Deploying Splunk

- Deployment models with Splunk's Validated Architecture
- Storing data on disk
- Licensing
- Configuration files
- Apps and add-ons

Deployment Models

The Splunk Data Pipeline

Input

- Forwarded data, uploaded data, network data, scripts

Parsing

- Examines the data, adds metadata

Indexing

- Data divided into events. Writes the data to the disk in "buckets"

Searching

- User interaction with the data

Splunk Platform

The Splunk Platform

Splunk Cloud Platform

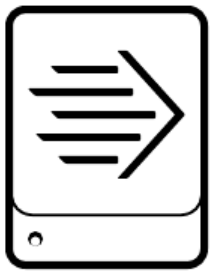
Splunk Enterprise

Splunk Enterprise

Splunk Enterprise

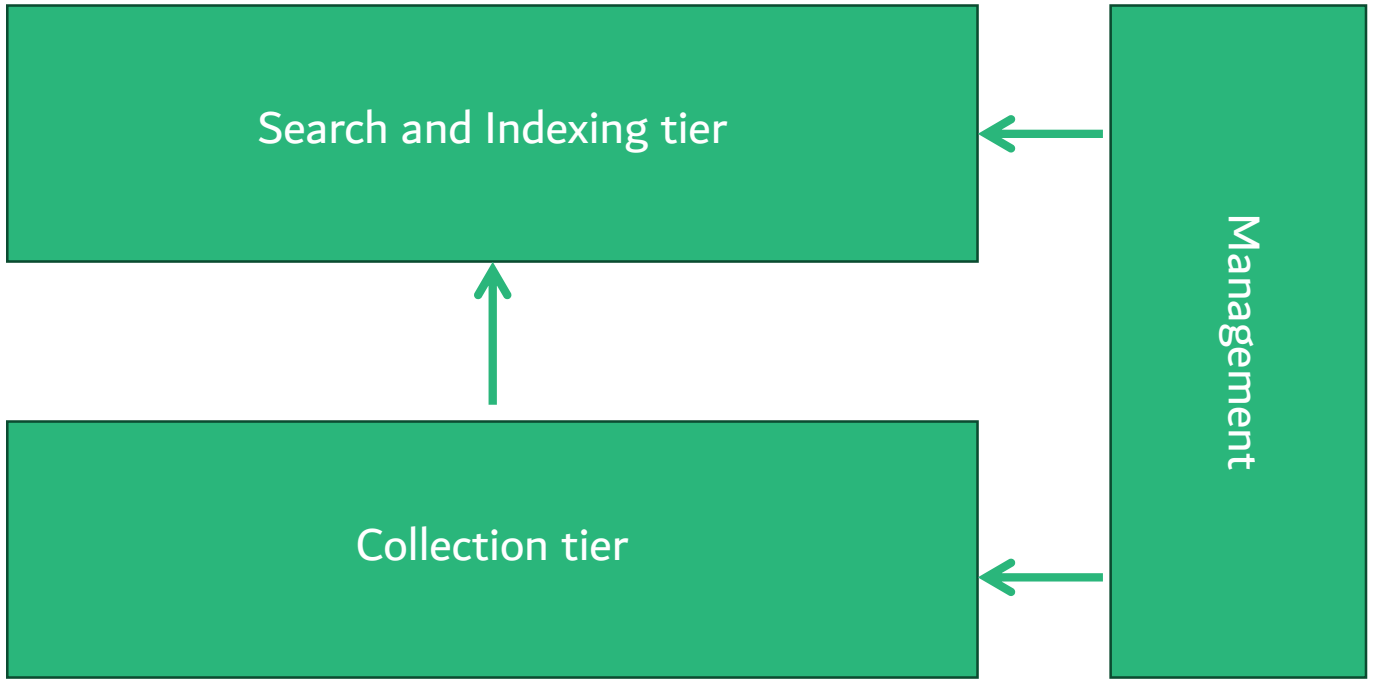
Search Head

Indexer



Forwarder

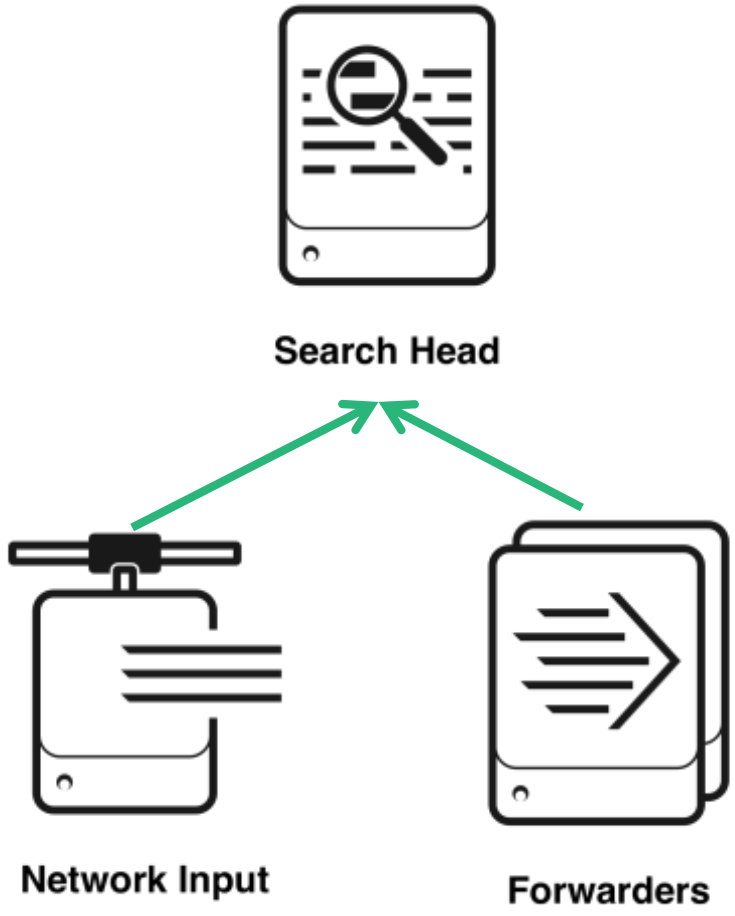
Splunk Enterprise S1 Architecture



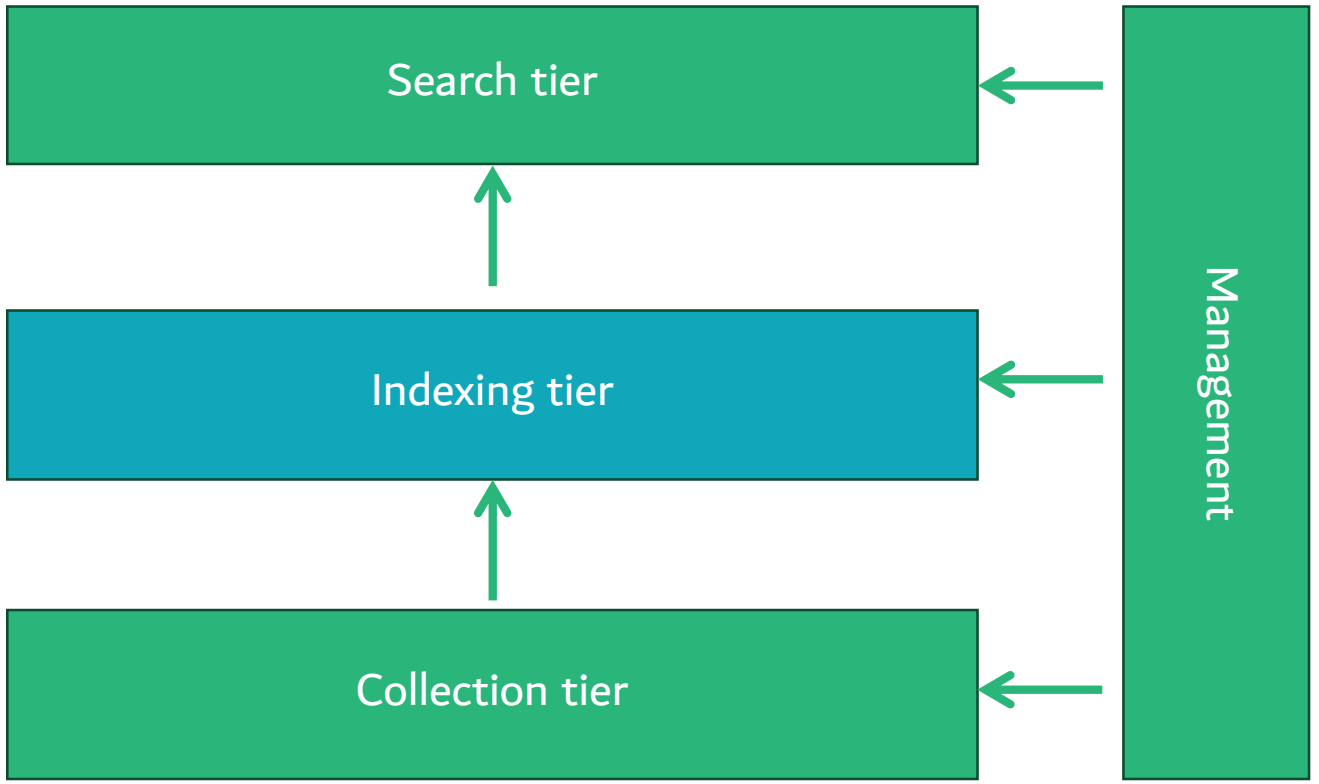
Splunk Enterprise

S1 (Single Server)

- Search Head and Indexer **are the same Splunk component**
- Daily data ingest 500GB
- Small number of users
- Splunk recommends S1 for non-critical data



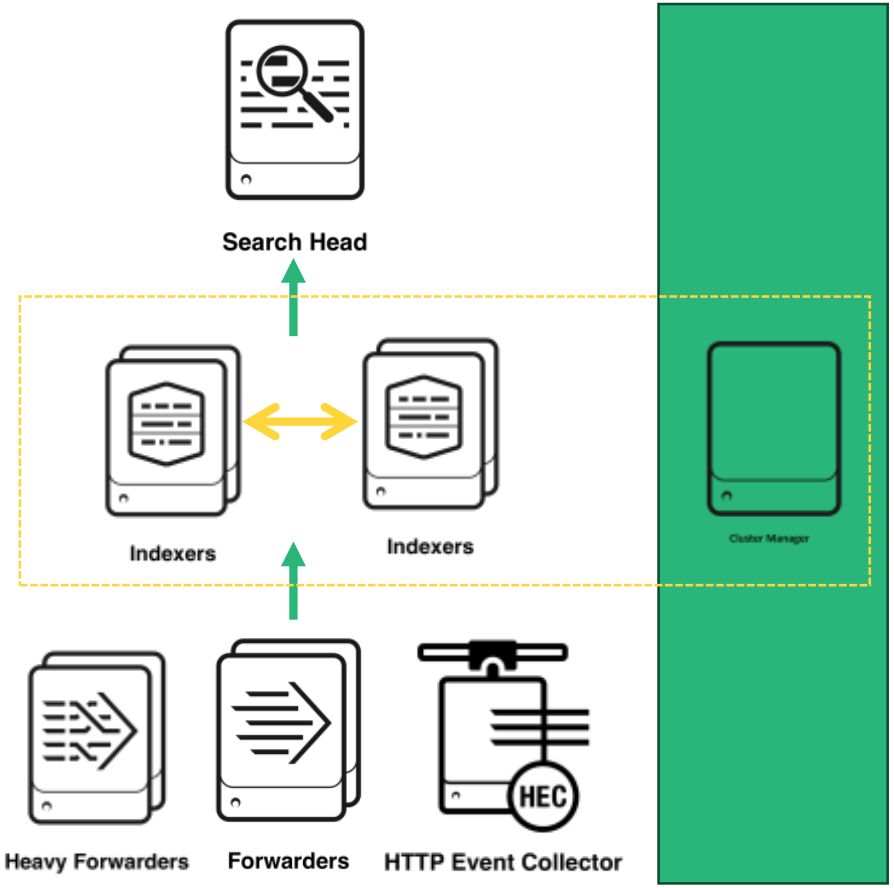
Splunk Enterprise C1 Architecture



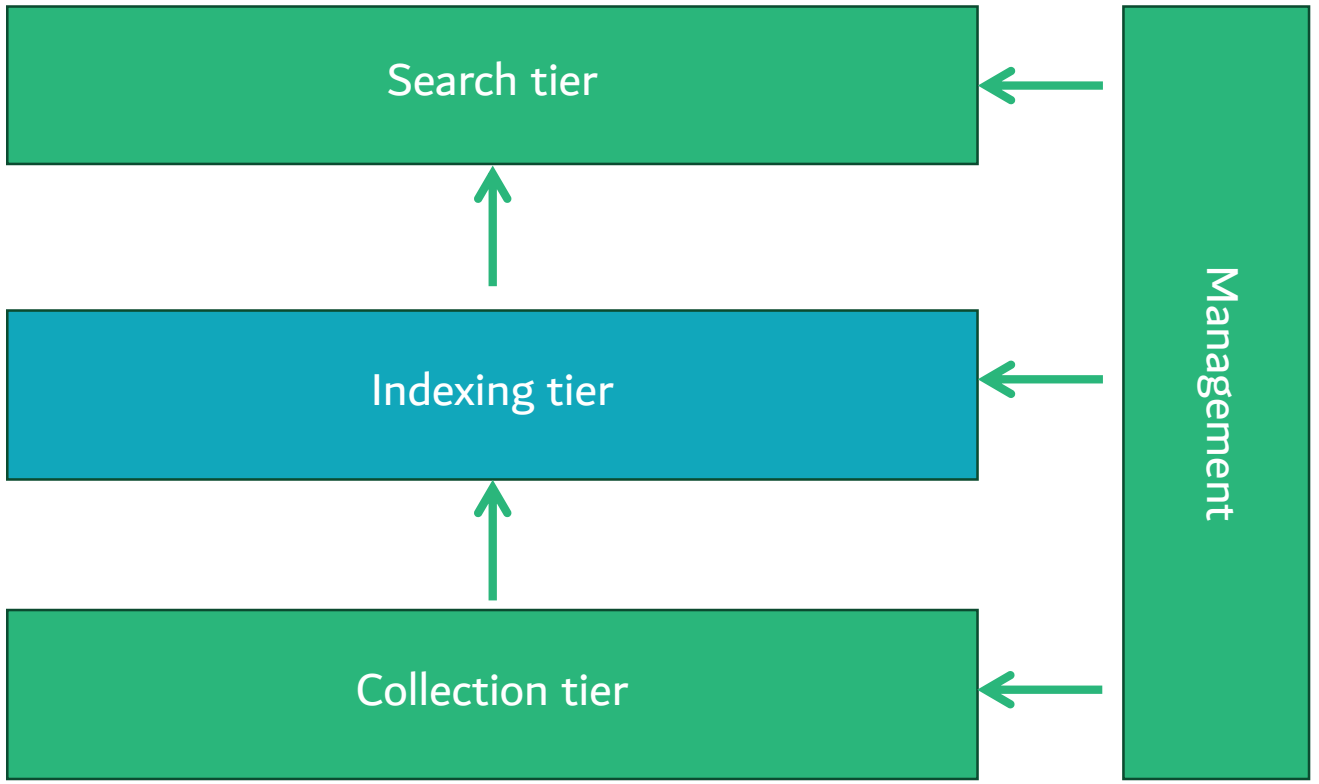
Splunk Enterprise

C1/C11 (Distributed Clustered Deployment – Single Site)

- One or more stand-alone search heads
- An indexer cluster with data replication
- Multiple, load balanced collection inputs



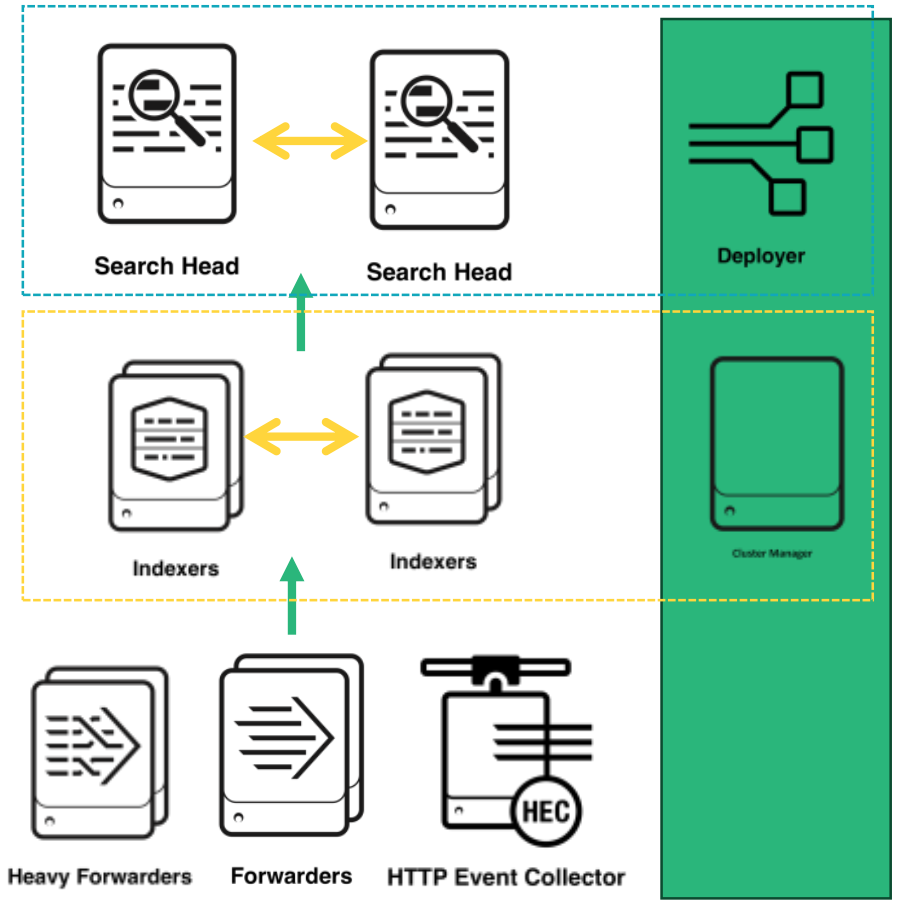
Splunk Enterprise C3 Architecture



Splunk Enterprise C3 Architecture

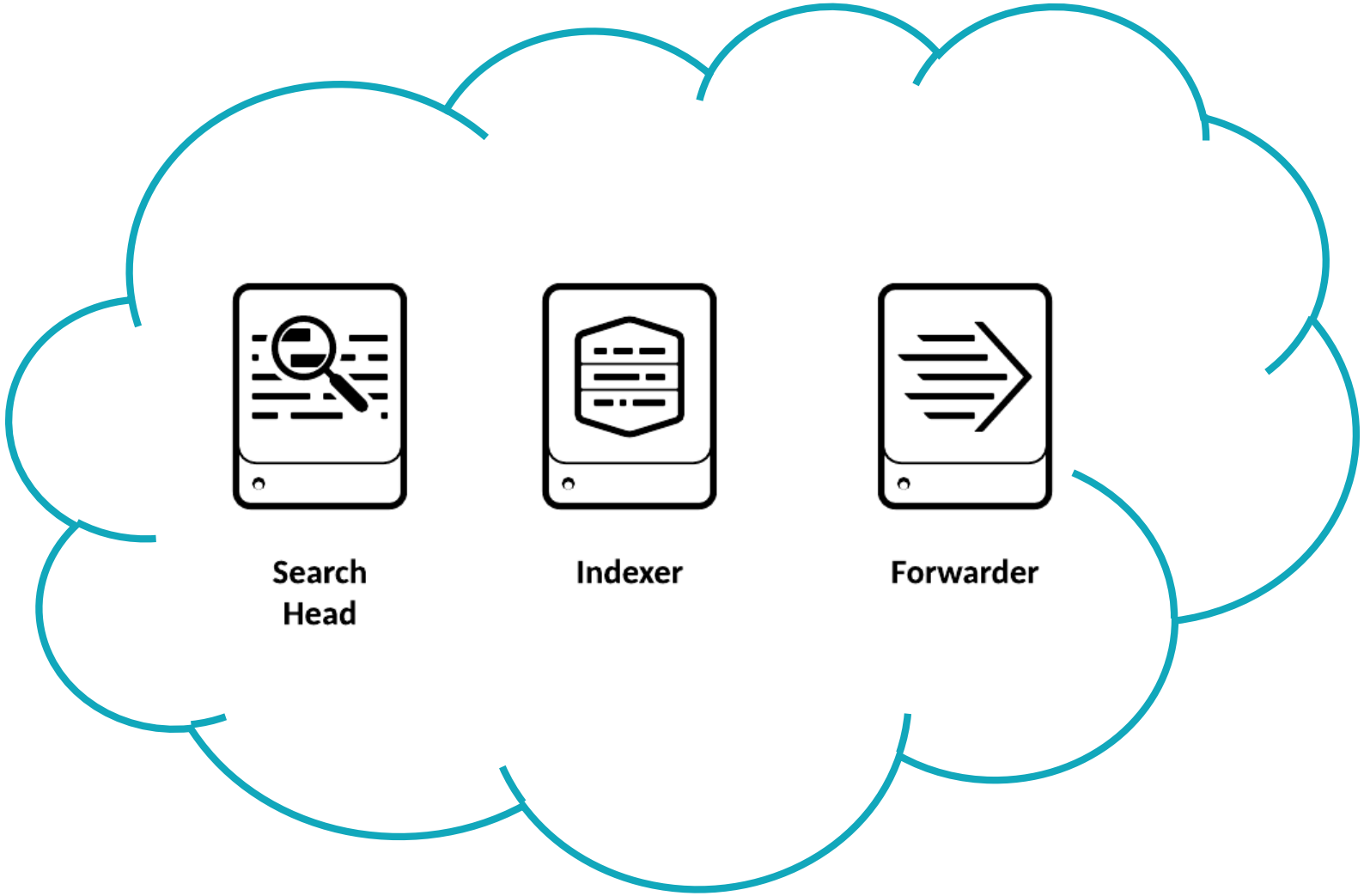
C3/C13 (Distributed Clustered Deployment with SHC – Single Site)

- Search head cluster
  - Deployer
- Indexer cluster
  - Cluster Manager

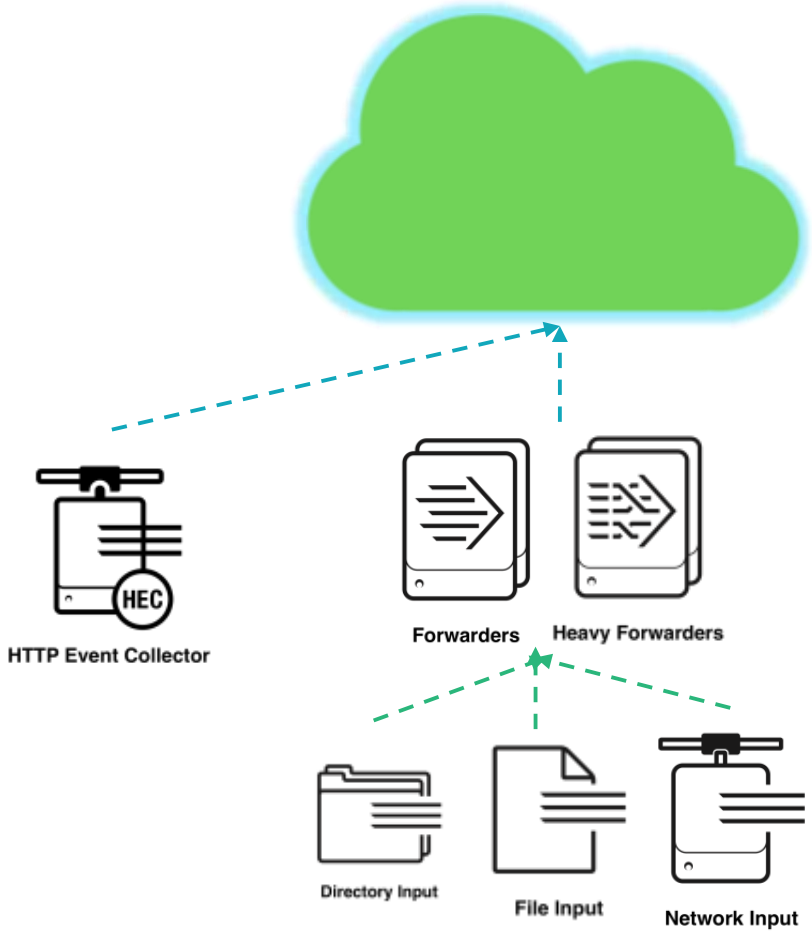


Splunk Cloud Platform

Splunk Cloud Platform



Splunk Cloud Platform Architecture



Subscription Types

- Splunk Cloud is subscription based
- Workload-based subscription is the default
- Ingest-based subscription

Apps in Splunk Cloud

- Only vetted and compatible apps
- Some apps can be self installed through the app browser; others require a support ticket to be submitted
- Private apps are supported, but are vetted by Splunk

How Splunk Stores Data

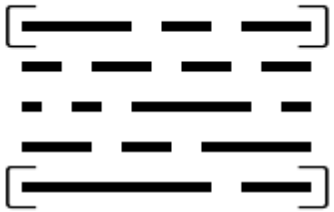
How Splunk Stores Data

Index

- A repository for Splunk data
- Splunk transforms incoming data into events, and stores it in indexes
- An event is a single row of data

host	source	sourcetype	ip_address
web1.company.com	/var/log/	cisco_syslog	10.250.117.14
web2.company.com	udp:514	syslog	10.250.117.17

How Splunk Stores Data

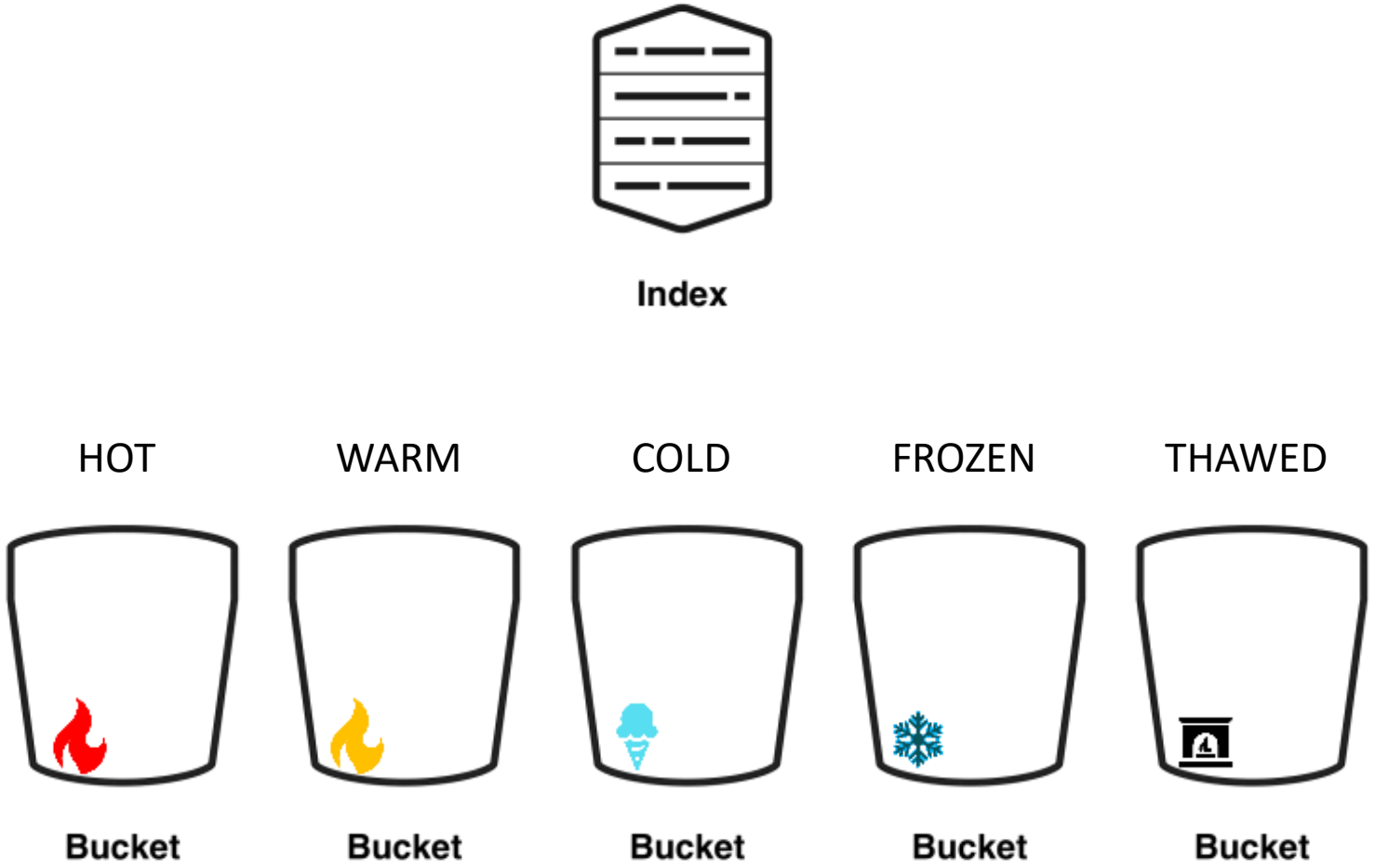


Event

- A single row of data, made up of fields
- Fields are key=value pairs
- Splunk adds default fields to all events
  - `_time`
  - `index`
  - `host`
  - `source`
  - `sourcetype`

```
[Fri Sep 09 10:42:29.902022 2011] [core:error]
[pid 35708:tid 4328636416] [client 72.15.99.187]
File does not exist: /usr/local/apache2/htdocs/favicon.ico
```

How Splunk Stores Data



How Splunk Stores Data

`$SPLUNK_HOME/var/lib/splunk/defaultdb/db/*`

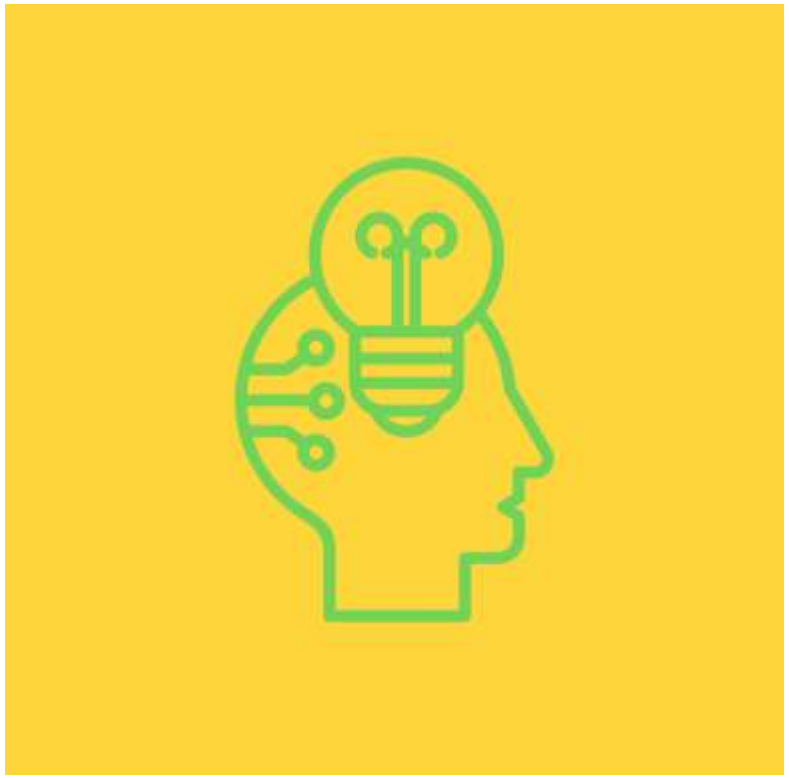
`$SPLUNK_HOME/var/lib/splunk/defaultdb/db/*`

`$SPLUNK_HOME/var/lib/splunk/defaultdb/colddb/*`

(Location that you specify for archival purposes)

`$SPLUNK_HOME/var/lib/splunk/defaultdb/thaweddb/*`

SmartStore



- Allows you to use remote object stores like AWS S3
- Most data resides on remote storage while the indexer maintains a local cache (hot buckets)

Splunk Enterprise Licensing

Two Types of Enterprise Licensing

Volume-based

Infrastructure

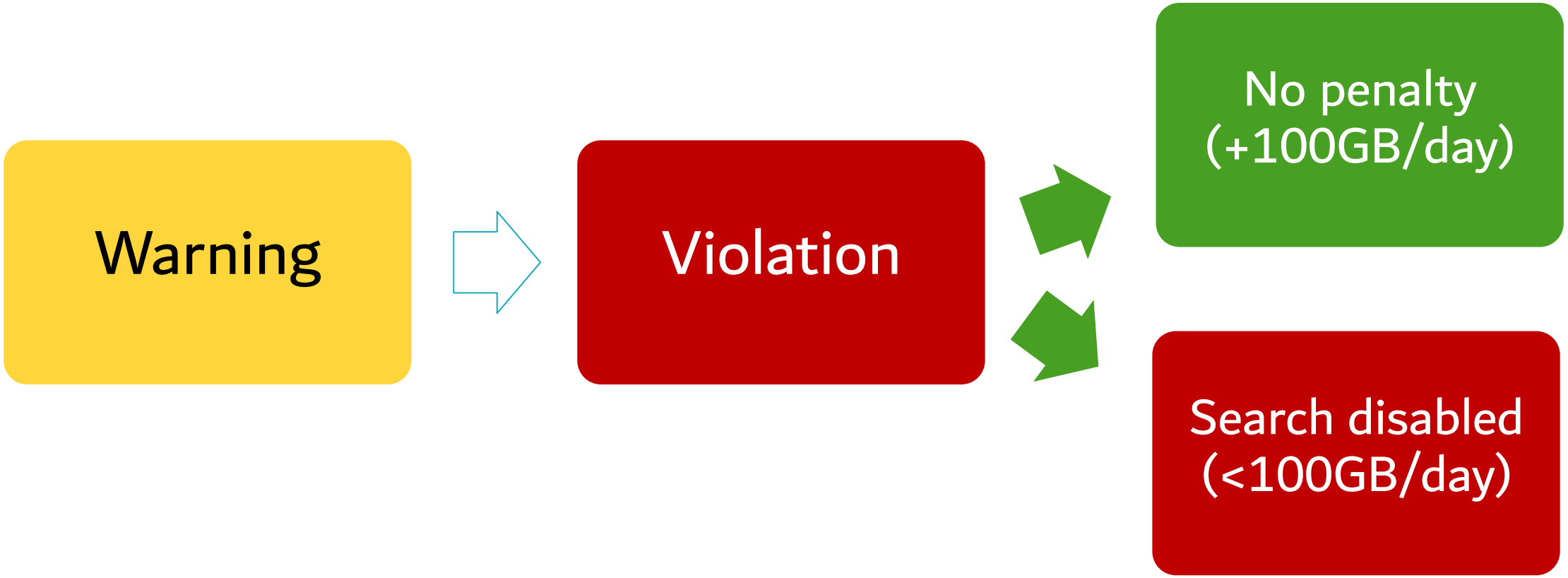
Volume-based Licensing

- Based on you data indexed per day, not data stored
- Daily indexing volume is measured from midnight to midnight by the clock on the license manager

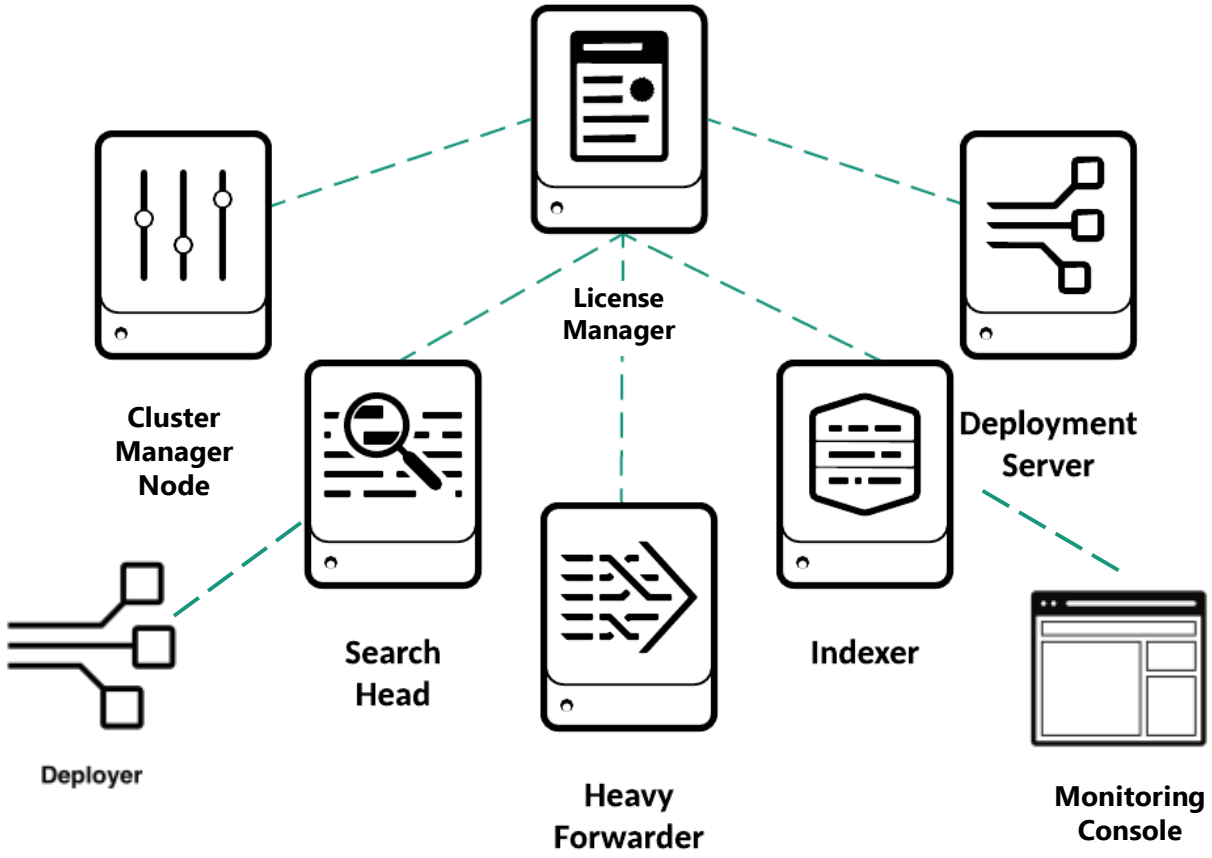
Infrastructure Licensing

- Based on virtual CPUs (vCPU)
  - Physical or logical core, virtual core, etc.
  - Splunk uses the CPUs reported by the OS
- All search heads and indexers count towards vCPU capacity

Splunk Volume-based Licensing



Which components need a license?

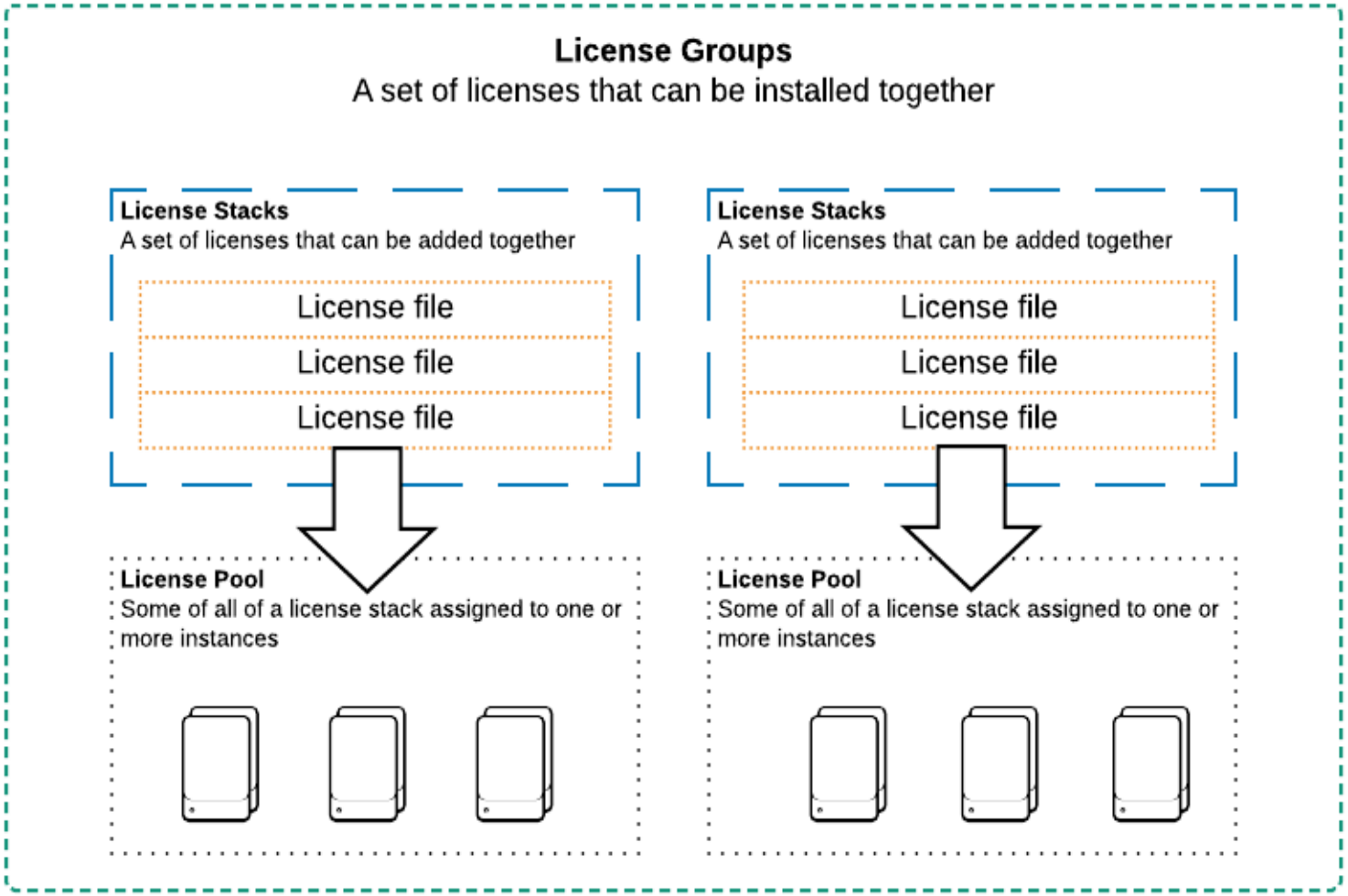


License Pooling

- License pools are created from license *stacks*
- Pools are sized for specific purposes
- Managed by the license manager
- Indexers and other Splunk Enterprise components are assigned to a pool

Pool

Splunk Licensing



Configuration Files

A Note on Configuration Files

- Everything Splunk does is governed by configuration files
- Configuration files are stored in `/etc`, and they have the `.conf` extension
- Configuration files are layered
  - You can have `.conf` files that have the same name in different directories
  - Splunk determines which one to use based on the current app
- The `/etc/<app>/default` directory contains preconfigured versions of .conf files
- The `/etc/<app>/local` directory is where custom configurations are stored

Configuration File Structure

```
[Stanza]
Attribute = Value

[Stanza]
Attribute = Value
```

Splunk Apps

Splunk Apps

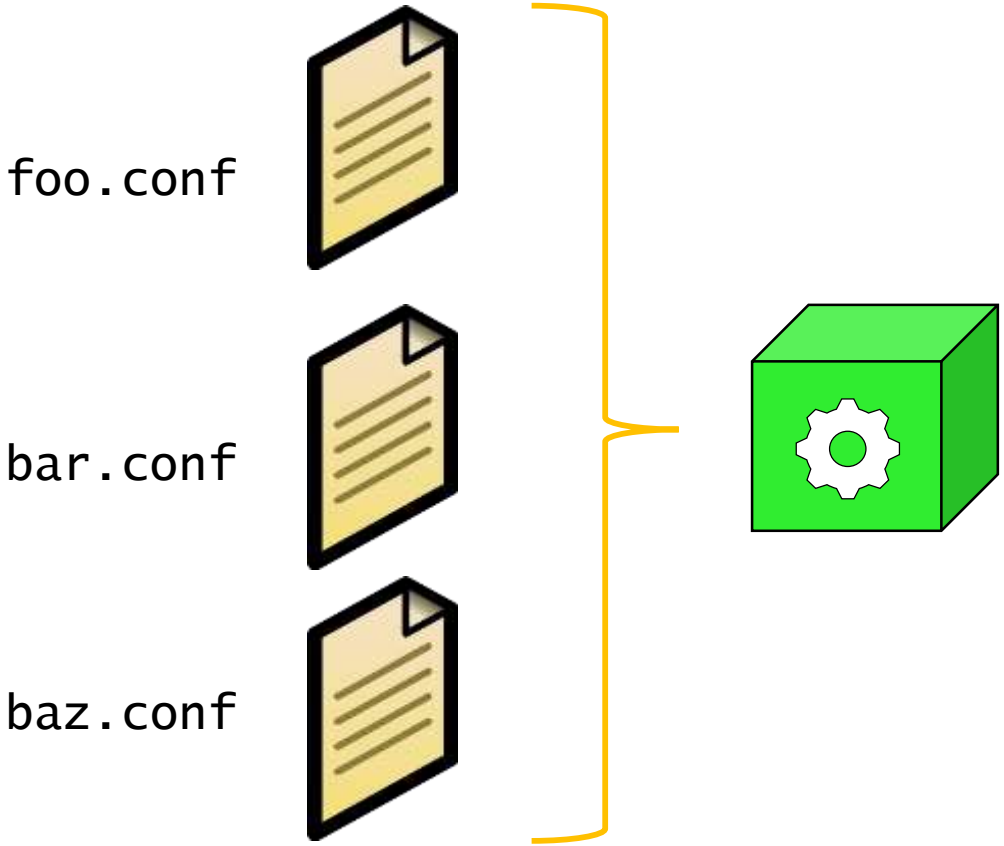
Apps

- Visualization
- Analysis
- Reports & dashboards
- User interface

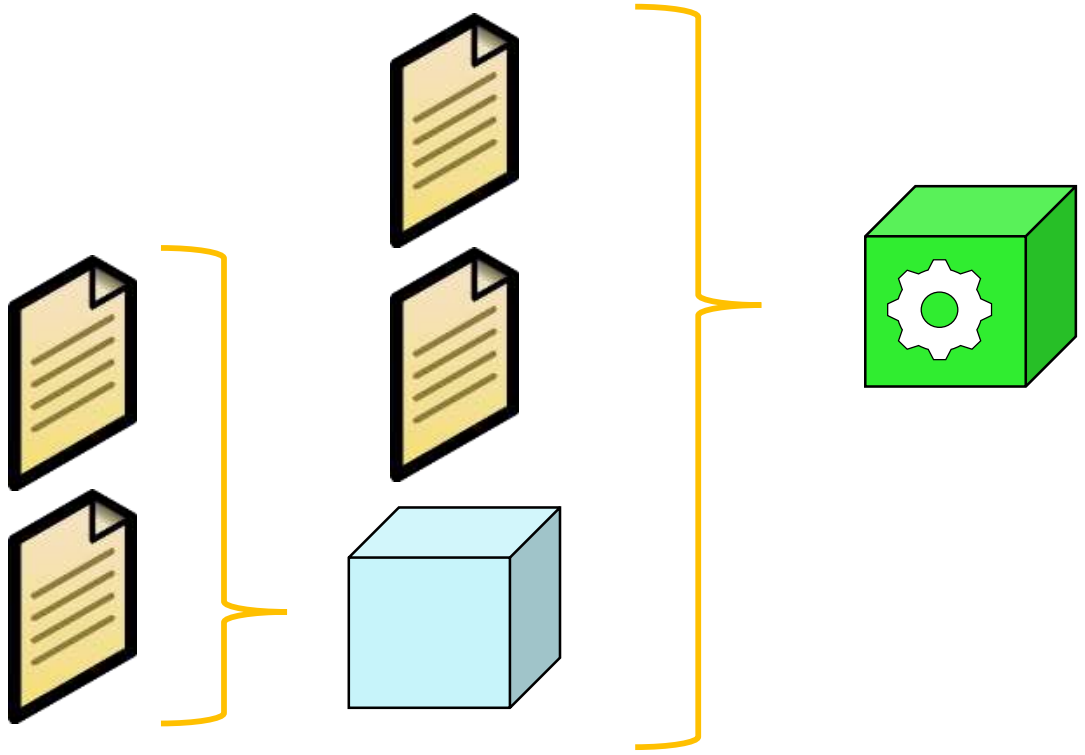
Add-ons

- Data enrichment
- Tags
- Data models
- Datasets

Splunk Apps



Splunk Apps



Summary

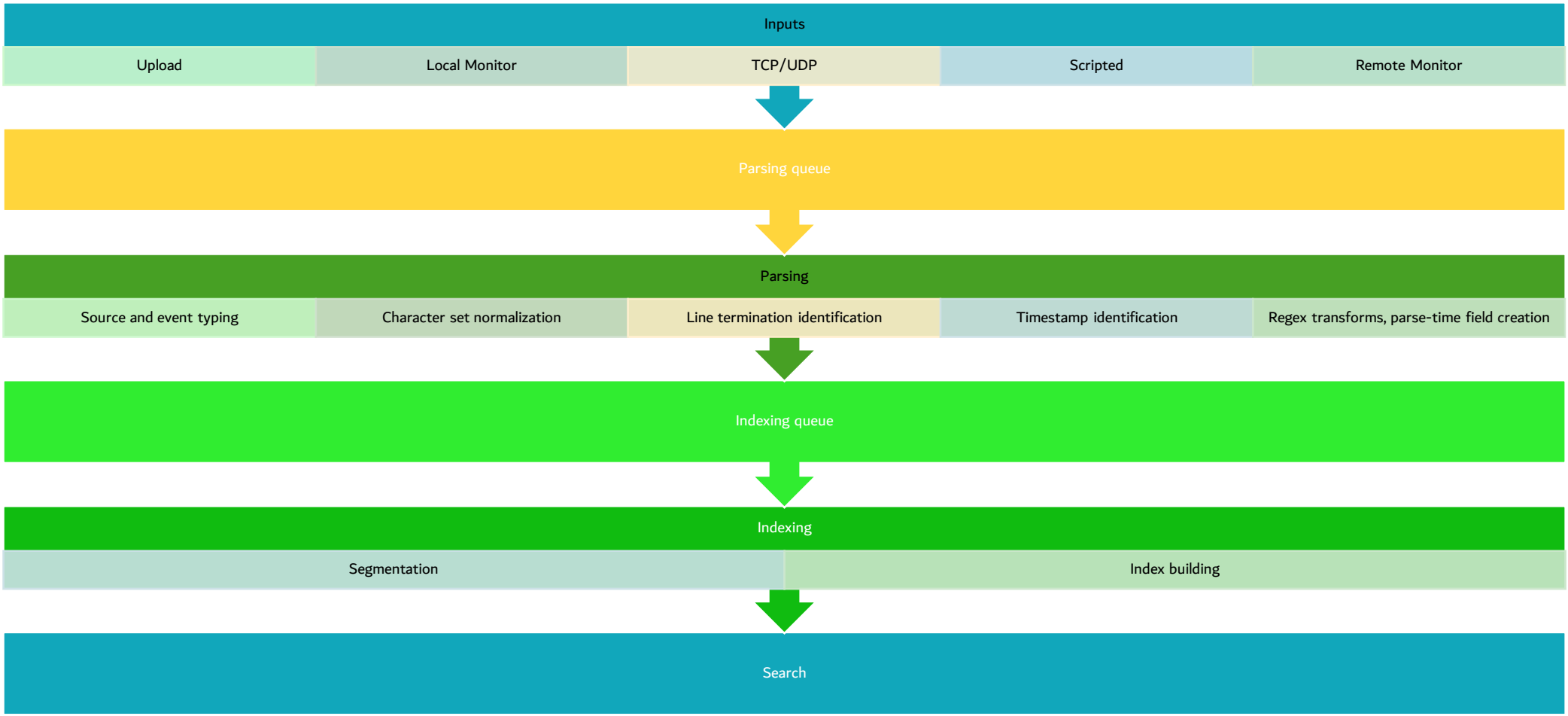
- Splunk deployment topologies and when to use them
- Indexes and storing data
- Licensing
- Extending Splunk with apps and add-ons

Getting Data in

- Describe the basic settings for an input
- List Splunk forwarder types
- Configure a forwarder
- Add an input to a Universal Forwarder using the CLI

Splunk Inputs

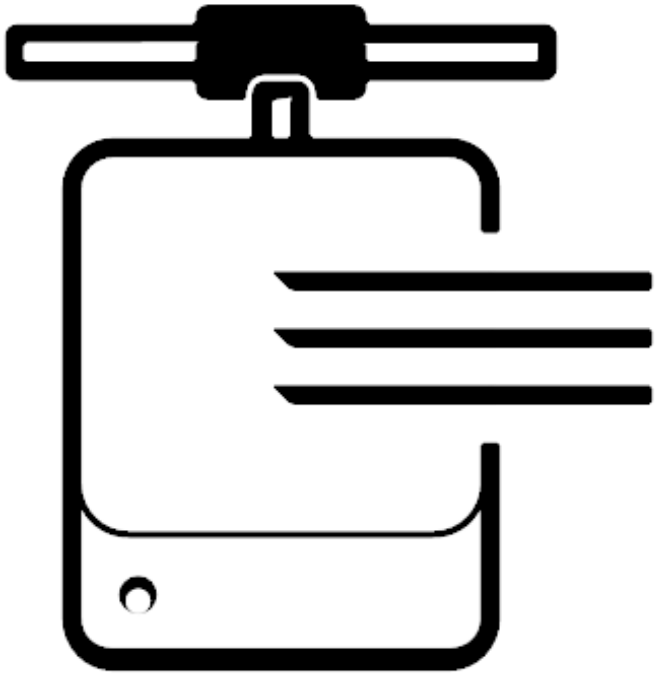
The Splunk Data Pipeline



General Input Categories

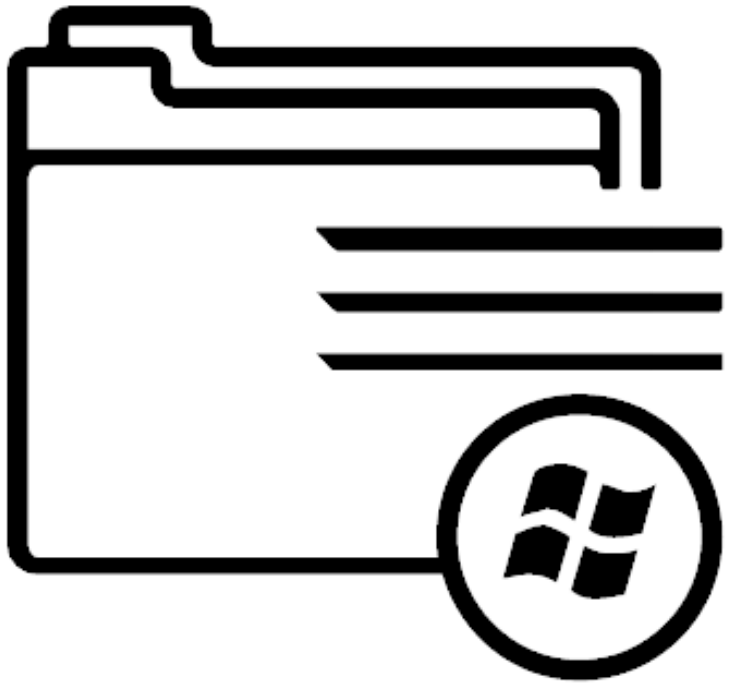
- File and directory inputs
  - Monitor files and directories
    - Locally and remotely
    - Monitor compressed files
  - Upload
    - Upload files to Splunk
    - Used for one-time analysis
  - MonitorNoHandle
    - Available for Windows hosts only
    - Monitors files and directories that the system rotates automatically

General Input Categories



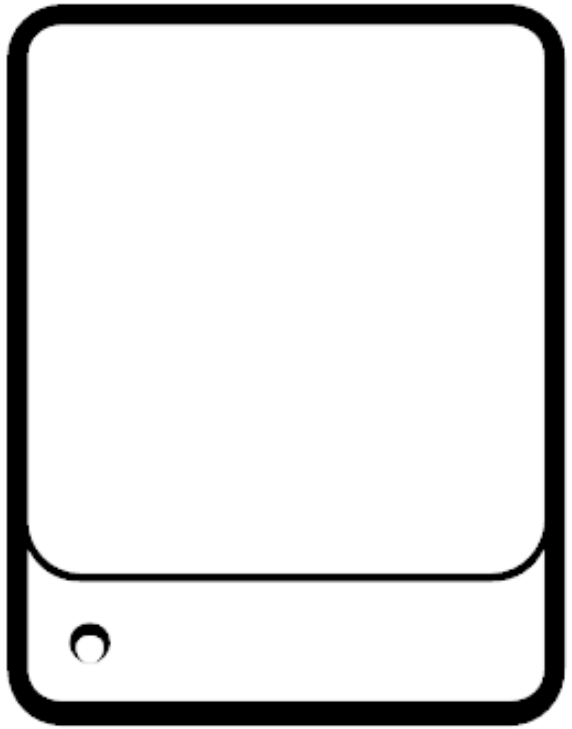
- Network inputs
  - Data from TCP and UDP
    - syslog
  - Data from SNMP events

General Input Categories



- Windows inputs
  - Windows event logs
  - Registry
  - Active Directory
  - WMI
  - Performance monitoring (perfmon)

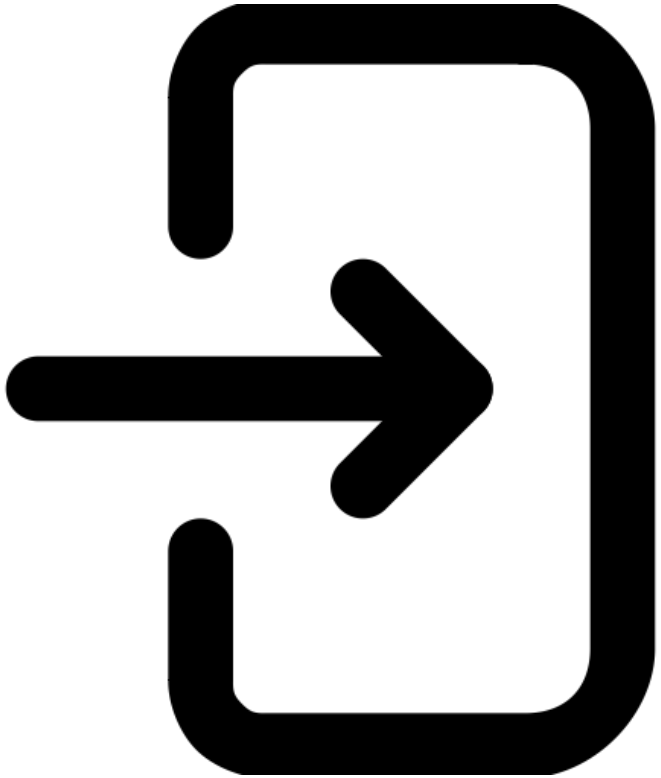
Other Data Sources



- Metrics
- Scripted inputs
- Modular inputs
- HTTP event collector

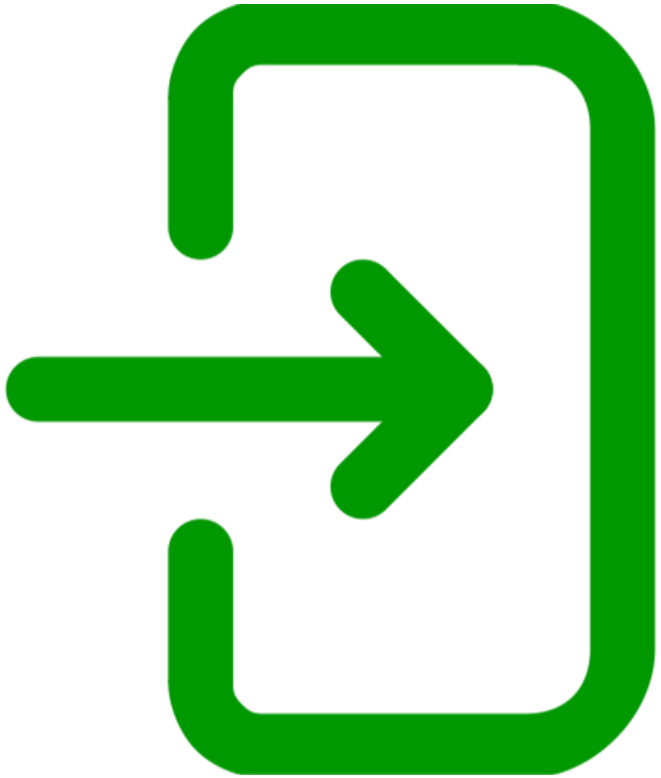
Basic Settings for an Input

Ways to Configure Inputs



- Through an app
  - Many apps have preconfigured inputs
- Splunk web
  - Settings > data inputs
  - Settings > add data
- CLI
  - `./splunk add monitor <path>`

Ways to Configure Inputs



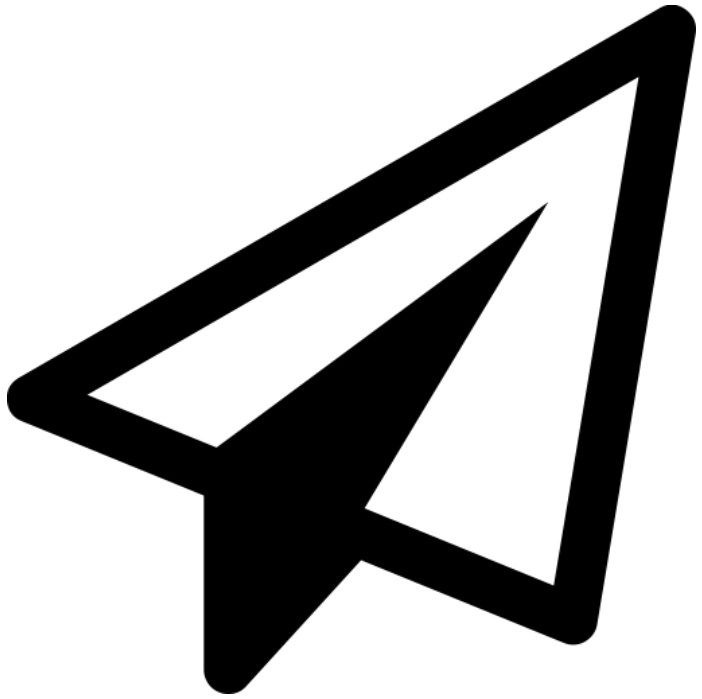
- Through `inputs.conf`
  - Add a stanza for each input
- Guided Data Onboarding (GDO)
  - Data input wizard

Splunk Forwarder Types

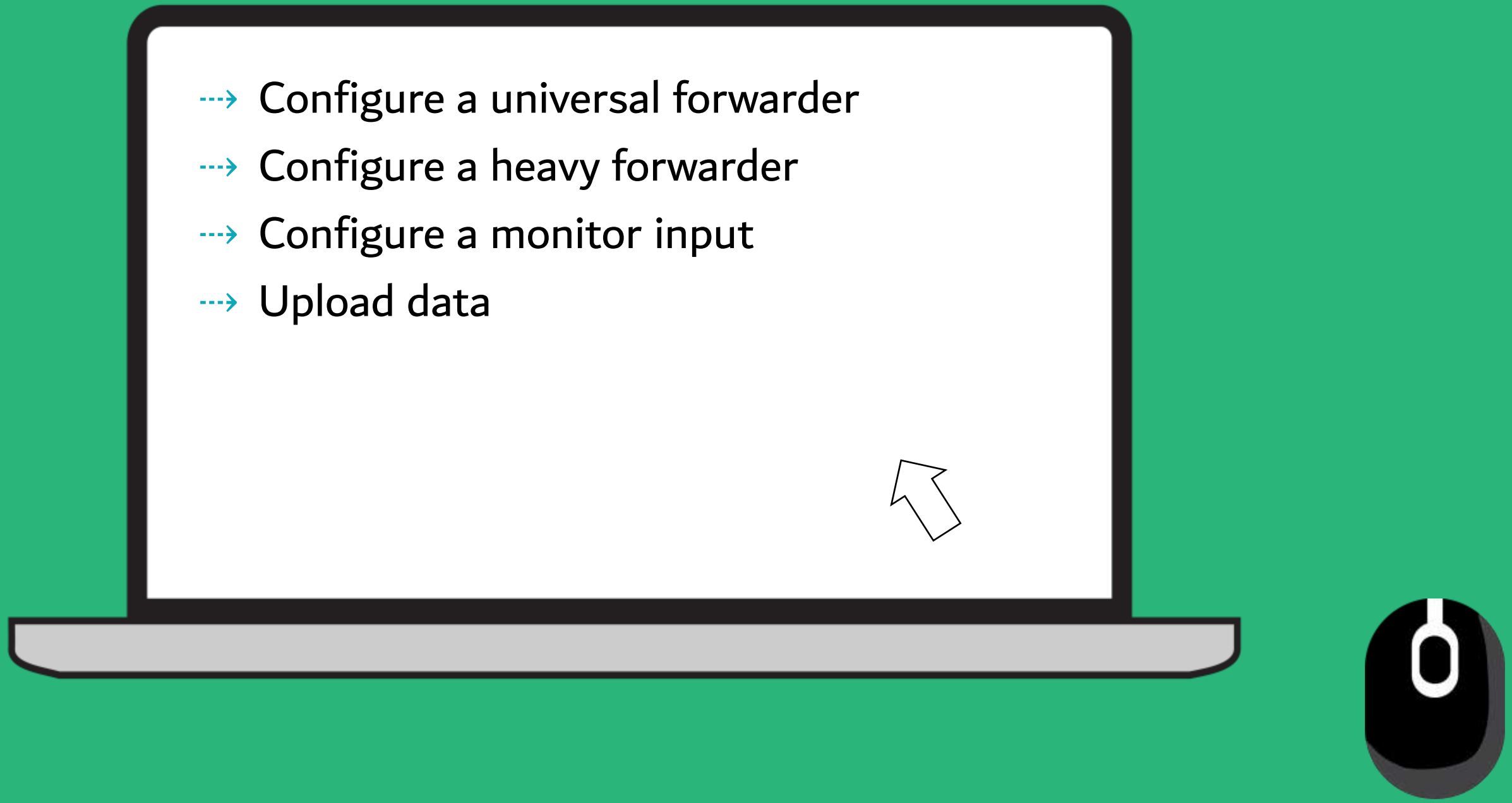
Universal

Heavy

Universal Forwarder Configuration Steps



- Configure receiving on a Splunk Enterprise instance
- Download and install the UF
- Start the UF
- Configure the UF to send data
- Configure the UF to collect data from the host system


Configure a universal forwarder
Configure a heavy forwarder
Configure a monitor input
Upload data

Summary

- Splunk data pipeline
- Configuring inputs
- Types of forwarders
- Configuring universal forwarders
- Configuring heavy forwarders

Searching in Splunk

- What is Search Processing Language (SPL)?
- The Anatomy of a Search
- Time
- Basic Search Processing Language (SPL)
- Fields and Field Extractions

What is Search Processing Language?

SPL encompasses all the search commands and their functions, arguments, and clauses. Its syntax was originally based on the Unix pipeline and SQL. The scope of SPL includes data searching, filtering, modification, manipulation, insertion, and deletion. – The Splexicon

What Does it Mean to Search in Splunk?

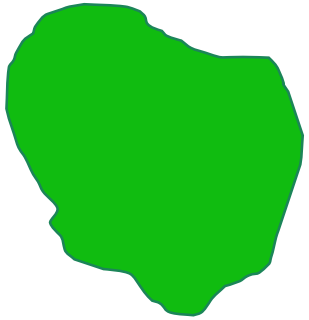
- It's the primary way users interact with data in Splunk.
  - Query
  - Calculate
  - Transform
  - Organize
  - Visualize
  - Manipulate

Search and Reporting App

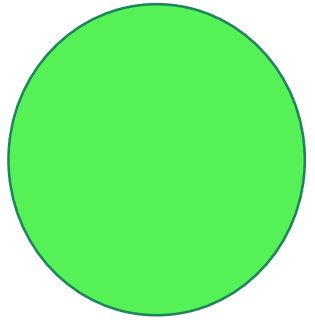
- Default app
  - Comes built-in to Splunk
- Primary way to search and analyze data in Splunk
  - Index data
  - Build reports and visualizations
  - Configure alerts
  - Create dashboards

The Anatomy of a Search

Big glob of events

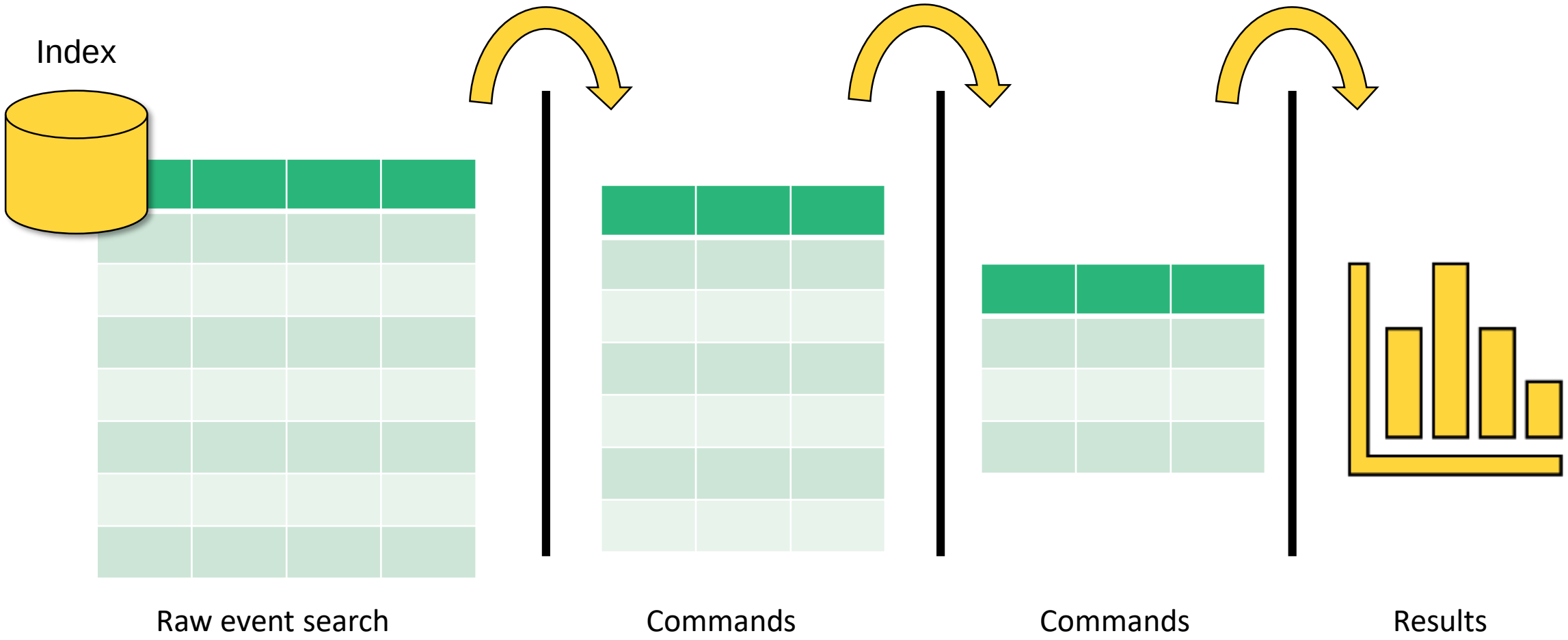


Intermediate glob of events



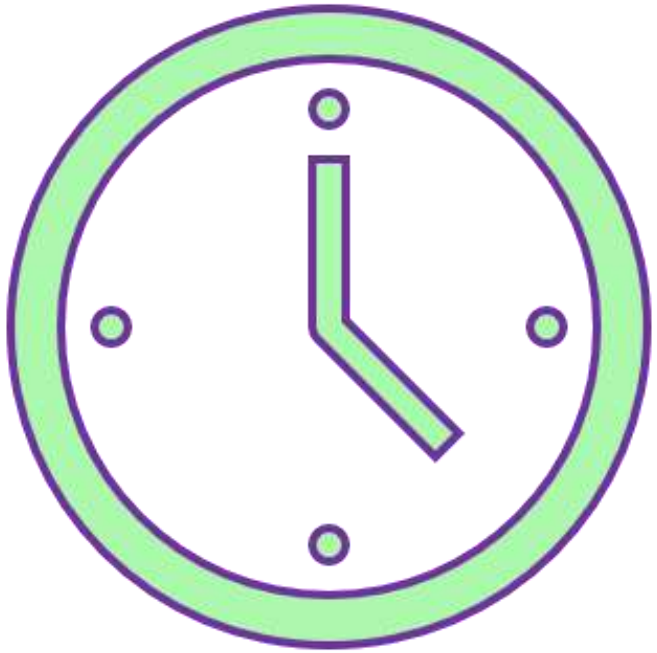
The events we want in the format we want

Pipe: Take the previous data, do something to it, then output it to the next step



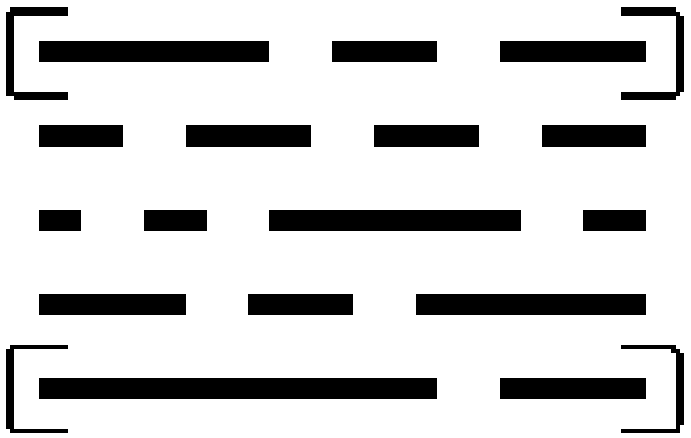
Time

Timestamps



- Timestamps are converted to UNIX time and stored in the `_time` field
- Splunk assumes that any data indexed is in the time zone of the Splunk instance

The _time Field



- A default, and essential, field
- Values in the _time field are stored in UNIX time
- In Splunk web, the _time field appears in human-readable format
- Use search commands to manipulate the time format

Looks for timestamps in the event data

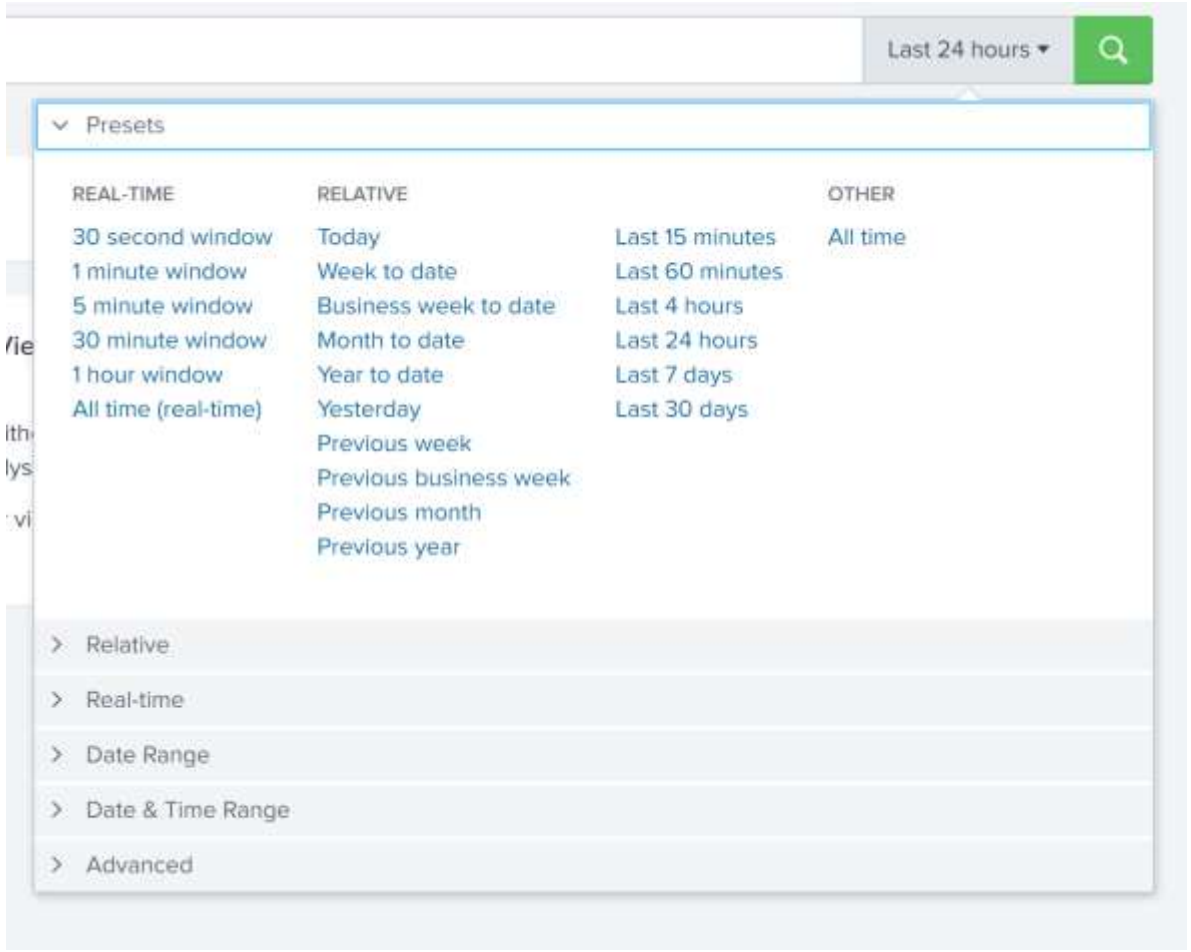
Source name or file name

File modification time

Current system time

The Time Range Picker

- Splunk uses time stamps for the time range picker
- The defaults for the time range picker are Real-Time, Relative, and Other
- We can also do some more granular things with time selection



Specify Absolute Time Ranges Using SPL

- Specify time ranges directly in your search
- For absolute ranges:

```
earliest=<%m/%d/%Y:%H:%M:%S> latest=<%m/%d/%Y:%H:%M:%S>
```

```
earliest=01/14/2021:16:32:00 latest=07/14/2021:21:00:00
```

Specify Absolute Time Ranges Using SPL

- Specify time ranges directly in your search
- For relative ranges:
  - + or – to indicate the offset
  - Number
  - Time unit (years `y`, quarters `q`, months `mon`, weeks `w`, days `d`, hours `h`, minutes `m`, seconds `s`)

`-30m`	`30 minutes ago`
`-7d`	`7 days ago`
`+1d`	`1 day from now`

Time Variables

- Format time using time variables
- Useful when evaluating time and specifying time in SPL

Variable	Description
%c	Date and time in the format of the server
%H	Hour (24-hour clock)
%I	Hour (12-hour clock)
%M	Minutes (00 – 59)
%p	AM or PM
%S	Seconds (00 – 59)

Time Variables

- Format dates using variables

Variable	Description
%F	Date in ISO 8601 format (yyyy-mm-dd)
%A	Full weekday name (Monday)
%d	Day of month (01 – 31)
%j	Day of year
%B	Full month name (January)
%m	Month (01 – 12)
%y	Year as a two-digit number (00 – 99)
%Y	Year as a four-digit number (yyyy)

Converting Time Using `strftime`

- We can convert time into the format we want during search time using and `eval` expression, `strftime`, and time variables on the `_time` field

```
| eval <new field> = strftime(<time field>, "<format>")
```

```
| eval New_Time = strftime(_time, "%I:%M, %p")
```

POP QUIZ!

It is currently Monday, January 1st, 1900 at 5:00 PM. How will the following string format this time and date?

String	Answer
%A, %B %d, %Y - %I:%M %p	

Basic Searching

Basic Searching

Broad search terms - metadata

- Index
  - `index = main, index = default`
- Host
  - `host = server.com, host = 192.168.1.1`
- Source , sourcetype
  - `source = /var/lib, sourcetype = csv`

Basic Searching

Broad search terms

- Keywords
  - `failed, error`
- Phrases
  - `“failed login”`
- Fields
  - Key value pairs
  - `user=user1.domain.com`
- Wildcards
  - `*ailed, fail*, user=*`
- Booleans
  - `AND, OR, NOT`

Basic Searching

Basic search commands

- `chart / timechart`
  - Returns results in tabular output for charting
- `rename`
  - Renames a specific field
- `sort`
  - Sorts results by specified fields
- `stats`
  - Statistics
- `eval`
  - Calculates an expression
- `dedup`
  - Removes duplicates
- `Table`
  - Builds a table with specified fields

Basic Searching

Constructing a basic search

Search Terms | Commands

```
host=myhost.lcl  source=hstlogs  user=* (message=fail* OR message=lock*)
| table _time user message
| sort -_time
```

Basic Searching

_time	user	message
2021-03-28-7:10:07	user1.domain.com	failed log on
2021-03-28-7:17:00	user1.domain.com	failed log on
2021-03-28-7:17:00	user1.domain.com	locked

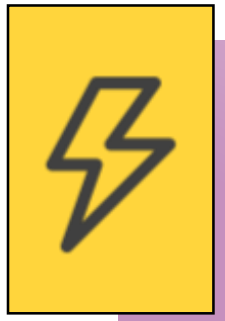
Fields

Fields

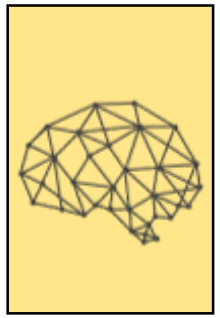
- Searchable key – value pairs
- Key = value
  - `user = user1`
  - `ip_addr = 192.168.1.1`
  - `message = error`
  - `host = websvr.com`

Field Discovery

- Splunk automatically discovers fields
  - Default fields
    - `host, source, sourcetype, _time,` etc.
  - Obvious key-value pairs
  - Field extractions



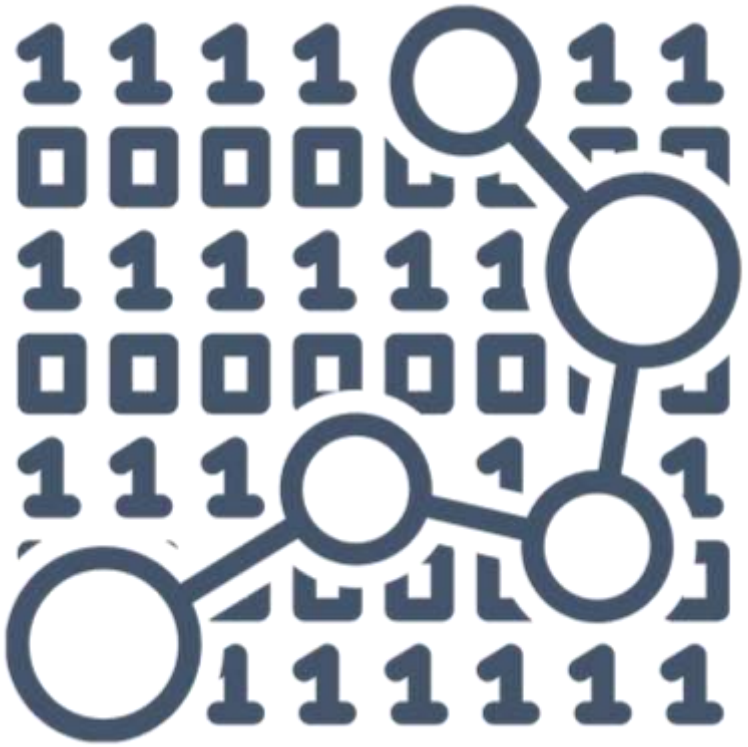
Fast



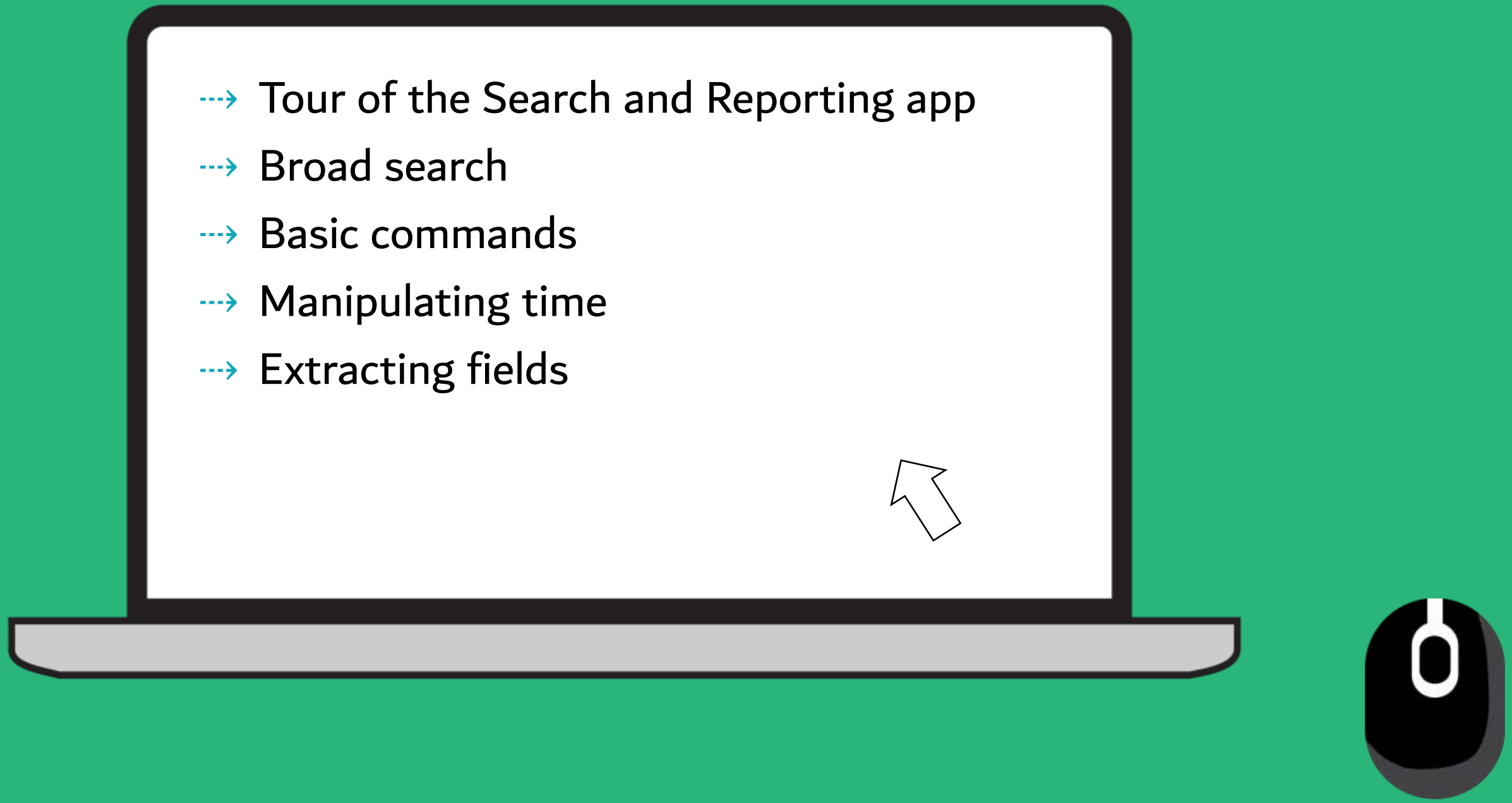
Smart

Verbose

Field Extractions



- Custom field extractions can be built using the Splunk field extractor
- Uses regular expressions (regex) to extract fields based on patterns


Tour of the Search and Reporting app
Broad search
Basic commands
Manipulating time
Extracting fields

Summary

- What SPL is
- The anatomy of a search
- Timestamps and dealing with time
- Fields

Visualizing Your Data

- The basics of visualization
- Modeling data using data models
- Reporting and alerting
- The Pivot tool

The Basics of Visualization

Success in data visualization does not start with data visualization.

Cole Nussbaumer Knaflic, Storytelling with Data: A Data Visualization Guide for Business Professionals

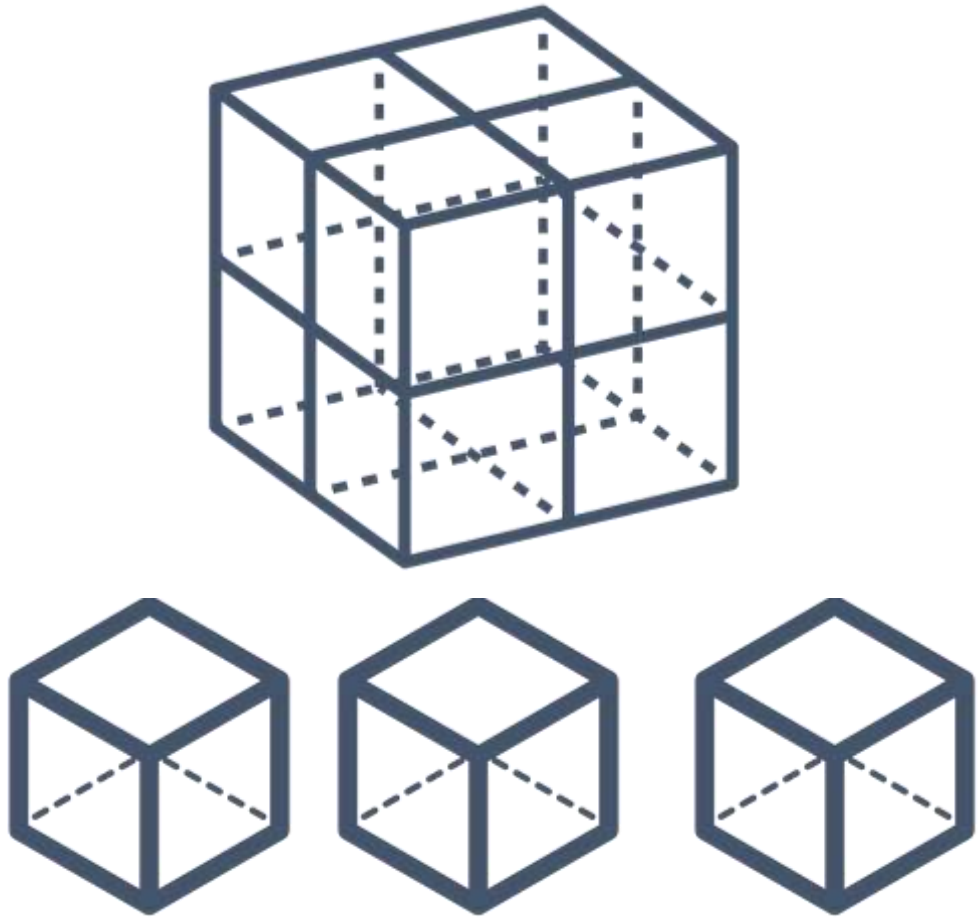
Why Visualize Data?

- Data alone is not very interesting to look at
- Non-technical people might not be interested in tabular data
- Visualizing data is more human centric
  - Humans are pattern recognizing machines
  - From an early age, we are taught to visualize data
  - Visualizations are art, and therefore have an emotional impact on us

Dimensions

Dimension

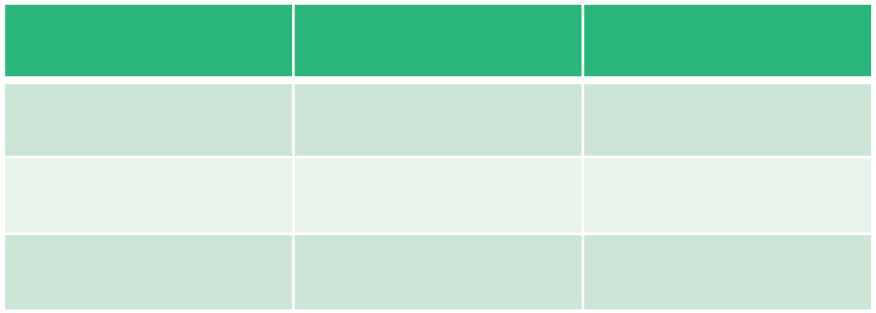
Something you want to measure

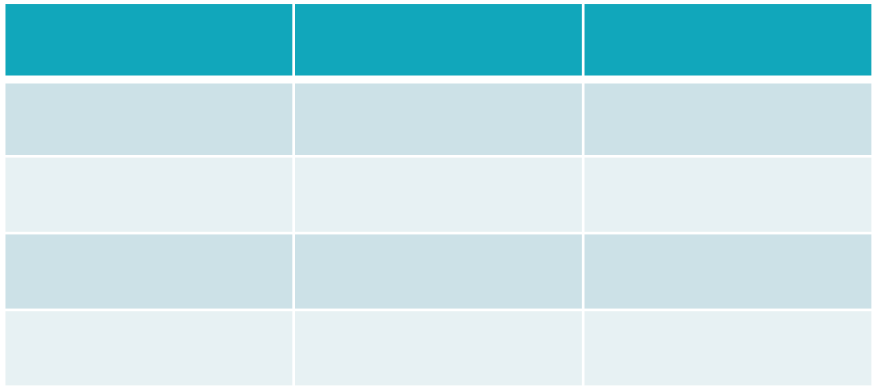


Dataset

Time/Date	Customer	Purchase Price

Data Model

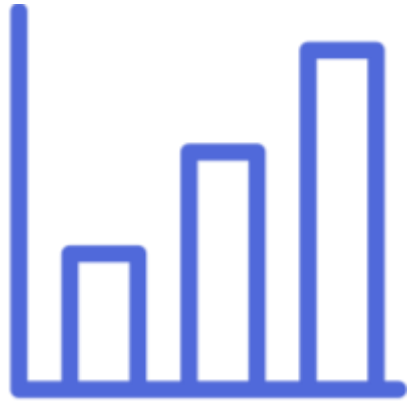


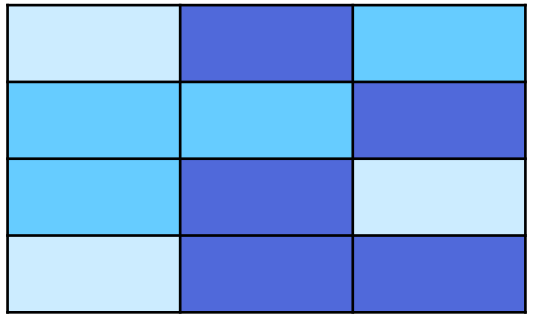


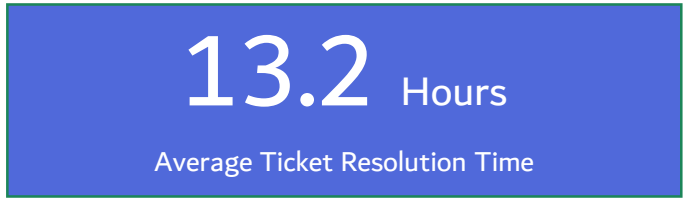
Data Visualization Context

- Understand your audience
  - Executives usually have different needs than individual contributors
  - Publicly available dashboards might display different data than internal-only reports and dashboards
- Understand your own goals
  - What is the message for which you are using this visualization?
  - Like the famous “author’s purpose” consider the PIE: do you want to persuade, inform, or entertain?

Types of Visualizations in Splunk







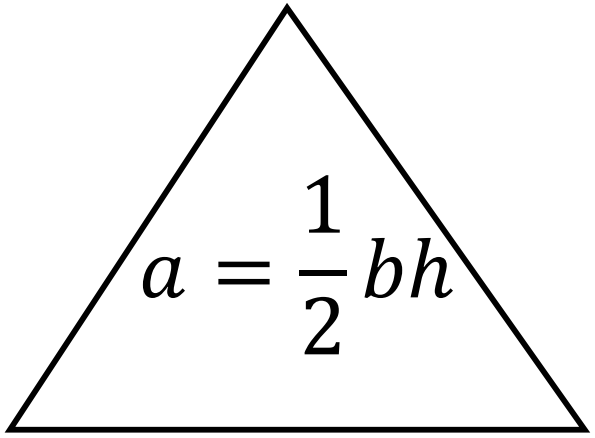
Data Models

Data Models

- Make machine data easier to use
- Simplify complex data through abstraction
- Group specific types of data

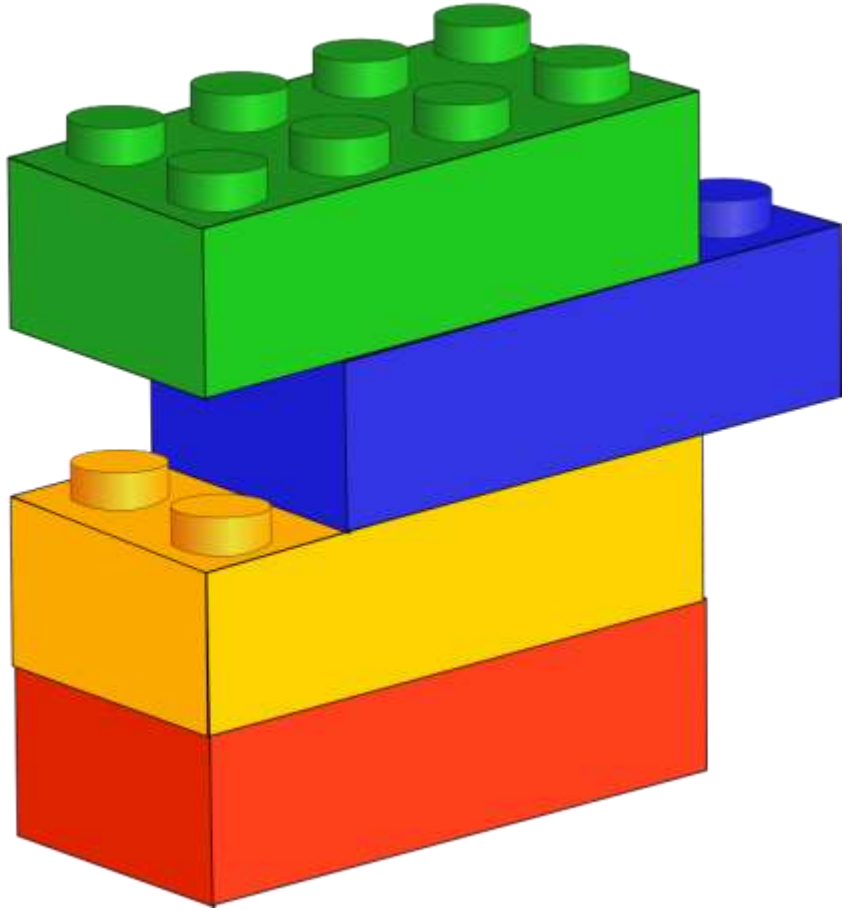
Data Models

We see data models all the time in the “real world.”



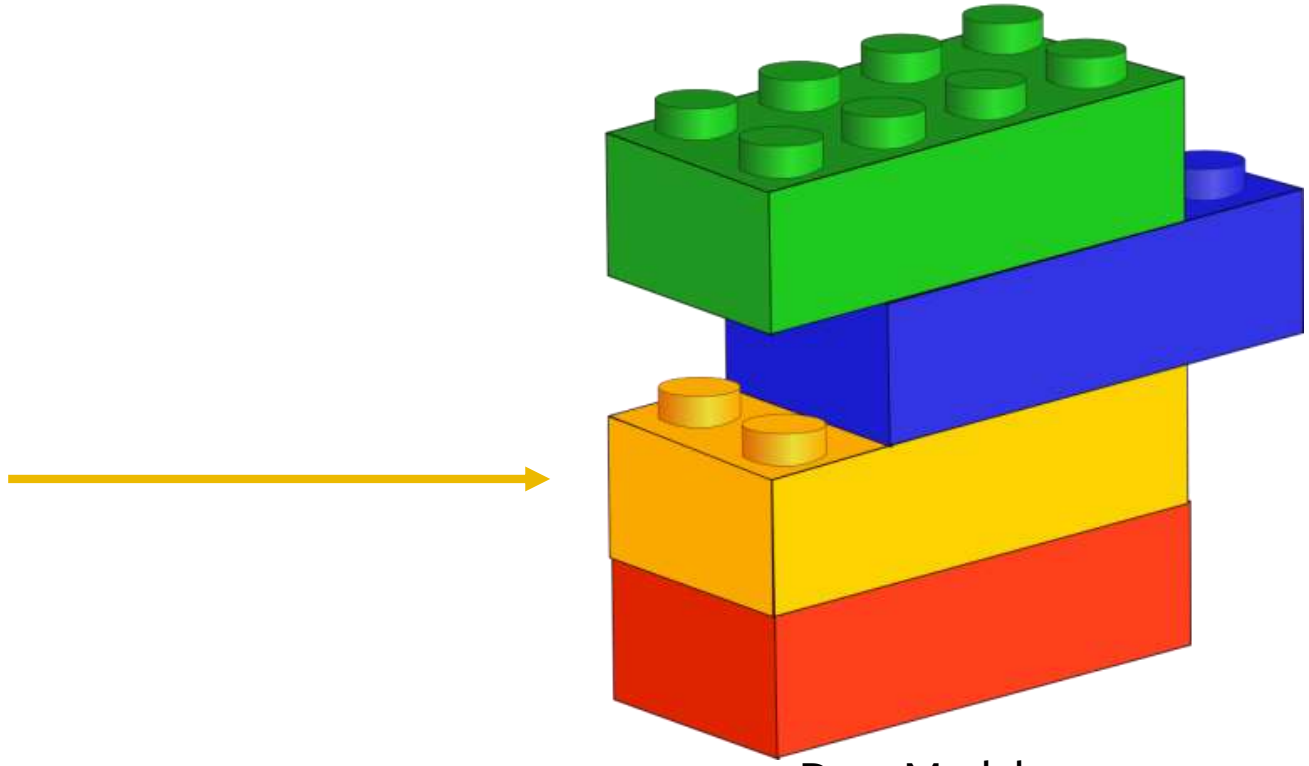
Data Models

- Stacks of datasets
- Datasets are stacks of knowledge objects
- Knowledge objects include
  - Saved searches, field extractions, tags, and more



Data Models

- Root Events
  - Event Constraints
  - Event Knowledge
  - Child Events
    - Constraints
    - Knowledge Objects
- Root Searches
  - Search Constraints
  - Search Knowledge
  - Child Searches
    - Constraints
    - Knowledge Objects



Data Model

Data Models

- **Events** – most commonly used
  - Event constraints – Must include an index constraint (`index=`)
- **Searches** – Spunk saved searches that include transforming commands, etc.
  - Search constraints – Constrained to the full search string (Must include an index constraint (`index=`))
- **Transactions** – Combine multiple events from one or many sources into a single event
  - Transaction constraints – must be legally formed transaction search

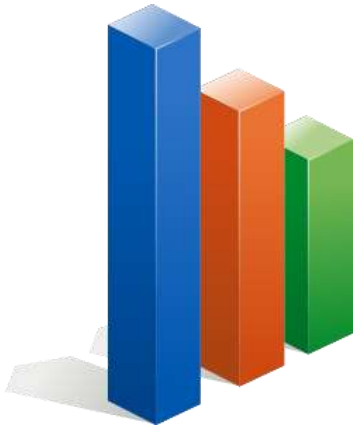
Data Models

- Fields
- Can be added to roots and children
- Children inherit all fields from their parent
  - Auto extracted – Splunk automatically discovered fields
  - Eval – A field generated as a result of an `eval` expression
  - Lookup – Fields that are the result of a lookup
  - Regular expression – Fields extracted by `regex`

Reporting and Alerting

Reporting and Alerting

- Reports and alerts are knowledge objects in Splunk
- To create reports and alerts, you need a Splunk Enterprise license
  - The free license disables these features



Reporting and Alerting

Reports

- Saved searches that can run on a schedule and perform an action
  - Send an e-mail to report consumers
  - Embed on a web page
  - Update a dashboard panel
  - Run a script

Reporting and Alerting

Reports

- Scheduled reports can run
  - Every hour
  - Every day
  - Every week
  - Every month
  - On a chron schedule that you define
- You can stagger the report running window
  - Useful if you have a lot of reports running at the same time

Reporting and Alerting

Alerts

- Can be scheduled or in real-time
- Triggered when the results of a search meet a specific condition that you define
  - For example, if the search `host=firewall1 user=* authentication=failed` returns anything, trigger an alert

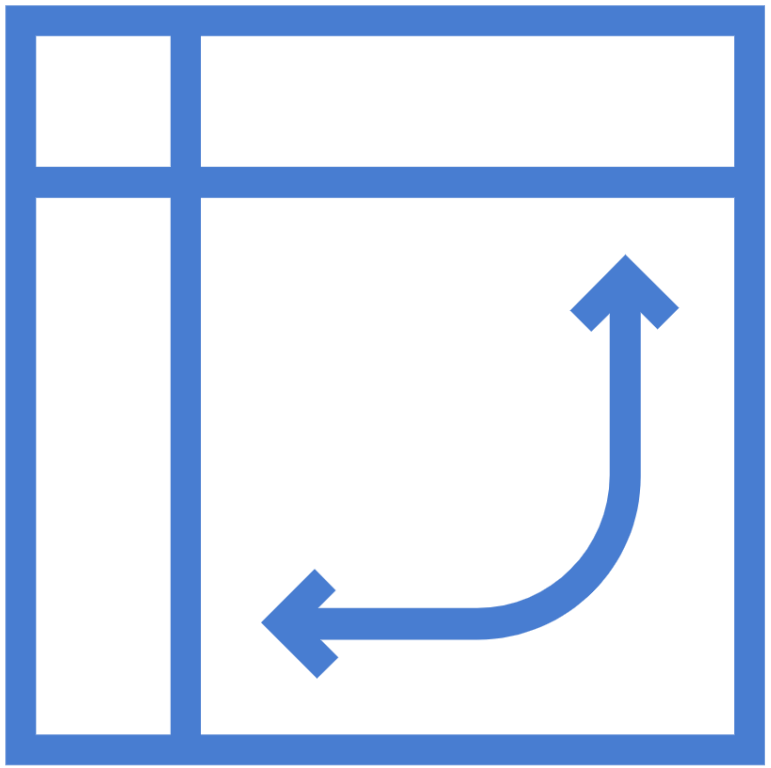
Reporting and Alerting

Alerts

- Alert actions can include
  - Send an email
  - Trigger a script
  - Use a webhook
  - List in triggered alerts
  - Use an app (like PagerDuty or Slack)

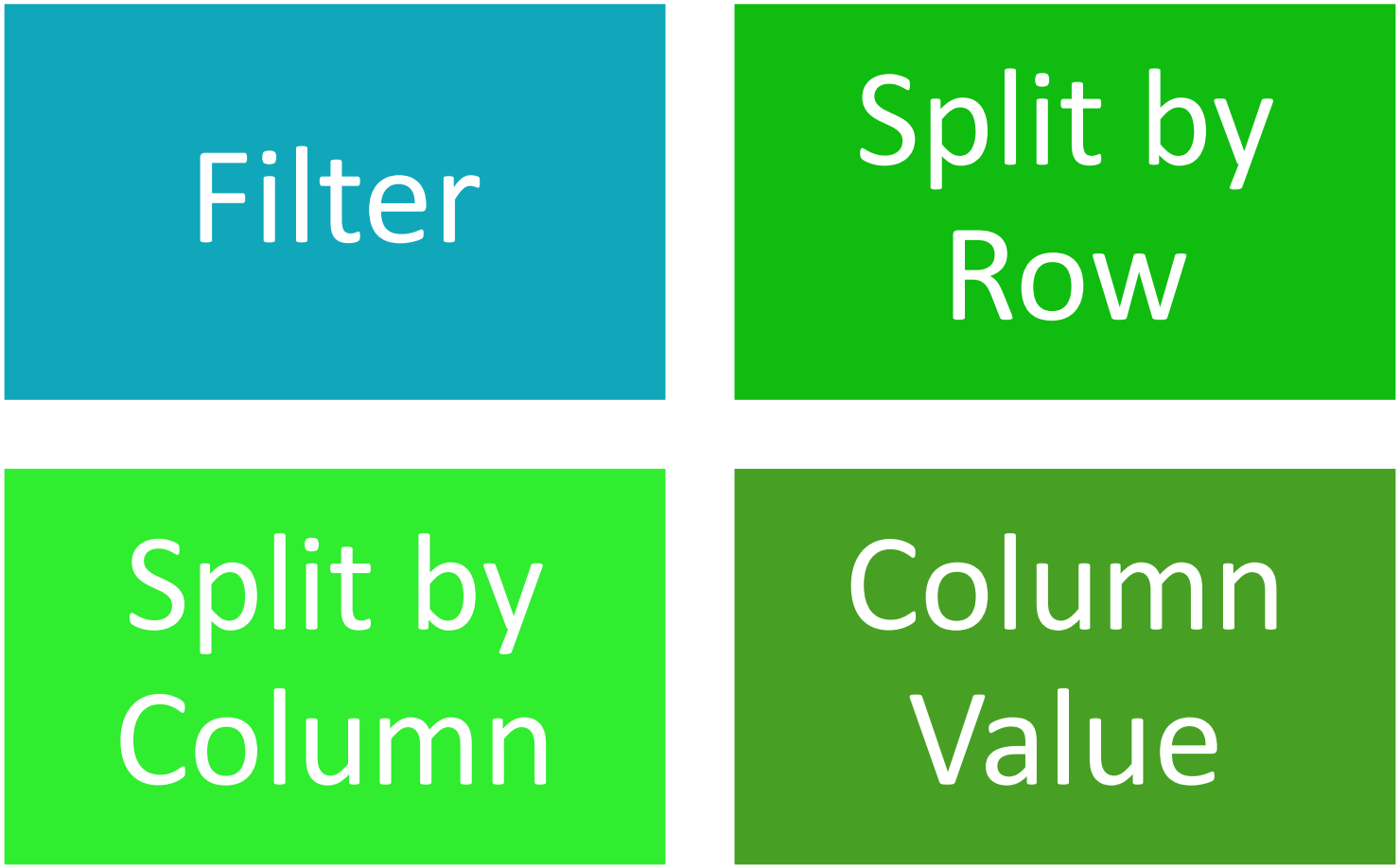
The Pivot Tool

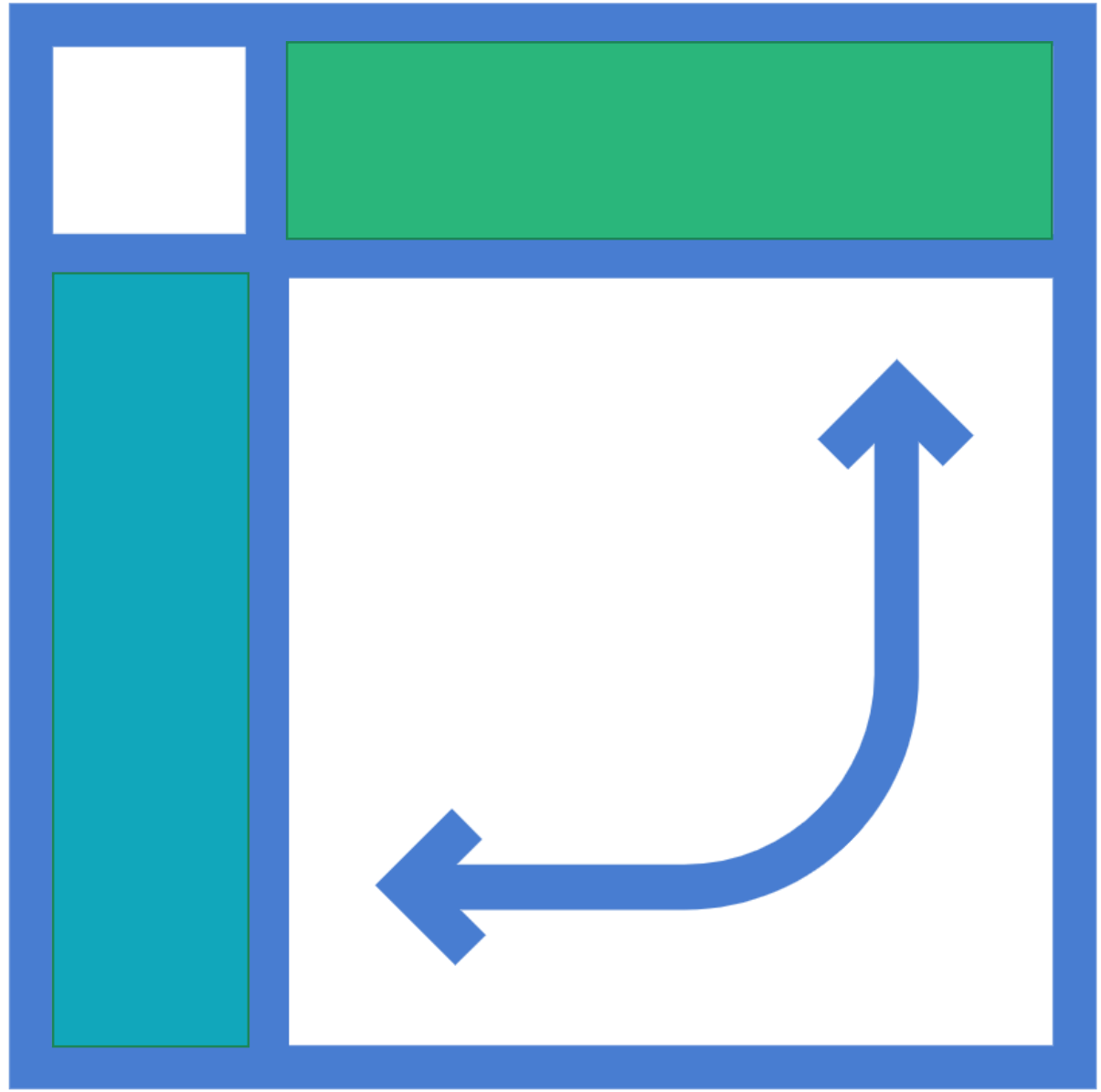
Why Use the Pivot Tool

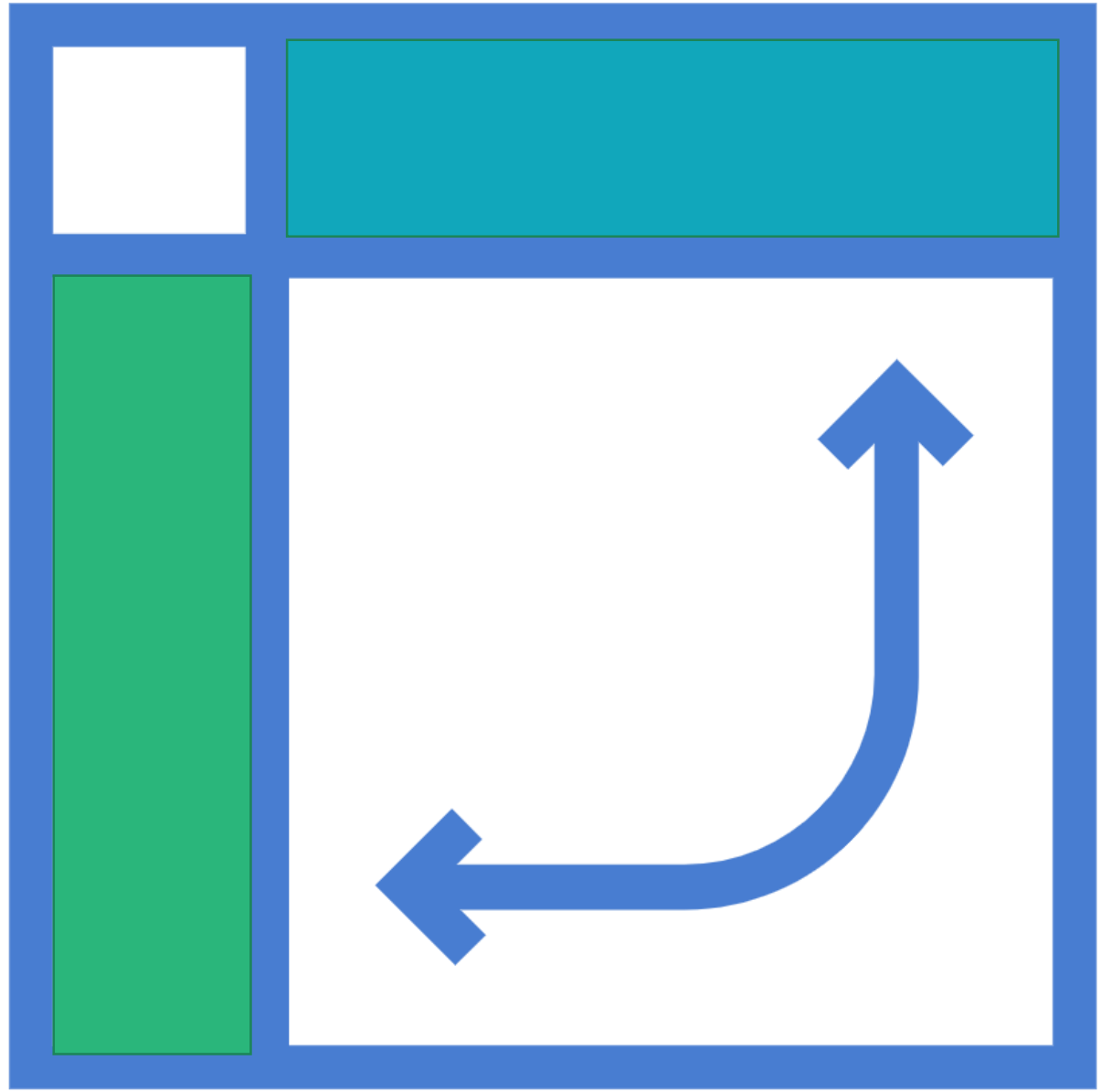


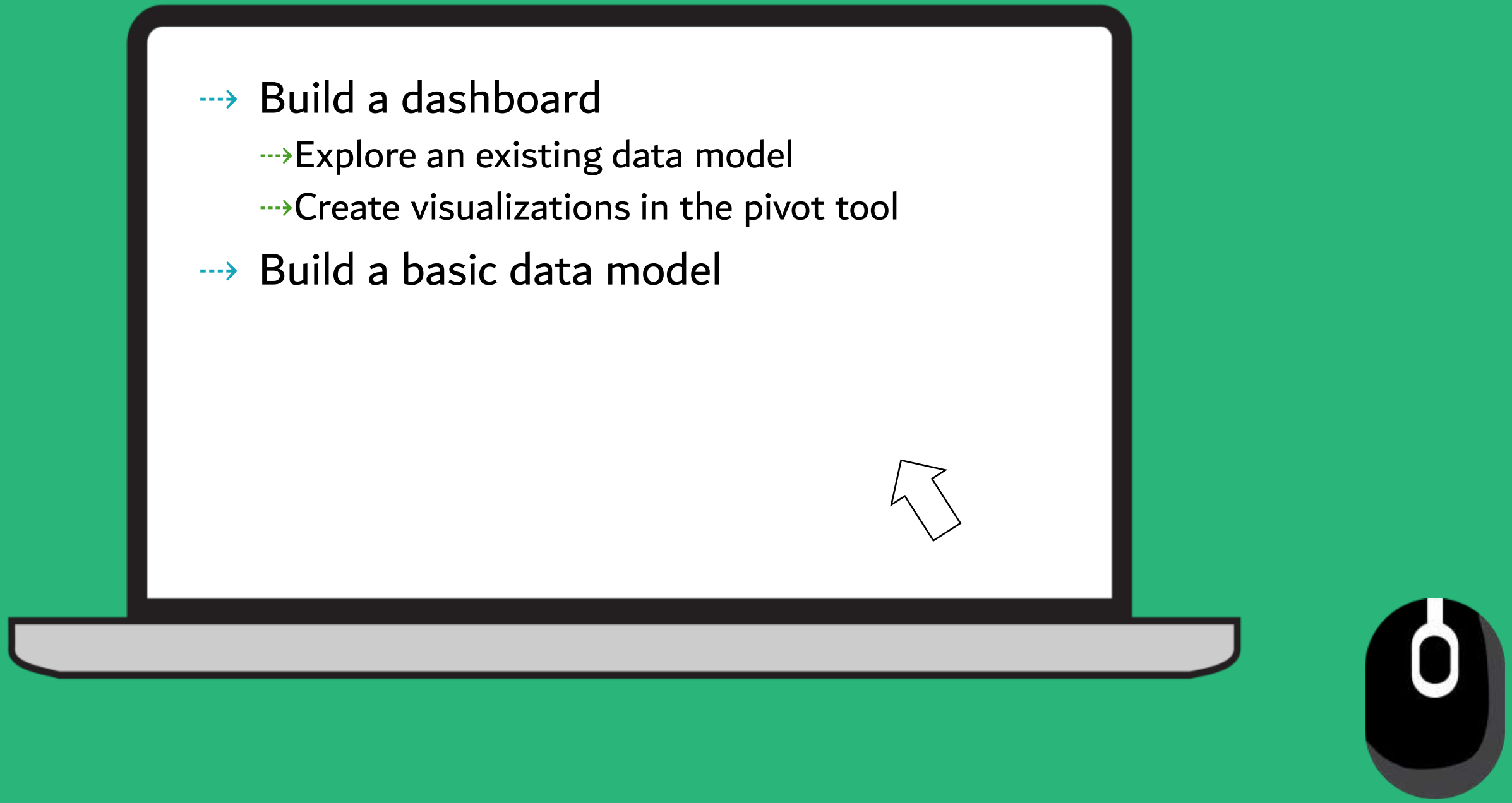
- Create dashboards, reports, and alerts without using SPL
- Provides a drag-and-drop interface to Splunk users
- Pivot functionality is built on data models

Basic Pivot Functions








Build a dashboard
Explore an existing data model
Create visualizations in the pivot tool
Build a basic data model

Summary

- Explored the pivot editor
- Built reports and alerts
- Built a data model