DiskBoss
DATA MANAGEMENT

File Classification

Version 14.1

Jan 2024

1   Product 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.
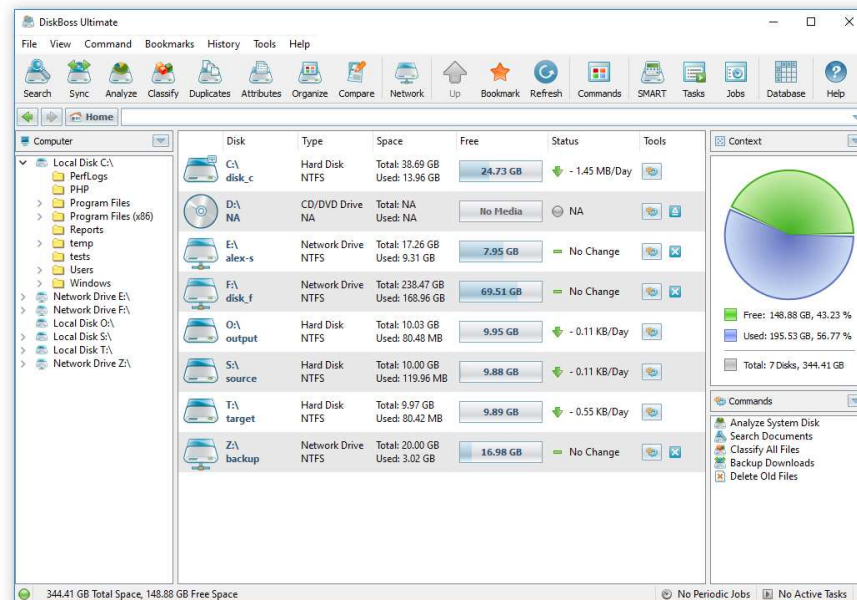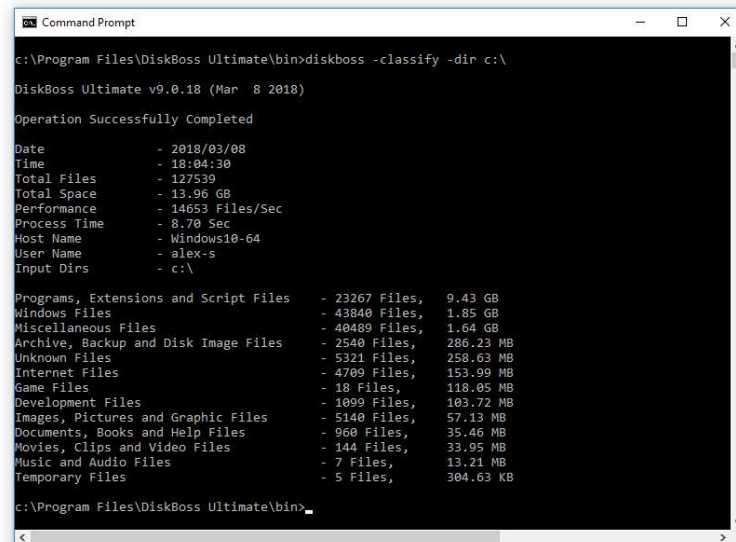
All disk space analysis and file management operations are integrated into a centralized and easy-to-use GUI application allowing one pre-configure analysis and file management operations as user-defined commands and execute any required command in a single mouse click using the DiskBoss GUI application 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, 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.
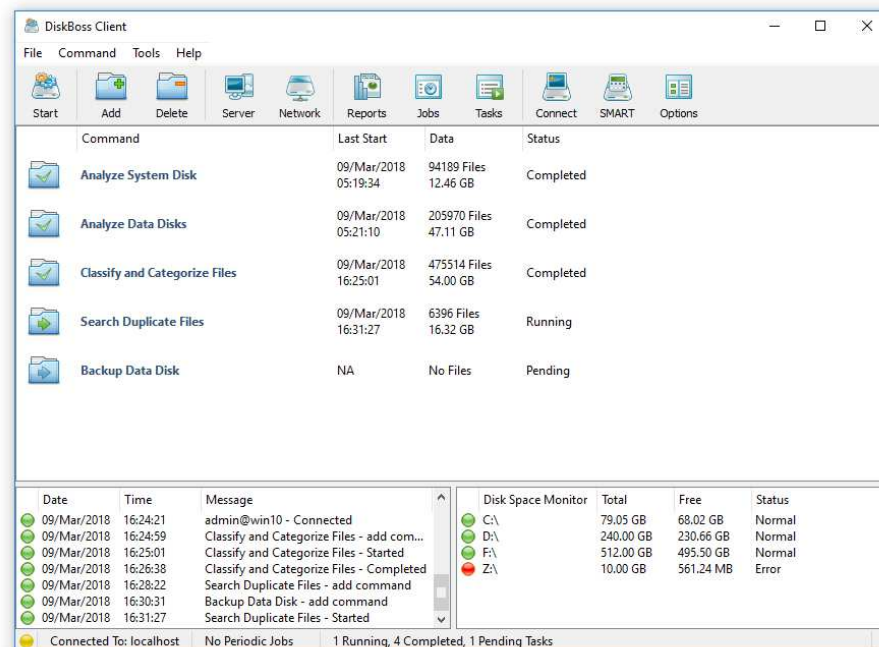
DiskBoss Server can be controlled locally or through the network using the DiskBoss client GUI application or the command line utility. DiskBoss Server provides the ability to pre-configure various types of disk space analysis and/or policy-based file management operations, schedule periodic jobs, save 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   File Classification

DiskBoss recognizes more than 3,500 types of files and is capable of classifying files and showing a categorized hierarchy of files sorted by the amount of the disk space used by each file category. DiskBoss includes a number of file classification plugins capable of classifying files by the file type, extension, user name, last access, modification and creation dates, etc. In order to classify files in one or more disks or directories, select the required directories in the DiskBoss file navigator and press the 'Classify' button.
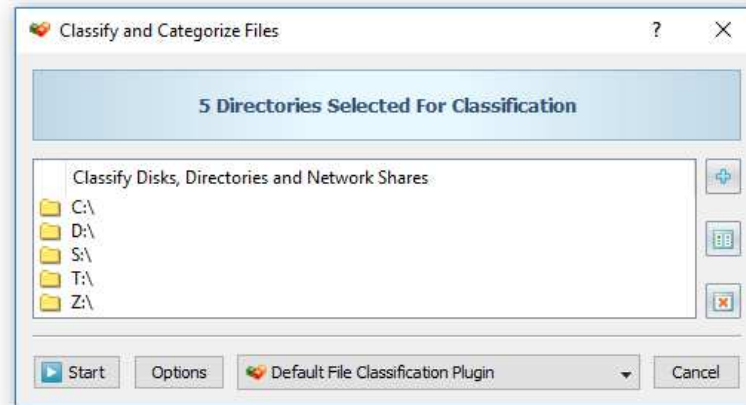

On the 'File Classification Plug-ins' dialog, select an appropriate file classification plug-in and press the 'Classify' button. DiskBoss will classify all the selected disks and directories and show the file classification results dialog displaying a categorized hierarchy of files sorted by the amount of the disk space used by each file category.
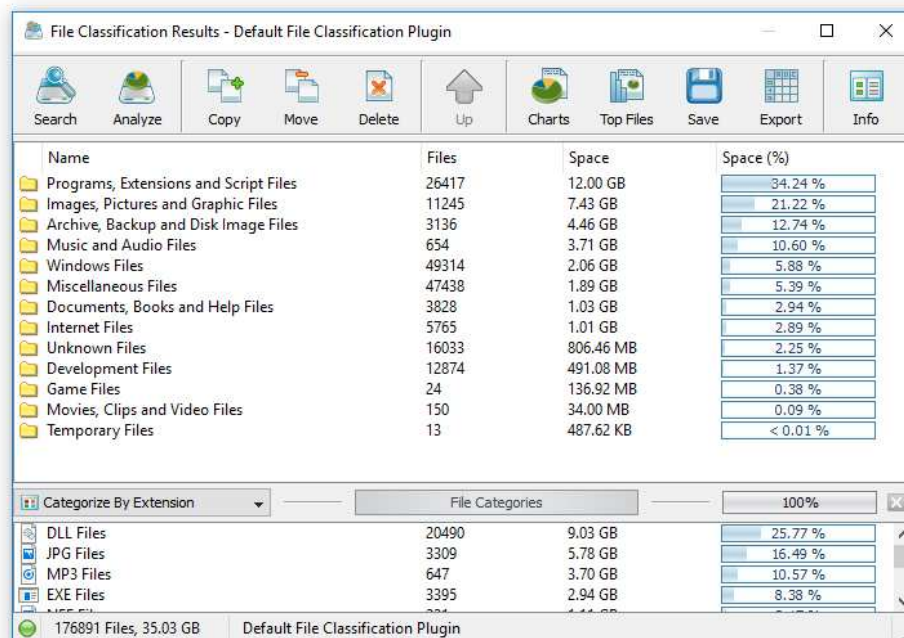
The file classification results dialog allows one to browse the file hierarchy, perform file management operations on files and file categories, generate various types of charts, export HTML, PDF, Excel, XML, text and CSV reports and submit file classification results to an SQL database. In addition, the user is provided with the ability to physically reorganize files on the disk according to the used file classification plug-in. Finally, power users are provided with the ability to design custom plug-ins using an open and easy-to-use XML-Based format.

3   File Filters and File Categories

The top view of the file classification results dialog shows the file categories and classes sorted by the amount of the used disk space and the bottom view shows the second-level file categories and filters, which may be used to filter files and categories displayed in the top view making it very easy to focus on specific types of files or file groups. DiskBoss allows one to categorize file classification results by the file extension, file size, user name, creation, last modification and last access dates.
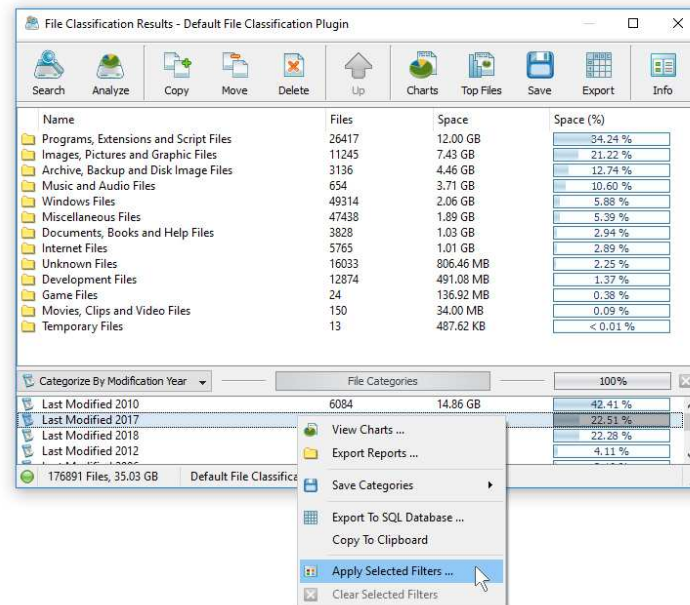

In order to change the current file categorization mode, click on the file categorization combo box located in the top-left corner of the second-level file categories view. In order to apply a second-level file filter, select one or more file categories in the filters view, press the right mouse button and select the 'Apply Selected Filters' menu item. DiskBoss will filter file classification results and display in the top view files matching the selected file filters and file categories containing these files.
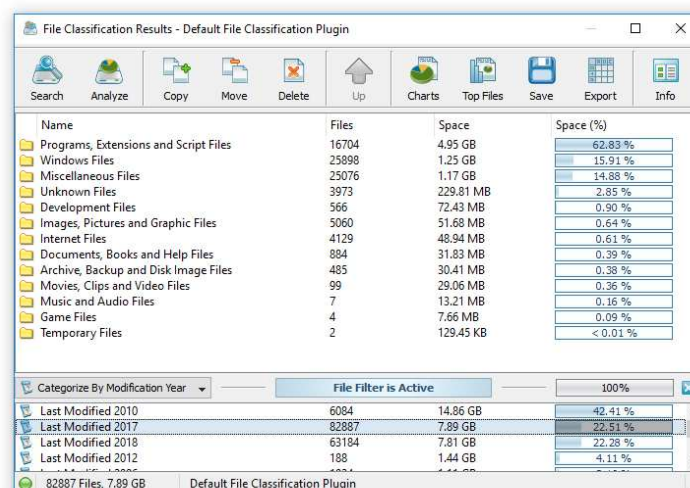
When a file filter is active, the file classification results dialog generates charts and exports HTML, PDF, Excel, XML, text and CSV reports for filtered files only. The user is provided with the ability to browse the filtered file hierarchy and perform file management operations in the usual way. In order to reset the file filter, press the right mouse button over the filters view and select the 'Clear Selected Filters' menu item.

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 file classification 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.
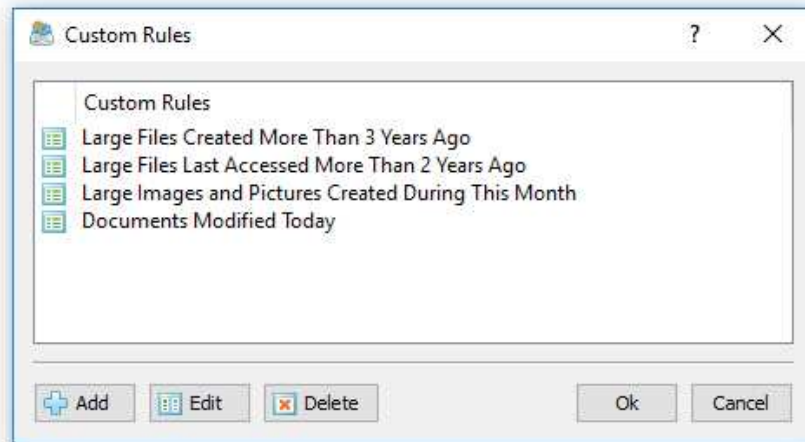

In order to add a new custom file filter, press the 'Add' button, enter a unique filter name and configure one or more file matching rules specifying which types or groups of files to display. In order to edit a custom file filter, select the filter in the custom file filters dialog and press the 'Edit' button.
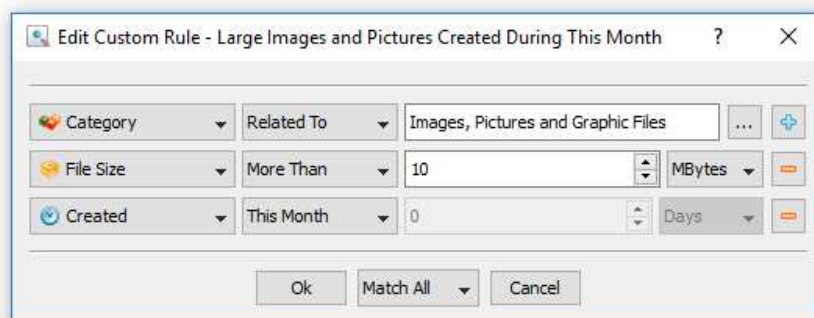
Once finished configuring custom file filters, close the filters dialog and DiskBoss will categorize the currently displayed file classification results and show the list of user-specified custom file filters sorted by the amount of data and the number of files matching the specified file filters. In order to filter the currently displayed file classification 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 file classification 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   File Classification Pie Charts

The DiskBoss file classification utility provides multiple types of statistical pie charts and timeline charts capable of showing the amount of disk space usage and the number of files per file category, file extension, file size, file owner, creation, modification and last access time. In order to open the charts dialog, press the 'Charts' button located on the file classification results dialog and select an appropriate chart type.
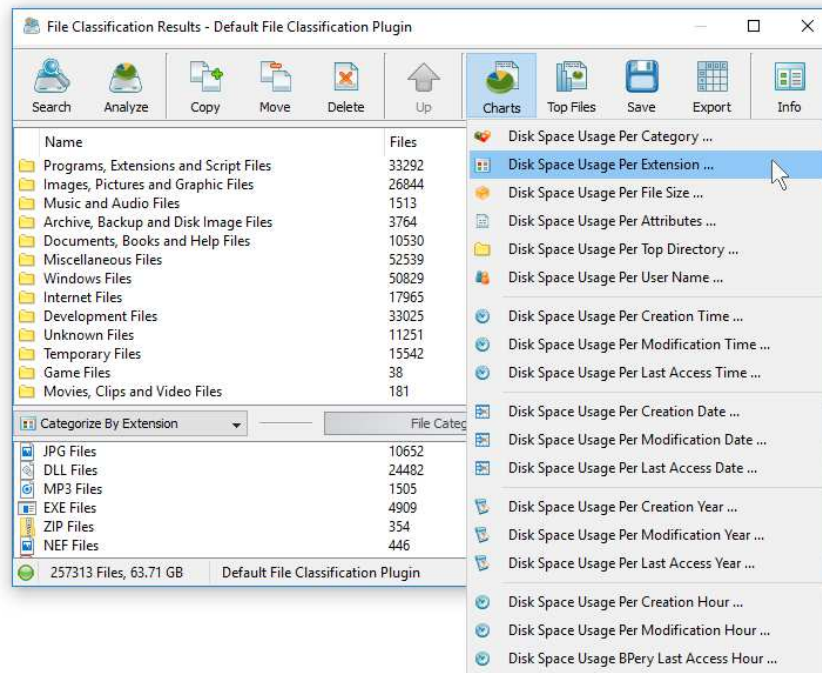
For example, in order to open a chart showing the disk space usage and the number of files per file category, press the 'Charts' button located on the file classification results dialog and select the 'Disk Space Usage Per Category' chart type.
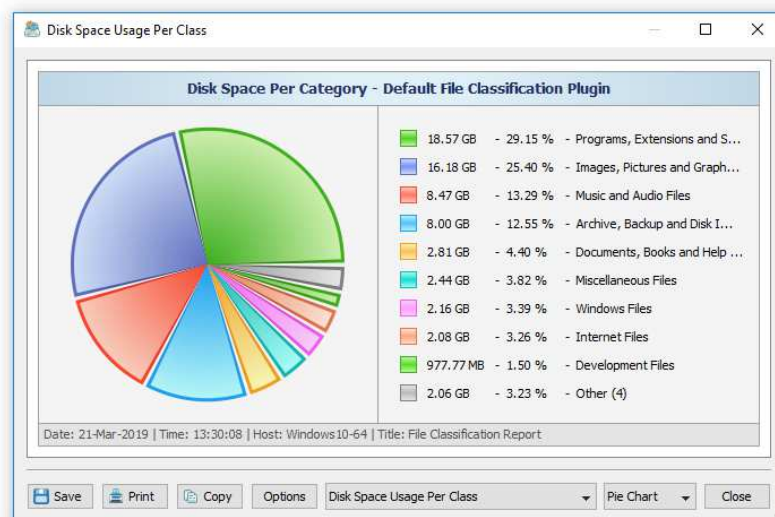
In addition, the charts dialog provides the user with the ability to export the chart data to a number of different formats, export the chart image to the BMP, JPEG and PNG image files, and copy the displayed chart image to the clipboard allowing one to easily integrate DiskBoss charts into user's documents 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 file classification utility is the ability to filter file classification 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 disk space usage per file extension for 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.
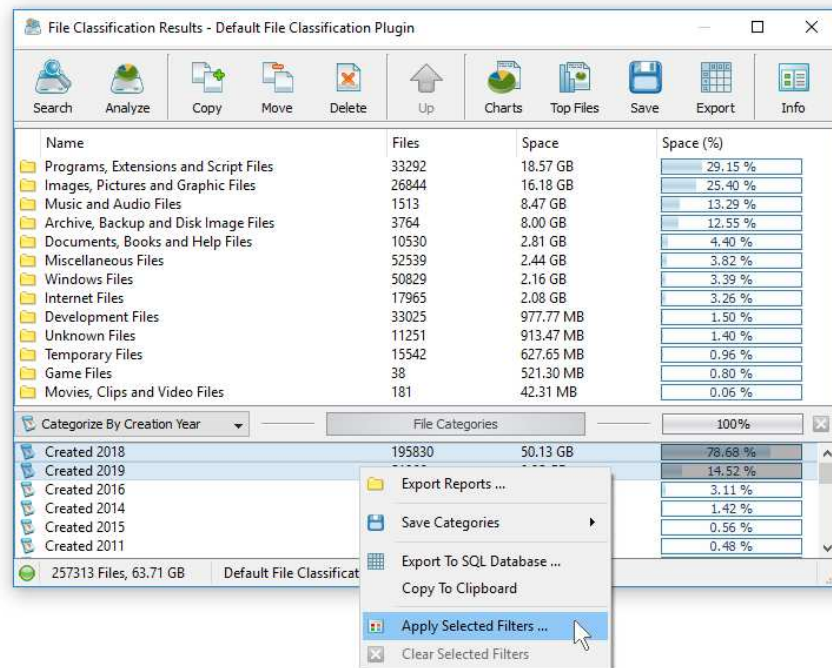
DiskBoss will filter the file classification results and display files that were modified during the selected years. Now, press the 'Charts' button located on the file classification results dialog and select the 'Disk Space Usage Per Extension' menu ite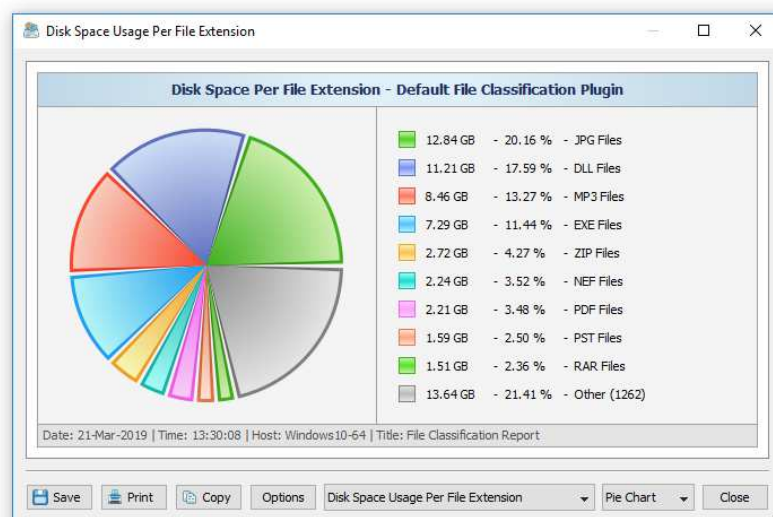m. DiskBoss will categorize the filtered file classification results and display a pie chart showing the amount of disk space usage per file extension for files that were modified during the selected years.

In a similar way, the user can filter file classification 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 disk space usage or the number of files according to user-specific needs.

Another powerful capability is the ability to display the disk space usage timeline charts for all or filtered file classification results. In the simplest case, press the 'Charts' button located on the file classification results dialog and select the 'Disk Space Usage Per Modification Year' menu item.
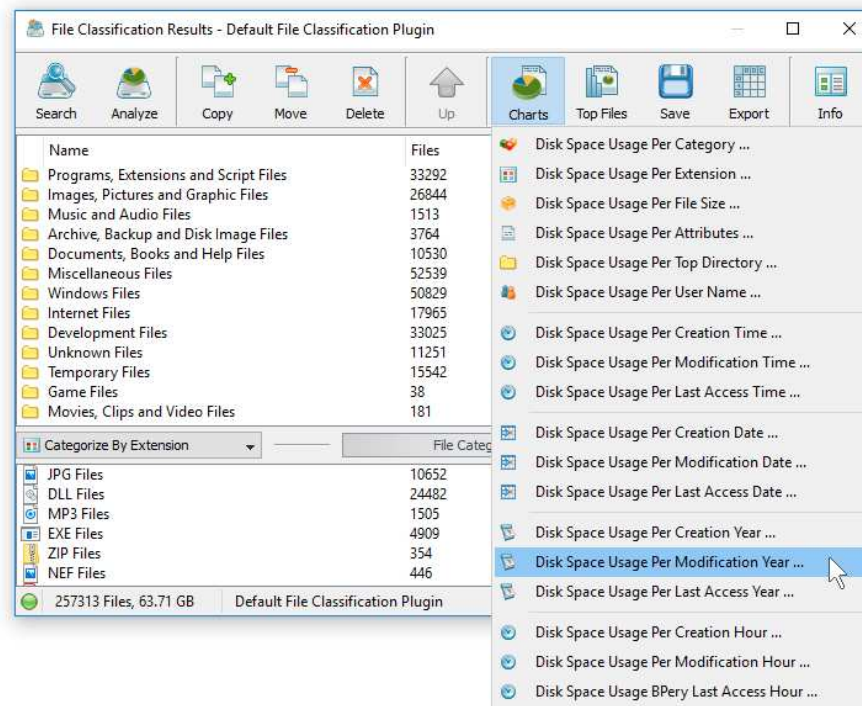
DiskBoss will display a time line chart showing how much data was modified per year. In order to display how many files were modified per year, set the chart mode to 'Number Of Files Per Modification Year'.

In addition, the user is provided with the ability to display timeline charts for filtered file classification results. For example, in order to display how many 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 main toolbar, and select the 'Disk Space Usage Per Creation Year' menu item.

9

5   Analyzing File Classification Results

DiskBoss provides multi-level file classification and disk space analysis capabilities allowing one to analyze file classification results and visualize physical locations of all files related to one or more specific file categories or file classes. These capabilities are especially useful when used in conjunction with second-level file categories and filters providing the user with the ability to analyze categorized and filtered file classification results and making it very easy to reveal locations of very specific files the user is looking for.
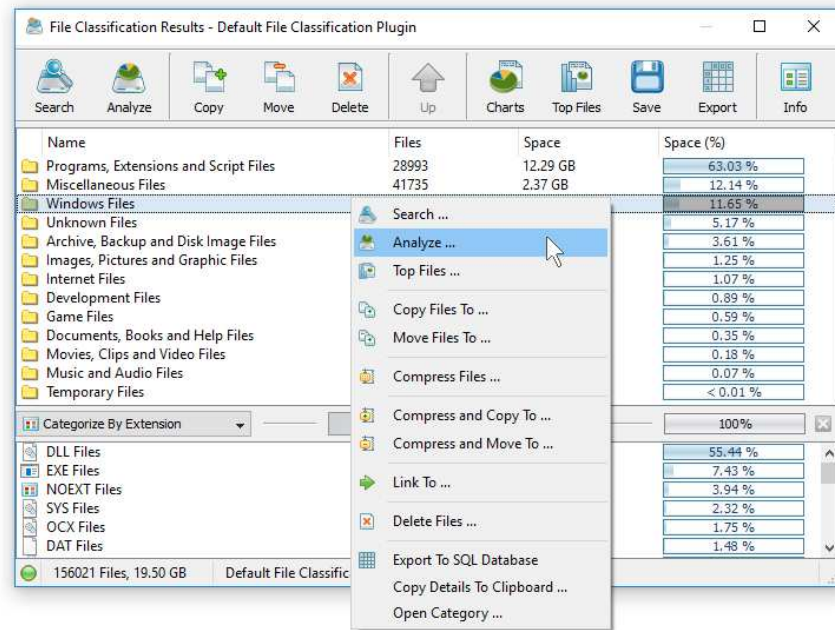
In order to analyze file classifications results, select one or more file classes or categories in the file classification results dialog, press the right mouse button and select the 'Analyze' menu item. DiskBoss will process the selected file categories and classes and display a file system hierarchy showing files related to the selected file categories or classes and the directories containing these files.
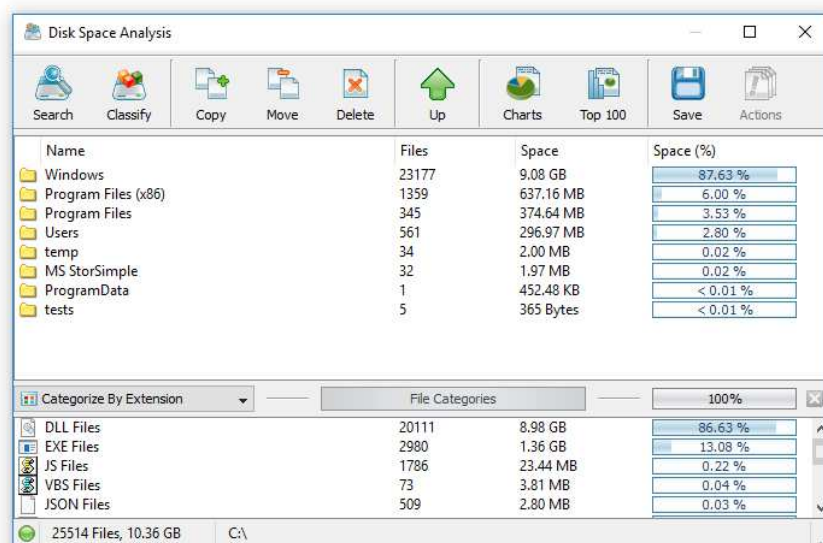
Analysis results are displayed in the standard disk space analysis dialog, which provides the user with the ability to generate pie charts, export HTML, PDF, Excel, XML, text and CSV reports, perform file management operations, etc.

6   Searching Files in File Classification Results

DiskBoss provides the ability to search files in file classification 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, classify one or more disks or directories and press the 'Search' button located on the file classification results dialog.
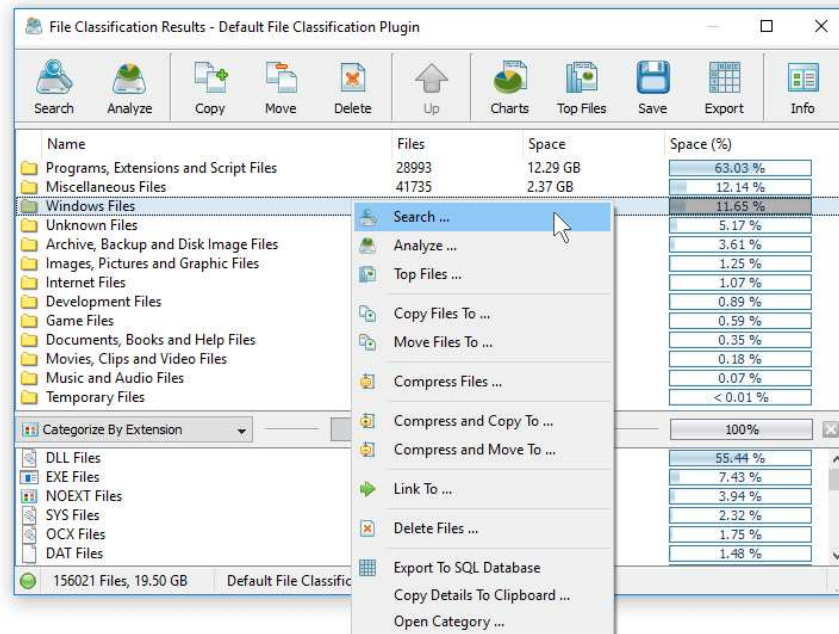
By default, the search operation will be performed on all file classification results. In order to search files in a specific file category, select the category in the file classification results view, press the right mouse button and select the 'Search Files' menu item. DiskBoss will search in the selected categories of files and display the file search results dialog showing a list of files matching the user specified criteria.
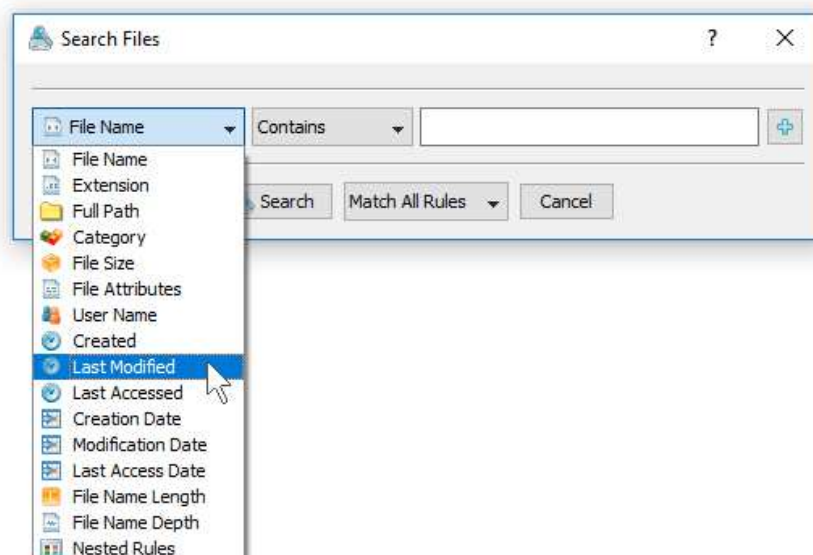
The search results dialog allows one to filter and categorize file search results, display various types of pie charts, copy, move and/or delete files, export file search results to a number of standard report formats including HTML, PDF, Excel, text, CSV and XML. In addition, advanced users are provided with the ability to export file search results to an SQL database.

7   File Management Operations

DiskBoss allows one to copy, move, compress and delete files and related to specific file categories. In order to perform a file operation on one or more file categories, select the required items in the file categories view, press the right mouse button and select an appropriate file management operation.
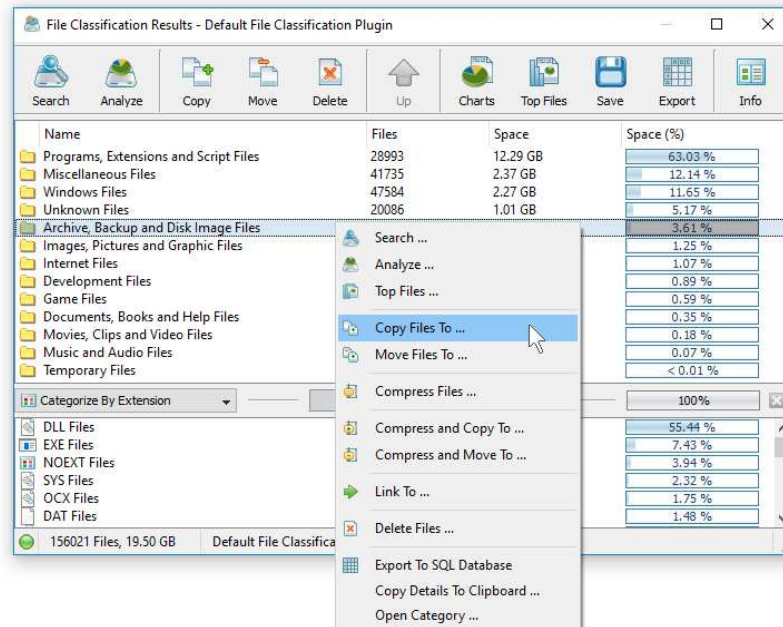
For copy and move operations, enter the destination directory and optionally press the 'Advanced Options' button to open the file copy options dialog. The copy options dialog allows one to control the file overwrite mode, file copy rules, performance tuning options and error handling options.
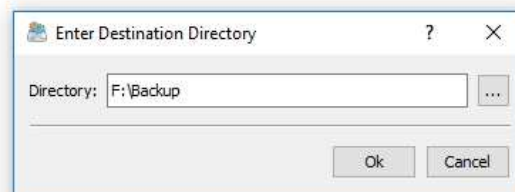
When copying files with an active file filter, DiskBoss automatically adds one or more file matching rules to the 'Rules' tab ensuring that only filtered files are copied to the destination directory. In order to delete files related to one or more file categories, select the required file categories in the categories view, press the right mouse button and select the 'Delete' menu item.
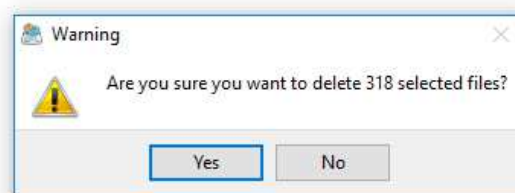
For example, in order to delete all types of document files, classify files in a disk or directory using the default file classification plug-in, select the 'Document Files' file category in the file classification results view, press the right mouse button and select the 'Delete' menu item.

8 Saving File Classification Reports

DiskBoss allows one to save file classification reports into a number of standard formats including HTML, PDF, Excel, XML, text and CSV. In the simplest case, perform a file classification operation and press the 'Save' button located on the file classification results dialog. On the save report dialog, select an appropriate report format, enter a report file name and press the 'Save' button.
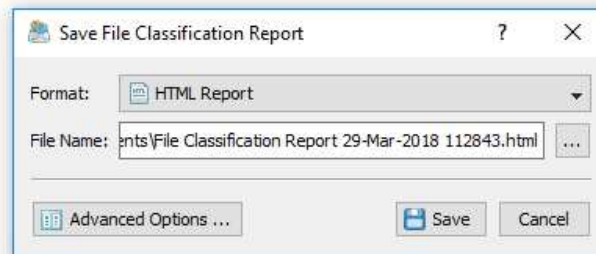
For the HTML, PDF, Excel, text, CSV and XML report formats, the user is provided with the ability to save a short summary report or a longer detailed report, which may be very long for large file systems containing millions of files. By default, DiskBoss will save a short, summary file classification report in the HTML report format, which will include a list of top-level file categories according to the selected file classification plug-in and a list of tables showing the disk space usage and the number of files per file extension, file attribute, user name, etc.

In addition, the user is provided with the ability to save file classification results to the DiskBoss native report format, which preserves all information related to each specific file classification 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 file classification results using external tools such as Microsoft Excel. In order to export file classification results to the Excel report format, perform a file classification operation, press the 'Save' button located on the file classification results dialog, select the 'Excel Summary' report format for a short summary report or the 'Excel Report' format for a detailed file classification report.
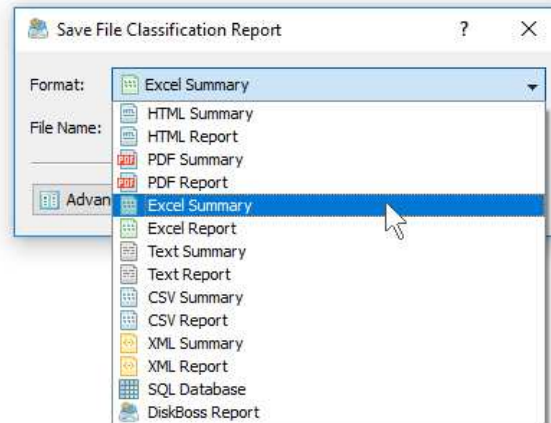
A summary Excel report will include a list of top-level file categories and a number of tables showing the used disk space and the number of files per file extension, file attribute, file creation time, last modification time, user name, etc.
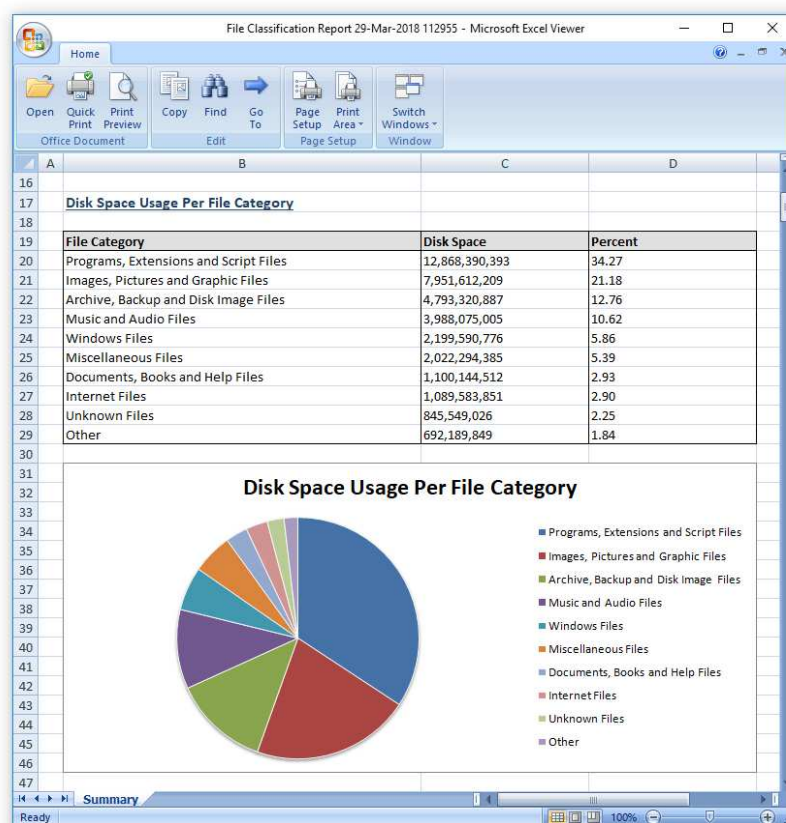
A detailed Excel report will include a list of file categories according to the currently selected file categorization mode and an hierarchy of file groups and file classes sorted by the amount of the used disk space, which may be very long for large file systems containing millions of files. In order to control how many hierarchy levels and how many files per level are exported in the detailed report, press the 'Advanced Options' button located on the 'Save Report' dialog and customize the file classification report according to your specific needs.

One of the most useful ways to export file classification 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 disk space usage and the number of files per file extension, file category, creation time, last modification time, user name, etc. In order to save file classification results to a PDF report file, press the 'Save' button located on the file classification results dialog and select the 'PDF Summary' report format for a short, summary report or the 'PDF Report' format for a detailed file classification report.
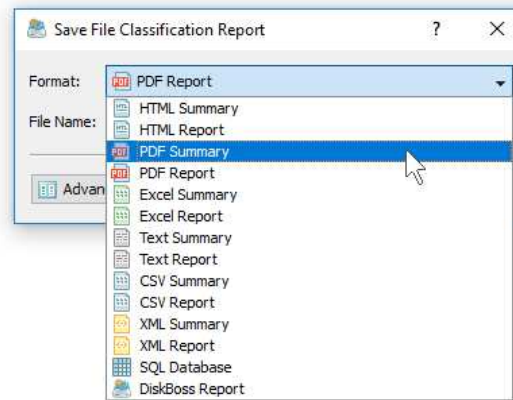
A summary PDF report will include a list of top-level categories of files according to the selected file classification plug-in sorted by the amount of the used disk space followed by a number of pie charts showing the disk space usage and the number of files per file extension, file attribute, file cre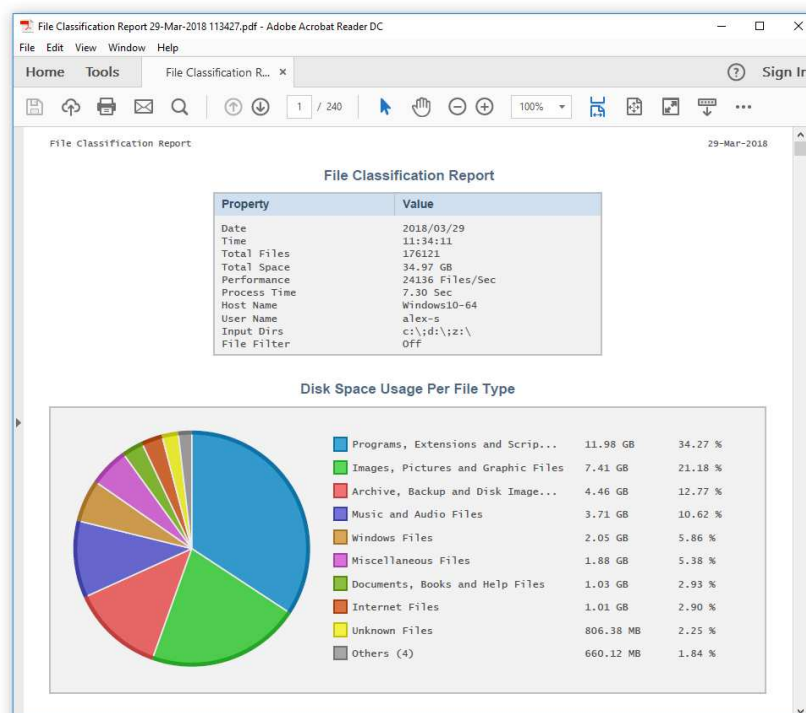ation time, last modification time, user name, etc. A detailed PDF report will include an hierarchy of file groups and file classes sorted according to the used disk space, which may be very long for large file systems containing millions of files.

In addition to the hierarchy of file types sorted by the used disk space, detailed PDF reports include pie charts showing the disk space usage per file category and the number of 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 file classification results by the file extension, the PDF report will display pie charts showing the used disk space and the number of files per file extension.

9   Batch File Classification Reports

DiskBoss allows one to save batches of file classification reports according to the currently selected categories of files with each report showing the disk space usage and the number of files for a corresponding category of files. For example, let's assume that we need to save file classification reports for the last 10 years with each report showing categories of files and the disk space usage for all files that were created during each year.
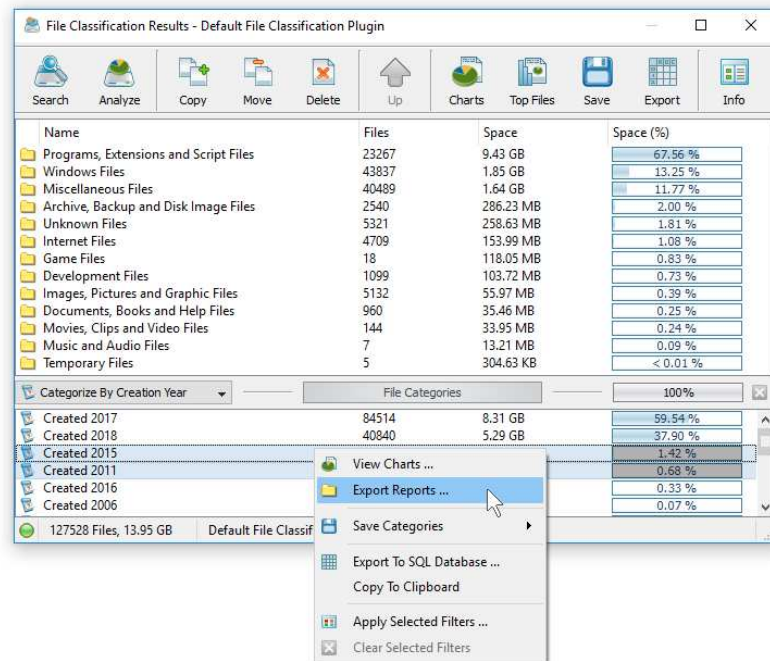
In order to generate such reports, set the file categorization mode to 'Categorize By Creation Year', select all years for which you need to generate analysis reports in the file categories view, press the right mouse button and select the 'Export Reports' menu item. On the 'Export Reports' dialog, select an appropriate report format and report mode, specify a destination directory to export reports to and press the 'Save' button.
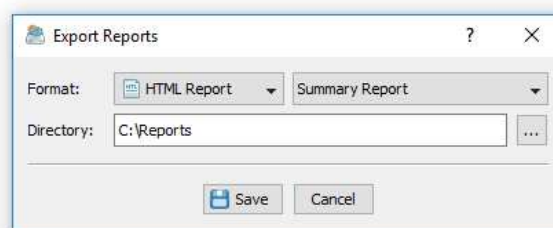
For each selected year, DiskBoss will filter file classification results by the file creation year and save an individual file classification report in the specified destination directory with each report file name and report title containing the file creation year.

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

For example, in order to export individual file classification 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 file classification reports with each report showing categories of files and the disk space usage per 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 categories of files sorted by the amount of used disk space and lists of files for each category sorted by the file size.
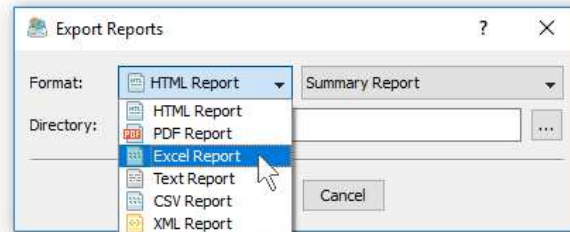
In addition, the batch reports dialog allows one to select one of the following report modes: summary report, compact report or full report. When the summary report mode is selected, each report will include a summary table and lists of file categories sorted by the amount of used disk space. When the compact report mode is selected, each report will include a summary table, lists of file categories and lists of top 100 files per file category. When the full report mode is selected, each report will include a summary table, lists of file categories and full lists of files per category sorted by the file size.
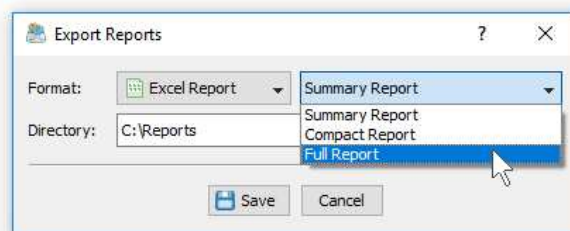
Once the export operation will be completed, the specified destination directory will contain a number of file classification reports according to the selected file categorization mode, selected categories of files and the selected report format.
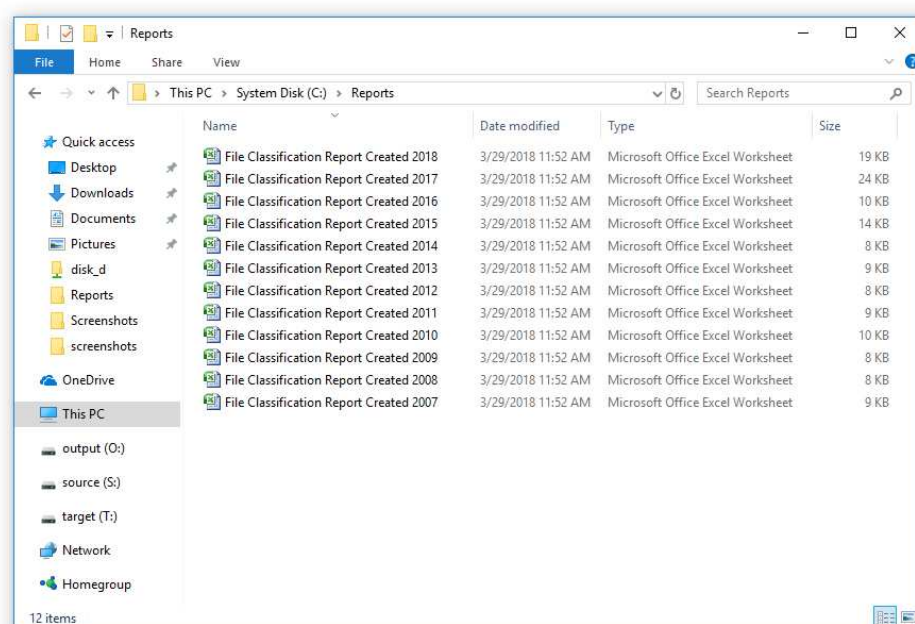
10 Exporting Reports to an SQL Database

IT professionals and storage administrators are provided with the ability to submit file classification reports from multiple servers and desktop computers to a centralized SQL database. A long-term history of file classification reports allows one to gain an in-depth visibility into disk usage trends across the entire enterprise. In order to submit a file classification report to an SQL database, select the 'SQL Database' report format and press the 'Save' button.
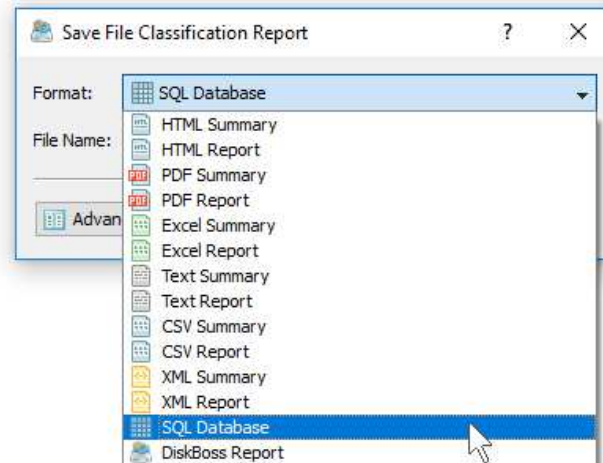
Before exporting reports to an SQL database, open the 'Options' dialog, enable the ODBC interface and specify the name of the data source, database user name and password to use to connect to the database.
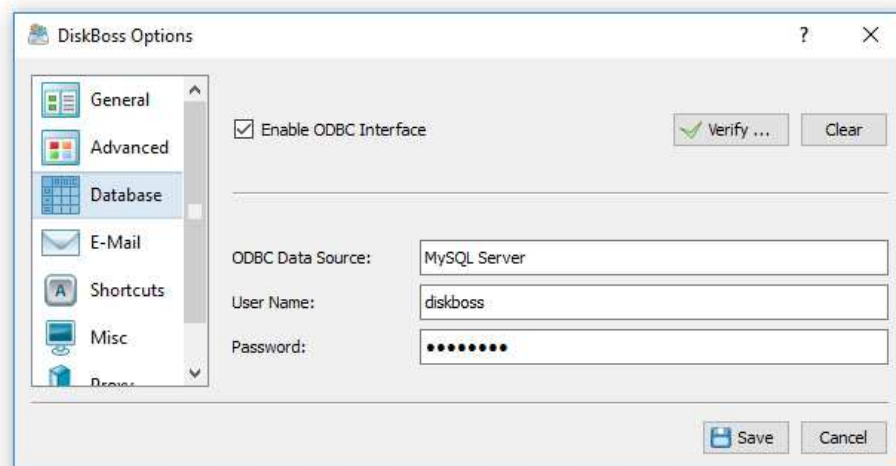
For each report in the database, DiskBoss displays the report date and time, the host name, the input directories, the total number of files and disk space and the report title. To open a previously saved report, just click on the report item in the database reports dialog.

11 Classifying 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 classify 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.
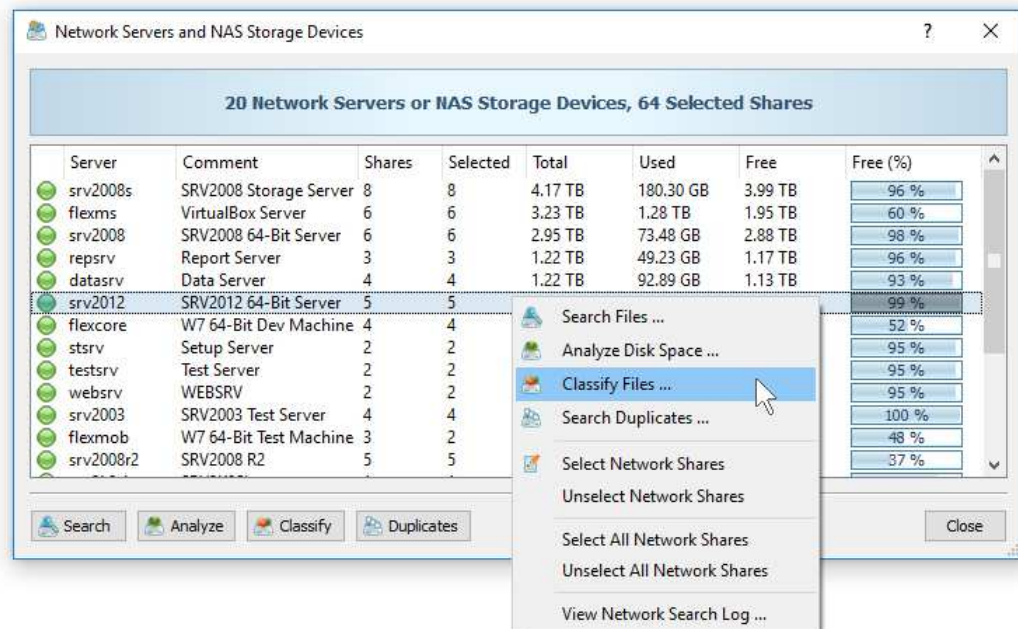
In order to discover all network servers and NAS storage devices on the network, press the 'Network' button located on the main toolbar and wait while DiskBoss will scan the network and show a list of detected network servers and NAS storage devices. In order to classify files in one or more servers or NAS storage devices, select the required servers and NAS storage devices and press the 'Classify' button.
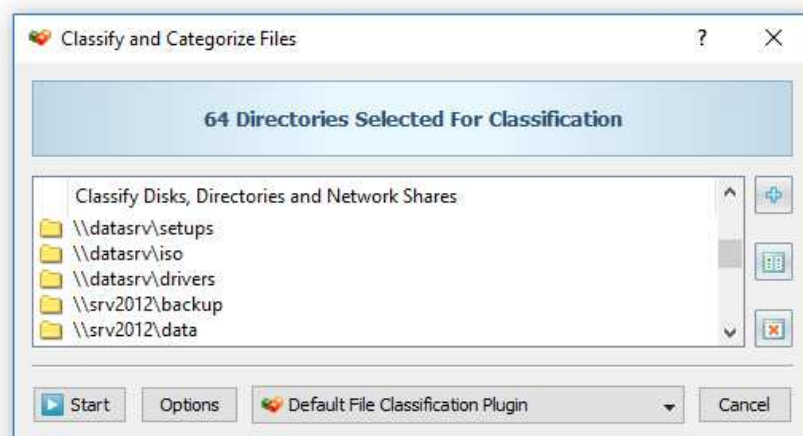
DiskBoss will show all accessible network shares hosted on the selected servers and NAS storage devices allowing one to classify files and save various types of file classification pie charts and reports. In addition, the user is provided with the ability to customize a large number of advanced file classification options allowing one to tune file classification operations for user specific needs and hardware configurations.

12 Batch File Classification Operations

DiskBoss Server and DiskBoss Enterprise provide the ability to execute one or more pre-configured file classification operations on all network servers and NAS storage devices on the network and generate an individual file classification report for each server and NAS storage device. In order to be able to use batch file classification operations, the user needs to pre-configure one or more file classification commands customized to generate file classification reports according to user-specific needs and requirements.
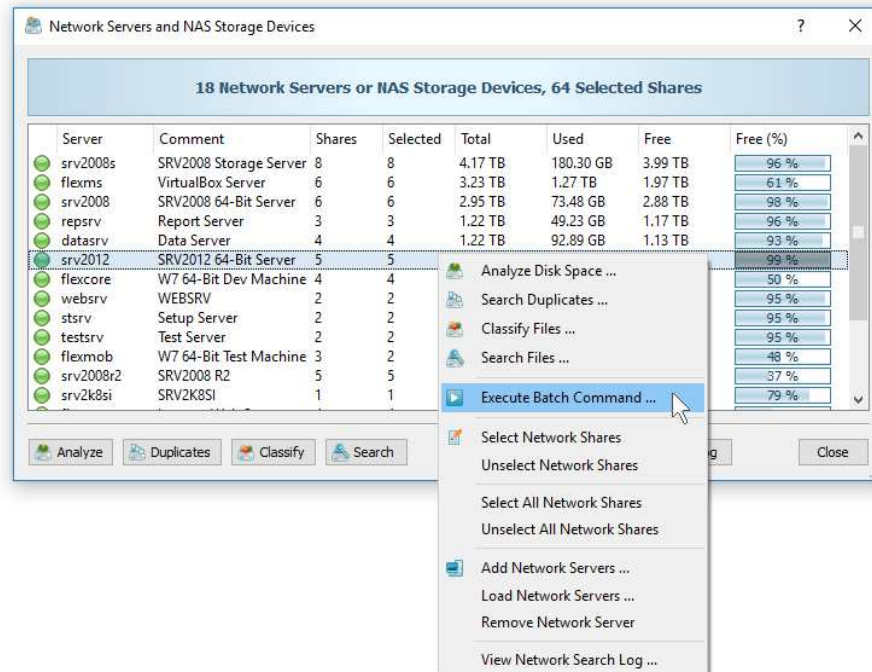
In order to start a batch file classification operation, press the 'Network' button located on the main toolbar, search all servers and NAS storage devices on the network, select one or more servers and NAS storage devices, press the right mouse button and select the 'Execute Batch Command' menu item.
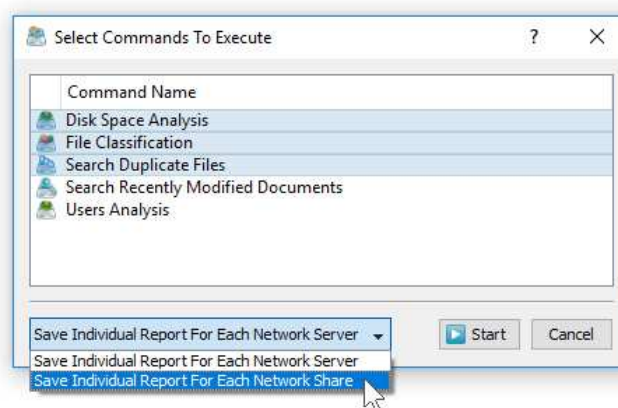
DiskBoss will display a list of pre-configured file classification commands allowing one to select one or more commands to be executed on all selected servers and NAS storage devices. In addition, the user is provided with the ability to select how to save file classification reports - for each server or for each network share. By default, all file classification 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 Classifying Specific File Types and File Groups

On of the most powerful capabilities of DiskBoss is the integration between file search and file classification capabilities allowing one to classify files matching one or more user-specified rules. For example, the user is provided with the ability to classify files with the file size greater than X MB that were last modified more than Y months ago or any other combination of user-specific file matching rules.
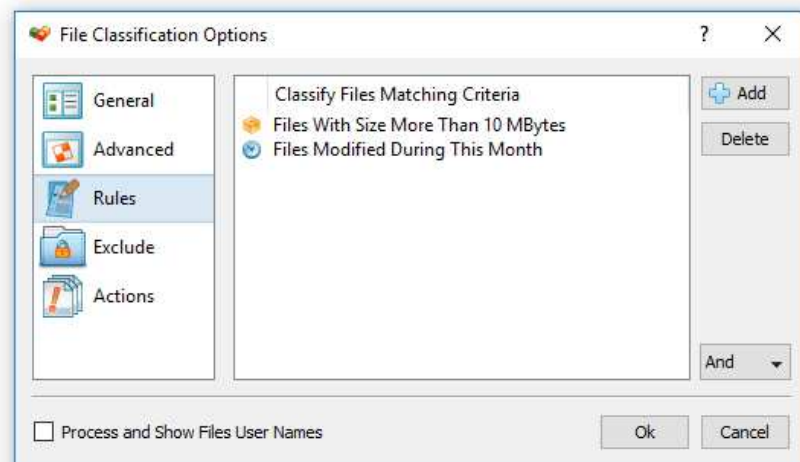
In order to add one or more file matching rules to a file classification operation, open the file classification options dialog, select the 'Rules' tab and press the 'Add' button. Once finished adding file matching rules, select an appropriate rules logic mode (AND/OR) and press the 'Ok' button. During the file classification process, DiskBoss will first evaluate all existing files using the specified file matching rules and classify matching files only. Files not matching the specified rules will be just skipped from the file classification process and not displayed in the file classification results dialog making it very easy to focus on files the user is looking for.
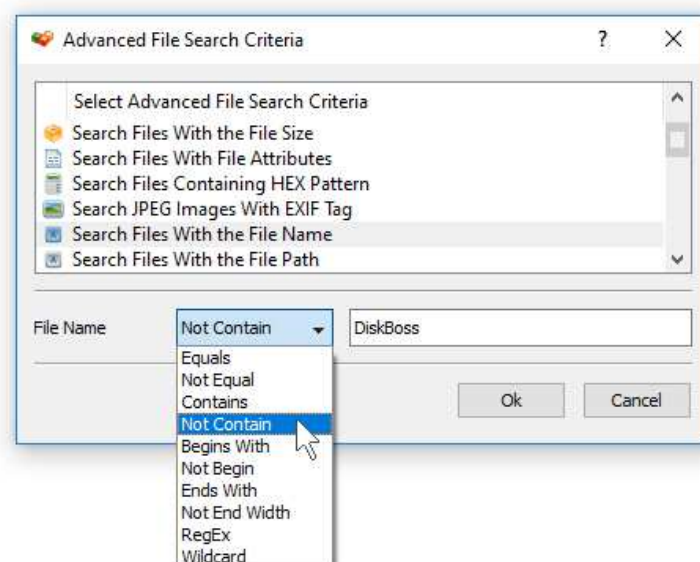
Another option is to exclude specific types of groups of files from the file classification process using one or more negative file matching rules. For example, in order to exclude files with a specific text pattern in the file name, add a file name file matching rule, select the 'Not Contain' rule operator and enter a text pattern to search for.

14 Excluding Subdirectories

Sometimes, it may be required to exclude one or more subdirectories from the file classification process. For example, if you need to classify 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.
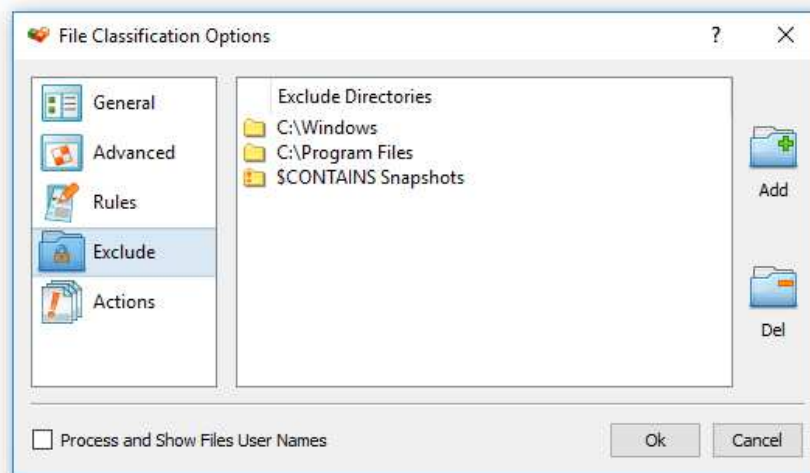
In order to add one or more directories to the exclude list, open the file classification command dialog, press the 'Options' button, select the 'Exclude' tab and press the 'Add' button. All files and subdirectories located in the specified exclude directory will be excluded from the file classification 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 File Organizing Operations

One of the most powerful capabilities of the DiskBoss file classification engine is the ability to physically organize files on the disk according to the selected file classification plug-in. File organizing may be performed manually on file classification results displayed in the file classification results dialog or automatically by selecting an automatic file organizing mode in the file classification options dialog.
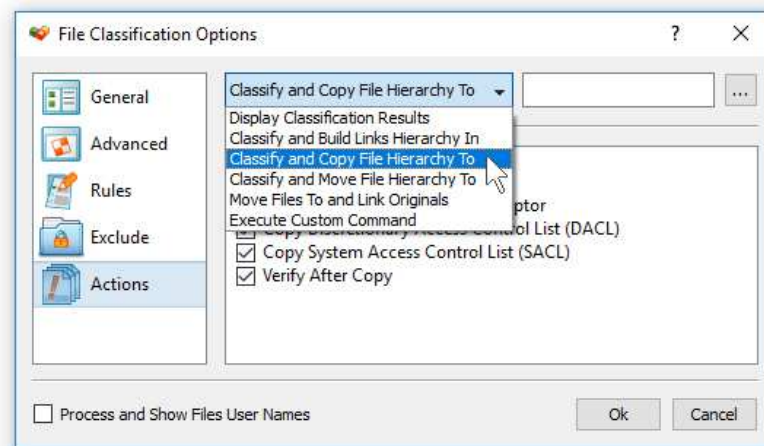
DiskBoss provides the user with the following advanced file organizing operations that may be performed on file classification results:

- **Build Hierarchy of Links** – in this mode DiskBoss will construct an organized hierarchy of directories and populate it with links to original files according to the selected file classification plug-in.
- **Copy File Hierarchy** – in this mode DiskBoss will construct an organized hierarchy of directories and populate it with copies of original files according to the selected file classification plug-in.
- **Move File Hierarchy** – in this mode DiskBoss will construct an organized hierarchy of directories and move all files into that hierarchy according to the selected file classification plug-in.
- **Move Files and Link Originals** – in this mode DiskBoss will construct an organized hierarchy of directories, move all files to that hierarchy according to the selected file classification plug-in and create links instead of original files pointing to files located in the organized file hierarchy.
- **Execute Custom Command** – in this mode DiskBoss will execute a custom command for each file passing the name of the file and the full file category as command line arguments allowing one to implement custom file organizing solutions.

By default, file classification results are displayed in the results dialog and the user is provided with the ability to perform these operations manually by selecting one or more file categories or classes in the file classification results dialog.

In addition, power computer users and IT professionals are provided with the ability to physically reorganize files on the disk by using one of the mentioned automatic file organizing modes. In an automatic file classification mode, DiskBoss will physically reorganize files on the disk according to the selected file classification plug-in without showing the file classification results dialog.

Automatic file classification modes are especially useful in conjunction with the DiskBoss command line utility allowing one to perform automatic file classification operations from shell scripts and batch files and integrate file classification capabilities in other products and solutions. Finally, DiskBoss provides the user with the ability to design custom file classification plugins using an open and easy-to-use XML-Based format making it possible to reorganize files into custom directory structures designed according to user-specific needs.

16 File Classification Plugins

By default, DiskBoss classifies and categorizes files using the 'Default File Classification Plug-in', which is capable of recognizing more than 3,000 of the standard file types and formats and allows one to organize them into a logical file hierarchy. In addition, a number of advanced file classification plug-ins are available for registered users of DiskBoss Ultimate and DiskBoss Server.
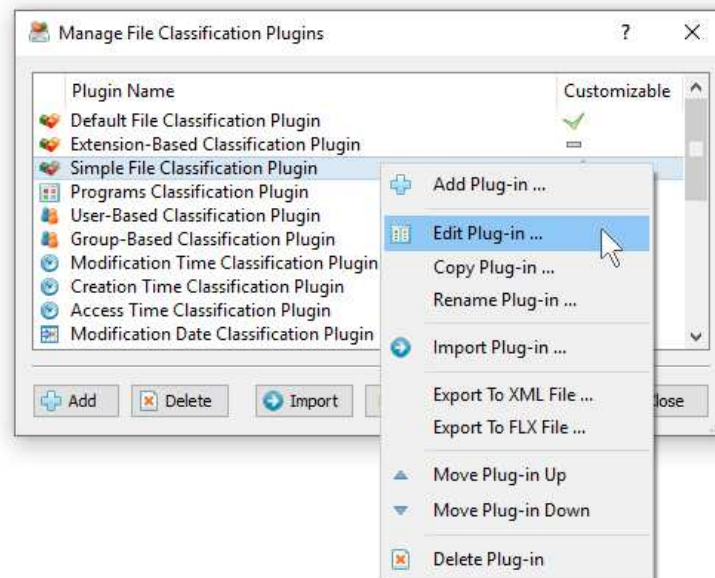
Finally, DiskBoss Ultimate provides the ability to add user-custom file classification plugins and customize the built-in file classification plug-ins allowing one to classify and categorize files according to the user-specific needs and requirements. In order to add a custom file classification plug-in, select the menu 'Tools - Manage Plugins', press the 'Add' button, specify a unique file classification plug-in name and then use the file classification plug-in designer dialog to create your custom file categories and file classes.
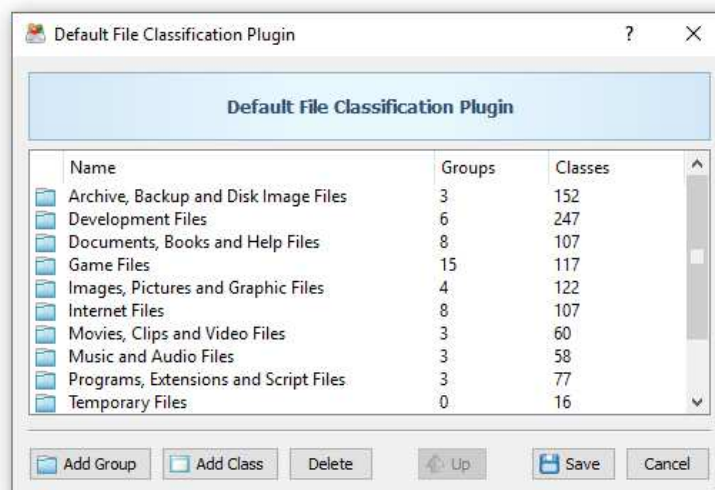
Another option is to customize a built-in file classification plug-in by adding support for one or more additional file type or file categories. In order to customize a built-in file classification plug-in, select the menu 'Tools - Manage Plugins', select the required file classification plug-in item, click the right mouse button and select the 'Copy Plug-in' menu, specify a new plug-in name and then double click on the newly copied plug-in item to open the plug-in copy in the file classification plug-ins designer dialog.

17 Analyzing File Categories Per Host

DiskBoss Server and DiskBoss Enterprise provide the ability to automatically detect all servers and NAS storage devices on the network, classify files in hundreds of servers and/or NAS storage devices via the network, submit reports to the reports database and display charts showing the disk space usage per server or NAS storage device across the entire enterprise.
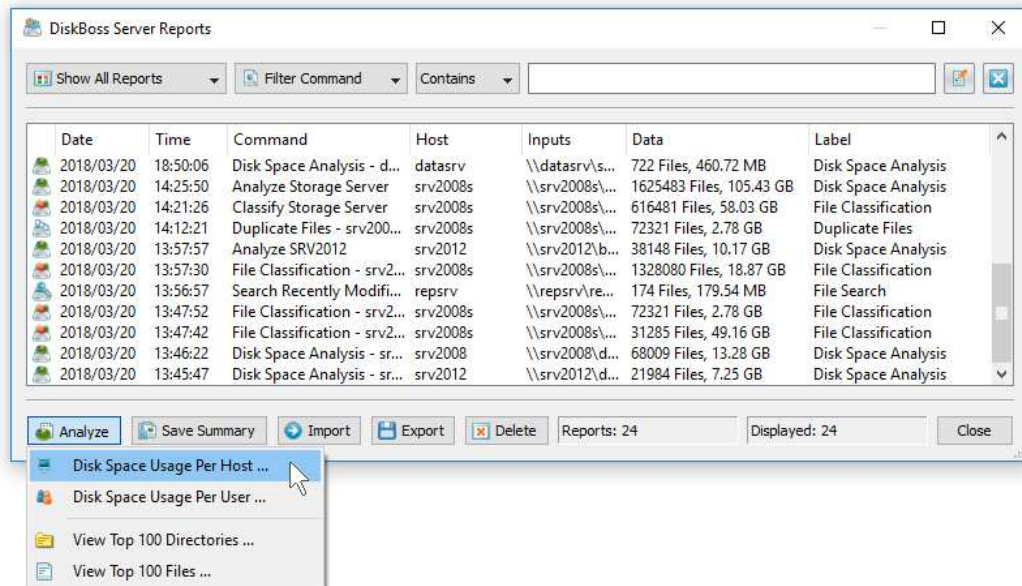
In order to analyze disk space usage per server, perform one or more file classification operations on multiple servers and/or NAS storage devices, open the 'Reports' dialog, press the 'Analyze' button and select the 'Analyze Disk Space Usage Per Host' menu item. DiskBoss will analyze all reports saved in the reports database and display the hosts analysis dialog showing the disk space usage per host.
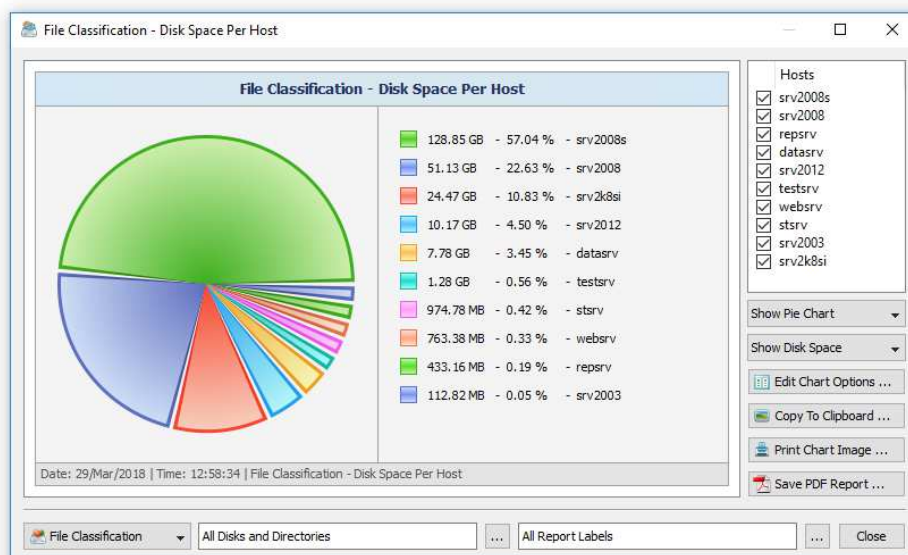
The hosts analysis dialog provides the ability to display pie charts and bars charts showing the disk space usage per server or NAS storage device according to file classification reports saved in the reports database. The user is provided with the ability to select the types of file classification operations 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.

18 Analyzing File Categories Per User

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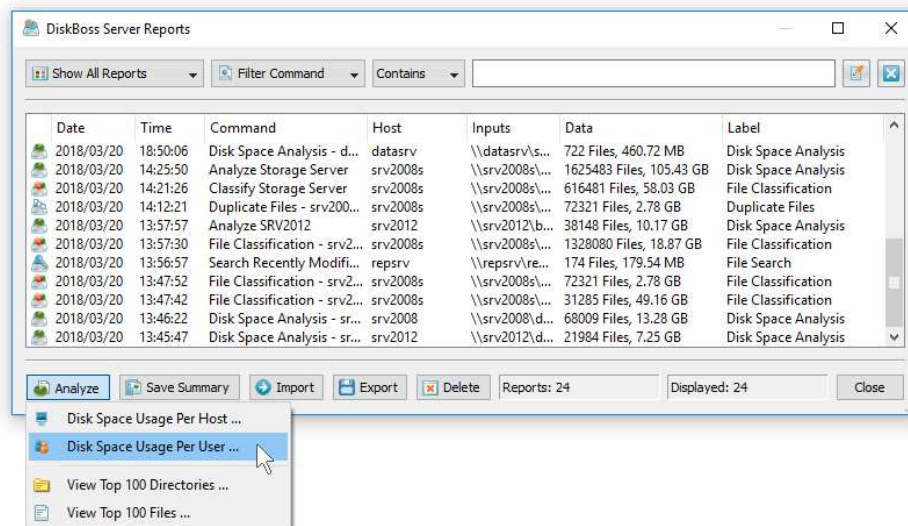

In order to analyze disk space usage per user, perform one or more file classification operations on multiple servers and/or NAS storage devices, open the 'Reports' dialog, press the 'Analyze' button and select the 'Analyze Disk Space Usage Per User' menu item. DiskBoss will analyze all reports saved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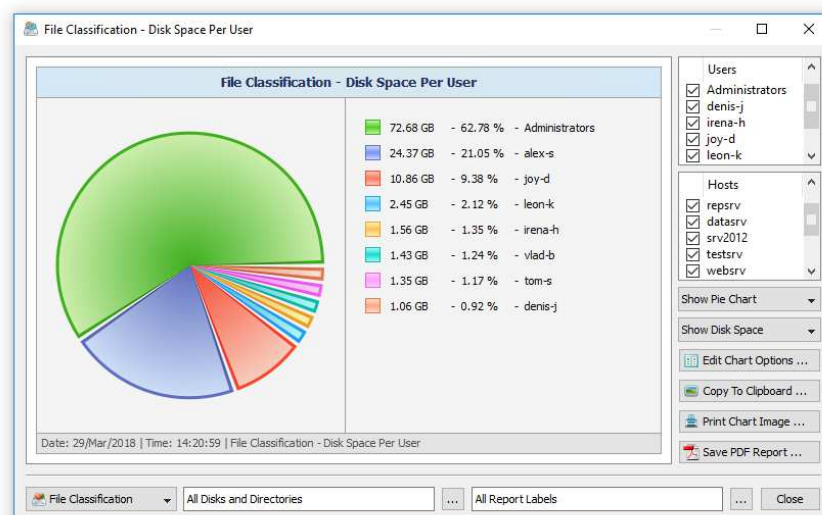

The users analysis dialog provides the ability to display pie charts and bars charts showing the disk space usage per user according to file classification reports saved in the reports database. The user is provided with the ability to select the types of file classification operations and file system 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 the disk space usage per user, the file classification operation should be configured to process and display files user names.

26

19 Advanced File Classification Options

DiskBoss provides a large number of advanced file classification options allowing one to customize and configure file classification operations for user-specific needs and custom hardware and storage configurations.
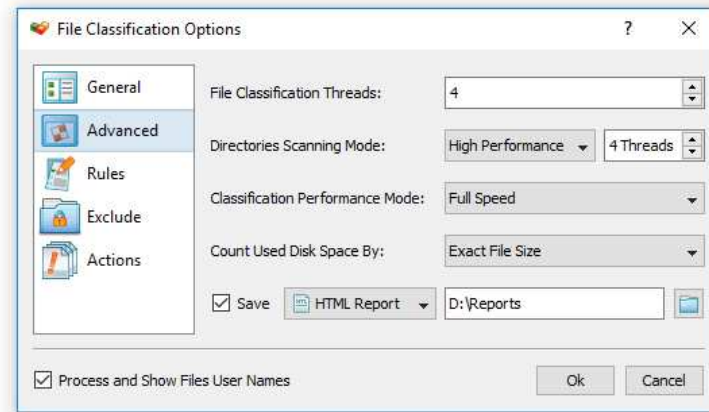
The file classification options dialog allows one to control the file scanning mode, the default report title, advanced reporting options, the default file scanning filter and enable/disable automatic report generation. In addition, the user is provided with the ability to classify specific file types, control performance tuning options, exclude specific directories from the file classification process and physically reorganize files on the disk.

19.1 Pre-Configured File Classification Commands

One of the most powerful and flexible capabilities of DiskBoss is the ability to pre-configure custom file classification operations as user-defined commands and execute such commands in a single mouse click using the DiskBoss GUI application or direct desktop shortcuts.
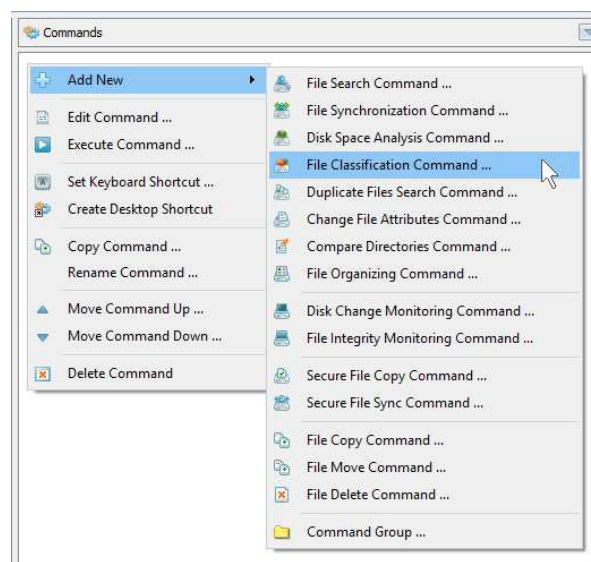
User-defined commands may be managed and executed through the commands dialog or the commands tool pane. In order to add a new command through the commands pane, press the right mouse button over the pane and select the 'Add New - File Classification Command' menu item. In order to execute a previously saved command, just click on the command item in the commands tool pane or create a direct desktop shortcut on the Windows desktop.

20 Classifying Files Using the Command Line Utility

In addition to the GUI application, DiskBoss Ultimate provides a command line utility capable of performing file classification operations and executing user-defined file classification commands from shell scripts and batch files. The DiskBoss command line tool is located in the '<ProductDir>/bin' directory.

Command Line Syntax:

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

This command classifies files in the specified disks, directories or network shares.

diskboss -classify -server <Host Name 1> [... <Host Name X> <Options>]

This command classifies files in all network shares in the specified servers.

diskboss -classify -network [Options]

This command classifies files in all network shares in all servers on the network.

diskboss -execute <User-Defined File Classification Command>

This command executes the specified user-defined file classification command.

Parameters:

-dir <Input Directory 1> [... <Input Directory X>]

This parameter specifies the list of input disks or directories to classify. In order to ensure proper parsing of input directories, directories containing space characters should be double quoted. By default, DiskBoss will generate a combined file classification report showing information about all processed disks, directories and network shares. In order to generate an individual report for each input directory, use the '**-batch**' command line option to enable the batch report generation mode.

-server <Host Name 1> [... <Host Name X>]

This parameter specifies the list of host names or IP addresses of servers or NAS storage devices that should be classified. DiskBoss will enumerate all network shares accessible in the specified servers or NAS storage devices, perform file classification and generate reports if required. By default, DiskBoss will generate a combined file classification report showing information about all processed network shares. In order to generate an individual report for each network share, use the '**-batch**' command line option to enable the batch report generation mode.

-network

In the network-wide file classification mode, DiskBoss will discover servers and NAS storage devices accessible on the network, enumerate all network shares available in all detected servers and NAS storage devices, perform file classification and generate reports if required. By default, DiskBoss will generate a combined file classification report showing information about all processed network shares. In order to generate an individual report for each network share, use the '**-batch**' command line option to enable the batch report generation mode.

Options:

–save_html_report [ReportFileName]

This option saves file classification results to an HTML report file. If no file name is specified, DiskBoss will automatically generate a file name according to the following template: diskboss_[date]_[time].html and save a report file in the user's home directory.

–save_csv_report [ReportFileName]

This option saves file classification results to a CSV report file. If no file name is specified, DiskBoss will automatically generate a file name according to the following template: diskboss_[date]_[time].csv and save a report file in the user's home directory.

–save_text_report [ReportFileName]

This option saves file classification results to a text report file. If no file name is specified, DiskBoss will automatically generate a file name according to the following template: diskboss_[date]_[time].txt and save a report file in the user's home directory.

–save_pdf_report [ReportFileName]

This option saves file classification results to a PDF report file. If no file name is specified, DiskBoss will automatically generate a file name according to the following template: diskboss_[date]_[time].pdf and save a report file in the user's home directory.

–save_xml_report [ReportFileName]

This option saves file classification results to an XML report file. If no file name is specified, DiskBoss will automatically generate a file name according to the following template: diskboss_[date]_[time].xml and save a report file in the user's home directory.

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

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

–workers <Number of Processing Threads>

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

–batch

Enables the batch report generation mode. In the batch report generation mode DiskBoss saves an individual report file for each input disk, directory or network share.

–v

This command shows the product's major version, minor version, revision and build date.

–help

This command shows the command line usage information.

21 File Classification Performance

DiskBoss is optimized for modern multi-core and multi-CPU systems and is capable of classifying 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 file classification operations for user-specific hardware and storage configurations.
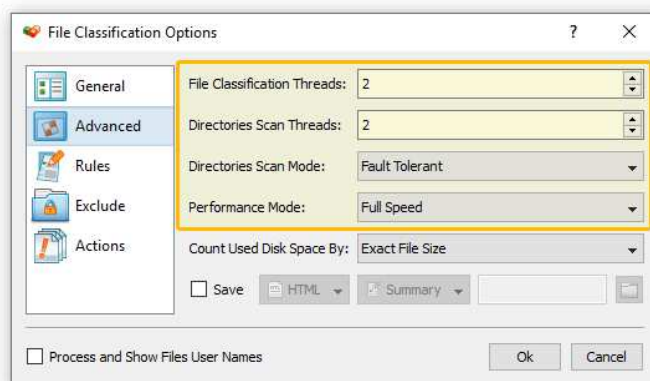

In order to customize the file classification performance optimization options, open the file classification operation dialog, press the 'Options' button and select the 'Advanced' tab. The 'File Classification Threads' option controls how many parallel threads are used to classify files. The 'Directories 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.
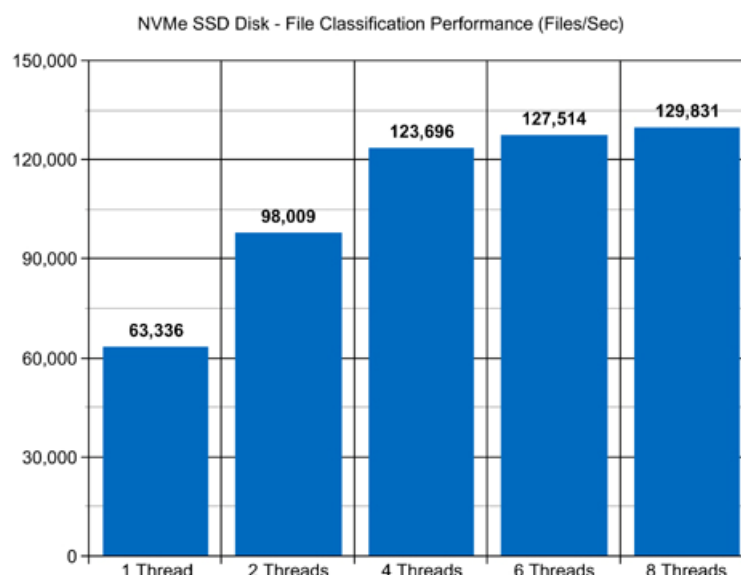

For example, when classifying files stored on a high-speed NVMe SSD disk, DiskBoss reaches up to 63,000 files/sec using a single file classification thread. With two parallel file classification threads, the performance scales up to 98,000 files/sec and with four parallel file classification threads, the performance increases up to 123,000 files/sec showing a very good multi-threaded performance scalability. With six processing threads the file classification performance reaches up to 127,000 files/sec and with eight processing threads the performance increases up to 129,000 files/sec, which makes it possible to classify and categorize 10 millions files within two minutes.

When classifying files stored on regular SATA SSD drives, which are significantly slower than NVMe SSD drives, the performance of the file classification process reaches up to 24,000 files/sec using a single file classification thread and scales up to 69,000 files/sec with four parallel file classification threads. With eight parallel file classification threads, the performance reaches up to 98,000 files/sec, which allows to classify and categorize huge numbers of files relatively fast.

SATA SSD - File Classification Performance (Files/Sec)
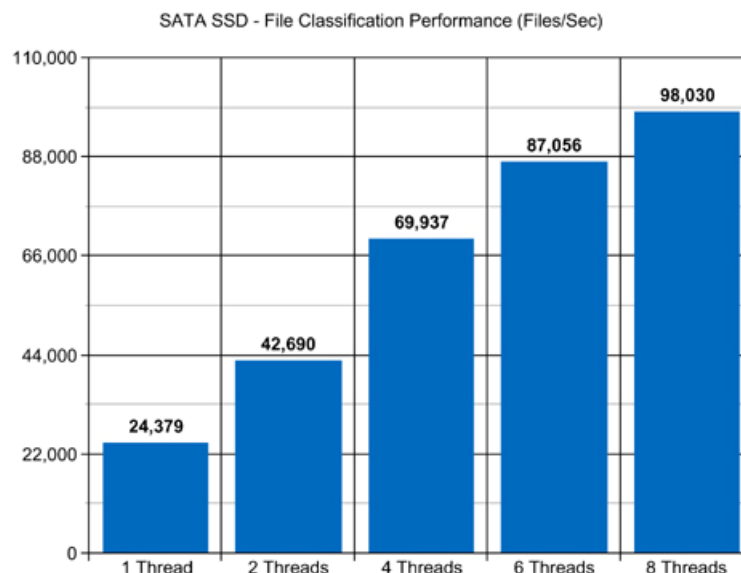
Classifying files stored on a NAS storage device via a network is more complicated because the user needs to take into account the speed and the latency of the network. If the computer, on which DiskBoss is installed, is connected to the NAS storage device via a high-speed, low-latency network, the performance of the file classification operations may reach up to 12,000 files/sec with one file classification thread, scale up to 49,000 files/sec with four parallel file classification threads and increase up to 82,000 files/sec with eight parallel file classification threads.

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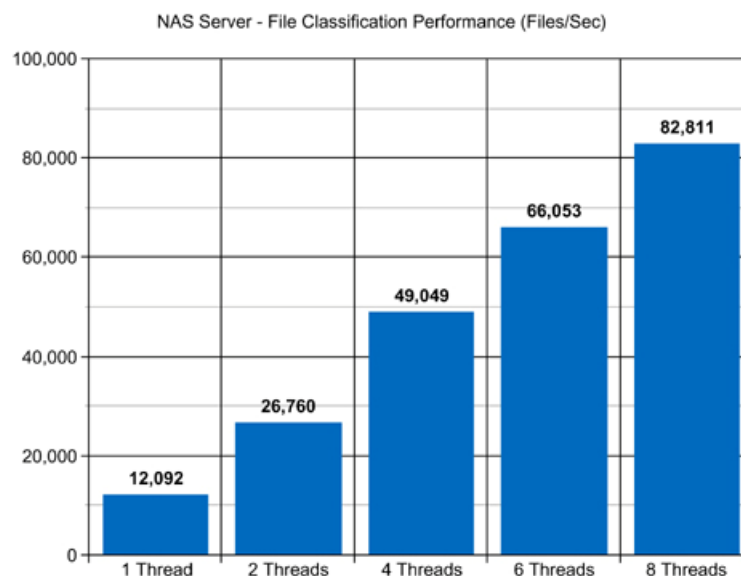
On the other hand, if DiskBoss will need to access network shares via the Internet or via a long-distance, high-latency network, the performance of the file classification operations will be relatively slow. One of the options to increase the performance of the file classification operations in such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

Classifying millions of files stored in one or more large NAS servers may be a very time consuming operation and one of the ways to speed-up the file classification process is to perform the file classification operation via the 2.5 Gigabit Ethernet network. With 2.5 Gigabit Ethernet the performance of the DiskSorter file classification operations continues to scale up to 80,000 Files/Sec with 8 parallel file classification threads.

**2.5 Gigabit File Classification Performance (Files/Sec)**
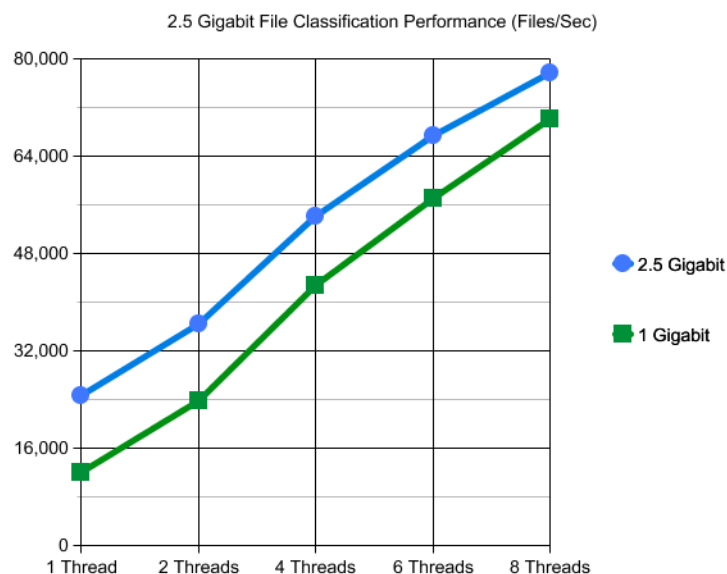
Due to a very wide adoption of laptops and NAS servers with built-in WiFi network interfaces, many users may consider classifying files stored in NAS servers via the wireless network. But, the latency of the wireless network is much higher and therefore it will take much more time to complete the file classification 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 perform the file classification operation via a wired network.

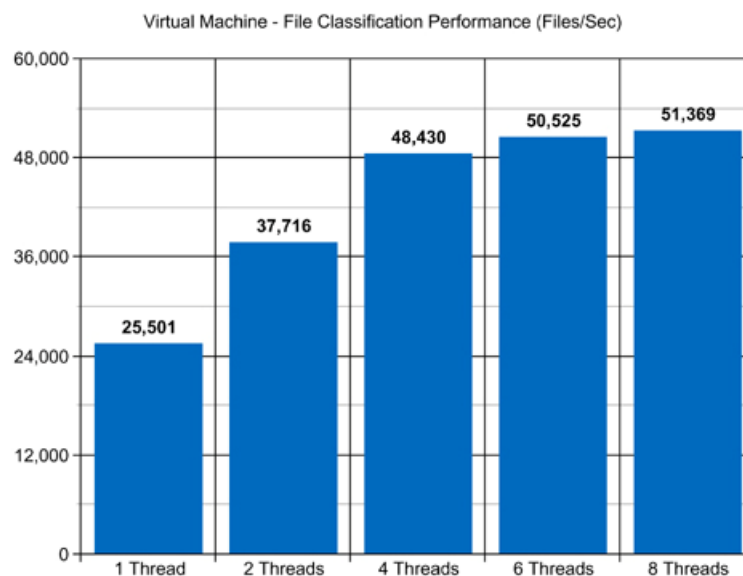**WiFi File Classification Performance (Files/Sec)**
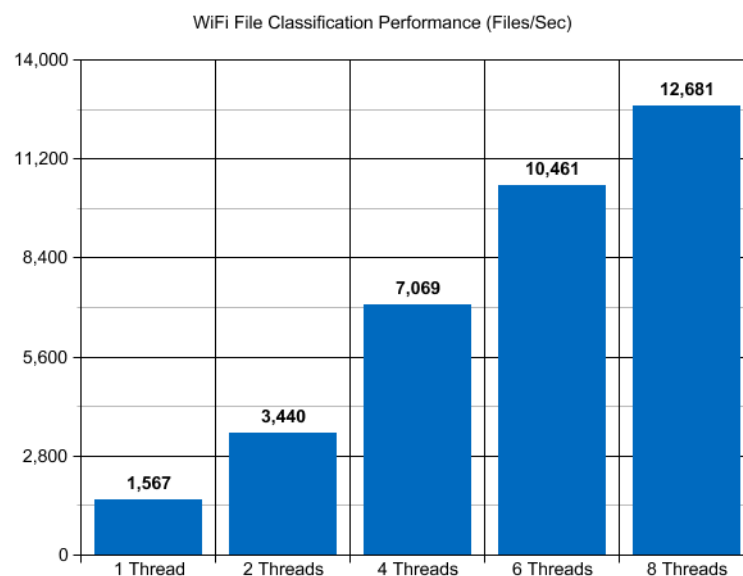
Based on our benchmarks, via a 5 GHz wireless network, DiskSorter reaches up to 1,500 Files/Sec with a single file classification thread and scales up to 12,000 Files/Sec with 8 parallel file classification threads, which is approximately 5 times slower compared to the standard Gigabit Ethernet and 7 times slower when compared to the 2.5 Gigabit Ethernet. So, if the user needs to classify up to 1 million files, it is possible to use a wireless network. On the other hand, if the user needs to classify a large NAS server with many millions of files, 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 classify and categorize files on a USB flash drive in order to free the used disk space. When classifying files stored on a USB flash drive, DiskBoss can reach up to 15,000 files/sec with a single file classification thread. With two parallel file classification threads, the performance increases up to 18,000 files/sec, which is good, but more than two file classification threads will slightly degrade the performance of the file classification operations.

USB Flash Disk - File Classification Performance (Files/Sec)
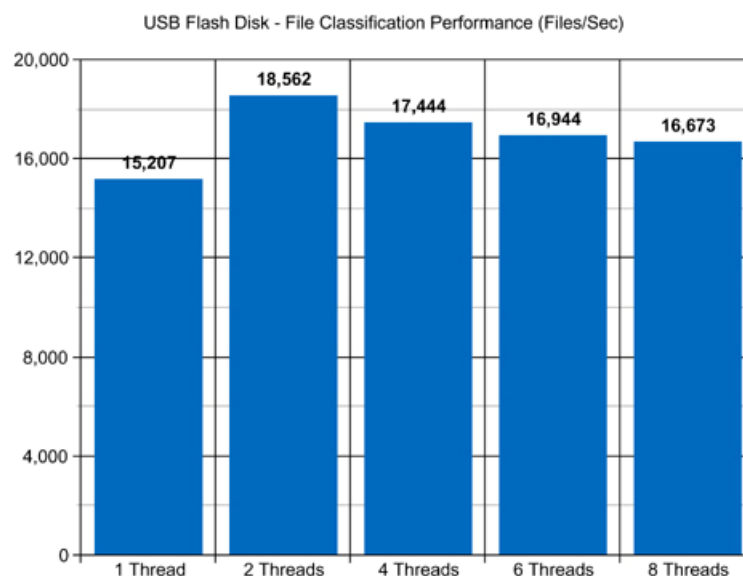
Today, modern IT environments widely deploy virtual servers and/or virtual workstations. Most of the popular virtualization platforms provide a high level of performance, but anyway some performance degradations are inevitable when a file classification operation is executed on a guest virtual machine compared to the same file classification operation executed directly on the host computer.

Virtual Machine - File Classification Performance (Files/Sec)

For example, when a virtual machine with 4 virtual CPUs is stored on an NVMe SSD disk and classifying files stored on a virtual local disk drive, which is physically stored on the same NVMe SSD disk, the performance of the file classification operations reaches up to 25,000 files/sec using a single file classification thread. With two parallel file classification threads, the performance of the file classification operations scales up to 37,000 files/sec and with four parallel file classification threads, the performance of the file classification operations increases up to 48,000 files allowing one to classify and categorize huge numbers of files relatively fast.