Validator

System Requirements
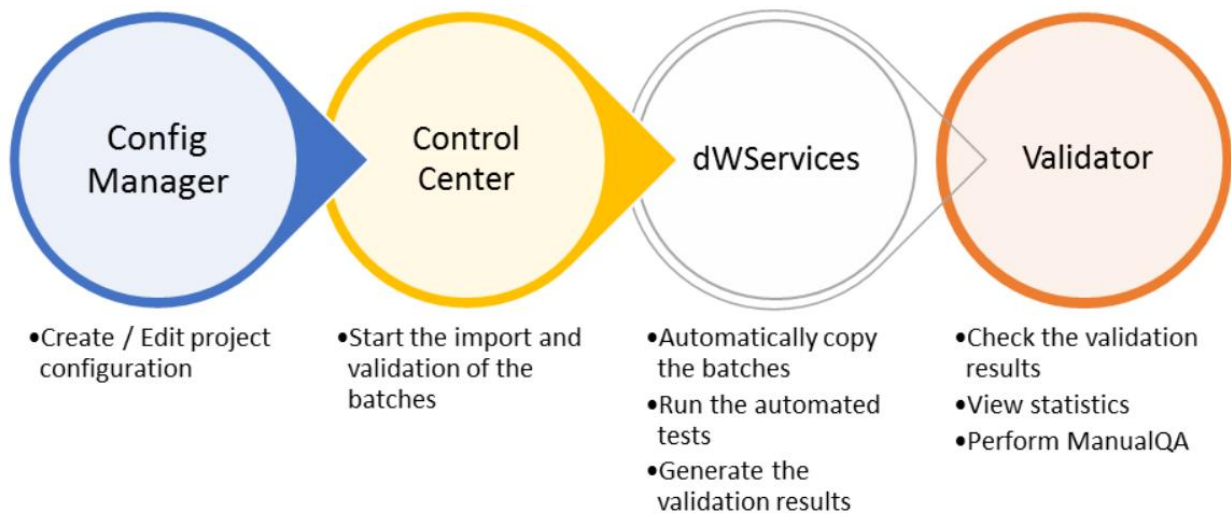
CCS

Validator

Table of Contents

1 Validator

Validator is a tools package used to automatically validate digital objects, such as image files, PDF files or XML data (METS, ALTO).

The validation is performed in two steps: Automated and Manual validation.

The Automated Validation is checking all files for specific parameters and integrity and ensures all data is of consistent quality.

The Manual Validation is performed by operators using the Validator user interface. They check samples for image quality, meta data, text quality or other parameters that cannot be validated by a machine (subjective criteria).

Config Manager
- Create / Edit project configuration

Control Center
- Start the import and validation of the batches

dWServices
- Automatically copy the batches
- Run the automated tests
- Generate the validation results

Validator
- Check the validation results
- View statistics
- Perform ManualQA
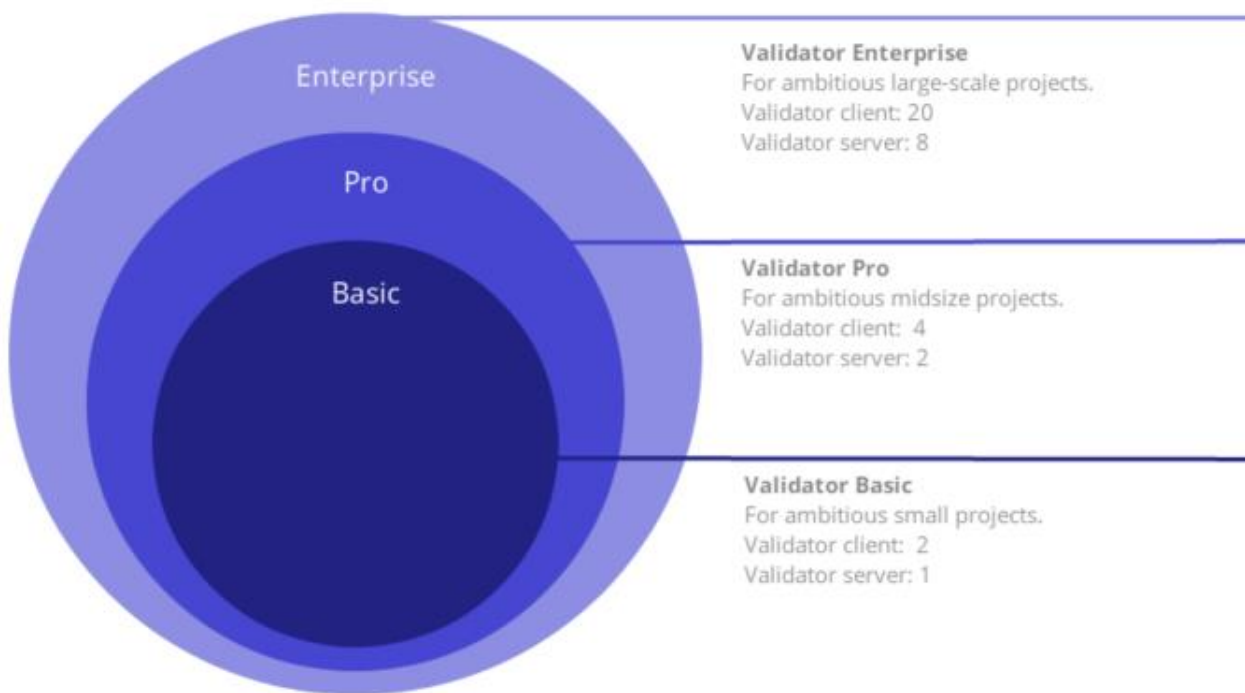
1.1 Introduction Validator editions

Validator comes in three different editions. Each edition is directed towards a specific field of application.

- Validator Basic: 1 processing server, 2 Validator clients, 1 configuration Manager, 1 Control center
- Validator Pro: 2 processing servers, 4 Validator clients, 1 configuration Manager, 1 Control center,
- Validator Enterprise: 8 processing servers, 20 Validator clients, 1 configuration Manager, 1 Control center

1.2 Overview Validator editions

Compare the Validator editions   ☑ Basic   ☑ Pro   ☑ Enterprise

Enterprise

Pro

Basic

Validator Enterprise
For ambitious large-scale projects.
Validator client: 20
Validator server: 8

Validator Pro
For ambitious midsize projects.
Validator client: 4
Validator server: 2

Validator Basic
For ambitious small projects.
Validator client: 2
Validator server: 1

2 Validator requirements

This chapter provides the technical requirements for Validator clients, Config manager and Control Center system requirements for a Validator installation including hardware, operating system and additional settings.

2.1.1 Operator workstation (Client, ControlCenter)

- The Validator user interface applications are installed on operator workstations (Client & ControlCenter for all editions). The monitoring interface named ControlCenter is installed on these machines to control and maintain the production of the Validator environment
- The intermediate verification tasks are performed by operators with the Validator Client. The ControlCenter monitors the background services and environment
- In general these machines only operate during office hours. No automated tasks or data transfers through the network are required here after hours.
- The parallel execution of Client's are limited by license keys (floating seat licenses)
- No production data is stored on these operator machines. Only the Validator binaries (program data) are stored locally on the machine and the local hard drives are just used for temporary data swapping, low resolution quality copies of the images, etc.

Validator Client / ControlCenter requirements:

CPU	Minimum: Intel® Core™ i3 Recommended: Intel® Core™ i5
RAM	Minimum: 4 GB RAM Recommended: 8 GB RAM
NIC	Minimum: 1Gbit/s
Screen resolution	Minimum: 1280 x 720 Recommended: 1920 x 1080
Local disk space	4 GB of local space required More than 1 GB free space for memory swapping
Operating system	Versions: Windows 10 Processor architecture: 64 bit recommended / 32 bit supported Version type: Professional or higher Language: English*
Domain	Workstation must be member of the Microsoft domain
Permission	Applications need to be executed on domain user account with permission for updating binaries on local machine in %ProgramFiles(x86)% folder - in case this is not allowed alternative solutions exists, but autoupdate feature is disabled and update procedure will be customized on additional costs.
Internet access	Internet access for validation of XML data is required for validation purposes. If no internet access is available, the configuration needs to be changed to local validation.
Firewall	MS Firewall needs to be disabled or opened for full ODBC and UNC connection. See Firewall (on page 11)

*)
other languages possible, but not tested

2.1.2 Validator processing server (Services)

- The processing Services are installed on processing servers; these are running 24/7 as Windows services
- The services are controlled by a separate application called "ControlCenter". This application is part of the Validator installation and can be started from any Validator server or operator workstation. The operators do not need direct access to the servers to control the services
- The amount of processing servers and total amount of Services are limited by license keys
- No production data is stored on these processing servers. Local hard drives only hold the installed Validator binaries and are used for temporary data swapping
- For support / administrative work, internal RDC (Remote Desktop Connection) access should be enabled for all servers

For high stability demands:

- The services request all processing data on demand from the file servers. To ensure effective operation, the file servers and processing servers should be closely connected to each other (best same [stacked] switch). No production data is stored locally
- Due to operation time of 24/7 and prevention of down-times, a redundant hardware design is recommended (power supply, hard drive [RAID-1], NIC [recommended: load balanced])
- In large production environments, BladeCenter technology is recommended. By default, this already provides remote management and monitoring interfaces and can send out alerts if a problem is detected

Validator Services requirements:

CPU	Minimum: Intel® Xeon® E3 (or similar) Recommended: Intel® Xeon® E5 (or similar)
RAM	Minimum: 8 GB RAM More RAM may be required for large-sized image processing (>DIN A2)
NIC	Minimum: 1 Gbit/s For high-demand environments 2 to 4 Gbit/s trunks are beneficial depending on entire environment and file server speed
Screen resolution	No local screens are required on these machines (except for administrative work by IT). Minimum: 1280 x 1024
Local disk space	4 GB of local space required More than 4 GB free space for memory swapping
Operating system	Versions: Windows Server 2012 / 2016 Processor architecture: 64 bit (required to address full memory) Version type: Standard or higher* Language: English **
Domain	Server must be member of the Microsoft domain.
Permission	Services need to be registered on domain user. This domain user needs local administrator rights to be able to operate as a local service. The machine / domain users need permanent full access to all Validator locations.
Firewall	MS Firewall needs to be disabled or opened for full ODBC and UNC connection. See Firewall (on page 11)

*) The Essentials edition of Windows 2012 R2 server is designed as a small business server and might cause problems, if used in a network with several other servers, running Validator services. Hence, we do not recommend the Essentials edition, but rather go for the Standard Edition of Windows 2012 R2.

**) Other languages possible, but not tested

3 Environmental requirements

Validator was developed for Microsoft environments as server- / client- application.

We outline in the following chapter the best-practice recommendations based on the experience of CCS with all of our clients.

3.1 Standard requirements

Finally the requirements for the individual roles and specific hints to be considered within a Validator environment are outlined in the following subchapters.

3.1.1 Domain server

The domain server is the administrative server running all Microsoft domain services. Herewith it is considered of all basic domain functionality including DHCP, DNS, NTP, etc. (DNS or DHCP services can also be operated by network routers of course.)

Following Validator production considerations:

- For small environments, the usage of existing domains is recommended (prevention of administrative overhead). For large environments the amount of accounts and production separation often make sense – best practice recommendation: define a sub-domain of existing organization domain.
- For operator monitoring (Validator log) and access permission control, each operator should get his personalized domain account.
- Also, local redundancy is recommended to prevent loss of the domain controller (power supply, local hard drive (RAID-1)), NIC (recommended: load balanced).
- Due to operation time of 24/7 and prevention of down-times a second domain controller is recommended or running on virtual machine with redundant hosts.

The following overview of requirements are just for an initial general idea. The concrete requirements for the version used should be checked on Microsoft website.

Domain server requirements:

CPU	Minimum: Intel® Xeon® E3 (or similar) Recommended: Intel® Xeon® E5 (or similar)	
RAM	Minimum: 4 GB RAM required (by Microsoft)	
NIC	Minimum: 1 Gbit/s	
Screen	No local screens are required on these machines (except for administrative work).	
OS	Versions: Processor architecture: Version type: Language:	Windows Server 2012 / 2016 32 bit (supported) 64 bit (recommended) Standard or higher only English OS supported

3.1.2 Database server

The SQL server is the heart of each Validator environment. The SQL server hosts administrative data but as well all validation results and tests configurations. The SQL server does not contain any image data or the whole description of each documents.

Without the databases no production is possible (single point of failure). Without the SQL server you cannot access any validated batch or get any information about the batch processing history. While for any batches access and for all history reporting the SQL server is needed, a high performant SQL Server response for the Clients and servers is mandatory for a smooth operation.

Following Validator production considerations:

- Due to operation time of 24/7 and prevention of down-times, the SQL server should be clustered or running on virtual machine with redundant hosts.
- Also hardware redundancy is recommended to prevent any loss of the SQL server (power supply, hard drive [RAID-1], NIC [recommended: load balanced]).
- Due to similar requirements for redundancy / fallback of the SQL server to the Domain Controller, it makes sense to set it up together on the domain controller.
- SQL Server Express is **not sufficient** due to amount of data stored and queries complexity, SQL Server Standard is required

The following overview of requirements is just for an initial, general idea. The concrete requirements for the version used should be checked on Microsoft website.

Database server requirements:

CPU	Minimum: Intel® Xeon® E3 (or similar) Recommended: Intel® Xeon® E5 (or similar)
RAM **HDD**	12 GB RAM recommended for combined server (8 GB for processing + 4 GB for domain services and database services) Stored data could be quite high, 500GB local HDD for SQL database is recommended
NIC	Minimum: 1 Gbit/s
Screen	No local screens are required on these machines (except for administrative work).
OS	Versions: Windows Server 2012 / 2016 Processor architecture: 32 bit (supported) and 64 bit (recommended) Version type: Standard or higher Language: English
Domain	SQL Server must be member of the Microsoft domain
SQL server	Version: MS SQL Server 2012 / 2014 /2016 Version type: Standard Edition or higher required (not Express, due to amount of data stored and queries complexity).
Firewall	MS Firewall needs to be disabled or opened for full ODBC and UNC connection access
Validator environment	All machines / domain user need permanent access to this SQL server

3.1.3 File server

Shares

INV: Input Validator - Batches to be validated. For mixed environments (dW and Validator) this should be different from IN.

LOWRESPOOL: Location for temporary low res quality images for Validator speedup. For mixed environments (dW and Validator) this should be different from POOL.

REPORTS: Default path for Validation reports. For mixed environments (dW and Validator) this should be different from OUT.

As File server(s) we consider any server hosting the volumes for the Validator shares.

This can be a Windows server with shared folders on UNC paths up to a professional NAS (network attached storage) for high-performance operations.

Following Validator production considerations:

All shares need to be accessible from any machine (Servers and operator machines) at any time.

> Please note: SAN storage, with assignment of storage to individual processing servers is not applicable (see requirements above).

If SAN is used, the server having assigned the storage capacities of the SAN will operate as file server and share these locations to all processing servers / Validator client machines.

The required storage capacities depend on the following facts:

- Batches size (number of documents/pages inside, type of images, etc.)
- Daily incoming data (batches to be validated)
- Turn-around-time for Validator processing / holding time validated batches

In case help for storage calculation is required, the above mentioned criteria needs to be provided / estimated as good as possible.

The speed of file servers depends mainly on the type and number of discs. For Validator operations, we can differentiate between two categories:

- REPORTS share: less traffic and more sequential writing (-> less fragmentation)
- LOWRESPOOL and INV: more parallel traffic and high fragmentation

Based on these categories, different kinds of storage can be used and costs can be reduced for large scale environments. High-end discs are not required for all storage.

Due to operation time of 24/7 and prevention of down-times, the file server(s) should fulfill highest uptime requirements.

Also, local redundancy is recommended to prevent loss of the file server(s) (power supply, hard drive [RAID-1], NIC [recommended: load balanced]).

File server requirements:

	Basic	Pro	Enterprise
Storage type	Microsoft File server Also possible Linux based storage		Recommended: NAS (Network Attached Storage) with integrated fallback-solutions
Disc types	Due to less parallel I/O large S-ATA discs will be sufficient, number of spindles not that impact as for large scale environments	Config Share, low res pool shares and input shares have high random I/O demands. Here high-speed storages are required (fiber channel / 15k-RPM-discs / smaller but more disk spindles) REPORTS share(s) have more sequential write/read operations and are less frequented.	
NIC	Minimum: 1 Gbit/s Recommended: network trunk of 2 Gbit/s	Minimum: network trunk of 2 Gbit/s Recommended: trunks with 4 Gbit/s	Minimum: 4 Gbit/s Recommended: 10 Gbit/s
Validator environment	CIFS share required (no delayed write, case-insensitivity) NTP (network time protocol) required		

3.1.4 Firewall

As this is an essential requirement for all Validator processing servers, Validator client machines, file servers and SQL servers, we have summarized the requirements for the firewall settings in this separate chapter.

CCS recommends having firewalls enabled for security reasons, but certain rules need to be configured according to the following table:

	Port	Source	Destination
File share access	TCP/445, TCP/139	processing server operator machine	<file server(s)>
MS SQL server	TCP/1433-1434 TCP/2382 (SQL Browser)	processing server operator machine	<SQL server>

Last updated: 06/29/2021

CCS Content Conversion Specialists GmbH

Weidestraße 134

22083 Hamburg, Germany

Phone: +49-(0)40-228582990

Fax: +49-(0)40-228582998

E-Mail: info@content-conversion.com

Website: www.content-conversion.com