filebeat详解

```
# List of prospectors to fetch data.
filebeat.prospectors:
#------------------------------ Log prospector 定义监控哪里的日志文件------------------------------
#指定文件的输入类型log(默认)或者stdin
- input_type: log
  paths:
      - /var/log/*.log
      #- c:\programdata\elasticsearch\logs\*
  #指定被监控的文件的编码类型，使用plain和utf-8都是可以处理中文日志的。
  #encoding: plain
  #在输入中排除符合正则表达式列表的那些行
  # Exclude lines. A list of regular expressions to match. It drops the lines that are
  # matching any regular expression from the list. The include_lines is called before
  # exclude_lines. By default, no lines are dropped.
  #exclude_lines: ["^DBG"]
  #包含输入中符合正则表达式列表的那些行（默认包含所有行），include_lines执行完毕之后会执行exclude_lines。
  # Include lines. A list of regular expressions to match. It exports the lines that are
  # matching any regular expression from the list. The include_lines is called before
  # exclude_lines. By default, all the lines are exported.
  #include_lines: ["^ERR", "^WARN"]
```

#忽略掉符合正则表达式列表的文件（默认为每一个符合paths定义的文件都创建一个harvester）

Exclude files. A list of regular expressions to match. Filebeat drops the files that

are matching any regular expression from the list. By default, no files are dropped.

#exclude_files: [".gz$"]

#向输出的每一条日志添加额外的信息，比如“level:debug”，方便后续对日志进行分组统计。默认情况下，会在输出信息的fields子目录下以指定的新增fields建立子目录，例如fields.level

Optional additional fields. These field can be freely picked

to add additional information to the crawled log files for filtering

#fields:

level: debug

review: 1

#如果该选项设置为true，则新增fields成为顶级目录，而不是将其放在fields目录下。自定义的field会覆盖filebeat默认的field。

Set to true to store the additional fields as top level fields instead

of under the "fields" sub-dictionary. In case of name conflicts with the

fields added by Filebeat itself, the custom fields overwrite the default

fields.

#fields_under_root: false

#可以指定Filebeat忽略指定时间段以外修改的日志内容，比如2h（两个小时）或者5m(5分钟)。

Ignore files which were modified more then the defined timespan in the past.

```
  # ignore_older is disabled by default, so no files are ignored by setting it to 0.
  # Time strings like 2h (2 hours), 5m (5 minutes) can be used.
  #ignore_older: 0
  #设定Elasticsearch输出时的document的type字段，也可以用来给日志进行分类
  # Type to be published in the 'type' field. For Elasticsearch output,
  # the type defines the document type these entries should be stored
  # in. Default: log
  #document_type: log
  #Filebeat以多快的频率去prospector指定的目录下面检测文件更新（比如是否有新增文件），如果设置为0s，则Filebeat会尽可能快地感知更新（占用的CPU会变高）。默认是10s。
  # How often the prospector checks for new files in the paths that are specified
  # for harvesting. Specify 1s to scan the directory as frequently as possible
  # without causing Filebeat to scan too frequently. Default: 10s.
  #scan_frequency: 10s
  #每个harvester监控文件时，使用的buffer的大小
  # Defines the buffer size every harvester uses when fetching the file
  #harvester_buffer_size: 16384
  #日志文件中增加一行算一个日志事件，max_bytes限制在一次日志事件中最多上传的字节数，多出的字节会被丢弃。
  # Maximum number of bytes a single log event can have
```

```
  # All bytes after max_bytes are discarded and not sent. The default is 10MB.
  # This is especially useful for multiline log messages which can get large.
  #max_bytes: 10485760
  #适用于日志中每一条日志占据多行的情况，比如各种语言的报错信息调用栈。这个配置的下面包含如下配置：
  ### Multiline options
  # Mutiline can be used for log messages spanning multiple lines. This is common
  # for Java Stack Traces or C-Line Continuation
  #多行日志开始的那一行匹配的pattern
  # The regexp Pattern that has to be matched. The example pattern matches all lines starting with [
  #multiline.pattern: ^\[
  #是否需要对pattern条件转置使用，不翻转设为true，反转设置为false
  # Defines if the pattern set under pattern should be negated or not. Default is false.
  #multiline.negate: false
  #匹配pattern后，与前面（before）还是后面（after）的内容合并为一条日志
  # Match can be set to "after" or "before". It is used to define if lines should be append to a pattern
  # that was (not) matched before or after or as long as a pattern is not matched based on negate.
  # Note: After is the equivalent to previous and before is the equivalent to to next in Logstash
  #multiline.match: after
  #合并的最多行数（包含匹配pattern的那一行）
  # The maximum number of lines that are combined to one event.
```

```
  # In case there are more the max_lines the additional lines are discarded.
  # Default is 500
  #multiline.max_lines: 500
  #到了timeout之后，即使没有匹配一个新的pattern（发生一个新的事件），也把已经匹配的日志事件发送出去
  # After the defined timeout, an multiline event is sent even if no new pattern was found to start a new event
  # Default is 5s.
  #multiline.timeout: 5s
  #如果设置为true，Filebeat从文件尾开始监控文件新增内容，把新增的每一行文件作为一个事件依次发送，而不是从文件开始处重新发送所有内容。
  # Setting tail_files to true means filebeat starts reading new files at the end
  # instead of the beginning. If this is used in combination with log rotation
  # this can mean that the first entries of a new file are skipped.
  #tail_files: false
  # Experimental: If symlinks is enabled, symlinks are opened and harvested. The harvester is openening the
  # original for harvesting but will report the symlink name as source.
  #symlinks: false
  #Filebeat检测到某个文件到了EOF之后，每次等待多久再去检测文件是否有更新，默认为1s
  # Backoff values define how aggressively filebeat crawls new files for updates
```

```
  # The default values can be used in most cases. Backoff defines how long it is waited
  # to check a file again after EOF is reached. Default is 1s which means the file
  # is checked every second if new lines were added. This leads to a near real time crawling.
  # Every time a new line appears, backoff is reset to the initial value.
  #backoff: 1s
  #Filebeat检测到某个文件到了EOF之后，等待检测文件更新的最大时间，默认是10秒
  # Max backoff defines what the maximum backoff time is. After having backed off multiple times
  # from checking the files, the waiting time will never exceed max_backoff independent of the
  # backoff factor. Having it set to 10s means in the worst case a new line can be added to a log
  # file after having backed off multiple times, it takes a maximum of 10s to read the new line
  #max_backoff: 10s
  #定义到达max_backoff的速度，默认因子是2，到达max_backoff后，变成每次等待max_backoff那么长的时间才backoff一次，直到文件有更新才会重置为backoff
  # The backoff factor defines how fast the algorithm backs off. The bigger the backoff factor,
  # the faster the max_backoff value is reached. If this value is set to 1, no backoff will happen.
  # The backoff value will be multiplied each time with the backoff_factor until max_backoff is reached
  #backoff_factor: 2
```

```
#-------------------------- Stdin prospector -------------------
------------
# Configuration to use stdin input
#- input_type: stdin
#========================= Filebeat global options
============================
#spooler的大小，spooler中的事件数量超过这个阈值的时候会清空发送出去（不论是否到达超时时间）
# Event count spool threshold - forces network flush if exceeded
#filebeat.spool_size: 2048
# Enable async publisher pipeline in filebeat (Experimental!)
#filebeat.publish_async: false
#spooler的超时时间，如果到了超时时间，spooler也会清空发送出去（不论是否到达容量的阈值）。
# Defines how often the spooler is flushed. After idle_timeout the
spooler is
# Flush even though spool_size is not reached.
#filebeat.idle_timeout: 5s
#记录filebeat处理日志文件的位置的文件
# Name of the registry file. If a relative path is used, it is
considered relative to the
# data path.
#filebeat.registry_file: ${path.data}/registry
#如果要在本配置文件中引入其他位置的配置文件，可以写在这里（需要写完整路径），但是只处理prospector的部分
# These config files must have the full filebeat config part
inside, but only
# the prospector part is processed. All global options like
spool_size are ignored.
```

```
# The config_dir MUST point to a different directory then where
the main filebeat config file is in.
#filebeat.config_dir:
#------------------------- Elasticsearch output ----------------
-------------
output.elasticsearch:
  # Boolean flag to enable or disable the output module.
  #enabled: true
  # Array of hosts to connect to.
  # Scheme and port can be left out and will be set to the
default (http and 9200)
  # In case you specify and additional path, the scheme is
required:  http://localhost:9200/path
  # IPv6 addresses should always be defined
as:  https://[2001:db8::1]:9200
  hosts: ["localhost:9200","localhost2:9200"]
  index: "filebeat-%{+yyyy.MM.dd}"



logging.to_files: true
logging.files:
  # Configure the path where the logs are written. The default is
the logs directory
  # under the home path (the binary location).
  #path: /var/log/filebeat
```