Welcome to *Splunk Fundamentals 2*

splunk> turn data into doing™

Let's Get Ready To Spluuuuuuuuuuuuuuuuunk!>

Mitch Fleischman
Senior Staff Instructor

+1 650.605.7549 m
mitchf@splunk.com
splunk.com

splunk>

splunk> ENTERPRISE CERTIFIED ARCHITECT

@ Splunk 7 years

Enterprise software experience 25+ years:

- Customer Relationship Management (CRM)
- Relational Database (RDB)
- Portal
- Business Process Management (BPM)
- Endpoint Management
- **Big Data / Data Analytics**

Roles:

- Programmer
- Project Lead
- Consultant
- Course Developer
- **Instructor**

Geography:

- New Jersey
- California
- Vermont
- **Metro DC / Arlington, VA**

Fundamentals 1

Module 1: Introducing Splunk

Module 2: Searching

Module 3: Using Fields in Searches

Module 4: Creating Reports and Dashboards

Module 5: Splunk's Search Language table, fields, rename, dedup, sort

Module 6: Transforming Commands top, rare, stats

Module 7: Creating and Using Lookups

Module 8: Creating Scheduled Reports and Alerts

Appendix B: Using Pivot Data Models and datasets

Fundamentals 2 (Part A): More Search Commands

Module 1: Beyond Search Fundamentals

Review / Index buckets / Performance tips / Search modes / Job Inspector

Module 2: Using Transforming Commands for Visualization

chart / timechart

Module 3: Using Trending, Mapping, and Single Value Commands

trendline (sma, ema, wma) / maps: iplocation, geostats, geom / single value visualizations / addtotals

Module 4: Filtering and Formatting Results

eval / search, where / fillnull

Module 5: Correlating Events

transaction

Fundamentals 2 (Part B): Knowledge Objects

Ease of use for SPL / ease of access (Field aliases, Tags, Data models, Dashboards) / **Re-usability** (Reports, Macros)
Provide **more context** for searches (Lookups, Field aliases, Tags) / **Greater insight** / depth of analysis (Event types, Workflow)

- Field extraction rex / erex not persisted (search syntax) / FX persists as a knowledge object
- Field alias - alias for field name (Username / cs_username => user)
  - Easier name to work with
  - Normalize different names from different sources (possibly to become CIM-compliant)
- Calculated Field - Persistence of an | eval bandwidthMB=round((sc_bytes / (1024*1024)), 2)
- Tag - Alias for field value
  - Easier value to work with (adldapsv1 => ldap)
  - Aggregate different values together (adldapsv1 / badgesv1 => IT)
- Eventtype - Categorize / classify results [optionally, tag them]
- Macro - Re-usable portions of search code
- Workflow - User interaction with search results (GET / POST / Search)
- Data model - GUI search (Pivot), normalize inputs (especially for CIM and ES), accelerate searches (tstats)

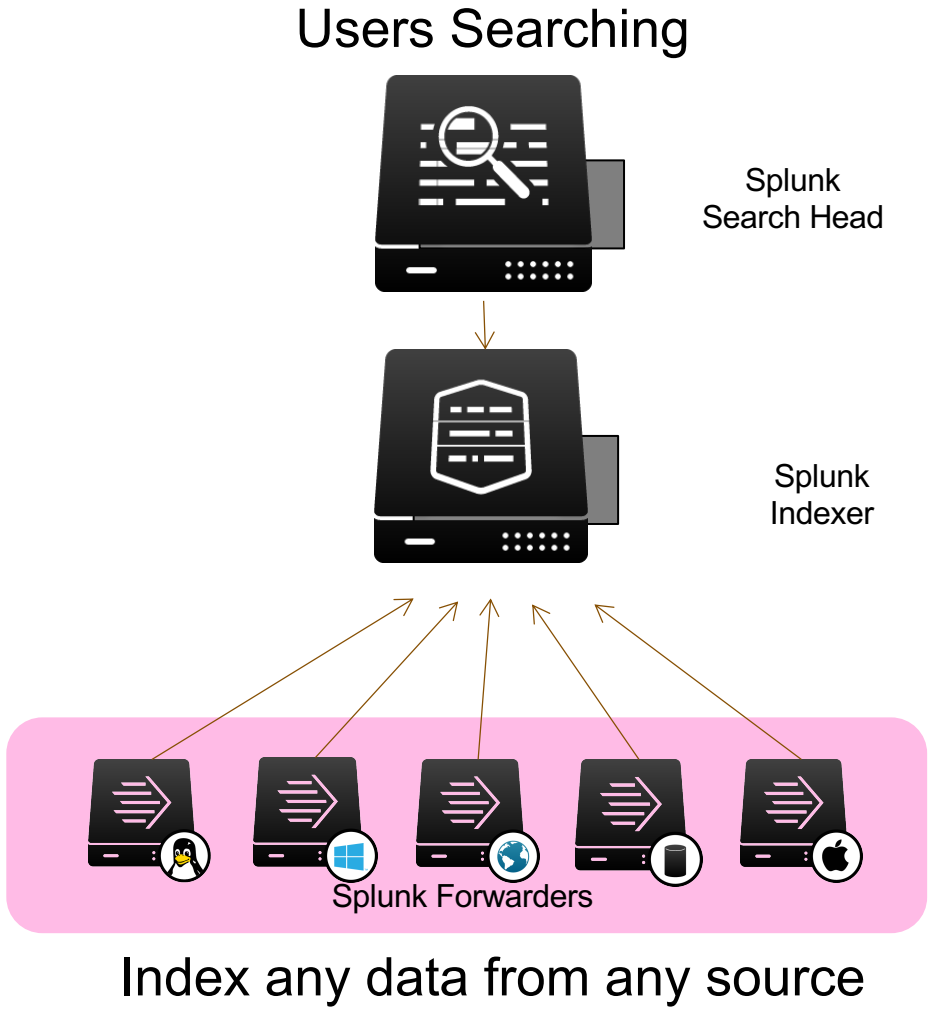
Fundamentals 3

- Module 1: Exploring Statistical Commands — appendpipe / streamstats / eventstats
- Module 2: Exploring eval Command Functions — tostring / tonumber / printf / strftime / strptime / replace / if / case
- Module 3: Exploring Lookups — Include, exclude / KV Store / External / DB Connect
- Module 4: Exploring Alerts — Lookups and Alerts / Webhook
- Module 5: Advanced Field Creation and Management — Field extractor / rex / erex
- Module 6: Self-Describing Data and Files — JSON and XML files: spath; Table-formatted (netstat, etc): multikv
- Module 7: Advanced Search Macros
- Module 8: Using Acceleration Options: Report Acceleration and Summary Indexing
- Module 9: Using Acceleration Options: Data Models and tsidx Files — tstats

Got Data?

- Computers
- Network devices
- Virtual machines
- Internet devices
- Communication devices
- Sensors
- Databases
- ***Any source***

- Logs
- Configurations
- Messages
- Call detail records
- Clickstream
- Alerts
- Metrics
- Scripts
- Changes
- Tickets
- ***Any data***



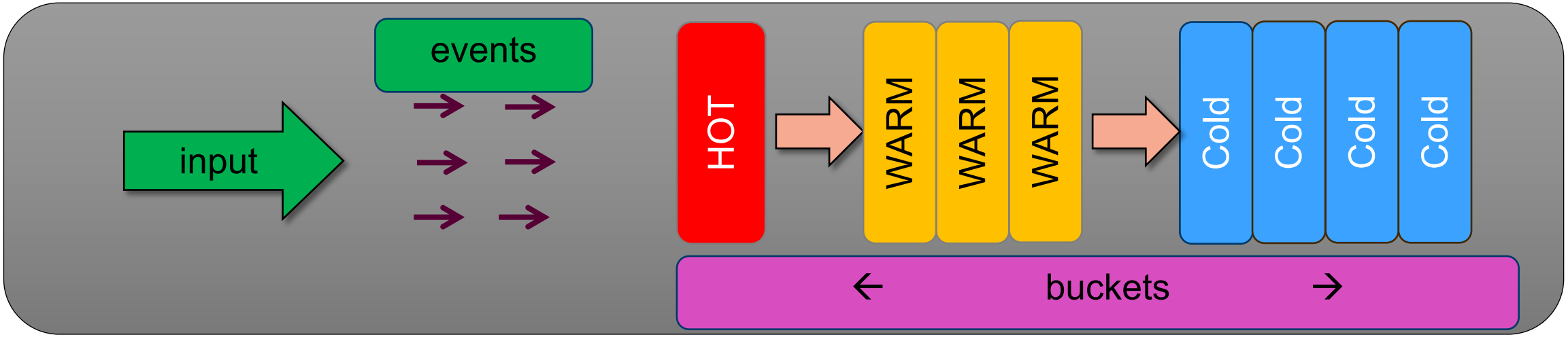
Index any data from any source

Setting the Standard for Operational Intelligence

- 2006 – 2008 Versions 1, 2, 3
  Tool: "Google for the data center"
- 2009 – 2011 Versions 4, 4.1, 4.2, 4.3
  "**Engine** for machine-generated data"
- 2012 – 2015 Versions 5, 6, 6.1, 6.2
  "**Platform** for operational intelligence"
- 2016 – Present Versions 6.3+, 7.x
  Enterprise Machine Data Fabric:
  "Accelerating digital transformation for business innovation"
- 2019 and beyond: "The **Data-to-Everything** Platform"

How Splunk Indexes Data

- Inputs (file / network port, etc) come into Splunk and are broken up into events with these attributes:
  - host: The machine where the data originated
  - source: For example, path to the file
  - sourcetype: Classification / categorization, for example web log: access_combined
  - timestamp: Epoch (Unix) time, UTC offset
  - index (main by default)
- A bucket (directory) has the raw data (compressed), a time-series index (.tsidx files), [bloomfilter]
  - Each bucket has an earliest and latest time for the events it contains
  - There are also metadata (.data) files that track `source`, `sourcetype`, and `host`



Splunk Web: *group1*.class.splunk.com

Name	Splunk Web User	Splunk Web Password
Ahmad Rasool	student1	Tysons1
Anil Rock	student2	Tysons2
Ann Hoang	student3	Tysons3
Aschalew Lingane	student4	Tysons4
Bradly Pardon	student5	Tysons5
Christine Tran	student6	Tysons6
Dee Bailey	student7	Tysons7
Evonne Goldman	student8	Tysons8
Harold King	student9	Tysons9
Jayman Gandhi	student10	Tysons10

Splunk Web: *group2.class.splunk.com*

Name	Splunk Web User	Splunk Web Password
Jhonathan Lora	student1	Corner1
Kelly Moan	student2	Corner2
Luis Rodriguez	student3	Corner3
Marcus Jones	student4	Corner4
Michael Stone	student5	Corner5
Rahul Patel	student6	Corner6
Robyn Swan	student7	Corner7
Ronald Richter	student8	Corner8
Spivey Williams	student9	Corner9
William Vollono	student10	Corner10

Search-time Operation Sequence

- Search-time operations are always applied in the same order
- Each operation can reference fields derived from operations that **precede** them
- No operation can reference fields that are derived by operations that **follow** them

1. Inline field extractions
2. Field extractions that use a field transform
3. Automatic key-value field extractions
4. Field aliases — can create a field alias from a field extraction
5. Calculated fields — can't create a field alias from a calculated field
6. Lookups
7. Event types
8. Tags

Note

Details are available at https://docs.splunk.com/Documentation/Splunk/latest/Knowledge/Searchtimeoperationssequence#

Acceleration Overview

Summary Indexing	• Store search results (typically a small subset of raw events) in indexes separate from the main indexes • The original acceleration technique - to speed up dashboard loading, panels load from a smaller index • Premium apps such as ES and ITSI use as repositories for app components such as notable events • Summary indexes reside on the Search Head! "Events" stored in the normal fashion – no additional metering (Admins configure Search Head Forwarding to the Indexer tier) • Can persist after "parent" events have been frozen Summary indexes have their own settings for controlling retention period and index size
Report Acceleration	• Store results of statistical (reporting) searches in indexes separate from the main indexes • Qualifying searches must use a reporting command (stats, top, etc) • All matching events from the search have to be "streaming" into the reporting command (no dedup, no transaction) • If these criteria can't be met, fallback is Summary Indexing - given a choice, use Report Acceleration • RA indexes reside on the Indexer, alongside the main index buckets. "Events" stored in the normal fashion – no additional license metering • Configurable retention window – but data always ages out with the corresponding main index buckets
Data Model Acceleration	• Store metadata (statistics), **not** search results, for the fields defined in the datamodel • Storage format is time-series index (tsidx) files, which are created on the Indexer – no additional metering • Perform searches directly against the metadata (tstats command) without opening the raw data • Huge performance gain ("cost" is maintenance processing and storage space) • Data ages out when the corresponding main index buckets ages out

Common Information Model (CIM)

The Splunk Common Information Model (Splunk_SA_CIM) provides a methodology to normalize data:

1. Data inputs can be structured (xml) or unstructured (tcp stream)
2. Inputs (raw data) are broken up into events
3. Events can be broken down into extracted fields and tags
4. Inputs represent a domain of interest:

    * authentication * email * network * web log * web proxy

Using CIM data models, you can test whether your fields have been normalized correctly. You can also use the models to generate searches.

Related domains have a core of common fields, regardless of vendor, for example, the data model Web (for Web logs and Web proxy) expects for Web logs:

*src *dest *bytes *bytes_in *bytes_out *status *url . . . etc

CIM (Continued)

- An event should have (at minimum) the following information:
  `<timestamp> name="<name>" event_id=<event_id> <key>=<value>`
- Any number of field key-value pairs are allowed:
  `2008-11-06 22:29:04 name="Failed Login" event_id=sshd:failure`
  `src_ip=10.2.3.4 src_port=12355 dest_ip=192.168.1.35 dest_port=22`
- The objective of normalization is to use the same names and values for equivalent events from different sources or vendors
  - Normalized data are able to present a unified view of a data domain across heterogeneous vendor data formats
  - Data is normalized when events from different products and vendors, formatted in different ways, have the same field values for the semantically equivalent events
- Reference

http://docs.splunk.com/Documentation/Splunk/latest/Knowledge/UnderstandandusetheCommonInformationModel

Resources

Splunk docs docs.splunk.com

Quick Reference Guide https://www.splunk.com/pdfs/solution-guides/splunk-quick-reference-guide.pdf

Dashboard Quick Ref Guide

https://www.splunk.com/pdfs/solution-guides/splunk-dashboards-quick-reference-guide.pdf

Splunk YouTube: http://splk.it/How-To

Download / Install Free Splunk: splunk.com/download

Sample data too (updated daily):
http://docs.splunk.com/Documentation/Splunk/latest/SearchTutorial/Systemrequirements#Download_the_tutorial_data_files

Pre-trained sourcetypes:
http://docs.splunk.com/Documentation/Splunk/latest/Data/Listofpretrainedsourcetypes?r=searchtip

Splunk Community

- Ask an expert: http://answers.splunk.com
  - Get answers to your questions from Splunk experts
- Hot wiki topics: http://wiki.splunk.com
  - Splunk and customer generated knowledge base
  - Includes best practices and how-tos
- Splunk blogs: http://blogs.splunk.com
  - Tips, tricks, latest events, and trends from Splunk insiders
- Splunk for developers: http://dev.splunk.com
  - Customize and extend the power of Splunk
- Splunk user groups: https://usergroups.splunk.com
  - Connect with like-minded Splunk professionals near you
- Splunk Slack community: http://splunk402.com

Splunk Training + Certification

1. Course **Certificate of Completion** Perform the labs!
2. **Splunk Certification** Program
   - Tracks
     https://www.splunk.com/en_us/training/faq-training.html

     Splunk Core Certified User
     Splunk Core Certified Power User
     Splunk Enterprise Certified Admin
     Splunk Enterprise Certified Architect
     Splunk Certified Developer
   - Program information handbook
     https://www.splunk.com/pdfs/training/Splunk-Certification-Candidate-Handbook.pdf
   - Exam registration
     https://www.splunk.com/pdfs/training/Exam-Registration-Tutorial.pdf
   - If you have further questions, send an email to certification@splunk.com

splunk> .conf19

.conf19
October 21-24, 2019

Splunk University
October 19-21, 2019

Las Vegas, NV
The Venetian Sands Expo

4 Days of Innovation

350 Education Sessions

20 Hours of Networking

conf.splunk.com

splunk>

Making machine data accessible, usable, and valuable to everyone.