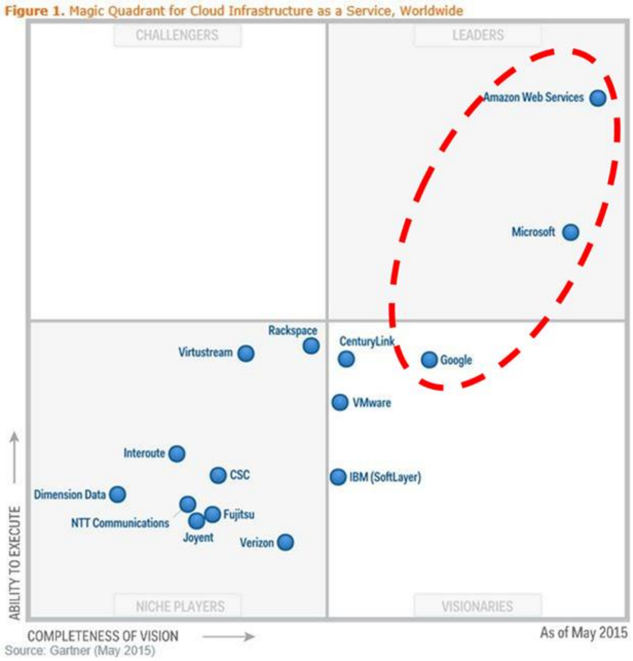
AWS ~~vs.~~ & Azure ~~vs.~~ & GCP

Determining Your Optimal Mix of Clouds

Introduction

Cloud vendor competition is heating up. 2015 has marked the shift of cloud services from early to mainstream adoption, especially within large enterprises. While AWS remains the clear leader in the market (as it has been since its inception in 2006), Azure is the fastest growing cloud provider, with triple digit growth in both 2014 and 2015. Google Cloud Platform (GCP), though far behind in market share, is still considered a top visionary by Gartner based on the comprehensiveness of their offering, go-to-market strategy, enhanced performance and global infrastructure.



Gartner 2015 Magic Quadrant for IaaS, with top three leaders circled
Source: Gartner

To better assist in the decision making process, this paper offers a concise breakdown of the three market leaders: Amazon Web Services (AWS), Google Compute Platform (GCP) and Windows Azure (Azure), who, according to Gartner, together hold the majority of the IaaS market in 2015.
Produced by Cloudyn, this eBook provides a high-level overview of the "Big 3" cloud giants, pointing out the strengths of each provider in both operational and financial terms. Towards the end, we will also explain why it's not a question of either/or, but rather which/how much by examining the benefits that can be reaped by employing a multi-cloud strategy in the enterprise.

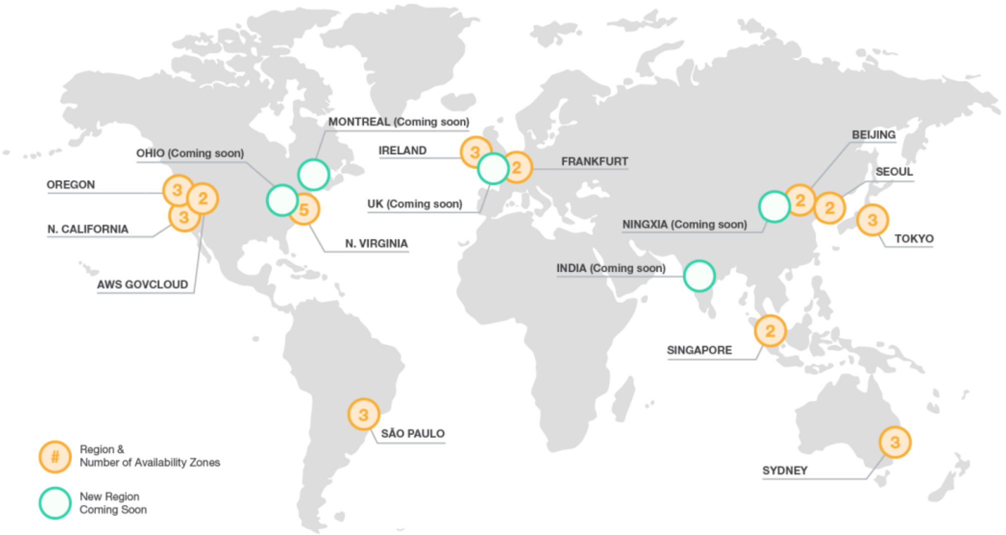
Cloud vendors: features and performance

Upon first glance, all cloud providers seem have the same offering: compute, storage, networking and other platforms as a service (databases, microservices, big data, APIs etc.). While this 40,000-foot view is true, it's a misrepresentation of the myriad services and differentiation each cloud provider has created in these categories. In this section, we will examine the offering of the three providers under each category while pointing out key differentiators.

Global Infrastructure Deployment

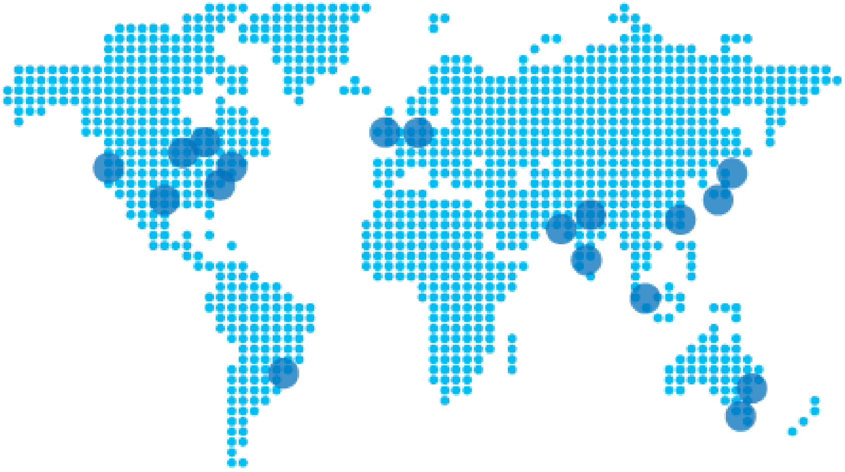
As we all know, cloud services don't actually appear out of thin air (or cloud). They are anchored in hyperscale data centers spanning between 100,000-1,000,000 square feet each (2-20 football fields), where tens of thousands of servers reside in densely populated racks, along with storage and networking appliances, as well as advanced systems for supplying power and HVAC to these data crunching behemoths.

Amazon has deployed data centers across the globe, as seen in their deployment map below. Each circle represents an AWS "region" (currently 12 regions with five more planned for 2016), which is then divided into "Availability Zones" (AZ), each of which is in reality a stand-alone data center (represented by the number in the circle). AZs are located far enough from each other so that the failure of an AZ doesn't affect the others, and close enough for zero-latency connectivity between them, giving users the sense of "one big data center" in each region.



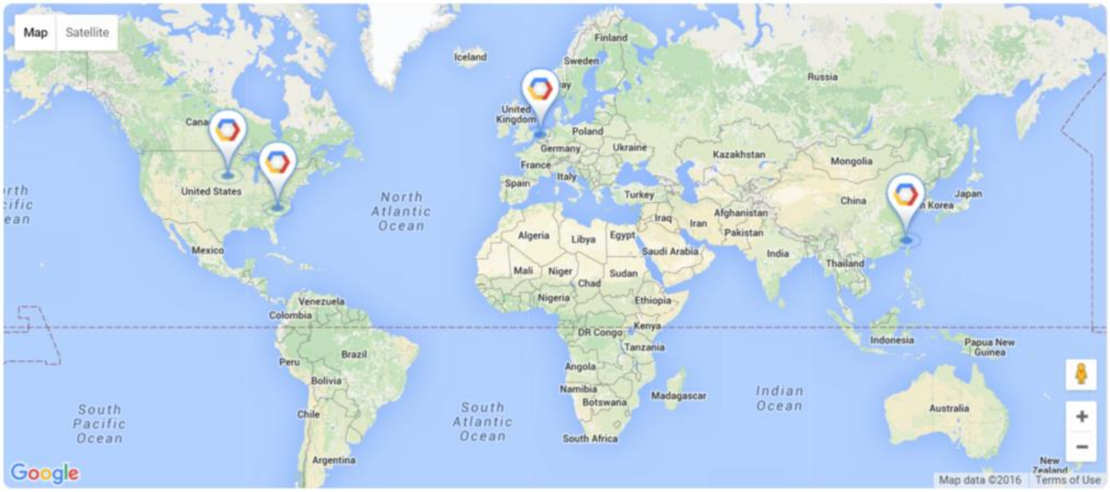
AWS Regions and Availability Zones in each region. Source: AWS

Microsoft has been quickly building more and more data centers all over the world in an effort to catch up with Amazon's vast geographical presence. From six regions in 2011, they currently have 22 regions, each of which contains one or more data centers, with five additional regions planned to open in 2016. While Amazon was the first to open a region in China, Microsoft preceded to open the India region at the end of 2015.



MS Azure Regions. Source: Microsoft

Google has the smallest footprint of the three providers, with four regions, comprised of 3-4 "zones" (data centers) each. Other data centers provide regional support against zonal failures and act as redundancy only. Google makes up for its geographical shortcomings with its global network infrastructure, which provides high-speed, low-latency connectivity between its data centers, both on a regional and interregional level (compared to public Internet connectivity with Amazon and Microsoft), as well as a large number of its own PoPs deployed in over 30 countries.



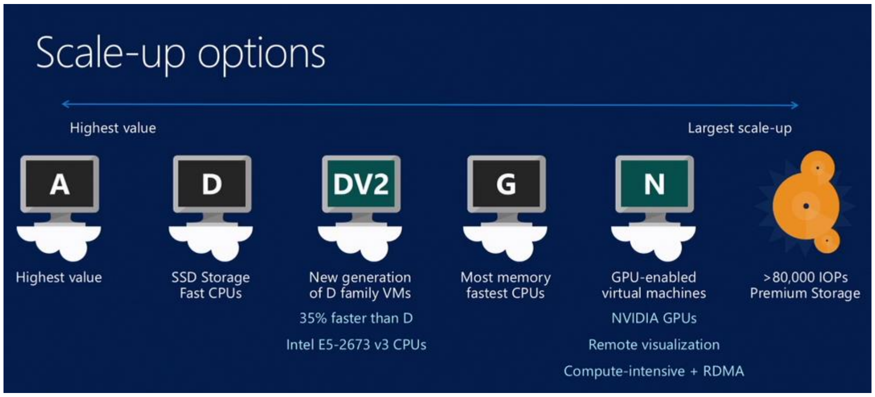
Google Cloud Platform Regions. Source: Google

Compute

One of the staples of IT is compute resources, or servers. In the cloud, physical (bare-metal) servers, in a process called “virtualization” are broken down into virtual machines (VMs), each of which acts as a server on its own. Each bare-metal server is equipped with software known as “hypervisor”, serving as the regulator of physical resources (CPU, RAM, network bandwidth etc.) between the VMs.

Amazon, with its *Elastic Compute Cloud (EC2),* has the largest flavor (server configuration) variety of all three providers. Its virtual machines are called “instances” and are divided into nine “instance families”, each of which serves a different purpose, with 2-5 instance sizes (typically 5) within each family. The families include general-purpose computing, CPU-optimized, RAM-optimized, storage-optimized and GPU-optimized families.

Microsoft on one hand has less variety in VM families compared to AWS, but on the other hand, much more flexibility with regards to machine size. Its families include *general-purpose*, *optimized machines* (better CPU, more RAM and more SSD storage), *performance-optimized* (even more than “optimized”) and *network-optimized* (32Gbps Infiniband networking).



Scale-up options

Highest value
Largest scale-up

A
Highest value

D
SSD Storage
Fast CPUs

DV2
New generation of D family VMs
35% faster than D
Intel E5-2673 v3 CPUs

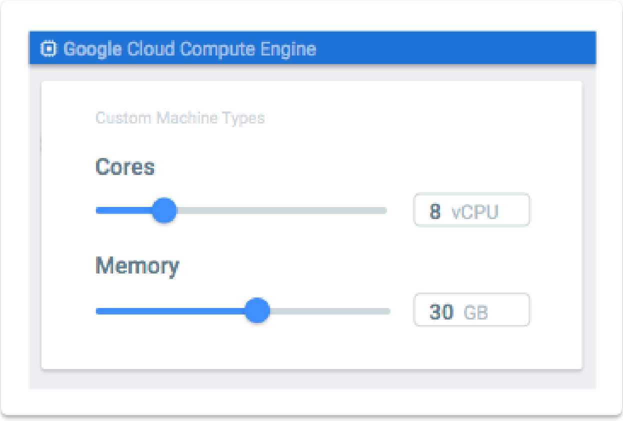
G
Most memory fastest CPUs

N
GPU-enabled virtual machines
NVIDIA GPUs
Remote visualization
Compute-intensive + RDMA

>80,000 IOPs
Premium Storage

MS Azure VM families. Source: Microsoft

On the surface **Google**’s offering seems to have the least amount of VM families under its *Google Compute Engine (GCE)*. The three families are *general-purpose, CPU-optimized and RAM-optimized,* with 5-6 sizes within each family. However, with their *“custom machines”* offering, they can be viewed as having the largest variety of VM families. This is achieved since the “*custom machines”* offering allows the user to define exactly how much CPU-power and how much RAM she requires on her machine. While custom machines provide the highest flexibility in choosing the right machine, one must also take into account the complexity involved in managing endless machine types. Local or ephemeral storage is also configurable, whereas with Amazon and Microsoft it’s defined in the machine spec.



Google Cloud Compute Engine

Custom Machine Types

Cores: 8 vCPU

Memory: 30 GB

GCP gives users the option to define their own custom machine configuration. Source: Google

Storage

Storage is another basic building block of IT. We will discuss storage service offerings from the “big three” cloud providers with regards to the two main types of storage: *Block* and *Object* storage.

Amazon’s block storage service is called “Elastic Block Storage” (EBS) and supports three types of persistent disks: Magnetic, SSD and SSD with provisioned IOPS. Maximum volume sizes range from 1TB for magnetic disks, up to 16TB for SSD disks.

Their world-renowned object storage service is “Simple Storage Service” (S3), with four different SLAs: *standard*, *standard - infrequent access, reduced redundancy and Glacier* (for archiving). All data is stored in one availability zone, unless manually replicated across AZs or regions.

Microsoft’s storage services are all referred to as *Blobs. Page Blobs and Disks* are Azure’s block storage service. It can be sourced as *standard* (magnetic) or as *Premium* (SSD), with volumes of up to 1TB.

Block Blobs (not to be confused with block storage) is Azure’s object storage service. Similar to Amazon, it’s offered in four different SLA levels: *Locally redundant storage (LRS)* where redundant copies of the data are stored within the same data center; *zone redundant storage (ZRS),* where redundant copies are stored in different data centers within the same region; and *geographically redundant storage (GRS)* which performs *LRS* on two distant data centers, for the highest level of durability and availability.

INDESCRIBABLE...
INDESTRUCTIBLE!
NOTHING CAN STOP IT!
THE BLOB
STEVEN McQUEEN ANETA CORSEAUT · EARL ROWE
JACK H. HARRIS · IRVIN S. YEAWORTH, JR. · THEODORE SIMONSON and KATE PHILLIPS

A completely different Blob

With **Google,** storage is organized a bit differently than the former two. Block storage does not get a category on its own, but is rather offered as an add-on to instances within GCE. There are two options for either magnetic or SSD volumes; however the IOPS count is fixed (compared to provisioned IOPS with AWS). Ephemeral (local) disks are fully configurable and are part of the block storage offering.

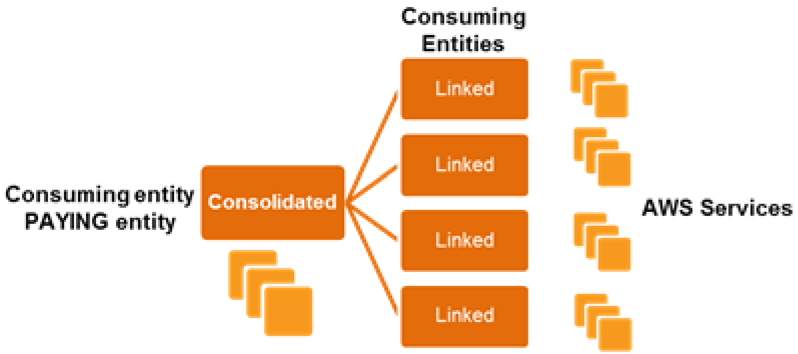
Object storage is called *Google Storage,* and divided into three classes: *Standard, Durable Reduced Availability* for less critical data (similar to RRS in S3) and *nearline*, which is for archives, but contrary to Amazon's *Glacier*, data starts streaming within a few seconds, not hours - more like *S3 standard - infrequent access.*

Cloud vendors: Billing and Pricing

After examining the different services offered by the cloud providers, we'll look into how they price their services. There are clear, inherent differences in the way the "big three" organize resources into accounts, and in the way they price their main services, specifically compute services.

Account and Billing Structure

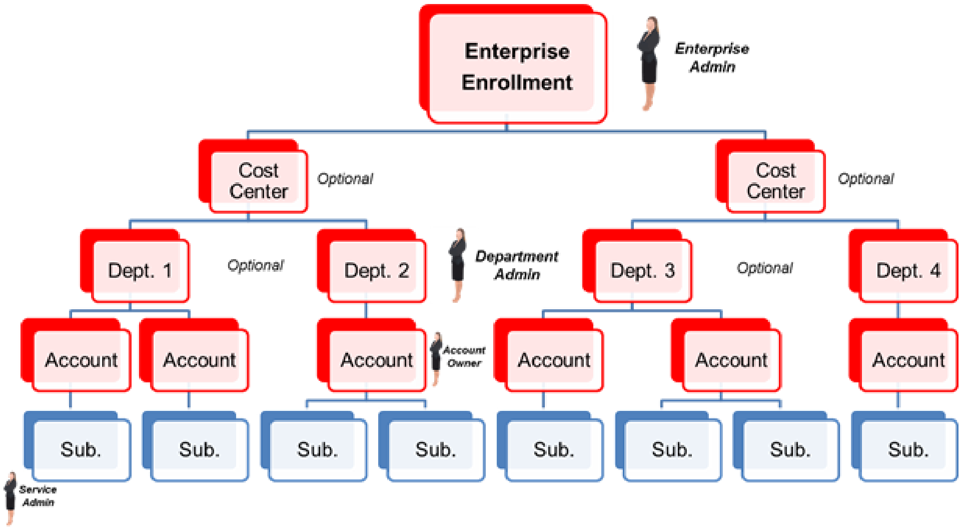
Amazon organizes resources under accounts. An account is a single billing unit under which cloud resources may be provisioned. Organizations with multiple AWS accounts, however, would like to receive one consolidated bill and not multiple separate bills. AWS enables this by creating *consolidated billing.* One of the accounts is designated as *consolidated account* and all other accounts are linked to it, hence *linked accounts.* The bill is then consolidated to include billing for all linked accounts and the consolidated account, which together are called *consolidated billing account family.*



AWS Account Structure

Microsoft employs a hierarchical approach towards accounts management. The *subscription* is the lowest in the hierarchy, and the only one, which actually provisions and consumes resources. An *account* manages multiple *subscriptions*. This may sound similar to the AWS account structure, however Azure *accounts* are management entities and don't actually consume resources themselves. For organizations without MS Enterprise Agreements

(EA), this is where the hierarchy ends. Those who have EAs, can enroll their EA in Azure, and manage all of the *accounts* under them, with optional *cost center* and *department* administrative hierarchies.



Azure Account Structure

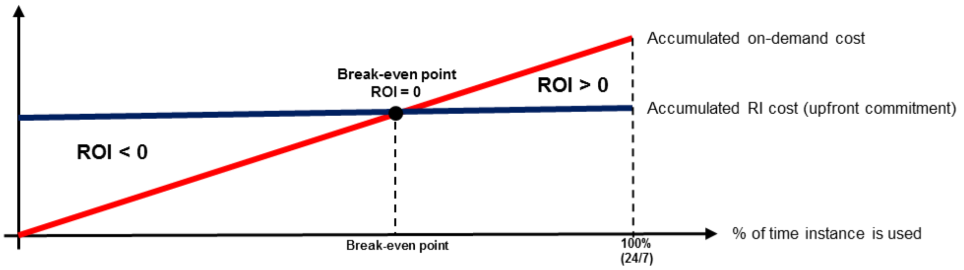
Google employs a flat hierarchy. The different resources are grouped under *"Projects"* (which are similar to AWS *accounts* or Azure *subscriptions*). There is no higher entity than *projects*, however multiple projects can be grouped under a *"consolidated billing account"*, similar to AWS consolidated billing. This billing account is not a consuming entity though, and cannot provision services, similar to Azure's *accounts*.

Pricing Models and Discounts

Cloud providers offer different pricing models and discounts for their services. Most of these complex pricing models and discounts revolve around compute services, while simple bulk discounts are usually used with all other services. There are two objectives to this. First, the providers find themselves in a highly competitive market and would like to lock in their users to long-term commitments. The second involves an interest to maximize the utilization of their infrastructure, as every hour in which a VM is unused, represents a pure loss for the providers.

Amazon has the most complex and diversified pricing structure for its EC2 services:

- *On-demand*: Pay per hour of usage, with 1-hour billing granularity
- *Reserved Instances*: Commit upfront for 1 or 3 years of usage (payment for 24/7 usage over the term) in return for a discounted per-hour price (30-70%). Payment options include:
    - *All-upfront*: Pay for the whole commitment upfront - highest discount rate
    - *Partial-upfront*: Pay 50-70% of the commitment upfront, and the rest in monthly installments over the time of the reservation; slightly lower discount than all-upfront.
    - *No-upfront*: Pay nothing upfront, and monthly installments over the term of the reservation. Significantly lower discount than all-upfront.



AWS Reserved Instances provide a positive ROI only beyond a certain rate of usage

- *Scheduled Reserved Instances:* Commit upfront for 1 year of usage over recurring daily/weekly/monthly timeframes (as opposed to 24/7). Discounts amount to 5% during "peak hours" (Mon-Fri) and 10% during "off-peak hours" (Sat-Sun).
- *Spot Instances*: AWS puts up its unused capacity for a price bid. The "market" price is updated every 5 minutes and usually encompasses discounts of around 90% over on-demand. Amazon however can terminate these instances at any time with only a 2-minute warning.
- *Fixed Duration Spot Instances*: Instances which are provisioned for either one or six hours, at 25%-45% discount over on-demand pricing.

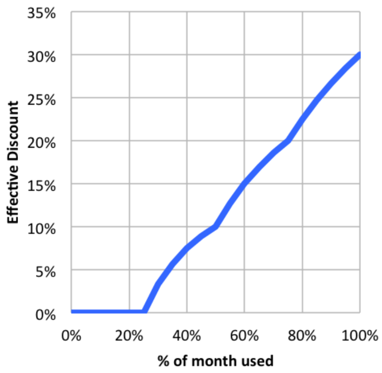
Pricing Model	Discount	Trade-Off
On-demand	-	-
Reserved Instance	30% - 70%	Upfront commitment
Scheduled RIs	5% - 10%	Upfront commitment (smaller than RI)
Spot Instance	80% - 90%	Can be terminated at any time by AWS with a 2-minute warning
Fixed Duration Spot	25% - 45%	Terminated after one or six hours

Summary of AWS pricing models

Microsoft, in sharp contrast to Amazon, has one main pricing model: on-demand, charged by the minute (compared to by the hour with AWS). Discounts are only offered for bulk monetary commitments, either through pre-paid subscriptions, which offer 5% discount on the bill, or through Microsoft's Enterprise Agreements (EAs), where much higher custom discounts can be applied in return for an upfront monetary commitment by the customer.

Google offers on-demand pricing, charged by the minute, and is the only provider of the three who doesn't require upfront commitment for receiving discounts. *Sustained use discount* is a pricing model, which retroactively discounts services, which were extensively used over the period of the billing month. The method is simple to understand, and no special action need be taken to enjoy its benefits. However, it does complicate the task of forecasting a project's cost, as it is unclear until the end of the month.

Usage Level (% out of the month)	**Sustained Usage Discount**
0-25%	0%
25%-50%	20%
50%-75%	40%
75%-100%	60%



Google Cloud Platform Sustained Use Discount

Summary and Takeaways

Companies preparing for cloud migration must do their homework to find the best course for them. Each vendor has its own sweet spot for particular deployments and it is up to the migrating company to figure out which KPIs are most important to them and select cloud vendors accordingly. Which one will fit and integrate with business needs best, as well as how it meets business' short term and long-term goals.

As the title of this paper suggests, choice of cloud provider is not an either/or question. With the growing maturity of public and private cloud platforms and as IaaS hits massive adoption, enterprises are realizing that relying on a single cloud provider is not a long lasting option. Issues like higher availability, vendor lock-in and leveraging competitive pricing push enterprises to find the optimal *mix* of clouds for their needs, rather than a single provider.

COMPANY OVERVIEW

Founded in 2011, Cloudyn is the leader in cloud monitoring and optimization. The company's industry award-winning SaaS solution delivers unprecedented insights into usage, performance, and cost, coupled with custom prescriptive actions for enhancing performance and reducing cloud spend .With more than 10,000,000 virtual instances monitored Cloudyn helps businesses select the right mix of cloud vendors, increase operational performance, reduce cloud costs to bring them under optimum control, and capitalize on customer choice. More than 2,400 customers use Cloudyn's technology worldwide including F500 industry leaders in aerospace, infrastructure, consumer online travel services, IT management consulting, and manufacturing . For more information, interested parties may visit **www.cloudyn.com.**