Assignment on: **18 Aug 2023**

- **Requirement:**
  There will be a concurrent write of 1 lakh of request and each request writing data of 20 KB into a particular disk. In a day, maximum request has 5 lakhs.
  So, we must estimate following things:
    - What is the disk required?
    - What is disk type required?
    - What should be the IOPS and throughput of that disk?
    - What is the storage queuing size for that request and what is the network bandwidth which is expected.
    - How much storage capacity is required for a month?

  Additional Step Required:

    - Cost Analysis for that infrastructure in Azure Cost Centre and find a better way to do it in a cheaper cost.
    - Find the default Latency of L1 cache, L2 cache, L3 cache compute, Redis cache trading, SQL Server if it's a random IOPs or sequential IOPS

- **Identify all the required parameters before finding the solution:**
  First, we have to calculate IOPS, Throughput for requirement and also calculate how much disk we required for daily basis and for a month.

    - IOPS = (Concurrent request per sec)
      = 100000 IOPS

    - Throughput = (IOPS * IO Size)
      = (100000 * 20 KB)
      = 2000000 KB
      = 1953.125 MBps
      = 1.907 GBps

    - Disk required for a day = (Daily 5 lakhs request processed * IO Size)
      = (500000 * 20 KB)
      = 9.54 GB per day

    - Disk required for a month = (Daily data stored * 30 days)
      = (9.54 GB * 30)
      = 286.10 GB per month

    - Queue Depth = (IPOS * Latency)
      = (100000 * 0.001)
      = 100

    - Calculate Network Bandwidth = (IOPS * IO Size)
      = (1 Lakh * 20 KB)
      = (100000 * 20 * 1024 bytes)
      = 2048000000 bytes/s
      = 16384000000 bits/s
      = 16.34 Gbps (Note: If you want the result in Gbps, divide by 1,000,000,000 (1 billion))

    - Calculating CPU core = (IOPS * 1 vCPU/10000 IOPS) + (1GBps Throughput * 1 vCPU) Where: 1 vCPU process 10000 IOPS and 1 GBps Throughput for SSD dick
      = (100000 * 1/10000) + (1.9 * 1)
      = 10 + 2 = 12 vCPU

- **Solution: (Selection single VM for given requirement)**

  - Virtual Machine Selection

Size	vCPU	Memory: GiB	Max data disks	Max uncached Premium SSD disk throughput: IOPS/MBps	Max burst uncached Premium SSD disk throughput: I OPS/MBps	Max uncached Ultra Disk and Premium SSD V2 disk throughput: IOPS/MBps	Max burst uncached Ultra Disk and Premium SSD V2 disk throughput: IOPS/MBps	Max NICs	Network bandwidth	Price	Region
Standard_E32bs_v5	32	256	32	88000/2500	120000/4000	117920/2500	160000/4000	8	16000	$1,740.32/month	West US 3

  - Disk Selection

Disk type	Instance	Count	Size	Max IOPS	Max throughput (MB/s)	Region	Price
Ultra Disk Storage	- - -	1	512 GiB	120000	2048	East US	$6,734.22/month
						Total Price:	$8474.54 / month

Total Price: 8474.54 / month

  - If we use Premium SSD V2 with combination of Premium SSD then cost will get reduce

Disk type	Instance	Count	Size	Max IOPS	Max throughput (MB/s)	Region	Price
Premium SSD V2	- - -	1	256 GiB	80000	1000	East US	$499.16/month
Premium SSD	P30	4	1024 GiB	5000 X 4	200 x 4	East US	$540.68/month
						Total Price	$2780.16/month

Total Price: 8474.54 / month

Note: If we want to select VM and Disk Storage and try to find the best process then please use this link: https://azure.microsoft.com/en-us/pricing/vm-selector/

- Default Latency of L1 cache, L2 cache, L3 cache, Radis cache, SQL Server Random IOPS or sequential IOPS

Cache type	Default latency
L1 cache	1-2 nanoseconds
L2 cache	3-5 nanoseconds
L3 cache	7-10 nanoseconds
Radis cache	1-10 microseconds
SQL Server Random IOPS	50-100 microseconds
SQL Server Sequential IOPS	10-20 microseconds

Storage Queue Size of Queue Depth:

- The queue depth or queue length or queue size is the number of pending IO requests in the system. The value of queue depth determines how many IO operations your application can line up, which the storage disks will be processing.
- Queue Depth and multi-threading are closely related. The Queue Depth value indicates how much multi-threading can be achieved by the application. If the Queue Depth is large, application can execute more operations concurrently, in other words, more multi-threading. If the Queue Depth is small, even though application is multi-threaded, it will not have enough requests lined up for concurrent execution.
- Very high queue depth value also has its drawbacks. If queue depth value is too high, the application will try to drive very high IOPS. Unless application has persistent disks with sufficient provisioned IOPS, this can negatively affect application latencies.

Compare different type of disk with respect of IOPS, Throughput, Bandwidth etc and we are only considering Azure for this required

Disk type	Disk size	IOPS	Throughput	Bandwidth	Latency	Storage queue
Standard HDD	Up to 16TB	50-180	100-1000 MB/s	100-1000 Mbps	10-15 ms	64k
Standard SSD	Up to 16TB	500-6000	500-3000 MB/s	500-3000 Mbps	4-6 ms	64k
Premium SSD	Up to 32,767 GB	20,000-80,000	900-1,200 MB/s	900-1,200 Mbps	1-2 ms	64k
Ultra Disk	Up to 65,536 GB	160,000-400,000	4,000 MB/s	4,000 Mbps	0.1-0.2 ms	256k

More details information about individual Disk Type:

Ultra-Disk		
Disk-Size (GiB)	IOPS	Throughput (MB/s)
4	1,200	300
8	2,400	600
16	4,800	1,200
32	9,600	2,400
64	19,200	4,000
128	38,400	4,000
256	76,800	4,000
512	153,600	4,000
1,024-65,536	160,000	4,000

Important Points:

- Ultra-disks are designed to provide sub millisecond latencies and target IOPS, and throughput described in the preceding table 99.99% of the time.
- Ultra-disks can't be used as OS disks, they can only be created as empty data disks.
- Azure Site Recovery doesn't support ultra-disks. In addition, only un-cached reads and un-cached writes are supported. Snapshots for ultra-disks are available but have additional limitations.

Premium SSD v2		
Disk-Size (GiB)	IOPS	Throughput (MB/s)
1 GiB to 6 GiB	3,000	125 MB/s
7 GiB to 160 GiB	500 IOPS/GiB + 3,000	125 MB/s + 0.25 MB/s/IOPS
161 GiB to 64 TiB	80,000	1,200 MB/s

Important Points:

- All Premium SSD v2 disks have a baseline IOPS of 3000 that is free of charge. After 6 GiB, the maximum IOPS a disk can have increases at a rate of 500 per GiB, up to 80,000 IOPS.
- Premium SSD v2 is suited for a broad range of workloads such as SQL server, Oracle, MariaDB, SAP, Cassandra, Mongo DB, big data/analytics, and gaming, on virtual machines or stateful containers.
- Premium SSD V2 disk can’t be used as an OS disk and Currently, Premium SSD V2 disks can’t be attached to VMs in Availability Sets.

Premium SSD														
Premium SSD sizes	**P1**	**P2**	**P3**	**P4**	**P6**	**P10**	**P15**	**P20**	**P30**	**P40**	**P50**	**P60**	**P70**	**P80**
Disk size in GiB	4	8	16	32	64	128	256	512	1,024	2,048	4,096	8,192	16,384	32,767
Base provisioned IOPS per disk	120	120	120	120	240	500	1,100	2,300	5,000	7,500	7,500	16,000	18,000	20,000
Base provisioned Throughput per disk	25 MB/s	25 MB/s	25 MB/s	25 MB/s	50 MB/s	100 MB/s	125 MB/s	150 MB/s	200 MB/s	250 MB/s	250 MB/s	500 MB/s	750 MB/s	900 MB/s
Max burst IOPS per disk	3,500	3,500	3,500	3,500	3,500	3,500	3,500	3,500	30,000*	30,000*	30,000*	30,000*	30,000*	30,000*
Max burst throughput per disk	170 MB/s	170 MB/s	170 MB/s	170 MB/s	170 MB/s	170 MB/s	170 MB/s	170 MB/s	1,000 MB/s*	1,000 MB/s*	1,000 MB/s*	1,000 MB/s*	1,000 MB/s*	1,000 MB/s*

Important Points:

- Capacity, IOPS, and throughput are guaranteed when a premium storage disk is provisioned. For example, if you create a P50 disk, Azure provisions 4,095-GB storage capacity, 7,500 IOPS, and 250-MB/s throughput for that disk. Your application can use all or part of the capacity and performance. Premium SSDs are designed to provide the single-digit millisecond latencies, target IOPS, and throughput described in the preceding table 99.9% of the time.

Standard SSD														
Standard SSD sizes	**E1**	**E2**	**E3**	**E4**	**E6**	**E10**	**E15**	**E20**	**E30**	**E40**	**E50**	**E60**	**E70**	**E80**
Disk size in GiB	4	8	16	32	64	128	256	512	1,024	2,048	4,096	8,192	16,384	32,767
Base IOPS per disk	Up to 500	Up to 500	Up to 500	Up to 500	Up to 500	Up to 500	Up to 500	Up to 500	Up to 500	Up to 500	Up to 500	Up to 2,000	Up to 4,000	Up to 6,000
***Expanded IOPS per disk**	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Up to 1,500	Up to 3,000	Up to 6,000	Up to 6,000	Up to 6,000	Up to 6,000
Base throughput per disk	Up to 60 MB/s	Up to 60 MB/s	Up to 60 MB/s	Up to 60 MB/s	Up to 60 MB/s	Up to 60 MB/s	Up to 60 MB/s	Up to 60 MB/s	Up to 60 MB/s	Up to 60 MB/s	Up to 60 MB/s	Up to 400 MB/s	Up to 600 MB/s	Up to 750 MB/s
***Expanded throughput per disk**	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Up to 150 MB/s	Up to 300 MB/s	Up to 600 MB/s	Up to 750 MB/s	Up to 750 MB/s	Up to 750 MB/s

Important Points:

- Azure standard SSDs are optimized for workloads that need consistent performance at lower IOPS levels. They're an especially good choice for customers with varying workloads supported by on-premises hard disk drive (HDD) solutions. Compared to standard HDDs, standard SSDs deliver better availability, consistency, reliability, and latency. Standard SSDs are suitable for web servers, low IOPS application servers, lightly used enterprise applications, and non-production workloads.

- Virtual Machine Details:
  - In Azure, Virtual Machine categories 6 different categories:

Type	Sizes	Description
General purpose	B, Dsv3, Dv3, Dasv4, Dav4, DSv2, Dv2, Av2, DC, DCv2, Dpdsv5, Dpldsv5, Dpsv5, Dplsv5, Dv4, Dsv4, Ddv4, Ddsv4, Dv5, Dsv5, Ddv5, Ddsv5, Dasv5, Dadsv5, DCasv5, DCadsv5	Balanced CPU-to-memory ratio. Ideal for testing and development, small to medium databases, and low to medium traffic web servers.
Compute optimized	F, Fs, Fsv2, FX	High CPU-to-memory ratio. Good for medium traffic web servers, network appliances, batch processes, and application servers.
Memory optimized	Esv3, Ev3, Easv4, Eav4, Epdsv5, Epsv5, Ev4, Esv4, Edv4, Edsv4, Ev5, Esv5, Edv5, Edsv5, Easv5, Eadsv5, Mv2, M, DSv2, Dv2, ECasv5, ECadsv5	High memory-to-CPU ratio. Great for relational database servers, medium to large caches, and in-memory analytics.
Storage optimized	Lsv2, Lsv3, Lasv3	High disk throughput and IO ideal for Big Data, SQL, NoSQL databases, data warehousing and large transactional databases.

GPU	NC, NCv2, NCv3, NCasT4_v3, NC A100 v4, ND, NDv2, NGads V620, NV, NVv3, NVv4, NDasrA100_v4, NDm_A100_v4	Specialized virtual machines targeted for heavy graphic rendering and video editing, as well as model training and inferencing (ND) with deep learning. Available with single or multiple GPUs.
High performance compute	HB, HBv2, HBv3, HBv4, HC, HX	Our fastest and most powerful CPU virtual machines with optional high-throughput network interfaces (RDMA).