

# Taking High Performance Computing to the Cloud: Windows HPC and Windows Azure

Wenming Ye  
Technical Evangelist  
Microsoft Corp.

# Agenda

## ▶ Goals

- ▶ Understand on-premises HPC vs in Windows Azure
- ▶ Learn what's best suited for burst to Azure from HPC
- ▶ See how to scale up and out in the Cloud

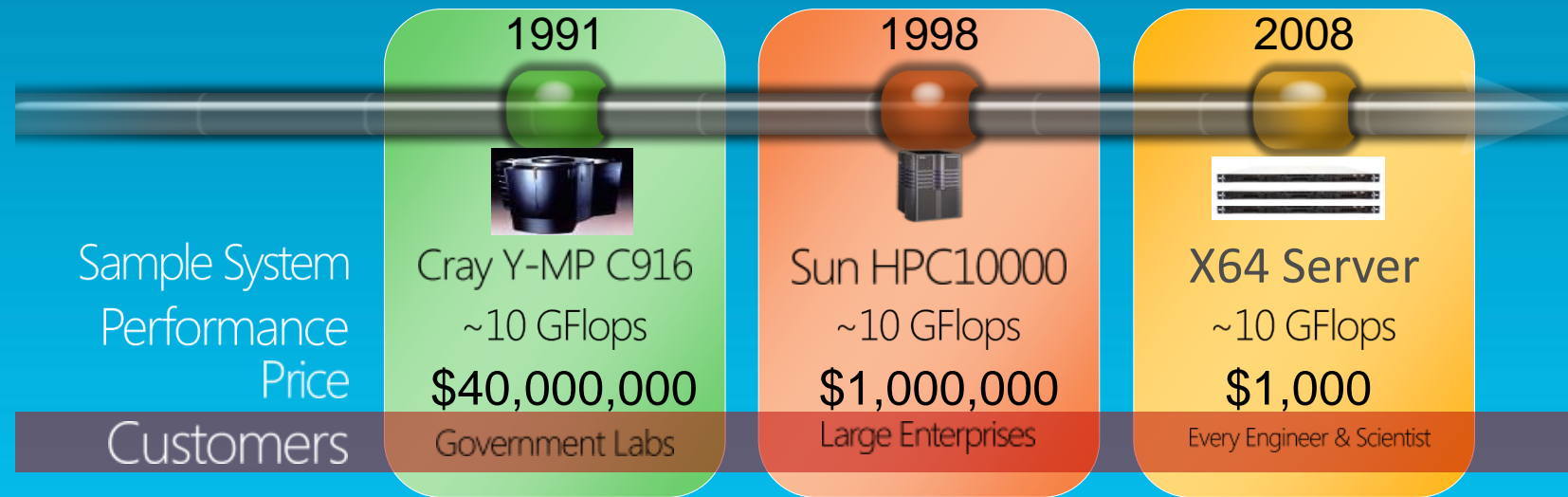
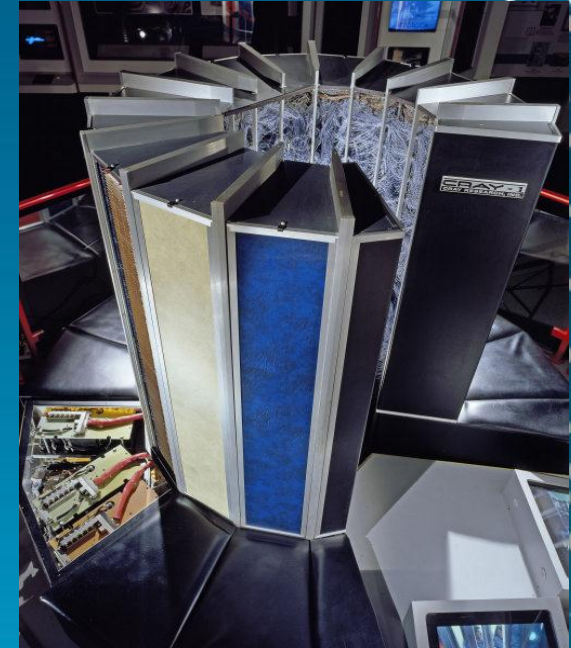
## ▶ How we'll get there

- ▶ Introduction
- ▶ Applications in Azure
- ▶ Applications on-premises
- ▶ Azure and HPC integration
- ▶ Summary

# Introduction

# What is High Performance Computing?

- ▶ Solve large computational problems
- ▶ HPC is best described as **massively parallel**
- ▶ High bandwidth, low latency, cluster-based
- ▶ Becoming **critical** part of R&D for businesses
- ▶ Traditionally required specialized staffing



# Zero-One-Infinity Goal for Computational Workload

- ▶ 0: Your Initial Investment or barrier for entry
- ▶ 1: Utilization Efficiency of your Application (100%)
- ▶  $\infty$ : Scale your Application to as many cores as possible

# Barrier to entry for technical workloads

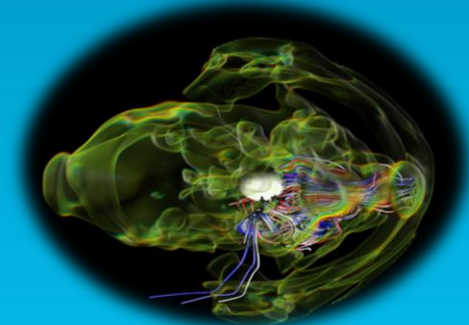
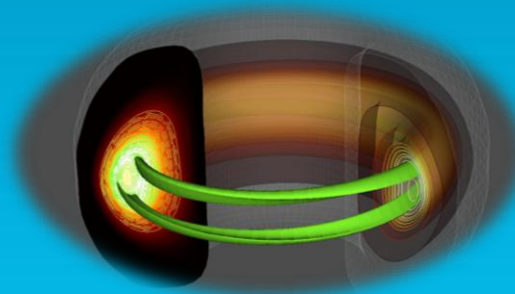
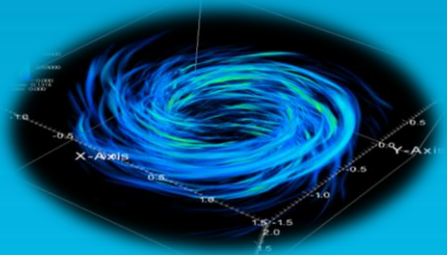
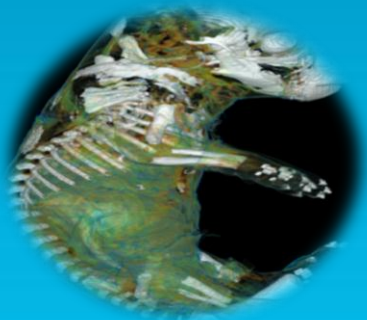
- Advances in technical computing and research are powered by access to computing resources

Fast, powerful hardware cheaper  
more readily available, but...

Windows Azure lowers the startup and  
administrative costs to almost nothing



How Do we take existing technical workloads onto azure?



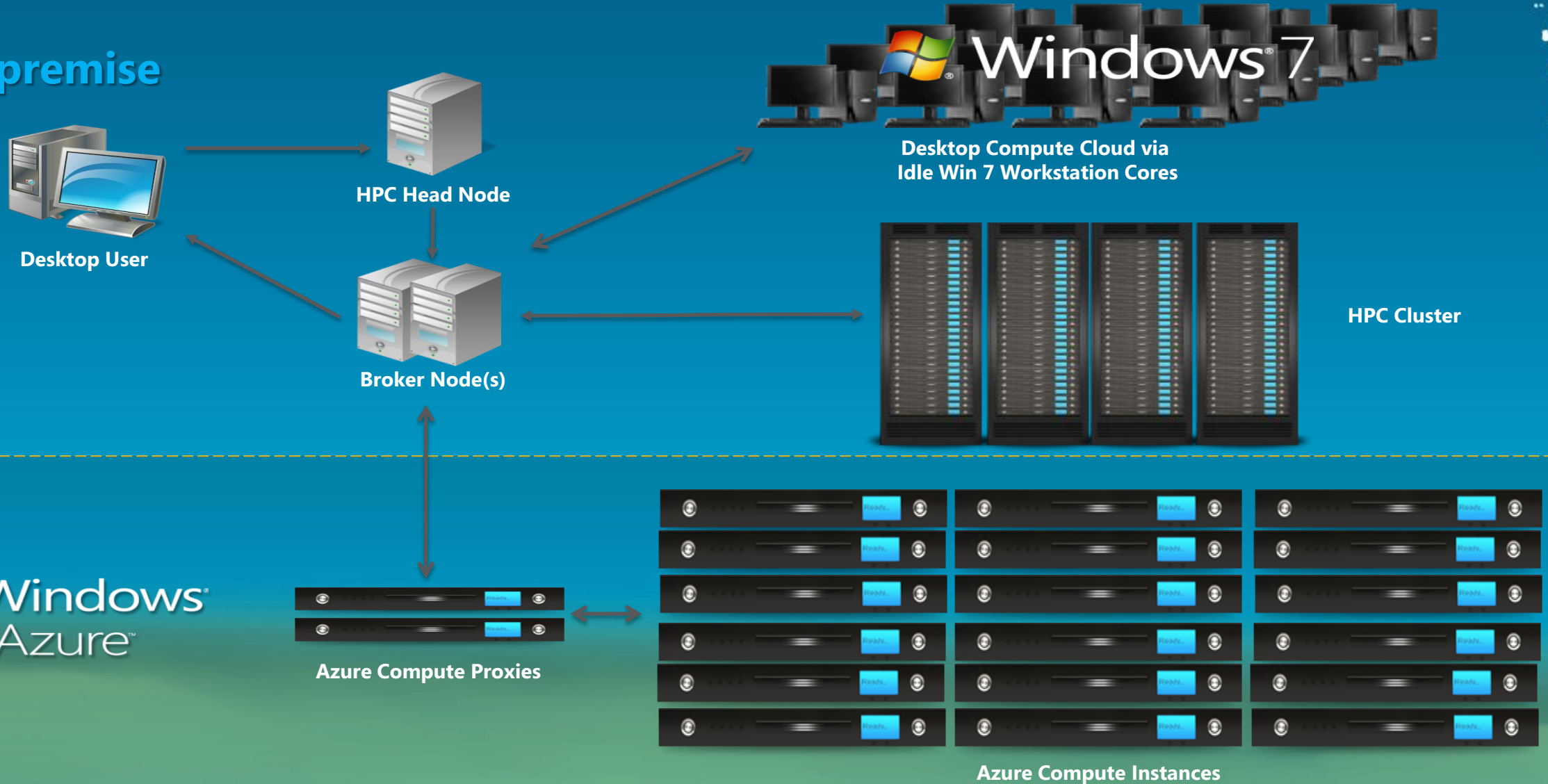
# My Cray CX1 Cluster vs. Windows Azure



	My CRAY CX1	Windows Azure Only
<b>Initial cost:</b>	Planning, Hardware, Space, Power, Cooling, \$10000s	Get an account.
<b>Cost of ownership:</b>	Maintenance, Staff, Power, Backups, Internet, Storage, Down time, Disk failures, Paying for idle computing!	Pay as you go, limited only by your budget
<b>Suitable Applications:</b>	MPI problems require high speed network. HPC Batch Scheduler System	Applications without much traffic. No HPC batch submission engine
<b>Node interconnects:</b>	20 Gbps Infini-band!	1 Gbps Ethernet
<b>As an officemate:</b>	No!	Silent and Invisible
<b>CPU &amp; Memory Hardware:</b>	Dedicated 8 nodes with 8 cores &16gb	N+1 nodes 8 cores &16gb
<b>Access to My data:</b>	LAN Connection	Internet Connection first upload LAN Connection if already in blob

# Windows HPC + Windows Azure Burst!

On-premise

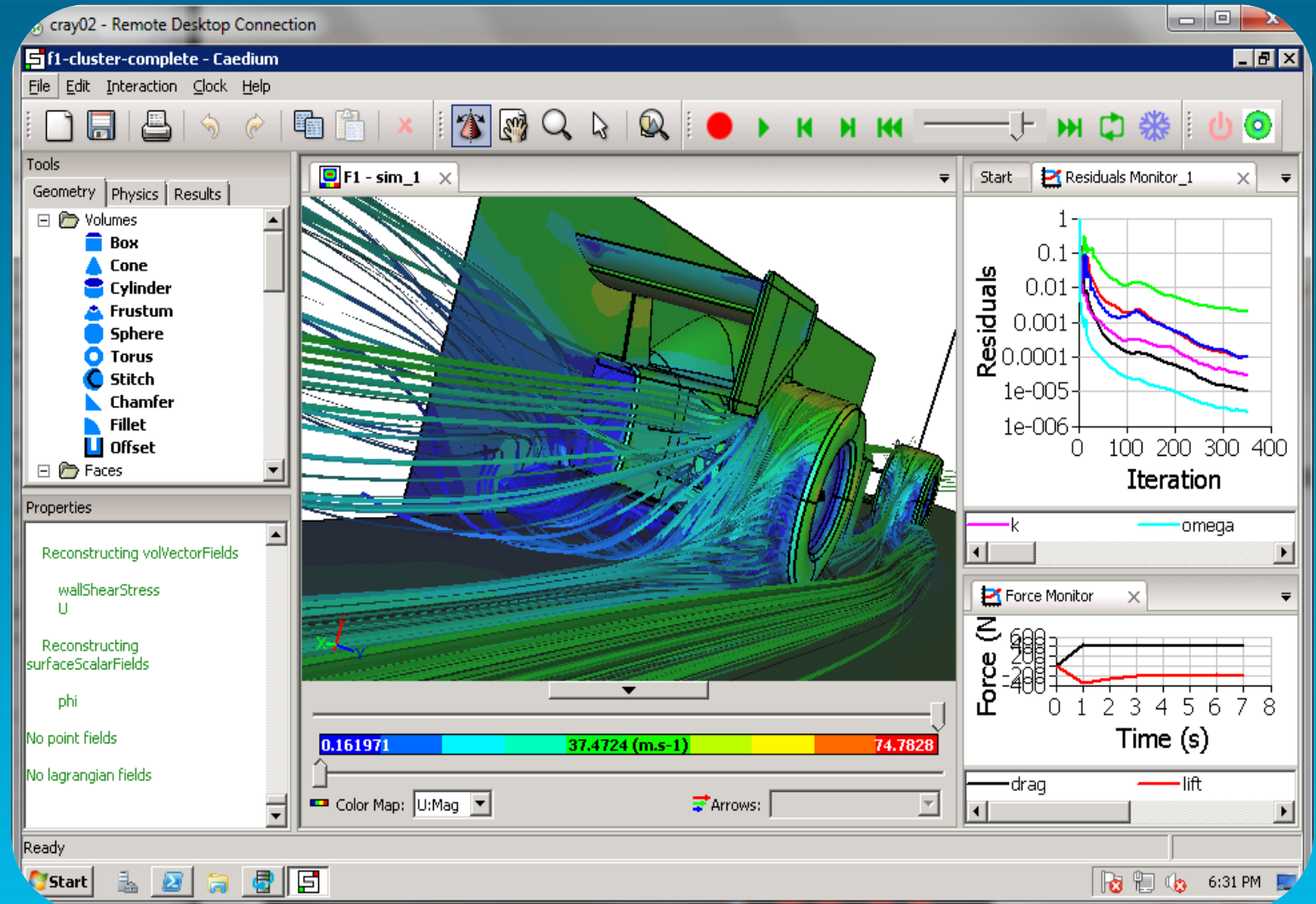
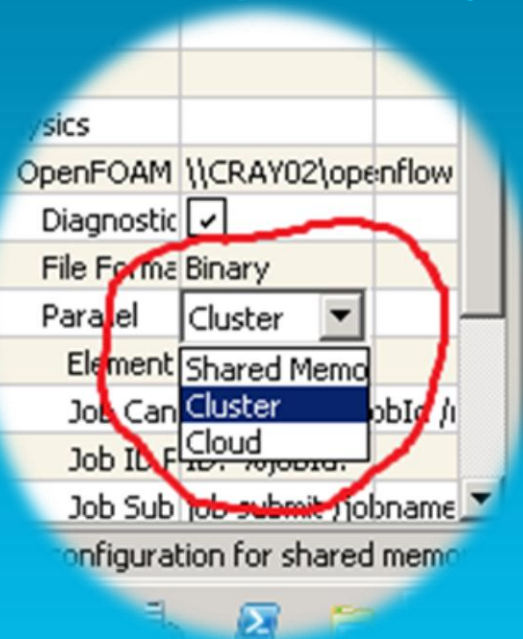


# Simplicity of Microsoft HPC Compute Intensive Paradigms

Run on:

- ▶ Client
- ▶ Cluster
- ▶ Cloud

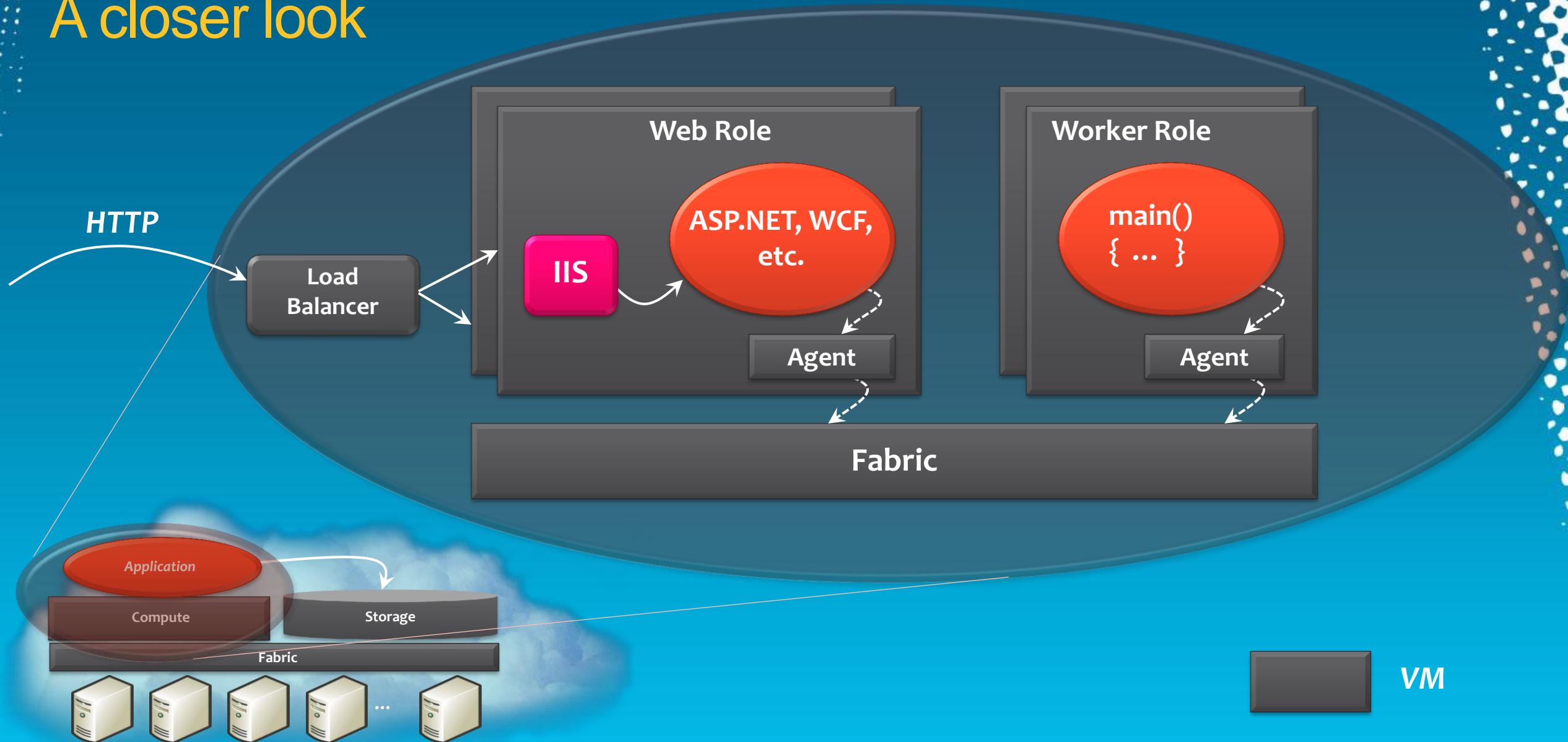
Transparently



# Applications in Azure

# Windows Azure Services

## A closer look



# TaskParallel Application: Traveling Salesman

## Resources

- ▶ One Azure instance (multi-core)

## Limitations and challenges

- ▶ Scaling beyond a single Node
- ▶ Scheduling is manual

demo

# Applications on-premises

# MPI Application: OpenFoam

## Computational fluid dynamics



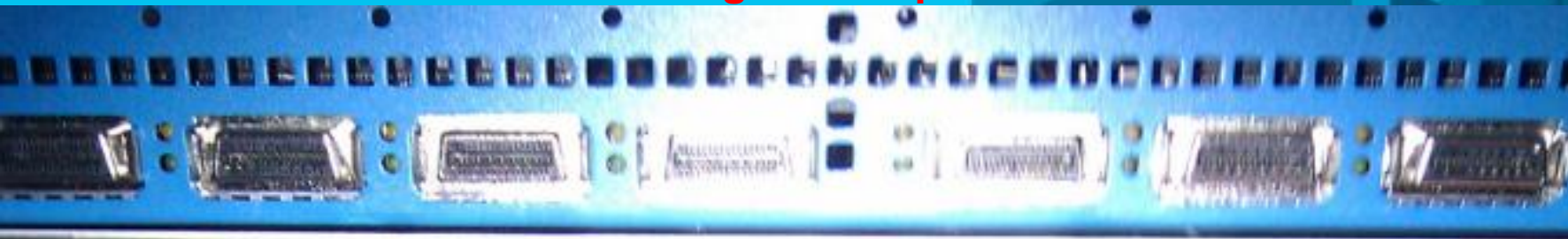
### Resources

- ▶ Dedicated nodes
- ▶ InfiniBand network

### Limitations and challenges

- ▶ Cost of ownership
- ▶ Fixed number of resources

**Infiniband running at 20Gbps!**



*demo*

# MPI Application: Weather Forecast

## Resources

- ▶ Dedicated nodes
- ▶ InfiniBand network
- ▶ External data source

## Limitations and challenges

- ▶ Cost of ownership
- ▶ Fixed number of resources



*demo*

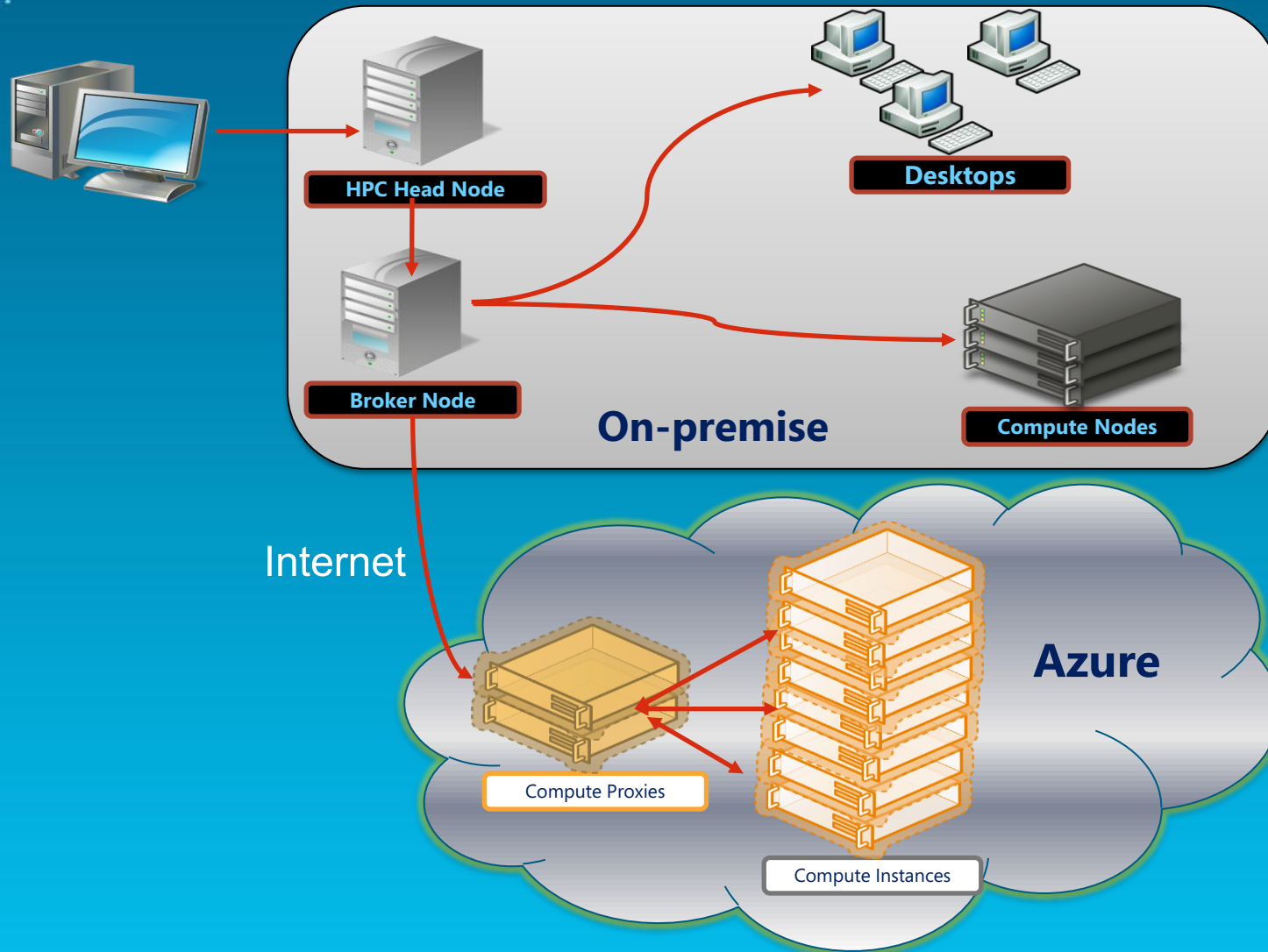
# Azure and HPC integration

- ▶ Embarrassingly parallel, no data
- ▶ Embarrassingly parallel + data
- ▶ Embarrassingly parallel + complex data
- ▶ Sharing HPC results through Azure

# Windows HPC on Azure

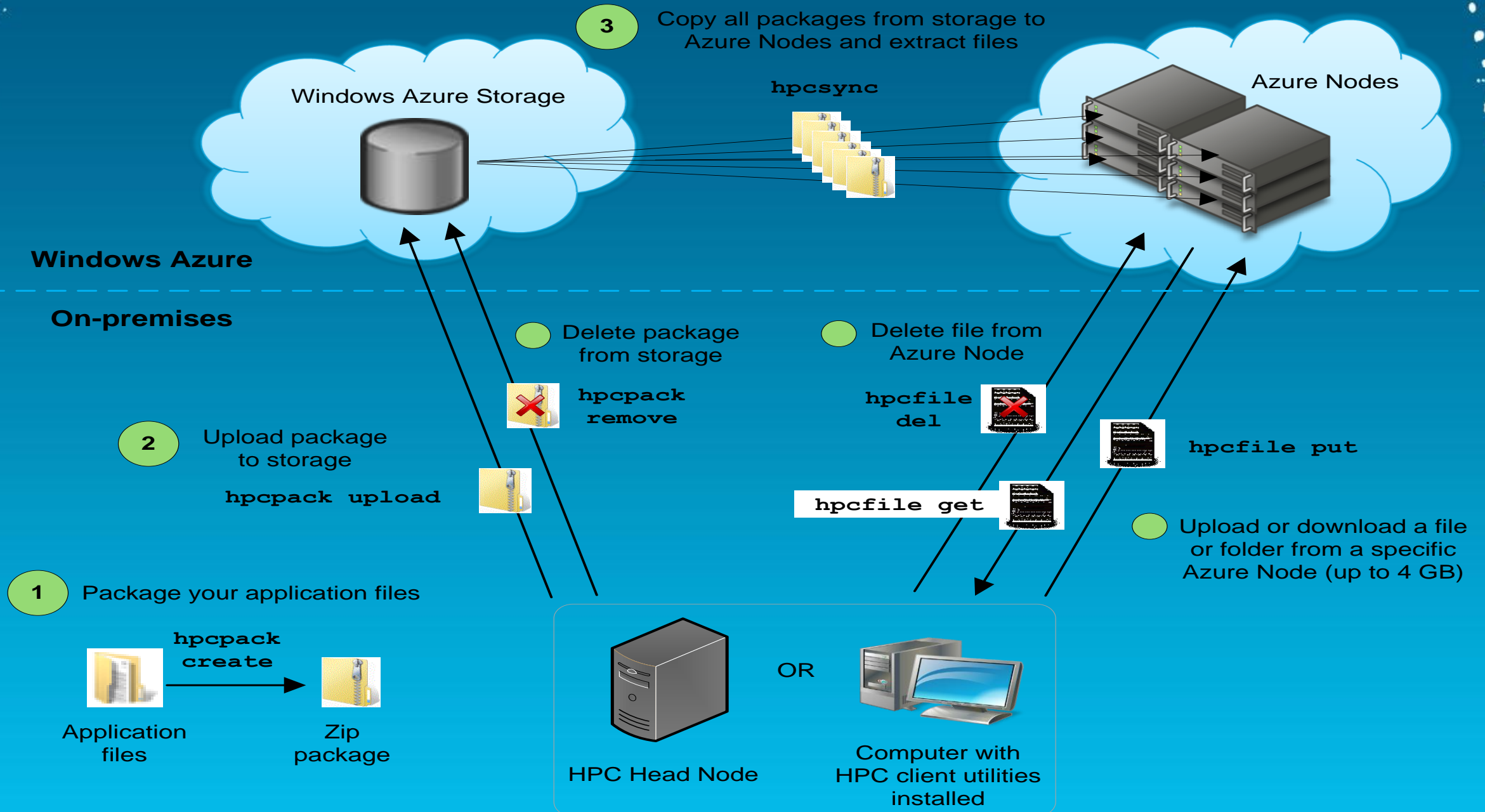
Mixed Mode Deployment On-premise and Azure

## Compute Nodes On-Premise and in Azure Simultaneously



- “Burst” into cloud on-demand while keeping control over data and corporate policies
- Pay only for what you use
- A stepping stone to hybrid and public clouds.
- Dynamically adjust how much runs on-premise and in the cloud
- Azure computer instances include both worker role and VM role

# Moving files to and from Azure Nodes in a Windows HPC 2008 R2 cluster



# SOA app: Asian Options

Embarrassingly parallel, no data

## Work

- ▶ No code change!

## Resources

- ▶ Azure & Dedicated Node instances
- ▶ HPC head node

## Benefits

- ▶ Flexible resources
- ▶ Scaling and scheduling handled by HPC

*demo*

# Parametric app: Movie Rendering

## Embarrassingly parallel + data

### Work

- ▶ Data staging

### Resources

- ▶ All Azure Node instances
- ▶ HPC head node

### Benefits

- ▶ Flexible resources
- ▶ Scaling and scheduling handled by HPC

*demo*

# An example : Better Faster Product Design

## Shape Optimization with Sculptor

- ▶ Cut design times
- ▶ Cut design costs
- ▶ Provide better products
- ▶ Geometric optimization
- ▶ New trend in design

NASA

TOSHIBA



*DENSO*

**HONDA**

speedo 

# Latin HyperCube – compute what you can afford

The screenshot displays the Optimization Control Center software interface. The main window shows a table of experimental results for a Latin HyperCube design. The table has columns for Experiment (Exp), State, Type, Index, and various Design Variables (DV1, DV2, DV3, DV4, DV5, DV6, DV7, DV8, DV9). The table lists 24 experiments, each with a unique index and corresponding DV values. A smaller inset window shows a detailed view of the Analysis Variables tab, listing the DVs and their values, minimums, and maximums.

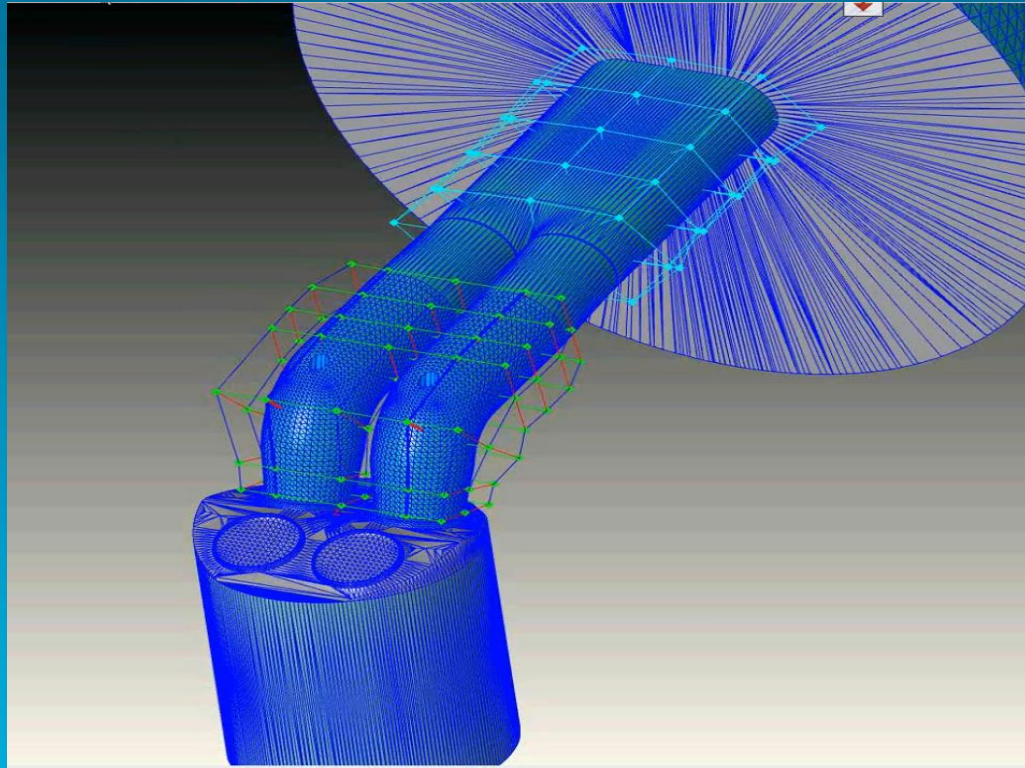
Exp	State	Type	Index	DV1	DV2	DV3	DV4	DV5	DV6	DV7	DV8	DV9
0	0	OLH_000	20	0	0.000252525	0	0	-0.00146465	0	-0.00378788	0.00479798	0
1	0	OLH_001	21	0	-0.000353535	0	0	0.000151515	0	0.00247475	0.00489899	0
2	0	OLH_002	22	0	0.002777778	0	0	0.00267677	0	-0.00489899	0.000757576	0
3	0	OLH_003	23	0	-0.000252525	0	0	0.00388889	0	-0.000555556	-0.005	0
4	0	OLH_004	24	0	-0.00368687	0	0	0.00186869	0	0.00308081	0.00419192	0
5	0	OLH_005	25	0	-0.00419192	0	0	0.00207071	0	0.00136364	0.00126263	0
6	0	OLH_006	26	0	-0.00116162	0	0	0.0019697	0	-0.00318182	-5.05051e-05	0
7	0	OLH_007	27	0	0.000858586	0	0	0.00227273	0	-0.00409091	0.0039899	0
8	0	OLH_008	28	0	0.00439394	0	0	0.000454545	0	-0.00308081	-0.00176768	0
9	0	OLH_009	29	0	0.00237374	0	0	5.05051e-05	0	-0.00166667	0.005	0
10	0	OLH_010	30	0	-0.00449495	0	0	-0.00419192	0	-0.00146465	0.000454545	0
11	0	OLH_011	31	0	-0.00459596	0	0	-0.00126263	0	0.00378788	0.00318182	0
12	0	OLH_012	32	0	0.00368687	0	0	0.00116162	0	-0.000454545	-0.00388889	0
13	0	OLH_013	33	0	-0.00247475	0	0	-0.000959596	0	0.00449495	-0.00257576	0
14	0	OLH_014	34	0	0.00308081	0	0	-0.00156566	0	-0.00429293	0.00257576	0
15	0	OLH_015	35	0	0.00449495	0	0	0.00318182	0	0.000252525	-0.000151515	0
16	0	OLH_016	36	0	0.000656566	0	0	0.00419192	0	-0.00207071	0.00338384	0
17	0	OLH_017	37	0	-0.005	0	0	-0.000151515	0	-0.00106061	-0.00106061	0
18	0	OLH_018	38	0	0.00176768	0	0	0.00489899	0	0.00106061	0.00207071	0
19	0	OLH_019	39	0	0.00166667	0	0	-0.0039899	0	0.000454545	0.00378788	0
20	0	OLH_020	40	0	0.00388889	0	0	0.00439394	0	0.0019697	-0.00338384	0
21	0	OLH_021	41	0	5.05051e-05	0	0	-0.00358586	0	0.00429293	0.00106061	0
22	0	OLH_022	42	0	0.0039899	0	0	-0.000757576	0	0.0039899	0.00237374	0
23	0	OLH_023	43	0	0.00126263	0	0	-0.000454545	0	-0.00479798	-0.00186869	0
24	0	OLH_024	44	0	0.00186869	0	0	0.00156566	0	0.00419192	0.00368687	0

#	ID/DV	Value	Minimum	Maximum
0	<input type="checkbox"/> 1) DV1	0	-0.005	0.005
1	<input checked="" type="checkbox"/> 1) DV2	0.003421	-0.005	0.005
2	<input type="checkbox"/> 1) DV3	0	-0.005	0.005
3	<input type="checkbox"/> 1) DV6	0	-0.005	0.005
4	<input checked="" type="checkbox"/> 1) DV5	0.001316	-0.005	0.005
5	<input type="checkbox"/> 1) DV4	0	-0.005	0.005
6	<input checked="" type="checkbox"/> 2) DV7	-0.002368	-0.005	0.005
7	<input checked="" type="checkbox"/> 2) DV8	-0.003947	-0.005	0.005
8	<input type="checkbox"/> 2) DV9	0	-0.005	0.005

- Latin hyperCube sampling
- ▶ A Statistical method
- ▶ Know your cost up front
- ▶ Know you time to solution
- ▶ More design insight
- ▶ Best statistical accuracy

# Automotive Intake Manifold Design

- ▶ 3% performance enhancement vs. over 60 humans working for months to get less than 0.6%.



- ▶ Change shape of intake
- ▶ Improve flows
- ▶ Better gas mileage
- ▶ Better performance

Courtesy of Ford Motor Company

# Sharing results: Earthquake simulation

MPI on-premises + data shared in Azure

## Work

- ▶ Transfer data to SQL Azure
- ▶ Visualization app

## Resources

- ▶ HPC cluster
- ▶ SQL Azure
- ▶ Visualization app hosted in Azure

## Benefits

- ▶ Share results



demo

# Summary

# Nine Tips for Best Practices

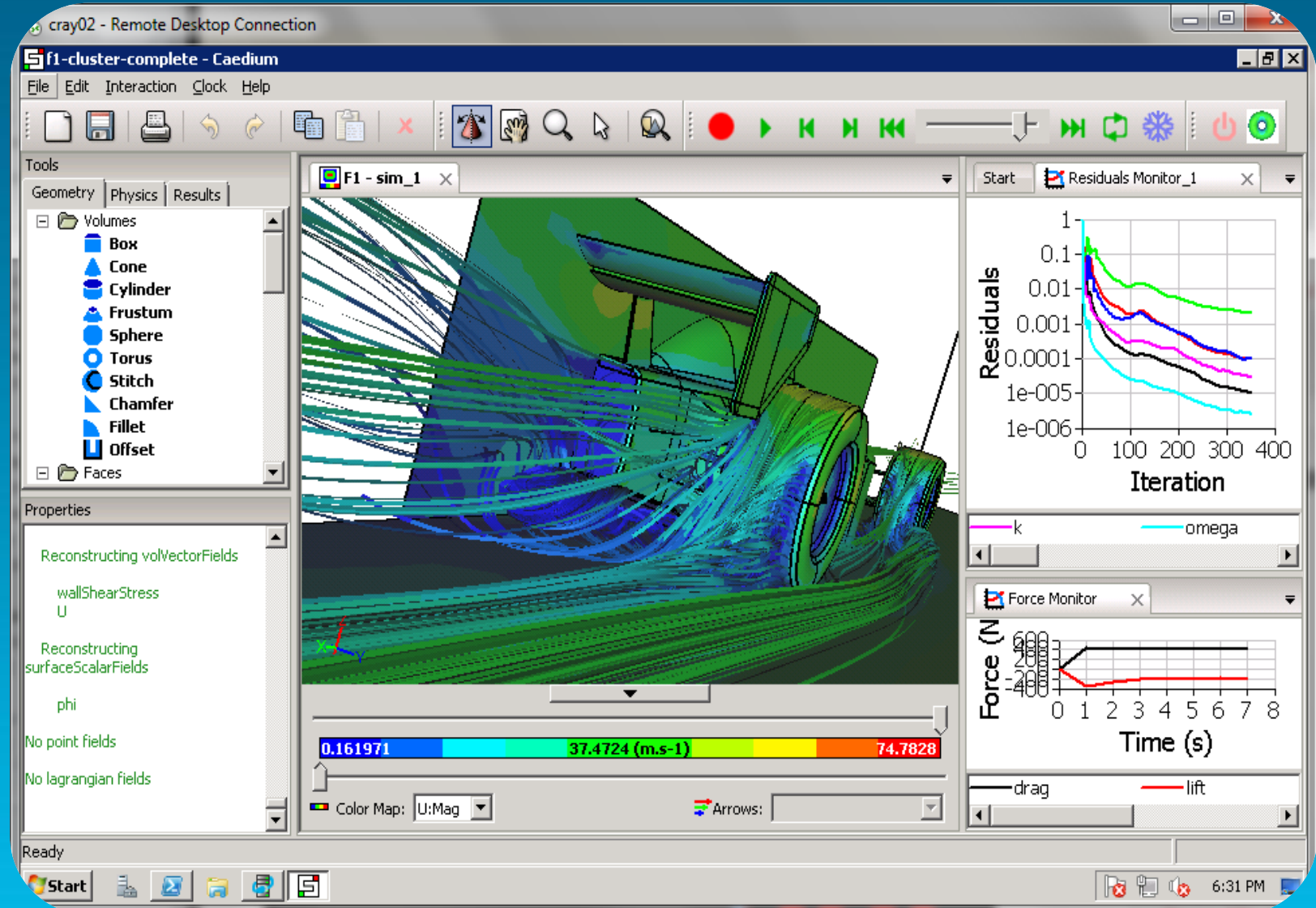
- ▶ Always consider total cost of ownership when making decisions
- ▶ Use .Net Parallel Libraries to take advantage of multi-core on Azure
- ▶ Use HPC Azure burst solution for scaling out to multi-node
- ▶ Embarrassingly parallel CPU intensive jobs are best suited for Azure
- ▶ Parametric sweep engine is a powerful yet simple way to scale
- ▶ Have multi-user design considerations for HPC Apps as a service
- ▶ Hide cloud from your users by providing a transparent experience
- ▶ Minimize data transfer between on-premises and the cloud
- ▶ Be creative and focus on what you can do on Azure today

# The Microsoft Windows Computing Experience

Run on:

- ▶ Client
- ▶ Cluster
- ▶ Cloud

Transparently

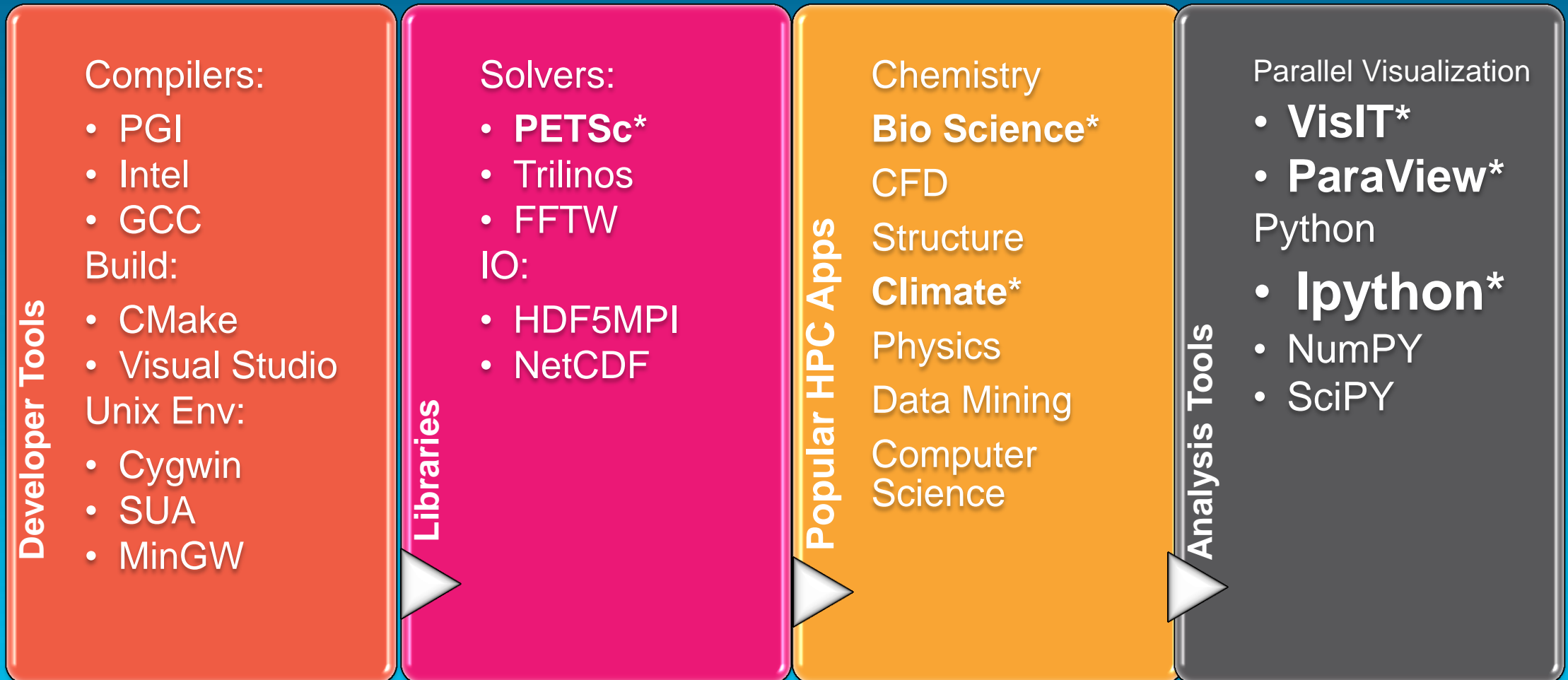


Physics  
OpenFOAM \\CRAY02\openflow  
Diagnostic ☒  
File Format Binary  
Parallel Cluster  
Element Shared Memo  
Job Can Cluster  
Job ID F  
Job Sub job submit jobname  
configuration for shared memo

# Developer Benefits of Windows

- ▶ Easier binary distribution
  - ▶ Compile once, run everywhere
  - ▶ from laptop, to desktop, to Windows HPC, to Windows Azure
- ▶ Better tools, first class developer support
  - ▶ Visual Studio and other tools make development easier
- ▶ More potential new users and developers
  - ▶ Run on World's most popular OS
  - ▶ Reveal issues by exercising code on commercial platform

# Building a Healthy App Ecosystem to Support HPC and Azure



# HPC Server 2008 R2 Roadmap



July 2010

Dec 2010

June 2011

## Microsoft HPC Server 2008 R2

- Excel on HPC clusters
- Schedule jobs to workstations
- Business critical SOA
- Scale to 1000+ nodes out-of-the-box
- Dual boot clusters
- Diskless compute nodes
- Extensible diagnostics and troubleshooting framework
- Flexible, customizable color-coded Heat map

## Microsoft HPC Server 2008 R2 SP1 & LINQtoHPC CTP

- Integration with Windows Azure:
  - “Burst” SOA and Parametric sweep jobs to Azure as worker roles.
  - All compute nodes in Azure with head-node only on-premise.
- User activity scheduling for workstation nodes
- C# cluster SOA debugger for VS 2010.

## Microsoft HPC Server 2008 R2 SP2\*

- LINQ to HPC for large scale data analytics
- MPI runtime in Windows Azure
- Azure VM role
- New job scheduler policies
  - Lend/Borrow resource pools
  - Over/under subscribe nodes
- Common data for SOA Apps
- Multi-domain support for workstation nodes
- Web Portal for job submission and job status
- Smart Card authentication

# The Matrix can not be told, **try** it yourself!

- ▶ **Subscribe** to my blog <http://blogs.msdn.com/hpctrekker>
- ▶ **Read** [Application Models and Data Considerations](#)
- ▶ Download SP2 Beta from <http://connect.microsoft.com>
- ▶ Get a [free Trial of Azure for 30 days](#)
- ▶ Download and [try out the Azure Burst Samples](#)
- ▶ Implement your own solution and let us know!

# Track Resources

▶ My Blog: <http://blogs.msdn.com/hpctrekker>

▶ [The Azure Burst Samples and White paper](#)

▶ [Windows HPC ResourceKit](#)

▶ Windows HPC Server [TechCenter](#)

# Track Resources

- ▶ Cloud Power - <http://www.microsoft.com/cloud/>
- ▶ Private Cloud - <http://www.microsoft.com/privatecloud/>
- ▶ Windows Server - <http://www.microsoft.com/windowsserver/>
- ▶ Windows Azure - <http://www.microsoft.com/windowsazure/>
- ▶ Microsoft System Center - <http://www.microsoft.com/systemcenter/>
- ▶ Microsoft Forefront - <http://www.microsoft.com/forefront/>

The Microsoft logo is centered on a blue background. The logo consists of the word "Microsoft" in a white, bold, sans-serif font, followed by a registered trademark symbol (®). The background features a blue gradient with a white, pixelated, abstract shape on the left side and a geometric pattern of blue triangles on the bottom right.

© 2011 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.