Splunk Notes

Monday, June 25, 2018 7:41 PM

- Cyber Security Analyst II, Security-to-Business Integration 90%> data is machine data in org.
- Machine data is not always structured or generated by web servers.
- Knowledge objects classify, add enrichment, and normalize, saving reports.
- Alerts allow you to auto respond.
- Indexers, search heads, and forwarders.
  - Indexers label to source type, and events with time stamps, stored in idex
  - Search heads handle requests, and distribute requests to indexers, they perform the acutal searches. Search heads will consolidate and enrich before returning to users. Dashboards, reports, visualizations are provided by search heads.
  - Forwarders consume data and forward to indexers, little resources and reside on the machines they operate.
- Single instance deployment, handles input, parsing, indexing, searching of data. Perfect for proof of concept, learning, personal use, small department sized environments.
  - Production environment would need to split into multiple instances, adding forwarders, adding multiple indexers and search heads.
  - Search heads and indexers can be clustered.
- Three main roles:
  - Admin
    - Install apps and create knowledge objects
  - Power
    - Create and share knowledge objects for users on an app and do realtime searches
  - Users
    - Only see own knowledge objects and those shared with them
- Splunk enterprise has 2 apps
  - Home app
    - Launch, custom dashboard
  - Search and report app
    - We will do most stuff here with power user role.
- Admins can add data to splunk
- In most prod environments, forwarders will be the main source of data
- Source types - splunk uses source types to categorize the data being indexed
- Save source type lets you choose name, description, category, app.
- App - system allows it to be used system wide.
- Hostname is the name of machine that this data originates
- Indexes are directories where data will be stored
- Separate indexes will make search more efficient
  - Allows you to limit roles for access
  - Retain data for different time intervals
    - Custom retention policies
- Upload
  - One time file upload
- Monitor
  - Allows you to monitor files and directories, http event collector, tcp/udp, scripts.
  - Files and directories
    - Continuously monitor or index once
    - You can blacklist or whitelist specific files in directories.
- Forwarders are out of scope for this source.

- 7 main components to search & reporting
  - Spunk bar - switch between apps, edit account, system level messages, settings, monitor search jobs, find help
  - App bar - navigate application
  - Search bar- run searches
  - How to earch panel, links to documentation and tutorial
  - What to search, summary of data indexed
    - Data summary button, breakdown by host, source, sourcetypes
    - Sourcetypes
      - Classification of data
    - Sources
      - Origination of data
    - Host
      - FQDN or ip address of originator
  - Search history
    - View and rerun past searches
- Searches
  - Failed - select last 7 days to limit by time, best practice
  - Save as menu - to save search to knowledge objects, search results tab, search mode selector, timeline.
  - Event list shows results, field list on side bar
  - Events tab will show results and fields extracted
  - Patterns tab allows you to see patterns in your data
  - Statistics or visuals, they will be displayed in statistics or visualization tabs
    - If not, instant pivot, quick reports, and search commands will be displayed
  - Commands that create statistics and visualizations are called transforming commands.
    - Transform into data tables
  - By default, a search job will remain active for 10 mins
    - After, splunk will have to rerun to show results
  - Shared search jobs will remain active for 7 days.
  - Export icon allows you to export in raw, csv, xml, json format.
  - 3 search modes
    - Smart, fast, verbose
    - Fast
      - Cutting down on field info, field discovery is disabled, only returning default field info and fields required to fulfill your search
    - Verbose
      - Returns as much field and info as possible
    - Smart
      - Toggle behavior based on type of search.
  - You can select time ranges by clicking and dragging on the timeline
    - You can zoom in and out of a selection
    - Splunk will use the old job when you zoom in, but will need to run a new search job to return newly selected events when you zoom out.
- Events
  - Text searched for is highlighted, reverse chronological order, time is normalized based on what you set in your user account.
  - Selected fields shown at bottom of events, host, source, sourcetype is the default for this serach
  - Rolling over text will highlight it, allow you to add that to search, exclude from search, or launch new search. You can click on it again to remove from search if you have added it.

- Event actions and field actions are out of scope.

- Search Everything
  - Adding an astrisk after fail will return fail, failure, and failed. Booleans with NOT, OR, AND as long as they are capped.. Like FAILED OR PASSWORD, if no boolean is used, AND is implied (e.g. failed NOT password)
  - Order of boolean ops: 1. NOT 2. OR 3. AND, unless parentheses are used (e.g. password AND (failed OR fail)
  - Quotes can be used for exact phrases, backslash to escape the quotes. (e.g. \"rocky\" so you can use them in the search.)
- Field sidebar, selected fields are most important to you
  - Interesting fields have values in 20% of events
  - # number
  - a text
  - Adding a field persists for subsequent searches
  - =, !=. > can be used with fields.
  - != can be used similarly to NOT
  - Field values are not case sensitive
  - Field names are case sensitive
  - Wildcards can be used with fields
- Best practices
  - Using time filters - most efficient way to search
  - Using specific filters (inclusion is better than exclusion)
  - Filtering early is optimal
  - Realtime searches - 10 mins ago until now.
  - -30m for 30 mins ago, s for seconds, m for minutes, h for hours, d for days, w for weeks, mon for months, y for years, @ to round down, -30m@h run at 9:37 would be run at 9
  - Earliest=-2h latest =-1h will specify the range in the search bar
  - 01/08/2018:12:00:00 can also be used
  - Using indexes
    - Segregates data
    - Makes searches more efficient
    - Limits access
    - Index=[name] to search
    - Can use multiple indexes with OR
    - Wildcards can be also used
- The splunk search language
    - Search terms
    - Commands - blue
      - Tell splunk what to do with results, including charts, stats, formatting
    - Functions - purple
      - Explains how we want to chart, compute and evaluate results
    - Arguments - green
      - Variables we want to apply to function
    - Clauses
      - How we want results grouped or designed
  - Hotkeys
    - Cntrl or command key plus \ will make a new line for the pipe
  - Left to right search flow
    - Splunk will go left to right, once you pipe something out it is no longer available to search
- SPL Fundamentals
  - Fields command
    - Include or exclude fields from results, to exclude use negative before the names of the

fields.
        - Internal fields will always be displayed unless you use - _raw and _time
        - Field extraction is one of the most costly parts of searching in splunk
        - Field inclusion happens before field extraction and can improve performance
        - Field exclusion happens after field extraction, only affecting displayed results
    - Table command
        - Retains searched data in a tabulated format
        - Table [field names]
        - Each row are events
    - Rename command
        - Rename fields
        - Rename [field] as "[new name]", you can rename multiple fields with a space inbetween
        - You need to use new field names further down pipe
    - Dedup command
        - Removes events with duplicate values
        - Dedup [field name] [field name 2]
    - Sort command
        - Dispalys results in ascending or descending order
        - Sort [fioeld name] [field name 2]
        - + is ascending
        - - is decending
        - Sort - sale_price vendor
            - That space after - will apply the decending to both, removing it will only apply it to sale_price
        - Using the limit arument
            - Adding limit=20 will limit it to the first 20 results
- Transforming commands
    - Order search results into a data table for statistical purposes
    - Need them to transfrom results into visualizations
        - Top Command
            - Finds most common values of a given field
            - Will automatically present count and precent columns
            - Limit clause can be added
            - Can use for multiple fields
            - Top Vendor limit=5 showperc=False will remove percentages for the top 5 vendors, countfield="Number of Sales" will rename count field, useother=True will group into other.
            - You can use by to show multiple things.. So
                - Top product_name by Vendor limit=3 countfield="Number of Sales" showperc="False" would give you the top 3 products per vendor, without percents
            - Otherstr changes name of "other"
        - Rare command, same option, but shows least common values
        - 10 results are shown by default for both top and rare
- Stats command
    - Show statistics
    - Count
        - Number of events matching
            - Stats count as "total Sells by Vendors" by product_name, sale_price
            - You can add a count like
                - Stats count(action) as ActionEvents, only events with actions are counted
            - Adding a comma lets you do another count column

- Distinct count (dc)
    - unique values for a field
    - Stats distinct_count(product_name) as "Blah" by sale_price
- Sum
    - Sum of numerical values
    - stats sum(price) as "Gross Sales" by product_Name
    - Stats count as "units sold" sum(price) as "Gross Sales" by product_name
        - This command needs to be in the same pipe
- Average (avg)
    - Avg value of numerical value fields
    - stats avg(sale_price) as "avg sell price" by title
- List
    - Lists all values of fields
    - stats list(Asset) as "company assets" by employee
- Values
    - Unique values of a given field
    - stats values(s_hostname) by cs_username
- Lab notes
    - Top clientip is 87.194.216.51
    - Top product id is WC-SH-G04
    - API file used least bandwidth

- Reports and Dashboards
    - If a search returns statistical values it can be viewed as a chart
    - Save as report, give a title, description, and if we want to display time range picker.
    - Power users can change permission, run as owner or user, use user for more sensitive info
    - Edit schedule will let you run scheduled reports
    - Acceleration will let you run things faster, with a smaller index
    - Quick reports are available from the sidebar
        - Table, bar or both options
    - Dashboard is a collection of reports (panels)
    - Time ranges will only work for inline searches.
    - All roles can create reports
    - Charts can be based on numbers, time, or location
    - Lab Notes
        - 100 attempts from 73.202.92.7 for 403 pages
- Pivots and Datasets
    - Pivots are easier to use, without using searches
    - Pivots can be saved as dashboard panels
    - Adding child data model objects is like the "AND" Boolean in the Splunk Search Language
    - Data models are knowledge objects that provide the data structure that drives pivots
        - Created by admin and power roles
        - Data model is made of datasets
            - Represented as tables
            - Can be made by admin and power roles
    - Settings, data model, pivot
        - Here you will see the hierarchy of datasets
        - Default time range is all time for pivots
        - Filters are available, and row and column selectors
        - Pivot sidebar is used to visualize data
        - Can add to dashboard or save as reports
    - Instant pivot
        - You can pivot without having a data model
        - Clicking pivot after a search will initiate this

        - You get to pick the fields in the data model
        - The instant pivot button is displayed in the stats and visualizations tab when a non-transforming search is run.
    - Datasets will let you see the datasets.
        - Field names are column headers, event data as rows
        - Summarize fields at the top will give you info for each field name and values
        - Explore menu will let you visualize with pivot or investigate in search
- Lookups
    - Tying data not in an index.
    - You can use its field in search, and it shows up in field sidebar
    - External data used by a lookup can come from sources like scripts, CSV files, and geospatial data.
    - Lookup is categorized as a dataset
    - To keep from overwriting existing fields with your Lookup you can use the outputnew clause.
    - Two steps to setup lookup file:
        - Define lookup datable
        - Define lookup
        - You can configure it to run automatically
        - Lookup field values are case-sensitive by default
    - Settings -> Lookups, click new lookup table file
        - Select destination app
        - Choose file, choose filename
    - Inputlookup [filename] to display data from a file
    - Settings -> Lookups -> new lookup definitions
        - Destinations app
        - Name
        - File-based
        - Select csv file
            - Your first row is the field names
        - Time based lookups
            - Field that represents time, you can create a time based lookup
        - Advanced
            - Minimum matches, max matches, default matches, cases senitive, batch index query will group index queries for better perfomance, match type, filter lookup
    - Lookup [name of lookup] code as status OUTPUT code as "Http Code", description as "Http Description" | table host, "Http Code", "Http Description"
    - lookup products.csv productId OUTPUT product_name

    - Automatic lookup
        - Destination app
        - Name
        - Select lookup table
        - What data to apply to, "access_Combined"
        - Input field - code = status
        - code = Code
        - description = Description
    - Additional lookup options
        - Populate lookup table with search results
        - Define lookup based on external script or command
        - Use splunk DB connect app to do lookups based on external databases
        - Use geospatial lookups to create queries that can be used to generate choropleth map visualizations
        - Populate events with kv store fields

- Scheduled reports and alerts
  - Can trigger an action each time it runs
  - Useful for weekly or monthly reports
    - Dashboards
    - Automatically sending emails
  - Create a report
    - Search
    - Create a report
    - Schedule a report after creating the report
    - Select the schedule
  - Running concrrent reports can put a demand on your system hardware, even if everything is configured to recommended specs
  - You can set priority to put less strain on your deployment
    - Only available to admin users
    - Default higher and highest are the options
    - Schedule window sets timeframe to run report
      - If busy it will be delayed within acceptable window
    - Include schedule window if you're ok with delay and it doesn't have to run at a specific time, you can also set this to auto to let splunk decide the best
    - You can also choose actions to run
      - Log event to splunk receiver endpoint
      - Output results to lookup
      - Output results to telemetry endpoint
      - Run a script
      - Send email
        - Enter addressee, priority, subject (with tokens like $name$), message, and what to include.
      - Webhook (post to external url)
      - And you can manage and browse for more actions depending on user role (admins can do this).
    - Managing scheduled report
      - You can edit search, permissions, schedule, acceleration, summary indexing.
      - You can also disable, clone, embed, move, or delete reports.
        - Embedding - anyone with access to web page can see embedded report
        - Emebedded report will not show data until scheduled report is run. Once it is enabled we will not be able to edit attributes of that report.
- Alerts
  - Based on searches that run on scheduled intervals or in real-time
  - Notify you when the results of a search meet defined conditions.
  - Triggers once searches are completed
  - You can still edit defining searches after an alert is created.
  - Alerts can send emails and be shared to all apps

  - Can:
    - List in interface
    - Log events
    - Output to lookup
    - Send to telemetry endpoint
    - Trigger scripts
    - Send email
    - Use a webhook or run a custom alert
  - Creating alert
    - Save as -> alert

        - Set title
        - permissions
            - (Shared in app will allow all users of app to see, but by default all have read access and admin have write access to alert),
        - alert type
            - (scheduled or real time)
            - Scheduled allows you to set a schedule and time range for the search to be run
            - Use cron expression or time range
            - Real time - search continuously in the background
                - Can place more overhead on performance
        - Trigger alert when
            - Per result - number of results, number of hosts, number of sources, custom.
        - Throttle
            - Suppressed based on time intervals or field values
        - Trigger Actions (different based on user role)
            - Add to triggered alerts
            - Log event
                - Sent to splunk deployment for indexing
                - Set event, source, sourcetype, host and index
            - Output results to lookup
                - Create or update CSV lookup table, append or replace
            - Output results to telemetry endpoint
                - Set a name, input field, data type, categories, and opt-in
            - Run a script
                - Trigger a shell script or bash file, officially deprecated and use a custom action instead
            - Send emails
                - Where to send, priority, subject, message, include, and type
                - You can use tokens
            - Webhooks,
                - Callbacks on web resource
                - Ticket in support app or message in chatroom
            - Custom action
                - You can build your own alert actions
- Activity -> Triggered alerts and alerts menu in search and reporting app to edit alerts