



Demystifying Newforma Indexing

David Dark

Development Liaison

September 15, 2015

About David Dark

David Dark – *Development Liaison*

ddark@newforma.com

[linkedin.com/in/davidrdark](https://www.linkedin.com/in/davidrdark)



Agenda

1. Indexing History
2. Definitions
3. Architecture Overview
4. Index Lifecycle
5. Lifecycle breakdown
6. Troubleshooting
7. Guided Tour
8. Q&A
9. Additional Resources
10. Contact Info

Indexing History

Pre-7.0

- Microsoft Windows Indexing
- Black Box

7.0

- Newforma built indexing
- Allows closer interaction with the Newforma Indexing API

8.0 & 9.0

- Query optimization
- Improved end-user performance

Indexing History (cont.)

10.0

- Introduced additional filter services
- Separately installable for remote offices

11.0

- Changed from Filter Service to Work Service
- Added Indexing for EDMS
- Moved index admin UI to the Servers tab
- Enhanced filtering performance

11.5.7

- Enhanced scanning performance

Definitions

Index

noun

Location where the searchable content is stored

verb

To gather searchable content and save it

synonyms: filter

Definitions (cont.)

- Scope – The container definition of where the data lives. For example \\myserver\proj\0001\
- Scan – Traversing a scope or sub-container for new or updated items
- Filter – The act of gathering searchable content from an item
- IFilter – Third party library to assist in extraction of searchable content

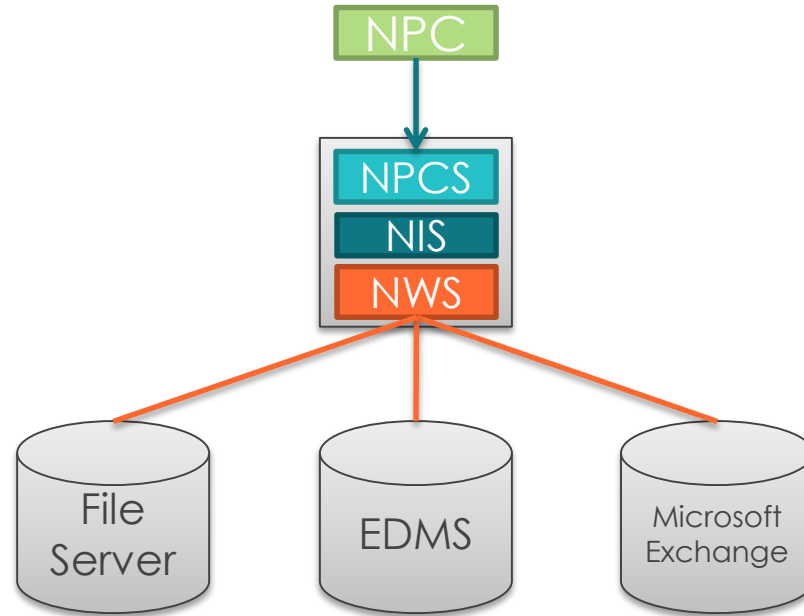
Definitions (cont.)

- Namespace – The root share of a scope.
For example \\myserver\proj\
- Scope Scan Modes
 - Full Scan – Scan and filter everything (default mode)
 - Additive Scan – Add only new content, don't remove anything (not widely used)
 - No Scan – Stop scanning the scope, but retain existing

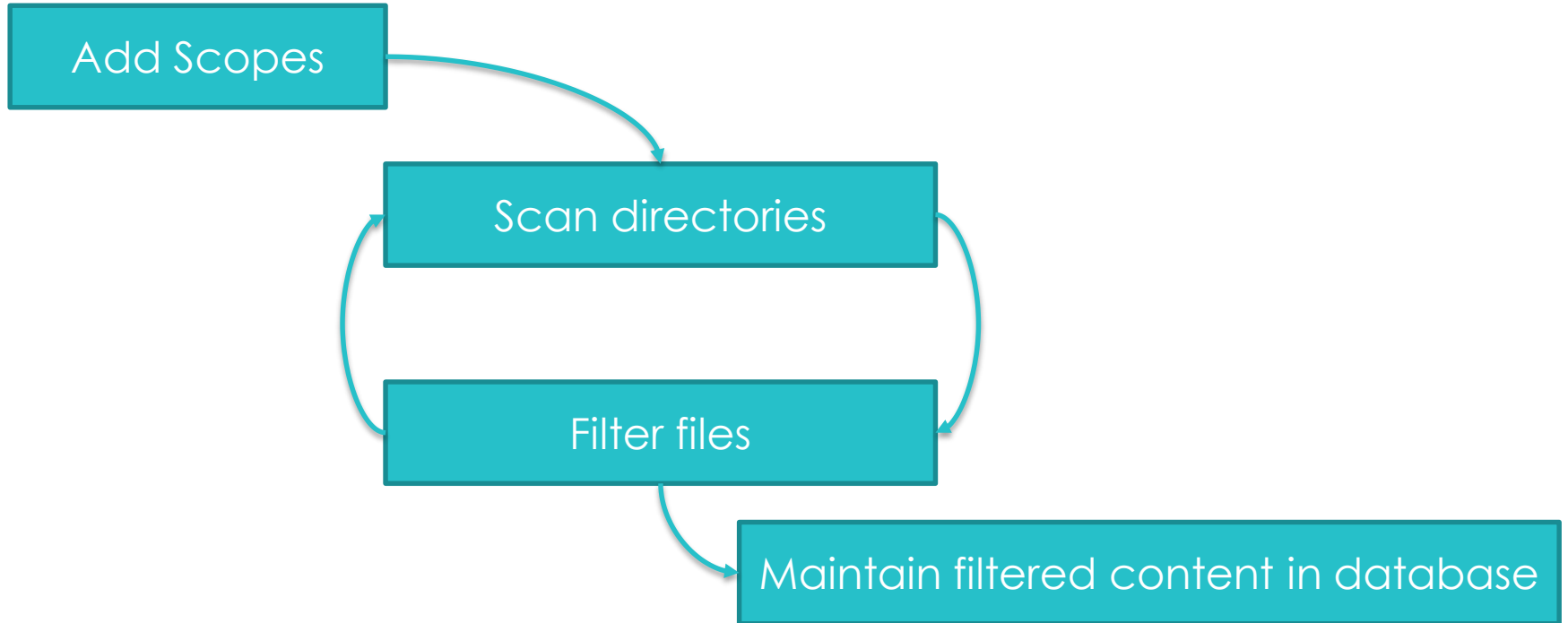
Definitions (cont.)

- Indexing Service – Controls the list of scopes and maintains the stored searchable content within the `nis` database
- Work Service – Scans scopes for new and updated content and filters for searchable content
- Search – All of these parts working together to provide results in Project Center

Architecture Overview



Index Lifecycle



Scopes are Added When...

- Creating a project with a project folder
- Adding a project folder to an existing project
- Repathing project folders to a new location
- Repinning projects to another server

How Scanning is Prioritized

1. Recently added or changed items within Project Center get added to rescan queue in near real-time
2. If a scope has changed within the last 25 hours, we will scan it every two hours for changes within it
3. If scope has not changed within the last 25 hours, we will scan it
4. Actively used projects are OCRed and Rendered daily as a part of project maintenance

What is Filtered

Files filtered

- DWG file indexing – [active layer](#)
- OCR – PDF files only
- Supported file types – [supported file types](#)
- Additional file types – [using third party IFilters](#)

Archived Projects – Typically only once per day

Offline Projects – Scopes are set to 'No Scan' mode

Files are Filtered when...

- Creating, adding, renaming, moving and deleting files within Project Center
- Filing email using
 - File in Project from Outlook
 - Newforma Items to File project folder in Outlook
- Dragging & dropping files into Project Center
- Sending transfers with record copies

When to Intervene

- Indexing errors appear for the server
- End user reports they can't find a file
- Takes days for a file to show up in search
- Indexing failure rate over 5%

How to Intervene

Check System Performance

- Disk Queue Length – Anything over 2 is a red flag
- Enough RAM – There should be at least 2GB free
- CPUs at 100% – Should fluctuate between 40-80%
- Scope over a WAN – Files in scope >10ms away

Reprocess the item manually

- Rescan a directory – Find new/updated items
- Re-index a directory – Filter the new/updated items

Troubleshooting Tips

Break down the process into each step of the indexing lifecycle

- Added Scope
 - Confirm scope is listed in the index
- Scan Directories
 - Verify scanning picked up the file using 'enumerate files'
- Filter File
 - Check the filter result, filter time, and filter reason
 - Use filter-tester to validate filtering process and content
- Maintain index
 - Confirm 'Search index built' statistic is at 100%

Troubleshooting Tips (cont.)

- View files that can't be indexed
 - Password-protected
 - Corrupt
 - Third-party IFilter cannot open them
 - Files marked 'Do not index' in Windows
 - Newforma Service Account doesn't have adequate permission
- Server reporting indexing errors
 - Unfilterable namespace
 - Outlook profile does not exist
 - Find broken project folders using Disk Utilization Report

What can be done to boost performance?

Add an additional Work Service

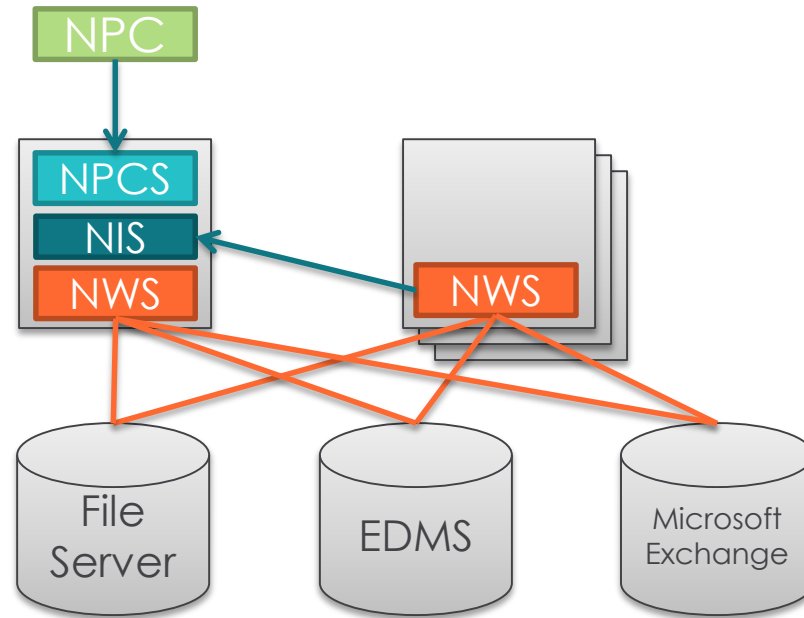
Archive or move projects Offline

Add more hardware resources to the Index server

Add a Project Center server to balance load

Email services@newforma.com for a server health check and more suggestions on optimization

Architecture with Additional Work Services



GUIDED TOUR

Frequently Asked Questions

- How much space does the index take?
 - 1-2% of file size
- What is nightly maintenance?
 - Removes scopes and indexed information from database for project folders and projects of a removed project
- How do I ensure files have been removed from Index?
 - Search for it
 - Use 'Enumerate files' from the Search Query tab

LIVE Q&A

Additional Resources

Newforma Customer Community

- <https://customercommunity.newforma.com>

Knowledge Base Articles

- [Deploying Newforma Work Service](#)
- [High Indexing Failure Rates](#)
- [Clean up Index from unintentionally added files](#)
- [Disable Indexing on a specific folder](#)
- [MSG files fail due to Exchange Connectivity Failure](#)

Contact Info

David Dark – *Development Liaison*

ddark@newforma.com

[linkedin.com/in/davidrdark](https://www.linkedin.com/in/davidrdark)

Services Team – *optimization*

services@newforma.com

Support Team – *break fix*

support@newforma.com

