



The screenshot displays the 'earlyCASE Analysis' web application interface. At the top, it shows the 'Login' area with the earlyCASE logo and the tagline 'where eDiscovery STARTS'. The main header includes 'Early Case Assessment and Analytics' and the date 'Monday, August 17, 2009 12:45:41 PM'. The interface is organized into several sections:

- Client Information:** Client Name: sjsi, Matter #/Name: #46464.09, Attorney: sjsi sjsi, Date Loaded: 8/17/2009 10:30:55 AM, Amount of Data: 18.68 (Mb Analyzed) / 276 (Total Files).
- Budgets / Costs:** Includes 'Assumptions', 'Budgets and Timelines', and '26(f) Report*'. Sub-sections include 'Other Costs and Expenses', 'Backup Tapes / Costs', 'Extrapolated Costs', and 'Attorney Review Costs'.
- Summary / General:** Includes 'Sizes and Counts (Expanded)', 'Estimated Page Counts', 'Custodian Summary', 'Matter', 'Warnings and Errors', 'Charts and Graphs', and 'Filter Summary*'. There is also an 'Edit Data' section with buttons for 'Matter', 'Assumptions', 'File Types', 'Page Assumptions', and 'Filters*'. An 'Additional Reports*' button is also present.
- Files:** A list of report categories including 'Top 25 File Types', 'Office File Types', 'File Type Summary', 'DeDup - File Type Summary*', 'Known System Files**', 'Generally "Included" File Types', 'Generally "Excluded" File Types', 'Unknown File Types', 'Images Summary (tiff, jpg)', 'File Author Summary', 'File Type Pivot', and 'File Inventory with MD5*'. A 'Duplicate Files and Emails' section includes 'Duplicates (Custodian)*', 'Duplicates (Global)*', 'Duplicate Files (Loose)*', and 'Duplicate Emails*'. There is also a 'Collection / Images' section with 'Data Collection Summary*', 'EnCase, AFF, DD Image Info*', and 'Compare Run 1 and Run 2'.
- Email:** Includes 'Email "To:" Summary', 'Email "From:" Summary', 'Conversation Summary', 'Attachment Summary', 'PST's and NSF's Analyzed', and 'Other Mail Containers'. A 'Containers' section includes 'Summary of Container Files' and 'Container Files (Details)*'.
- Dates:** Includes 'Top 25 File Dates', 'File Date Summary', 'File Dates Pivot', and 'Email Dates Summary'. A 'Folders / Media' section includes 'Folders with > 100 Files', 'Folder Inventory (ALL)', and 'Folders with System Files**'. An 'Exports' section includes 'Export Email and file tables to CSV*', 'Microsoft Excel Export*', '.XML Export of Data (2007)*', 'Detail List of System Files**', 'Detail List of Duplicate Files*', and 'Filter Scripts- Dates, NIST...*'.

At the bottom right, there is a NIST MD5# logo and technical information: TCP/IP Address: 24.99.218.58, NSRL Version: 2.44, Schema Version: .506j. A note at the bottom right states '* = Professional Runs Only'.

The following is a brief introduction to the reports available in earlyCASE.

Budgets / Costs

1. **Assumptions** – Outlines the assumptions about the data, the cost to process & review it. The assumptions directly impact many of the reports. If you are going to conduct what if scenarios with different sets of assumptions, make sure you print the assumptions report and attach it to the reports generated based on those assumptions.
2. **Budgets & Timelines** - This report contains the summary of the amounts of data, applies the assumption to this data and calculates the budgets and time required to handle the sample analyzed, but also extrapolates what the larger population of data will look like based on the sample. This report is an excellent report to use when changing assumptions and effect that those

Reports available in earlyCASE

(v506)

changed assumption have on the project budget. This report should be used in conjunction with the sizes and counts expanded report.

3. **26(f) report** – This report is a Rollup of 6 of the other reports in earlyCASE. The intent of this report is to provide a snapshot of the data in a matter, the assumptions, etc. – all without disclosing anything that would be privileged or confidential. As such this report is careful to only include summary type information useful in communicating and negotiating with the other side in a matter about the ESI / eDiscovery.
4. **Other Costs and Expenses** – This report takes the cost assumptions and applies it to the data that was analyzed to estimate the costs in this area for both the analyzed sample as well as the larger (extrapolated) set of data. These costs include items like data collection, litigation support, project management, hosting costs, etc.
5. **Backup Tape Costs** – Backup tape related costs are reported here and are NOT rolled into the budget for a project as generally backup tape is treated as inaccessible information. This report identifies these costs based on the assumptions provided. If backup tapes are to be handled in a matter these costs need to be manually added into your budget.
6. **Extrapolated Costs** – This report extrapolates the size and counts of the larger data set size based on the sample that was analyzed. It applies the filter cull assumptions provided and extrapolates the anticipated costs pre and post filtering for the larger set of data anticipated.
7. **Attorney Review Cost** – This report shows the costs anticipated on reviewing the extrapolated set of data. This report uses the assumption provided to arrive at cost of attorney review as well as the number of full time attorneys that would be required to complete the review in the time frame specified.

Summary / General

8. **Sizes and Counts Expanded** – This report provides a summary view of how the data expanded both in size and counts. It breaks this down by run and by piece of media. This report also identifies the number and percentage of what was on that media that was a duplicates of things already processed. Included on this report is a summary of the containers, what was extracted from the containers and a anticipated size of what you would pay to process and review based on the actual data you analyzed.
9. **Estimated Page Counts** – This report uses the page assumptions table to estimate the number of pages that will be created if these documents are tiffed as a part of producing them. This report is at best a guess as no two document populations are the same. To change the assumptions as to the average number of pages per document, MB or GB edit the appropriate file type entry in the page assumptions table.
10. **Custodian Summary** – This report will show you the number of files and emails per custodian. In addition to counts this report will also show you the expanded size per custodian.
11. **Matter** – This report is the information you entered on the first screen when you created a new matter. This information can be updated and does affect the other reports. The amount of anticipated raw data (ESI) and how long you have to review it are pivotal to the other reports.
12. **Warnings & Errors** – Any warnings of errors encountered in the processing of the ESI by earlyCASE will be reported here. It is a good idea to check this report. Password protected files, corrupt files, unsupported file compression, all are reported here.
13. **Charts and Graphs** – earlyCASE provides charts and graphs of the file types, containers and other key pieces of information displayed in chart and graph form.

14. **Filter Summary** – This report uses the filter record (configurable) to show you the number of documents and emails that can be filtered out prior to processing as well as during processing based on the filter rule you defined. This report will also show you the number of known system files in the collection as well as file types which are generally excluded. This report requires that you have installed the NIST database on the machine that you are running the reports from. Run this report prior to running the “Filter Scripts – Dates, NIST...” which creates the scripts to do the filtering of the collection.

Files

15. **Top 25 File Types** – Generally the most common file types seen in a population of documents make up the majority of the count as well as size. This report identifies the most frequent file types, the count, the aggregate size along with totals for what the top 25 file type represent in count, size and percentage of the total document population. It is not uncommon to see the top 25 file types represent over 85% of the total population and it a good starting place to understand both the loose files as well as the files types that were attached to emails.
16. **Office File Types** – This report looks at just the “Common” office types of documents against the larger population of documents and provides counts, and aggregate sizes by file type for the normal Office Document types. This report should be used in conjunction with the Top 25 file types report to form a picture of what files are predominant and meaningful.
17. **File Type Summary** – This report will show you all of the file types seen in this ESI (including email attachments), a description of the file type, how many there are, the aggregate size of all of the files of this type, and the date range (last modified) of these files. The report is sorted by the count with the largest (highest #) first. This report should also be used in conjunction with the office files report.
18. **De-Dup – File Type Summary - (Professional Run Required)** This report is very similar to the file type summary – BUT with shows the counts and sizes based on a global deduplication of the documents. It will show you all of the file types seen in this ESI (including email attachments), a description of the file type, the DE Duplicated count of how many there are, the aggregate size of the DE Duplicated files of this type, and the date range (last modified) of these files. The report is sorted by the count with the largest (highest #) first. This report also includes totals at the report of the report – Count and Size.
19. **Known System Files** –
20. **Generally Included File Types** – This report compares the file types observed in the data analyzed against a database (Filetypes) to summarize what the sizes and counts of the generally reviewed file types would look like. This differs from the top 25 file types and office file types in that you can customize the file types which show on this report
21. **Generally Excluded File Types** - This report compares the file types observed in the data analyzed against a database (Filetypes) to summarize what the sizes and counts of file types which are generally NOT processed or reviewed. This differs from the top 25 file types and office file types in that you can customize the file types which show on this report
22. **Unknown File Types** – This report summarizes the file types, counts and sizes for any file type which is unknown / not defined in the FileTypes table. This report should be checked to see if there are file types which the generally included and generally excluded file types reports did not pickup.

23. **Images Summary (tiff, jpg)** – This report shows the types of images (pictures, etc) along with the aggregate size and date range. Depending on the matter pictures (images) may not be useful to process, OCR, and review. This report identifies the types and counts to assist in the budget and planning process of dealing with them if needed.
24. **File Author Summary** – This report relies on the metadata in the documents and will summarize by document author how many documents they authored, the aggregate size and the date range. For this report to be useful the applications installed / configured on the machine which generated the documents must have been set up with the users name and not some generic information. If you use this report make sure you validate it, as many machines do not have this information configured and hence it is not in the document metadata.
25. **File Type Pivot** – Pivot tables allow you to select / deselect filetypes which are then charted or visible in table form. This is useful in seeing the impact size and count of filtering file types from the population. This is an advanced Microsoft Excel Function and is well worth learning more about.
26. **File Inventory with MD5** – This report is useful when you intend to turn over a population of native files to another party and you want to provide a complete inventory of what you are giving them which includes the hash values of every document. This report will be very long!

Duplicate Files and Emails

27. **Duplicates (Custodian) – (Professional Run Required)** This report provides a summary by custodian of the number of duplicate emails and duplicate files. If both MD5 and SHA1 were generated it will show the counts by hash type as well. In addition this report will summarize across all of the custodians the total number of duplicate emails and duplicate files. For more information on how we determine duplicate emails please refer to this document on our website: <https://www.earlycase.com/resources/earlyCASE%20Detecting%20Duplicate%20Emails.pdf>
28. **Duplicates (Global) - (Professional Run Required)** This report provides a summary across all of the custodians of the number of duplicate emails and duplicate files. If both MD5 and SHA1 were generated it will show the counts by hash type as well. In addition this report lists the duplicate documents with there MD5 hash and the count of how many duplicates there are. This section is sorted by the count descending (largest # first) and is useful to spot check the duplicates as well as see what documents were duplicated the most. For more information on how we determine duplicate emails please refer to this document on our website: <https://www.earlycase.com/resources/earlyCASE%20Detecting%20Duplicate%20Emails.pdf>
29. **Duplicate Files (Loose)** – This report will show you the global duplicate files in the collection which are loose (ie not attachments or within a container). This files can generally be de-duplicated prior to processing. The “Filter Scripts – Dates, NIST...” creates scripts to filter the global duplicate files which are loose.
30. **Duplicate emails – (Professional Run Required)** This report can be fairly long in a large population of emails. It shows the MD5 hash of the email message, the sent date, the subject and provides totals at the end of the report. This report should generally **NOT** be provided to the other side in a matter as the subject lines may contain privileged or confidential information. For more information on how we determine duplicate emails please refer to this document on our website: <https://www.earlycase.com/resources/earlyCASE%20Detecting%20Duplicate%20Emails.pdf>

Email

31. **Email "To:" summary** – This report shows you who the custodians have been sending email "TO", the counts and the date ranges of those emails. It can be useful in identifying additional custodians as well in being able to justify eliminating a custodian from what is processed and reviewed.
32. **Email "From:" summary** - This report shows you who the custodians have been receiving email "FROM", the counts and the date ranges of those emails. It can be useful in identifying additional custodians as well in being able to justify eliminating a custodian from what is processed and reviewed.
33. **Conversation Summary** – This report summarizes the number of emails (including responses and forwards) in an email thread as well as the date range of the messages in that thread. This report should generally **NOT** be provided to the other side in a matter as the subject lines may contain privileged or confidential information.
34. **Attachment Summary** –
35. **PST & NSF Analyzed - (Professional Run Required for NSF)** This report will show you a summary of the types of email containers processed, their aggregate size, how many emails were processed from those containers and the aggregate size of the messages. This report also shows by custodian and container how many messages and the aggregate size of the messages. Lotus Notes (NSF) processing requires that a Professional earlyCASE run was done.
36. **Other EMail Containers** – This report identifies less common email container files that are not processed by earlyCASE and that may require additional work to extract and review the messages. The objective of this report is to make you aware that you have some uncommon / less common email mailboxes in the population of data.

Containers

37. **Summary of Container Files - (Professional Run Required)** This report summarizes by custodian the containers observed as well as summarizing the container files (by type), provides a description of that container type along with the number of that type of container, the aggregate size of the container, and the date (last modified date) ranges of those containers. Totals of all the containers are also on this report.
38. **Container File Details - (Professional Run Required)** This report summarizes by custodian the containers observed as well as providing details on what was extracted from each container.

Collection / Images

39. **Data Collection Summary** – This report shows the details on drive images that were created using the new earlyCASE "Collect" feature. This includes the Hash values of the images, the machine information, who and when the image was created, etc.
40. **Encase, DD, AFF Image Info - (Professional Run Required)** This report details the drive images seen in the data that was processed by earlyCASE, the size of the image, the date of the image and the number of files that the image contains. It also provides the details about the image - when it was created, by who, the machine information, and any notes entered by the examiner when the image was created.
41. **Compare Run 1 and Run 2** – This report compares the files and emails analyzed in the first and second run of a matter to summarize what the differences are in the 2 runs. This is useful when you have 2 images of the same hard drive and you need to understand what (if anything) has changed. It also is helpful to validate a drive image against the source drive to insure that the image has everything.

Dates

42. **Top 25 File Dates** – The top 25 file dates report summarizes the population of documents by year and month and order the list based on the year / month with the most documents. This is useful in understanding the distribution of the documents by date. It be useful in seeing how the data you have relates to the dates of interest in the matter.
43. **File Date Summary** – This can be a very long report in that it shows by date and time the number of documents along with the aggregate size of the documents for that date time. This report is useful when you are looking for a specific document based on a narrow date range.
44. **File Date Pivot** - Pivot tables allow you to select / deselect date ranges which are then charted or visible in table form. This is useful in seeing the impact size and count of filtering date ranges from the population. This is an advanced Microsoft Excel Function and is well worth learning more about.
45. **Email Dates summary** – Summaries by Year and Month the number of emails and their aggregate size. It is useful in determining date cutoffs based on the ESI you have against the request. Allows you to quickly isolate blocks of emails by data that are clearly outside the date range at issue.

Folders / Media

46. **Folders > 100 Files** – This report shows the folders (and path) of any folder which has over 100 files in it. This is useful in checking for temporary locations, system storage which may be in the collection of data but really does not need to be processed and reviewed. This report should be used with the Folder Inventory ALL report to form a complete picture of where documents were stored and folders / paths that can be filtered out.
47. **Folder Inventory ALL** - This report shows the counts by folders (and path) and is ordered by the folder with the most files to the folders with the least files. This is useful in understanding where the bulk of the files came from and potentially identifying paths which really do not need to be processed and reviewed. This report can be pretty long, and you may want to focus on just the first 3 or 4 pages of this report.
48. **Folders with System Files** - This report will show you the folders which are found to contain system files along with the counts for those folders. In many cases folders with a lot of known system files can be excluded from processing all together.

Exports

49. **Export Email and File tables to CSV** – Use this feature to export the data in the access database to standard CSV files. The Email Details, File Details, and Folder Details are exported with this option.
50. **Microsoft Excel Export** - Use this feature to export the data in the access database to standard Microsoft Excel Spreadsheets. The Email Details, File Details, and Folder Details are exported with this option. Be aware that MS Excel has a limit if 64,000 rows, so larger matters will overflow the limits of Excel and you should use the Export to CSV report.
51. **XML Export of Data** – Users who have installed a full copy of Microsoft Access 2007 have the option of exporting the data into XML (instead of CSV). Because XML is very verbose you should only use this if the CSV option will not work for you.
52. **Detail List of System Files** – This report will show you all of the know system files in the collection you analyzed.
53. **Detail List of Duplicate Files** –This report will show you all of the global duplicate files in the collection you analyzed.

Reports available in earlyCASE

(v506)

54. **Filter Scripts – Dates, NIST...** - This will create filter scripts based on the filter rule you defined. The scripts are stored in "C:\L9\". These are batch files which can be run from DOS. They are created with a ".TXT" file extension. Check these carefully, and change the file extension to ".BAT" to run them. Additional information about filtering is available in the resources section of our website.

Additional Reports

55. **File MetaData** – This report shows select MetaData fields for the files analyzed by earlyCASE. The intent of this report is to display that metadata is being extracted. It is very verbose and should not be printed.
56. **Email MetaData** - This report shows select MetaData fields for the Email analyzed by earlyCASE. The intent of this report is to display that metadata is being extracted. It is very verbose and should not be printed.

Reports available in earlyCASE

(v506)

Alphabetical List of the currently available reports:

1. 26(f) report
2. Assumptions
3. Attachment Summary
4. Attorney Review Cost
5. Backup Tape Costs
6. Budgets & Timelines
7. Charts and Graphs
8. Compare Run 1 and Run 2
9. Container File Details
10. Conversation Summary
11. Custodian Summary
12. Data Collection Summary
13. De-Dup – File Type
14. Detail List of Duplicate Files
15. Detail List of System Files
16. Duplicate Email “To:” summary
17. Duplicate Files (Loose)
18. Duplicates (Custodian)
19. Duplicates (Global)
20. Email “From:” summary
21. Email Dates summary
22. Email MetaData
23. Encase, DD, AFF Image Info
24. Estimated Page Counts
25. Export Email and File tables to CSV
26. Extrapolated Costs
27. File Author
28. File Date Pivot
29. File Date Summary
30. File Inventory with MD5
31. File MetaData
32. File Type Pivot
33. File Type Summary
34. Filter Scripts – Dates, NIST...
35. Filter Summary
36. Folder Inventory ALL
37. Folders > 100 Files
38. Folders with System Files
39. Generally Excluded File Types
40. Generally Included File Types
41. Images Summary (tiff, jpg)
42. Known System Files
43. Matter
44. Microsoft Excel Export
45. Office File Types
46. Other Costs and Expenses
47. Other EMail Containers
48. PST & NSF
49. Sizes and Counts Expanded
50. Summary of Container Files
51. Top 25 File Dates
52. Top 25 File Types
53. Unknown File Types
54. Warnings & Errors
55. XML Export of Data