Microsoft Intelligent Cloud

Ulrich (Uli) Homann

Distinguished Architect

Microsoft Cloud and Enterprise Engineering

ulrichh@microsoft.com

Why Cloud for transformation?

EFFICIENCY

Fast time to value
Pay for what you need
Hybrid scenarios

+

AGILITY

Speed
Flexibility
Elastic scale

+

DIFFERENTIATION

Unique customer experiences
New business models
IoT & feedback loop scenarios

500+

New releases in the last 12 months

Azure IoT Suite

Azure Site Recovery: Protect VMWare and Physical Servers in Public Preview

Azure Backup Generally Available

Azure API Management Premium simplifies high availability and massive scale for APIs

ExpressRoute for Office 365

Azure Active Directory Dynamic Membership For Groups

Automatic Password Change for Social Media Shared Accounts

Compute-Intensive A10 and A11 Virtual Machine Instances

Remote Desktop app for Windows Phone support for Gateway and Remote Resources

Informatica Cloud Agent availability in Linux and Windows Virtual Machines

Azure DocumentDB Hadoop Connector

Azure HDInsight support for more VM sizes

>90,000

New Azure customer subscriptions/month

>1.5 Million

SQL Databases running on Azure

4 out of 5

of the Fortune 500 have Office 365

>80%

Of Fortune 500 Companies Use Microsoft Cloud

2 Trillion

Messages per week processed by Azure IoT

90 Trillion

Storage Objects in Azure

>551 Million

Users in Azure Active Directory

>40%

Revenue from Start-ups and ISVs

Hyper-scale footprint

28

CLOUD REGIONS

Cloud regions around the world

Platform Services

Security & Management

- Portal
- Azure Active Directory
- Azure AD B2C
- Multi-Factor Authentication
- Automation
- Scheduler
- Key Vault
- Store/ Marketplace
- VM Image Gallery & VM Depot

Services Compute

- Cloud Services
- Service Fabric
- Batch
- RemoteApp

Integration

- Storage Queues
- BizTalk Services
- Hybrid Connections
- Service Bus

Media & CDN

- Media Services
- Content Delivery Network (CDN)

Web and Mobile

- Web Apps
- API Apps
- Mobile Apps
- Logic Apps
- API Management
- Notification Hubs

Developer Services

- Visual Studio
- Azure SDK
- VS Online
- App Insights

Data

- SQL Database
- Data Warehouse
- DocumentDB
- Redis Cache
- Azure Search
- Storage Tables

Analytics & IoT

- HDInsight
- Machine Learning
- Stream Analytics
- Data Lake
- Data Factory
- Event Hubs
- Data Catalog
- IoT Hub
- Mobile Engagement

Hybrid Operations

- Azure AD Health Monitoring
- AD Privileged Identity Management
- Domain Services
- Backup
- Operational Analytics
- Import/Export
- Azure Site Recovery
- StorSimple

Infrastructure Services

OS/Server Compute

- Virtual Machines
- Container Service

Storage

- BLOB Storage
- Azure Files
- Premium Storage

Networking

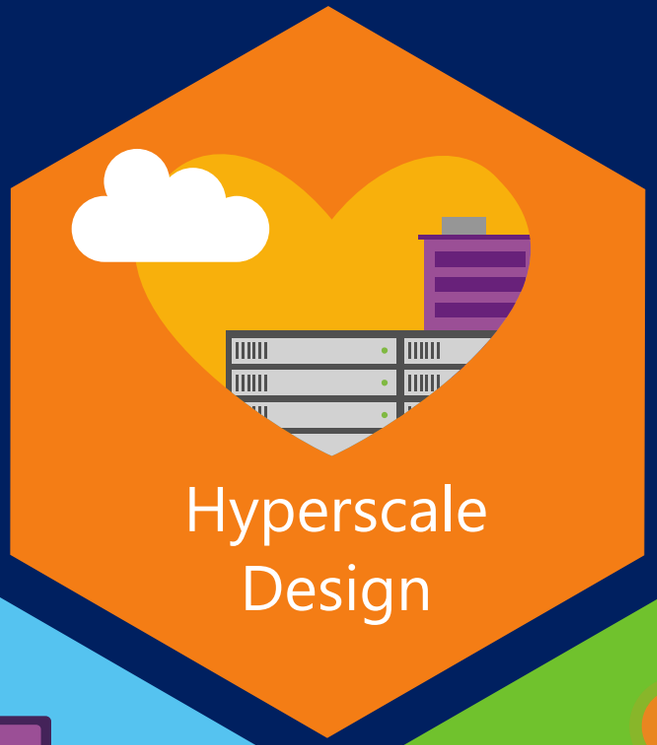
- Virtual Network
- Load Balancer
- DNS
- Express Route
- Traffic Manager
- VPN Gateway
- App Gateway

Datacenter Infrastructure (28 Regions)

Microsoft Azure - Big Compute

http://azure.microsoft.com/solutions/**big-compute**

Big Compute Philosophy



Managing risk and compliance

Large European bank optimized compute by moving to the cloud

Challenge

- Pursue innovation internally and within the market.
- Reduce compute costs while continuing growth.

Strategy

- Dynamic clusters in Azure for compute grids
- Azure Batch for new application

Results

- More efficient risk and compliance infrastructure
- Increased agility for dynamic markets
- Reduced cost

Molecular Modeling in Microsoft Azure

By one of the largest chemical companies in the world

Challenge

- Reduce compute costs while continuing new product development
- Custom HPC applications that require low-latency network

Strategy

- Lift and shift cluster and applications to Azure
- Reduce time to solution by running a second instance

Results

- Greater ability to meet tight deadlines
- Make best use of in-house compute resources

Reservoir Simulation as a Service

By the largest oilfield services company in the world

Challenge

- Serve customers with zero datacenter investment strategy
- Be first to market with cloud service
- Support larger models, new software, and optimization
- Need for specialized hardware and tight security

Strategy

- Validate Azure HPC scalability and performance
- Start with managed clusters
- Prove it to customers

Results

- Engineer working with Petrel not aware they're using cloud
- Capabilities are available on a subscription-based model

Schlumberger

Simulation and Analytics in Azure

Business transformation in Motorsports

Challenge

- Make better decisions on car setup and race strategy through simulation and data analysis
- Desktop to garage to racetrack

Strategy

- CFD in Azure to model airflow
- Azure Machine Learning and Cortana Analytics for race strategy

Results

- CFD results in half the time
- Highly accurate models for fuel and tires
- Tools for the simulator and at race day

Flexibility and Choice

Microsoft ♥ Linux

Job Scheduler Agnostic

- Commercial: Altair PBS, Univa Grid Engine, IBM LSF, Adaptive Computing Moab
- Open Source: SLURM, SGE, Torque
- Financial Services: Tibco Data Synapse, IBM Platform Symphony
- Microsoft Hybrid: Microsoft HPC Pack (Windows and also now Linux)

- Deploy via template (see our samples)
- Connect network with VPN or Express Route
- Scripts to scale up and down

- Partner tools: Cycle Computing, UberCloud, Elastacluster, Altair, Bright

HPC Scaling in Azure

Compute Intensive VM's: A8-9 (A10-11)

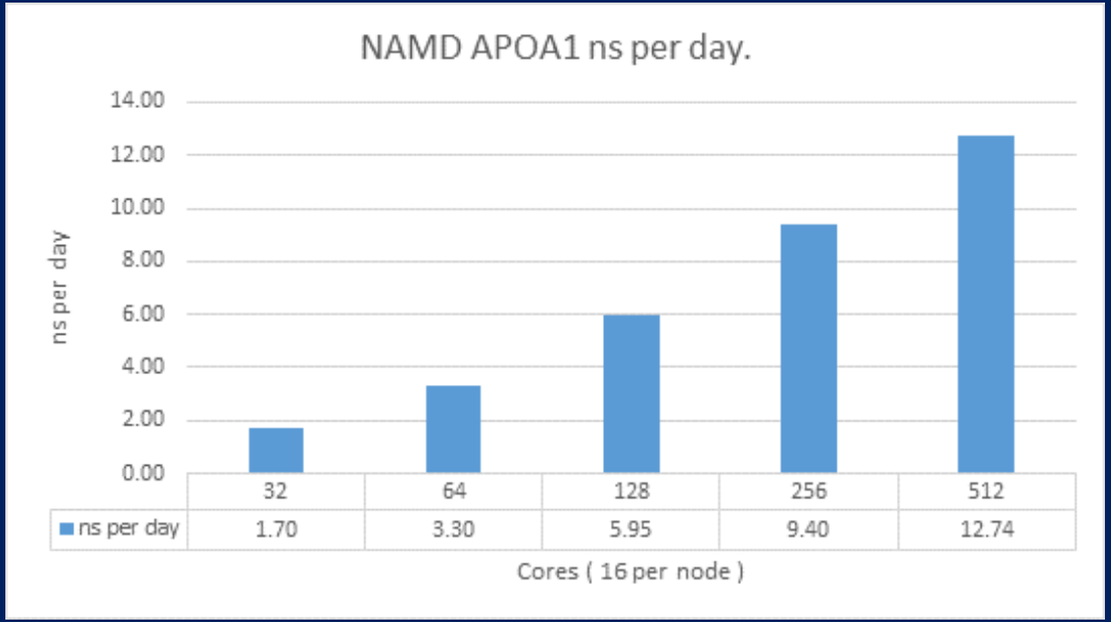
Hardware designed for HPC

- High CPU: 2x8 core processors per node, Sandybridge E-2670 at **2.6 GHz**
- High Memory: **128 GB**, 1600MHz DDR3
- Fast Interconnect: **QDR InfiniBand** for intra deployment traffic, 10gigE for standard Azure traffic and internet access
- Scratch storage: 2 TB per node
- Available in 8 core/56 GB and 16 core/112 GB instances
- RDMA for Linux and Windows
- Available in 7 regions

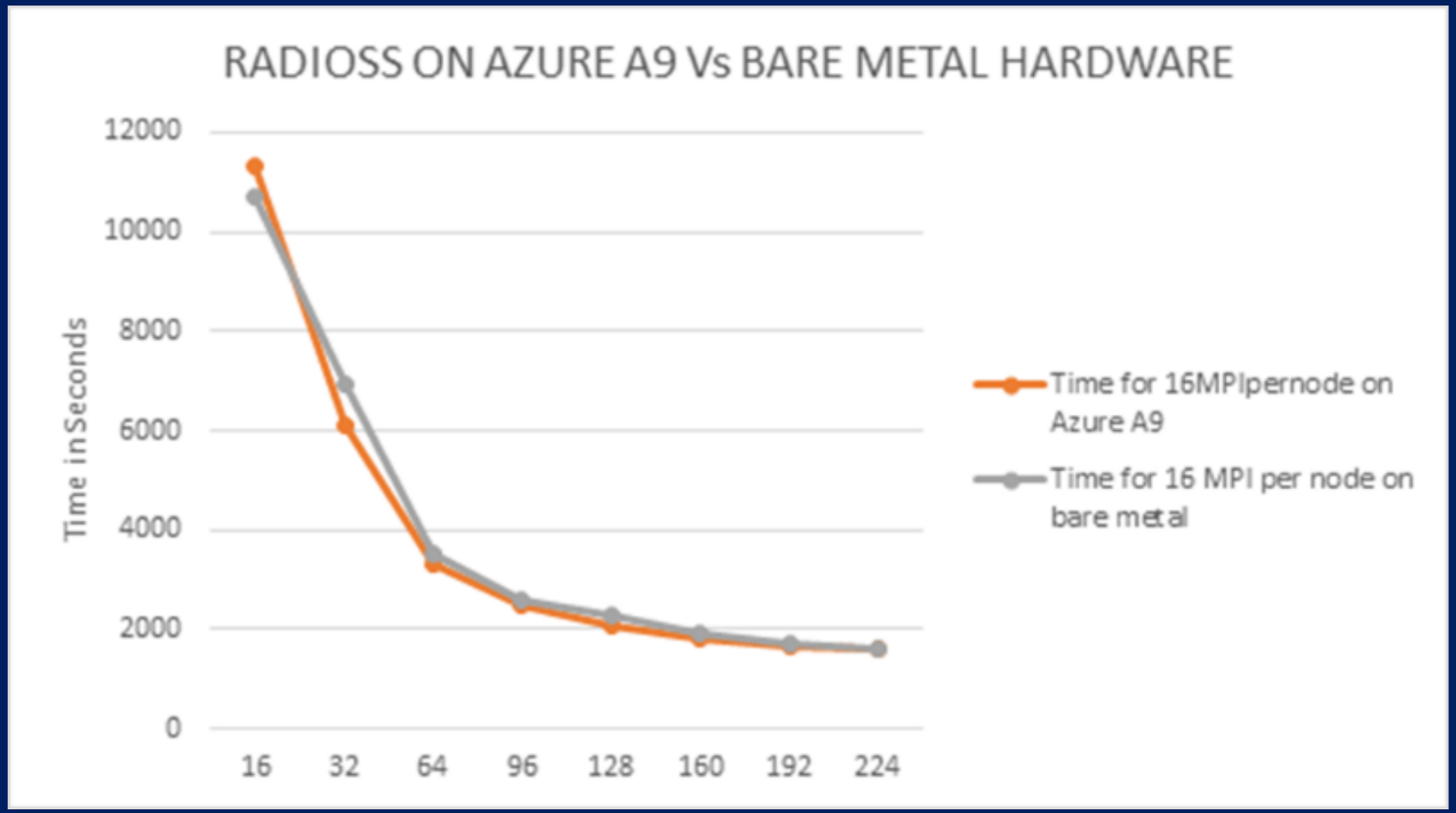
Next generation coming soon

Bare Metal Equivalent Performance

- ~2.5-3.1 microsecond latency
- >3GB/sec non blocking
- 90% efficiency on Linpack
- Example: linear scaling on NAMD

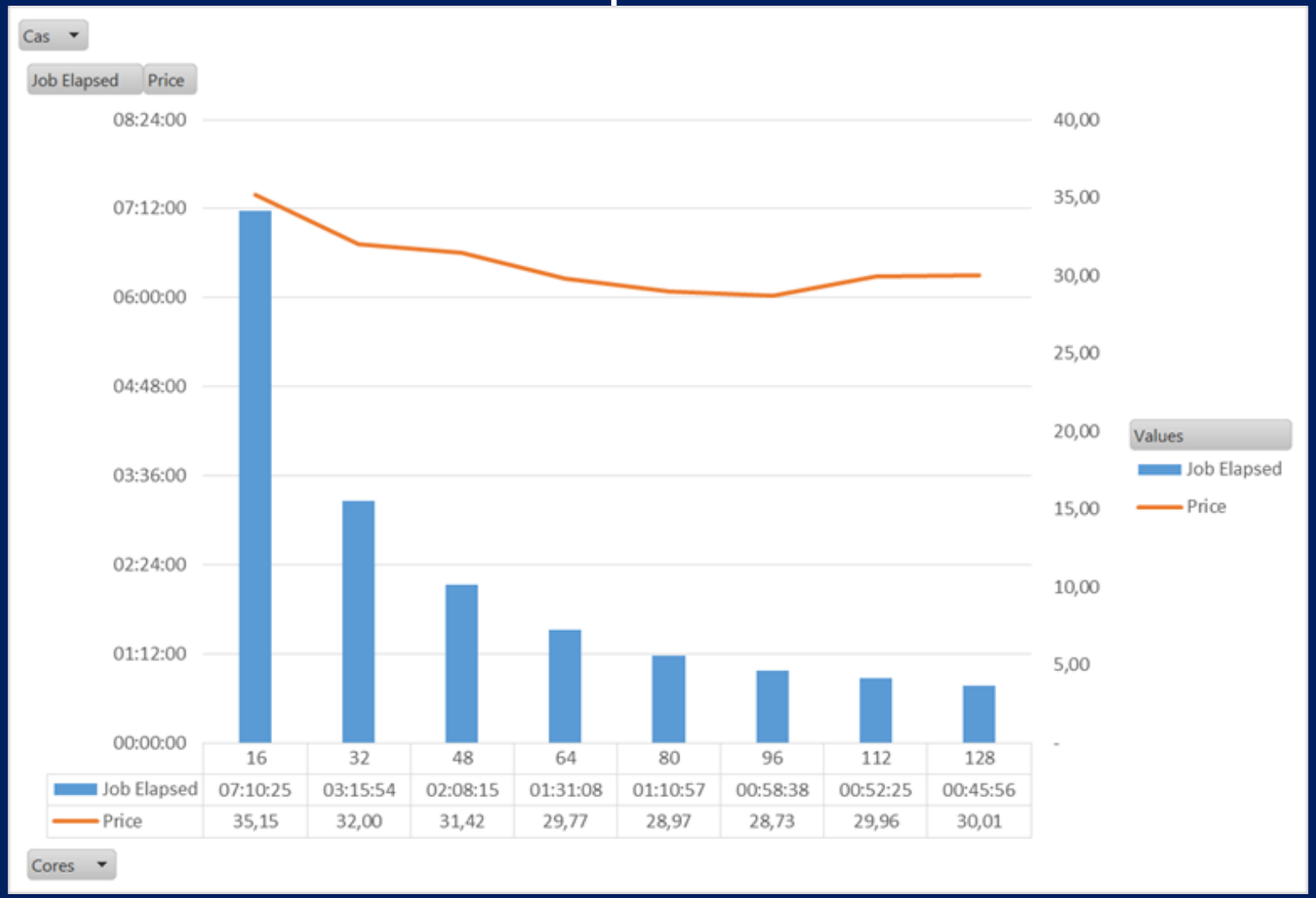


No Penalty for Virtualization



https://azure.microsoft.com/en-us/blog/availability-of-altair-radioss-rdma-on-microsoft-azure/

Customer Impact



	16	32	48	64	80	96	112	128
Job Elapsed	07:10:25	03:15:54	02:08:15	01:31:08	01:10:57	00:58:38	00:52:25	00:45:56
Price	35,15	32,00	31,42	29,77	28,97	28,73	29,96	30,01

- Customer CFD model run on 1-8 A9 VMs using RDMA
- Very good scalability: Shrink compute time from 430 min to 45 min at same cost for compute
- Elapsed time includes fetch input from blob, un-compress, run solver, compress results, write blob
- Many customers don't have 256 spare capacity available
- ISV licensing is a challenge—per core or per hour?

GPUs to Complete the End-to-End

- Accelerated Desktop Applications (OpenGL and DirectX)

- Cloud-based Streaming and Gaming
- Video Processing / Encoding Workloads

- GPU Compute (CUDA and OpenCL) - single and multiple machine workloads
- Computational Network Toolkit (CNTK)

GPU VM Offerings (N-Series)

Size/Component	N1	N3	N11	N12	N13	N23
CPU Cores (E5-2690v3)	6	24	6	12	24	24
RAM	56 GB	224 GB	56 GB	112 GB	224 GB	224 GB
SSD	~0.5 TB	~2.0 TB	~0.5 TB	~1.0 TB	~2.0 TB	~2.0 TB
Network	Azure Network	Azure Network	Azure Network	Azure Network	Azure Network	Azure Network + Dedicated RDMA Backend
GPU Resources	1 x M60 GPU (1/2 Physical Card)	4 x M60 GPU (2 Physical Cards)	1 x K80 GPU (1/2 Physical Card)	2 x K80 GPUs (1 Physical Card)	4 x K80 GPUs (2 Physical Cards)	4 x K80 GPUs (2 Physical Cards)

Azure Batch

Cloud-enable applications

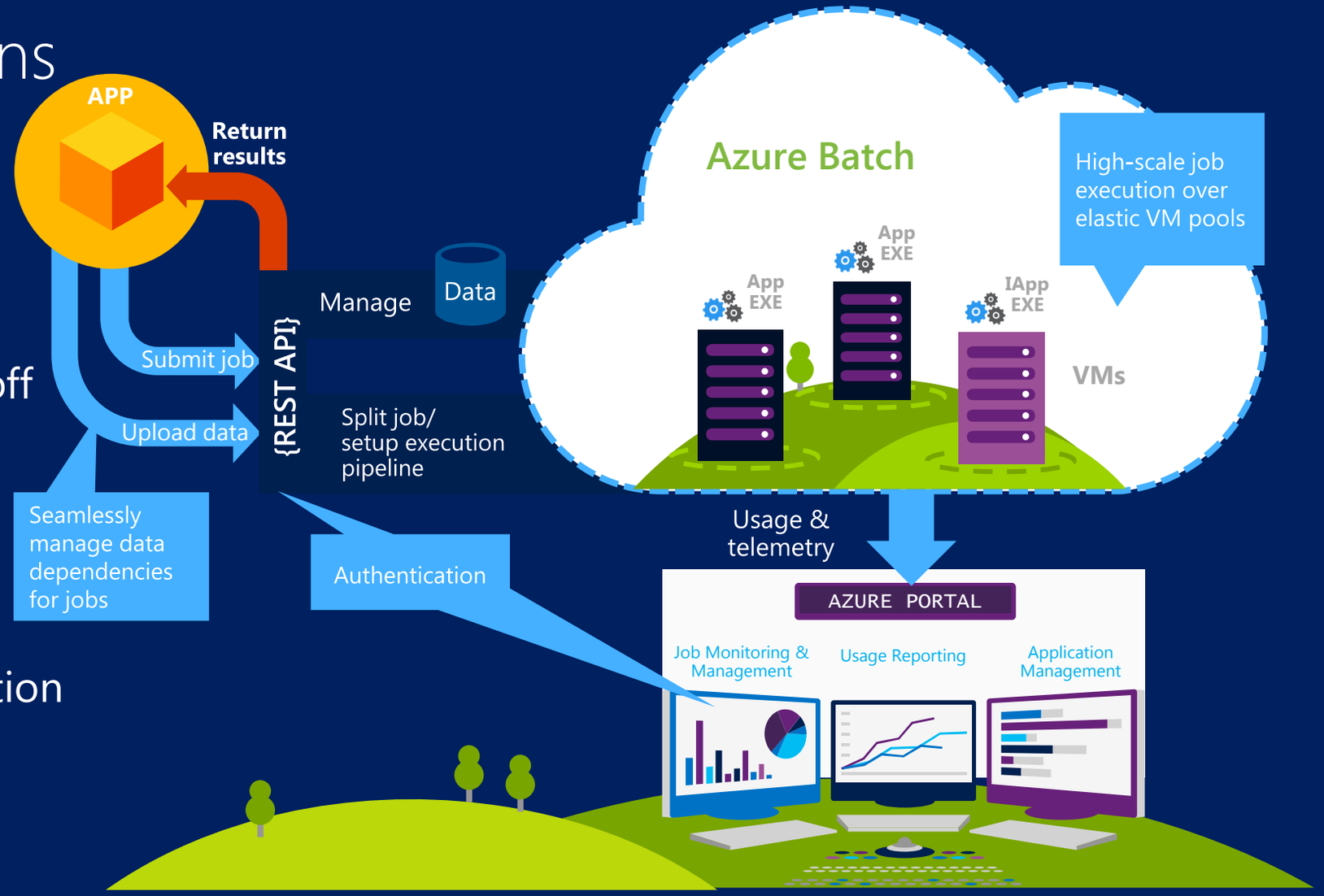
Run the applications you use on workstations and clusters today

At scale on demand

Batch starts a pool of compute VMs when submit a job, and turns them off when you're done

Managed delivery

Control who can access it, how many resources they can use, and ensure requirements such as encryption are met



Compute + Data Workflows