

The newer EDGAR Log File Data Set files (2020-present) contain the following variables:

Variable	Description
<code>_time</code>	The request's date-time field in the format <code>yyyy-mm-ddThh:mm:ss.sss-HHMM</code> (e.g. "2022-05-10T23:59:59.998-0400") where the HHM is the offset in hours and minutes from UTC time.
<code>uri_path</code>	Root relative path of <code>/Archives/edgar/data/</code> asset requested. For EDGAR archive assets, the fourth level folder is the integer representation of the CIK (leading zeros removed for the 10 digit CIK) and the fifth level folder is the filing's accession number with the dashes removed

The older EDGAR Log File Data Set files (2003-2017) contain the following variables:

Variable	Description
<code>ip</code>	This variable provides the first three octets of the IP address with the fourth octet obfuscated with a 3 character string that preserves the uniqueness of the last octet without revealing the full identity of the IP (###.###.###.xxx). For example, all fourth octets of 150 will have the same three character string across all files.
<code>date</code>	Apache log file date (yyyy-mm-dd)
<code>time</code>	Apache log file time (hh:mm:ss)
<code>zone</code>	Apache log file zone
<code>cik</code>	SEC Central Index Key (CIK) associated with the document requested
<code>accession</code>	SEC document accession number associated with the document requested
<code>doc</code>	This variable provides the filename of the file requested including the document extension. If the filename is missing and only the file extension is present, then the filename is the document accession number.
<code>code</code>	Apache log file status code for the request
<code>size</code>	document file size
<code>idx</code>	takes on a value of 1 if the requester landed on the index page of a set of documents (e.g., index.htm), and zero otherwise
<code>noreferrer</code>	takes on a value of one if the Apache log file referrer field is empty, and zero otherwise
<code>noagent</code>	takes on a value of one if the Apache log file user agent field is empty, and zero otherwise
<code>find</code>	numeric values from 0 to 10, that correspond to whether the following character strings /[\$string]/ were found in the referrer field – this could indicate how the document requester arrived at the document link (e.g., internal EDGAR search): a. <code>\$find=0;</code> b. <code>if(\$referrer=~m/.*(action)=getcompany)/{\$find=1};</code> c. <code>if(\$referrer=~m/.*(action)=getcurrent)/{\$find=2};</code> d. <code>if(\$referrer=~m/.*(Find\+Companies)/{\$find=3};</code> e. <code>if(\$referrer=~m/.*(cgi\bin\srch\edgar)/{\$find=4};</code> f. <code>if(\$referrer=~m/.*(EDGARFSCClient)/{\$find=5};</code> g. <code>if(\$referrer=~m/.*(cgi\bin\current)/{\$find=6};</code> h. <code>if(\$referrer=~m/.*(Archives\edgar)/{\$find=7};</code> i. <code>if(\$referrer=~m/.*(cgi-bin\viewer)/{\$find=8};</code> j. <code>if(\$referrer=~m/.*(.*\index)/{\$find=9};</code> k. <code>if(\$find==0 && \$referrer ne "-" && \$referrer ne ""){\$find=10};</code>
<code>crawler</code>	This variable takes on a value of one if the user agent self-identifies as one of the following webcrawlers or has a user code of 404. Below are the actual Perl regular expressions used: a. <code>if(\$agent=~m/(wget Googlebot polybot Yahoo!\s*Slurp spider robot perl python lwp crawler)/i){\$crawl=1};</code> b. <code>if(\$code==404){\$crawl=1};</code>

browser	<p>This variable is a three character string that identifies potential browser type by analyzing whether the user agent field contained the following /[text]/. Below are the actual Perl regular expressions used:</p> <ul style="list-style-type: none"> a. if(\$agent=~m/MSIE/){\$browser="mie"}; b. if(\$agent=~m/Firefox/){\$browser="fox"}; c. if(\$agent=~m/Safari/){\$browser="saf"}; d. if(\$agent=~m/Chrom/){\$browser="chr"}; e. if(\$agent=~m/Seamonk/){\$browser="sea"}; f. if(\$agent=~m/Opera/){\$browser="opr"}; g. if(\$agent=~m/(DoCoMo KDDI Cricket Vodaphone)/){\$browser="oth"}; h. if(\$agent=~m/Windows\s*NT/){\$browser="win"}; i. if(\$agent=~m/Mac\s*OS/i){\$browser="mac"}; j. if(\$agent=~m/Linux/i){\$browser="lin"}; k. if(\$agent=~m/iPhone/){\$browser="iph"}; l. if(\$agent=~m/iPad/){\$browser="ipd"}; m. if(\$agent=~m/Android/){\$browser="and"}; n. if(\$agent=~m/(BB10 PlayBook BlackBerry)/){\$browser="rim"}; o. if(\$agent=~m/(IEMobile Windows\s*CE Windows\s*Phone)/){\$browser="iem"}.
----------------	---