External Storage and IBM i
Sizing and modelling

Jana Jamsek
Advanced technical skills, Europe

Agenda

- IBM i architecture and external storage
- Storage systems that connect to IBM i
- Steps for sizing and modelling Storage systems for IBM i
- Disk Magic
- Collecting IBM i performance data
- Decide the peaks
- Sizing - decide the initial storage configuration for Disk Magic
- Modelling with Disk Magic
- Demo of Disk Magic
IBM i Single level storage, Handling I/O operations

Main memory

Disk space

Disk drives or LUNs

Page swap, close files, journals, etc.

Page fault

4 KB pages

One set of virtual addresses for memory and disk space

IBM i Single level storage, Handling I/O operations

Spreading IBM i objects accross disk units - LUNs

IBM i objects

User profiles

Database files

Job descriptions

Documents

Etc.

Disk units or LUNs

IBM i storage management spreads IBM i objects evenly across the available disk units or LUNs
Storage Manager – Handling IO operations

Storage management directories

Map disk and sector to a virtual address

Page tables

Virtual addresses

Pages

Internal disk

LUN on external storage

IO flow – Storage system connected with VIOS-NPIV

POWER

Virtual I/O server

Server virtual FC adapter

Device driver

POWER Hypervisor

FC adapter

IBM i client

Client virtual FC adapter

FC adapter

SAN

Storage system

Host port

Host port

Cache

LUN

Disk drives
IBM i and Storage systems

Note: The shown connections with VIOS refer to either VSCSI, NPIV or both.

DS8000 Enterprise disk system

The IBM POWER processor has been behind the success of IBM enterprise storage beginning with the Enterprise Storage Server in 1999.

DS8000 builds on a market-proven, reliable code base!
**DS8870 models**

- **DS8870 Model 961 Enterprise model**
  - Dual 4-core, dual 8-core, or dual 16-core processor complex
  - 64 GB to 1 TB cache
  - 8 Gb/s host and device adapters
  - 2.5" Enterprise SAS-2 drives

- **DS8870 Model 961 Business Class model**
  - Dual 2-core processor complex
  - 16 GB or 32 GB cache
  - 8 Gb/s host and device adapters
  - 2.5" Enterprise SAS-2 drives

- **DS8870 Model 96E Expansion frame**
  - 8 Gb/s host and device adapters
  - 2.5" Enterprise SAS-2 drives

**DS8870 Components**

- DS8870 front view
- DS8870 back view

- Disk drives
- HMC
- DC-UPS Power supplies
- POWER7 processor based servers
- I/O Enclosures with adapters
**Disk Drives in DS8870**

- 146 GB 15 K RPM – Small Form Factor (SFF)
- 300 GB 15 K RPM - SFF
- 600 GB 10 K RPM - SFF
- 900 GB 10 K RPM - SFF
- 3 TB 7.2 K RPM nearline SAS drives – Large Form Factor (LFF)
- 400 GB SSD

- RAID support:
  - SFF disk drives support RAID-5, RAID-6 and RAID-10
  - SSD support RAID-5
  - LFF disk drives (nearline disk drives) support RAID-6
  - RAID protection is enabled on Array site of 8 disk drives

**DS8870 Graphical User Interface (GUI)**
IBM Storwize V7000 – Responding to Our Customers

- Requests for SAN Volume Controller
  - SVC with internal disk drives
    - Robust and reliable virtualized storage product with RAID integration and simplified system management (only one interface for everything)
  - Hide SVC virtualization complexity
  - Easy to use graphical user interface

- Requests for midrange disk
  - Easy system management
  - Enterprise attributes
  - Simplified data migration
  - Robustness, high availability
  - Strong interoperability
  - Application integration
  - Low cost

IBM Storwize V7000 – Responding to Our Customers

IBM Storwize V7000 at a Glance

Enterprise level software capabilities

- RAID 0, 1, 5, 6, 10
- Storage virtualization
- Non-disruptive data migration
- New user interface
- Thin provisioning
- Easy Tier
- FlashCopy
- Remote copy
- Disaster recovery automation

Modular Hardware Building Blocks

Control enclosures
Expansion enclosures

Enclosure front View
12 x 3.5” LFF Drive Bays Model 2076-112/212/312

24 x 2.5” SFF Drive Bays Model 2076-124/224/324

© 2013 IBM Corporation
Storwize V7000 Drive Options

<table>
<thead>
<tr>
<th>Drive Type</th>
<th>Drive Sizes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 2076 124 / 224 / 324: 2.5&quot; Small Form Factor (SFF)</td>
<td>SSD: 300GB (MLC) 200-400GB (MLC)</td>
</tr>
<tr>
<td></td>
<td>10K RPM SAS: 300, 450, 600GB 900GB</td>
</tr>
<tr>
<td></td>
<td>15K RPM SAS: 146GB 300GB</td>
</tr>
<tr>
<td></td>
<td>7.2K RPM NL-SAS: 1TB 1.2TB</td>
</tr>
</tbody>
</table>

No restrictions on mixing of drive types within the same enclosure at this time.

<table>
<thead>
<tr>
<th>Drive Type</th>
<th>Drive Sizes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 2076 112 / 212 / 312: 3.5&quot; Large Form Factor (LFF)</td>
<td>7.2K RPM NL-SAS: 2TB 3TB 4TB</td>
</tr>
</tbody>
</table>

Storwize V7000 Components

Control Enclosure Rear View 2076-112/124/312/324

Bottom node canister. Notice the unit is identical to the top canister but installed upside down.

The power supply unit or PSU for the control enclosure includes batteries to protect the cache in the node canisters. Notice the units are identical except the unit on the right is installed upside down.

Top node canister.

Expansion Enclosure Rear View 2076-212/224

Bottom expansion canister. Notice the unit is identical to the top canister but installed upside down.

The power supply unit or PSU for the expansion enclosure does not include batteries and thus are not interchangeable with control enclosure PSUs. Notice the units are identical except the unit on the right is installed upside down.

Top expansion canister.
Sizing and modelling External storage for IBM i

Collecting data

Decide the peaks

Disk drives
RAID level
Capacity

Decide initial Storage configuration for Disk Magic

Host adapters
IBM i FC adapters
Cache

Model with Disk Magic

Disk Magic

- Modelling tool to help estimate disk subsystem performance
- Developed and maintained by the company IntelliMagic
- Models IBM, HP, HDS and EMC storage
- IntelliMagic closely collaborates with IBM development and performance teams
- Standard tool for sizing and modelling IBM Storage systems
- Supports modeling of IBM i with Storage systems since 2003
- Disk Magic can be obtained from the following web sites:
  - IBMers:
    [https://w3-03.sso.ibm.com/sales/support/ShowDoc.wss?docid=SSPCQ048068H83479I86](https://w3-03.sso.ibm.com/sales/support/ShowDoc.wss?docid=SSPCQ048068H83479I86)
  - Business partners:
- Customers can purchase the Disk Magic license from Intellimagic
IBM i performance data for sizing and Disk Magic modelling

- Collection Services data
  - Collected with 5 minutes intervals
  - Collected during 3 consecutive days
  - Collected also during intensive end-of-month jobs, if needed

- Needed Performance reports
  - System report / section Disk Utilization
  - Resource report / section Disk Utilization
  - Component report / section Disk Activity

- Performance reports must be in English language

- Performance reports must be transferred in .txt form to the PC with Disk Magic

- If a particular period of the day is important for the customer, create reports containing data of this period only
Collecting IBM i Collection Services data with Performance Tools (5770-PT1)

Creating reports using Performance Tools

Example: Creating System report / section Disk utilization for a period 12:00 to 12:30
Collecting IBM i Collection Services data without Performance Tools

- CFGPFRCOL (Configure Performance Collection)
- STRPFRCOL (Start Performance Collection)
- ENDPFRCOL (End Performance Collection)

Example: Configuring performance collection with 5 minutes intervals

Creating reports without Performance Tools

- PRTSYSRPT (Print System Report)
- PRTRSCRPT (Print Resource Report)
- PRTCPTRPT (Print Component Report)

Example: Creating System report / Disk utilization from 4. November 00:00 to 8:00
Transfer the reports to .txt file on your PC

Use IBM i Operations Navigator

Collecting data

Decide the peaks

Disk drives

RAID level

Capacity

Decide initial Storage configuration for Disk Magic

Model with Disk Magic

Cache

IBM i FC adapters

Host adapters
Decide for the peaks - 1

Decide for the peaks - 2

Recommend: Size for both peaks
Collecting data

Decide the peaks

Disk drives
RAID level
Capacity

Storage configuration for Disk Magic

Cache
IBM i FC adapters
Host adapters

Model with Disk Magic

Hard Disk Drives

- Requirements for capacity
- Guidelines for disk arms
Guidelines for RAID level

- RAID-10 provides better resiliency
- RAID-10 provides generally better performance:
  - RAID-5 results in 4 disk operations per write – higher penalty
  - RAID-10 results in 2 disk operations per write – lower penalty
- RAID-10 requires more capacity
- In DS8000 use RAID-10 when:
  - There are many random writes
  - Write cache efficiency is low
  - Huge workload
- In Midrange storage and Storwize V7000 use RAID-10 when:
  - There are more than 50% writes

Guideline - Number of DS8000 ranks of 8 disk drives for an IBM i workload

<table>
<thead>
<tr>
<th>Max host IO/sec for HDD rank with IOP-less adapters</th>
<th>Host iops at 70% Read</th>
<th>Host iops at 50% Read</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAS DDM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAID-5 15 K RPM</td>
<td>940</td>
<td>731</td>
</tr>
<tr>
<td>RAID-10 15 K RPM</td>
<td>1253</td>
<td>1116</td>
</tr>
<tr>
<td>RAID-6 15 K RPM</td>
<td>723</td>
<td>526</td>
</tr>
</tbody>
</table>

Example:
- IBM i workload with 10000 IO/sec, 70% reads, we are planning 15 K RPM disk drives in RAID-5
- Calculation for the ranks: 10000 / 940 = app 10 ranks
Guideline - Number of V7000 disk drives for an IBM i workload

Quick calculation for the number of disk drives for an IBM i workload

<table>
<thead>
<tr>
<th>Maxhots IO/sec per disk drive</th>
<th>Host iops at 70% Read</th>
<th>Host iops at 50% Read</th>
</tr>
</thead>
<tbody>
<tr>
<td>15 K RPM disk drive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAID-1 or RAID-10</td>
<td>138</td>
<td>122</td>
</tr>
<tr>
<td>RAID-5</td>
<td>96</td>
<td>75</td>
</tr>
<tr>
<td>10 K RPM disk drive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAID-1 or RAID-10</td>
<td>92</td>
<td>82</td>
</tr>
<tr>
<td>RAID-5</td>
<td>64</td>
<td>50</td>
</tr>
</tbody>
</table>

Example:

- IBM i workload with 3000 IO/sec, 50% reads, we are planning 15 K RPM disk drives in RAID-10
- Calculation for the disk drives: 3000 / 122 = app 24 disk drives in RAID-10

Guidelines for cache size in external storage

- To understand the needed cache size determine the type of workload
  - Random or sequential
  - Read/write ratio
  - Is the workload cache friendly

- Modelling with Disk Magic

- Rough guidelines for DS8800
  - 10 to 20 TB capacity: 64GB cache
  - 20 to 50 TB: 128 GB cache
  - > 50 TB: 256 GB to 1TB cache

- V7000:
  - Fixed cache size of 16 GB per node pair
LUNs – number and size

- The more LUNs the better performance
- Make sure that sufficient disk arms are provided to support the LUNs
- Guideline for sizes: 2 to 4 LUNs per disk drive
- Use one size of LUNs (or minimum 2 sizes) for an IBM i ASP
- Consider minimal capacity of LoadSource
  - Minimal size for LoadSource in IBM i V7.1 is 17 GB
  - We expect bigger minimal size in future releases
- Best practice
  - Minimal size about 40 GB
  - Maximal size about 150-200 GB
  - Minimal number of LUNs per IBM i LPAR: 6
- Consider SCSI command tag queuing Queue depth (max. concurrent I/O operations to a LUN)

Queue depth for IBM i LUNs

<table>
<thead>
<tr>
<th>SCSI command tag queuing - queue depth</th>
<th>LUN in DS8000</th>
<th>LUN in SVC / V7000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Native connection with IOP-based adapters</td>
<td>1</td>
<td>n/a</td>
</tr>
<tr>
<td>Native connection with IOP-less adapters</td>
<td>6</td>
<td>16</td>
</tr>
<tr>
<td>VIOS_NPIV connection</td>
<td>6</td>
<td>16</td>
</tr>
<tr>
<td>VIOS vsccsi connection</td>
<td>32</td>
<td>32</td>
</tr>
</tbody>
</table>
Sizing FC adapters in IBM i or in VIOS

<table>
<thead>
<tr>
<th>Guideline for IO/sec per port / GB per port / MB/sec per port</th>
<th>4 Gb adapters, FC 5774/5276</th>
<th>8 Gb adapters, FC 5735/5273</th>
</tr>
</thead>
<tbody>
<tr>
<td>IO/sec at 70% utilization</td>
<td>10500</td>
<td>12250</td>
</tr>
<tr>
<td>GB per port for native connection</td>
<td>2800</td>
<td>3266</td>
</tr>
<tr>
<td>Assumed: Access Density = 1.5</td>
<td>With Dual path: 5600 per two ports</td>
<td>With Dual path: 6532 per two ports</td>
</tr>
<tr>
<td>Sequential workload MB/sec at 70% utilization</td>
<td>122 per port, if one port in adapter is used</td>
<td></td>
</tr>
<tr>
<td></td>
<td>104 per port, if both ports are used</td>
<td></td>
</tr>
<tr>
<td>Transaction workload MB/sec at 70% utilization</td>
<td>175</td>
<td></td>
</tr>
</tbody>
</table>

Sizing host adapters in external storage

Rules of thumb for HBAs in DS8800

- About 4 to 8 * ports in IBM i per one HA card in DS8800
- For good performance: The number of HA cards should be the same or bigger than the number of device adapters in DS8800
- At least one HA card per IO enclosure in DS8800

Guideline for V7000 host ports

- Maximum 50 0000 IO/sec per host port
- Maximum 300 MB/sec per host port
- Native connection with SAN switches or VIOS_NPIV connection:
  - Zone one IBM i port with two V7000 ports each from one node canister
  - For multipath use two or more IBM i ports each zoned this way. Recommended: use ports from different adapters for multipath
- Native connection without SAN switches:
  - For resiliency: assign the LUNs to two IBM i ports each connected to different V7000 node canister
  - For Multipath: Use two IBM i adapters. Connect one port from each IBM i adapter to node canister 0, and the other port to canister 1. Assign the LUNs to all 4 IBM i ports.
Data layout

- Spreading workloads across all components and resources of the storage system, maximizes the utilization of the hardware components. However, it is always possible when sharing resources that performance problems may arise due to contention on these resources.

- To protect critical workloads you should isolate them minimizing the chance that non-critical workloads can impact the performance of critical workloads. Isolation of workloads is most easily accomplished where each ASP or LPAR has its own managed storage pool. This ensures that you can place data where you intend.

- I/O activity should be balanced between the two DS8000 processor complexes, or the two nodes or controllers on the SVC / Storwize V7000.

- Use only IBM i LUNs on any storage pool (rather than mixed with non-IBM i).

- If you mix production and development workloads or many small workloads in storage pools make sure that the customer understands that this may impact production performance.

Guidelines for VIOS resources

- Recommended: Use IBM i Workload Estimator to estimate needed VIOS resources

- Rule of thumb: 1 – 2 CPU per VIOS

- Rule of thumb: 8 GB memory in VIOS

- FC adapters in VIOS:
  - For 4Gb adapters calculate about 10500 IO/sec per port
  - For 8Gb adapters calculate about 12250 IO/sec per port

- With VIOS_NPIV
  - Maximum one virtual FC adapter from one IBM i LPAR can be mapped to a port in adapter in VIOS
  - Up to 64 virtual FC adapters, each from different IBM i LPAR can be mapped to the same port in adapter in VIOS

- To calculate the number of adapters in VIOS you can also use the measurements of MB/sec listed in foil 41
Solid State Drives with IBM i

- Hybrid disk pool and Easy Tier
- An IBM i ASP with application data resides on SSD
- The IBM i skew level is typically flat due to IBM i object orientated architecture
- Therefore IBM i might need some more SSD in the hybrid pool than other workloads
- The IBM i methods for data relocation can be used with DS8000, but can not be used with V7000

Typical IBM i Skew level and STAT

Data resides on SSD

Collecting data

Decide the peaks

Decide initial Storage configuration for Disk Magic

Model with Disk Magic
Modelling with Disk Magic – inserting IBM i reports -1

Start Disk Magic

Select New SAN Project

Modelling with Disk Magic – inserting IBM i reports -2

Select Open iSeries Automated ...

- Select the folder with IBM i reports
- Click Open
Modelling with Disk Magic – inserting IBM i reports -3

- Select “iSeries PT Reports”
- Select the reports to use for modelling

Use “Select all” and “Process”

Modelling with Disk Magic – selecting the peaks

- Select a peak by clicking the header of the column, click Add Model for each peak
- We recommend to select the peaks by IO/sec, MB/sec, Write MB/sec and Average
- After the models are added, click Finish
Modelling with Disk Magic – start modelling ext. Storage

- Disk Magic creates a model for each peak
- Click on "ES81" to start modelling external storage for a particular peak

Modelling with Disk Magic – Base (present conf.) -1

- Insert the current configuration of external storage or internal disks
Modelling with Disk Magic – Base - 2

- On interface tab insert the number and type of adapters used
- On iSeries Disk tab insert the number and type of presently used disk drives, create an extent pool

Modelling with Disk Magic – Base - 3

- On iSeries Workload tab select the created extent pool, specify the type of adapters, Base
- Note: Service Time and Wait time for the base are automatically inserted from IBM i reports

Automatically inserted from reports
Modelling with Disk Magic – new configuration - 1

- On General tab insert the storage system you want to model
- On Interface tab specify the number and type of adapters in IBM i and in storage system

Modelling with Disk Magic – new configuration - 2

- On iSeries Disk tab choose the disk type, RAID level and capacity of the new configuration and add new disk pool
- On iSeries Workload tab select the pool for new configuration, if needed change the number or size of LUNs, Solve
Modelling with Disk Magic – new configuration - 3

- At solving the model, Disk Magic calculates predicted response times.
- Compare present and predicted Service time and Wait time.

Modelling with Disk Magic – utilizations, history

- Click Utilizations to show the utilizations of modelled configuration. Utilizations that exceed the threshold are marked in red.
- Clicking History shows the modelled configurations. You may rename them with more descriptive names. Restore any configuration you want to look to, or to change and solve again.
Modelling with Disk Magic – modeling growth - 1

- Click Graph
- Choose the data to model growth for, graph type and the range of growth, Plot

Modelling with Disk Magic – modeling growth - 2

- Predicted service time and utilizations with IO growth

Utilization Overview (Series) | Total I/O Rate (I/Os per second)
---|---

<table>
<thead>
<tr>
<th>Utilizations</th>
<th>Amber Threshold</th>
<th>Red Threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processor (for I/O)</td>
<td>66%</td>
<td>90%</td>
</tr>
<tr>
<td>Processor (for Compression)</td>
<td>66%</td>
<td>90%</td>
</tr>
<tr>
<td>Average Disk</td>
<td>66%</td>
<td>90%</td>
</tr>
<tr>
<td>Highest HDD</td>
<td>66%</td>
<td>90%</td>
</tr>
<tr>
<td>Average HA</td>
<td>66%</td>
<td>90%</td>
</tr>
<tr>
<td>Back End Interface</td>
<td>66%</td>
<td>90%</td>
</tr>
<tr>
<td>Average Host Interface</td>
<td>66%</td>
<td>90%</td>
</tr>
</tbody>
</table>
DEMO
Disk Magic modelling of V7000 for IBM i

Backup charts
Response times in IBM i

- **Application response time:**
  - The response time of an application transaction. This time is usually critical for the customer.

- **Duration of batch job:**
  - Duration of batch job: Batch jobs usually run during the night; the duration of a batch job is critical for the customer, because it must be finished before regular daily transactions start.

- **Disk response time: service time + wait time**
  - **Service time:** processing I/O operation
  - **Wait time:** potential I/O queuing on the IBM i host

Disk drives – additional considerations and best practise

- **Number of disk drives in native, VIOS_NPIV and VIOS VSCSI connection**
  - The sizing guidelines and calculations for disk drives are the same regardless if the host connection is native, VIOS_NPIV or VIOS VSCSI.

- **Sizing for big block sizes (transfer sizes)**
  - By experiences, big block sizes have significant impact on performance. Therefore we recommend to size with additional 25% - 30% of disk drives.

- **Implementation recommendations:**
  - create vdisks in Striped mode (default)
  - extent size 256 MB (default)
Sizing for Metro Mirror or Global Mirror links

- Obtain the write rate in MB/sec
  - Use queries to IBM i collection services data
  - Or, use performance reports and calculate writes/sec * blksize

- Based on highest write rate calculate needed bandwidth as follows:
  - Assume 10 bits per byte for network overhead
  - Assume a maximum 80% utilization of the network
  - Apply 10% uplift factor to the result to account for peaks in the 5 minutes intervals

- If the compression of devices for remote links is known you may apply it. If it is not known you may assume a 2:1 compression

Sizing for FlashCopy

- GUI preset Snapshot – Flashcopy without background copying
- GUI preset Clone - Flashcopy with background copying
- Reasons for potential performance impact on production workload during FlashCopy relation:
  - At the beginning, every write operation to source volume or to target volume trigs copy from source to target; later copying usually decreases due to writes to the areas that have been already overwritten
  - A certain amount of read operations to target volume is done from source
- To minimize performance impact on production system it is important to provide sufficient disk drives to the FlashCopy target
- Rough guideline for the number of disk drives for FlashCopy target:
  - Calculate 150 writes/sec per 10 K RPM disk drive
  - Calculate 100 writes/sec per 15 K RPM disk drive
- With Thin provisioned FlashCopy target volumes, ensure sufficient capacity in the disk pool
  - The more write operations are done to either source or target LUNs, the more capacity is needed
DEMO
Disk Magic modelling of V7000 for IBM i