Where do I configure my Splunk settings?

In many environments there are a lot of different Splunk servers performing different roles. For example:

- Light Forwarders
- Forwarders
- Indexers
- Search Heads
- Summarizers

When we want Splunk to do something, we can find out which configuration file, what settings, and what values to set in the Administration Manual. However it is not always clear which server the settings need to be on, especially for indexing data, and especially with the `props.conf` and `transforms.conf` file settings.

Phases of the Splunk data life cycle

To understand this, we first have to understand the different stages of the data life cycle in Splunk. These main phases for the purposes of understanding configuration are: This topic is also in the docs : http://docs.splunk.com/Documentation/Splunk /latest/Admin/Configurationparametersandthedatapipeline

Input

The **Input** phase acquires the raw data stream from its source and annotates it with source-wide *keys*. The *keys* are values that apply to the entire input source overall, and includes the host, source, and sourcetype of the data. The keys may also include values that are used internally by Splunk such as the character encoding of the data stream, and values that can control later processing of the data, such as the index into which the events should be stored.

During this phase, Splunk does not look at the contents of the data stream, so key fields must apply to the entire source, and not to individual events. In fact, at this point, Splunk has no notion of individual events at all, only a stream of data with certain global properties.

Structured Data parsing

Since splunk 6, some source can be parsed for structured data (like headers, or json) and be populated at the forwarder level. see http://docs.splunk.com/Documentation /Splunk/6.1.2/Data/Extractfieldsfromfileheadersatindextime#Forwa... Those setting have to be on the forwarders (and indexers if they monitor files)

Parsing

The **Parsing** phases looks at, analyzes, and transforms the data. The parsing phase has many sub-phases:

- Breaking the stream of data into individual lines
- Identifying, parsing, and setting time stamps
- Annotating individual events with metadata copied from the source-wide source, host, sourcetype, and other keys
- Transforming event data and metadata according to Splunk regex transform rules

Indexing

The **Indexing** phase takes the events as annotated with metadata and after transformations and writes it into the search index.

Search

Search is probably easier to understand and distinguish from the other phases, but configuration for search is similar to and often combined with that for input and parsing.

Other phases

A couple of other phases and sub-phases:

- Routing
- Jobs
- Expiration

also govern the data life cycle, but for the sake of simplification will not be discussed in this article.

Which Splunk servers go with which phases

Here are how some common Splunk server configurations correspond to these phases:

Universal/Light Forwarder	→	Indexer
Input	→	Parsing, Indexing, Search

==========

Heavy Forwarder	→	Indexer
Input, Parsing	→	Indexing, Search

==========

Universal/Light Forwarder	→	Indexer	→	Search Head
Input	→	Parsing, Indexing	→	Search

==========

Universal/Light Forwarder	→	Heavy Forwarder	→	Indexer
Input	→	Parsing	→	Indexing, Search

==========

Universal/Light Forwarder	→	Heavy Forwarder	→	Indexer	→	Search Head
Input	→	Parsing	→	Indexing	→	Search

Which configuration parameters go with which phases

This is a non-exhaustive list of which configuration parameters go with which phase. By combining this information with an understanding of which server a phase occurs on, you can determine which server particular settings need to be made on.

Input

- inputs.conf
- props.conf
  - sourcetype
  - CHARSET
  - NO_BINARY_CHECK
  - detect_trailing_nulls
  - CHECK_METHOD
  - CHECK_FOR_HEADER
  - EVENT_BREAKER_ENABLE (UF v6.5+)
  - EVENT_BREAKER (UF v6.5+)
  - FIELD_NAMES
  - PREFIX_SOURCETYPE
  - INDEXED_EXTRACTIONS, also called Structured Data Header Extraction (version 6+) http://blogs.splunk.com/2013/10/18/iis-logs-and-splunk-6/
  - LEARN_SOURCETYPE, LEARN_MODEL
- wmi.conf
- regmon-filters.conf

Parsing

- props.conf
  - TRUNCATE, LINE_BREAKER, SHOULD_LINEMERGE, BREAK_ONLY_BEFORE_DATE, and all other line merging settings
  - TZ, DATETIME_CONFIG, TIME_FORMAT, TIME_PREFIX, and all other time extraction settings and rules
  - TRANSFORMS* which includes per-event queue filtering, per-event index assignment, per-event routing. Applied in the order defined
  - SEDCMD*
  - MORE_THAN*, LESS_THAN*
- transforms.conf
  - stanzas referenced by a TRANSFORMS* clause in props.conf
  - LOOKAHEAD, DEST_KEY, WRITE_META, DEFAULT_VALUE, REPEAT_MATCH
- datetime.xml

Indexing

- props.conf
  - SEGMENTATION*
- indexes.conf

- segmenters.conf
- multikv.conf

Search

- props.conf (note that this is the order in which these occur)
  - rename
  - EXTRACT*
  - REPORT*
  - KV_MODE
  - FIELDALIAS*
  - EVAL* (version 5+)
  - LOOKUP*
- transforms.conf
  - stanzas referenced by a REPORT* clause in props.conf
  - filename, external_cmd, and all other lookup-related settings
  - FIELDS, DELIMS
  - MV_ADD
- lookup files in the lookups folders
- search and lookup scripts in the bin folders
- search commands and lookup scripts
- savedsearches.conf
- eventtypes.conf
- tags.conf
- commands.conf
- alert_actions.conf
- macros.conf
- fields.conf
- transactiontypes.conf
- multikv.conf

Other

There are some settings that don't work well in a distributed server Splunk environment. These tend to be exceptional and include:

- props.conf
  - CHECK_FOR_HEADER, LEARN_MODEL, maxDist. These are created in the parsing phase, but they require generated configurations to be moved to the search phase configuration location.

Note with 6.1 props.conf might have to go on the UFs. From dev: With 6.1, the structured data props.conf are happening at monitoring time therefore the props.conf has also to be on the forwarders.