**Microsoft**

# HPC & Big Compute
# with Microsoft Azure

Karl Podesta
Global Black Belt Team – HPC & Big Compute specialist
kapodest@microsoft.com
http://microsoft.com/hpc

# Mission

Our mission is to enable researchers, engineers, analysts, designers, developers, and data scientists to achieve radically better results and faster answers to complex problems, by making it easy to do simulation and parallel computing in the cloud, at hyper-scale.

Azure empowers these technical experts to think at 10x or 100x the current scale, to work faster, better, and in new ways that before the cloud we could only dream about.
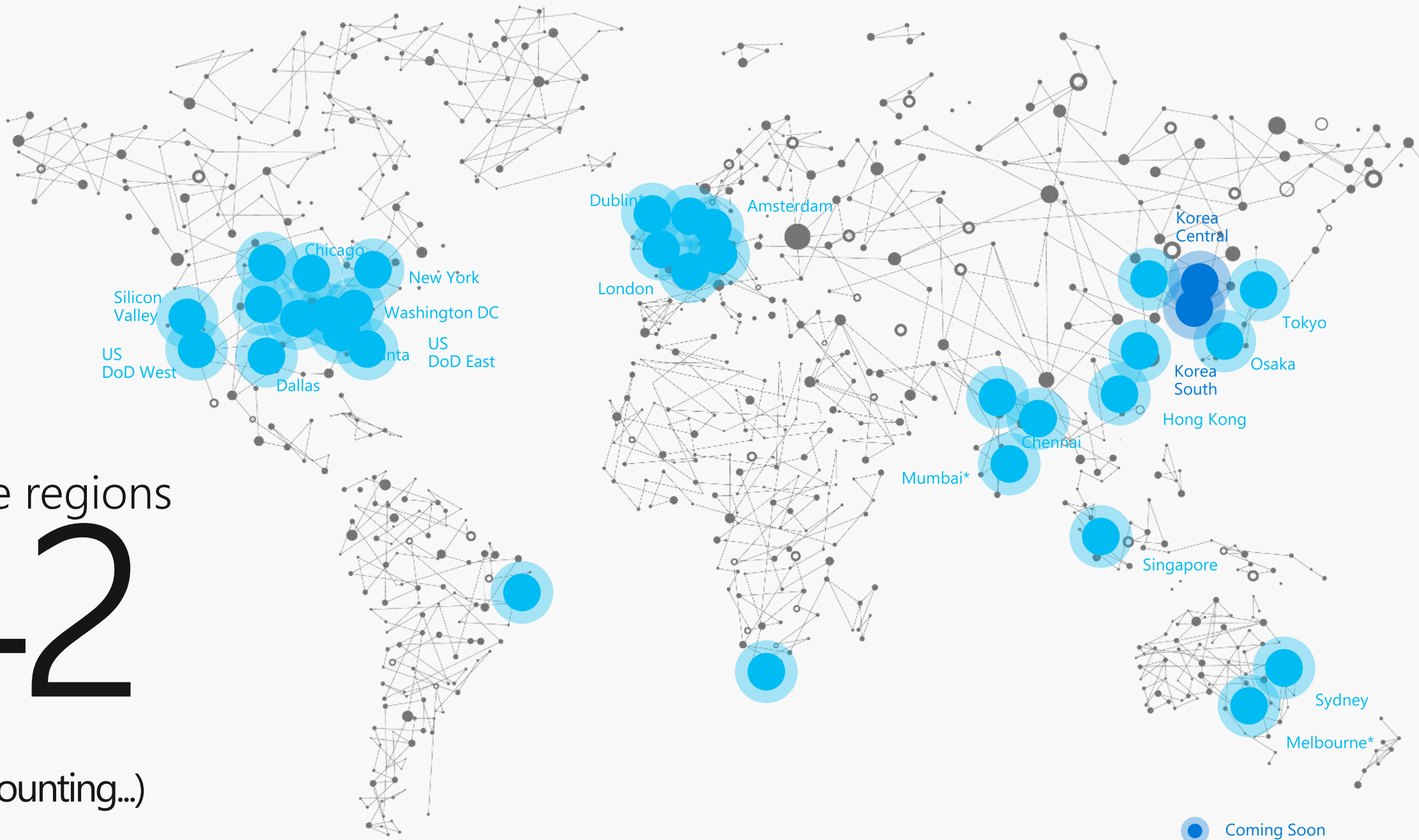
Microsoft

# Introduction to Azure

Azure regions

42

(and counting...)

Silicon Valley
US DoD West
Chicago
New York
Washington DC
Dallas
Atlanta
US DoD East

Dublin*
London
Amsterdam

Korea Central
Korea South
Tokyo
Osaka
Hong Kong

Mumbai*
Chennai

Singapore

Sydney
Melbourne*

Coming Soon

# 1.6
## million miles of fiber

Datacenter

CDN Locations

Edge Node

Internet Exchange

Terrestrial Network

Subsea Network
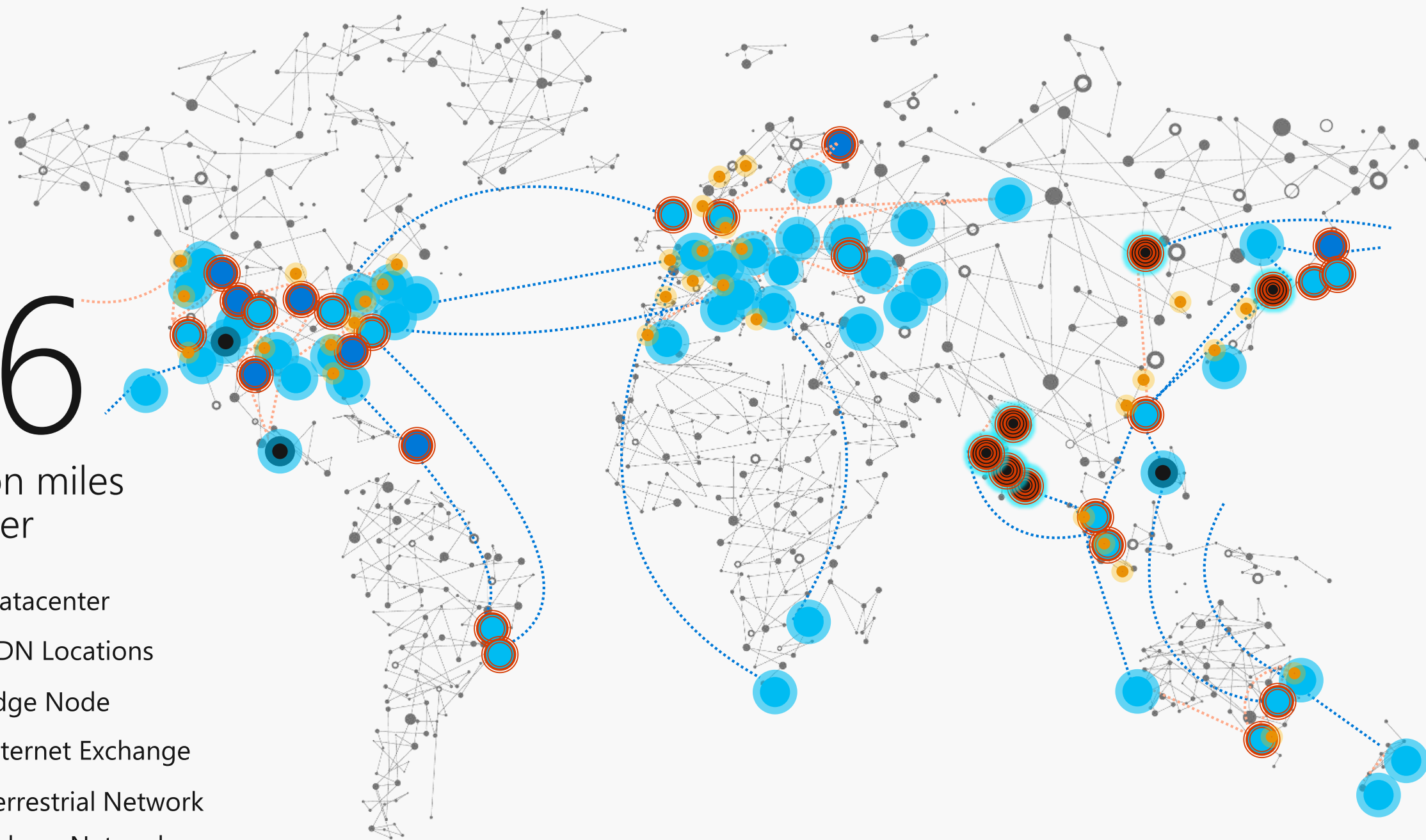
# ExpressRoute

Private connection to Azure
High-throughput
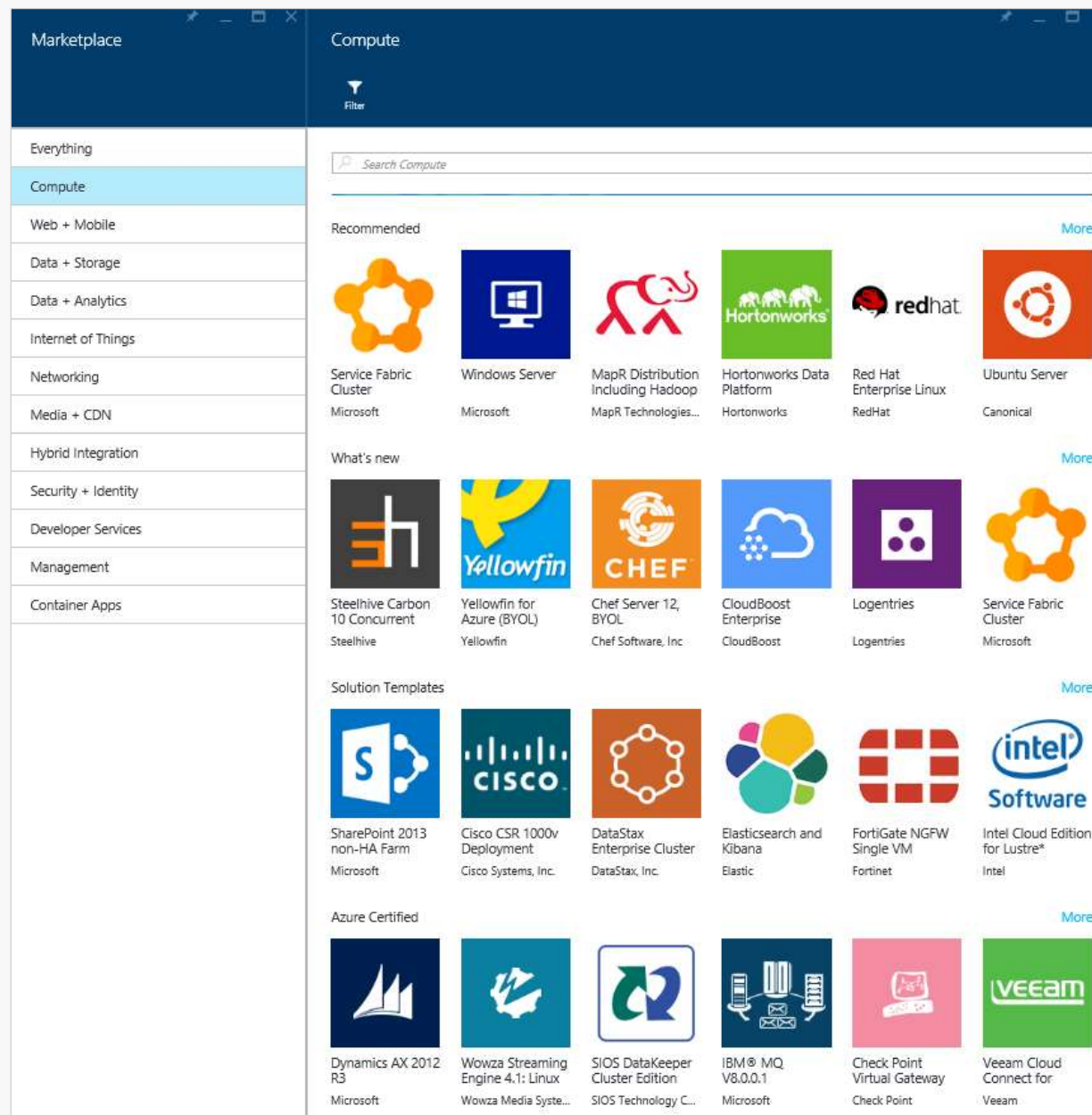Predictable  performance

Seattle

Silicon
Valley

Los Angeles

Chicago

Dallas

New York

Washington DC

Atlanta

Sao Paulo

Dublin*

London

Amsterdam

Mumbai*

Chennai

Hong Kong

Tokyo

Osaka

Singapore

Sydney

Melbourne*

# >4000

## Certified, pre-configured solutions in Azure Marketplace

**Marketplace**

- Everything
- Compute
- Web + Mobile
- Data + Storage
- Data + Analytics
- Internet of Things
- Networking
- Media + CDN
- Hybrid Integration
- Security + Identity
- Developer Services
- Management
- Container Apps

**Compute**

Filter

Search Compute

### Recommended — More

| | | | | | |
|---|---|---|---|---|---|
| Service Fabric Cluster | Windows Server | MapR Distribution Including Hadoop | Hortonworks Data Platform | Red Hat Enterprise Linux | Ubuntu Server |
| Microsoft | Microsoft | MapR Technologies... | Hortonworks | RedHat | Canonical |

### What's new — More

| | | | | | |
|---|---|---|---|---|---|
| Steelhive Carbon 10 Concurrent | Yellowfin for Azure (BYOL) | Chef Server 12, BYOL | CloudBoost Enterprise | Logentries | Service Fabric Cluster |
| Steelhive | Yellowfin | Chef Software, Inc | CloudBoost | Logentries | Microsoft |

### Solution Templates — More

| | | | | | |
|---|---|---|---|---|---|
| SharePoint 2013 non-HA Farm | Cisco CSR 1000v Deployment | DataStax Enterprise Cluster | Elasticsearch and Kibana | FortiGate NGFW Single VM | Intel Cloud Edition for Lustre* |
| Microsoft | Cisco Systems, Inc. | DataStax, Inc. | Elastic | Fortinet | Intel |

### Azure Certified — More

| | | | | | |
|---|---|---|---|---|---|
| Dynamics AX 2012 R3 | Wowza Streaming Engine 4.1: Linux | SIOS DataKeeper Cluster Edition | IBM® MQ V8.0.0.1 | Check Point Virtual Gateway | Veeam Cloud Connect for |
| Microsoft | Wowza Media Syste... | SIOS Technology C... | Microsoft | Check Point | Veeam |

# Azure covers 71 compliance offerings

Azure has the deepest and most comprehensive compliance coverage in the industry

## Global

| ISO 27001 | ISO 27018 | ISO 27017 | ISO 22301 | SOC 1 Type 2 | SOC 2 Type 2 | SOC 3 | CSA STAR Self-Assessment | CSA STAR Certification | CSA STAR Attestation |

## US Gov

| Moderate JAB P-ATO | High JAB P-ATO | DoD DISA SRG Level 2 | DoD DISA SRG Level 4 | DoD DISA SRG Level 5 | SP 800-171 | FIPS 140-2 | Section 508 VPAT | ITAR | CJIS | IRS 1075 |

## Industry

| PCI DSS Level 1 | CDSA | MPAA | FACT UK | Shared Assessments | FISC Japan | HIPAA / HITECH Act | HITRUST | GxP 21 CFR Part 11 | MARS-E | IG Toolkit UK | FERPA | GLBA | FFIEC |

## Regional

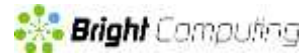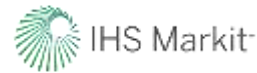| Argentina PDPA | EU Model Clauses | UK G-Cloud | China DJCP | China GB 18030 | China TRUCS | Singapore MTCS | Australia IRAP/CCSL | New Zealand GCIO | Japan My Number Act | ENISA IAF | Japan CS Mark Gold | Spain ENS | Spain DPA | India MeitY | Canada Privacy Laws | Privacy Shield | Germany IT Grundschutz workbook |

# Open and integrated
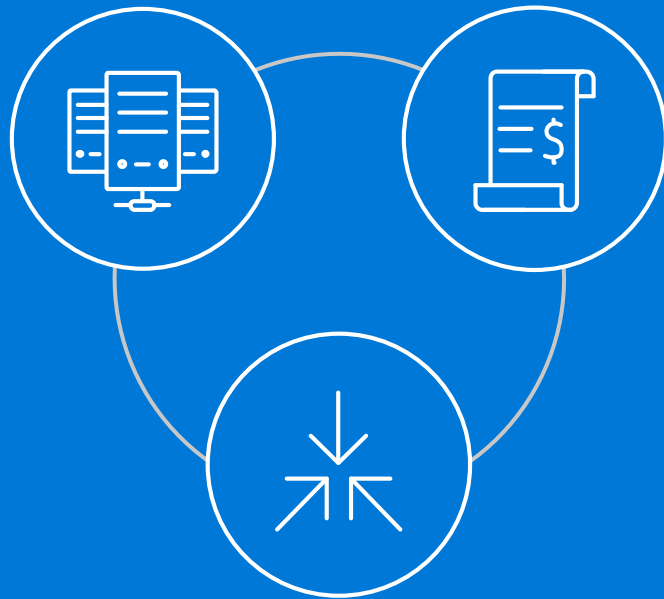
## Industry specific

### Life sciences

CYCLE COMPUTING · Bright Computing · BizData · teradici · Workspot · rescale

### Finance
IHS Markit · Willis Towers Watson · kinetica · ACTIVEVIAM · Milliman · excelian · rescale

### Manufacturing
Altair · Schlumberger · d3VIEW · JSOL Corporation · HALLIBURTON · Dassault Systèmes · OpenFOAM · CD-adapco · SIEMENS · esi · PBS Works · ANSYS · MSC Software · SIMULIA · LSTC · UberCloud
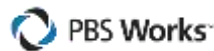
## Software and hardware

### Languages
node js · R · python · Ruby · Java · php

### Operating systems
ubuntu · redhat · FreeBSD · CentOS · SUSE · Linux · Windows Server

### Infrastructure
intel · UberCloud · teradici · Mellanox Technologies · CYCLE COMPUTING · excelian · rescale · FUJITSU · Workspot · Bright Computing

# We Research and we Add Value

## Grants for non-profit research



## Microsoft Genomics Service



## Azure Notebooks
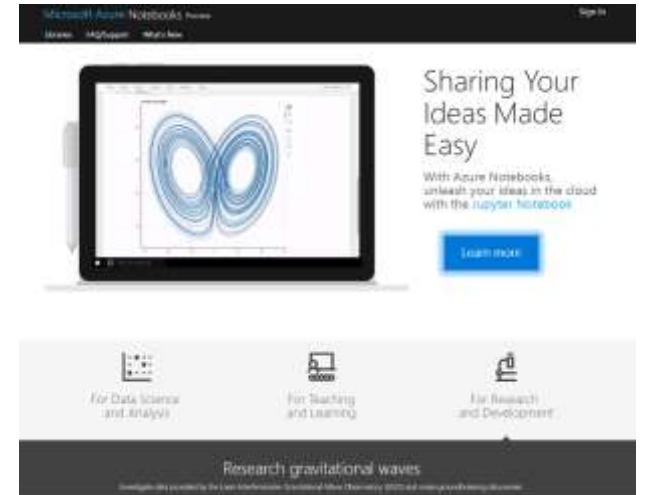


## We will bring Quantum to everyone

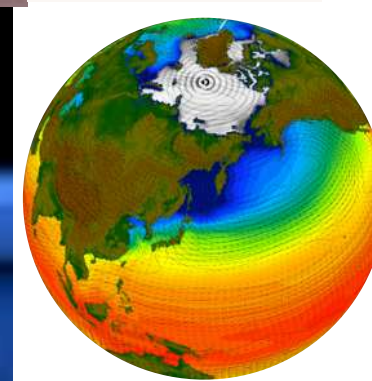

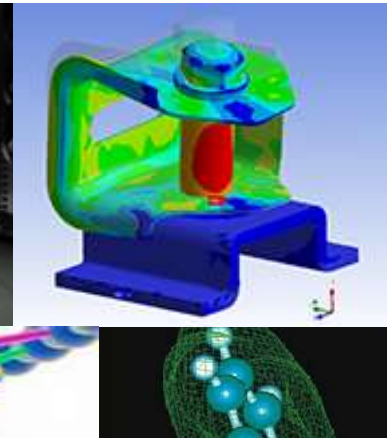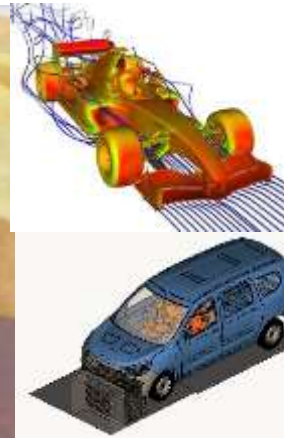## Top contributors on GitHub; Solutions/Recipes shared
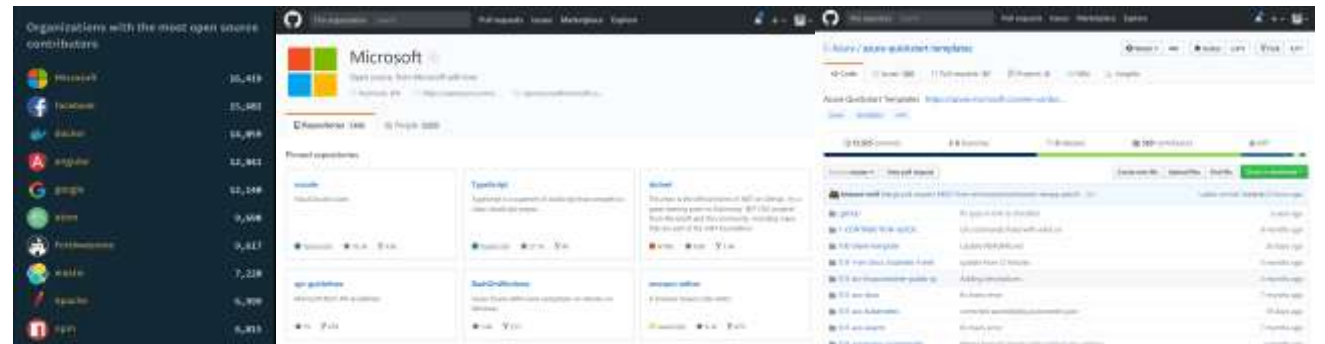
# What is High Performance Computing (HPC)?
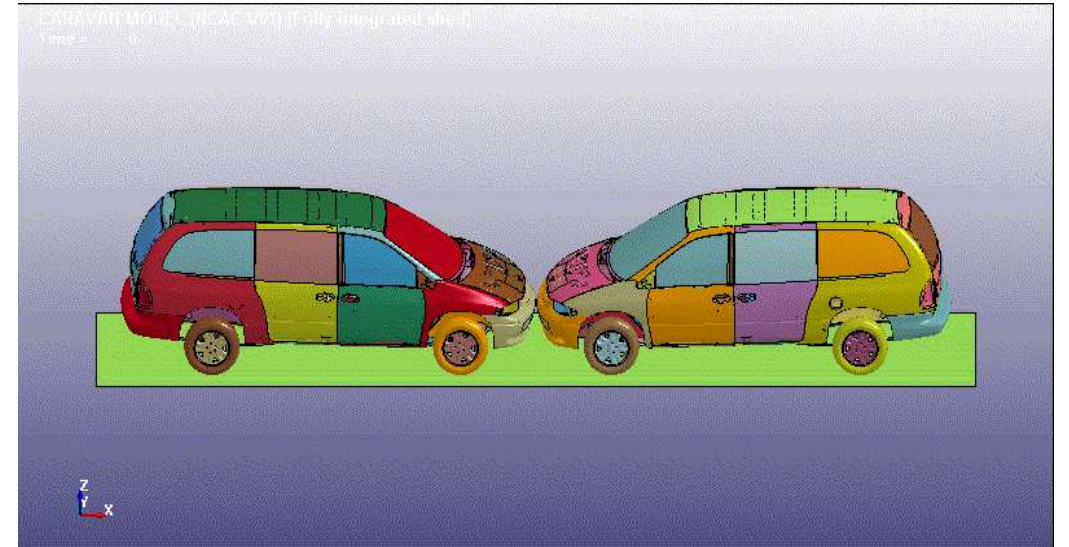
# Heard about Big Data?  … What about Big Compute?
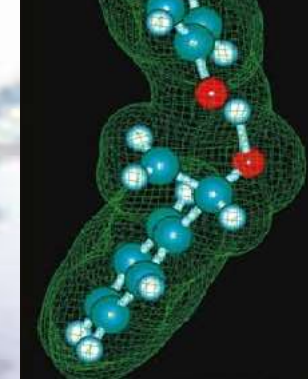


How can our banks understand risk?  Can we have safer cars?
How will global weather change affect us?  Can we find & cure disease?
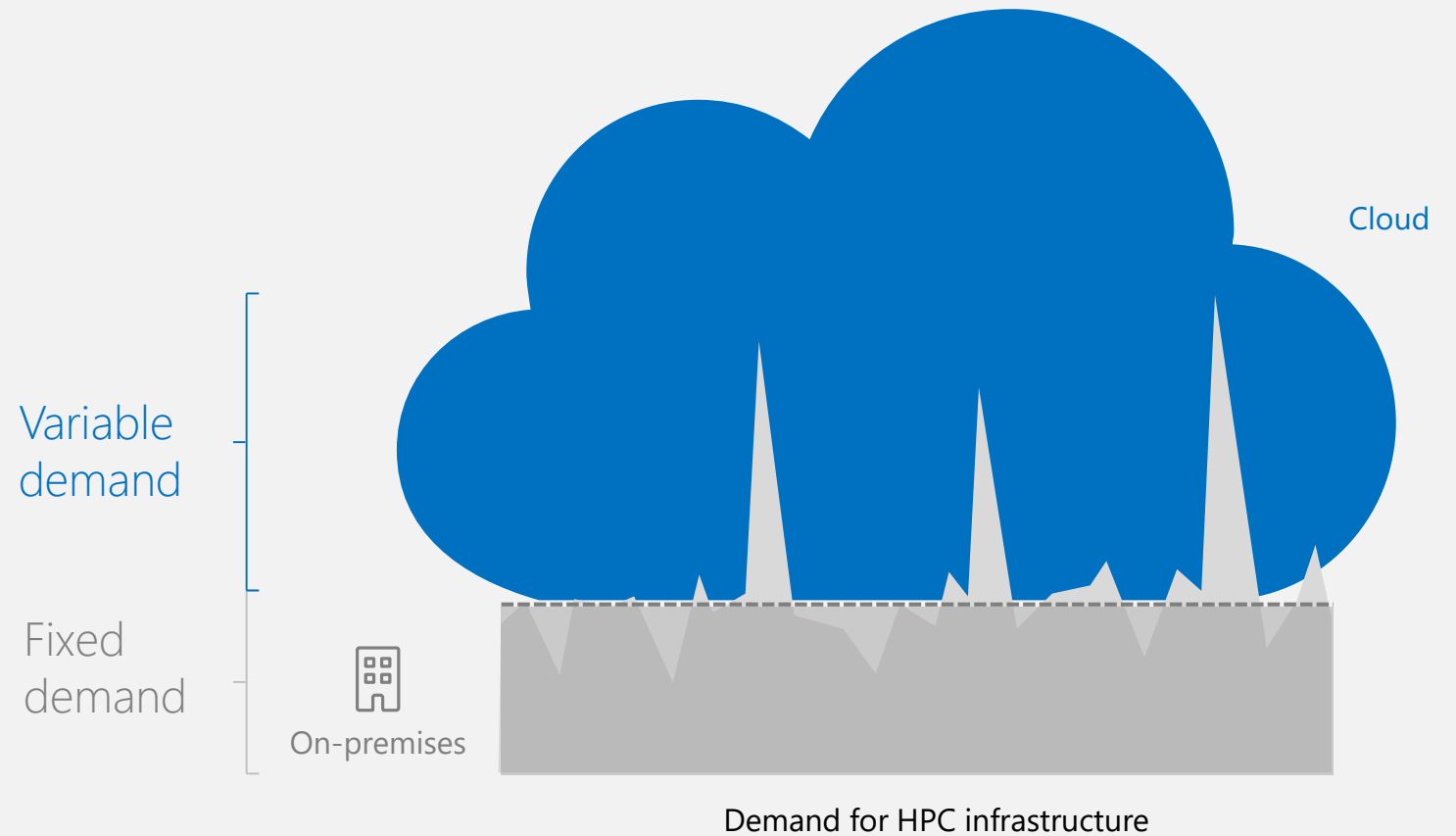
# HPC is about doing even more…

With HPC you can run more tests, analyze more scenarios, tackle more complex problems, explore new solutions, get faster answers…

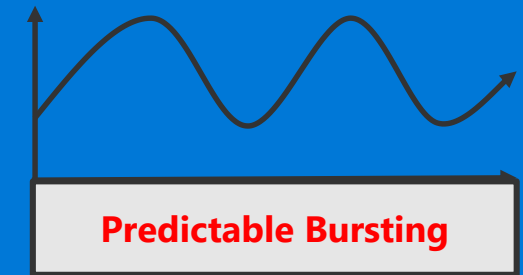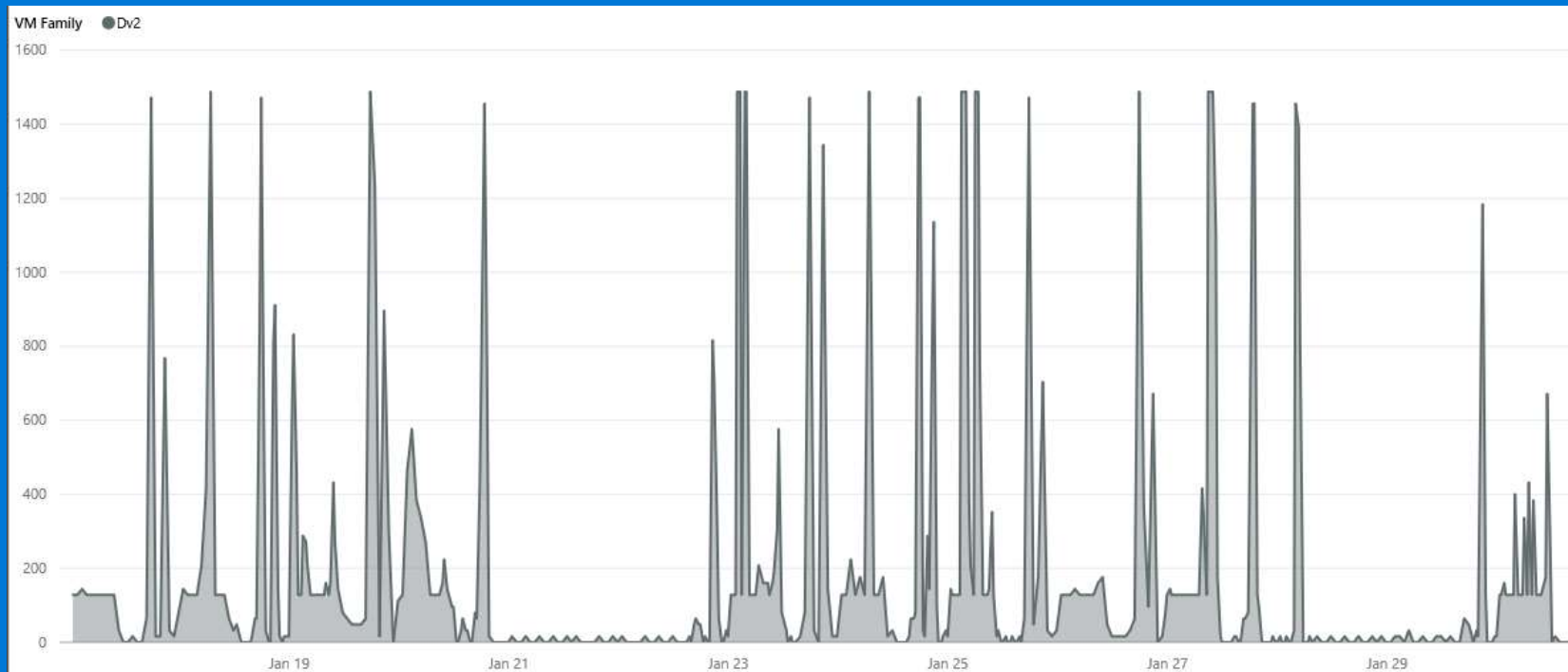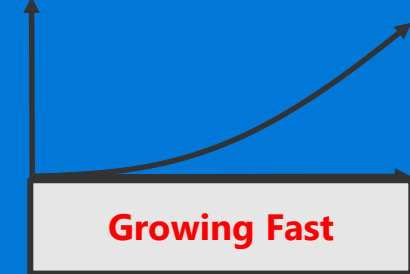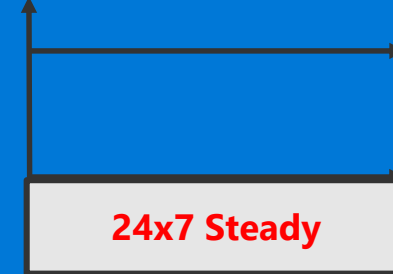The new HPC Opportunity

Expand your HPC environment to the Cloud

Cloud

Variable demand

Fixed demand

On-premises

Demand for HPC infrastructure

# Elasticity
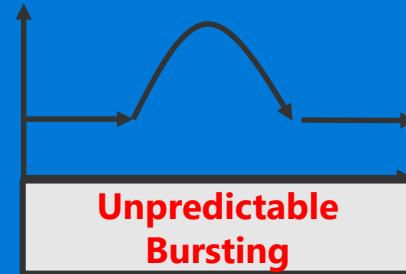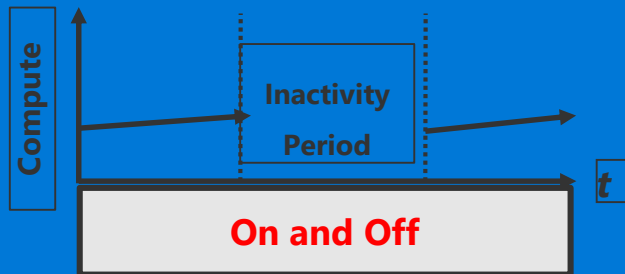
# What would you do with 100x the scale?

## Do more with hyper-scale:

- Service more users
- Run more projects
- Get results faster

- Run larger simulations
- Explore new insights (e.g., "What if?")

## Remove current limitations:

- Modify more parameters
- Analyze more complex models
- Visualize larger results

- Run more iterations
- Generate higher fidelity results
- Simulate longer periods of time
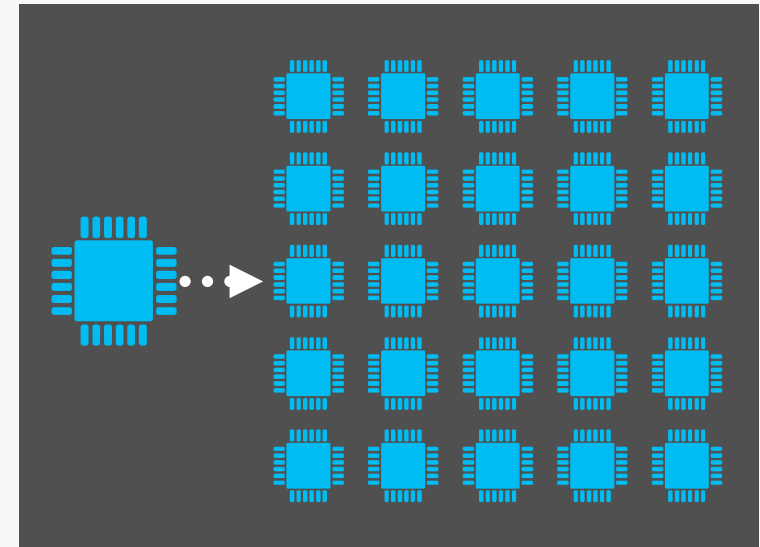
# Cloud value for Big Compute (HPC)
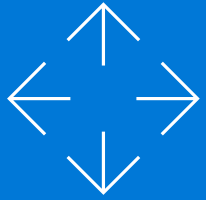
Scale

Elasticity

Pay for use

Reach & locality

Hardware selection

No infrastructure

Known costs

# Why InfiniBand RDMA matters?



Nodes with Ethernet Vs A9 run time for crash models/jobs

Lack of RDMA (InfiniBand)

Azure A9 nodes MPI RDMA
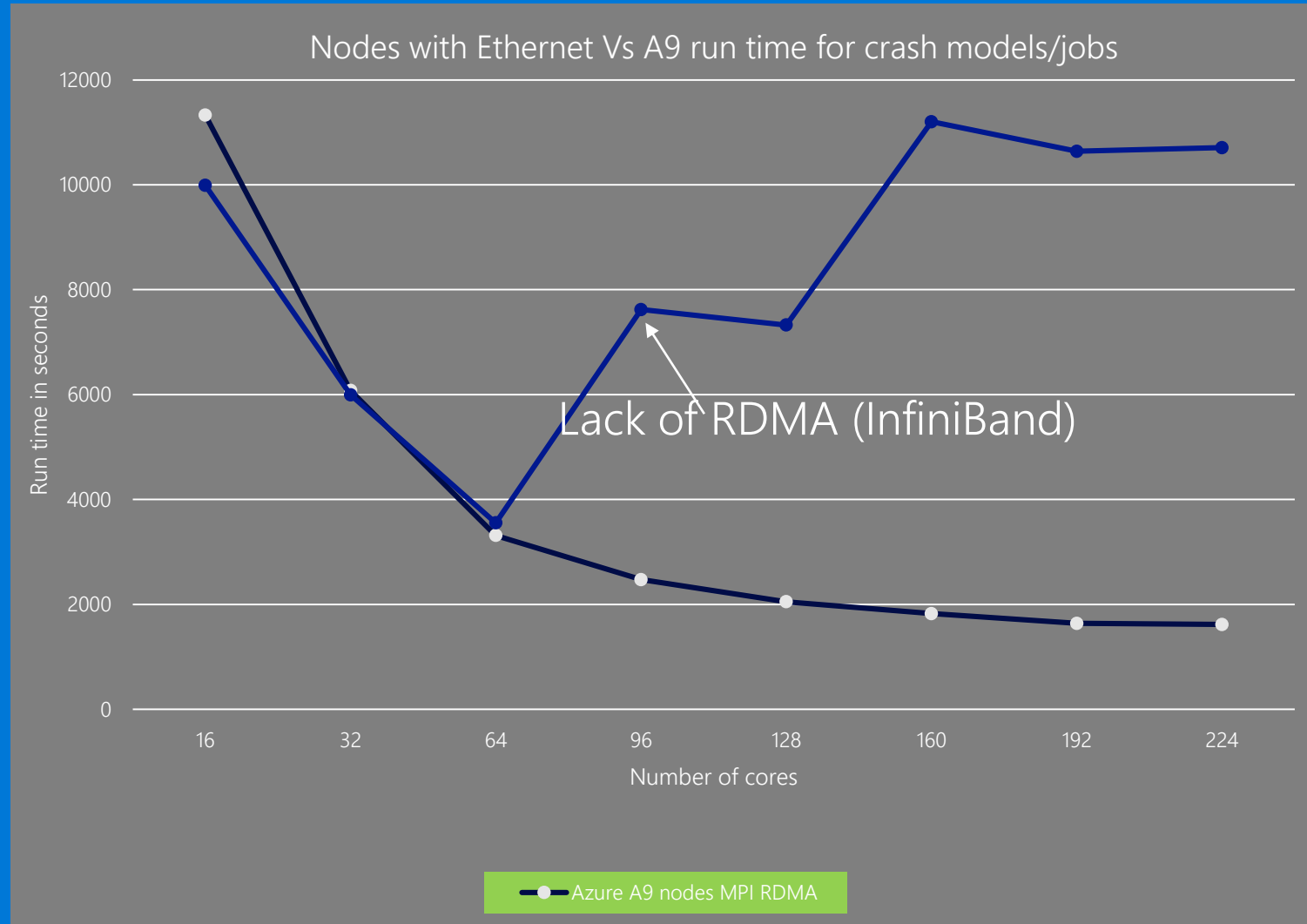
# STAR-CCM+ Scalability to 1024 cores



Average Elapsed time per iteration
STAR-CCM+ Le Mans 100 Millions Cell Model



Scalability
STAR-CCM+ Le Mans 100 Millions Cell Model

— Azure A9
--- Ideal (100% scaling)

https://azure.microsoft.com/en-us/blog/availability-of-star-ccm-on-microsoft-azure/

# Fluent Benchmark



Fluent 280M Cell Open Race Car Model Benchmark Scaling on Microsoft Azure A9

# Fluent Benchmark



Fluent 140M Cell F1 Race Car Model Benchmark Scaling on Microsoft Azure A9

# Fluent Benchmark



Fluent 71 million Cell Combustor Model Benchmark Scaling on Microsoft Azure A9

# Level Set: Cloud 101

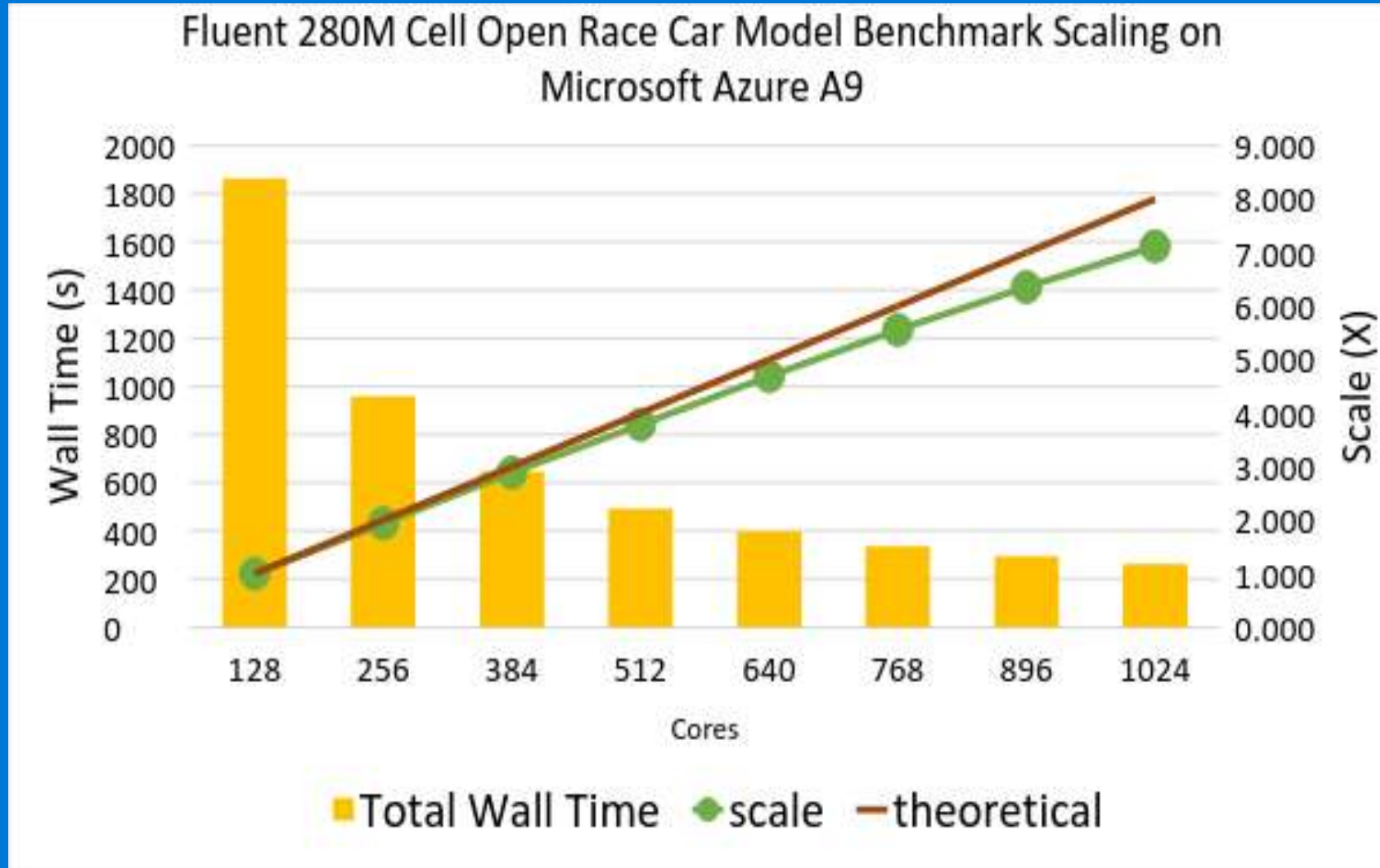| On Premises | Infrastructure (as a Service) | Platform (as a Service) | Software (as a Service) |
|---|---|---|---|
| Applications | Applications | Applications | Applications |
| Data | Data | Data | Data |
| Runtime | Runtime | Runtime | Runtime |
| Middleware | Middleware | Middleware | Middleware |
| O/S | O/S | O/S | O/S |
| Virtualization | Virtualization | Virtualization | Virtualization |
| Servers | Servers | Servers | Servers |
| Storage | Storage | Storage | Storage |
| Networking | Networking | Networking | Networking |

You Manage
Vendor Manages

# Application types in HPC

## Embarrassingly parallel:
- Applications do not communicate
- May share common store & data
- May have dependencies
- E.g. Monte Carlo simulations, transcoding, rendering

## Tightly coupled:
- Applications communicate; mainly use MPI
- Requires low latency, high bandwidth networking for scale
- E.g. car crash simulation, fluid dynamics, AI training

# Azure Big Compute – Solution Architectures
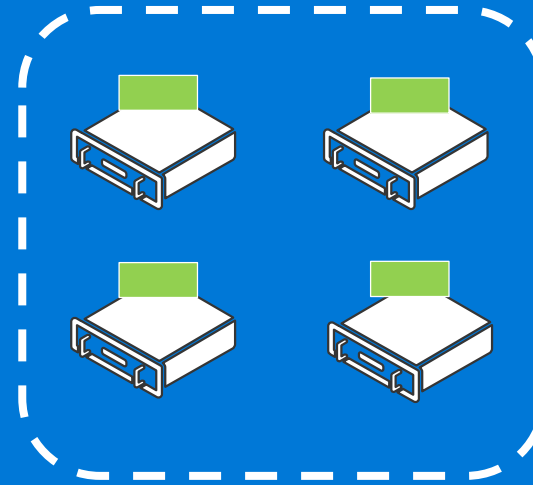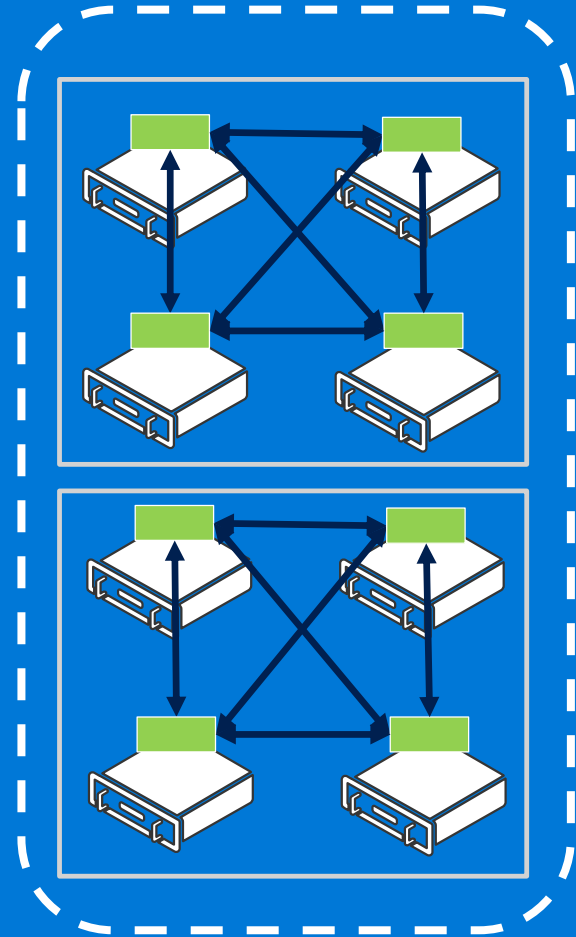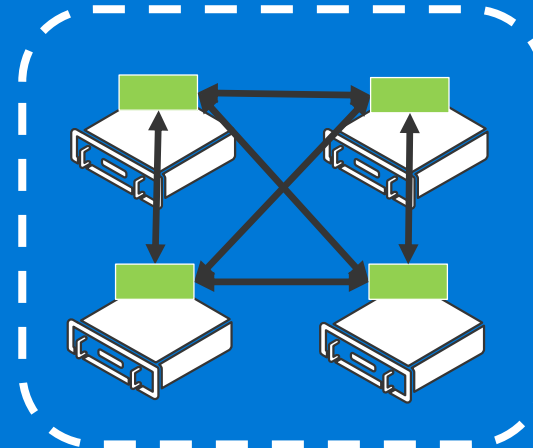
## Self-managed / Hybrid

### Cloud burst

On-premises / Compute nodes in the cloud

### Cluster on the cloud
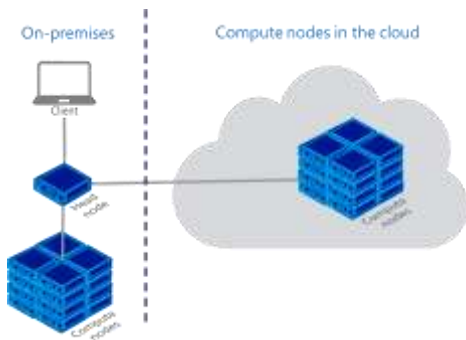
On-premises / All HPC resources in the cloud

### Remote Workstations

On-premises / Remote workstations on the cloud

NV NV
NV
NV NV

## Fully-managed

### HPC as a service

On-premises / All HPC resources in the cloud

Azure Batch
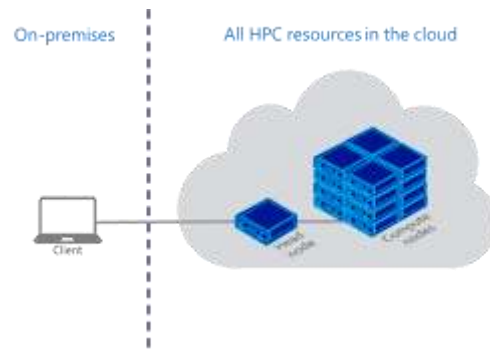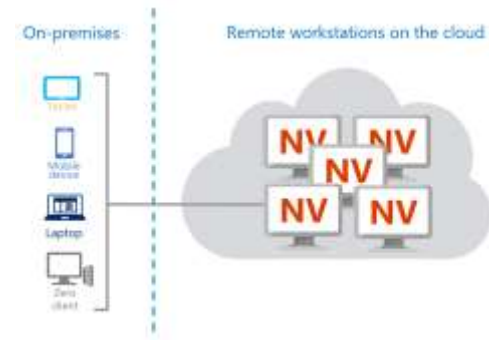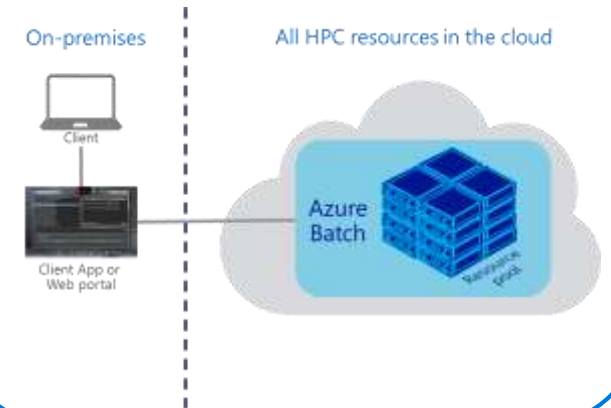
... and HPC is just one building block ... (150+ services in Azure, including:)

Import/ Export    ExpressRoute    Azure Network    Azure Storage    StorSimple    IoT    Stream Analytics    Machine Learning    Data Lake    HDInsight    Data Factory
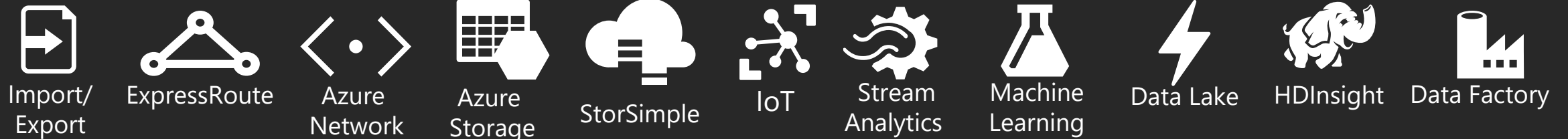
# Next Generation HPC – thinking differently with public cloud

"A cluster (cloud!) for every HPC workload"
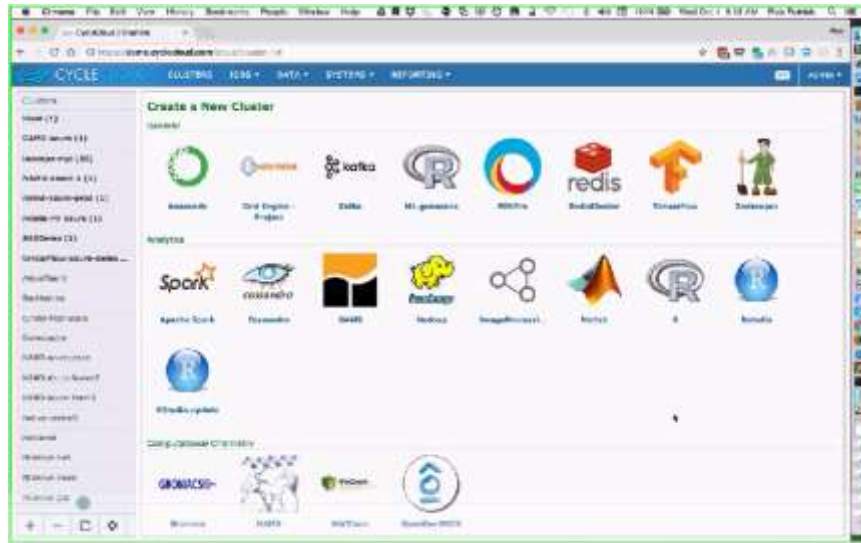


"HPC is one building block"
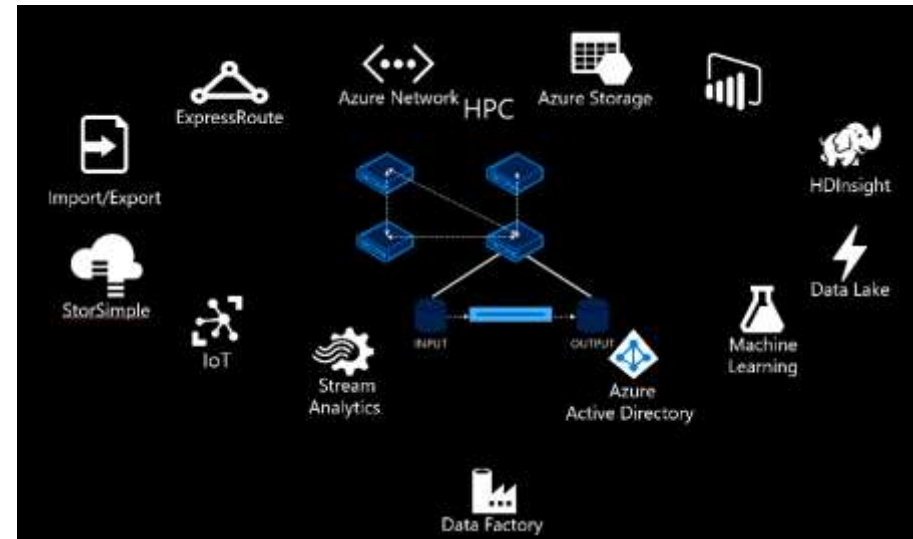


"HPC Infrastructure as code"
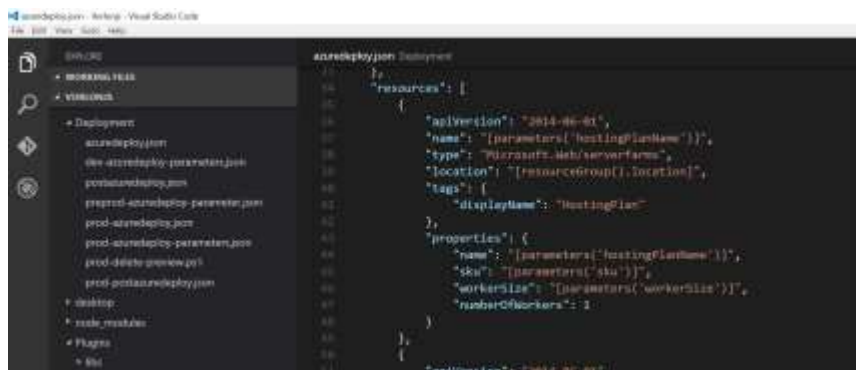


"AI, ML, Data Analytics"

# Build in intelligence with AI: our Cognitive Services API



**Vision**

From faces to feelings, allow apps to understand images and video

*Show me what is in the image*

**Speech**

Hear and speak to users by filtering noise, identifying speakers, and understanding intent

*Convert this text to speech please...*

**Language**

Process text and learn how to recognize what users want

*Play today's conference call...*

**Knowledge**

Tap into rich knowledge amassed from the web, academia, or your own data

*Top publications in AI...*

**Search**

Locate relevant information among billions of web pages, images, videos, and news with Bing APIs

*Fraud prevention results...*

Policy violation: Jackhammer unstable orientation detected in the workshop

Now Playing

11/29/2016 Customer Call

Natural Language Processing

Intent: PlayCall
Content: Customer#
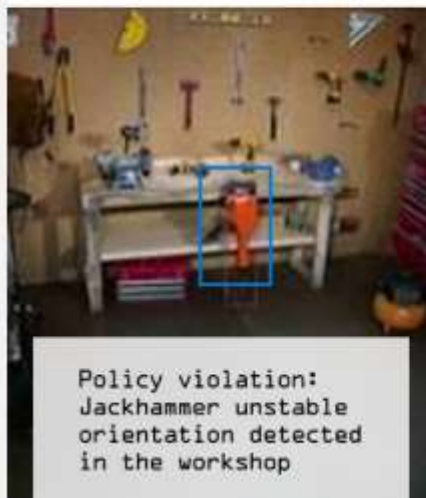DateTime.date: today

Genetic Algorithms in Search, Optimization Learning
1989, David E Goldberg
Cited 18,910 times

Induction of Decision Trees
1986, Machine Learning
J R Quinlan
decision tree  expert system  machine learning  compute
Cited 4,019 times

Outline of a New Approach to the Analysis Systems and Decision Processes
1973, IEEE Transactions on Systems, Man, and Cybernetics, iss
Lotfi A Zadeh (University Of California Berkeley)
fuzzy set  complex systems  artificial neural network  com
machine learning  artificial intelligence  computer science
Cited 2,602 times

How ID theft victims' in
nment documents or benefits
Credit card fraud
Phone or utilities fraud
Bank fraud  8.
Employment-related fraud  4.8%
Attempted identity theft  4.8%
Loan fraud  4.4%
Other
sion Consumer Sentinel Network Data Book,

# Hardware capabilities
## VM sizes

**A** — General purpose

**D** — SSD storage & fast CPUs

**Dv2** — New generation of D family VMs

**G** — High memory and large SSDs

**Ls** — Low-latency and extra-large SSDs

**NC** — NVIDIA GPUs K80 compute *P40 & P100 in preview

**NV** — NVIDIA GPUs M60 visualization

**H** — Fastest CPU IB connectivity

**D Dv2 F G S** — Premium storage

**F** — Compute intensive

Specialized infrastructure

All are/can be used for HPC

CRAY
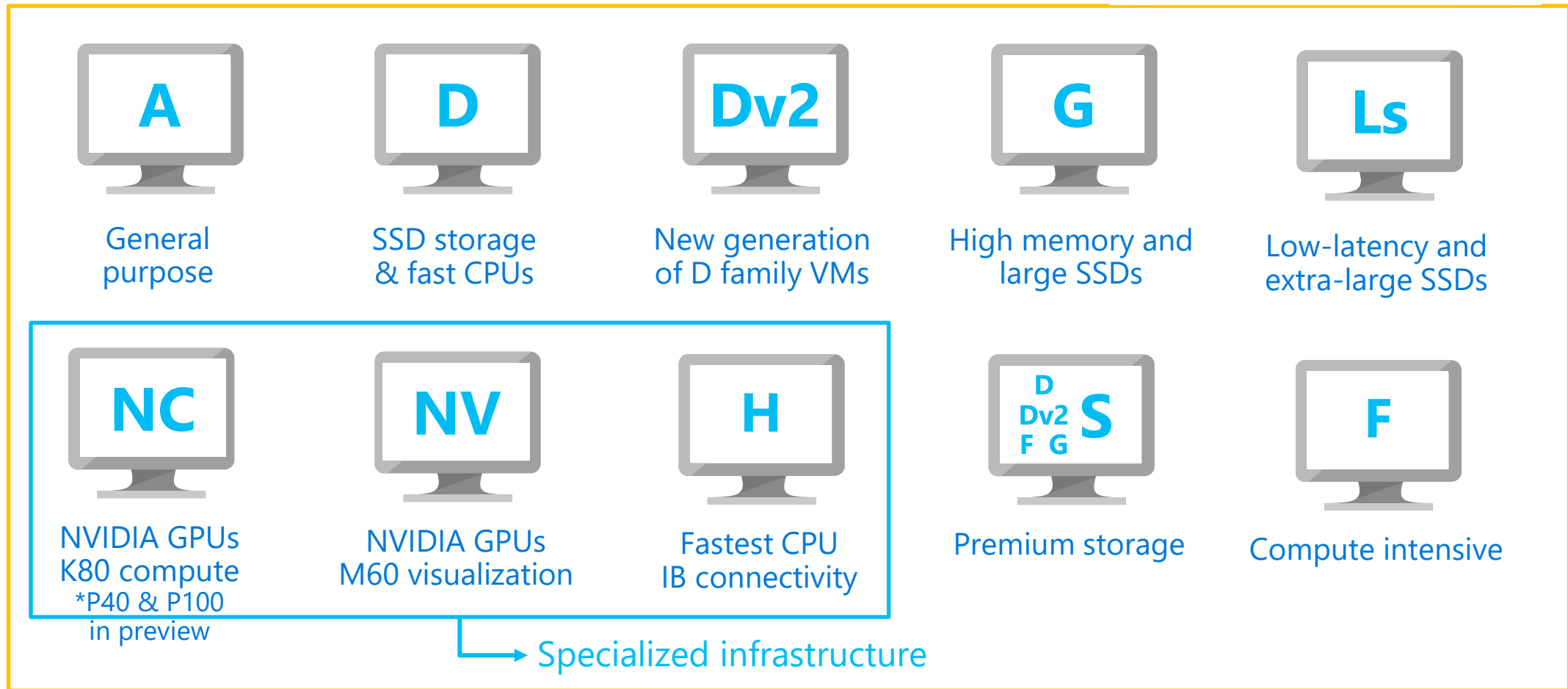
# No-compromise HPC and AI VMs

**H**
- Up to 16 cores, 3.2 GHz E5-2667 V3 Haswell processor
- Up to 224 GiB DDR4 memory
- FDR InfiniBand (56 Gbps, 2.6 microsecond latency)
- 2 TB of local SSD

**ND**
- Up to 4 NVIDIA Pascal P40 GPUs
- Up to 24 cores
- Up to 448 GiB memory
- Up to 3 TB of local SSD
- FDR InfiniBand

**NC**
- Up to 4 NVIDIA Tesla K80 GPUs
- Up to 24 cores
- Up to 224 GiB memory
- Up to 1440 GiB of local SSD
- FDR InfiniBand

**NV**
- Up to 4 NVIDIA Tesla M60 GPUs
- Up to 24 cores
- Up to 224 GiB memory
- Up to 1440 GiB of local SSD

**NC$_v$2**
- Up to 4 NVIDIA Pascal P100 GPUs
- Up to 24 cores
- Up to 448 GiB memory
- Up to 3 TB of local SSD
- FDR InfiniBand

CycleCloud | Clusters ✕ | | Rob

← → C ⟳ ⌂ | 🔒 Secure | https://demo.cyclecloud.com/home | ☆ 🔴 🛡 🔵 🔒 📋 ⬛ 🔵 ⋮

# CYCLECLOUD | CLUSTERS | JOBS ▾ | DATA ▾ | SYSTEMS ▾ | REPORTING ▾ | ✉ | ADMIN ▾

## Clusters

**blast (1)**

**GAMS-azure (1)**

**lammps-mpi (65)**

**NAMD-azure-1 (1)**

**namd-azure-prod (2)**

**nvidia-ml-azure (1)**

**tensorflow-azure-demo ...**

AnsysFluent

Backtesting

condor-high-scale

DemoLustre

NAMD-azure-jason

NAMD-Azure-Jason-2

NAMD-azure-Team2

nat-us-central1

nvidia-ml

RCensus-aws

RCensus-azure

RCensus-gcp

sge-azure-benchmark

➕ ➖ 🗗 ⚙

## lammps-mpi

⊙ Terminate     State **Started** at 6/20/17 11:19 AM (up 105d 19h 49m)
✏ Edit     Nodes **1** ready, **64** acquiring
👥 Share     Scalesets **1** configuring
🔄 Refresh     Size **1** instance, **16** cores (**$2.14** per hour)
    Usage **806.8 core-hours** (~**$52**) in the last 24 hours
    Alerts 🔔 Create new alert

Show: Active ▾ Instances ▾ by MachineType ▾

1

0

06:10    06:20    06:30    06:40    06:50    07:00

| Nodes | Arrays | Azure Scalesets |
|-------|--------|-----------------|

View: Template ▾ 🔄    Actions ▾        🔍 Search

| Template ▾ | Nodes | Cores | Status | Last Message |
|------------|-------|-------|--------|--------------|
| master | 1 | 16 | 🟩 | .... |
| execute | 64 | 1024 | 🟨 | Allocating nodes |

View: Details ▾    Show Detail   Edit   Connect   Actions ▾     🔍 Search

| Name ▲ | Type | Status | Cores | Zone | Host Name | UpTime |
|--------|------|--------|-------|------|-----------|--------|
| execute-1 | Standard_H16mr | Acquiring | 16 | .... | 10.8.0.9 | |
| execute-10 | Standard_H16mr | Acquiring | 16 | .... | 10.8.0.10 | |
| execute-11 | Standard_H16mr | Acquiring | 16 | .... | 10.8.0.21 | |
| execute-12 | Standard_H16mr | Acquiring | 16 | .... | 10.8.0.33 | |
| execute-13 | Standard_H16mr | Acquiring | 16 | .... | 10.8.0.46 | |
| execute-14 | Standard_H16mr | Acquiring | 16 | .... | 10.8.0.58 | |
| execute-15 | Standard_H16mr | Acquiring | 16 | .... | 10.8.0.69 | |
| execute-16 | Standard_H16mr | Acquiring | 16 | .... | 10.8.0.74 | |
| execute-17 | Standard_H16mr | Acquiring | 16 | .... | 10.8.0.75 | |

🗗 🔄    Show Detail     🔍 Search

| Time ▾ | Message |
|--------|---------|
| 7:06 AM | ℹ Added 64 nodes of type execute to cluster lammps |
| 9/28/17 9:20 AM | ℹ Node execute-6 in cluster lammps-mpi finished st |
| 9/28/17 9:19 AM | ℹ 8 nodes in cluster lammps-mpi finished starting |
| 9/28/17 9:19 AM | ℹ 4 nodes in cluster lammps-mpi finished starting |
| 9/28/17 9:19 AM | ℹ 2 nodes in cluster lammps-mpi finished starting |
| 9/28/17 9:19 AM | ℹ Node execute-11 in cluster lammps-mpi finished s |
| 9/28/17 9:12 AM | ℹ Added 16 nodes of type execute to cluster lammps |
| 9/19/17 5:44 AM | ℹ Node execute-25 in cluster lammps-mpi finished s |
| 9/19/17 5:43 AM | ℹ 3 nodes in cluster lammps-mpi finished starting |
| 9/19/17 5:43 AM | ℹ 2 nodes in cluster lammps-mpi finished starting |
| 9/19/17 5:43 AM | ℹ 2 nodes in cluster lammps-mpi finished starting |
| 9/19/17 5:43 AM | ℹ 7 nodes in cluster lammps-mpi finished starting |
| 9/19/17 5:42 AM | ℹ 4 nodes in cluster lammps-mpi finished starting |
| 9/19/17 5:42 AM | ℹ 5 nodes in cluster lammps-mpi finished starting |
| 9/19/17 5:42 AM | ℹ 2 nodes in cluster lammps-mpi finished starting |
| 9/19/17 5:42 AM | ℹ 4 nodes in cluster lammps-mpi finished starting |
| 9/19/17 5:41 AM | ℹ 4 nodes in cluster lammps-mpi finished starting |
| 9/19/17 5:41 AM | ℹ 6 nodes in cluster lammps-mpi finished starting |
| 9/19/17 5:41 AM | ℹ 4 nodes in cluster lammps-mpi finished starting |
| 9/19/17 5:41 AM | ℹ 3 nodes in cluster lammps-mpi finished starting |
| 9/19/17 5:40 AM | ℹ 6 nodes in cluster lammps-mpi finished starting |
| 9/19/17 5:40 AM | ℹ 3 nodes in cluster lammps-mpi finished starting |

# Azure Batch – Concepts

## Before....

| User application or service | |
|---|---|
| Get and manage VMs | Install task applications |
| Manage and authenticate users | Start the tasks |
| Move task input and output | Queue tasks |
| Task failure? Task frozen? | Scale up and down |
| PaaS Cloud Services | IaaS Virtual Machines |
| Hardware | |

## After...

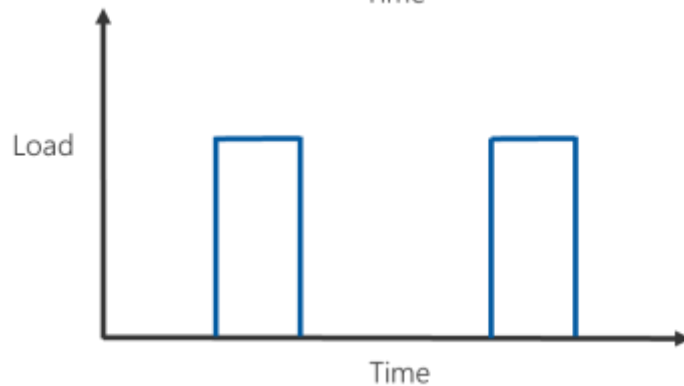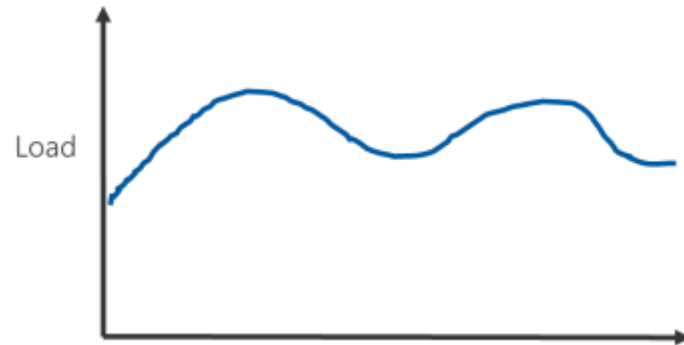| User application or service |
|---|
| **Azure Batch** App lifecycle, job dependencies, data movement, task rescheduling, user management & authorization / VM management and job scheduling |
| PaaS Cloud Services — IaaS Virtual Machines |
| Hardware |

- Code directly against this service (it's an API, and it's free)
- Let Azure Batch manage the resources & scheduling
- Use any resources (any VM instance, GPUs, RDMA/MPI, etc)
- Popular with developers, software vendors (ISVs)

- Create "Pools", submit "Jobs" made up of "Tasks"
- Mix & match, dynamically resize, include Low Pri VMs to save!

# Elasticity & Scale

What would you do with 100,000 cores? - Big compute at global scale

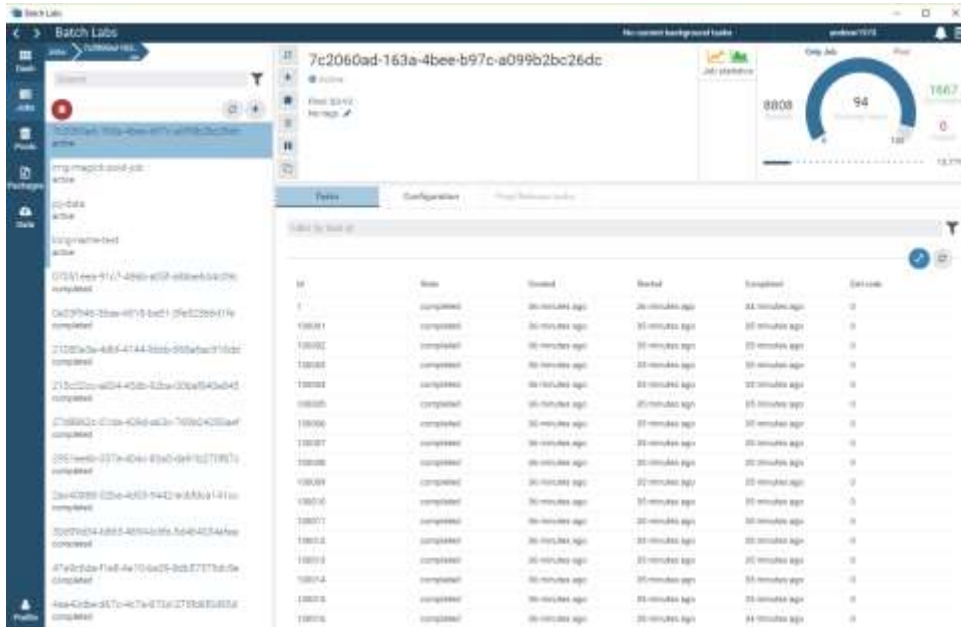https://azure.microsoft.com/en-us/blog/what-would-you-do-with-100000-cores-big-compute-at-global-scale/



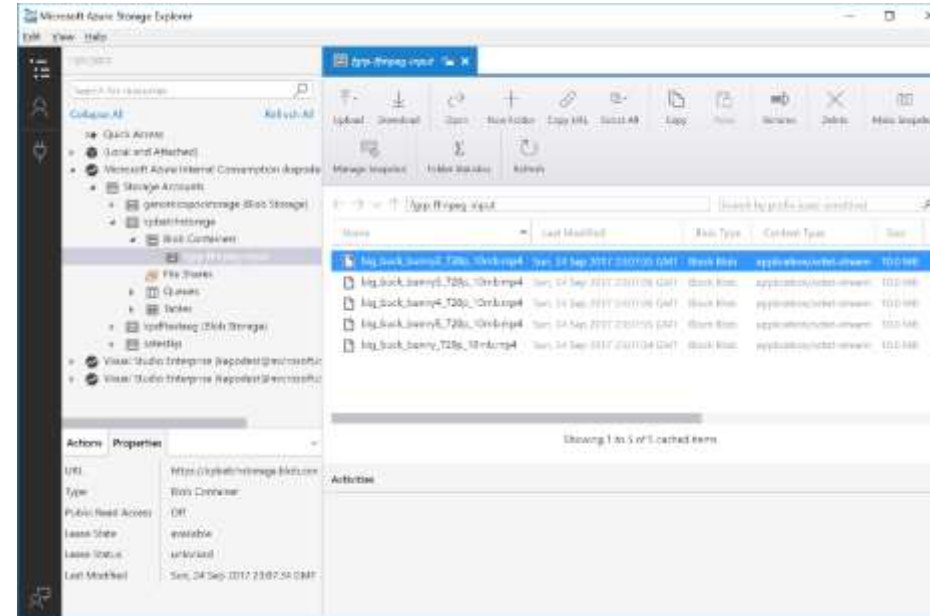https://blogs.endjin.com/2015/07/spinning-up-16000-a1-virtual-machines-on-azure-batch/

# Azure Batch – Lab setup

## Azure Batch Labs



## Azure Storage Explorer



Batch AI Lab:
https://github.com/azurebigcompute/CloudWorkshops/blob/master/BatchAIWorkshop.md

Batch with FFMPEG Lab:
https://github.com/azurebigcompute/Labs/tree/master/Azure%20Batch%20Masterclass%20Labs

# Some Links to get started!

Azure Big Compute (Labs, Solutions, Templates) - https://github.com/azurebigcompute

Azure Pricing Calculator - https://azure.microsoft.com/en-us/pricing/calculator/

Azure Documentation - https://docs.microsoft.com/en-us/azure/

Code samples - https://azure.microsoft.com/en-us/resources/samples/

Templates - https://azure.microsoft.com/en-us/resources/templates/