DiskBoss

DATA MANAGEMENT

Duplicate Files Search

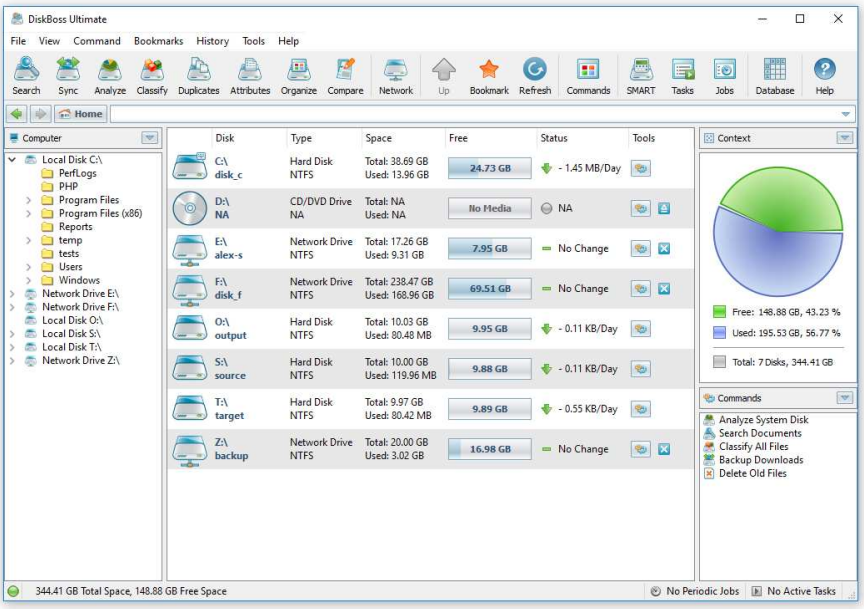
Version 16.3

Jan 2026

www.diskboss.com
info@flexense.com
Flexense Ltd.

1 DiskBoss Overview

DiskBoss is an automated, policy-based data management solution allowing one to analyze disks, directories and network shares, classify and categorize files, search and cleanup duplicate files, perform automated file management operations according to user-defined rules and policies, synchronize disks, directories and network shares, compare directories and files, perform bulk file delete and secure data wiping operations, detect unauthorized changes in files and directories, etc.



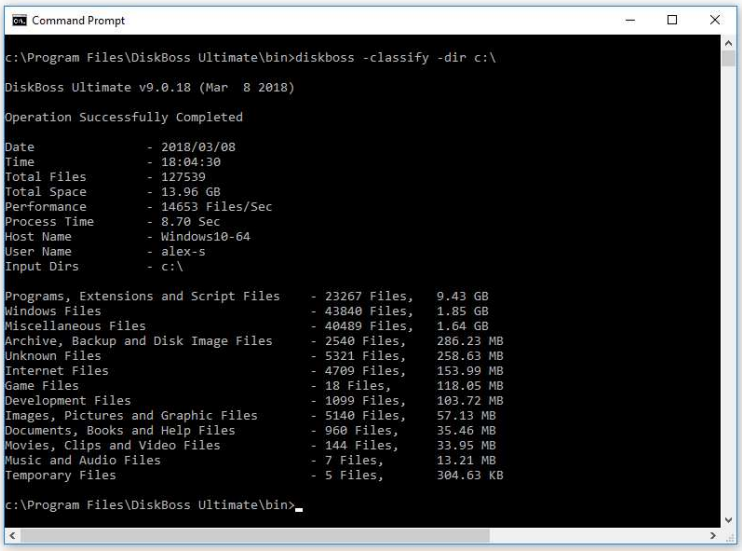
All disk space analysis and file management operations are integrated into a centralized and easy-to-use GUI application allowing one pre-configure analysis and file management operations as user-defined commands and execute any required command in a single mouse click using the DiskBoss GUI application or direct desktop shortcuts.

- Disk Space Analysis
- File Classification and Organizing
- Duplicate Files Search and Cleanup
- Bulk File Delete and Secure Data Wiping
- Automated, Policy-Based File Management
- Real-Time Disk Change Monitoring
- High-Speed File Synchronization
- Secure File Transfer Operations
- File Integrity Monitoring

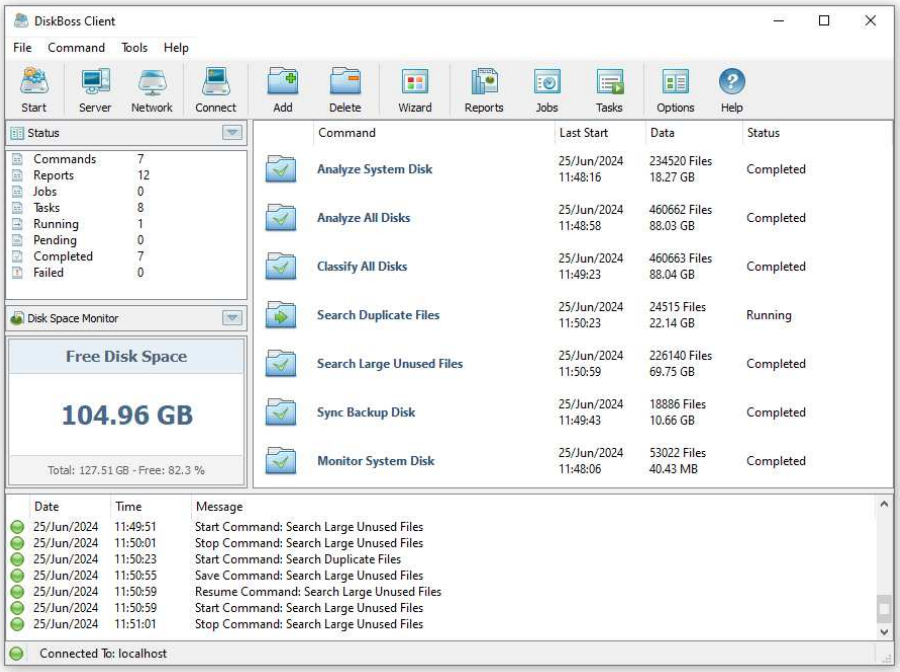
DiskBoss allows one to generate various types of pie charts and save HTML, PDF, Excel, JSON, text, CSV and XML reports for all types of disk space analysis, file classification and file search operations. The user is provided with the ability to categorize and filter analysis and file classification results and perform file management operations on categories of files.

IT administrators are provided with extensive SQL database integration capabilities allowing one to submit disk space analysis, file classification, duplicate files search and disk change monitoring reports into an SQL database. Reports from multiple servers and NAS storage devices may be submitted to a centralized SQL database allowing one to display charts showing the used disk space, file categories and duplicate files per user or per server and providing an in-depth visibility into how the disk space is used, what types of files are stored and how much space is wasted on duplicate files across the entire enterprise.

In addition to the DiskBoss GUI application, IT and storage administrators are provided with the DiskBoss command line utility, which can be used to execute all types of analysis and file management operations from batch files and shell scripts. The command line utility provides an extensive set of command line options allowing one to execute various types of disk space analysis, file synchronization, data migration and bulk file delete operations pre-configured for user-custom needs and hardware configurations.



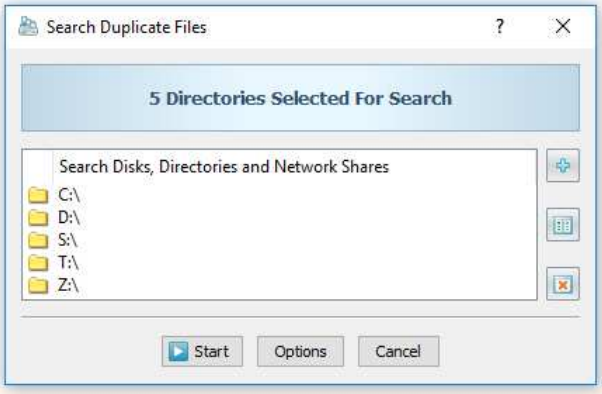
Finally, IT professionals and enterprise customers are provided with DiskBoss Server – a server-based product version, which runs in the background as a service and is capable of performing all type of disk space analysis and file management operations in a fully automatic and unattended mode according to a user-specified schedule.



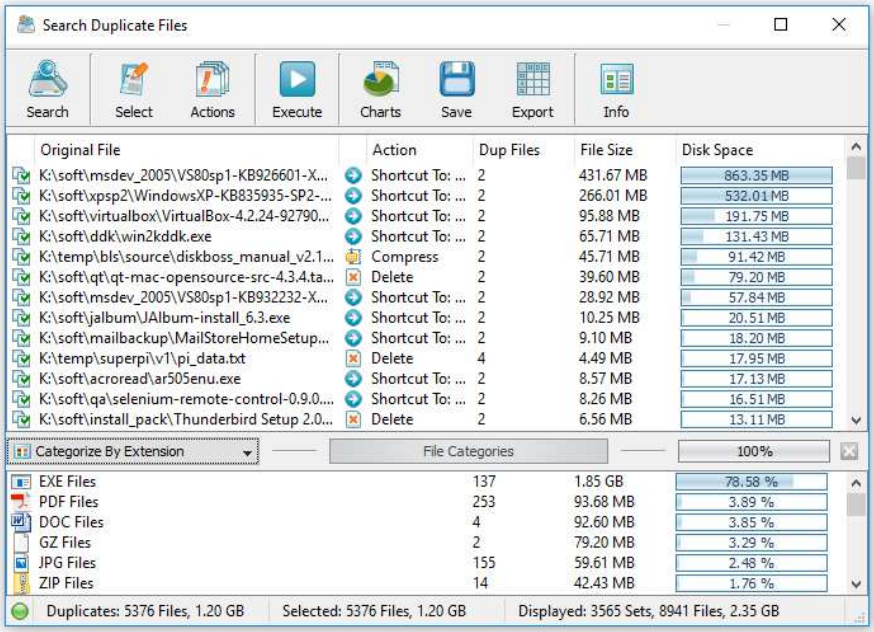
DiskBoss Server can be controlled locally or through the network using the DiskBoss client GUI application or the command line utility. DiskBoss Server provides the ability to pre-configure various types of disk space analysis and/or policy-based file management operations, schedule periodic jobs, save analysis reports into a number of different formats, export analysis results to an SQL database, periodically synchronize disks, directories and network shares and monitor critical disks and directories for unauthorized changes.

2 DiskBoss Duplicate Files Search

DiskBoss includes a built-in duplicate files finder allowing one to search duplicate files, generate various types of charts showing duplicate disk space, remove duplicate files and save reports into a number of different formats. In order to search duplicate files in one or more disks or directories, select the required directories in the DiskBoss file navigator and press the 'Duplicates' button located on the main toolbar. DiskBoss will scan the selected disks and directories and display a dialog showing the list of detected duplicate files.



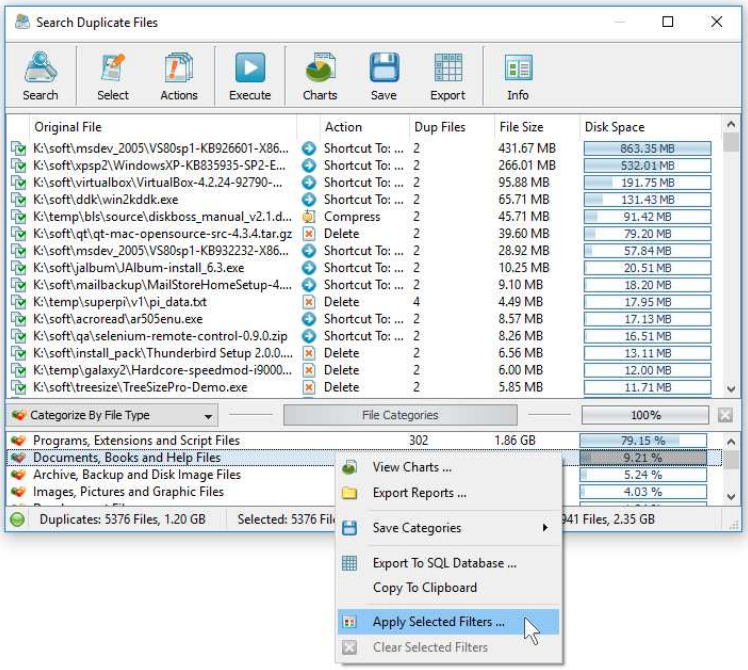
For each duplicate file set, DiskBoss shows the name of the original file, the number of duplicate files in the set, the size of each file in the set, the amount of wasted disk space and the currently selected duplicates removal action. In order to see all duplicate files related to a set, click on the set item in the set list.



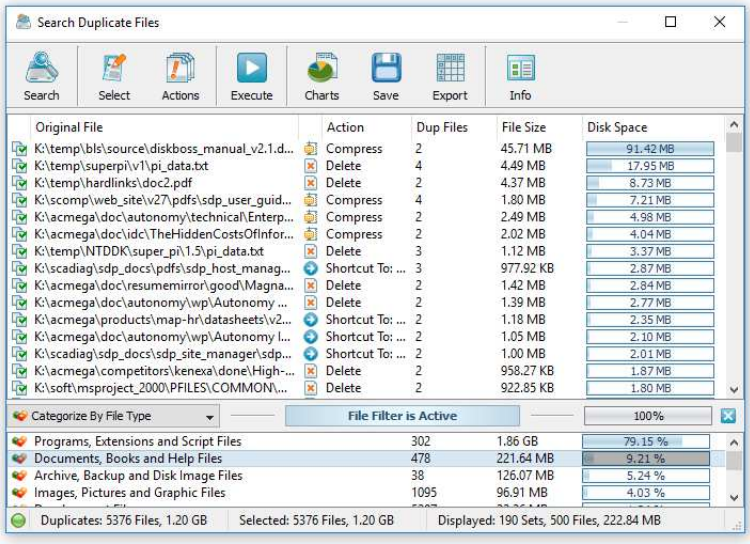
The duplicate set dialog shows all duplicate files related to the set and allows one to select the original file, the duplicate files and the duplicates removal action. In order to select a file as the original, select the file item, press the right mouse button and select the 'Set as Original File' menu item. In order to see more information about a file, just click on the file item in the file list. Once finished selecting the duplicate files, use the removal actions combo box located in the bottom-left corner of the dialog to select an appropriate duplicates removal action.

3 File Categories and File Filters

The DiskBoss duplicate files finder allows one to categorize and filter duplicate files by the file type, extension, category, size, user name, etc. In order to change the current duplicate files categorization mode, click on the file categories combo box located in the top-left corner of the categories view.

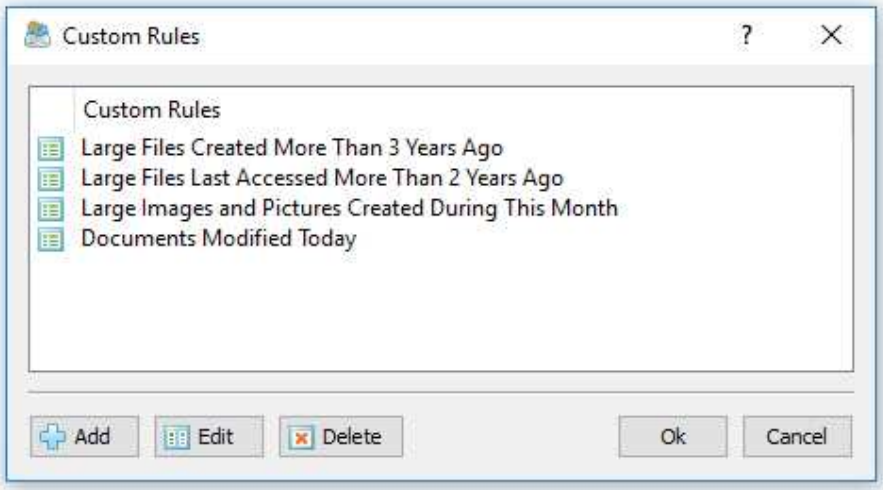


The user is provided with the ability to apply multiple file filters, display specific types of duplicate files and apply duplicate files removal actions to or export reports showing filtered duplicate files only. In order to set one or more file filters, select an appropriate type of file categories in the categories combo box, select one or more file filters in the filters view, press the right mouse button and select the 'Apply Selected Filters' menu item.

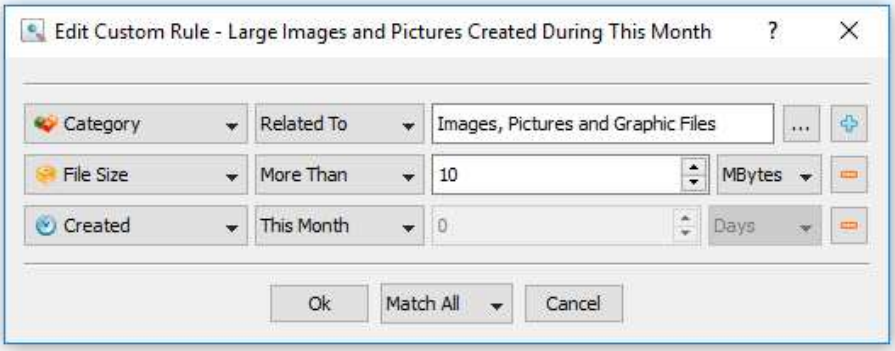


With active file filters, DiskBoss shows duplicate files matching the selected filters, exports reports showing matching files only and significantly simplifies selection of duplicates removal actions for specific file types or file categories. In order to clear the selected file filters, just press the 'Clear' button located on the right side of the categories selector.

In addition to the built-in file filters, DiskBoss provides the ability to pre-configure one or more custom file filters according to user-specific needs and requirements and then categorize and filter the duplicate files search results using to the user-custom file filters. In order to add a custom file filter, select the custom file filters view, press the right mouse button over the filters view and select the 'Edit Custom Rules' menu item.



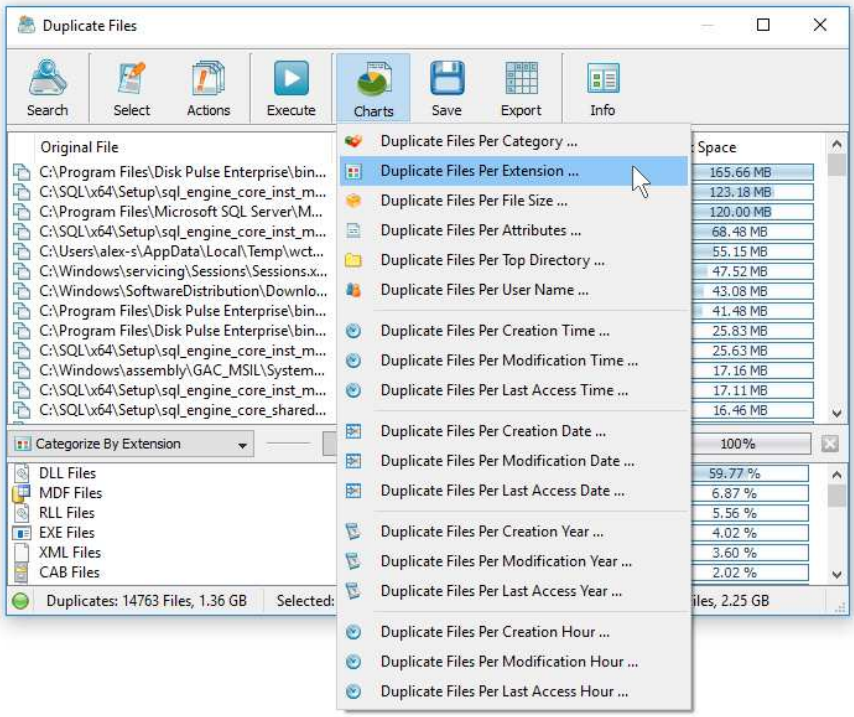
In order to add a new custom file filter, press the 'Add' button, enter a unique filter name and configure one or more file matching rules specifying which types or groups of files to display. In order to edit a custom file filter, select the filter in the custom file filters dialog and press the 'Edit' button.



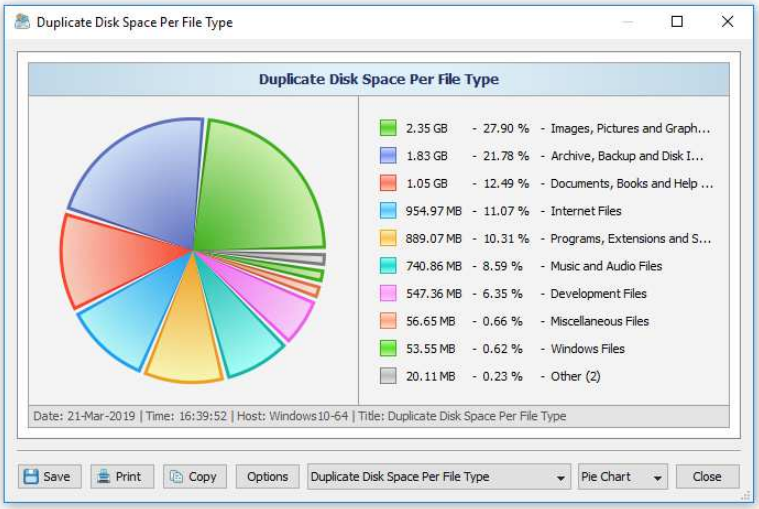
Once finished configuring custom file filters, close the filters dialog and DiskBoss will categorize the currently displayed duplicate files search results and show the list of user-specified custom file filters sorted by the amount of data and the number of files matching the specified file filters. In order to filter the currently displayed duplicate files search results by one or more custom file filters, select the required custom file filters in the filters view, press the right mouse button and select the 'Apply Selected Filters' menu item. DiskBoss will filter the duplicate files search results according to the selected custom file filters and display the amount of data and the number of files matching the selected file filters.

4 Duplicate Files Statistics and Pie Charts

The DiskBoss duplicate files finder provides multiple types of statistical pie charts and timeline charts capable of showing the amount of duplicate disk space and the number of duplicate files per directory, file extension, file type, file size, file owner, creation, modification and last access time. In order to open the charts dialog, press the 'Charts' button located on the duplicate files search results dialog and select an appropriate chart type.

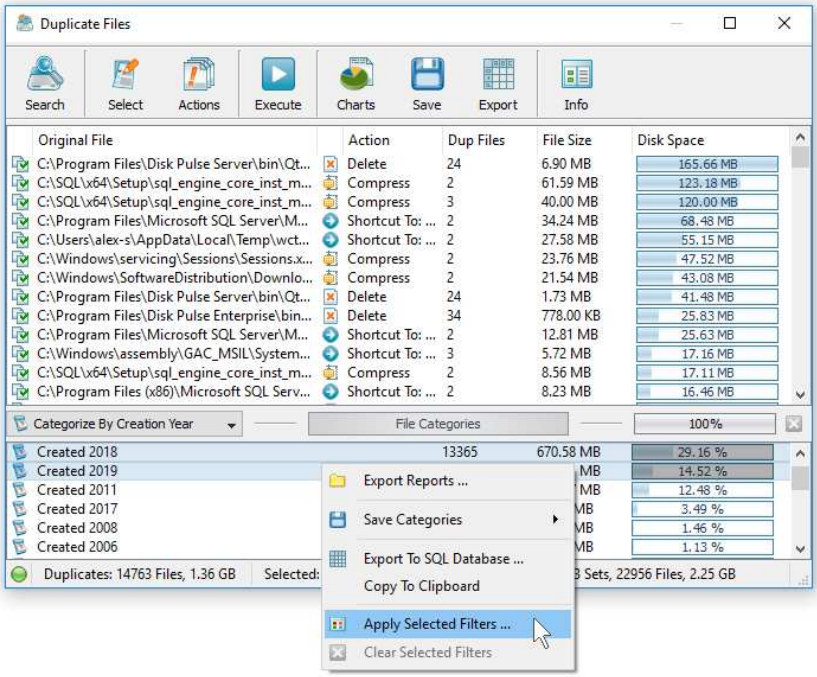


For example, in order to open a chart showing the duplicate disk space and the number of duplicate files per file category, press the 'Charts' button located on the duplicate files search results dialog and select the 'Duplicate Files Per Category' chart type.

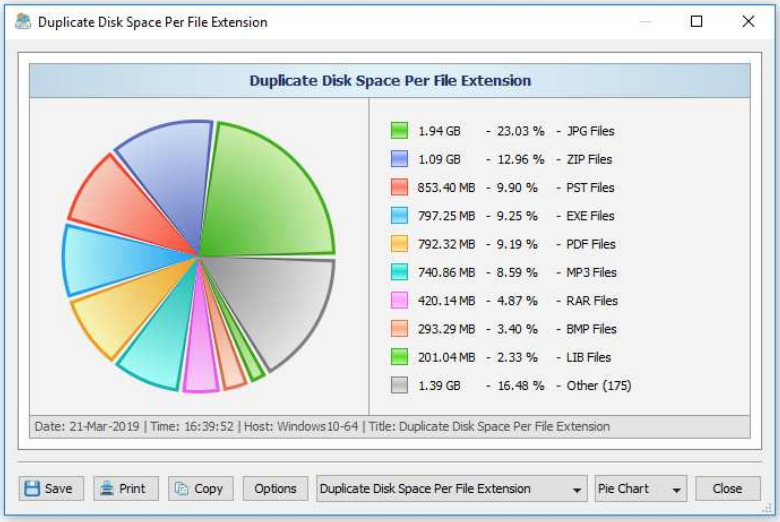


In addition, the charts dialog provides the user with the ability to export the chart data to a number of different formats, export the chart image to the BMP, JPG and PNG image formats and copy the displayed chart image to the clipboard allowing one to easily integrate DiskBoss charts into user's documents and presentations. In order to customize the chart's description, press the 'Options' button and specify a custom chart title, host name, date or time.

One of the most powerful features of the DiskBoss duplicate files finder is the ability to filter duplicate files search results using one or more user-specified filters and then categorize filtered results and display statistical pie charts. For example, in order to display a pie chart showing the amount of duplicate disk space per file extension for duplicate files that were modified during the last two years, set the file categorization mode to 'Categorize By Modification Year', select one or more years to show the information for, press the right mouse button over the file filters view and select the 'Apply Selected Filters' menu item.

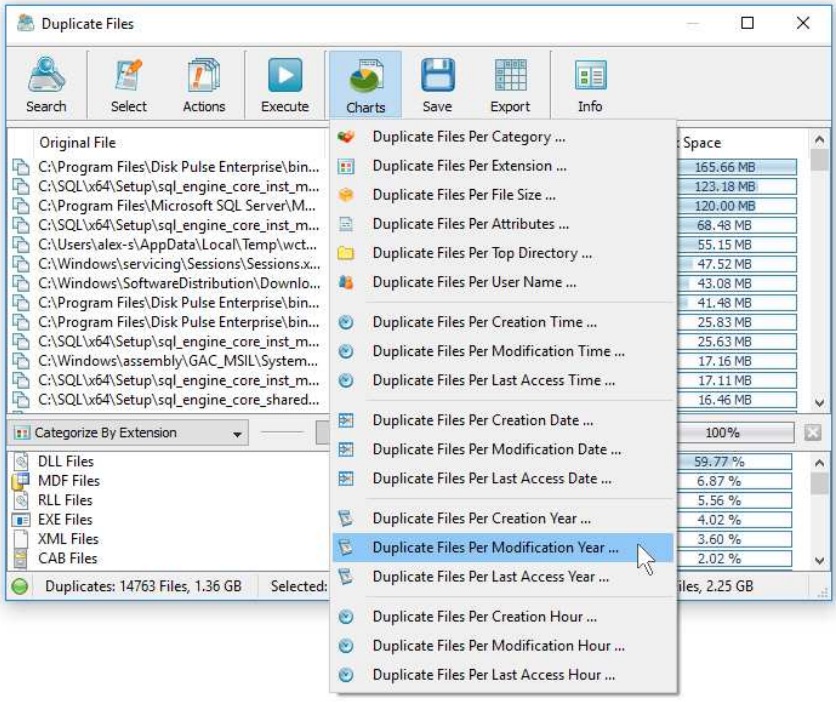


DiskBoss will filter the duplicate files search results and display duplicate files that were modified during the selected years. Now, press the 'Charts' button located on the duplicate files search results dialog and select the 'Duplicate Files Per Extension' menu item. DiskBoss will categorize the filtered duplicate files search results and display a pie chart showing the amount of duplicate disk space per file extension for duplicate files that were modified during the selected years.

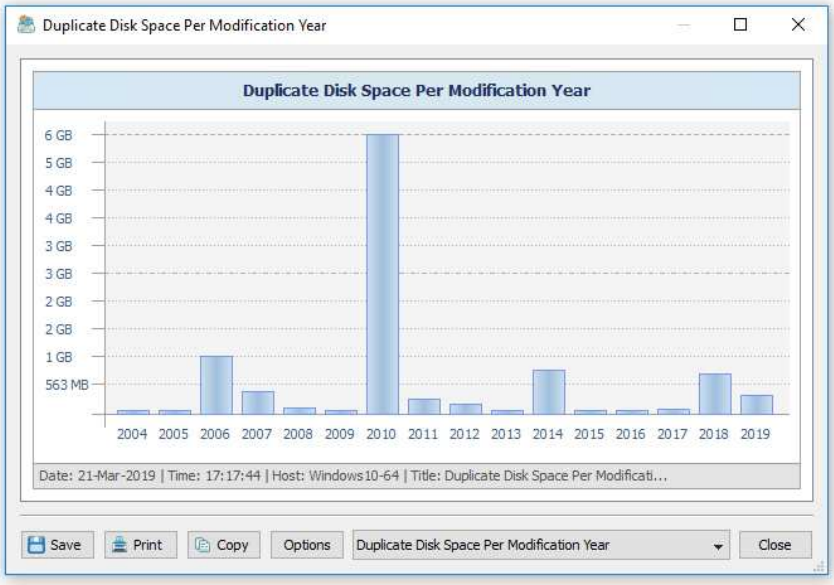


In a similar way, the user can filter duplicate files search results by the file type, size, user name, file attributes, creation, modification and last access dates and then display various types of charts showing the duplicate disk space or the number of duplicate files according to user-specific needs.

Another powerful capability is the ability to display the duplicate disk space timeline charts for all or filtered duplicate files search results. In the simplest case, press the 'Charts' button located on the duplicate files search results dialog and select the 'Duplicate Files Per Modification Year' menu item.



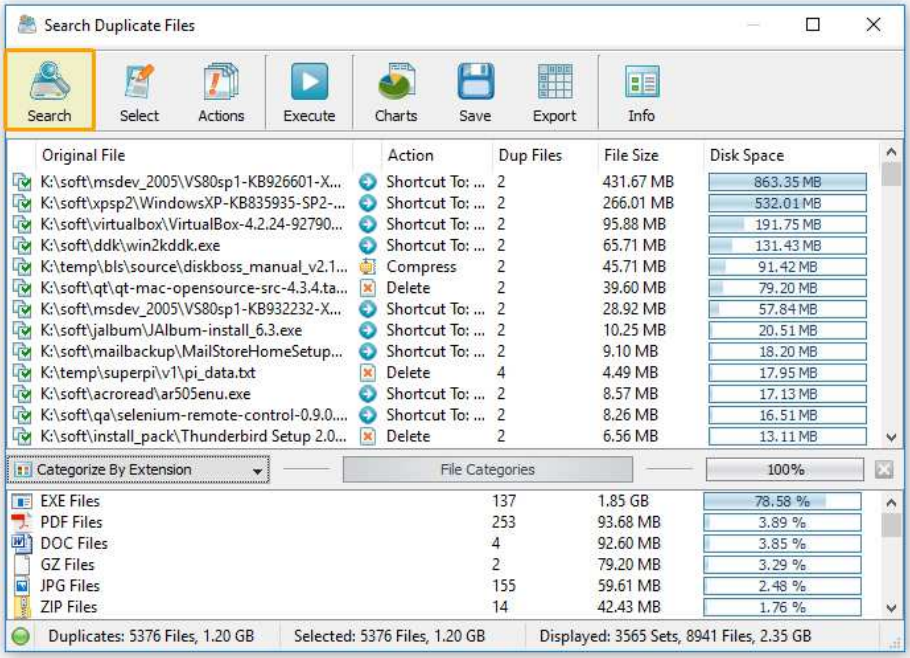
DiskBoss will display a time line chart showing how much duplicate data was modified per year. In order to display how many duplicate files were modified per year, set the chart mode to 'Number Of Duplicates Per Modification Year'.



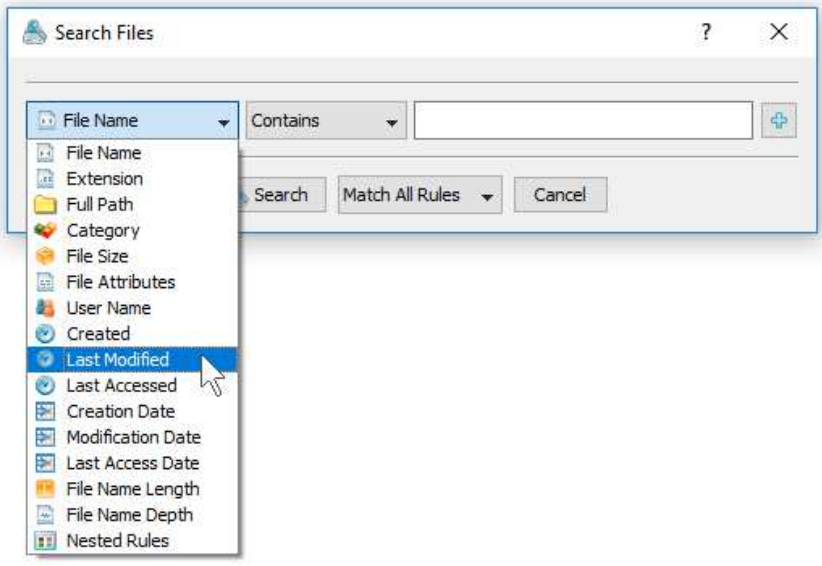
In addition, the user is provided with the ability to display timeline charts for filtered duplicate files search results. For example, in order to display how many duplicate documents were created per year, set the file categorization mode to 'Categorize By File Type', select the 'Documents, Books and Help Files' file category, press the right mouse button and select the 'Apply Selected Filters' menu item. Now, press the 'Charts' button located on the duplicate files search results dialog, and select the 'Duplicate Files Per Creation Year' menu item.

5 Searching Files in Duplicate Files Search Results

DiskBoss provides the ability to search files in duplicate files results by the file name, extension, full path, file category, file size, file attributes, creation, last modification and last access dates. In order to start a file search operation, search duplicate files in one or more disks or directories and press the 'Search' button located on the main toolbar.



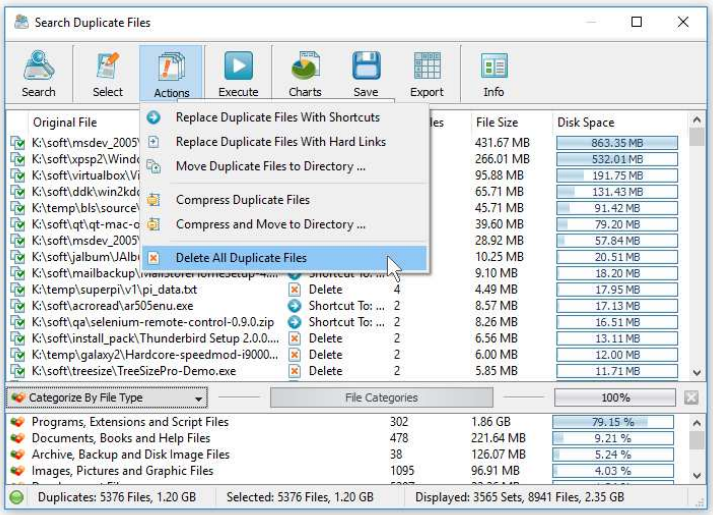
By default, DiskBoss will search duplicate files matching the user-specified rules in all duplicate file sets in the current duplicate files report. Once the file search operation is completed, DiskBoss will display the search results dialog showing all duplicate files matching the rules.



The search results dialog allows one to filter and categorize file search results, display various types of pie charts, copy, move and/or delete files, export file search results to a number of standard report formats including HTML, PDF, Excel, JSON, text, CSV and XML. In addition, advanced users are provided with the ability to export file search results to an SQL database.

6 Selecting Duplicate Files Removal Actions

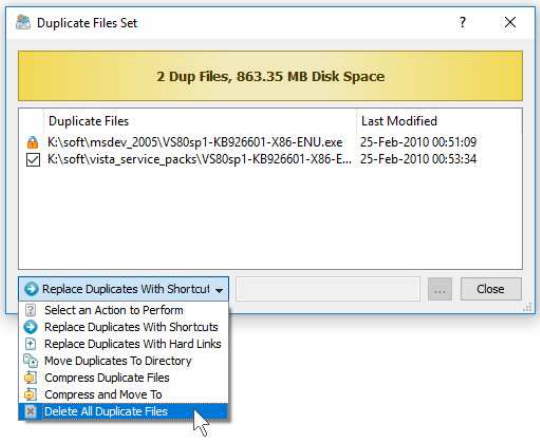
The DiskBoss duplicate files finder allows one to delete duplicate files, move duplicate files to another directory, replace duplicate files with shortcuts pointing to the original file, replace duplicate files with hard links, compress duplicate files, compress and move duplicate files to another directory.



In order to select a specific duplicates removal action for one or more sets of duplicate files, select the sets in the set list, press the right mouse button and select an appropriate duplicate files removal action.

WARNING: There are many duplicate files in the Windows system directory, which are important for proper operation of the operating system. Removal of duplicate files located in the Windows system directory may permanently damage the operating system and render the computer completely non-functional.

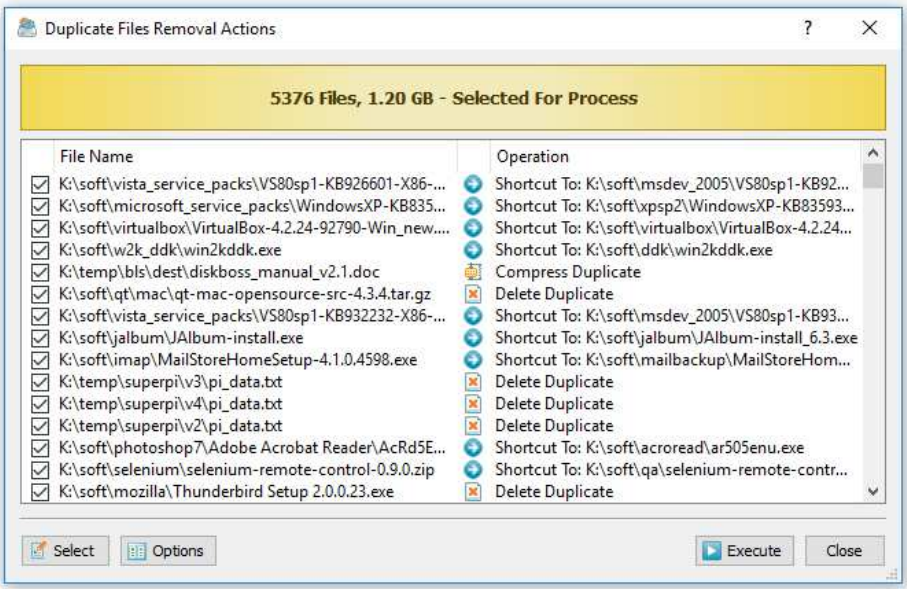
By default, DiskBoss selects the oldest file in each set as the original file and all other files in the set as duplicates. In order to change that, select one or more sets, press the right mouse button and select the 'Select Oldest Files as Duplicates' menu item.



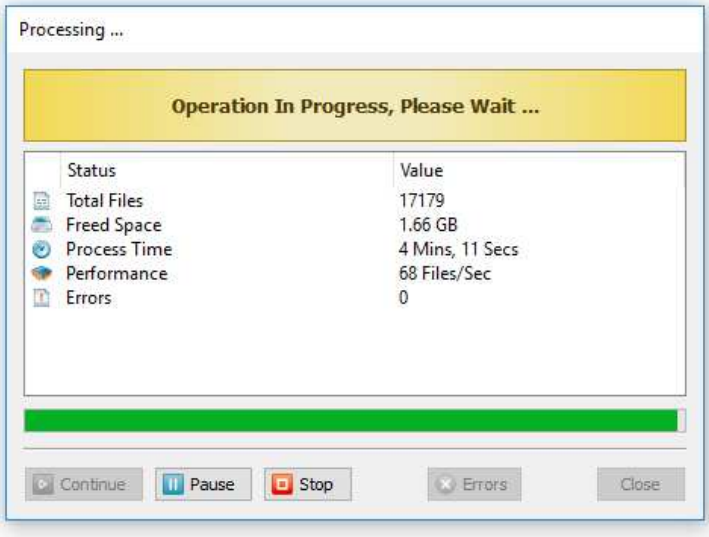
Alternatively, open the duplicate files set dialog, select any arbitrary duplicate file in the set as the original file, select an appropriate duplicate files removal action that should be executed for this specific duplicate files set and select one or more duplicate files in the set that the removal action should be applied to.

7 Executing Duplicate Files Removal Actions

Once finished selecting duplicates and removal actions, press the 'Execute' button to see the duplicate files removal actions preview dialog. The duplicates removal actions preview dialog shows the selected duplicate files and removal actions that will be executed and allows one to review and manually confirm each specific action before execution.



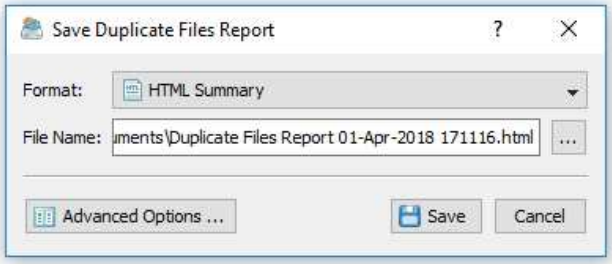
The operating system and other system applications may have a large number of duplicate files located in various system directories. These duplicate files may be very important for proper operation of the operating system and other system applications and it is highly dangerous to remove these duplicate files. To be on the safe side, use the duplicates removal actions only for your own documents, music files, videos, etc.



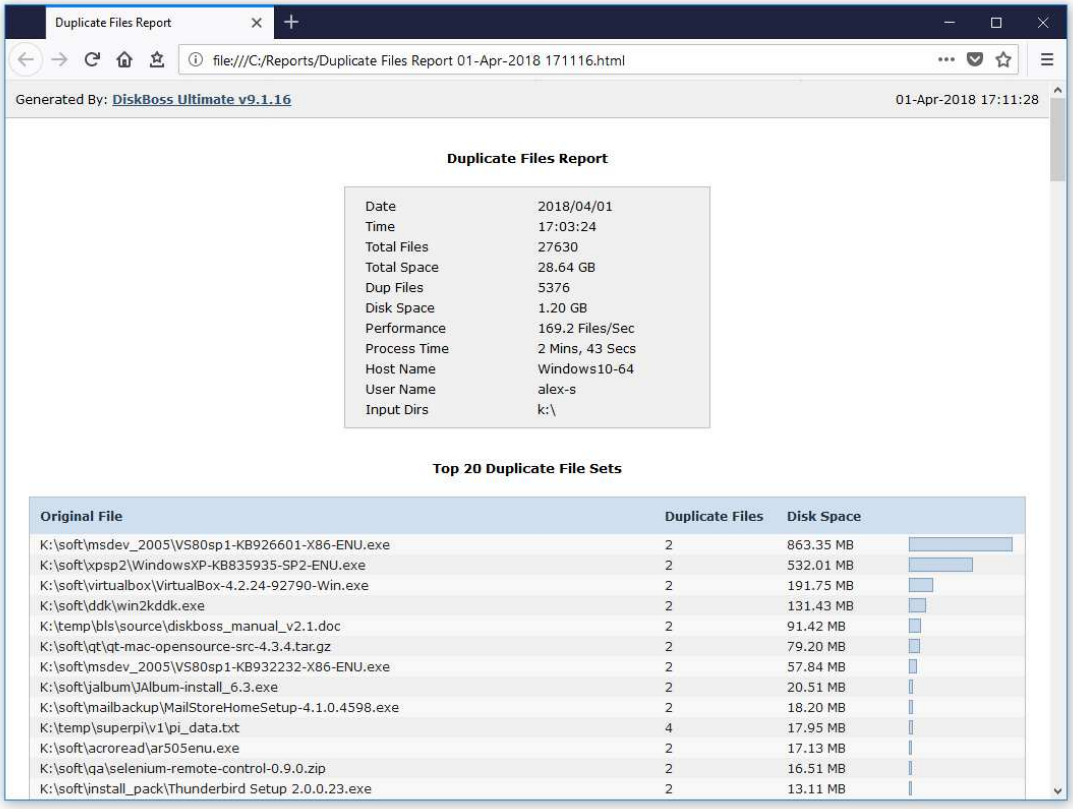
In order to execute the selected duplicates removal actions, press the 'Execute' button located in the bottom-right corner of the 'Preview' dialog. DiskBoss will process the selected duplicate files and execute the specified duplicates removal actions.

8 Saving Duplicate Files Reports

DiskBoss allows one to save duplicate files search reports into a number of standard formats including HTML, PDF, Excel, JSON, XML, text and CSV. In the simplest case, perform a duplicate files search operation and press the 'Save' button located on the duplicate files search results dialog. On the save report dialog, select an appropriate report format, enter a report file name and press the 'Save' button.

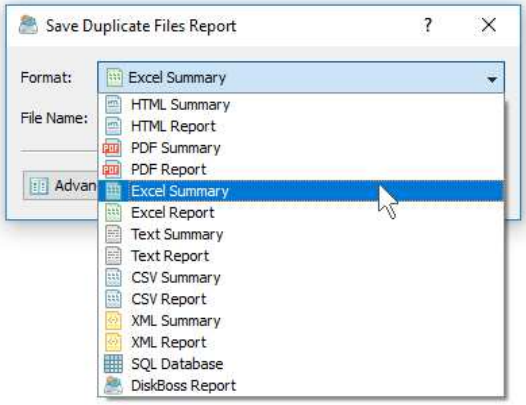


For the HTML, PDF, Excel, text, CSV and XML report formats, the user is provided with the ability to save a short summary report or a longer detailed report, which may be very long for large file systems containing millions of files. By default, DiskBoss will save a short, summary duplicate files search report in the HTML report format, which will include a list of top 20 duplicate file sets sorted by the amount of duplicate disk space and a list of tables showing the amount of duplicate disk space and the number of duplicate files per file extension, file type, top-level directory, user name, etc.

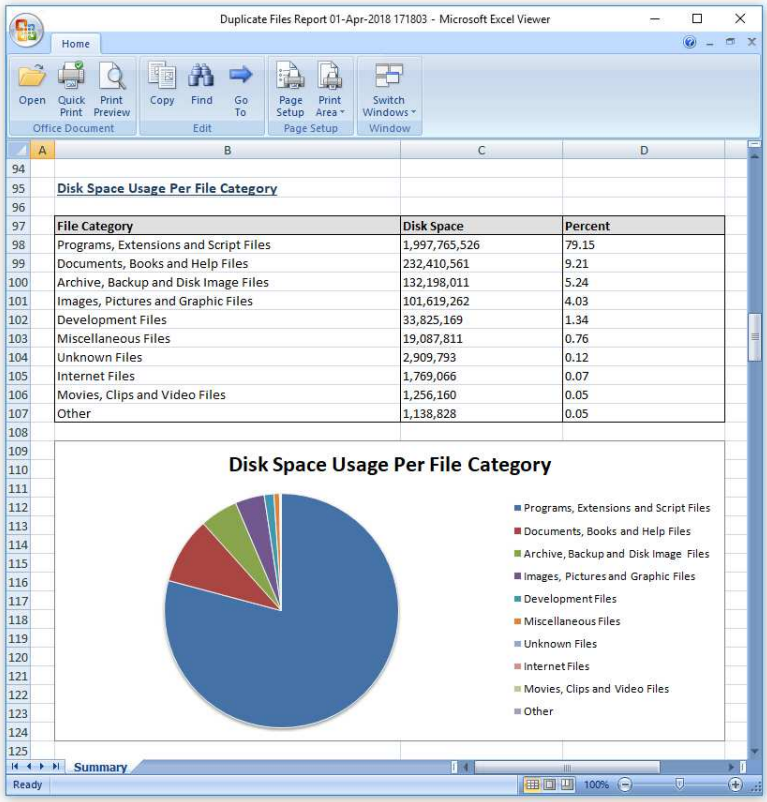


In addition, the user is provided with the ability to save duplicate files search results to the DiskBoss native report format, which preserves all information related to each specific duplicate files search operation and may be loaded at any time just by clicking on a report file in the DiskBoss file navigator.

Sometimes, it may be required to perform additional analysis of duplicate files search results using external tools such as Microsoft Excel. In order to export duplicate files search results to the Excel report format, perform a duplicate files search operation, press the 'Save' button located on the duplicate files search results dialog, select the 'Excel Summary' report format for a short summary report or the 'Excel Report' format for a detailed duplicate files search report.

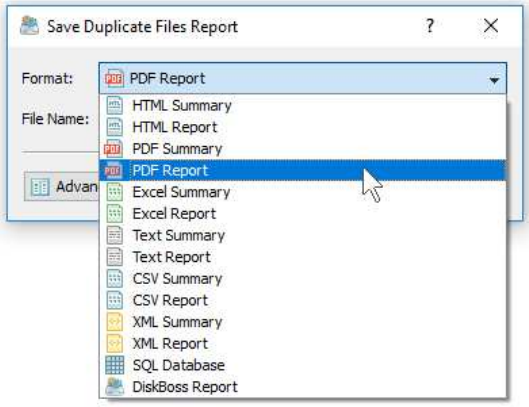


A summary Excel report will include a list of top 20 duplicate file sets sorted by the amount of duplicate disk space and a number of tables showing the amount of duplicate disk space and the number of duplicate files per file extension, file category, file creation time, last modification time, top-level directory, user name, etc.

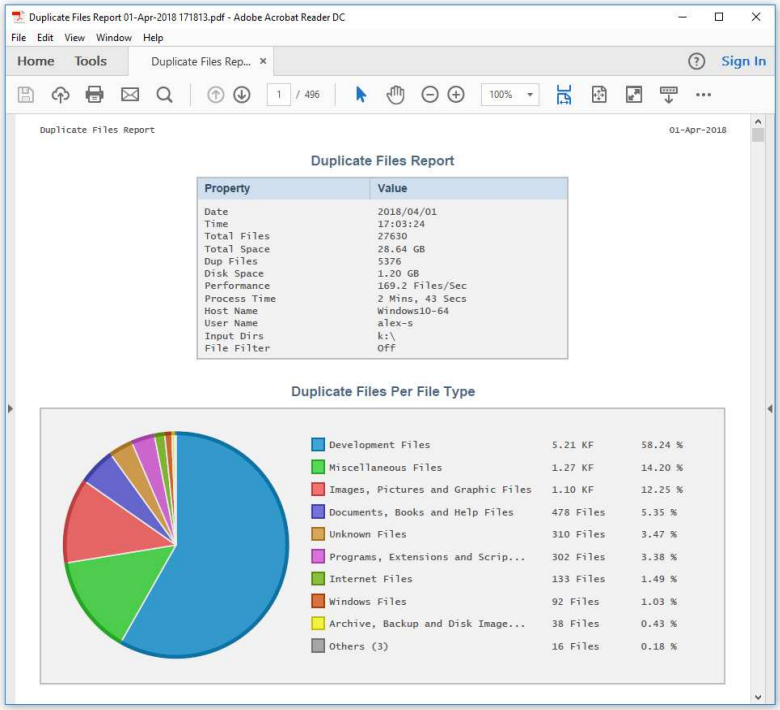


A detailed Excel report will include the list of duplicate file sets sorted by the amount of the duplicate disk space followed by lists of duplicate files in each set, which may be very long for large file systems containing millions of files. In order to control how many duplicate file sets are exported in the detailed report, press the 'Advanced Options' button located on the 'Save Report' dialog and customize the duplicate files search report for your specific needs.

One of the most useful ways to export duplicate files search results is to use the PDF summary or the PDF report formats. Both of these report formats include various types of graphical pie charts showing the amount of duplicate disk space and the number of duplicate files per file extension, file category, creation time, last modification time, user name, etc. In order to save duplicate files search results to a PDF report file, press the 'Save' button located on the duplicate files search results dialog and select the 'PDF Summary' report format for a short, summary report or the 'PDF Report' format for a detailed duplicate files search report.



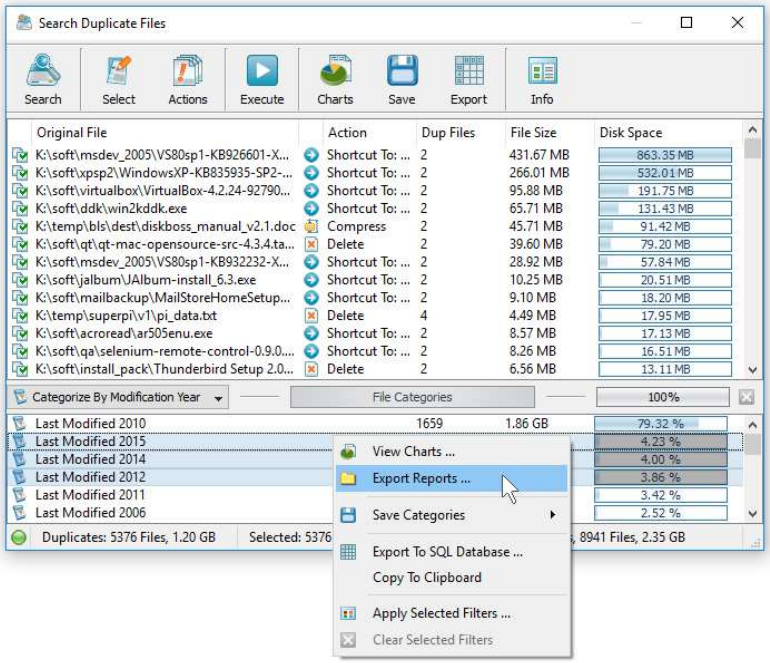
A summary PDF report will include a list of top 20 duplicate file sets sorted by the amount of the duplicate disk space followed by a number of pie charts showing the amount of duplicate disk space and the number of duplicate files per file extension, file category, file creation time, last modification time, user name, etc. A detailed PDF report will include a list of duplicate file sets sorted by to the amount of duplicate disk space followed by lists of duplicate files in each set, which may be very long for large file systems containing millions of files.



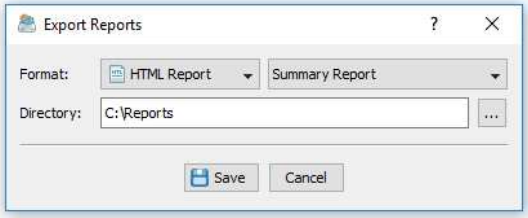
In addition to the list of duplicate file sets sorted by the amount of duplicate disk space, detailed PDF reports include pie charts showing the duplicate disk space per file category and the number of duplicate files per file category according to the currently selected file categorization mode. For example, if the second-level file categories mode is set to categorize duplicate files search results by the file extension, the PDF report will display pie charts showing the amount of duplicate disk space and the number of duplicates per file extension.

9 Batch Duplicate Files Search Reports

DiskBoss allows one to save batches of duplicate files search reports according to the currently selected categories of files with each report showing duplicate files and the amount of duplicate disk space for a corresponding category of files. For example, let's assume that we need to save duplicate files search reports for the last 10 years with each report showing duplicate files and the duplicate disk space for all files that were created during each corresponding year.



In order to generate such reports, set the file categorization mode to 'Categorize By Creation Year', select all years for which you need to generate duplicate files reports in the file categories view, press the right mouse button and select the 'Export Reports' menu item. On the 'Export Reports' dialog, select an appropriate report format and report mode, specify a destination directory to export reports to and press the 'Save' button.

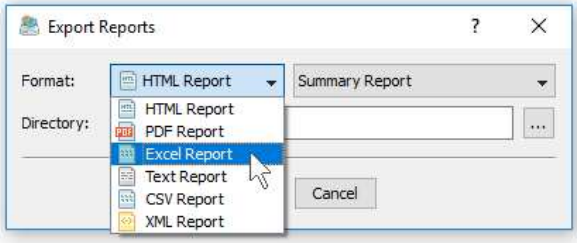


For each selected year, DiskBoss will filter duplicate files search results by the file creation year and save an individual duplicate files search report in the specified destination directory with each report file name and report title containing the corresponding file creation year.

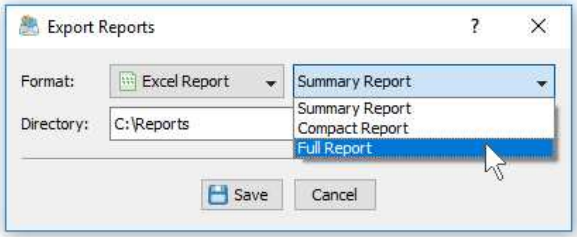
In the same way, the user can generate batches of reports showing duplicate files and the disk space usage per user, last modification date, last access date, file attributes, file size, top directory or custom categories of files pre-configured using flexible file matching rules.

For example, in order to export individual duplicate files search reports for all users, set the file categorization mode to 'Categorize By User Name', select all the required users, press the right mouse button and select the 'Export Reports' menu item. Now, the destination directory will contain a series of duplicate files search reports with each report showing duplicate files and the disk space usage for the corresponding user.

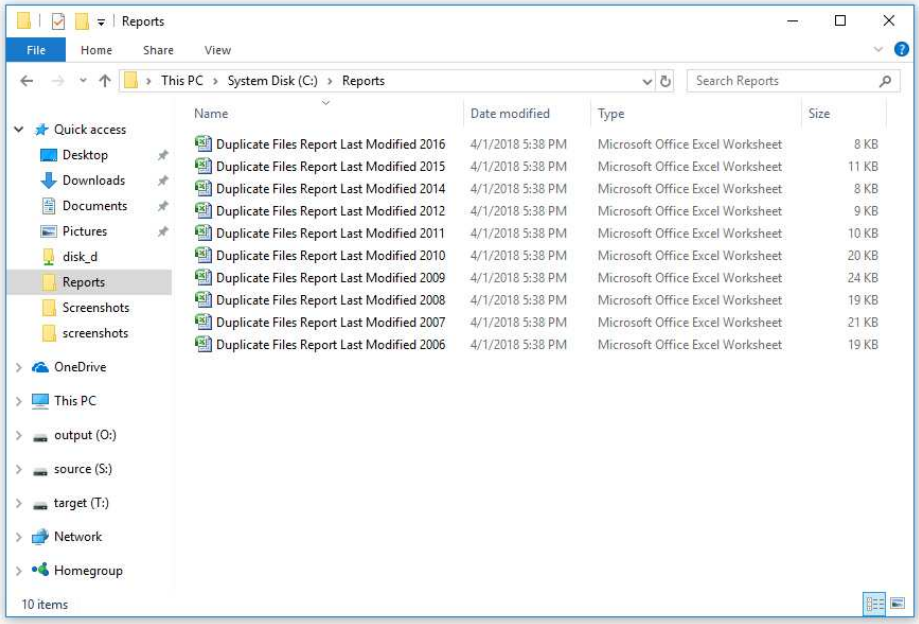
The batch reports dialog provides the ability to export reports to a number of standard report formats including: HTML, PDF, Excel, text, CSV and XML. Depending on the selected report format and report mode, each report will include a summary table showing statistics about the report, a table showing the list of duplicate file sets sorted by the amount of duplicate disk space and lists of duplicate files for each duplicate file set.



In addition, the batch reports dialog allows one to select one of the following report modes: summary report, compact report or full report. When the summary report mode is selected, each report will include a summary table and a list of duplicate file sets sorted by the amount of the used disk space. When the compact report mode is selected, each report will include a summary table and a list of top 100 duplicate file sets sorted by the amount of duplicate disk space. When the full report mode is selected, each report will include a summary table, a full list of duplicate file sets and full lists of duplicate files for each duplicate file set.

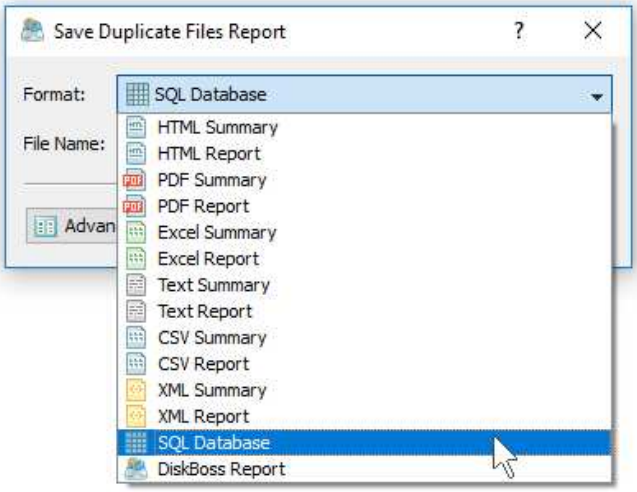


Once the export operation will be completed, the specified destination directory will contain a number of duplicate files search reports according to the selected file categorization mode, selected categories of files and the selected report format.

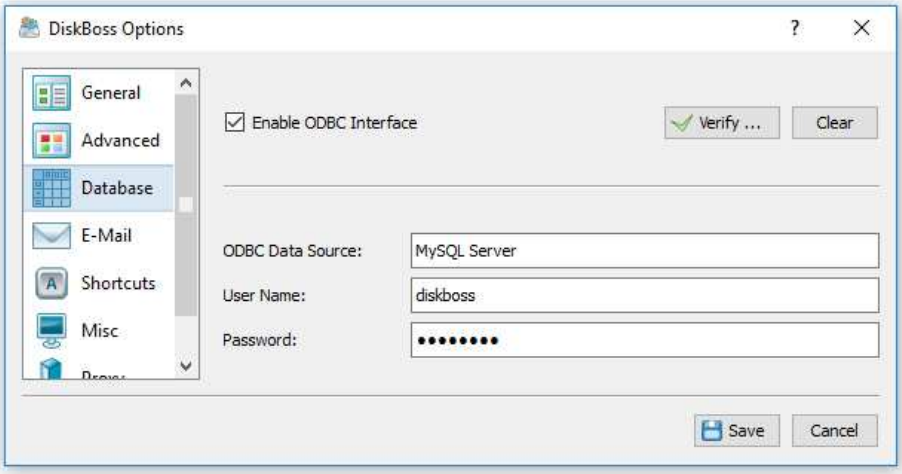


10 Exporting Reports to an SQL Database

IT professionals and storage administrators are provided with the ability to submit reports listing duplicate files detected on multiple storage systems, servers and desktop computers to a centralized SQL database enabling system and storage administrators to gain an in-depth visibility into amounts of duplicate files and wasted disk space across the entire enterprise.



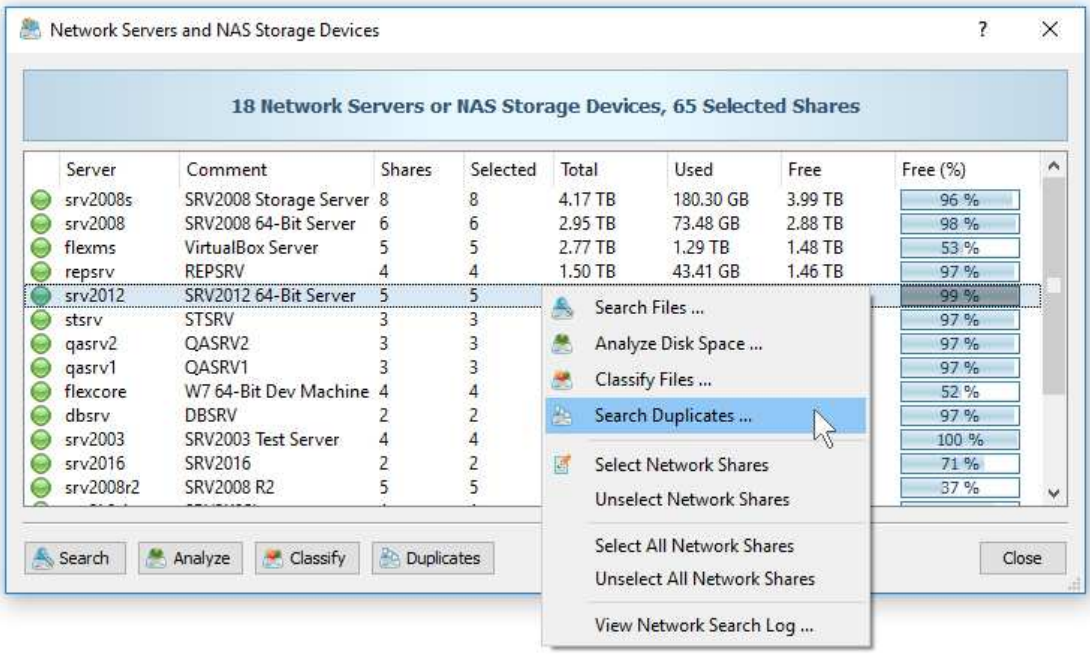
In order to submit a report to an SQL database, press the 'Save' button located on the duplicate files search results dialog toolbar, select the 'SQL Database' report format and press the 'Save' button. Before exporting a report to an SQL database, the user needs to open the options dialog, enable the ODBC interface and specify the name of the ODBC data source, the database user name and password to use for database export operations.



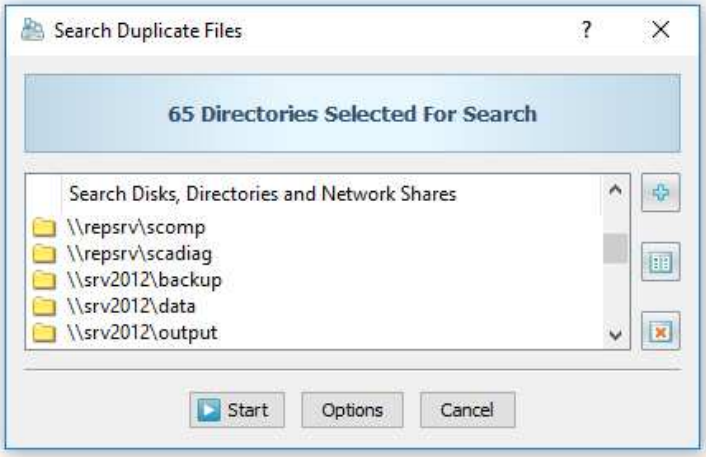
For each report in the database, DiskBoss shows the report date and time, the name of the host computer the operation was performed on, disks and directories that were processed, the total amount of disk space and the number of files that were processed and the report title. In order to open a report, just click on the report item in the report list.

11 Search Duplicate Files in Servers and NAS Devices

DiskBoss allows one to scan the network, discover network servers and NAS storage devices, automatically detect all accessible network shares and search duplicate files in hundreds of network servers and NAS storage devices. In addition, the user is provided with the ability to export the list of detected servers and NAS storage devices (including lists of network shares for each server) into HTML, PDF, text and Excel CSV reports.



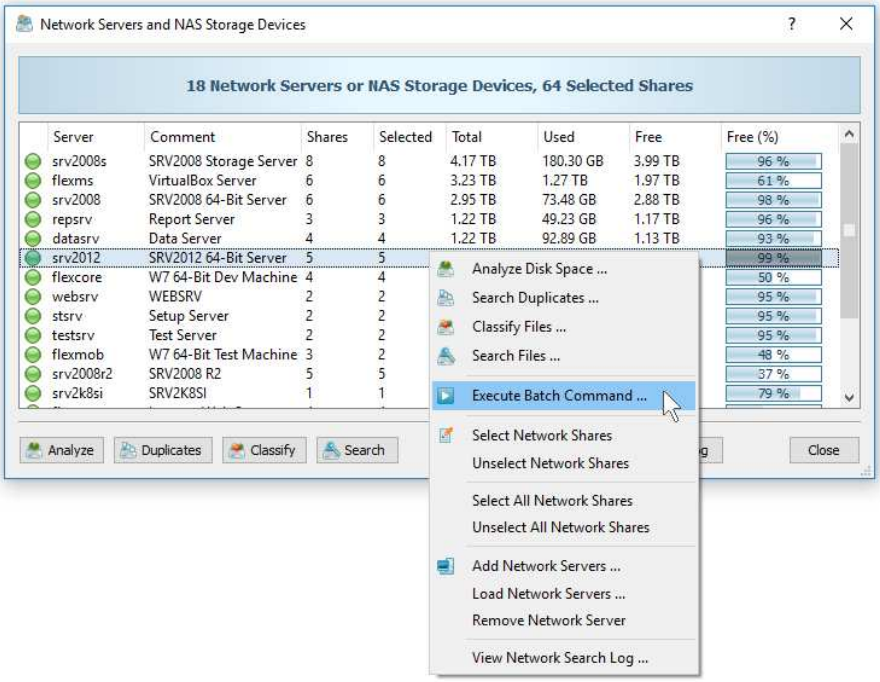
In order to discover all network servers and NAS storage devices on the network, press the 'Network' button located on the main toolbar and wait while DiskBoss will scan the network and show a list of detected network servers and NAS storage devices. In order to search duplicate files in one or more servers or NAS storage devices, select the required servers and NAS storage devices and press the 'Duplicates' button.



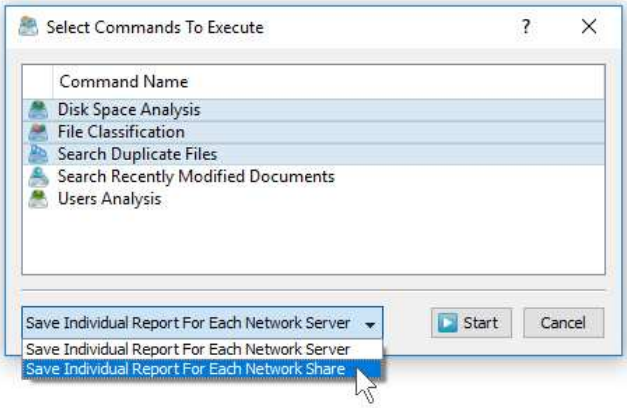
DiskBoss will show all accessible network shares hosted on the selected servers and NAS storage devices allowing one to search duplicate files and save various types of duplicate files pie charts and reports. In addition, the user is provided with the ability to customize a large number of advanced duplicate files search options allowing one to tune duplicate files search operations for user specific needs and hardware configurations.

12 Batch Duplicate Files Search Operations

DiskBoss Server and DiskBoss Enterprise provide the ability to execute one or more pre-configured duplicate files search operations on all network servers and NAS storage devices on the network and generate an individual duplicate files search report for each server and NAS storage device. In order to be able to use batch duplicate files search operations, the user needs to pre-configure one or more duplicate files search commands customized to generate duplicate files search reports according to user-specific needs and requirements.



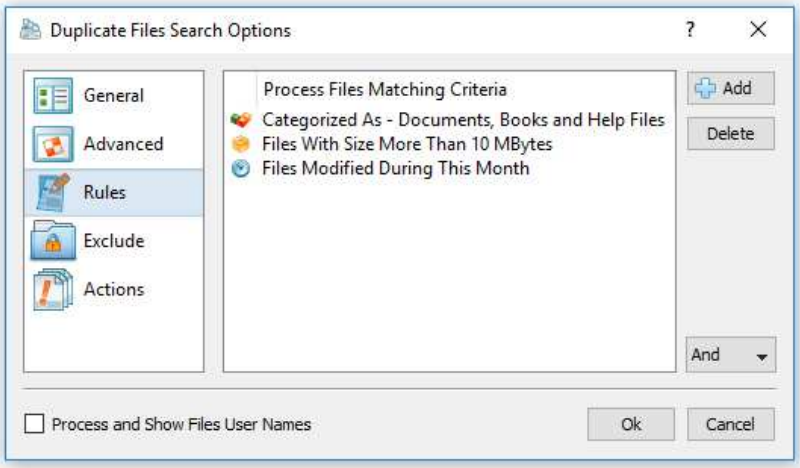
In order to start a batch duplicate files search operation, press the 'Network' button located on the main toolbar, search all servers and NAS storage devices on the network, select one or more servers and NAS storage devices, press the right mouse button and select the 'Execute Batch Command' menu item.



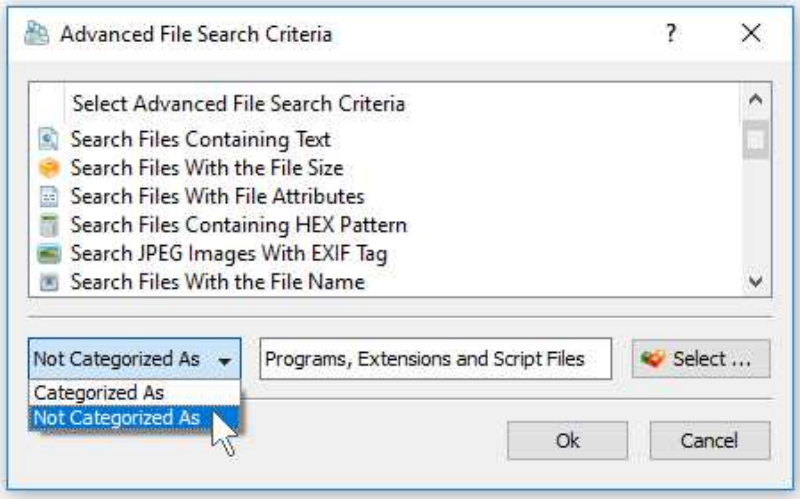
DiskBoss will display a list of pre-configured duplicate files search commands allowing one to select one or more commands to be executed on all selected servers and NAS storage devices. In addition, the user is provided with the ability to select how to save duplicate files search reports - for each server or for each network share. By default, all duplicate files search reports will be saved in the DiskBoss internal reports database allowing one to open each report, review results, generate various types of pie charts and export reports into a number of standard formats including HTML, PDF, Excel, text, CSV and XML.

13 Searching Specific Types of Duplicate Files

One of the most powerful capabilities of DiskBoss is the ability to search specific types of duplicate files according to one or more user-specified file matching rules. Files not matching the specified rules, will be just skipped from the duplicate files search process.



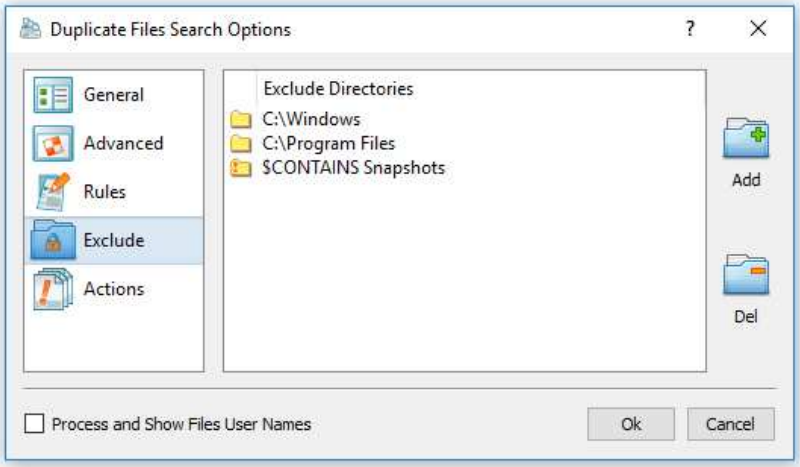
In order to add one or more file matching rules to a duplicate files search operation, open the command dialog, select the rules tab and press the 'Add' button located on the right side of the dialog. Once finished adding file matching rules, select an appropriate rules logic and press the 'Save' button.



Another option is to exclude specific types of duplicate files from the search process using one or more negative file matching rules. For example, in order to exclude all types of programs and executable files from the duplicate files search process, add a file category rule, select the 'Not Categorized As' rule operator and select the 'Programs and Executable Files' file category.

14 Excluding Subdirectories

Sometimes, it may be required to exclude one or more subdirectories from the duplicate files search process. For example, if you need to search duplicate files in a disk excluding one or two special directories, you may specify the whole disk as an input directory and add the directories that should be skipped to the exclude list. By default, in order to prevent accidental deletion of critical system files, DiskBoss automatically adds the operating system directory to the list of exclude directories in all duplicate files search commands.



In order to add one or more directories to the exclude list, open the duplicate files search command dialog, press the 'Options' button, select the 'Exclude' tab and press the 'Add' button. All files and subdirectories located in the specified exclude directory will be excluded from the duplicate files search process. In addition, advanced users are provided with a number of exclude directories macro commands allowing one to exclude multiple directories using a single macro command.

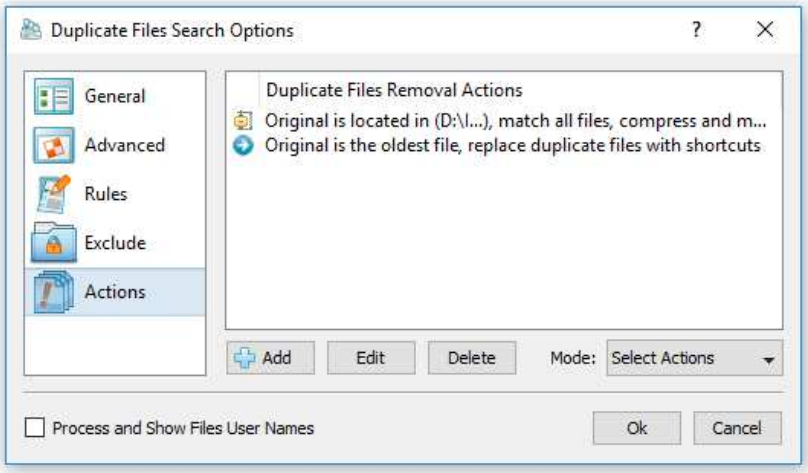
DiskBoss provides the following exclude directories macro commands:

- **$BEGINS <Text String>** - this macro command excludes all directories beginning with the specified text string.
- **$CONTAINS <Text String>** - this macro command excludes all directories containing the specified text string.
- **$ENDS <Text String>** - this macro command excludes all directories ending with the specified text string.
- **$REGEX <Regular Expression>** - this macro command excludes directories matching the specified regular expression.
- **$DIRLIST <File Name>** - this macro command excludes all directories listed in the user-specified text file.
- **$FILELIST <File Name>** - this macro command excludes all files listed in the user-specified text file.

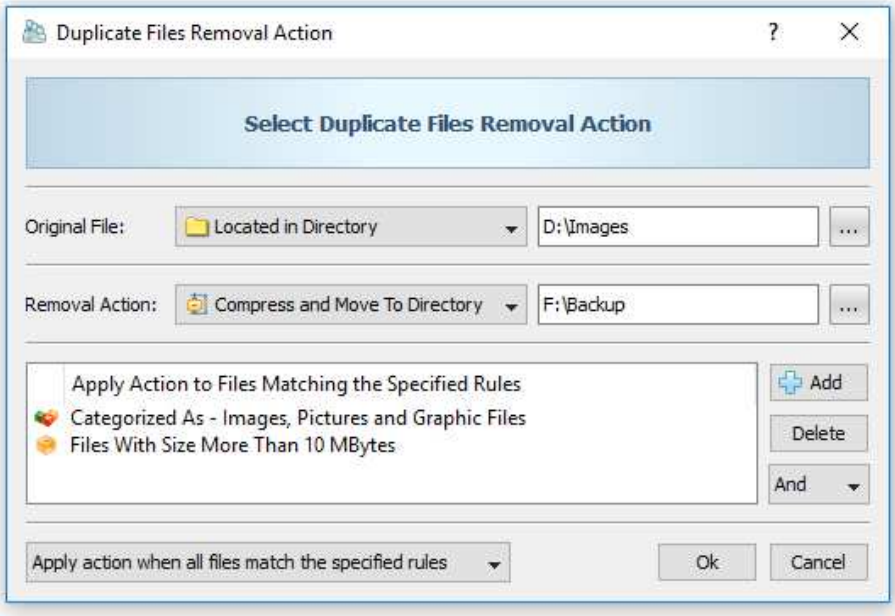
For example, the exclude macro command '$CONTAINS Temporary Files' will exclude all directories with 'Temporary Files' in the full directory path and the exclude macro command '$REGEX \.(TMP|TEMP)$' will exclude directories ending with '.TMP' or '.TEMP'.

15 Automatic Duplicate Files Removal Actions

DiskBoss Ultimate and DiskBoss Server provide the user with the ability to automatically execute one or more duplicate files removal actions for files matching user-specified rules. In order to define one or more automatic duplicate files removal actions, open the duplicate files search command dialog, select the 'Actions' tab and press the 'Add' button.



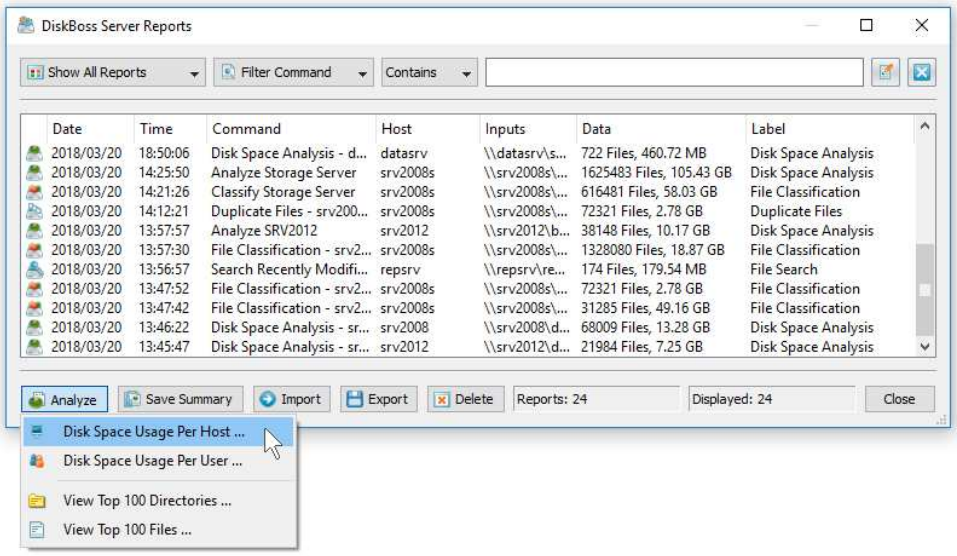
On the 'Action' dialog select the original file detection mode, an appropriate duplicates removal action and specify one or more file matching rules defining files the action should be applied to. During runtime, DiskBoss will process detected duplicate files, apply the specified file matching rules, detect the original file and execute the duplicates removal actions for files matching the specified rules and policies.



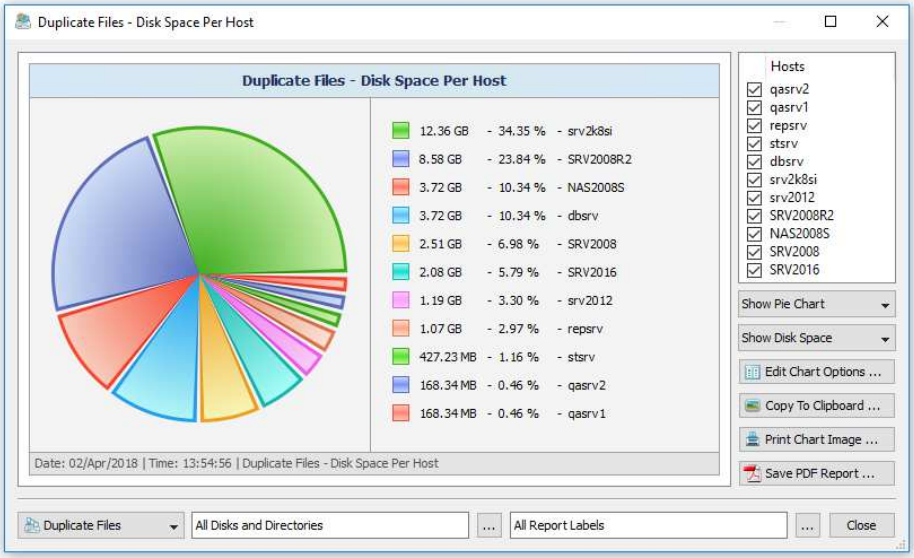
By default, DiskBoss executes automatic duplicates removal actions in the 'Auto-Select' mode, which selects the specified actions and displays the duplicates removal actions preview dialog allowing one to review and manually confirm each specific action. After testing the duplicate files search command in the preview mode, change the actions mode to 'Execute' to automatically execute the specified duplicates removal actions without showing the actions preview dialog.

16 Analyzing Duplicate Files Per Host

DiskBoss Server and DiskBoss Enterprise provide the ability to automatically detect all servers and NAS storage devices on the network, search duplicate files in hundreds of servers and/or NAS storage devices via the network, submit duplicate files search reports to a centralized report database and display charts showing the number of duplicate files and the amount of duplicate disk space per server or NAS storage device across the entire enterprise.



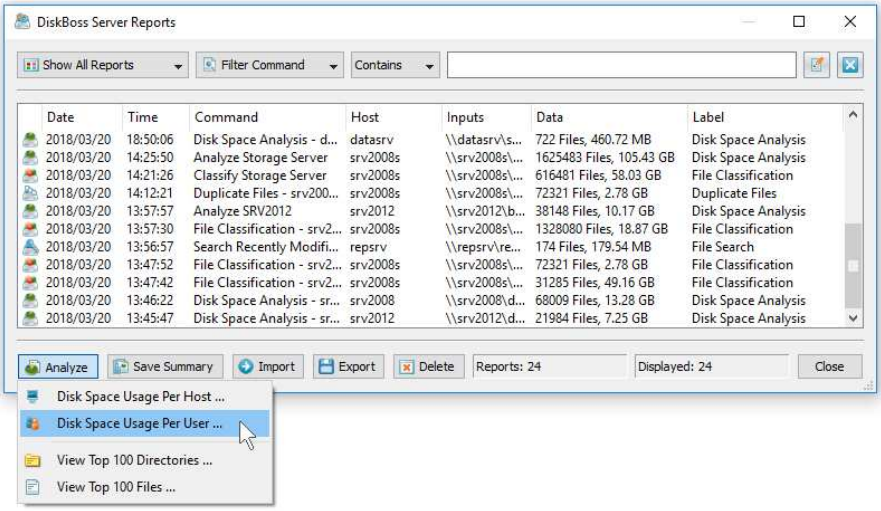
In order to analyze duplicate files per server, perform one or more duplicate files search operations on multiple servers and/or NAS storage devices, open the 'Reports' dialog, press the 'Analyze' button and select the 'Analyze Disk Space Usage Per Host' menu item. DiskBoss will analyze all reports saved in the reports database and display the hosts analysis dialog showing the number of duplicate files and the amount of duplicate disk space per host.



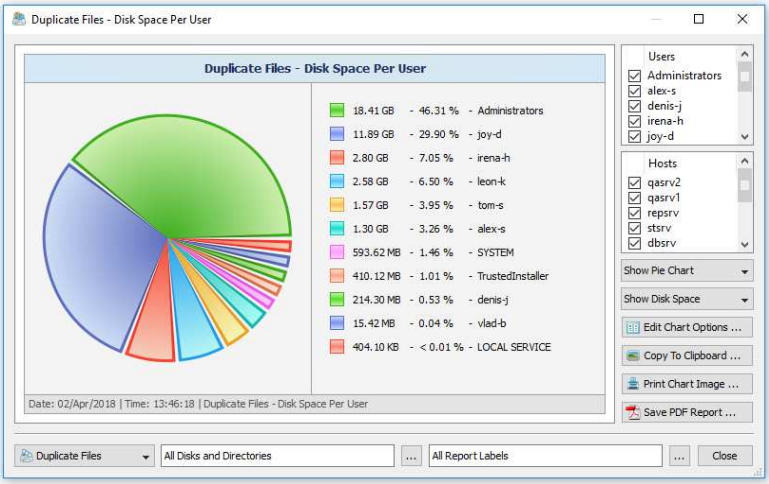
The hosts analysis dialog provides the ability to display pie charts and bars charts showing the number of duplicate files and the amount of duplicate disk space per host according to duplicate files search reports saved in the reports database. The user is provided with the ability to select the types of duplicate files search operations and file system locations to analyze, edit the chart header and footer, copy the chart image to the clipboard and export graphical PDF reports including pie charts.

17 Analyzing Duplicate Files Per User

DiskBoss Server and DiskBoss Enterprise provide the ability to automatically detect all servers and NAS storage devices on the network, search duplicate files in hundreds of servers and/or NAS storage devices via the network, submit duplicate files search reports to the reports database and display charts showing the number of duplicate files and the amount of duplicate disk space per user across the entire enterprise.



In order to analyze duplicate files per user, perform one or more duplicate files search operations on multiple servers and/or NAS storage devices, open the 'Reports' dialog, press the 'Analyze' button and select the 'Analyze Disk Space Usage Per User' menu item. DiskBoss will analyze all reports saved in the reports database and display the users analysis dialog showing the number of duplicate files and the amount of duplicate disk space per user.

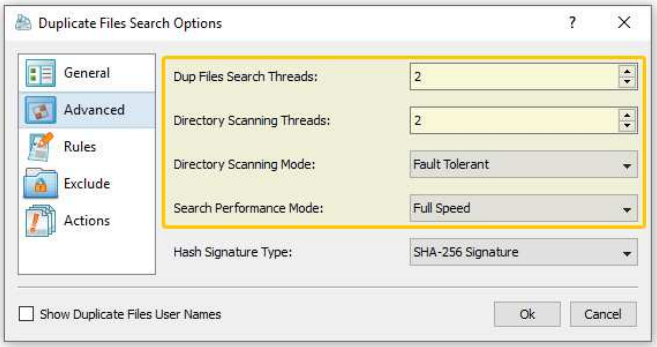


The users analysis dialog provides the ability to display pie charts and bars charts showing the number of duplicate files and the amount of duplicate disk space per user according to duplicate files search reports saved in the reports database. The user is provided with the ability to select the types of duplicate files search operations and file system locations to analyze, edit the chart header and footer, copy the chart image to the clipboard and export graphical PDF reports including pie charts.

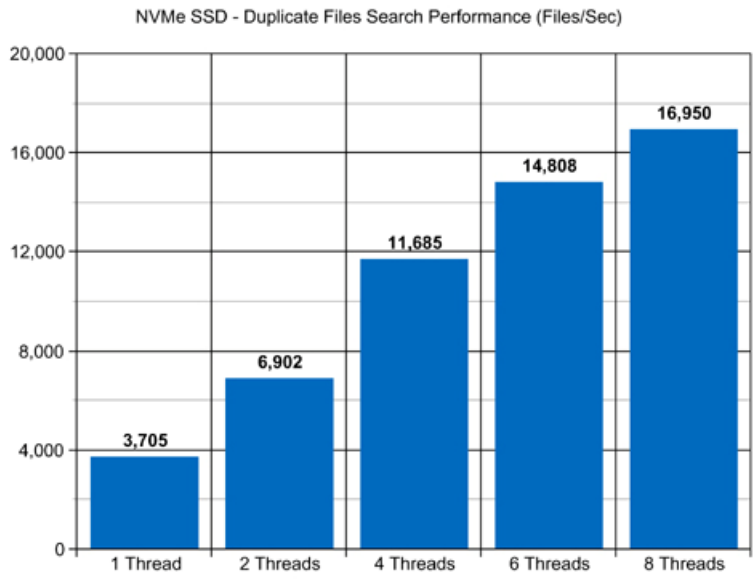
IMPORTANT: In order to be able to display duplicate files per user, the duplicate files search operation should be configured to process and display files user names.

18 Duplicate Files Search Performance

DiskBoss is optimized for modern multi-core and multi-CPU systems and is capable of searching duplicate files stored on multiple disks, directories or network shares in parallel using all CPUs installed in the computer. DiskBoss provides a number of different performance optimization options allowing one to tune the duplicate files search operations for user-specific hardware and storage configurations.

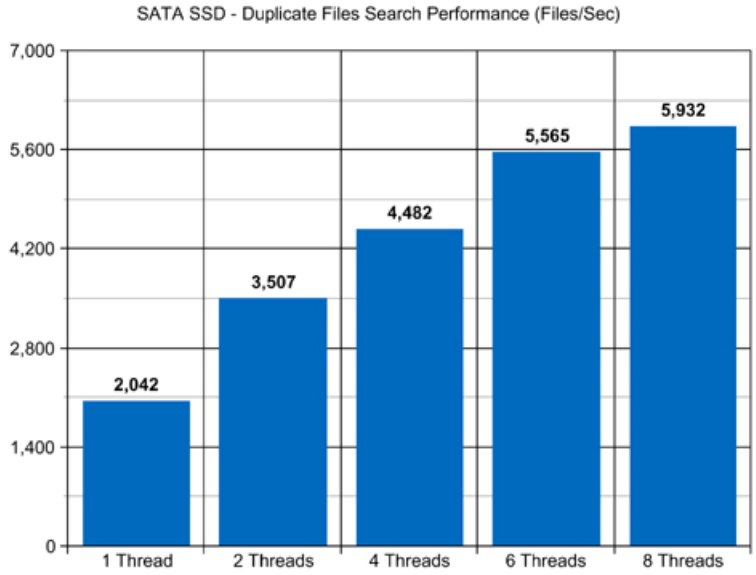


In order to customize the duplicate files search performance optimization options, open the duplicate files search operation dialog, press the 'Options' button and select the 'Advanced' tab. The 'Dup Files Search Threads' option controls how many parallel threads are used to search duplicate files. The 'Directories Scanning Threads' option controls how many parallel threads are used to scan input disks, directories and network shares. In the 'Fault-Tolerant' directory scanning mode, DiskBoss uses an individual processing thread for each input disk, directory or network share, but limits the maximum number of parallel scanning threads to the specified value. In the high-performance directory scanning mode, DiskBoss always uses the specified number of parallel directory scanning threads even when processing a single input disk, directory or network share.

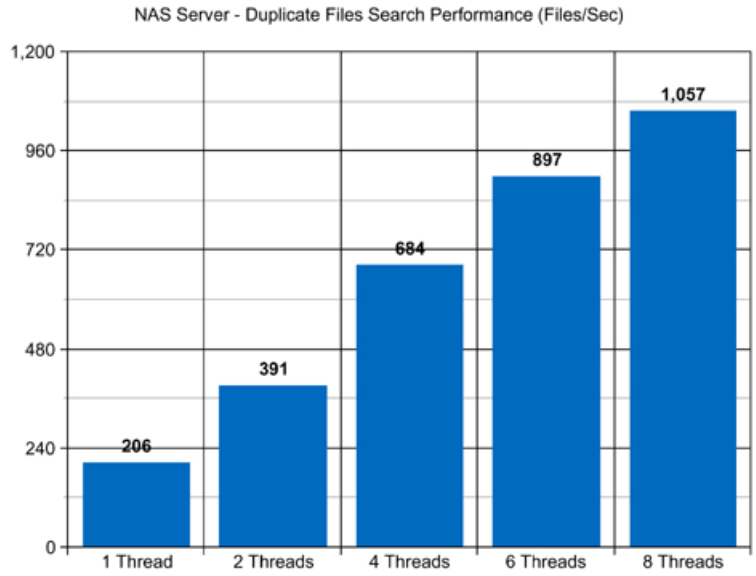


For example, when searching duplicate files stored on a high-speed NVMe SSD disk, DiskBoss reaches up to 3,000 files/sec using a single search thread. With two parallel search threads, the performance scales up to 6,000 files/sec and with four parallel search threads, the performance increases up to 11,000 files/sec showing a very good level of multi-threaded performance scalability. With six processing threads the duplicate files search performance reaches up to 14,000 files/sec and with eight processing threads the performance increases up to 16,000 files/sec allowing one to quickly process large numbers of files and identify how many files are duplicates and how much duplicate disk space these files are using.

When searching duplicate files stored on regular SATA SSD drives, which are significantly slower than NVMe SSD drives, the performance of the duplicate files search process reaches up to 2,000 files/sec using a single process thread and scales up to 4,400 files/sec with four parallel duplicate files search threads. With eight parallel threads, the performance reaches up to 5,900 files/sec, which allows to process large numbers of files relatively fast.

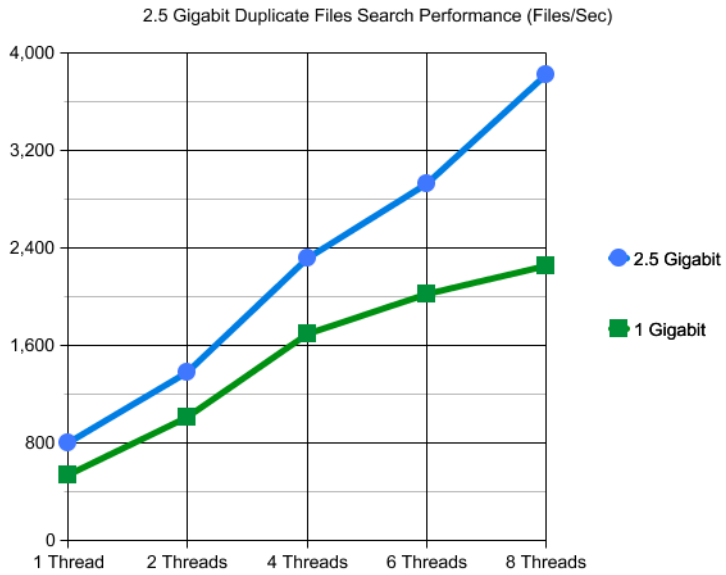


Searching duplicate files stored on a NAS storage device via a network is a more complicated task because the user needs to take into account the speed and the latency of the network. If the computer, on which DiskBoss is installed, is connected to the NAS storage device via a high-speed, low-latency network, the performance of the duplicate files search operations may reach up to 200 files/sec with one duplicate files search thread, scale up to 684 files/sec with four parallel search threads and increase up to 1,057 files/sec with eight parallel duplicate files search threads.

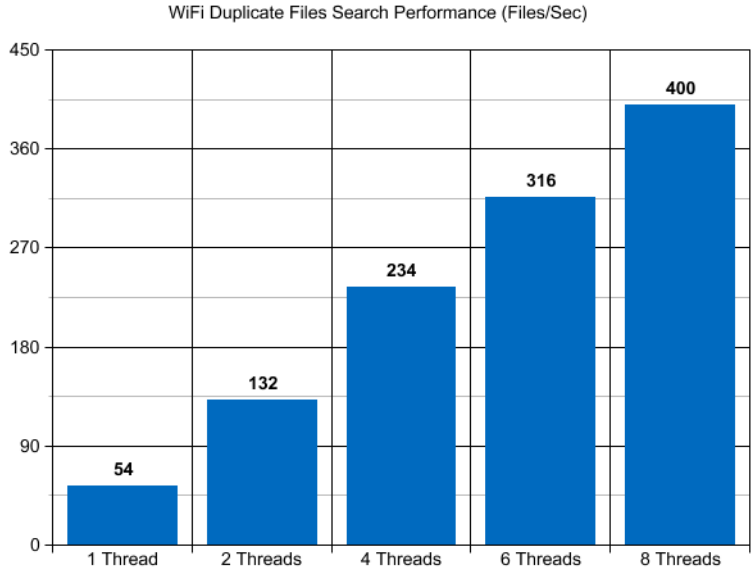


On the other hand, if DiskBoss will need to access network shares via the Internet or via a long-distance, high-latency network, the performance of the duplicate files search operations will be relatively slow. One of the options to increase the performance of the duplicate files search operations in such configurations is to set the 'High-Performance' directory scanning mode and increase the number of parallel duplicate files search threads to 16 or even 32 disregarding how many CPUs are actually installed on the computer.

Searching duplicate files stored in one or more NAS servers may be a very time consuming operation and one of the ways to speed-up the duplicate files search process is to use a 2.5 Gigabit Ethernet network. With 2.5 Gigabit Ethernet the performance of the DiskBoss duplicate files search operations continues to scale up to 3,800 Files/Sec with 8 parallel duplicate files search threads, which represents a 69% improvement compared to the standard Gigabit Ethernet.

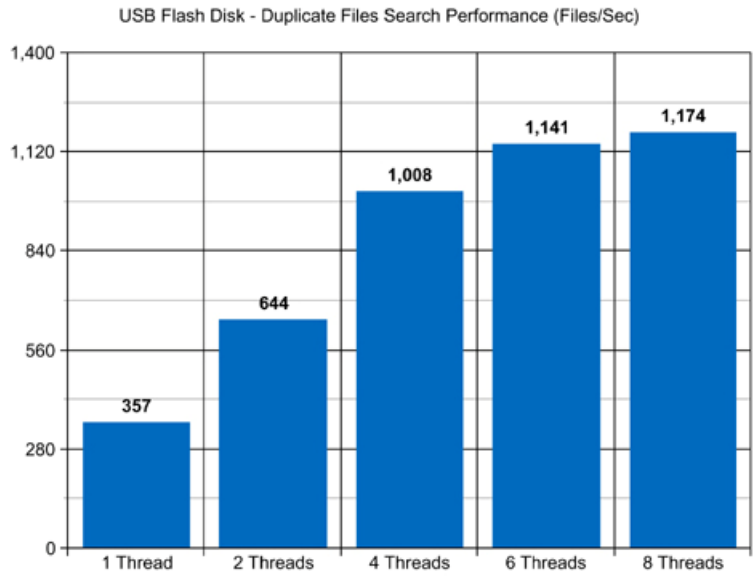


Due to a very wide adoption of laptops and NAS servers with built-in WiFi network interfaces, many users may consider searching duplicate files stored in NAS servers via the wireless network. But, the latency of the wireless network is much higher and therefore it will take much more time to complete the duplicate files search operation via the wireless network. The question is how much longer the user will need to wait and if it will save any significant amount of time to search duplicate files via a wired network.

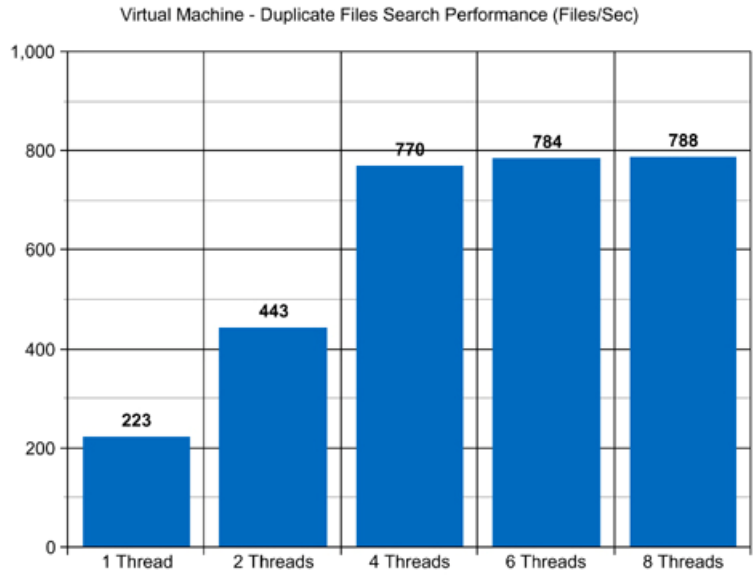


Based on our benchmarks, via a 5 GHz wireless network, DiskBoss reaches up to 54 Files/Sec with a single duplicate files search thread and scales up to 400 Files/Sec with 8 parallel duplicate files search threads, which is approximately 6 times slower compared to the standard Gigabit Ethernet and approximately 10 times slower when compared to the 2.5 Gigabit Ethernet. So, if the user needs to search duplicate files in a NAS server with 100,000 files or more, a low-latency Gigabit Ethernet or 2.5 Gigabit Ethernet is required.

Modern USB flash drives provide plenty of the storage space and are reasonably fast allowing one to store vast amounts of data for backup purposes. Sometimes, it may be required to search duplicate files on a USB flash drive in order to free the used disk space. When searching duplicate files stored on a USB flash drive, DiskBoss can reach up to 357 files/sec with a single search thread. With two parallel search threads, the performance increases up to 644 files/sec, with four parallel threads the performance increases up to 1,008 files/sec and with eight parallel duplicate files search threads the performance scales up to 1,174 files/sec.



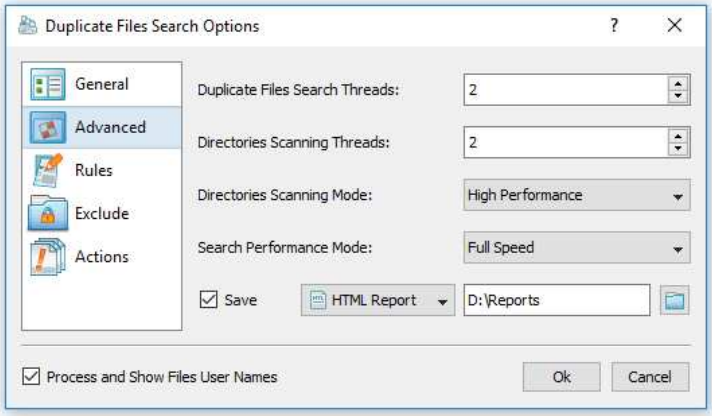
Today, modern IT environments widely deploy virtual servers and/or virtual workstations. Most of the popular virtualization platforms provide a high level of performance, but depending on the target hardware and software platforms, significant performance degradations are inevitable when a duplicate files search operation is executed on a guest virtual machine compared to the same duplicate files search operation executed directly on the host computer.



For example, when a virtual machine with 4 virtual CPUs is stored on an SSD disk and searching duplicate files stored on a virtual local disk drive, which is physically stored on the same SSD disk, the performance of the duplicate files search operations reaches up to 223 files/sec using a single search thread. With two parallel search threads, the performance of the duplicate files search operations scales up to 443 files/sec and with four parallel search threads, the performance of the duplicate files search operations increases up to 770 files/sec.

19 Advanced Duplicate Files Search Options

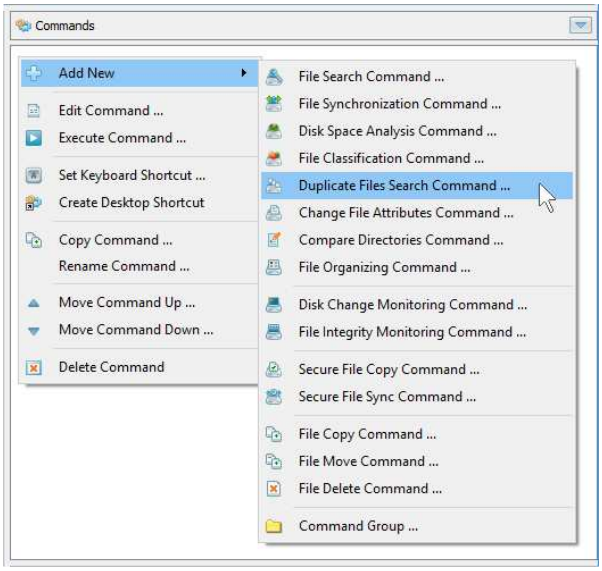
The DiskBoss duplicate files finder provides a large number of advanced options allowing one to customize duplicate files search operations for user-specific hardware and storage configurations. The 'General' tab allows one to control the file signature type, the file scanning mode, the maximum number of duplicate file sets to display in the results dialog.



The 'Advanced' tab provides the ability to intentionally slow down the duplicate files search process in order to minimize the potential performance impact on running production systems. The 'Exclude' tab allows one to define one or more subdirectories to be excluded from the duplicate files detection process.

20 Pre-Configured Duplicate Files Search Commands

One of the most powerful and flexible capabilities of DiskBoss is the ability to pre-configure custom duplicate files search operations as user-defined commands and execute such commands in a single mouse click using the DiskBoss GUI application or direct desktop shortcuts.



User-defined commands may be managed and executed through the commands dialog or the commands tool pane. In order to add a new command through the commands pane, press the right mouse button over the pane and select the 'Add New - Duplicate Files Search Command' menu item. In order to execute a previously saved command, just click on the command item in the commands tool pane or create a direct desktop shortcut on the Windows desktop.

21 Using the DiskBoss Command Line Utility

In addition to the DiskBoss GUI application, DiskBoss Ultimate provides a command line utility allowing one to execute duplicate files detection and removal operations from batch files and shell scripts. The command line tool is located in the **'<ProductDir>\bin'** directory.

Command Line Syntax:

diskboss -duplicates -dir <Input Directory 1> [... <Input Directory X> <Options>]

Parameters:

-dir < Directory 1> [... < Directory X> -file <File 1> ... <File 2>]

This parameter specifies the list of input directories or files to process. In order to ensure proper parsing of command line arguments, directories and file names containing space characters should be double quoted.

Options:

-signature_type <MD5 | SHA1 | SHA256>

This parameter sets the type of algorithm used to calculate signatures of files. By default, DiskBoss uses the SHA256 algorithm.

-exclude_dir <Exclude Directory 1> [... <Exclude Directory X>]

This parameter specifies the list of directories that should be excluded from processing. In order to ensure proper parsing of command line arguments, directories containing space characters should be double quoted.

-filter <FileFilter>

This parameter sets the directory search filter (default *.*).

-workers <WorkingThreadCount>

This parameter sets the number of working threads to process files. DiskBoss is optimized for Multi-Core and Multi-CPU computers and is capable of distributing the workload to an unlimited number of CPUs. By default, DiskBoss processes files with one working thread.

-max_dup_set <MaxNumberOfDuplicateSets>

This parameter sets the maximum number of duplicate file sets to report about. By default, DiskBoss will report about up to 1000 duplicate file sets sorted by the amount of wasted storage space.

-min_wasted_space <MinWastedStorageSpace>

This parameter sets the minimum amount of wasted storage space to report about. By default, DiskBoss will report about duplicate file sets wasting at least 1 MBytes of storage space.

-save_html_report | save_csv_report | save_text_report [ReportFileName]

This parameter saves a report file. If no file name is specified, DiskBoss will automatically generate a file name according to the following template:

diskboss_duplicates_[date]_[time].html

-v - This command shows the product's version, revision and build date.

-help - This command shows the command line usage information.