce, Amazon commerce, Google search, Google map, Facebook, Netflix
why: Data	transactions	Web content + transactions	Social (FB), Streaming (Netflix, YouTube)
How: Integration	CICS	SOA, ESB, BPM, Workflow engine	Micro Service Architecture, Node.js, Apogee
How: OLTP/OLAP	IMS on Parallel Sysplex	Data partitioning and sharing (DB2 eee, Oracle RAC)	NewSQL (e.g. Google Spanner)
How: Content Store & NoSQL		Enterprise content management	NoSQL (document store, key value store, graph store, wide column store)
How: Big Data & Analytics	Batch	SAS, SPSS	MapReduce, Stream, Graph analytics
How: Messaging	MQ	MQ Broker	RabbitMQ, ZMQ, ActiveMQ
How: Virtualization/Isolation	VM, PR/SM	Xen, kvm, VMware	Software Defined Environments, Container (e.g. Docker)
How: Backup & Disaster Recovery	Rare failure & often handled locally	Rare failure & often handled locally	Fail in place, handled globally
How: Compute	SMP	Client/Server, peer-to-peer, Clustering	Warehouse scale computing
How: Network	Channel/ESCON	Ethernet/SAN	Flat Network (Spine-Leaf, Spline)
How: Storage	Block (VSAM)	File	Object Store

At Scale Computing drives Deep Stack Optimization:

Micro Services (Application disaggregation) enable substantial improvement on flexibility, agility, and availability



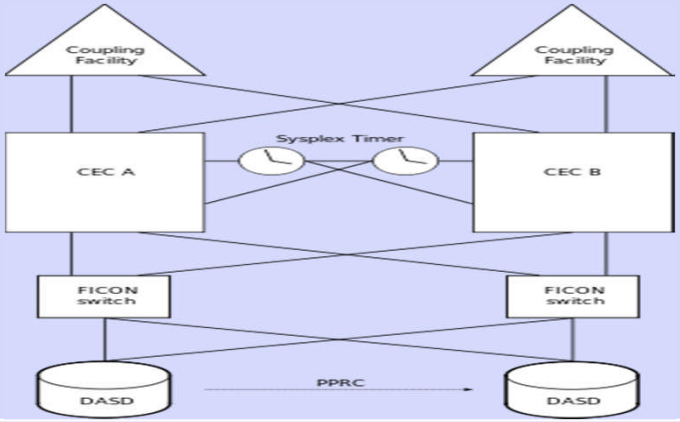
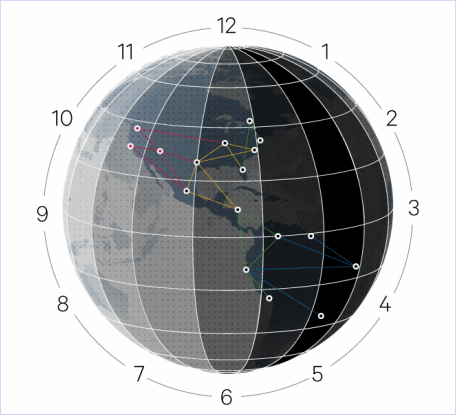
Large scale deployment demonstrated at Netflix now is driving many enterprises experimenting with both SoE and SoR deployments

NETFLIX

Downtime Per Year

	Monolithic Apps On-Premise	Monolithic Apps in the Cloud	Micro-Services in the Cloud
Downtime	3d15h36m	3d15h36m	5m
Software	99% Uptime Applications	99% Uptime Applications	99.999% Uptime Applications
Hardware	99.999% Uptime Infrastructure (specialized hardware)	99% Uptime Infrastructure (commodity hardware)	99% Uptime Infrastructure (commodity hardware)

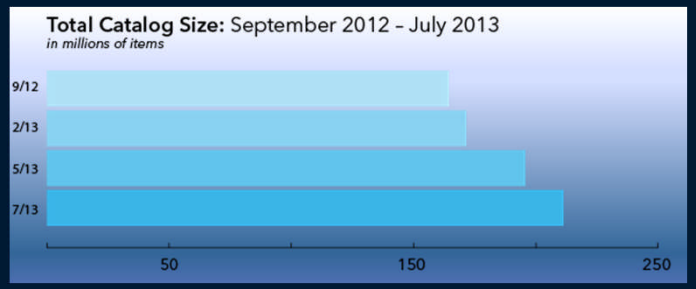
At Scale Computing drives Deep Stack Optimization:
Google Spanner vs. Parallel Sysplex

IBM Parallel Sysplex	Google Spanner
	
• Introduced in 1990 by IBM, allows up to **32 servers, each of which can have up to 64 LPARs** to cooperate with each other. • A common time source to synchronize all member systems' clocks. This can involve either a Sysplex timer (Model 9037), or the **Server Time Protocol** (STP) • Global Resource Serialization (GRS), which allows multiple systems to access the same resources concurrently, serializing where necessary to ensure exclusive access • Based on synchronous data mirroring technology that can be used on mainframes 200 km (120 miles) apart.	• **Scalable, globally-distributed, synchronously-replicated database:** Designed to scale up to millions of machines across hundreds of datacenters and trillions of database rows • **Replication used to achieve global availability and geographic locality:** Aiming for 99% HA and 50 ms latency with datacenters currently up to 100 ms (~20,000 km) apart, Is fault tolerant to large scale outage; Leveraging Alcatel Lucent optical products to interconnect datacenters. • **Currently operational and supports Google's advertising business F1.** • Leverages hardware features like GPS and Atomic Clocks

At Scale Computing drives Deep Stack Optimization:

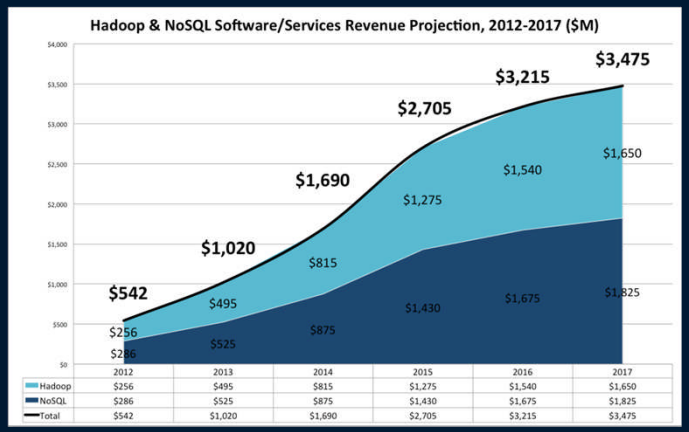
Enterprises embrace NoSQL and Object Store due to flexibility and agility resulting from dynamic schema (vs. fixed schema required by traditional SQL DB)

NoSQL environment was originally driven by Google BigTable and Amazon commerce catalog

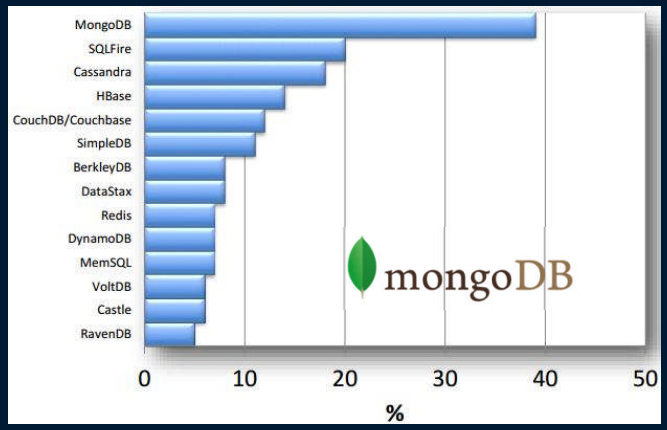


- Amazon commerce catalog reached 200M items as of July 2013
- 36M items were sold on Amazon during cyber Monday, 2013
- Amazon catalog is adding 175K items a day.
- Average # of items in a supermarket = 40K

Hadoop + NoSQL are expected to reach 3.4B by 2017



1000+ enterprise customers including Cisco, EA, eBay, Ericsson, Forbes, Intuit, LexisNexis, SAP and Telefonica. Among the Fortune 500 and Global 500, MongoDB already serves top companies from FSS, electronics, M&E, Retail, Telcos, Tech, and Healthcare.



At Scale Environment leads to the emerging focus on *Software Defined Environments*

		Mainframe era	Distributed Computing era	At Scale era
why	Application	SAP, Saber	Client/Server, 3 tier architecture	Salesforce, Amazon commerce, Google search, Google map, Facebook, Netflix
why	Data	transactions	Web content + transactions	Social (FB), Streaming (Netflix, YouTube)
How	Integration	CICS	SOA, ESB, BPM, Workflow engine	Micro Service Architecture, Node.js, Apogee
How	OLTP/OLAP	IMS on Parallel Sysplex	Data partitioning and sharing (DB2 eee, Oracle RAC)	NewSQL (e.g. Google Spanner)
How	Content Store & NoSQL		Enterprise content management	NoSQL (document store, key value store, graph store, wide column store)
How	Big Data & Analytics	Batch	SAS, SPSS	MapReduce, Stream, Graph analytics
How	Messaging	MQ	MQ Broker	RabbitMQ, ZMQ, ActiveMQ
How	Virtualization/Isolation	VM, PR/SM	Xen, kvm, VMware	Software Defined Environments, Container (e.g. Docker)
How	Backup & Disaster Recovery	Rare failure & often handled locally	Rare failure & often handled locally	Fail in place, handled globally
How	Compute	SMP	Client/Server, peer-to-peer, Clustering	Warehouse scale computing
How	Network	Channel/ESCON	Ethernet/SAN	Flat Network (Spine-Leaf, Spline)
How	Storage	Block (VSAM)	File	Object Store

Software Defined Computing enables agile and flexible composition of systems through *Software Defined Environments*

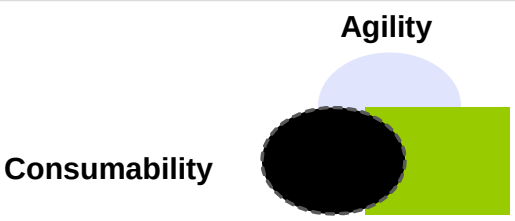
Workload Abstraction | Resource Abstraction | Mapping to Resource | Continuous Optimization

Workload Abstraction
Based on pattern and
functional and non-functional requirements

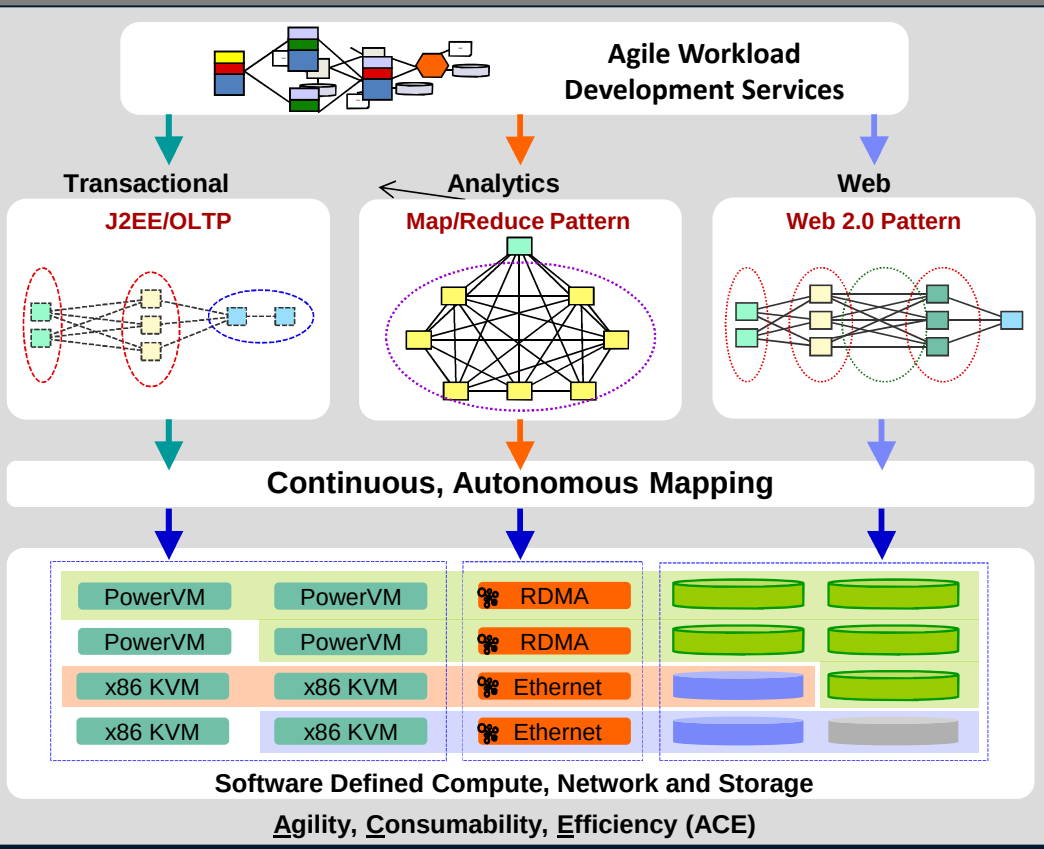
Resource Abstraction
Semantically rich abstractions of heterogeneous resource capabilities & system components

Mapping to resource
Map requirements to potential system architectures

Continuous Optimization
Autonomously construct available system architecture to optimize workload outcome



Software Define Environments

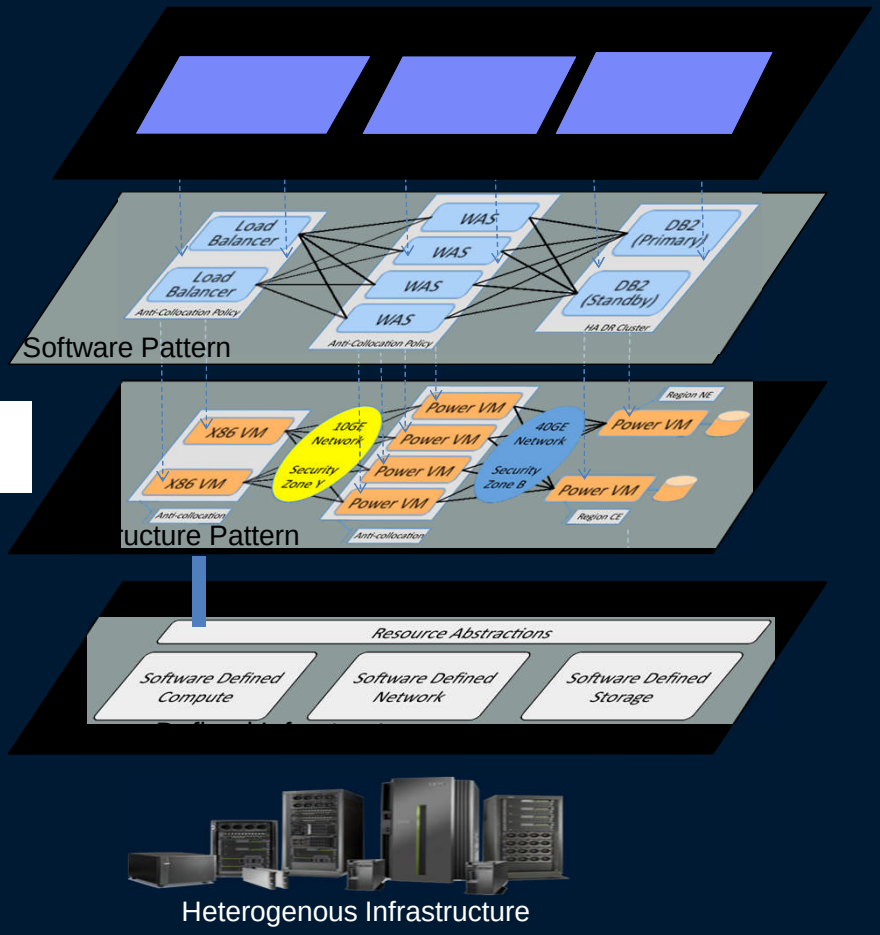


Workload-Aware Orchestration and Optimization

Best practices are captured as Software Defined Environment patterns

Workload Abstraction | **Resource abstraction** | **Mapping to resource** | **Continuous Optimization**

Simplified
Responsive
Adaptive



Heterogenous Infrastructure

Significant Degree of Automation and Virtualization

Solution Definition: Define business needs and define solution elements and building blocks

Software Pattern: Links solution to infrastructure leveraging best practices and expertise

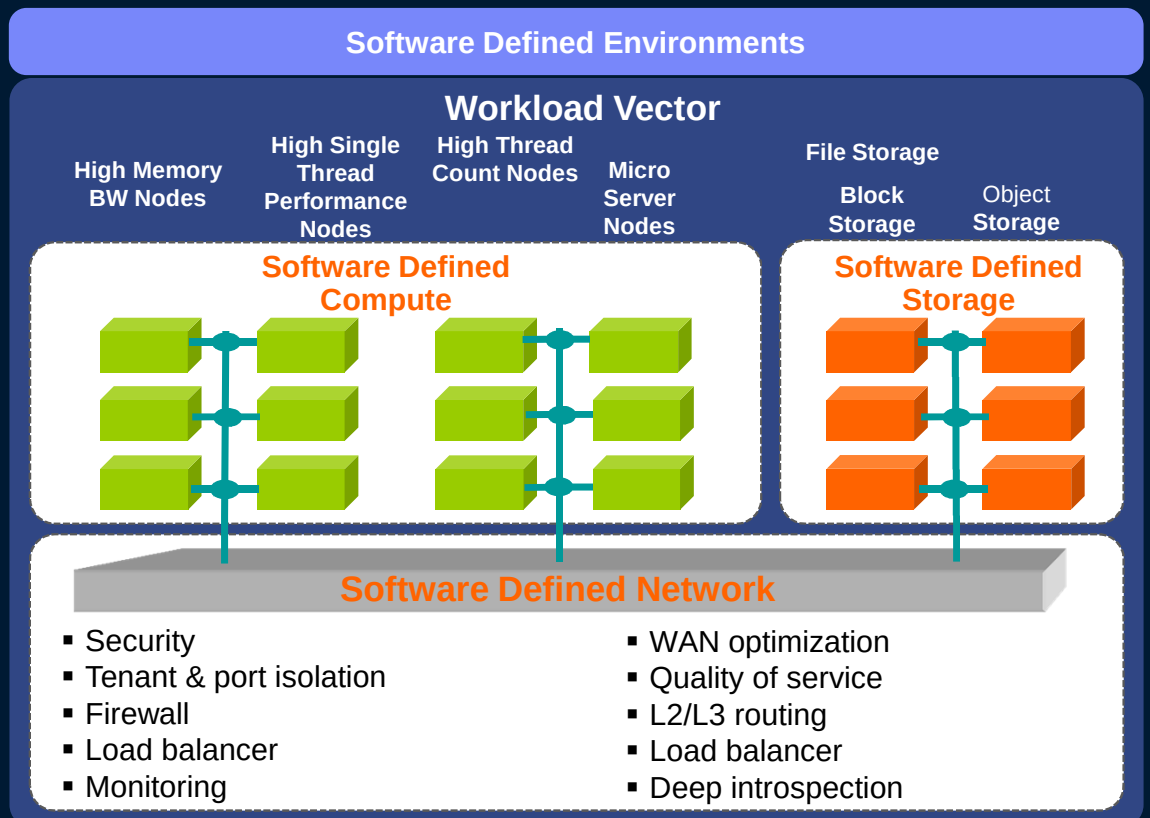
Infrastructure Pattern: Maps software pattern to optimal infrastructure based

Software Defined Infrastructure: Automatically orchestrate deployment and analytics-based optimization of the infrastructure resources

Unified control plane dynamically optimized for workload using heterogeneous resources

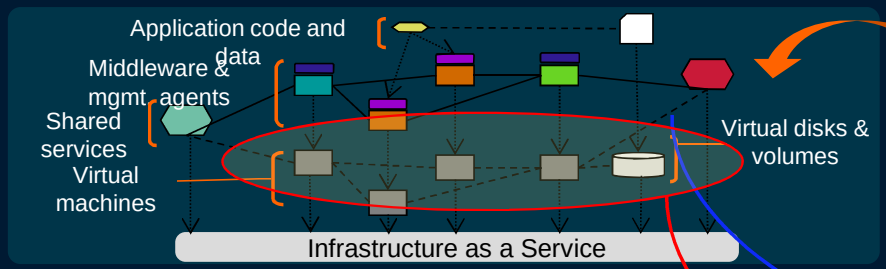
Workload Abstraction | **Resource abstraction** | Mapping to resource | Continuous Optimization

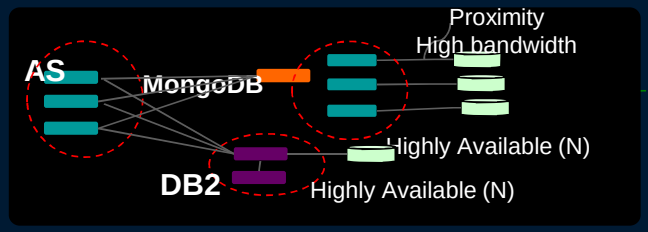
- Abstraction of heterogeneous resources requires compute & storage resources defined according to workload characteristics
- Examples of ***workload vectors***:
  - **Compute**:
    - High single thread performance: 2-socket P8 with SCM
    - High mem BW: 8-socket P8 DCM/AMC
    - Strong graphics/vector: Sandybridge with Nvidia Kepler GPU
  - **Storage:**
    - File/block/object
    - High IOPs storage
- These workload vectors are interconnected by network resources which specify connectivity, latency, bandwidth, isolation, …
- These characteristics can be discovered in advance and can be continuously revised by calibrating against established benchmarks: tpcC, SAP2D, specWeb, specJbb

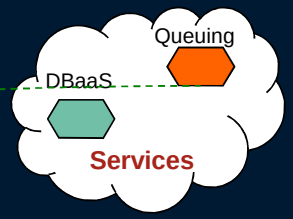


Abstracted workload mapped to workload vectors. Deployment & operation is managed by a proactive “governing system”

Workload Abstraction | **Resource abstraction** | **Mapping to resource** | **Continuous Optimization**





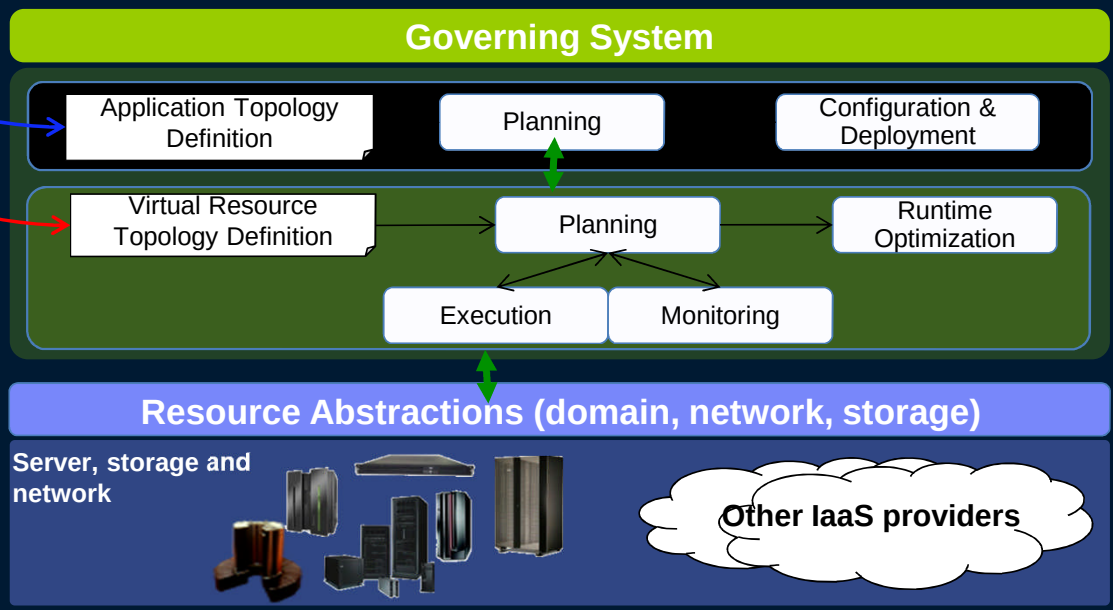


Governing System

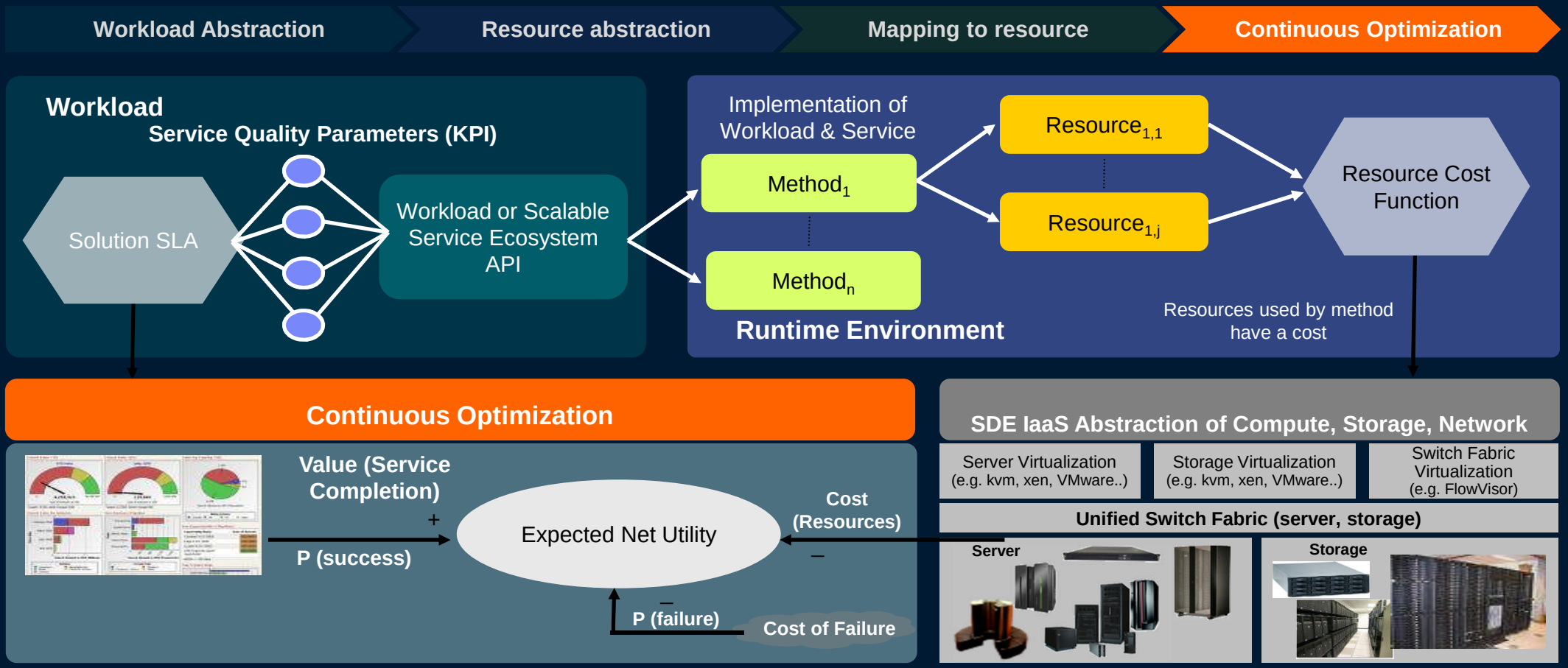
Workload-aware full life-cycle management of workloads: deployment, configuration, update, state tracking

Continuous optimized mapping of workload components to resource vectors, combining workload objectives and system objectives

Rich API enables cooperation with middleware and services.



Unified Control Planes within Software Defined Environments **continuously** evaluate & select methods which **optimize** expected net utility for a given service at that moment



Source: Darpa Mission Oriented Resilient Cloud

At scale applications and services introduce new breed of infrastructure and deep stack optimization.

	Mainframe era	Distributed Computing era	At Scale era
why			
Application	SAP, Saber	Client/Server, 3 tier architecture	Salesforce, Amazon commerce, Google search, Google map, Facebook, Netflix
Data	transactions	Web content + transactions	Social (FB), Streaming (Netflix, YouTube)
How			
Integration	CICS	SOA, ESB, BPM, Workflow engine	Micro Service Architecture, Node.js, Apogee
OLTP/OLAP	IMS on Parallel Sysplex	Data partitioning and sharing (DB2 eee, Oracle RAC)	NewSQL (e.g. Google Spanner)
Content Store & NoSQL		Enterprise content management	NoSQL (document store, key value store, graph store, wide column store)
Big Data & Analytics	Batch	SAS, SPSS	MapReduce, Stream, Graph analytics
Messaging	MQ	MQ Broker	RabbitMQ, ZMQ, ActiveMQ
Virtualization/Isolation	VM, PR/SM	Xen, kvm, VMware	Software Defined Environments, Container (e.g. Docker)
Backup & Disaster Recovery	Rare failure & often handled locally	Rare failure & often handled locally	Fail in place, handled globally
Compute	SMP	Client/Server, peer-to-peer, Clustering	Warehouse scale computing
Network	Channel/ESCON	Ethernet/SAN	Flat Network (Spine-Leaf, Spline)
Storage	Block (VSAM)	File	Object Store

At Scale Computing drives Deep Stack Optimization

At scale service providers (Amazon, Google, Facebook) customized their infrastructure stack to maximize efficiency and minimize cost

Custom Servers
- Built by ODMs to AWS specifications.
- Specialized for specific workloads
- Moving hot software kernels to hardware

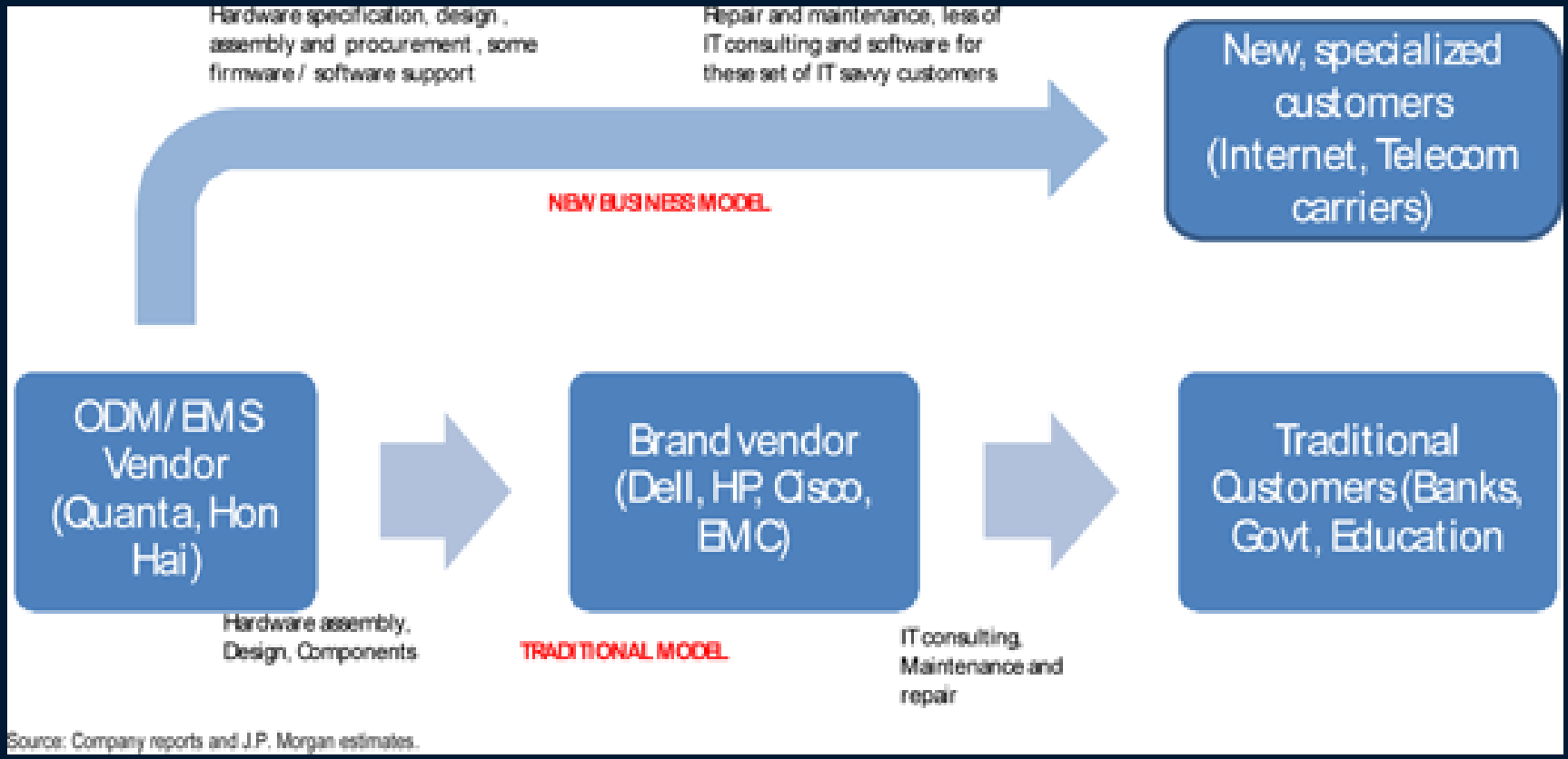
Custom Storage
- High density JBOD chassis.
- Optimized for AWS – lower power, higher density, lower cost)

Custom Network
- Custom routers and protocol stack.
- Dedicated metro-area and long-haul fiber.

Custom Power
- Negotiated power purchasing agreements
- Custom high voltage sub-stations

http://perspectives.mvdirona.com

Working directly with ODMs by fast growing large CSPs have started to refactor the traditional value chain for servers, storage, and switches



ODM Direct is now accounting for the lion's share of the server, storage, and switch/router growth (mostly attributed to the fast growth of CSPs) where the overall growth is relatively flat

WW Server Market (IDC 4Q14)

Vendor	4Q14 Revenue	4Q14 Market Share	4Q13 Revenue	4Q13 Market Share	4Q14/4Q13 Revenue Growth
1. HP	$3,894.5	26.8%	$3,831.8	26.9%	1.6%
2. Dell	$2,431.2	16.7%	$2,172.3	15.2%	11.9%
3. IBM	$1,986.4	13.7%	$3,820.2	26.8%	-48.0%
4. Lenovo	$1,105.9	7.6%	$130.4	0.9%	748.3%
5. Cisco	$769.5	5.3%	$646.1	4.5%	19.1%
ODM Direct	$1,192.1	8.2%	$907.3	6.4%	31.4%
Others	$3,156.5	21.7%	$2,756.7	19.3%	14.5%
Total	**$14,536.1**	**100%**	**$14,264.7**	**100%**	**1.9%**

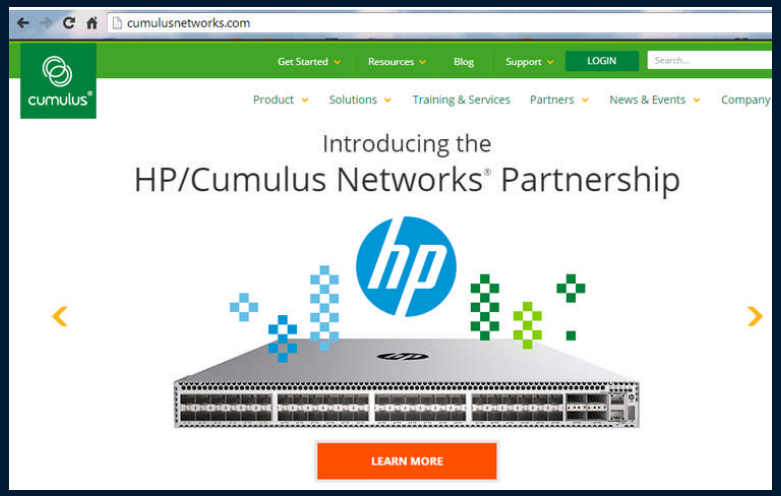
WW Storage Market (IDC 4Q14)

Vendor	4Q14 Revenue	4Q14 Market Share	4Q13 Revenue	4Q13 Market Share	4Q14/4Q13 Revenue Growth
1. EMC	$2,352	22.2%	$2,276	23.1%	3.3%
2. HP	$1,456	13.8%	$1,389	14.1%	4.8%
3. Dell*	$952	9.0%	$905	9.2%	5.2%
3. IBM*†	$951	9.0%	$1,248	12.7%	-23.8%
5. NetApp	$764	7.2%	$791	8.0%	-3.5%
ODM Direct	$1,357	12.8%	$973	9.9%	39.4%
Others	$2,740	25.9%	$2,281	23.1%	20.2%
All Vendors	**$10,571**	**100.0%**	**$9,864**	**100.0%**	**7.2%**

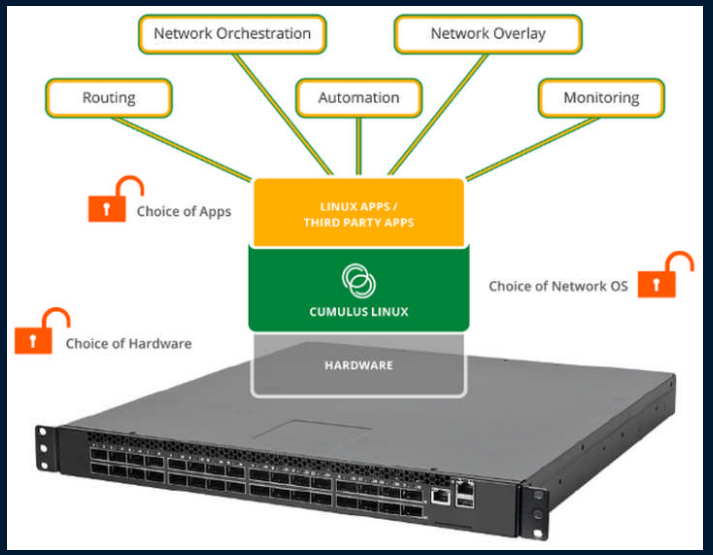
Traditional OEMs (HP, Juniper, Dell) are taking notice and starting to develop partnership with ODMs in creating the new Britebox market segment as an alternative to pure ODM/OEM

HP And Foxconn announce Cloudline for CSP during OCP summit (03/10/2015)

HP will ship ODM switches from Accton running cumulus network OS (02/23/2015)



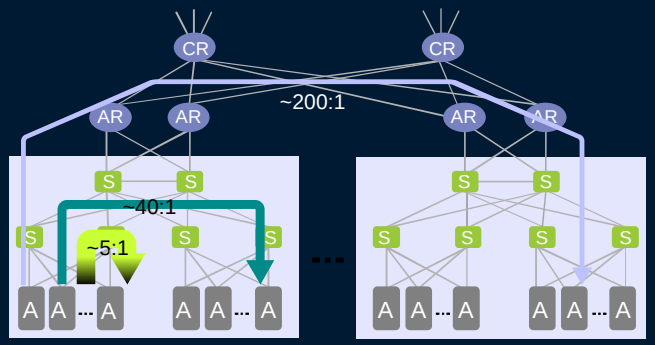
Dell started to drive Open Network Switches (with Cumulus, Bigswitch) during 2014



Juniper introduces OCX1100 (Juno on ODM white box) in Dec, 2014

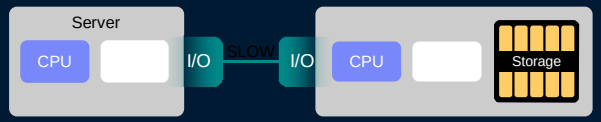
Evolution towards datacenter scale computing

Modern analytic workloads create high east-west datacenter traffic



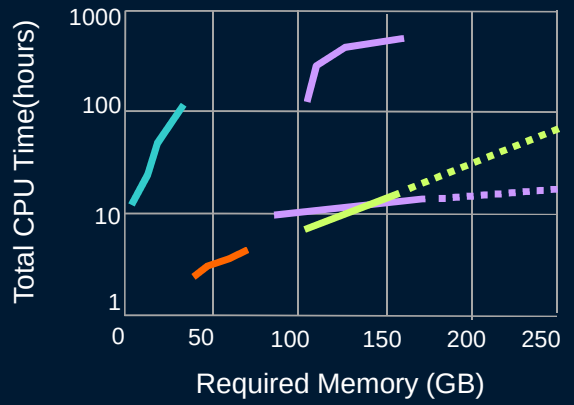
Modern analytic workloads often require large, low latency storage

- Remotely attached storage incur long latency and throughput bottleneck



- Locally attached SSD & storage could be inflexible and expensive

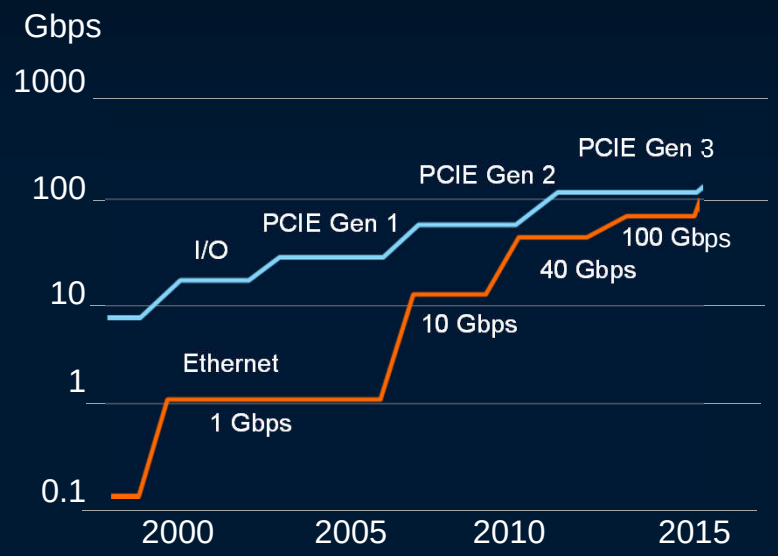
Modern analytic workloads often have wide spectrum of memory requirements



Composable systems take advantage of rapid progress on network speed and acceleration

High bandwidth network and interconnect speed is expected to be comparable to PCIe speed by 2015-2017

Network compared with System I/O



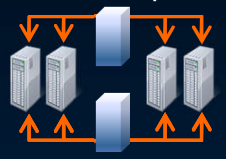
Increased focus on east-west traffic accelerate adoption of 2-tier (spine-leaf) and 1-tier DCN architectures

Network Design Choices

2-Tier Leaf-Spine

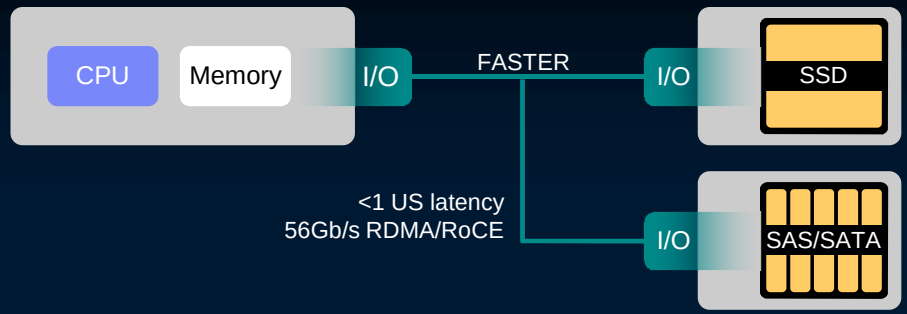
- Optimized for Scale & Growth – Cloud Model
- One network for all Apps / Tenants
- All nodes are equi-distant: 3-hops

1-Tier Spline



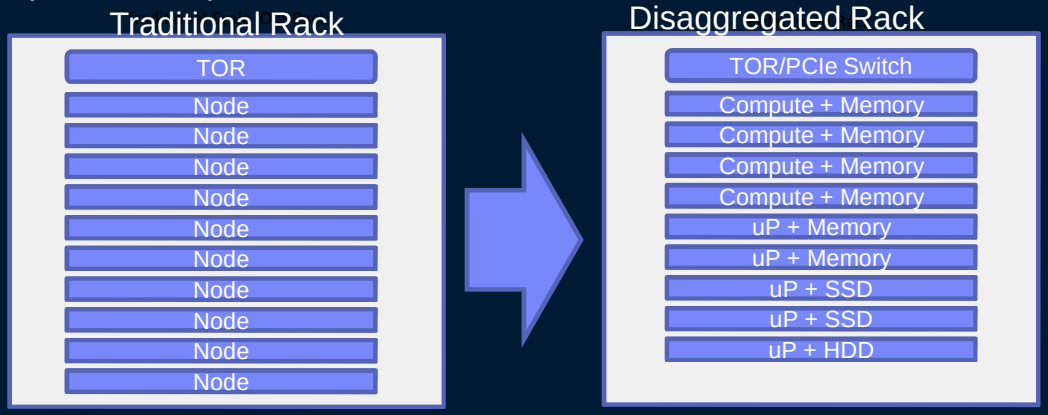
- Optimized for smaller clusters
- One network per Application
- All nodes are directly connected: 1 Hop

High speed network enables storage disaggregation with zero penalty to performance

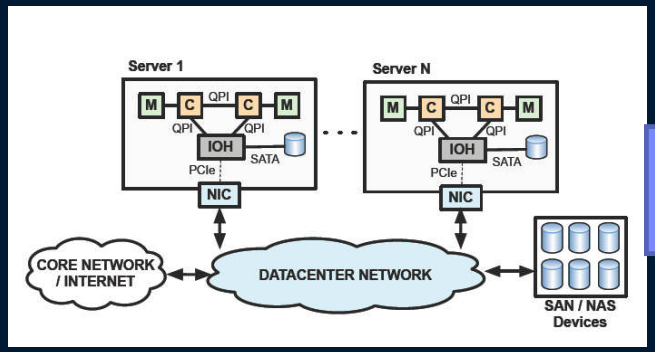


Systems composed from disaggregated datacenter offer scale of economy and flexibility in adapting system resources to rapidly changing workloads

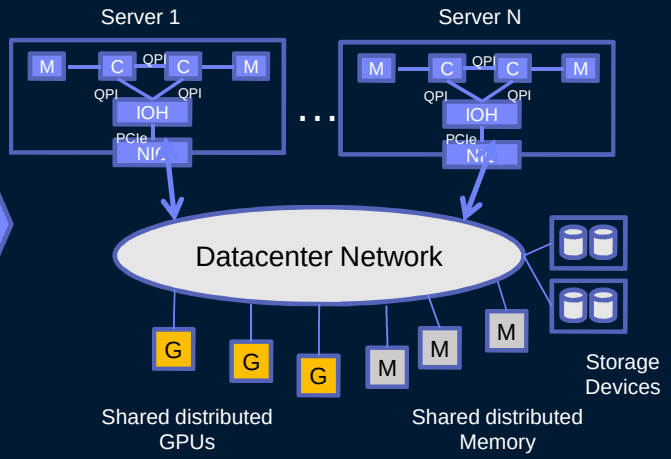
- Systems of insight workload drive substantial increase on east-west traffic
- Dynamic workload requirements and availability of higher network bandwidth enables disaggregated datacenter scale systems, resulting in agile reconfigurability and improved cost performance.



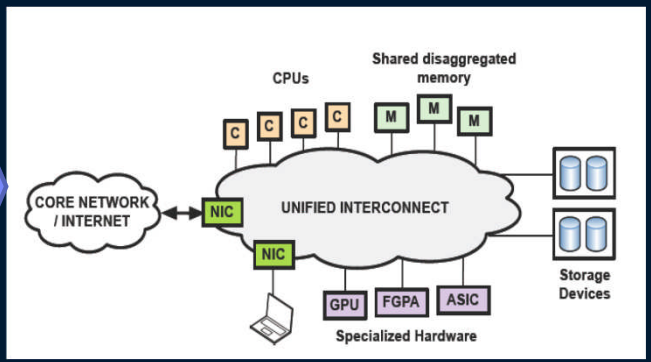
Traditional Datacenter



Partially Disaggregated Datacenter



Fully Disaggregated Datacenter



What's Next: Datacenter as a Computer

Enabled by significant reduction in cost of bandwidth and virtualization advances

Datacenter Scale "Computer"

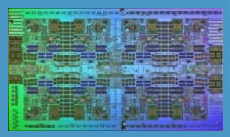
Intelligent Datacenter Infrastructure Management (DCIM) & Unified Control Plane

- SDC: Secure and lightweight container with support for heterogeneous environment including VM and bare metal.
- SDS: intelligent at scale data/object centric service.
- SDN: intelligent and agile orchestration for optimal quality of service and security.

Self-Tune & Self-Optimized, Fail-in-Place

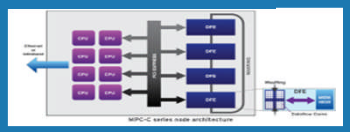
Resource Abstractions for Composable Systems

High Throughput integration with accelerators through CAPI & NVLink



GPU (Genomics, Healthcare)

TMS SSD (FSS, IoT)



Maxeler FPGA Accelerator (FSS, Natural Resources)

High BW, Low Latency Network and Interconnect

Disaggregated Components

Building Blocks for Composable System

Self-tuning could achieve 75% of optimal performance within minutes

Disaggregated fully non-blocking spine-leaf data center network based on SDN is available now.

Flat network with > Tb/s cross-sectional BW and < 5 us latency

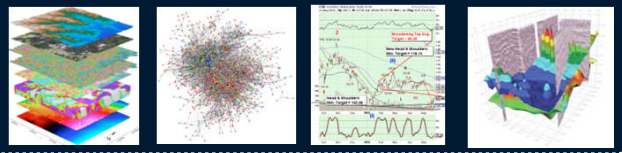
High bandwidth Si Photonics links for east–west direct connections rewired using optical switches

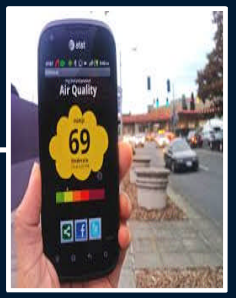
Holistic Energy efficient datacenter design

What's Next: Data and compute complexity of grand challenges define a unique category of at scale computing stack

Food Production

Grand challenges share similar characteristics of compute and data complexity and can potentially benefit from a common at scale computing stack





Environmental Quality

Sustainable Energy

Compute complexity (thread level concurrency, memory/cache, acceleratin, …)

Mount Sinai Analytics

JPMC Analytics (110K cores)

Google Search (5.9B search/day)

Facebook (1.4B MAU)

Goldman Sachs Risk Analytics

Bing Search

Salesforce

Netflix (2B hrs watched /month)

Workday

Instagram (55M upload/day)

Visa

JPMC

Twitter (400M/day)

Mastercard

Whatsapp (64B/day on April 1, 2014))

Amazon eCommerce (200M items)

YouTube (6B hrs watched /month)

Amex

Bharti

UPS

Walmart

AWS S3 (3T objects)

Dropbox

Note: Actual Location of each example is being refined.

Data complexity (data model, semantics, volume, velocity, uncertainty, data concurrency…)

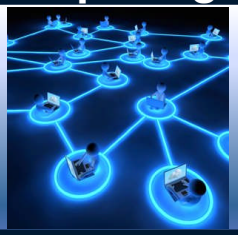
At Scale Computing: A New Computing Paradigm for Enterprise Computing

Successful computing paradigms emerged from at scale industry transformation, and differentiated through full stack optimization that includes applications, middleware, compute, storage, networking and programming models.

Mainframe Computing

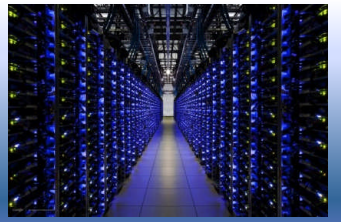
- Infrastructure: SMP,, channel, ESCON, Block Storage
- Middleware: TPF, IMS, CICS
- Application: Saber, SAP

Distributed Computing



- Infrastructure: client –server, scale out, TCP/IP, File storage
- Middleware: App Server, RDBMS, MQ Broker, SOA, ESB, BPM
- Application: 3-tier App

At Scale Computing



- Infrastructure: Warehouse Scale Computing, Flat network, Object Store
- Middleware: MapReduce, NoSQL, NewSQL, Micro Services, ZMQ,
- Application: FB, Google Search, Dropbox

Questions and Discussion

Please send your comments, questions & suggestions to
csli@us.ibm.com