Heraflux Technologies

Azure SQL

Storage Tuning

David Klee

1

About David Klee

Founder & Chief Architect
Heraflux Technologies

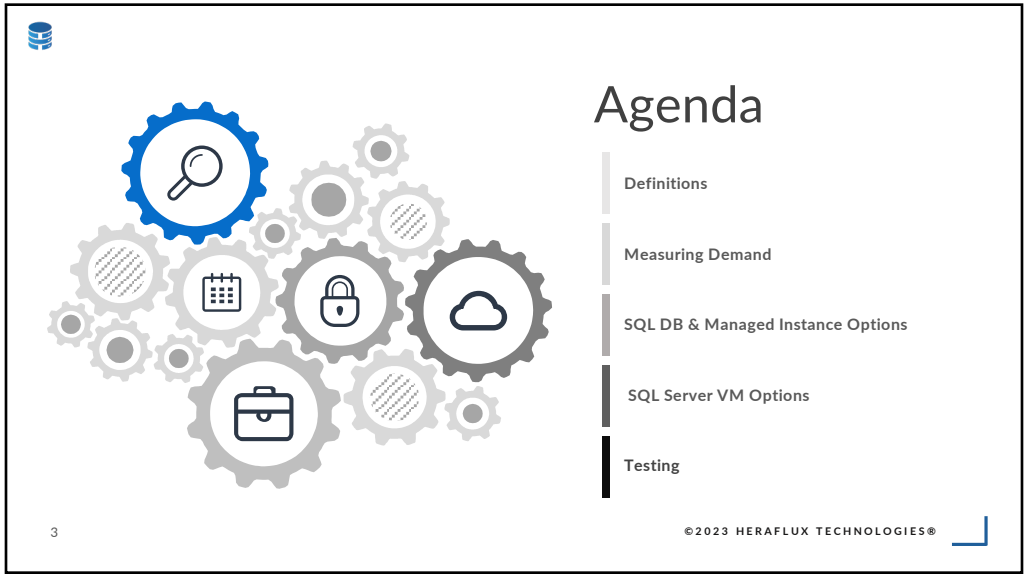
davidaklee
heraflux.com
davidklee.net
hfxte.ch/youtube

Microsoft® MVP Most Valuable Professional

vmware® vEXPERT

2

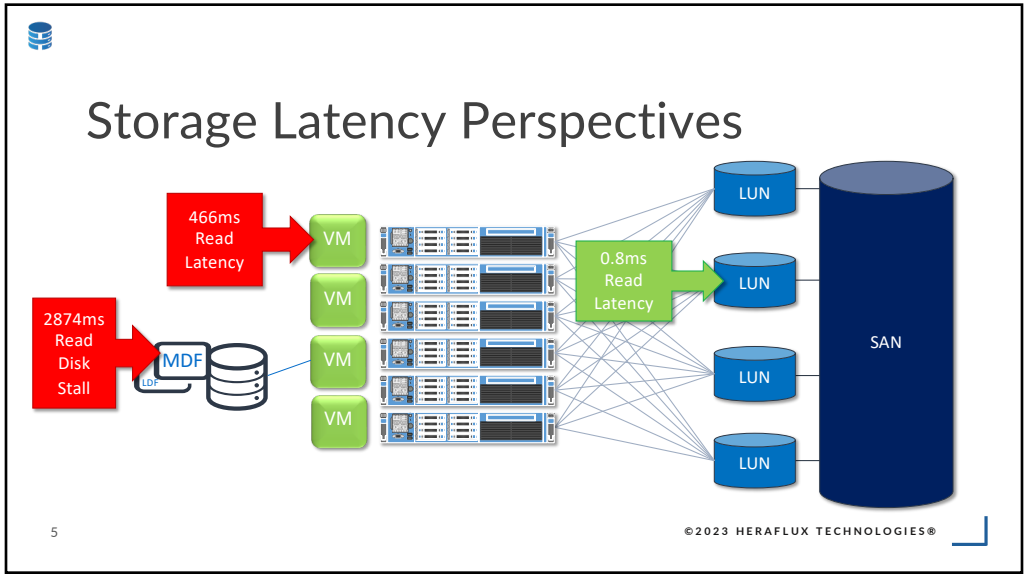
Do not organize...



3

4

Do not
organi...



5

6

Do not organize...

1 Definitions

7

7

Azure SQL

- SQL Server (boxed product)
- SQL Server (boxed product) in Azure VM
- SQL Managed Instance
- SQL Database
- *CosmosDB*
- *Lots of other cloudy services*

8

8

Do not organize...

LATENCY

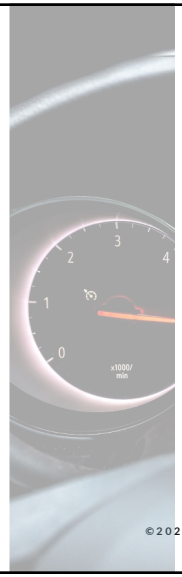
Most important for SQL Server. Round trip time to do something on disk. Milliseconds -> Microseconds.

I/O OPERATIONS PER SECOND

IOPs. Rate of demand. Higher is better. *

THROUGHPUT

IOPs x block size. MB/s. Monitor interconnect throughput.



Storage Performance

Speed matters, but low latency matters more for SQL Server.

9

9

1 IOP @ 4KB
1MB file = 250 IOPs

1 IOP @ 1MB
1MB file = 1 IOPs

Block Size

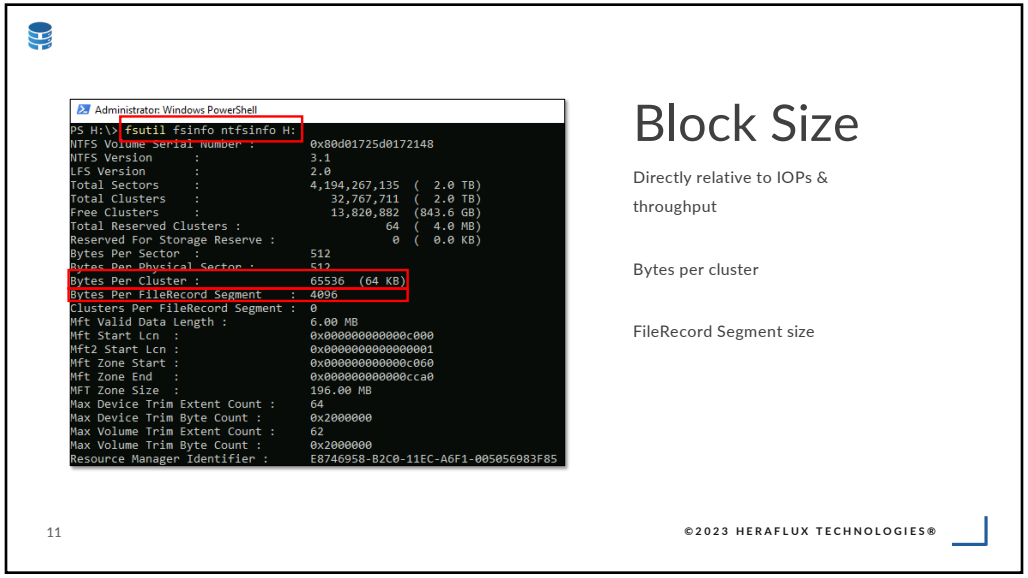
Directly relative to IOPs & throughput

IOPs to read a 1MB file?

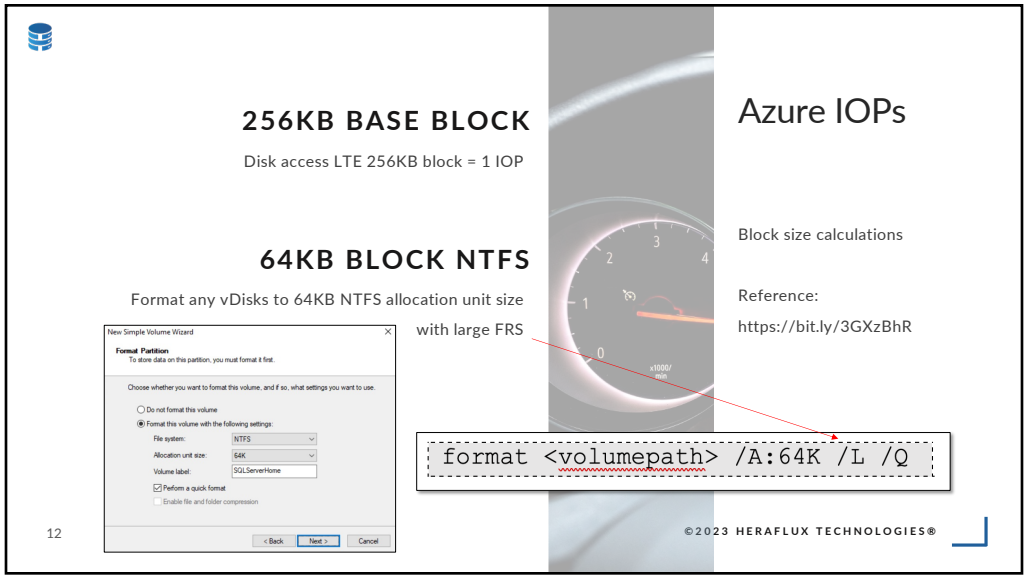
10

10

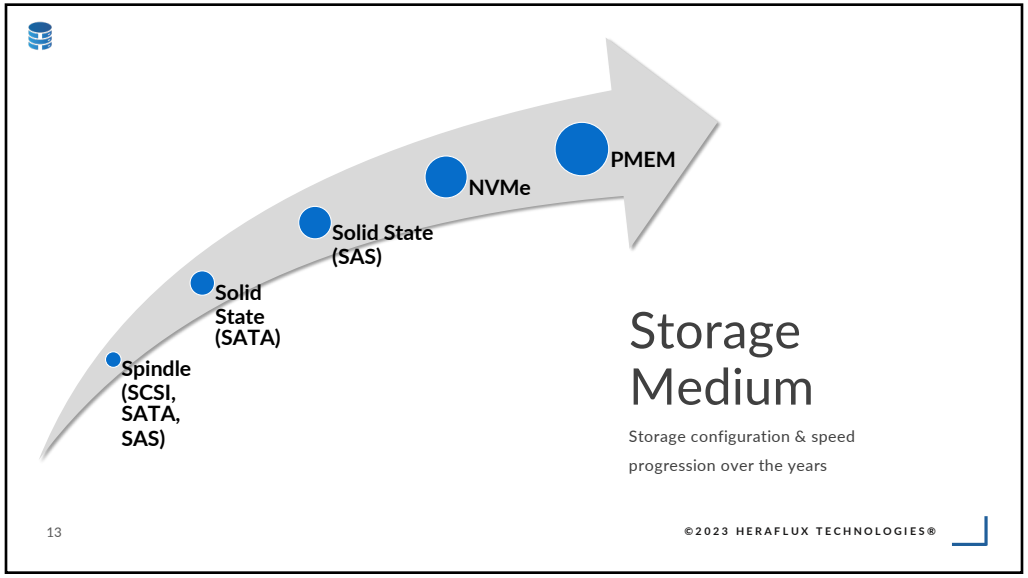
Do not organize...


Administrator: Windows PowerShell
PS H:\> fsutil fsinfo ntfsinfo H:
NTFS Volume Serial Number : 0x80d01725d0172148
NTFS Version : 3.1
LFS Version : 2.0
Total Sectors : 4,194,267,135 (2.0 TB)
Total Clusters : 32,767,711 (2.0 TB)
Free Clusters : 13,820,882 (843.6 GB)
Total Reserved Clusters : 64 (4.0 MB)
Reserved For Storage Reserve : 0 (0.0 KB)
Bytes Per Sector : 512
Bytes Per Physical Sector : 512
Bytes Per Cluster : 65536 (64 KB)
Bytes Per FileRecord Segment : 4096
Clusters Per FileRecord Segment : 0
Mft Valid Data Length : 6.00 MB
Mft Start Lcn : 0x000000000000c000
Mft2 Start Lcn : 0x0000000000000001
Mft Zone Start : 0x000000000000c060
Mft Zone End : 0x000000000000cca0
MFT Zone Size : 196.00 MB
Max Device Trim Extent Count : 64
Max Device Trim Byte Count : 0x2000000
Max Volume Trim Extent Count : 62
Max Volume Trim Byte Count : 0x2000000
Resource Manager Identifier : E8746958-B2C0-11EC-A6F1-005056983F85
Block Size
Directly relative to IOPs & throughput
Bytes per cluster
FileRecord Segment size
11
©2023 HERAFLUX TECHNOLOGIES®

11


256KB BASE BLOCK
Disk access LTE 256KB block = 1 IOP
64KB BLOCK NTFS
Format any vDisks to 64KB NTFS allocation unit size with large FRS
Azure IOPs
Block size calculations
Reference:
https://bit.ly/3GXzBhR
format <volumepath> /A:64K /L /Q
12
©2023 HERAFLUX TECHNOLOGIES®

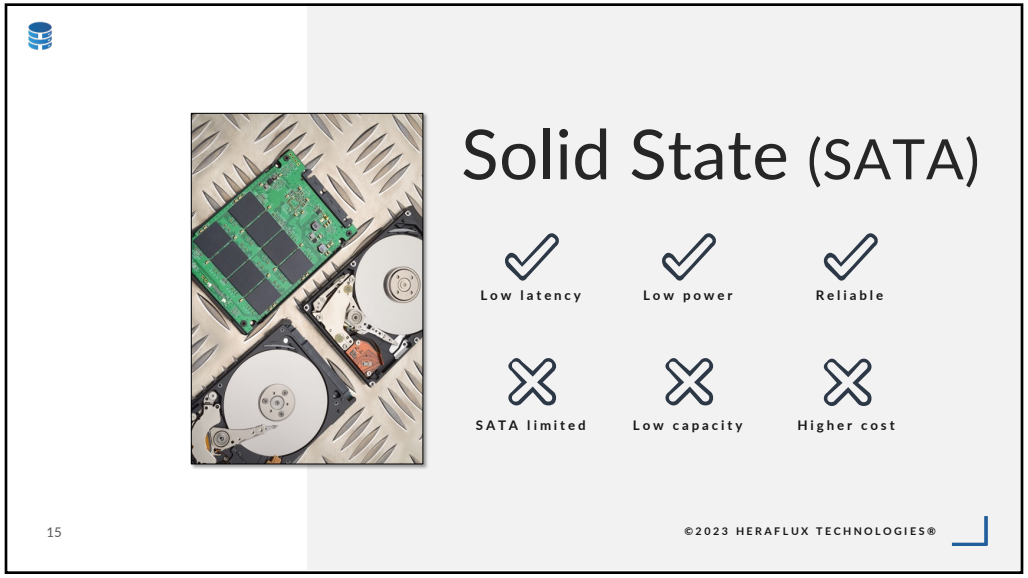
12



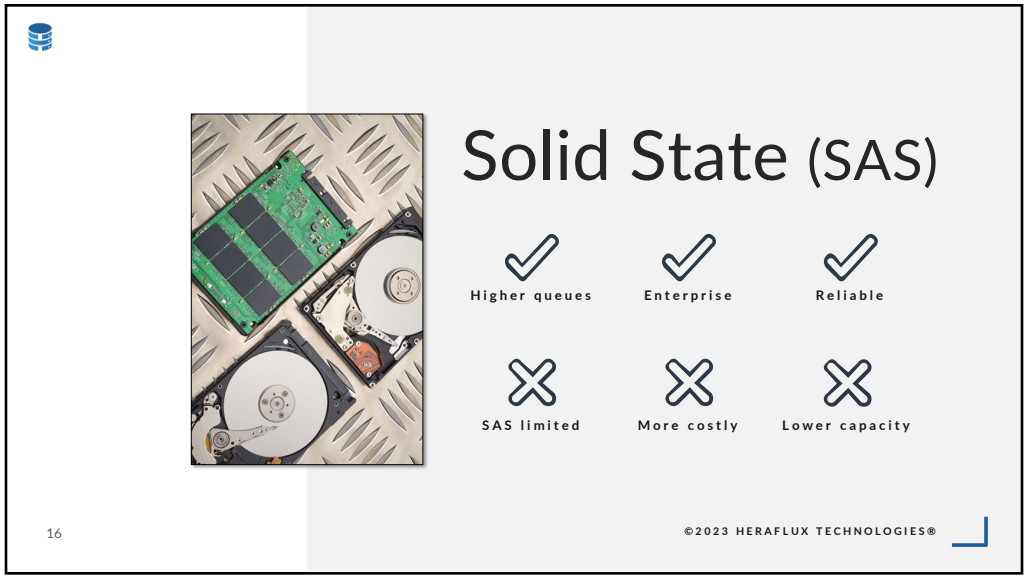
13

14

Do not organize...

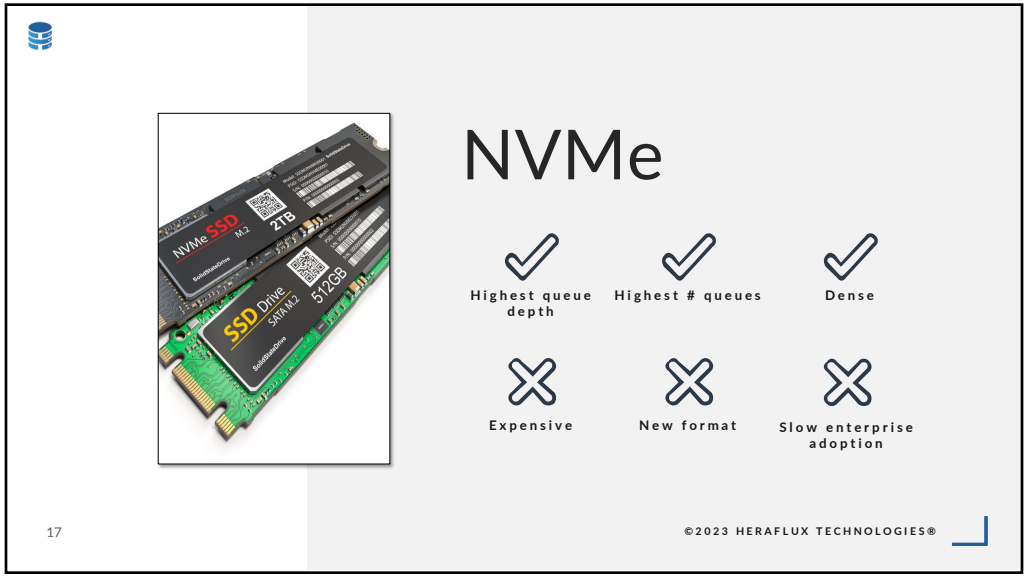


15



16

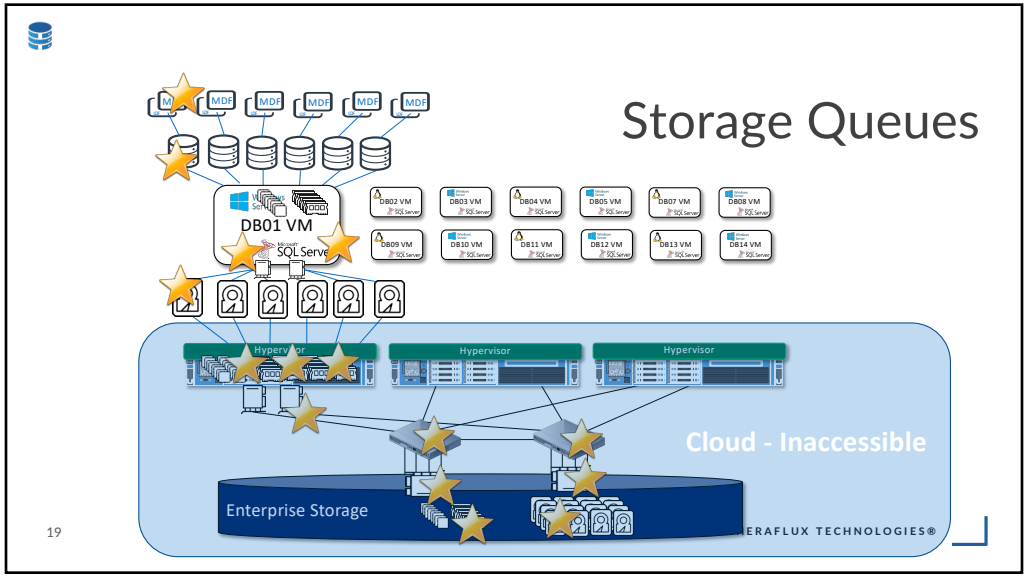
Do not
organize...



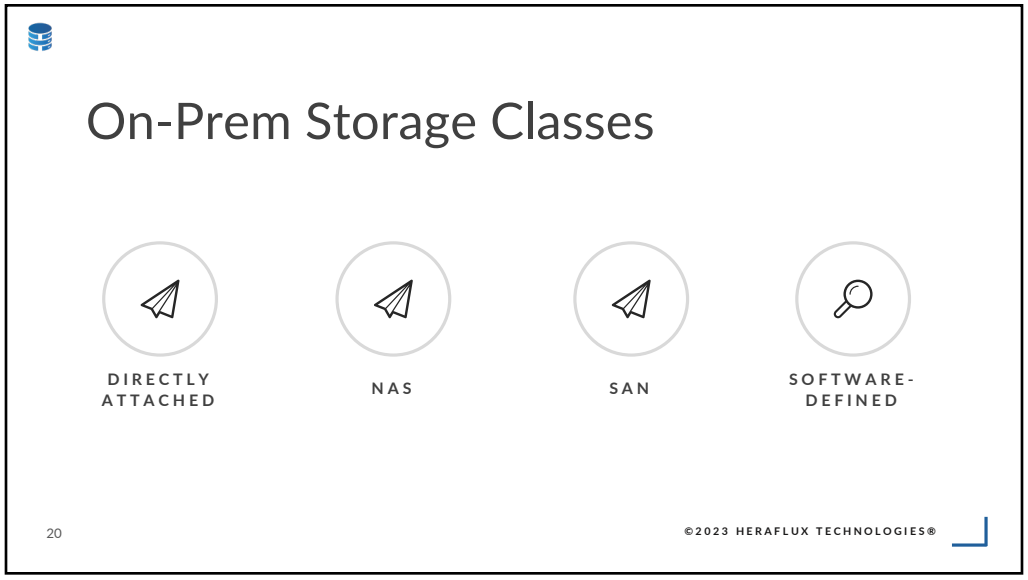
17

18

Do not organize...

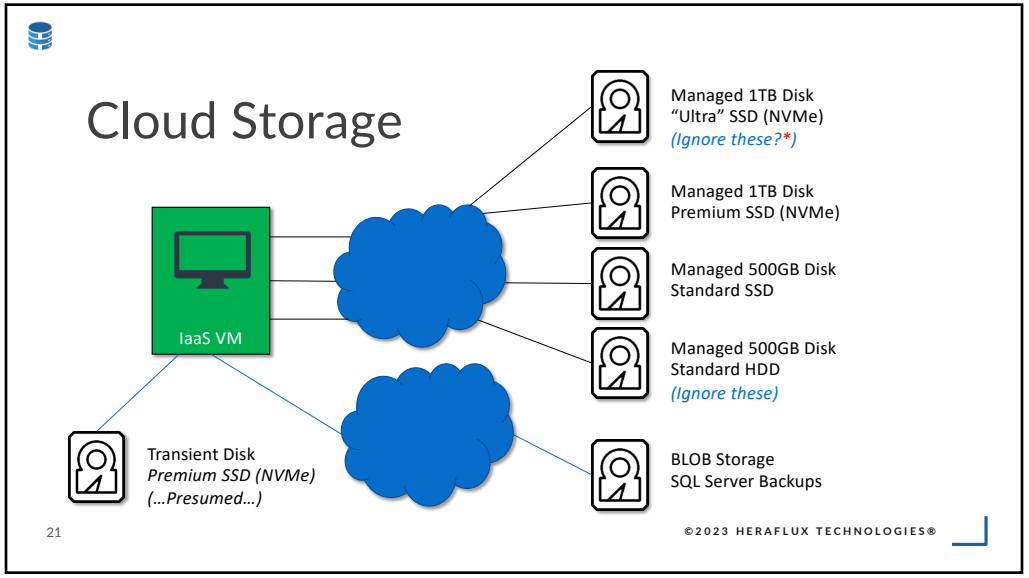


19

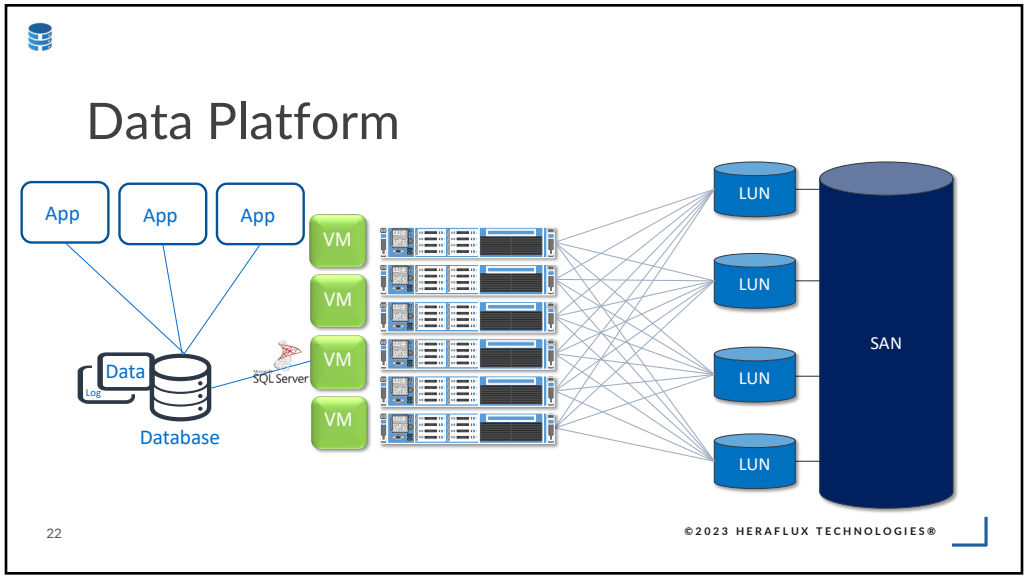


20

Do not organize...

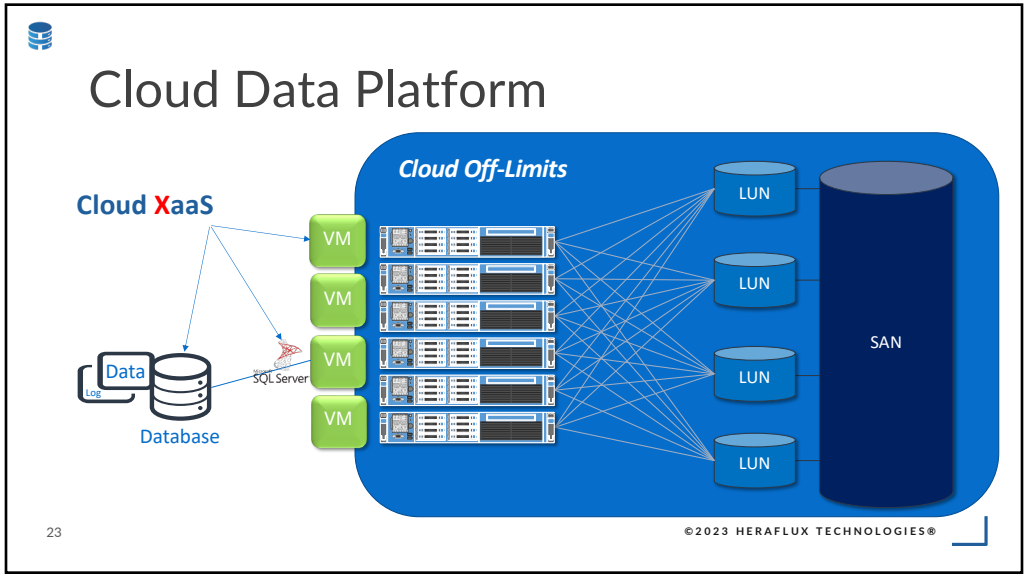


21



22

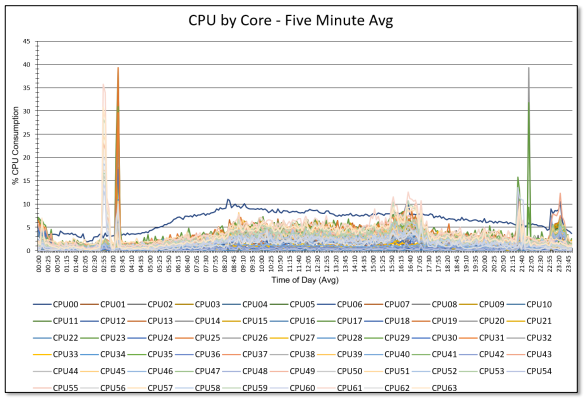
Do not
organize…



23

24

Do not
organi...



Baselines

- Identify consumption rates on allocated compute / storage
- "What are we actually using?"
- "How is the underlying platform responding to our demand?"
- "It's slow"
- Third-party tools
- Normal day vs. point-in-time

25

25

POWERSHELL BLGtoSQL

Load Perfmon data into SQL Server database for aggregation and analysis: github.com/Heraflux/BLGtoSQL

CAN RUN NIGHTLY

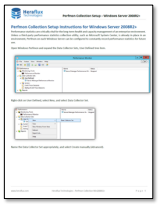
Can automate with Task Scheduler to run nightly

DATA IN RAW FORMAT

Can use T-SQL, PowerBI, SSRS, etc., for your own operational reporting and analysis

Windows Perfmon Analysis

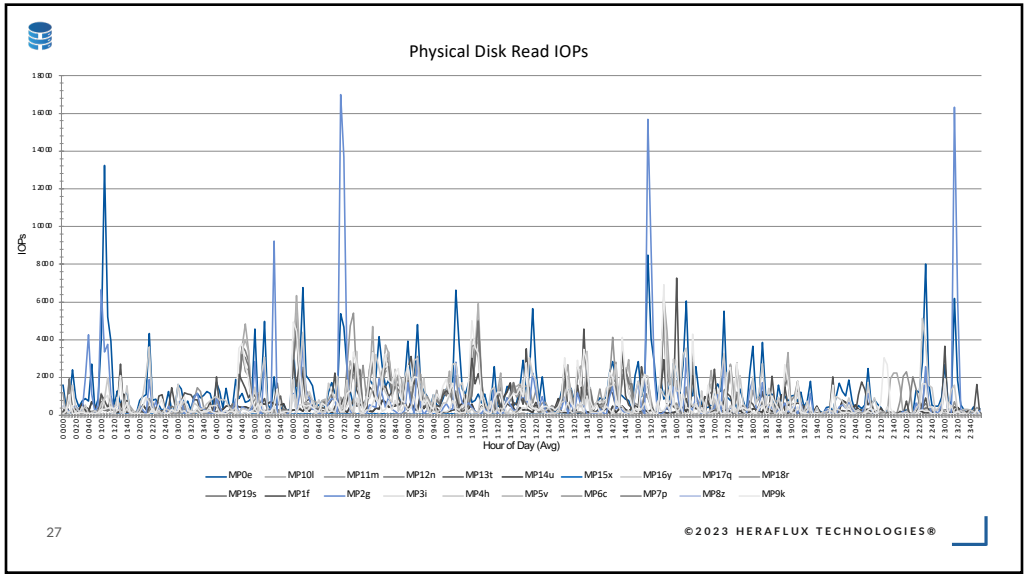
Load into SQL Server database



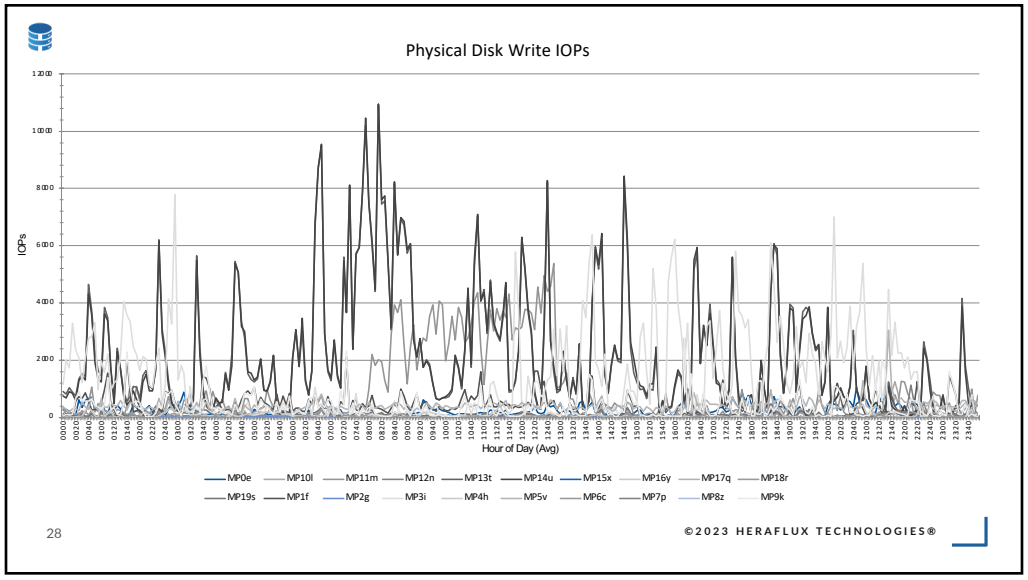
hfxte.ch/perfmon

26

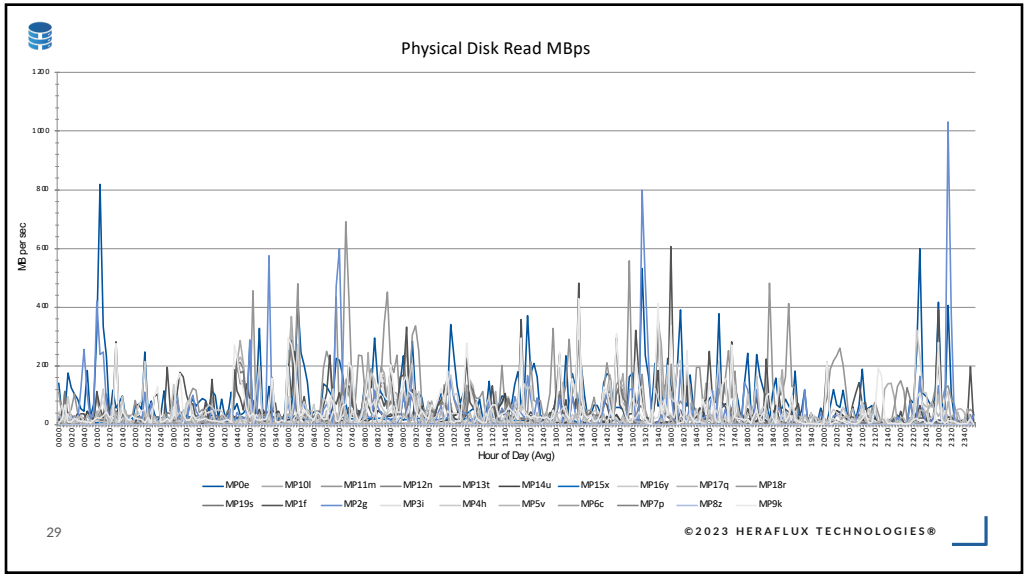
26


Physical Disk Read IOPs
IOPs
Hour of Day (Avg)
MP0e MP10l MP11m MP12n MP13t MP14u MP15x MP16y MP17q MP18r
MP19s MP1f MP2g MP3i MP4h MP5v MP6c MP7p MP8z MP9k
27
©2023 HERAFLUX TECHNOLOGIES®

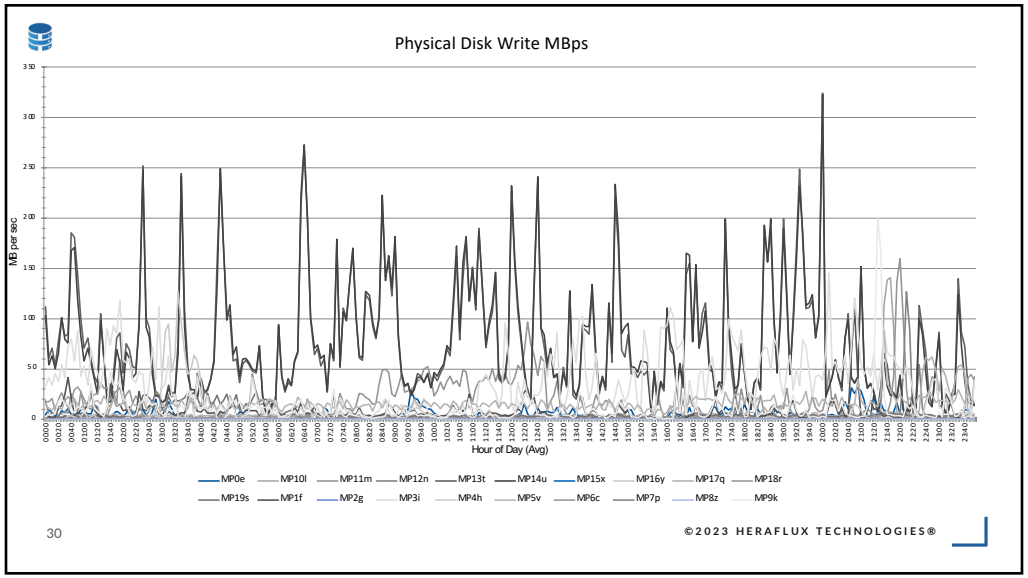
27


Physical Disk Write IOPs
IOPs
Hour of Day (Avg)
MP0e MP10l MP11m MP12n MP13t MP14u MP15x MP16y MP17q MP18r
MP19s MP1f MP2g MP3i MP4h MP5v MP6c MP7p MP8z MP9k
28
©2023 HERAFLUX TECHNOLOGIES®

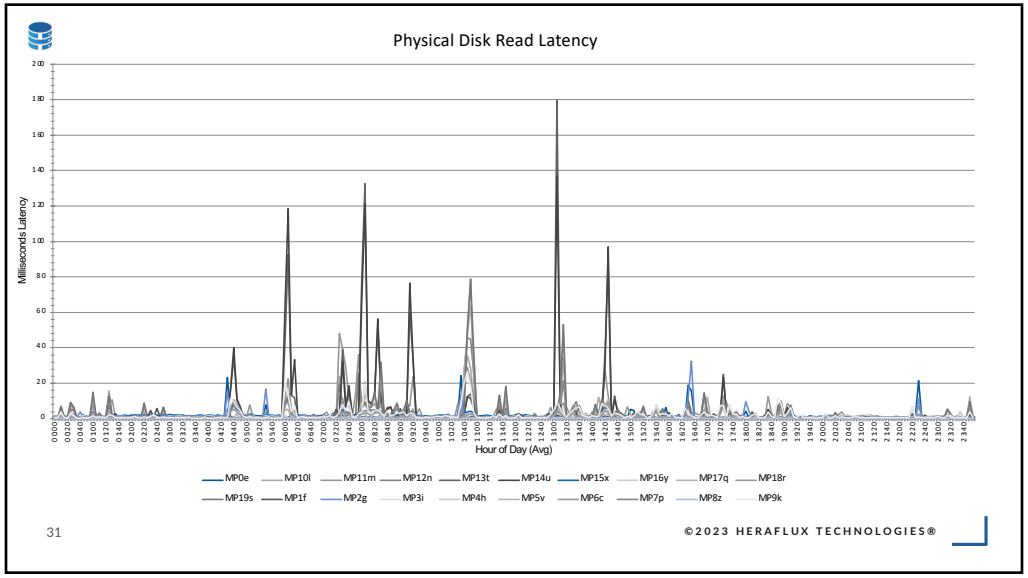
28


Physical Disk Read MBps
MB per sec
Hour of Day (Avg)
MP0e MP10l MP11m MP12n MP13t MP14u MP15x MP16y MP17q MP18r
MP19s MP1f MP2g MP3i MP4h MP5v MP6c MP7p MP8z MP9k
29
©2023 HERAFLUX TECHNOLOGIES®

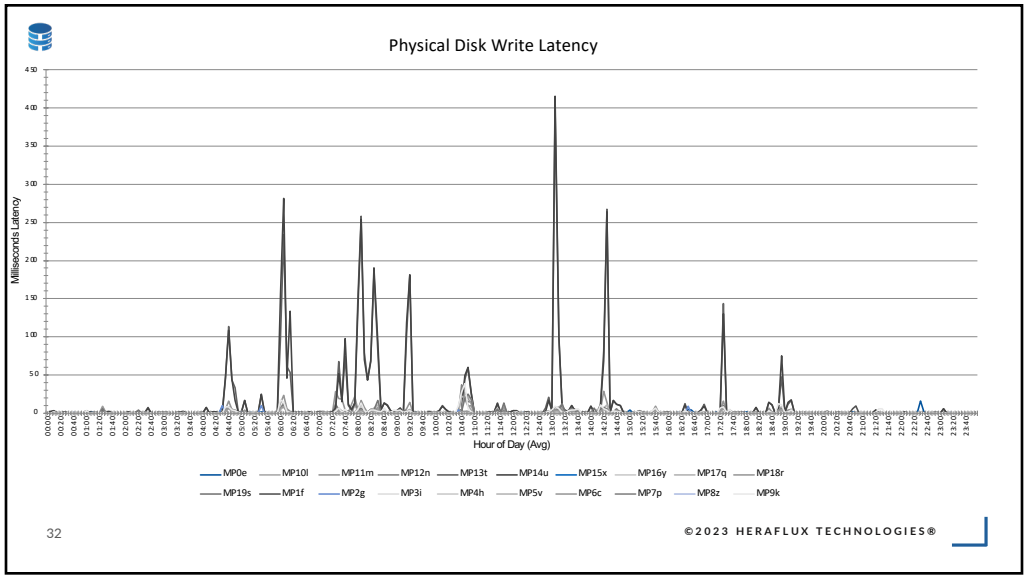
29


Physical Disk Write MBps
MB per sec
Hour of Day (Avg)
MP0e MP10l MP11m MP12n MP13t MP14u MP15x MP16y MP17q MP18r
MP19s MP1f MP2g MP3i MP4h MP5v MP6c MP7p MP8z MP9k
30
©2023 HERAFLUX TECHNOLOGIES®

30

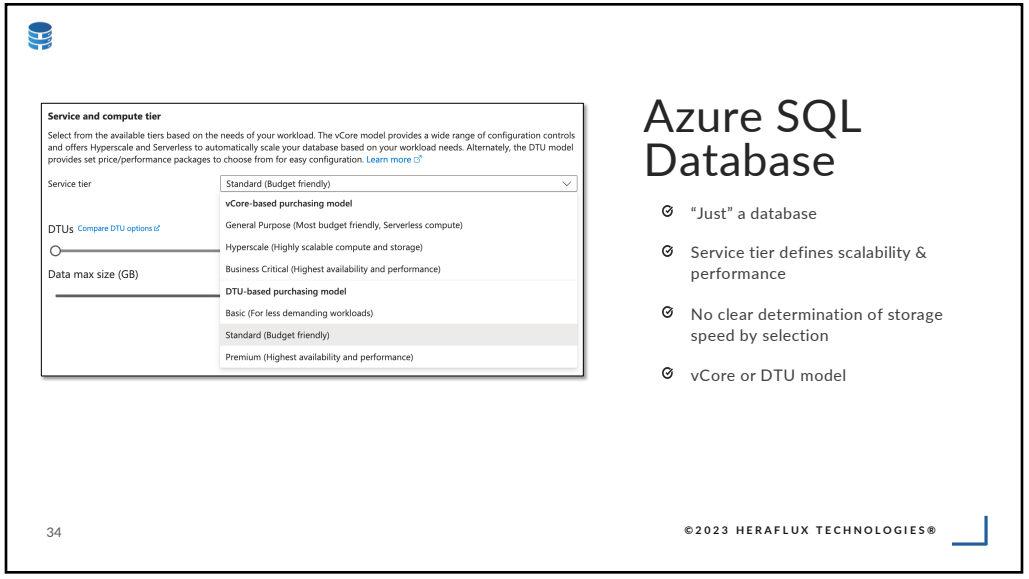

Physical Disk Read Latency
Milliseconds Latency
Hour of Day (Avg)
MP0e MP10l MP11m MP12n MP13t MP14u MP15x MP16y MP17q MP18r
MP19s MP1f MP2g MP3i MP4h MP5v MP6c MP7p MP8z MP9k
31
©2023 HERAFLUX TECHNOLOGIES®

31


Physical Disk Write Latency
Milliseconds Latency
Hour of Day (Avg)
MP0e MP10l MP11m MP12n MP13t MP14u MP15x MP16y MP17q MP18r
MP19s MP1f MP2g MP3i MP4h MP5v MP6c MP7p MP8z MP9k
32
©2023 HERAFLUX TECHNOLOGIES®

32

33



34

Do not organize...

DBaaS Storage

?? Disk
(Presumed dependent on deployment scale)

35

35

vCore Model

- Seems to be the "new" preferred model
- Underlying VM scale is listed
  - .... Sort of.

SQL hardware configuration
SQL database

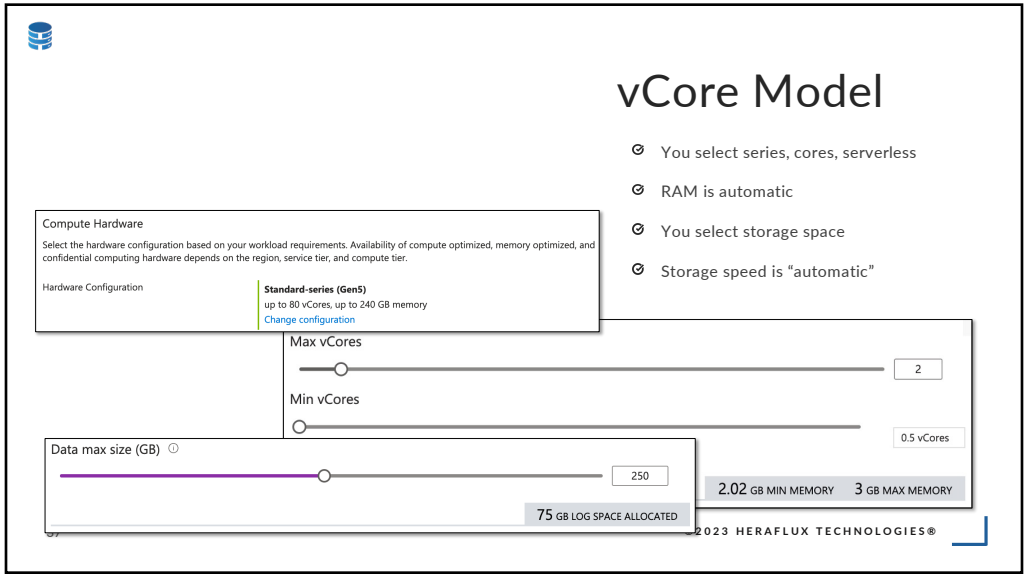
Available hardware configurations

Based on your workload requirements, select from available hardware configurations listed below.

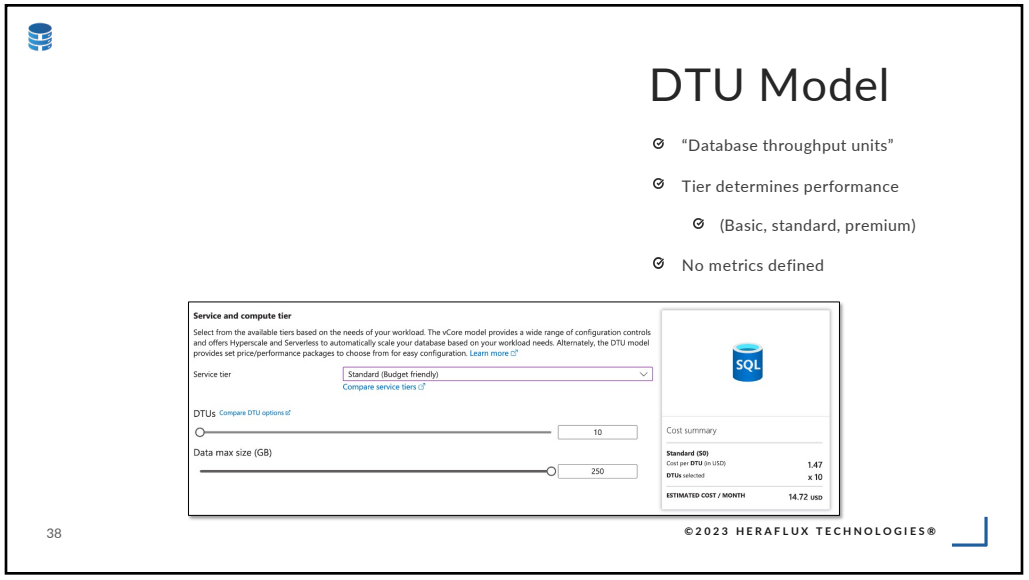
Configuration	Description	Max vCores	Max memory	Max storage	Compute cost / vCore / month
Standard-series (Gen5)	Balanced memory and compute	128	625 GB	4 TB	184.09 USD
M-Series	Memory optimized	--	--	--	--
FSv2-series	Compute optimized	--	--	--	--
DC-series	Enables confidential computing	--	--	--	--

36

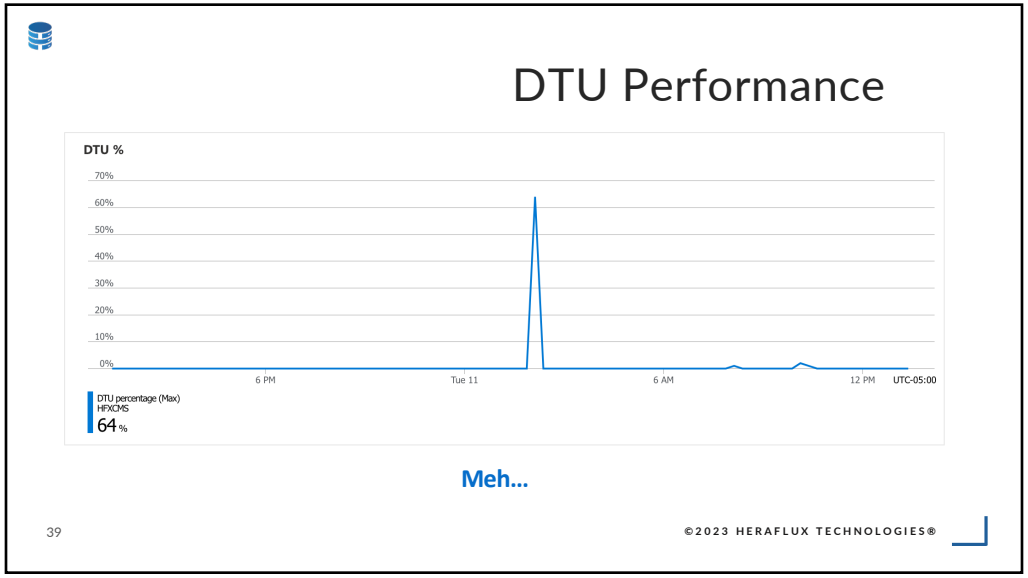
36


vCore Model
You select series, cores, serverless
RAM is automatic
You select storage space
Storage speed is “automatic”
Compute Hardware
Select the hardware configuration based on your workload requirements. Availability of compute optimized, memory optimized, and confidential computing hardware depends on the region, service tier, and compute tier.
Hardware Configuration
Standard-series (Gen5)
up to 80 vCores, up to 240 GB memory
Change configuration
Max vCores
2
Min vCores
0.5 vCores
Data max size (GB)
250
2.02 GB MIN MEMORY
3 GB MAX MEMORY
75 GB LOG SPACE ALLOCATED
©2023 HERAFLUX TECHNOLOGIES®

37


DTU Model
“Database throughput units”
Tier determines performance
(Basic, standard, premium)
No metrics defined
Service and compute tier
Select from the available tiers based on the needs of your workload. The vCore model provides a wide range of configuration controls and offers Hyperscale and Serverless to automatically scale your database based on your workload needs. Alternately, the DTU model provides set price/performance packages to choose from for easy configuration. Learn more
Service tier
Standard (Budget friendly)
Compare service tiers
DTUs Compare DTU options
10
Data max size (GB)
250
Cost summary
Standard (S0)
Cost per DTU (in USD)
1.47
DTUs selected
x 10
ESTIMATED COST / MONTH
14.72 USD
38
©2023 HERAFLUX TECHNOLOGIES®

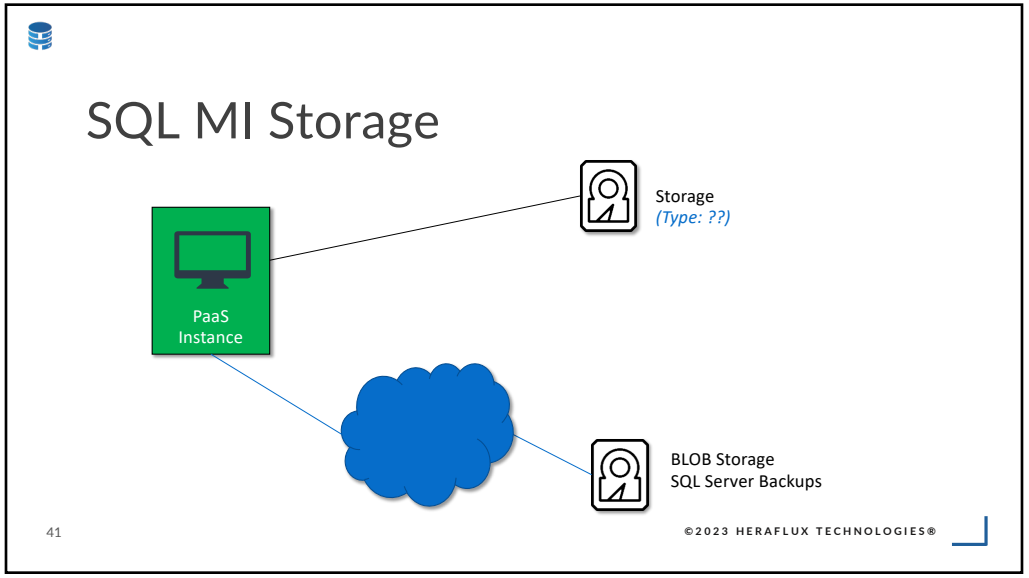
38


DTU Performance
DTU %
70%
60%
50%
40%
30%
20%
10%
0%
6 PM
Tue 11
6 AM
12 PM
UTC-05:00
DTU percentage (Max)
HFXCMS
64 %
Meh...
39
©2023 HERAFLUX TECHNOLOGIES®

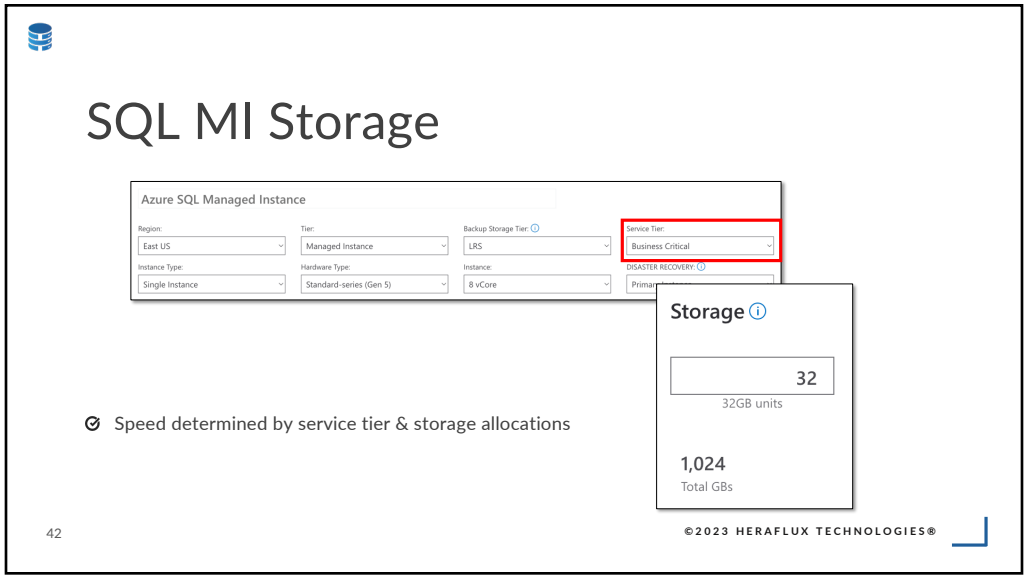
39


4
SQL MI
Options
?
40

40

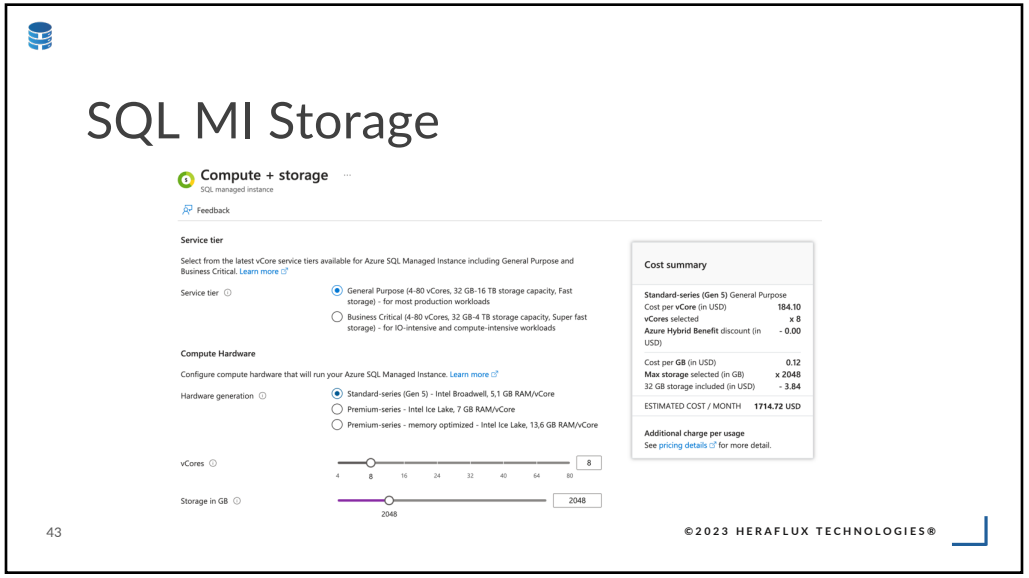


41



42

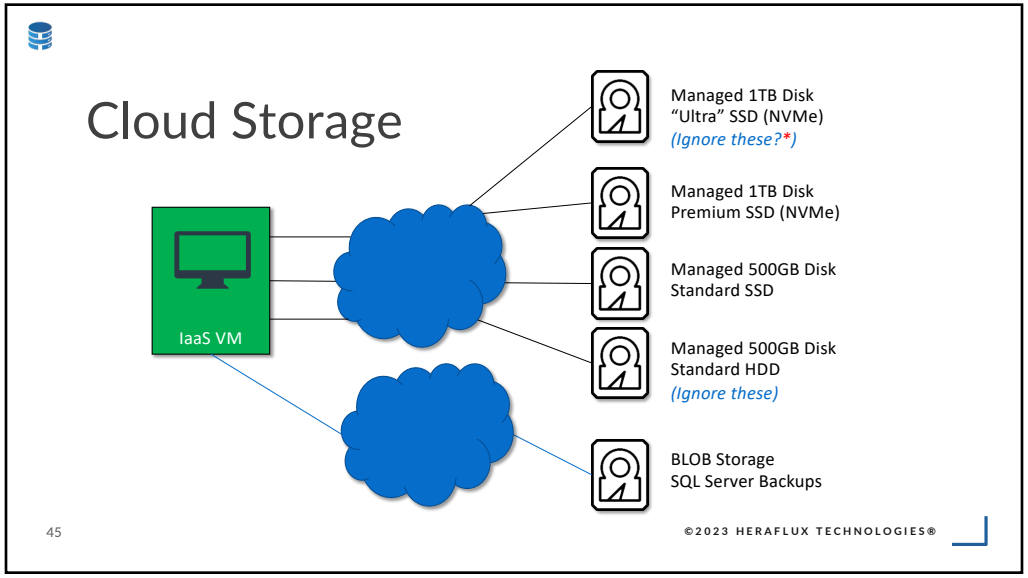
Do not
organize...


SQL MI Storage
Compute + storage
SQL managed instance
Feedback
Service tier
Select from the latest vCore service tiers available for Azure SQL Managed Instance including General Purpose and Business Critical. Learn more
Service tier
General Purpose (4-80 vCores, 32 GB-16 TB storage capacity, Fast storage) - for most production workloads
Business Critical (4-80 vCores, 32 GB-4 TB storage capacity, Super fast storage) - for IO-intensive and compute-intensive workloads
Compute Hardware
Configure compute hardware that will run your Azure SQL Managed Instance. Learn more
Hardware generation
Standard-series (Gen 5) - Intel Broadwell, 5,1 GB RAM/vCore
Premium-series - Intel Ice Lake, 7 GB RAM/vCore
Premium-series - memory optimized - Intel Ice Lake, 13,6 GB RAM/vCore
vCores
4 8 16 24 32 40 64 80
8
Storage in GB
2048
2048
Cost summary
Standard-series (Gen 5) General Purpose
Cost per vCore (in USD) 184.10
vCores selected x 8
Azure Hybrid Benefit discount (in USD) - 0.00
Cost per GB (in USD) 0.12
Max storage selected (in GB) x 2048
32 GB storage included (in USD) - 3.84
ESTIMATED COST / MONTH 1714.72 USD
Additional charge per usage
See pricing details for more detail.
43
©2023 HERAFLUX TECHNOLOGIES®

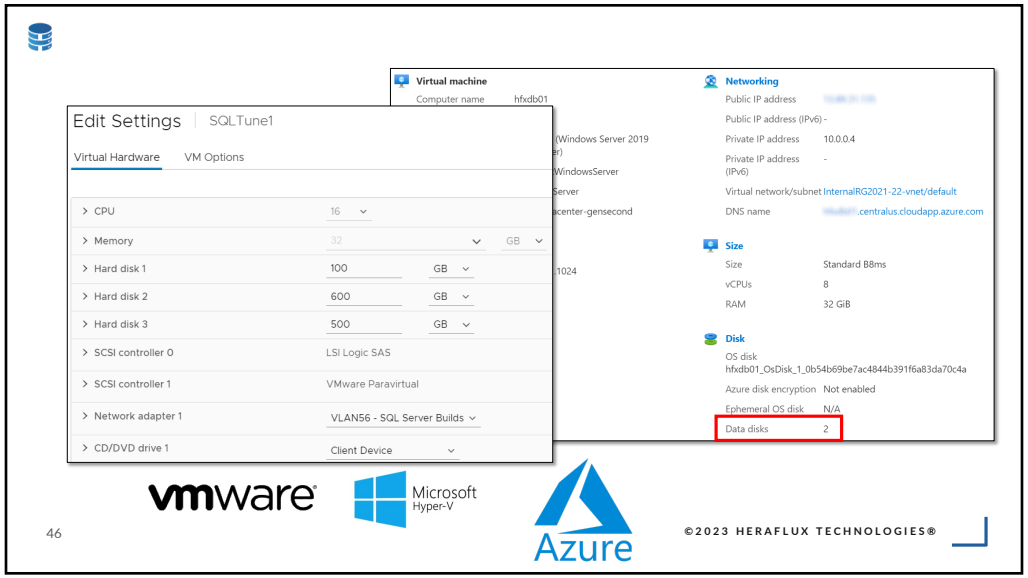
43


5
SQL VM Options
?
44

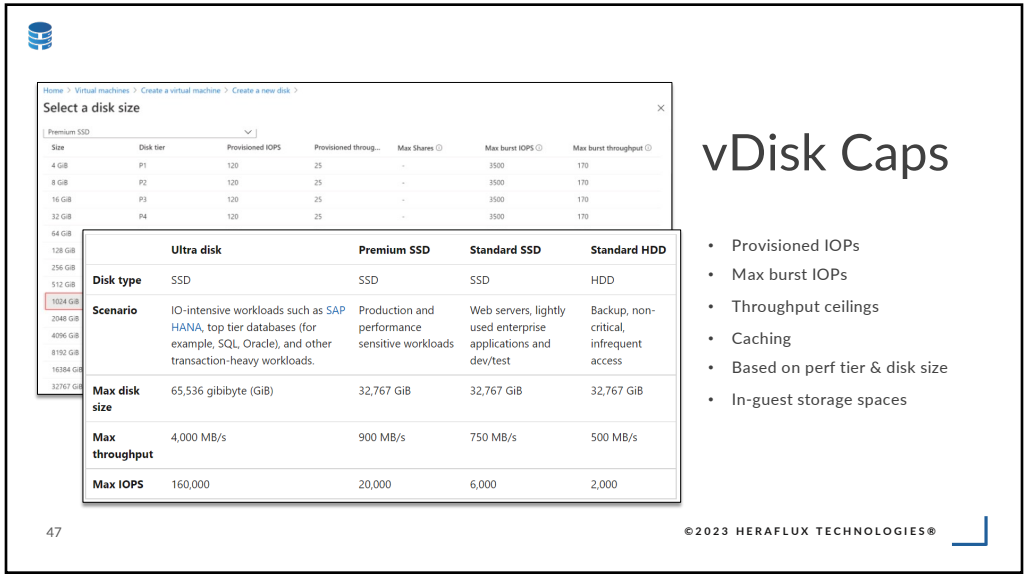
44


Cloud Storage
IaaS VM
Managed 1TB Disk
"Ultra" SSD (NVMe)
(Ignore these?*)
Managed 1TB Disk
Premium SSD (NVMe)
Managed 500GB Disk
Standard SSD
Managed 500GB Disk
Standard HDD
(Ignore these)
BLOB Storage
SQL Server Backups
45
©2023 HERAFLUX TECHNOLOGIES®

45


Edit Settings
SQLTune1
Virtual Hardware
VM Options
CPU 16
Memory 32 GB
Hard disk 1 100 GB
Hard disk 2 600 GB
Hard disk 3 500 GB
SCSI controller 0 LSI Logic SAS
SCSI controller 1 VMware Paravirtual
Network adapter 1 VLAN56 - SQL Server Builds
CD/DVD drive 1 Client Device
Virtual machine
Computer name hfxdb01
Networking
Public IP address
Public IP address (IPv6) -
Private IP address 10.0.0.4
Private IP address (IPv6) -
Virtual network/subnet InternalRG2021-22-vnet/default
DNS name .centralus.cloudapp.azure.com
Size
Size Standard B8ms
vCPUs 8
RAM 32 GiB
Disk
OS disk hfxdb01_OsDisk_1_0b54b69be7ac4844b391f6a83da70c4a
Azure disk encryption Not enabled
Ephemeral OS disk N/A
Data disks 2
vmware
Microsoft Hyper-V
Azure
46
©2023 HERAFLUX TECHNOLOGIES®

46



vDisk Caps

- Provisioned IOPs
- Max burst IOPs
- Throughput ceilings
- Caching
- Based on perf tier & disk size
- In-guest storage spaces

47

47

Size	vCPU	Memory: GiB	Temp storage (SSD) GiB	Max data disks	Max uncached disk throughput: IOPS/MBps	Max burst uncached disk throughput: IOPS/MBps[3]	Max NICs	Max network bandwidth (Mbps)
Standard_D2s_v5[1,2]	2	8	Remote Storage Only	4	3750/85	10000/1200	2	12500
Standard_D4s_v5	4	16	Remote Storage Only	8	6400/145	20000/1200	2	12500
Standard_D8s_v5	8	32	Remote Storage Only	16	12800/290	20000/1200	4	12500
Standard_D16s_v5	16	64	Remote Storage Only	32	25600/600	40000/1200	8	12500
Standard_D32s_v5	32	128	Remote Storage Only	32	51200/865	80000/2000	8	16000
Standard_D48s_v5	48	192	Remote Storage Only	32	76800/1315	80000/3000	8	24000
Standard_D64s_v5	64	256	Remote Storage Only	32	80000/1735	80000/3000	8	30000
Standard_D96s_v5	96	384	Remote Storage Only	32	80000/2600	80000/4000	8	35000

https://docs.microsoft.com/en-us/azure/virtual-machines/dv5-dsv5-series

VM Caps

- VM-level
- IOPs ceilings
- Throughput ceiling
- Cumulative for all vDisks

48

48

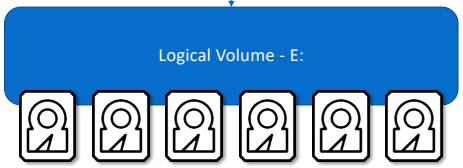
VM Tips

- TempDB on local ephemeral storage
  - Very low latency D:
- Data & logs on separate drives
- Host read-only caching for data file disks
- NO read caching for log file disks
- DO NOT ENABLE write caching – ever
- Disk bursting P1-P20 scale not for production
- 64KB NTFS allocation unit sizes

49

https://learn.microsoft.com/en-us/azure/azure-sql/virtual-machines/windows/performance-guidelines-best-practices-storage?view=azuresql

49



Storage Spaces

- Think software drive RAID
- Work around vDisk ceilings without provisioning lots of space
- Add space to pool as needed

50

50

Terms

- **Resiliency Types**
  - Simple (RAID-0)
  - Mirror (RAID-1)
  - Parity (RAID-5)
- **Storage Pool**
  - Group of physical disks in RAID setup
- **Virtual Disk**
  - Logical drive (not formatted)
- **Volume**
  - Formatted virtual disk
- **Column**
  - Arrangement of blocked storage
- Reference: learn.microsoft.com/en-us/windows-server/storage/storage-spaces/deploy-standalone-storage-spaces

51

51

Gotchya's

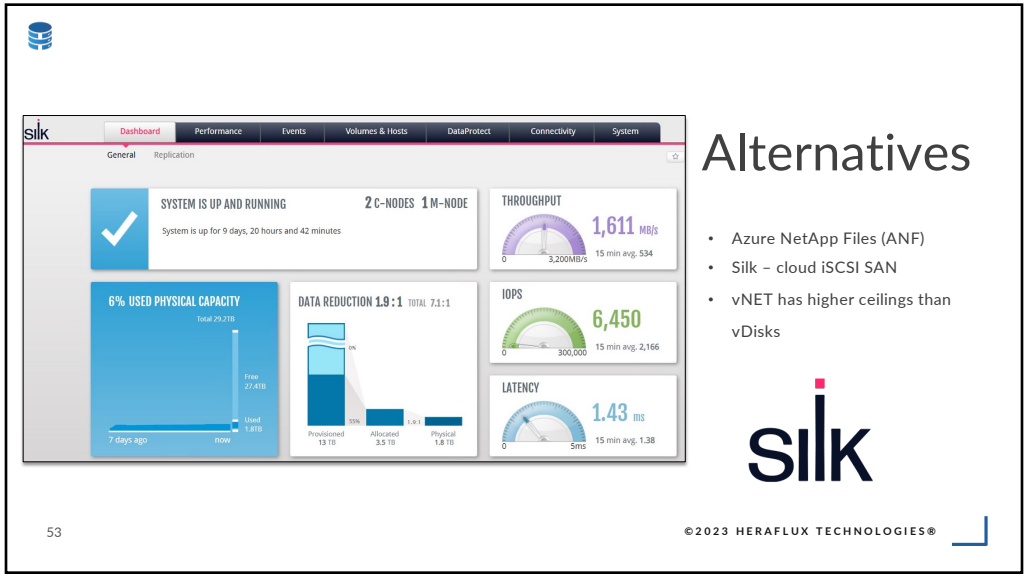
- Columns
- Create disk pool of two drives
- Add one more
- Can't add to pool
- Add one more
- Now can add both
- Reference: bit.ly/3LZU0FQ
- Max speed = sum of # disks in single column

```
PS C:\> get-virtualdisk

FriendlyName ResiliencySettingName FaultDomainRedundancy OperationalStatus HealthStatus   Size FootprintOnPool StorageEfficiency
------------ --------------------- --------------------- ----------------- ------------ ---- -------------- ----------------
VDisk01      Simple                0                     OK                Healthy      996 GB          997 GB            99.90%
```

```
PS C:\> get-virtualdisk | format-list

ObjectId                          : {1}\\WIN2TEST\root/Microsoft/Windows/Storage/Providers_v2\SPACES_VirtualDisk.ObjectId="{c97f53d7-
                                    5c3b-11ec-a6ec-806e6f6e6963}:VD:{1aa53a07-0a8e-453a-8ae8-9b1eca70d1ef}{cce72052-7163-40f5-8f65-074
                                    c8372cf99}"
PassThroughClass                  :
PassThroughIds                    :
PassThroughNamespace              :
PassThroughServer                 :
UniqueId                          : 5220E7CC6371F5408F65074C8372CF99
Access                            : Read/Write
AllocatedSize                     : 1069448856704
AllocationUnitSize                : 1073741824
ColumnIsolation                   : PhysicalDisk
DetachedReason                    : None
FaultDomainAwareness              : PhysicalDisk
FootprintOnPool                   : 1070520598528
FriendlyName                      : VDisk01
HealthStatus                      : Healthy
Interleave                        : 262144
IsDeduplicationEnabled            : False
IsEnclosureAware                  : False
IsManualAttach                    : False
IsSnapshot                        : False
IsTiered                          : False
LogicalSectorSize                 : 512
MaxIoBandwidth                    : 0
MaxIops                           : 0
MediaType                         : Unspecified
Name                              :
NameFormat                        :
NumberOfAvailableCopies           :
NumberOfColumns                   : 2
NumberOfDataCopies                : 1
NumberOfGroups                    : 1
```

52

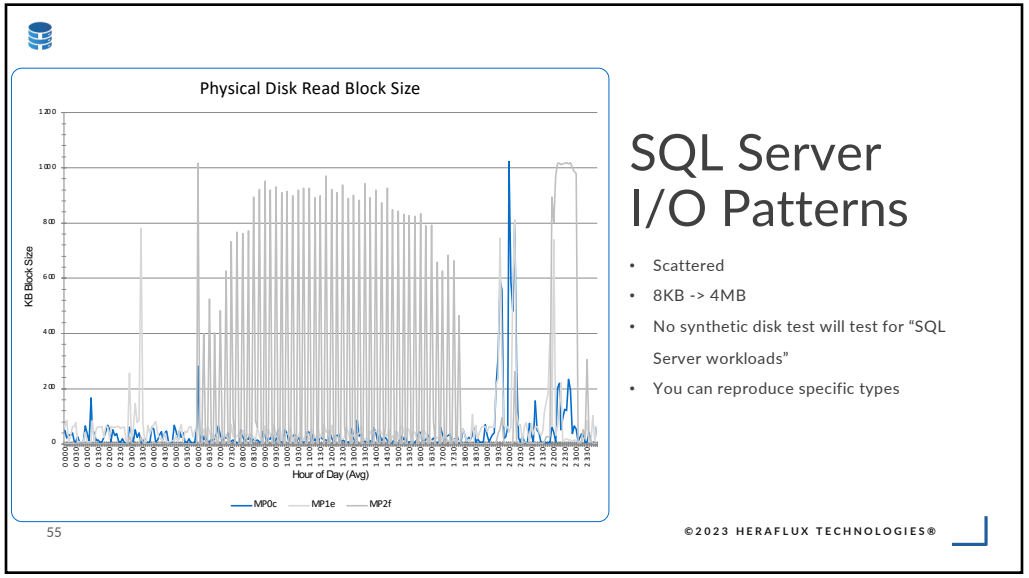
52


Dashboard
Performance
Events
Volumes & Hosts
DataProtect
Connectivity
System
General
Replication
SYSTEM IS UP AND RUNNING
2 C-NODES 1 M-NODE
System is up for 9 days, 20 hours and 42 minutes
THROUGHPUT
1,611 MB/s
15 min avg. 534
6% USED PHYSICAL CAPACITY
DATA REDUCTION 1.9 : 1
IOPS
6,450
15 min avg. 2,166
LATENCY
1.43 ms
15 min avg. 1.38
Alternatives
Azure NetApp Files (ANF)
Silk – cloud iSCSI SAN
vNET has higher ceilings than vDisks
53
©2023 HERAFLUX TECHNOLOGIES®

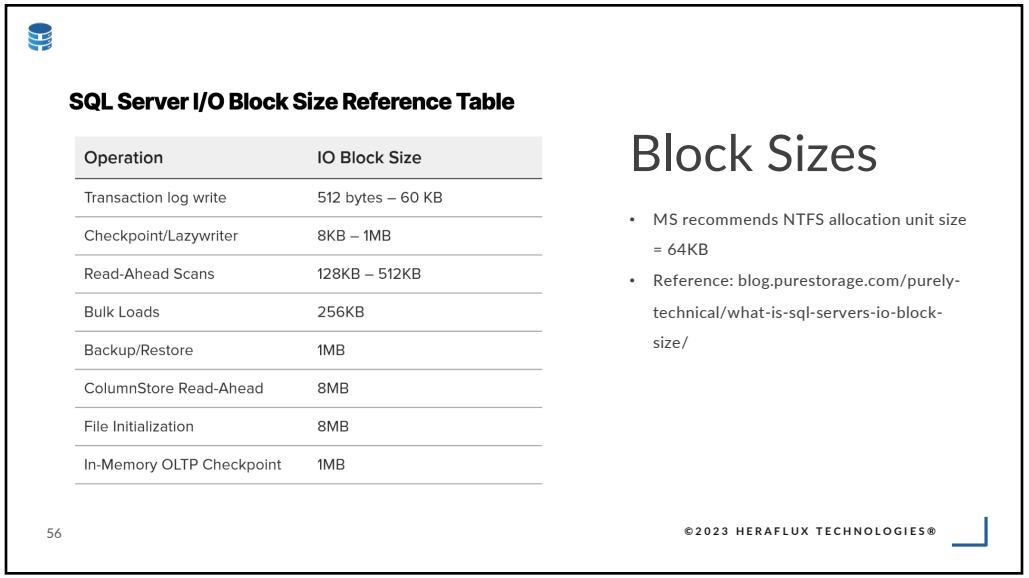
53


6
Testing
54

54

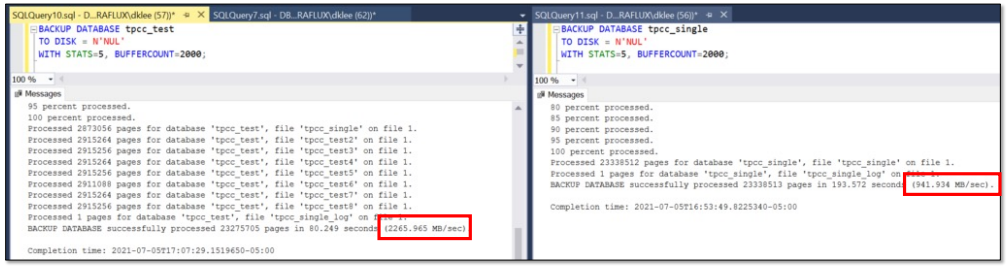

Physical Disk Read Block Size
KB Block Size
Hour of Day (Avg)
MP0c
MP1e
MP2f
SQL Server I/O Patterns
• Scattered
• 8KB -> 4MB
• No synthetic disk test will test for "SQL Server workloads"
• You can reproduce specific types
55
©2023 HERAFLUX TECHNOLOGIES®

55


SQL Server I/O Block Size Reference Table
Operation | IO Block Size
Transaction log write | 512 bytes – 60 KB
Checkpoint/Lazywriter | 8KB – 1MB
Read-Ahead Scans | 128KB – 512KB
Bulk Loads | 256KB
Backup/Restore | 1MB
ColumnStore Read-Ahead | 8MB
File Initialization | 8MB
In-Memory OLTP Checkpoint | 1MB
Block Sizes
• MS recommends NTFS allocation unit size = 64KB
• Reference: blog.purestorage.com/purely-technical/what-is-sql-servers-io-block-size/
56
©2023 HERAFLUX TECHNOLOGIES®

56

Disk READ Synthetic Test



After 8 Data Files | 1 Data File

57

57

DiskSpd

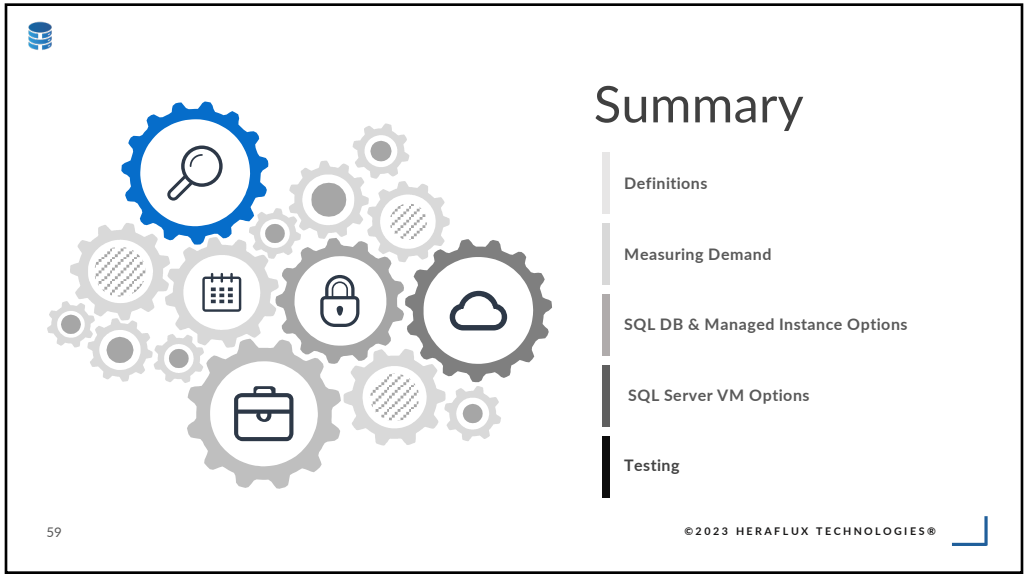
```
diskspd -b64K -d5 -h -L -o2 -t4 -r -w20 -c10G -z512 c:\diskspd\io.dat
```

- github.com/Microsoft/diskspd
- NOT A SQL SERVER WORKLOAD SIMULATOR
- But can generate specific patterns to try to mimic certain functions
- github.com/microsoft/diskspd/wiki/Command-line-and-parameters

Param	Description	Value
-b	Block Size	64KB
-d	Duration (seconds)	60+
-Sh	Disable caching (-h)	
-L	Measure latency stats	
-o	Operations per thread	2+
-t	Number of threads	Core count?
-r	Random I/O block alignment	4K
-w	Write/read percentage	20
-c	Target file size	10G = 10GB
-z	Randomization entropy	512

58

58



59

60

Do not
organize...

61