Parallels Remote Application Server and Microsoft Azure

Scalability and Cost of Using RAS with Azure

Contents

Introduction to Parallels RAS and Microsoft Azure 3

Scalability 4

Costs 18

Conclusion 21

CHAPTER 1

Introduction to Parallels RAS and Microsoft Azure

Deploying Parallels Remote Application Server (RAS) 15.5 on a Microsoft Azure cloud is a comprehensive desktop and application delivery solution (DAAS and SAAS) that lets you monitor and manage your entire infrastructure. RAS on Azure is fast to deploy, robust, scalable, and easy to manage, all for only about $6.62 per month per user.

This document presents an analysis of the cost and scalability of RAS deployed on Microsoft Azure.

Parallels RAS

Parallels RAS is a comprehensive virtual application and desktop delivery solution that allows your employees to use and access applications and data from any device. Seamless and easy to deploy, configure, and maintain, RAS supports both Microsoft RDS and major hypervisors.

By using RAS with Microsoft Azure, you can benefit from all the features of RAS while integrating any current use of other Microsoft technologies like Windows Server System Center and Hyper-V.

Microsoft Azure

Microsoft Azure is a collection of integrated cloud services that you can use to build, deploy, and manage applications through Microsoft's global network of datacenters. Azure has service level agreements (SLAs) that guarantee external connectivity at least 99.95% of the time. Using Azure lets you sidestep the cost of hardware and infrastructure for deploying RAS desktops and applications, providing the necessary resources for computing, networking, and storage.

CHAPTER 2

Scalability

Testing the scalability of Parallels RAS on Microsoft Azure

With Microsoft Azure, setting up virtual machines is quick and you can adjust them as your requirements change over time. VMs on Azure support all the RAS services needed for a deployment.

Note: VDI and HALB are currently unavailable but are being tested.

Parallels ran tests to evaluate using RAS 15.5 with Azure virtual machines. RAS workloads were evaluated on Azure A, D, and Dv2 series VM instance types. The Dv2 Series instances are a newer version of D-Series Standard instances. The Dv2 instance type offers more powerful 2.4 GHz Intel Xeon® E5-2673 v3 processors with Turbo Boost 2.0 technology that enables a maximum clock speed of 3.1 GHz. This newer instance series is approximately 35% faster than D-Series instances, while using the same memory and disk configurations as the D-Series.

The infrastructure VMs needed to deploy RAS—Publishing Agent, Active Directory and DNS servers, etc. were deployed primarily on D2v2 instances in the testing. The table below shows the configuration and hourly cost for a D2v2 instance type (based on Central U.S. pricing at the time of this writing).

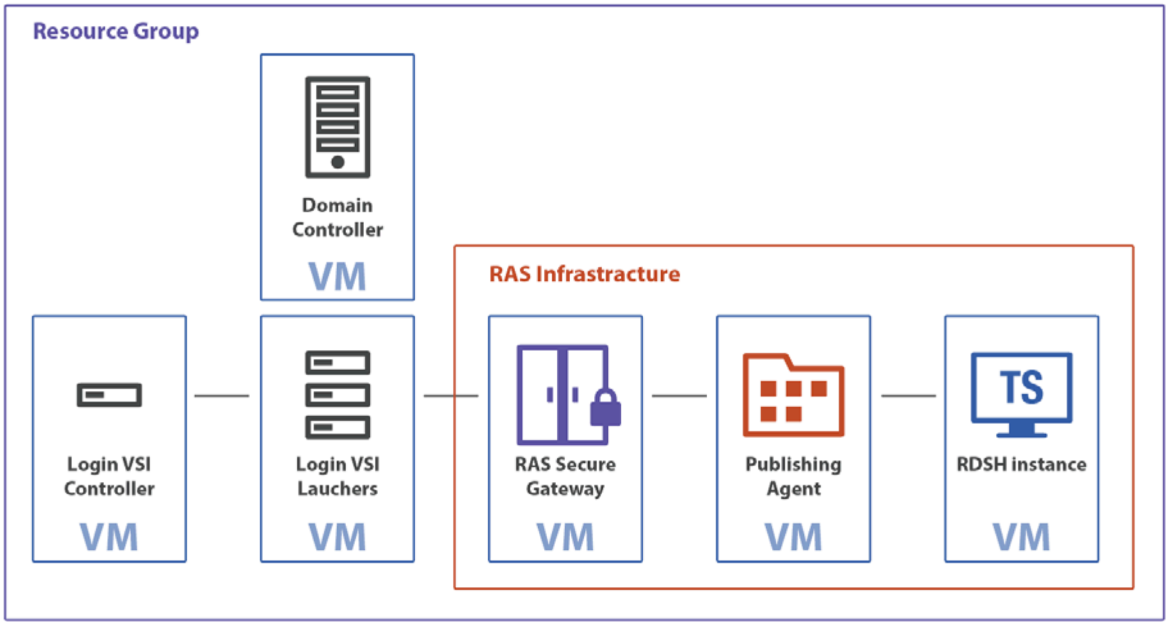
Instance	Virtual cores	RAM (GB)	Storage (GB)	Storage type	Price per hour
D2v2	2	7	100	4 data. 1 local SSD	$0.28

Configurations for scalability testing

For the RAS 15.5 scalability testing, the infrastructure VMs were configured with Microsoft Windows Server 2012 R2 on Azure instances as follows:

- Five infrastructure virtual machines on a D2v2 instance:
  - 1x VM containing a Login VSI controller and profile server
  - 4x VMs containing Login VSI launchers
- Infrastructure virtual machine on a D2v2 instance containing:
  - 1x dedicated RAS Publishing Agent
- Infrastructure virtual machine on a D2v2 instance containing:
  - 1x Active Directory controller and DNS server
- Infrastructure virtual machine on a D2v2 instance containing:
  - 1x RAS Secure Client Gateway

Creating a virtual machine on Azure also creates an Azure Resource Group container. All virtual machines in a Resource Group are siloed on the same virtual network.

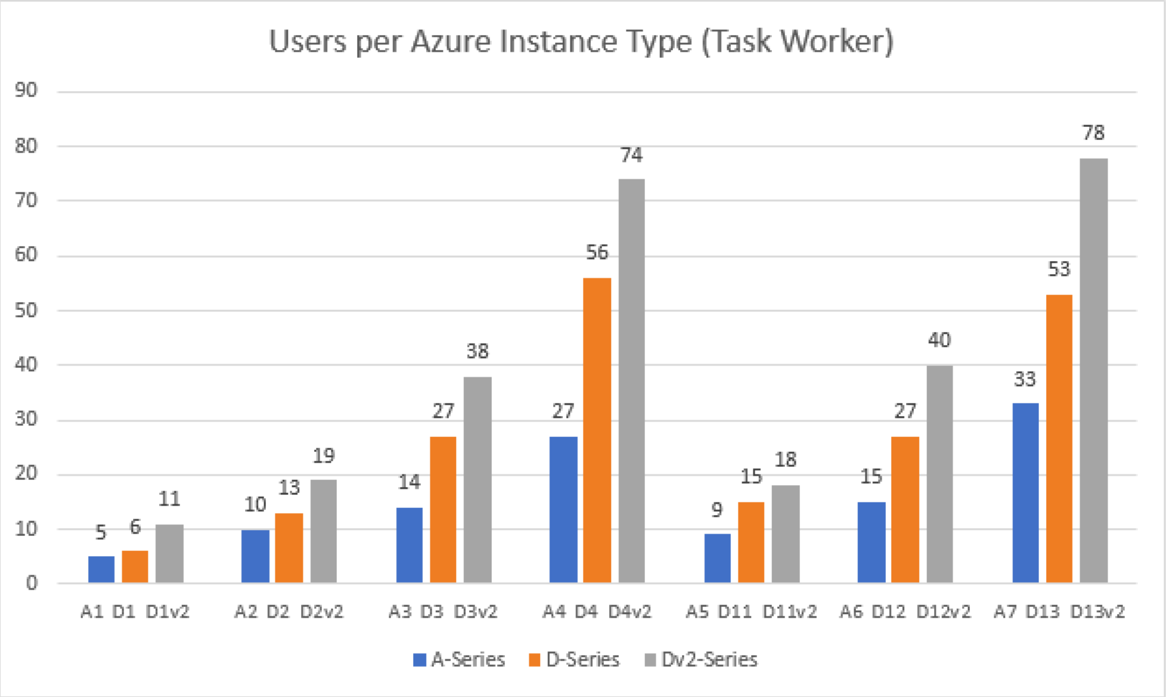


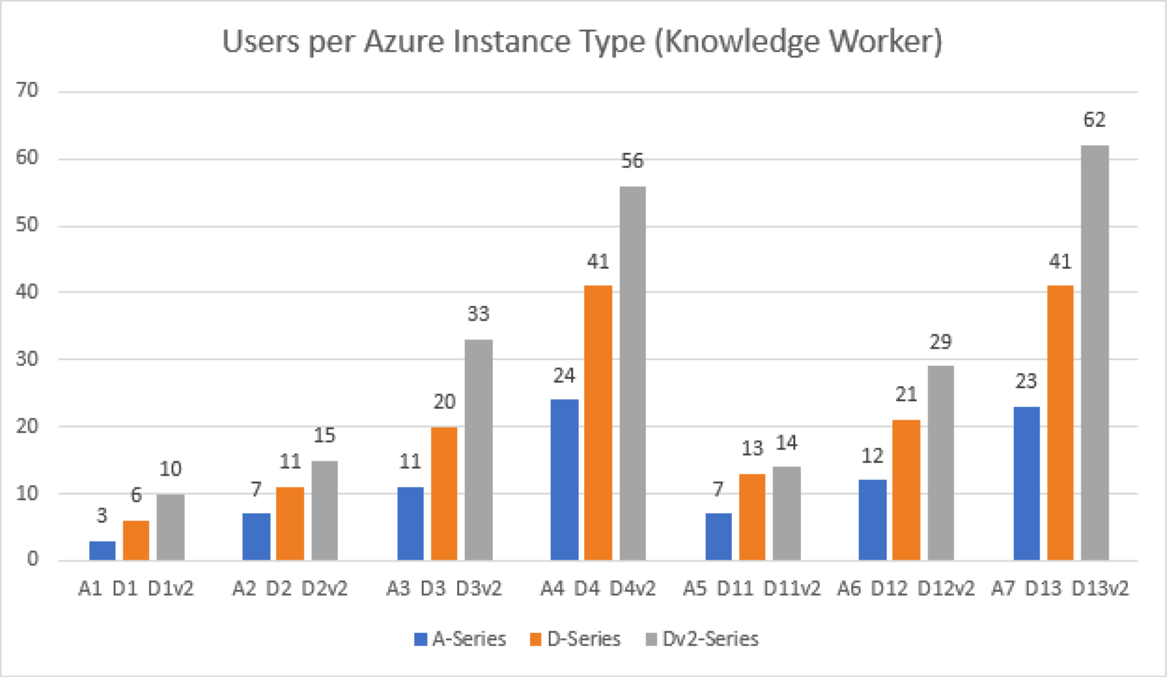
Architecture for RAS scalability testing on Azure.

In an Azure deployment, users connect through an RAS Windows client to access applications and desktops. Login VSI clients simulate user connections to the RAS server. As with the standard RAS architecture, Publishing Agents distribute the connections and set up service connection between end-users and Terminal Servers hosting applications.

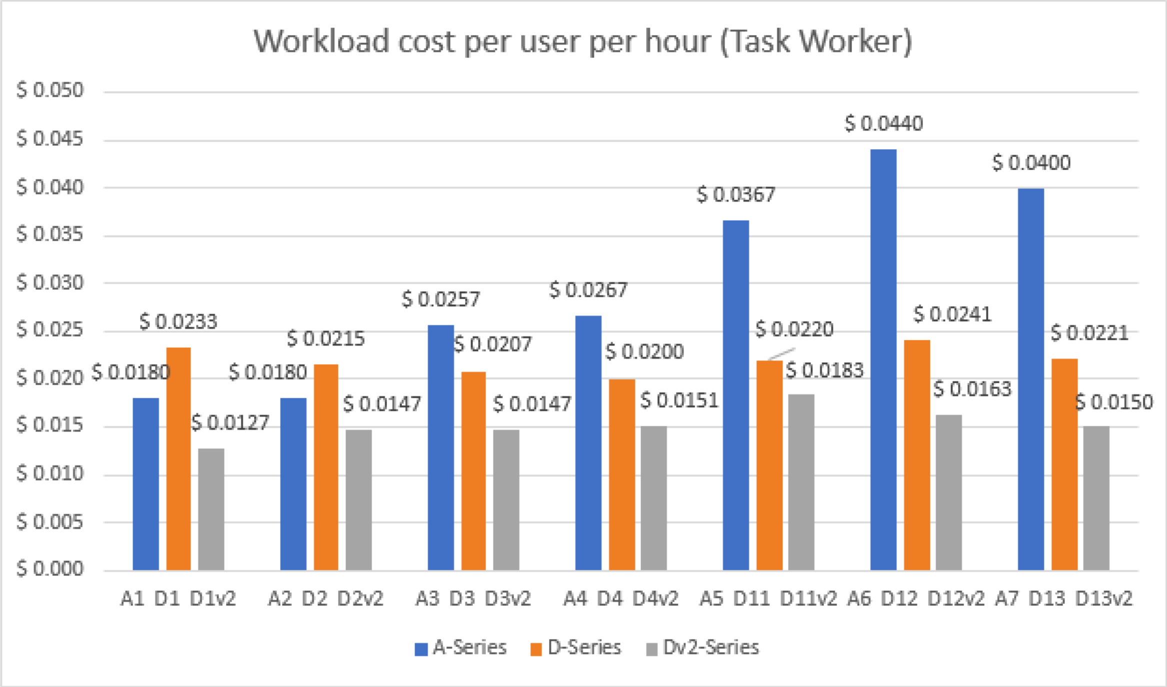
Findings

The following graphs show side-by-side comparisons of the maximum number of RAS user sessions supported by the A-Series, D-Series, and Dv2-Series VM instance type in single server scalability testing. As you can see, Dv2 VMs offer higher performance compared to the respective A-Series and D-Series VMs. For example, the D13v2 instance (with 8 vCPUs and 56 GB of RAM) hosted the highest densities, supporting 78 and 62 users under Task Worker and Knowledge Worker workloads respectively.

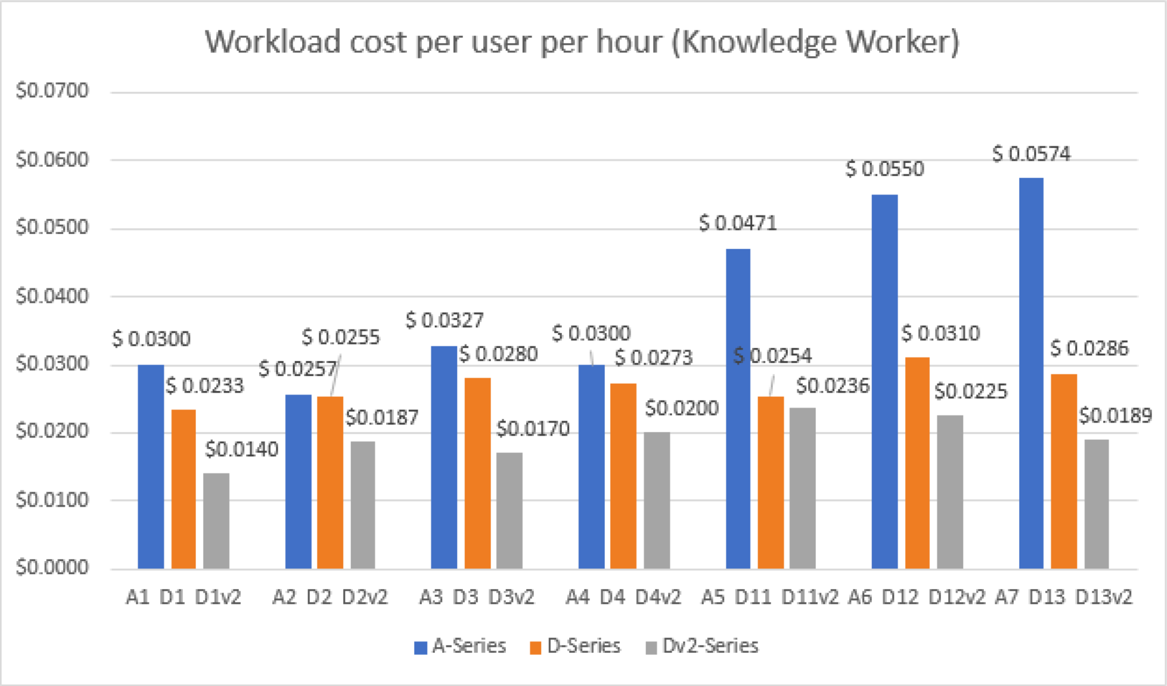




The following graphs compare the cost efficiency of each Azure instance type by Terminal Servers densities attained in single server testing. Pricing for Azure instances depends on region, instance type, and resources provided. Costs shown are based on Central U.S. pricing for standard VM instances, and include Microsoft Windows licensing.



.



The D3v2 instance offers the lowest cost per user at a good performance level (D1v2 has only a single core and 3.5 GB of RAM, with a VDI baseline greater than 2 seconds). Instance types A5, D11, and D11v2 through A7, D13 and D13v2 are configured to supply additional RAM resources and priced accordingly. In the testing, the density results showed no clear benefit from the extra memory available with these instance definitions. However, if users run applications that are particularly memory-intensive, there may be a benefit in deploying RAS 15.5 on a memory-intensive instance.

Testing process

In the scalability testing, Login VSI 4.1.12.8 was used to run a user load on the RAS shared desktops. Login VSI helps to gauge the maximum number of users that a desktop environment can support. Login VSI categorizes workloads as Task Worker, Knowledge Worker, Power Worker, and Office Worker.

It is important to note that while scalability testing is a key factor in understanding how a platform and overall solution perform, it should not be inferred as an exact measurement for real world production workloads. Customers looking to better assess how applications will perform in a RAS on Azure solution should conduct their own Login VSI scale testing using custom workload scripts.

Task Worker and Knowledge Worker workloads were selected for the testing and had the following characteristics:

- Task Worker Workload – includes segments with Microsoft Office 2013 Outlook, Excel, and Internet Explorer, Adobe Acrobat and PDF Writer. The Task Worker workload does not place a very severe demand on the environment and represents users that do not access the system very heavily.
- Knowledge Worker Workload – includes segments with Microsoft Outlook, Word, PowerPoint, and Excel; Adobe Acrobat, FreeMind, PhotoViewer, Doro PDF Writer and includes viewing of several 360p movies. The Knowledge Worker workload places a more severe

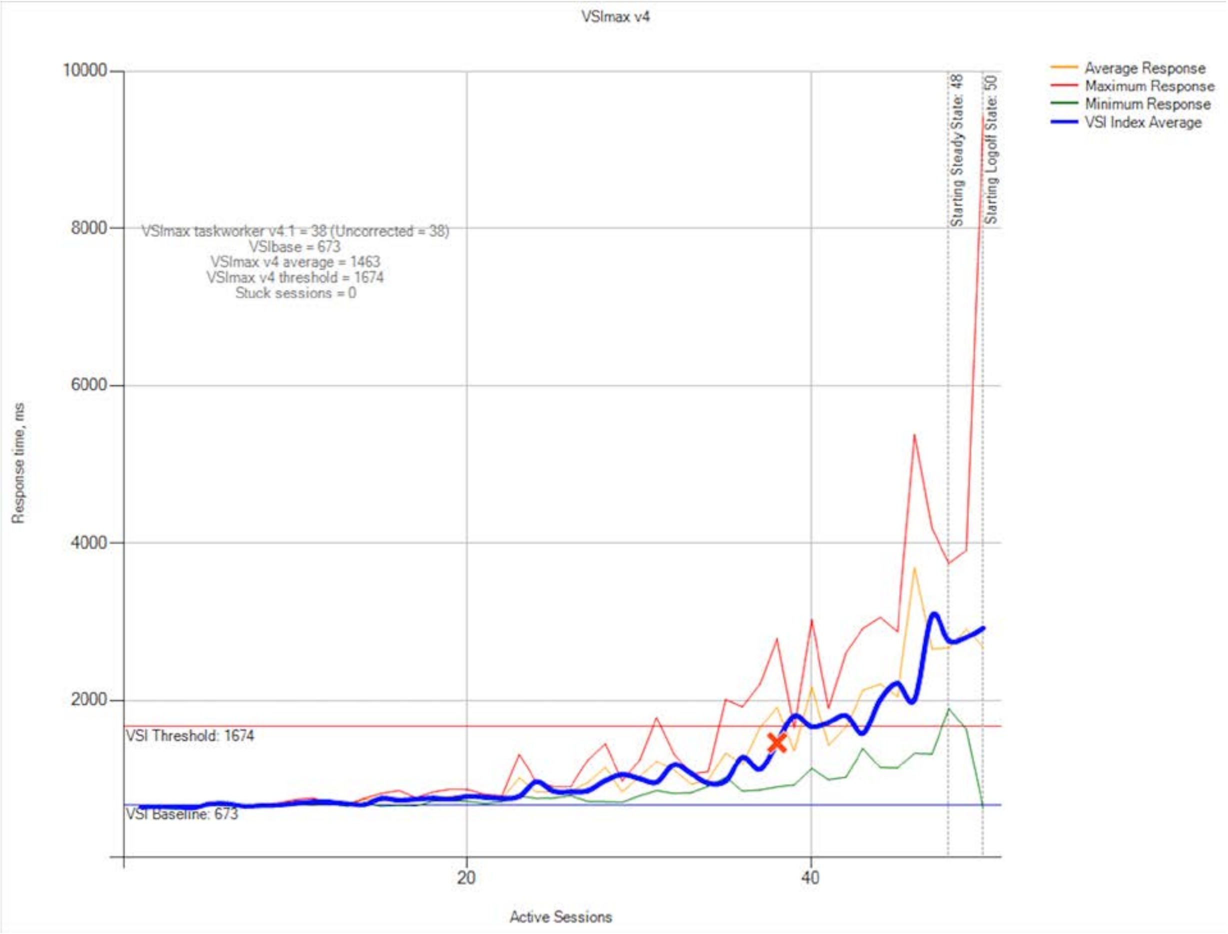
Azure A-Series, D-Series, and Dv2-Series VM instance types were tested. To to get a baseline showing the densities possible on each instance type in a series, the Login VSI client launchers were configured to go through Gateway server in proxy SSL mode. Performance was measured at user logon and virtual desktop acquisition (ramp-up), user workload execution (steady state), and user logoff. For consistent measurements showing when components were cached, each workload ran for 48 minutes before Login VSI performance was captured. VSI tests were repeated 3 times on each VM instance to get an average number of users that successfully ran the test.

The most cost-effective instance

The D3v2 instance turned out to be the most cost-effective for RAS 15.5. So the following pages show user density and performance metrics for the D3v2 instance type under the Login VSI Task Worker and Knowledge Worker workloads.

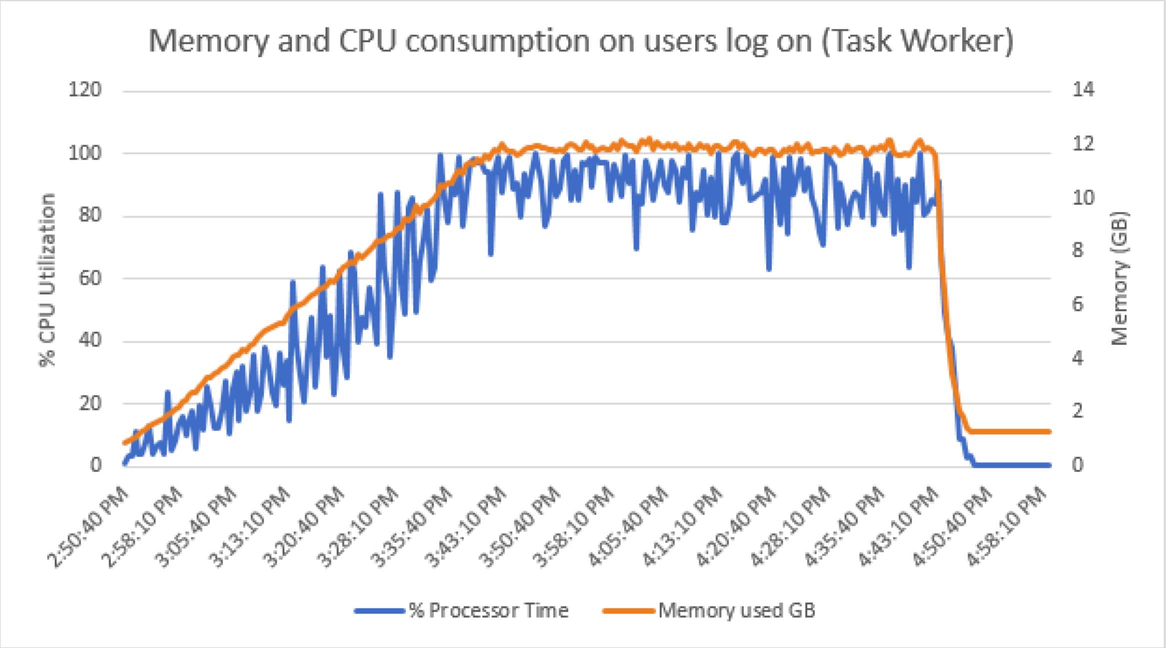
Task Worker workload findings

Following are test results for the D3v2 instance with the Task Worker workload. VSImax v4 (which indicates the maximum user density under a specific workload) is determined from the VSI Baseline and VSI Threshold metrics. VSI Baseline represents a pre-test Login VSI baseline response time measurement that is determined before the normal Login VSI sessions are sampled. The D3v2 instance shows a VSImax v4 density of 38 users running the Task Worker workload.

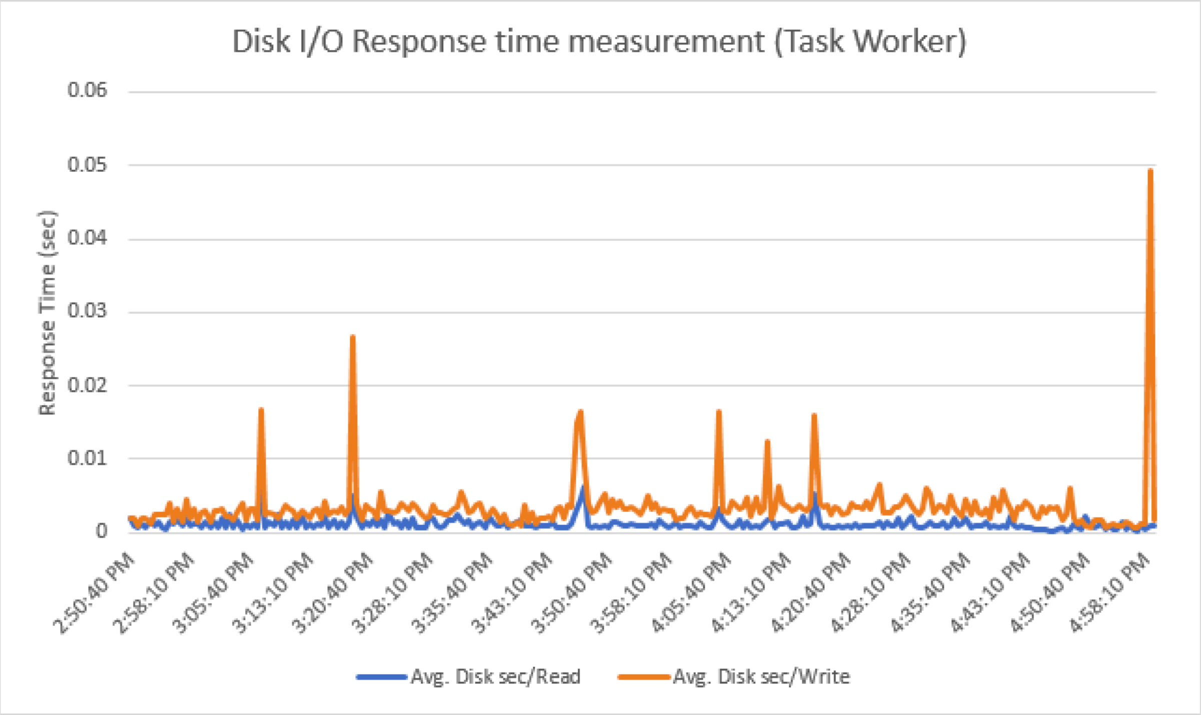


The following test results for CPU and memory consumption and disk I/O response times are helpful in evaluating performance under the test workload.

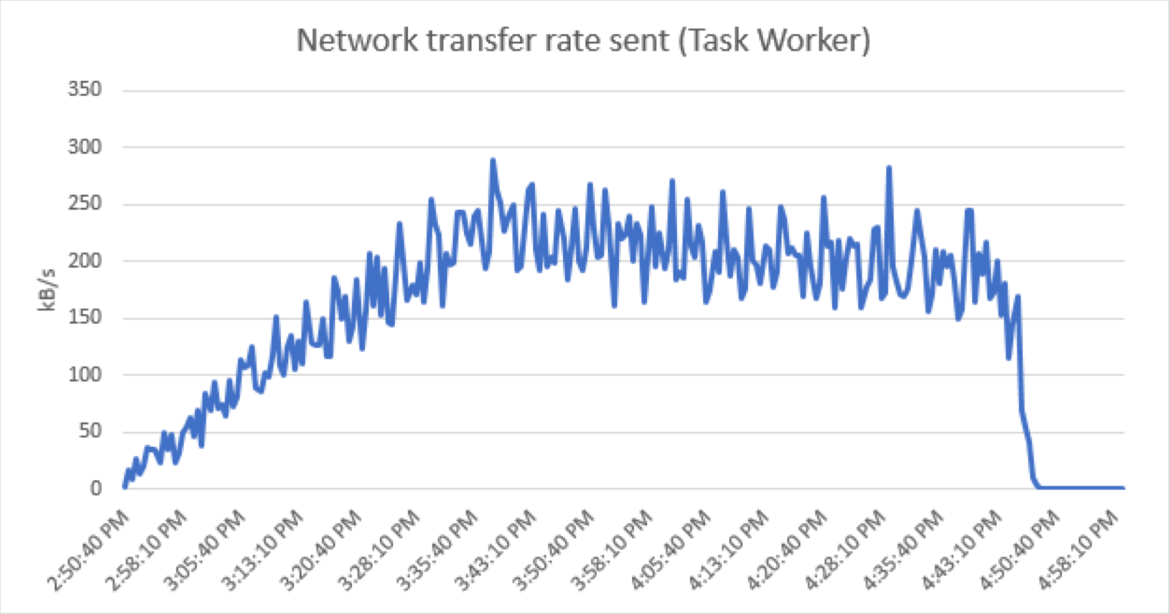
In this chart, as user load increases CPU and memory use peaks where the number of users approaches VSImax v4.



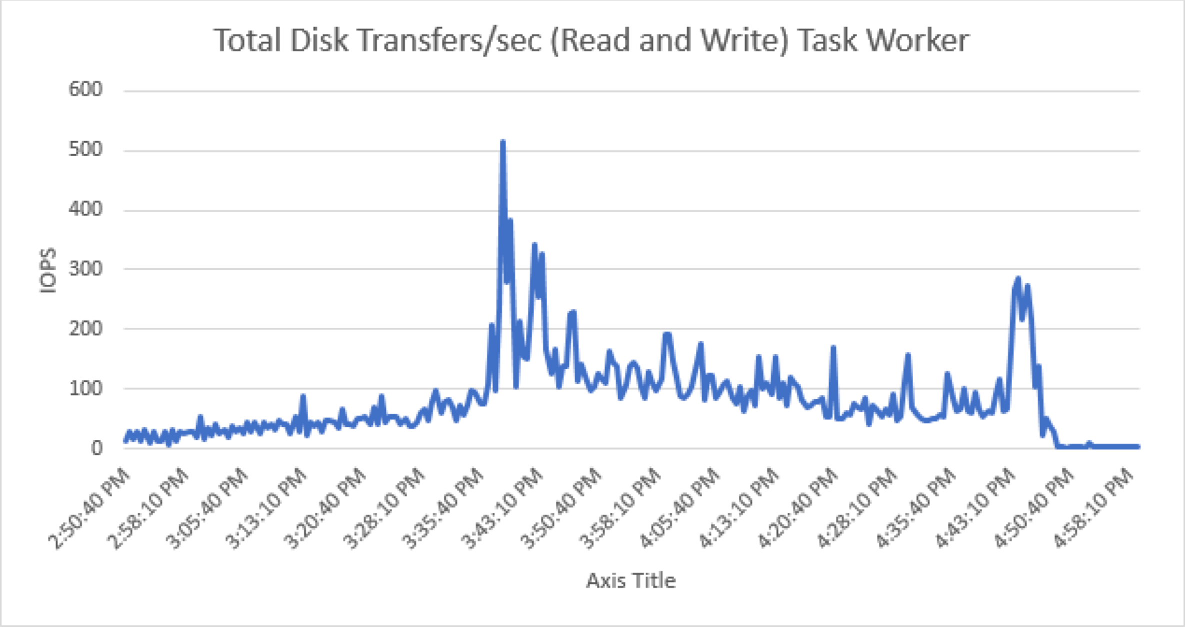
The write I/O response time averaged about 0.003491 sec. Read I/O response times averaged to 0.001223 sec.



The next two graphs show resource consumption for network transfers and disk I/O, both of which affect scalability and cost. The first graph shows networking transfer rates for data going out from Azure data centers. Microsoft charges for outbound data (and inbound data is free). For the Task Worker workload, the average outbound bandwidth at steady state is about 169 kB/s for our test group of 38 users. Therefore the outgoing transfer rate per user is about 4.44 kB/s (169/38= 4.44 kB/s).

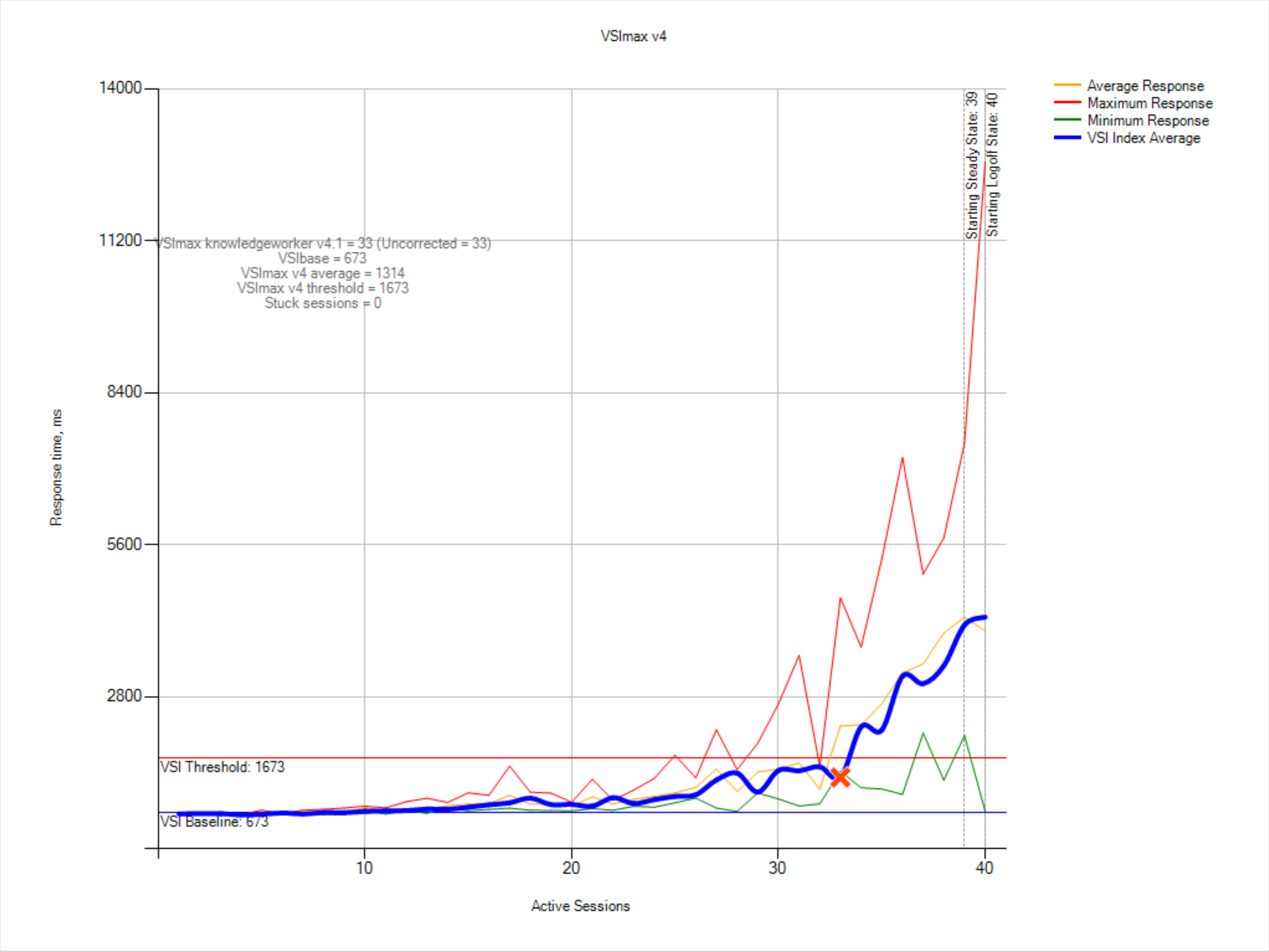


Microsoft also charges for disk transfers. The graph below shows disk transfer metrics. For the Task Worker workload, disk transfers during steady state averaged about 87 IOPS for the test group of 38 users, or about 2.3 IOPS per user. The peak transfer rate was 515 IOPS for 38 users, 14 IOPS per user. User profile data is recorded at logoff, generating disk transfer activity.



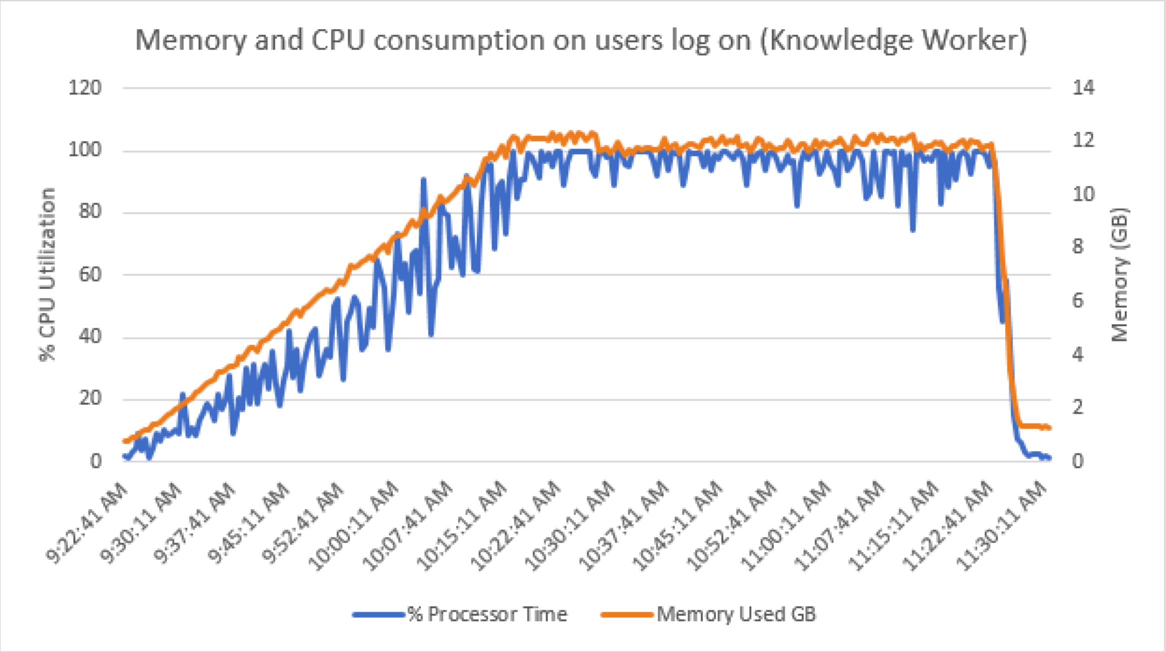
Knowledge Worker workload findings

The following test results are for the D3v2 instance with the Task Worker workload. The D3v2 instance supports a VSImax v4 of 33 users running the Knowledge Worker workload.

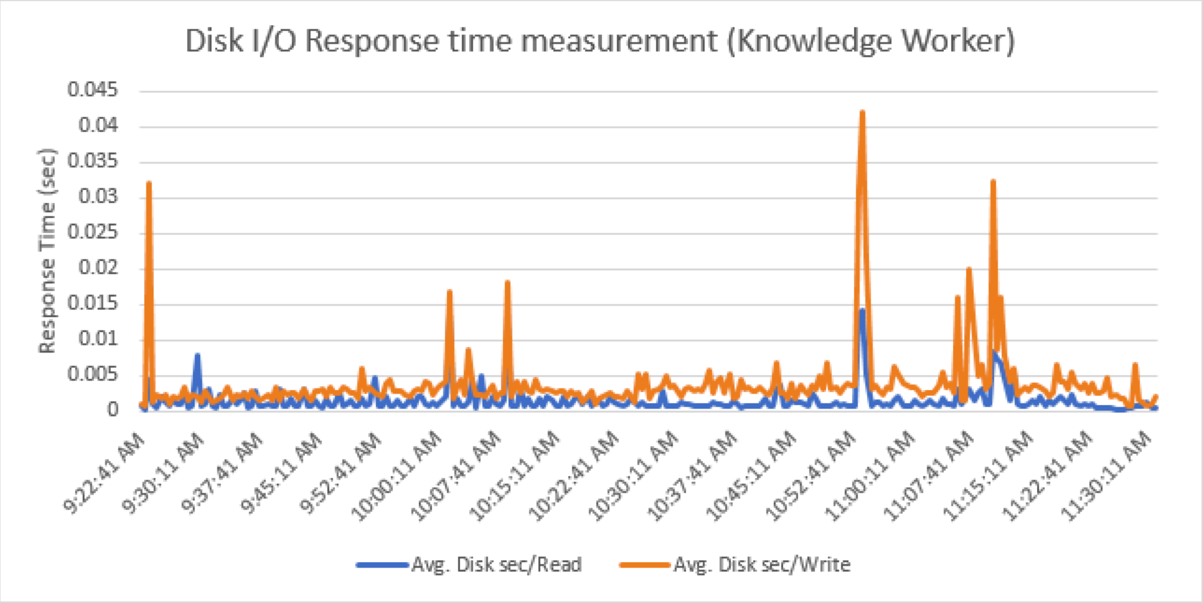


The next graphs show CPU and memory consumption and disk I/O response times for the Knowledge Worker workload, both of which are helpful to evaluate performance.

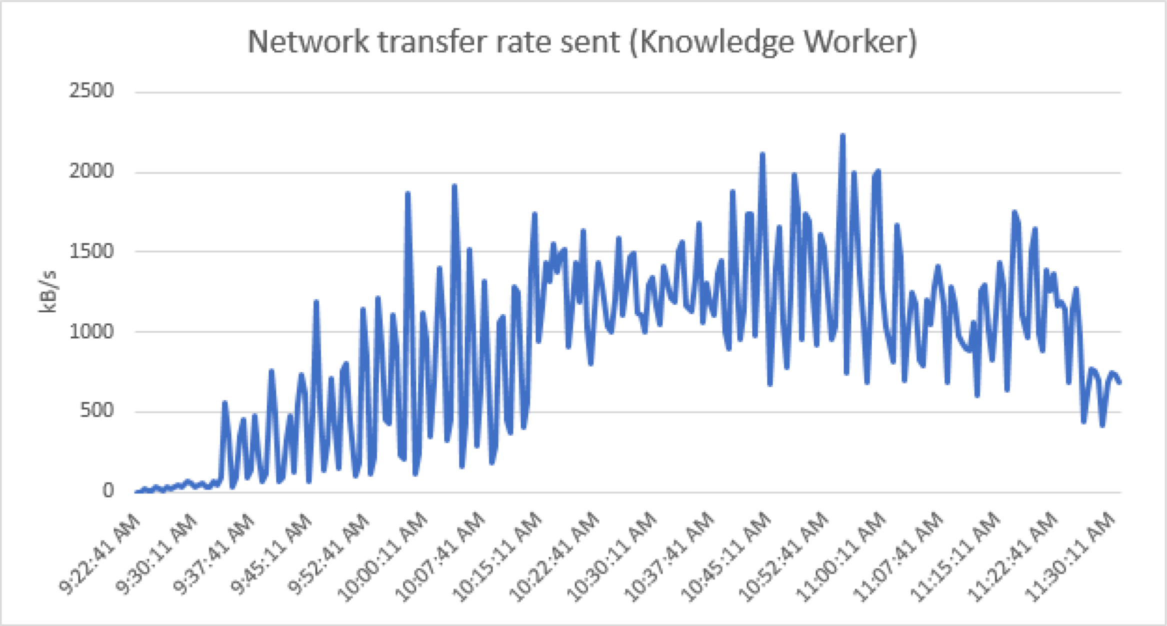
In this chart, as the user load increases, memory and CPU use peaks where the number of users approaches VSImax v4.



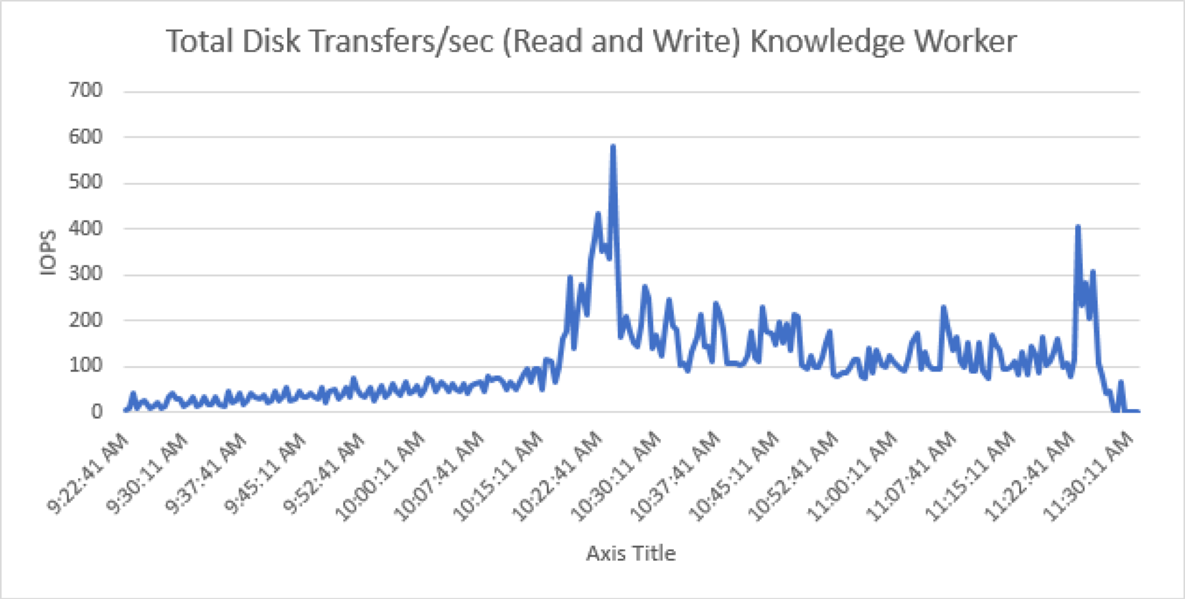
Write I/O response time for the Knowledge Worker workload averaged about 0.003883 sec while read I/O response times averaged to 0.001577 sec.



For network and disk use under the Knowledge Worker workload, the average outbound bandwidth at steady state is about 915 kB/s for our test workload of 33 users. So the the outgoing transfer rate per user was about 27.73 kB/s (915/33=27.73 kB/s).



For the Knowledge Worker workload, disk transfers during steady state averaged about104 IOPS for all 33 users or 3.2 IOPs per user (104/33=3.2). The peak value for disk transfer rate was 581 IOPS, or 17.6 IOPS per user.



CHAPTER 3

Costs

Using Microsoft Azure lets you sidestep the cost and complexity of developing the infrastructure needed for deploying RAS desktops and applications, as well as the considerable ongoing responsibility and costs to maintain that infrastructure. Instead, Microsoft provides all the necessary resources from its own datacenters.

This section describes how to estimate the cost of Infrastructure and workload VMs to deploy RAS on Microsoft Azure. The cost of Azure instances (which includes Windows licensing) is the main consideration in a budget estimate for RAS on Azure. Also to be considered are the cost of network and storage I/O, and storage used. Since RAS licensing costs are the same whether RAS is deployed on Azure or via an internal infrastructure, they aren't considered in our cost analysis.

The following cost analysis shows a monthly cost per user for both Task and Knowledge Worker workloads, based on current (as of this writing) Central U.S. pricing. Azure adjusts pricing once per month. For current pricing, go to https://azure.microsoft.com/en-us/pricing/.

Cost of Azure instances

The cost of Azure instances is the main consideration in a budget estimate for RAS on Azure. Pricing for Azure virtual machines differs by region depending on instance type and resources provided by each instance (see https://azure.microsoft.com/enus/pricing/details/virtual-machines/windows/).

Use of Azure is priced by the hour. Our cost analysis is based on users working 8 hours per day. With Azure you can reduce hourly instance costs by shutting down and deallocating virtual machines that aren't in use, but the following cost estimates (except for storage capacity) assume that all VMs are allocated and in use for 8 hours each work day. With that in mind, a D3v2 instance has a monthly cost of $138.88 and can support 38 RAS users under a Task Worker workload and 33 users under a Knowledge Worker workload, for a monthly cost of $3.66 for each Task Worker workload user and $4.20 for each Knowledge Worker workload user.

Cost of network use

Microsoft charges for data going out of Azure datacenters, but not for inbound data. Charges vary depending on the region providing services, and are tiered according to how much data is transferred each month. The monthly cost for the most expensive tier, zone1, is $0.09 per GB.

As found in our tests, the average user running a Task Worker workload uses network bandwidth at about 4.44 kbps. So for an 8-hour workday (excluding weekends), that's about 4 GB of network bandwidth per month. That's about $0.36 per month per user. For the Knowledge Worker workload, network use is approximately 18 kbps for each user, or about 16 GB for an 8-hour day. That's about $1.44 per month per user.

Cost of storage

The cost of I/O (read and write) operations to disk is $0.0036 per 100,000 transactions. A user with a Task Worker workload has about 2.3 IOPs, or about 2,053,440 transactions per month, again, assuming an 8-hour workday, for a cost of $0.08 per user. A Knowledge Worker workload calls for an average 3.2 IOPs per user, or about 2,856,960 IOPs per month. That's a cost of $0.11 a month per Knowledge Worker for storage.

Cost of storage capacity

Even when no users are active, Azure maintains persistent storage capacity and resources for the RAS infrastructure. Given that, analysis of charges for storage consumption is based on a 24-hour day. Azure provides various storage categories and options for redundancy. Storage pricing is tiered with lower rates for higher levels of consumption.

Estimated total costs

The table below shows approximate total costs per user (based on Central U.S. pricing) for both Task and Knowledge Worker workloads. Based on the D3v2 compute instance, the monthly cost per Task Worker workload user is about $6.62. The monthly cost for each user running a Knowledge Worker workload is about $8.27. Of course actual costs will vary, depending on region, instance infrastructure, and actual densities attained with real-world user workloads.

Cost per user per month	**Task Worker workload**	**Knowledge Worker workload**
D3v2 compute instance	$3.66	$4.20
Network utilization	$0.36	$1.40
Storage utilization	$0.08	$0.11
Storage capacity	$2.52	$2.56
Total	$6.62	$8.27

Remote Application Server Infrastructure VM costs on Azure

In addition to the cost of deploying VMs to support user workloads, a RAS deployment requires VMs to host infrastructure servers. The table below shows the approximate total cost per hour for each RAS infrastructure VM in Azure (based on Central U.S pricing).

Component	Instance type	Cost per hour
Publishing Agent	D2v2	$0.28
Secure Client Gateway	D2v2	$0.28
Terminal Server	D2v2	$0.28
Domain controller	D2v2	$0.28
Total		$1.12

CHAPTER 4

Conclusion

The RAS on Azure scalability results presented here should be used only as guidelines in configuring your Azure solution. Before making final sizing and deployment decisions, it is suggested that you run proof-of concept tests on different Azure instance types using your own workloads.

The Azure instance type that you select to deploy RAS workloads is the critical element that determines the user density and solution scalability, and in turn the cost-per-user for an Azure delivery model. Different instance types in Azure have advantages for specific workloads.