splunk>®

Enriching Data with Lookups

Document Usage Guidelines

- Should be used only for enrolled students
- Not meant to be a self-paced document, an instructor is needed
- Lab Exercise slides reference the hands-on lab exercise guide
- Do not distribute

Course Goals

- Define lookups
- Identify types of lookups
- Create a lookup
- Define a geospatial lookup
- Use an external lookup
- Define a KV Store lookup

Course Outline

- What is a Lookup?
- Create a Lookup
- Geospatial Lookups
- External Lookups
- KV Store Lookups
- Best Practices for Lookups

What is a Lookup?

Topic Objectives

- Define a lookup and the default lookup types
- Where lookups fall in the search-time operation sequence

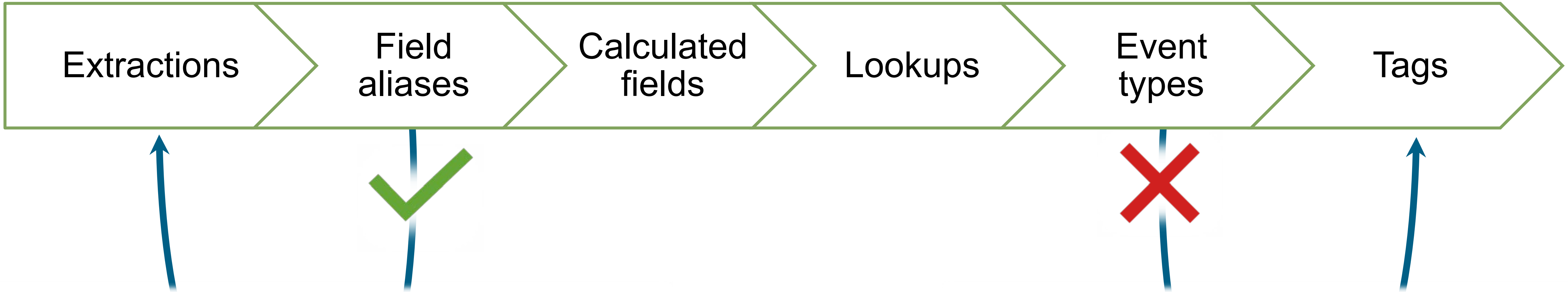
What is a Lookup?

- **Lookups** provide enrichment to your event data by appending fields from another data source (i.e. lookup output fields)
- Splunk provides four types of lookups by default

Lookup Type	Description
File-based	Populates your events with fields pulled from CSV files
External	Uses Python scripts or binary executables to append data
KV Store	Accesses key value pairs from a KV Store collection
Geospatial	References a KMZ or KML file

Search-time Operation Sequence

Search-time operations are always applied in the same order when generating the knowledge objects



Extractions

Field aliases

Calculated fields

Lookups

Event types

Tags

Each operation can reference knowledge objects derived from operations that **precede** them

No operation can reference knowledge objects derived by operations that **follow** them

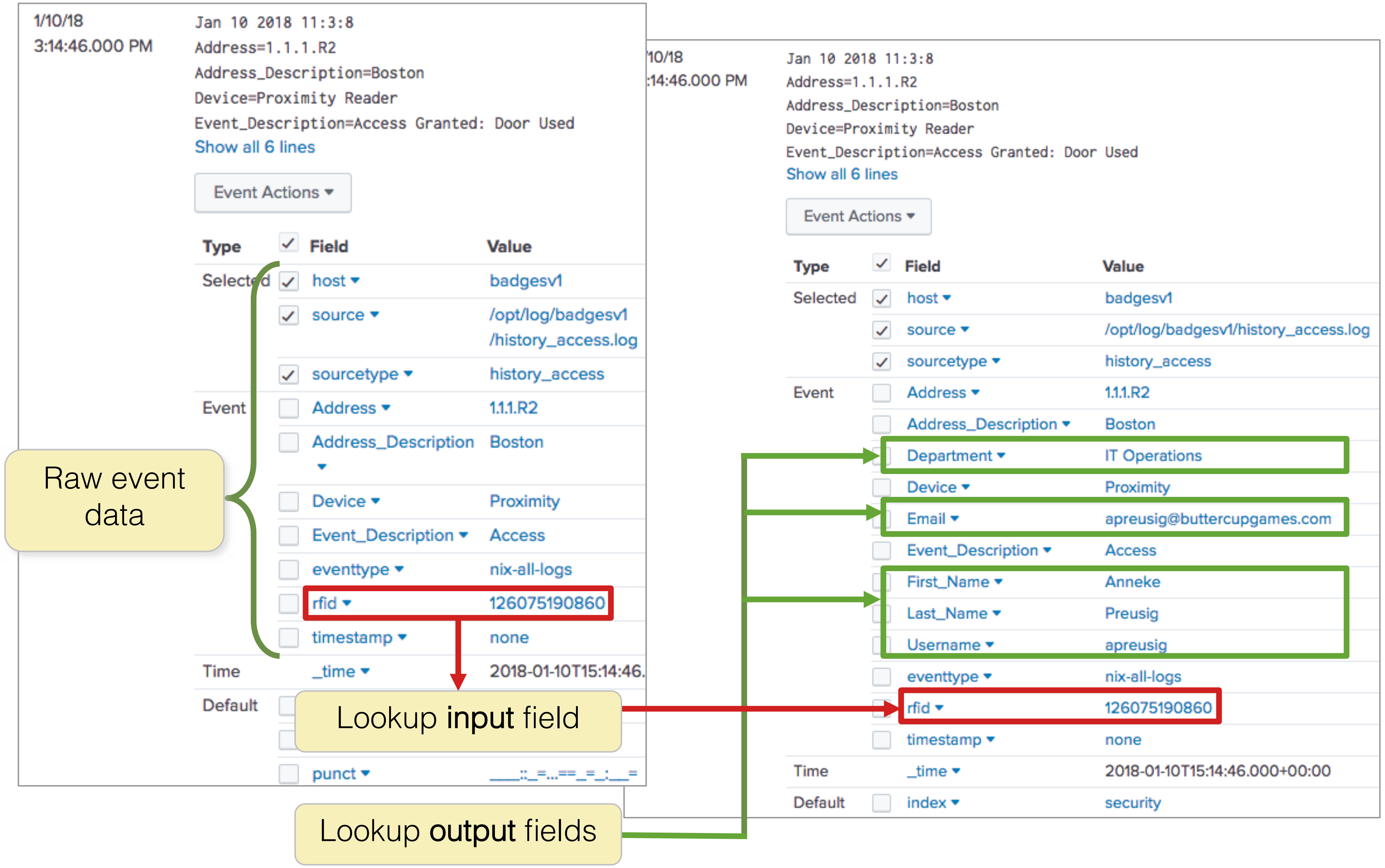
Create a Lookup

Topic Objectives

- Describe lookups at search time
- Use file-based lookups
- Examine a CSV lookup file
- Create a lookup
  - Upload a lookup table file
  - Define a lookup
  - Configure time-based lookup
  - Apply advanced lookup options
- Create and use an automatic lookup at search

Lookups at Search Time

- Sometimes static (or relatively unchanging) data is required for searches, but isn't available in the raw event data
- Lookups pull such data from standalone files at search time and add it to search results as field values



```
1/10/18             Jan 10 2018 11:3:8
3:14:46.000 PM      Address=1.1.1.R2
                    Address_Description=Boston
                    Device=Proximity Reader
                    Event_Description=Access Granted: Door Used
                    Show all 6 lines
```

Type	☑	Field	Value
Selected	☑	host ▾	badgesv1
	☑	source ▾	/opt/log/badgesv1 /history_access.log
	☑	sourcetype ▾	history_access
Event	□	Address ▾	1.1.1.R2
	□	Address_Description ▾	Boston
	□	Device ▾	Proximity
	□	Event_Description ▾	Access
	□	eventtype ▾	nix-all-logs
	□	rfid ▾	126075190860
	□	timestamp ▾	none
Time		_time ▾	2018-01-10T15:14:46.
Default	□		
	□		
	□	punct ▾	____::_=...==_=_:___=

Raw event data

Lookup **input** field

Lookup **output** fields

```
10/18               Jan 10 2018 11:3:8
:14:46.000 PM       Address=1.1.1.R2
                    Address_Description=Boston
                    Device=Proximity Reader
                    Event_Description=Access Granted: Door Used
                    Show all 6 lines
```

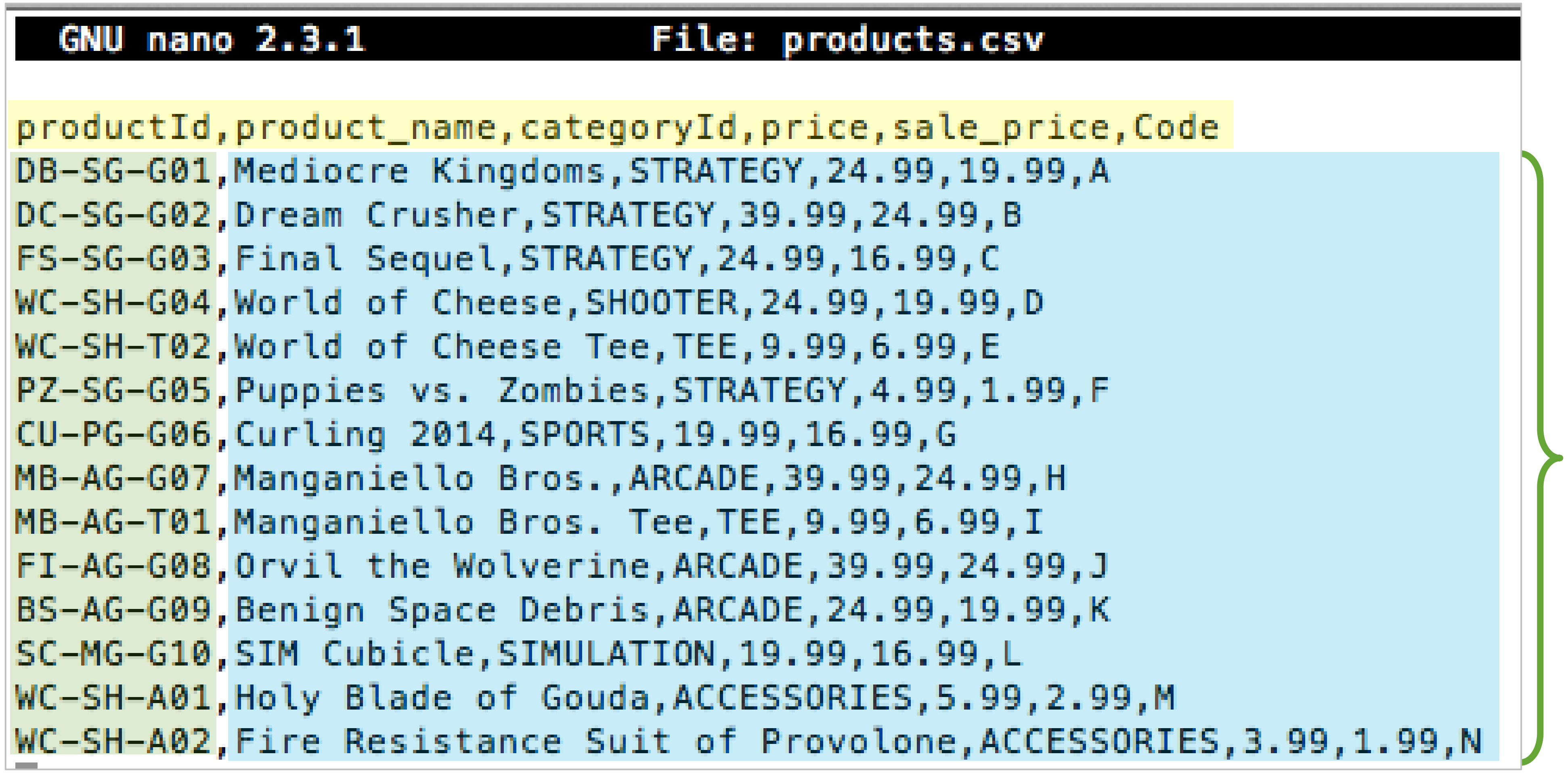
Type	☑	Field	Value
Selected	☑	host ▾	badgesv1
	☑	source ▾	/opt/log/badgesv1/history_access.log
	☑	sourcetype ▾	history_access
Event	□	Address ▾	1.1.1.R2
	□	Address_Description ▾	Boston
	□	Department ▾	IT Operations
	□	Device ▾	Proximity
	□	Email ▾	apreusig@buttercupgames.com
	□	Event_Description ▾	Access
	□	First_Name ▾	Anneke
	□	Last_Name ▾	Preusig
	□	Username ▾	apreusig
	□	eventtype ▾	nix-all-logs
	□	rfid ▾	126075190860
	□	timestamp ▾	none
Time		_time ▾	2018-01-10T15:14:46.000+00:00
Default	□	index ▾	security

Use a File-Based Lookup

- Lookups allow you to add more fields to your events such as:
  - Descriptions for HTTP status codes (“Not Found”, “Service Unavailable”)
  - Sale prices for products
  - Usernames, IP addresses, and workstation IDs associated with RFIDs
- After a lookup is invoked, lookup fields appear in the Fields sidebar and can be used in searches and reports
- Lookups can be invoked by the `lookup` command or configured to run automatically
- Lookup field values are case sensitive by default

A Sample CSV Lookup File

First row represents field names (header)



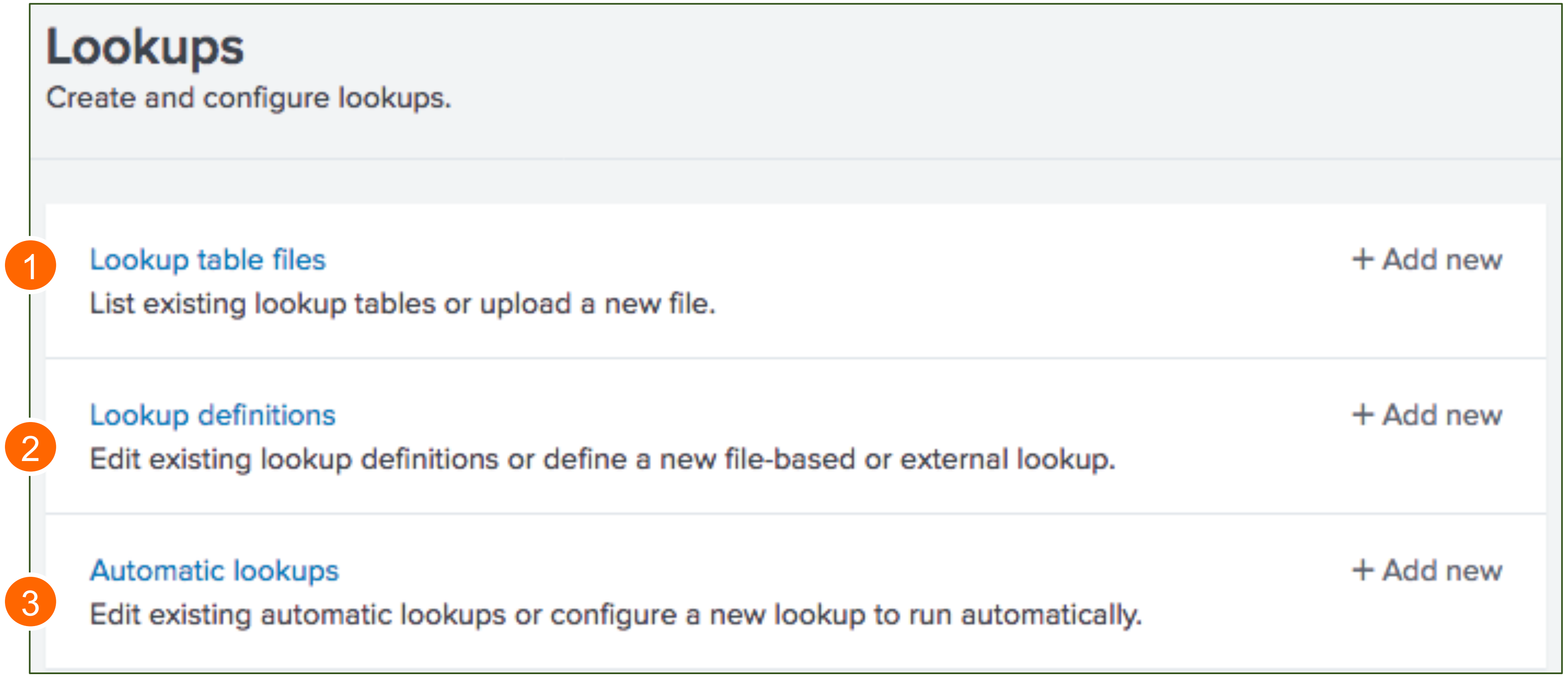
```
  GNU nano 2.3.1                File: products.csv

productId,product_name,categoryId,price,sale_price,Code
DB-SG-G01,Mediocre Kingdoms,STRATEGY,24.99,19.99,A
DC-SG-G02,Dream Crusher,STRATEGY,39.99,24.99,B
FS-SG-G03,Final Sequel,STRATEGY,24.99,16.99,C
WC-SH-G04,World of Cheese,SHOOTER,24.99,19.99,D
WC-SH-T02,World of Cheese Tee,TEE,9.99,6.99,E
PZ-SG-G05,Puppies vs. Zombies,STRATEGY,4.99,1.99,F
CU-PG-G06,Curling 2014,SPORTS,19.99,16.99,G
MB-AG-G07,Manganiello Bros.,ARCADE,39.99,24.99,H
MB-AG-T01,Manganiello Bros. Tee,TEE,9.99,6.99,I
FI-AG-G08,Orvil the Wolverine,ARCADE,39.99,24.99,J
BS-AG-G09,Benign Space Debris,ARCADE,24.99,19.99,K
SC-MG-G10,SIM Cubicle,SIMULATION,19.99,16.99,L
WC-SH-A01,Holy Blade of Gouda,ACCESSORIES,5.99,2.99,M
WC-SH-A02,Fire Resistance Suit of Provolone,ACCESSORIES,3.99,1.99,N
```

Input field: The `productId` field exists in both `access_combined` events and `.csv` file

Output fields: All other fields in the `.csv` file are searchable after the lookup is defined

Create a Lookup

1. Upload the file required for the lookup
2. Define the lookup type
3. Optionally, configure the lookup to run automatically



Lookups

Create and configure lookups.

1 Lookup table files — + Add new

List existing lookup tables or upload a new file.

2 Lookup definitions — + Add new

Edit existing lookup definitions or define a new file-based or external lookup.

3 Automatic lookups — + Add new

Edit existing automatic lookups or configure a new lookup to run automatically.

Add a New Lookup Table File

Settings > Lookups > Lookup table files > 1 New Lookup Table File

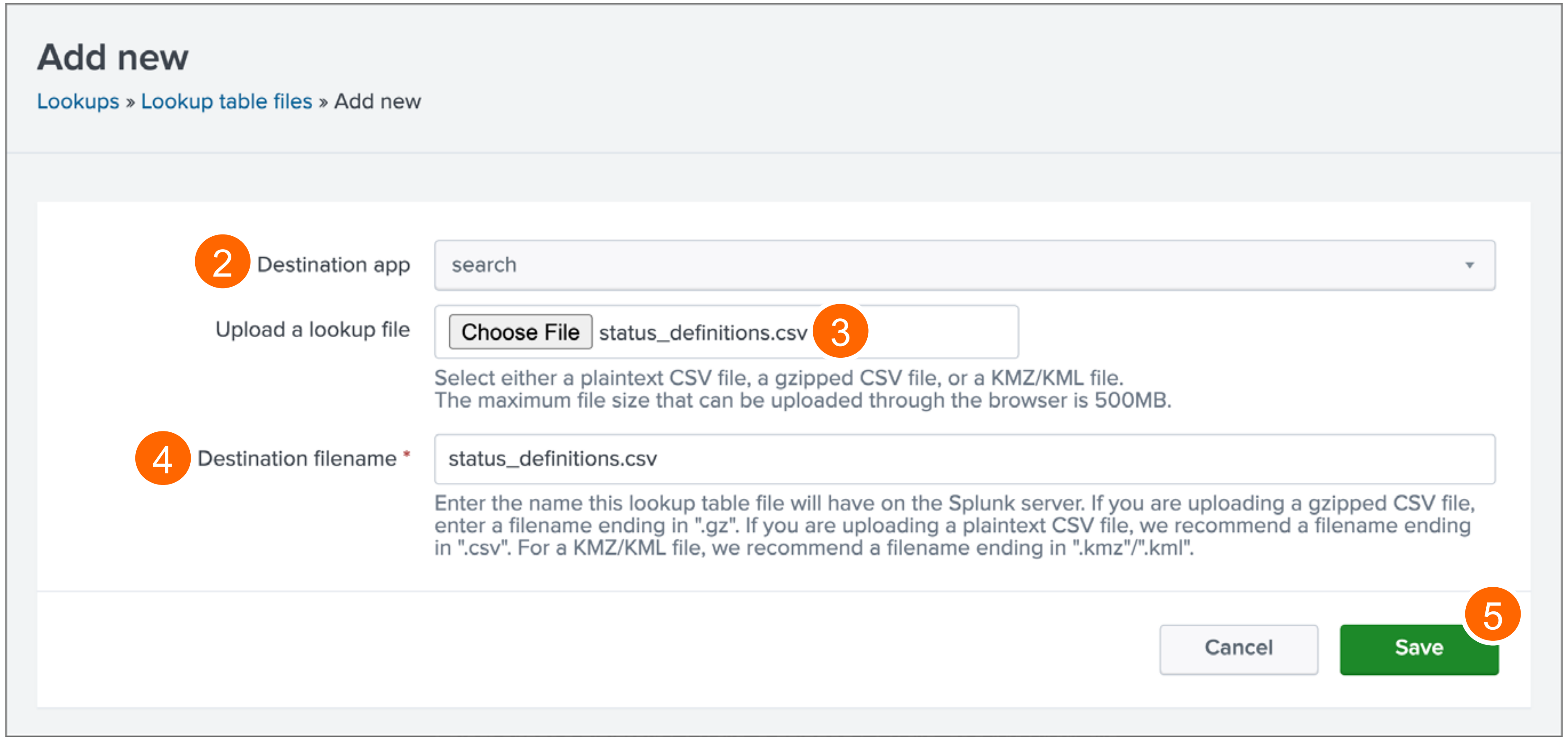
1 Click **New Lookup Table File**

2 Select a **Destination app**

3 Browse and select the file to use for the lookup table

4 Enter a name for the lookup file

5 Save



Add new

Lookups » Lookup table files » Add new

2 Destination app: search

Upload a lookup file: Choose File status_definitions.csv 3

Select either a plaintext CSV file, a gzipped CSV file, or a KMZ/KML file.
The maximum file size that can be uploaded through the browser is 500MB.

4 Destination filename *: status_definitions.csv

Enter the name this lookup table file will have on the Splunk server. If you are uploading a gzipped CSV file, enter a filename ending in ".gz". If you are uploading a plaintext CSV file, we recommend a filename ending in ".csv". For a KMZ/KML file, we recommend a filename ending in ".kmz"/".kml".

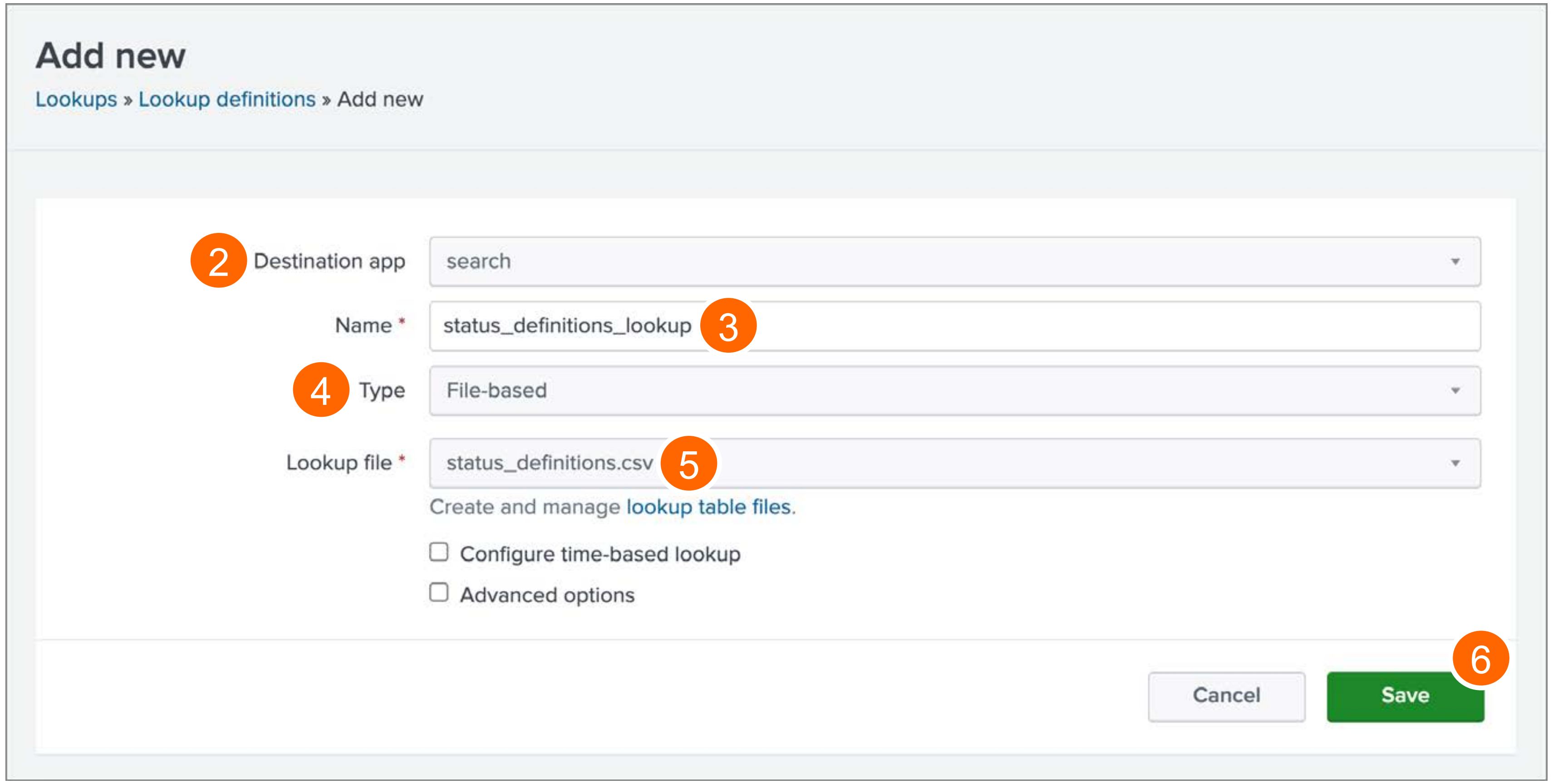
Note

For file-based lookups, the lookup should be a CSV file or a gzipped CSV file.

Create a Lookup Definition

Settings > Lookups > Lookup definitions> 1 **New Lookup Definition**

1. Click **New Lookup Definition**
2. Select a **Destination app**
3. **Name** the Lookup definition
4. For **Type**, select **File-based**
5. Browse and select the file to use for the lookup table
6. Save



Add new

Lookups » Lookup definitions » Add new

2 Destination app: search

Name *: status_definitions_lookup 3

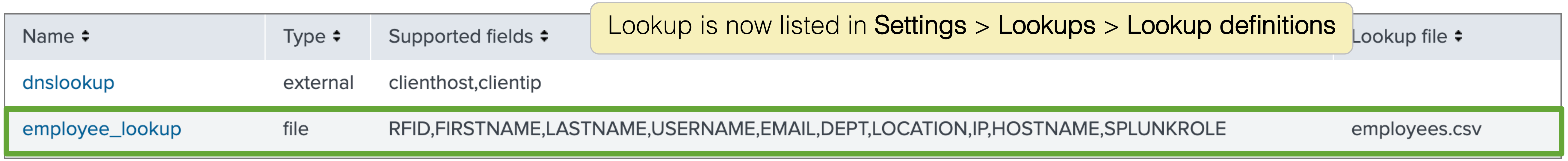
4 Type: File-based

Lookup file *: status_definitions.csv 5

Create and manage lookup table files.

□ Configure time-based lookup

□ Advanced options

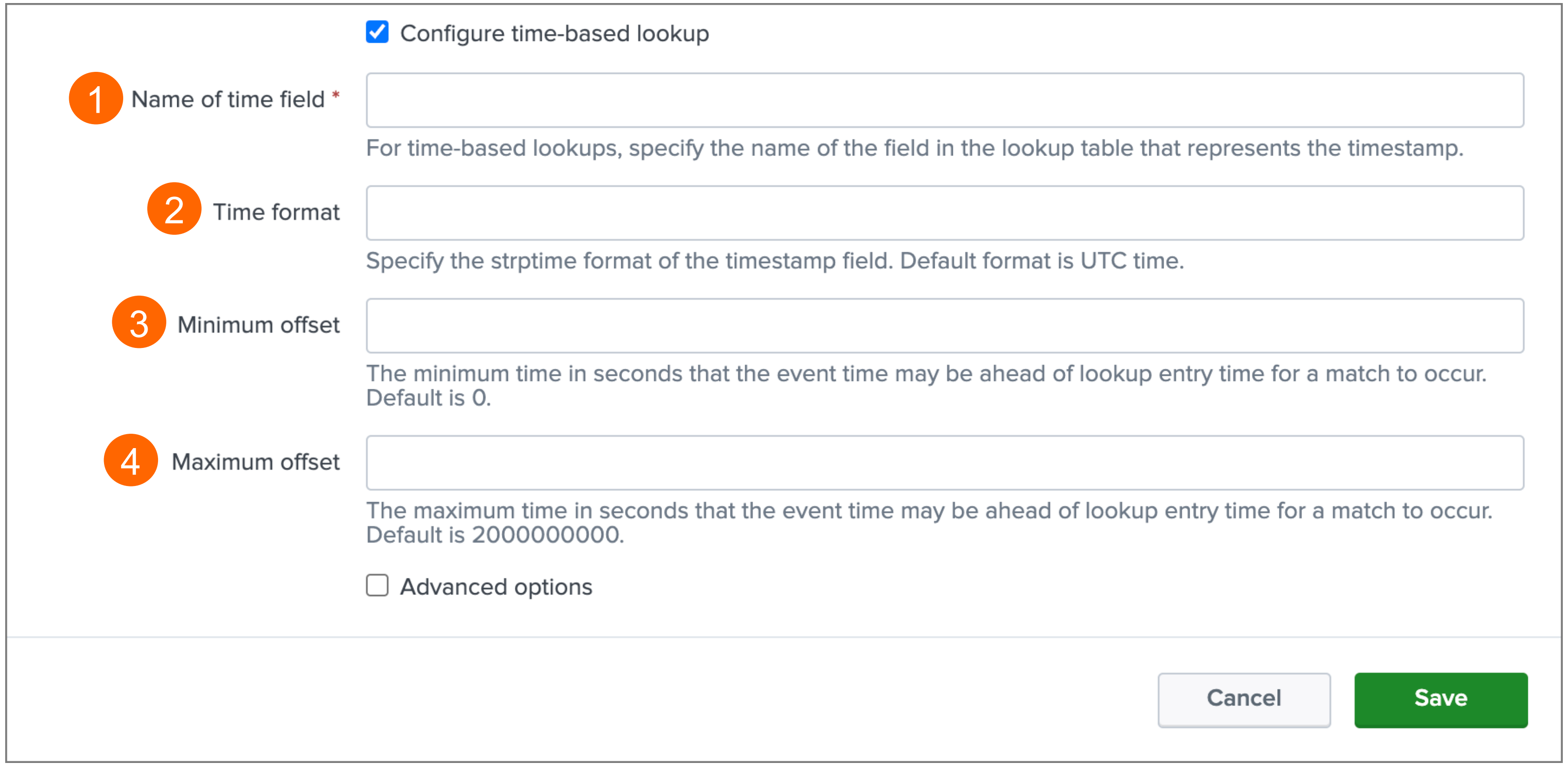


Lookup is now listed in **Settings** > **Lookups** > **Lookup definitions**

Name ⬍	Type ⬍	Supported fields ⬍	Lookup file ⬍
dnslookup	external	clienthost,clientip	
employee_lookup	file	RFID,FIRSTNAME,LASTNAME,USERNAME,EMAIL,DEPT,LOCATION,IP,HOSTNAME,SPLUNKROLE	employees.csv

Configure Time-based Lookup

1. Field in the record that represents the timestamp
2. Specifies the strptime() format of the time_field attribute (%s.%Q represents Unix epoch time in seconds and milliseconds)
3. Minimum amount of time in seconds that an event timestamp can be later than the record timestamp
4. Maximum amount of time in seconds that an event timestamp can be later than the record timestamp



☑ Configure time-based lookup

1 Name of time field *

For time-based lookups, specify the name of the field in the lookup table that represents the timestamp.

2 Time format

Specify the strptime format of the timestamp field. Default format is UTC time.

3 Minimum offset

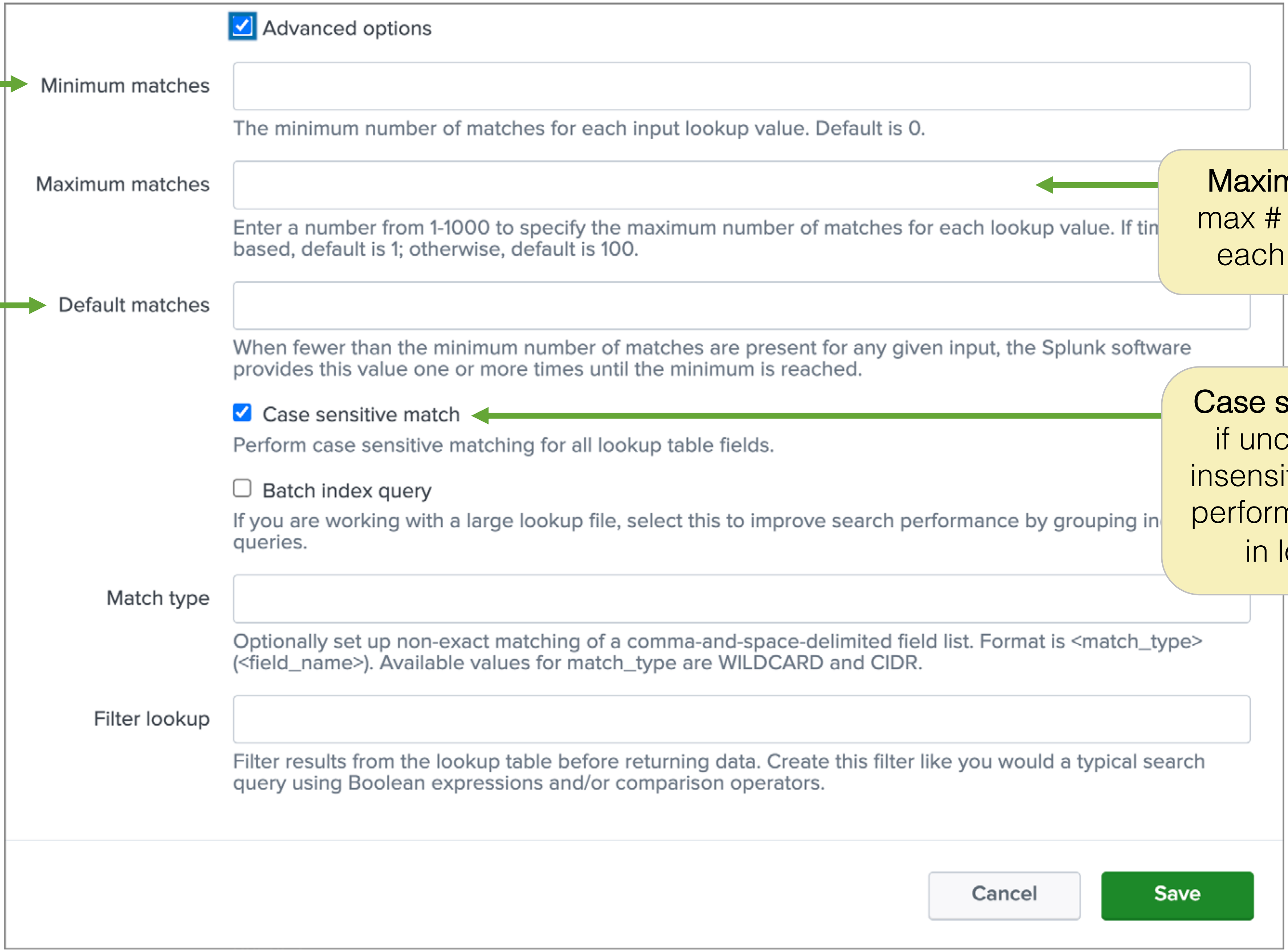
The minimum time in seconds that the event time may be ahead of lookup entry time for a match to occur. Default is 0.

4 Maximum offset

The maximum time in seconds that the event time may be ahead of lookup entry time for a match to occur. Default is 2000000000.

☐ Advanced options

Apply Advanced Lookup Options



☑ Advanced options

Minimum matches

The minimum number of matches for each input lookup value. Default is 0.

Maximum matches

Enter a number from 1-1000 to specify the maximum number of matches for each lookup value. If tin based, default is 1; otherwise, default is 100.

Default matches

When fewer than the minimum number of matches are present for any given input, the Splunk software provides this value one or more times until the minimum is reached.

☑ Case sensitive match

Perform case sensitive matching for all lookup table fields.

□ Batch index query

If you are working with a large lookup file, select this to improve search performance by grouping in queries.

Match type

Optionally set up non-exact matching of a comma-and-space-delimited field list. Format is <match_type> (<field_name>). Available values for match_type are WILDCARD and CIDR.

Filter lookup

Filter results from the lookup table before returning data. Create this filter like you would a typical search query using Boolean expressions and/or comparison operators.

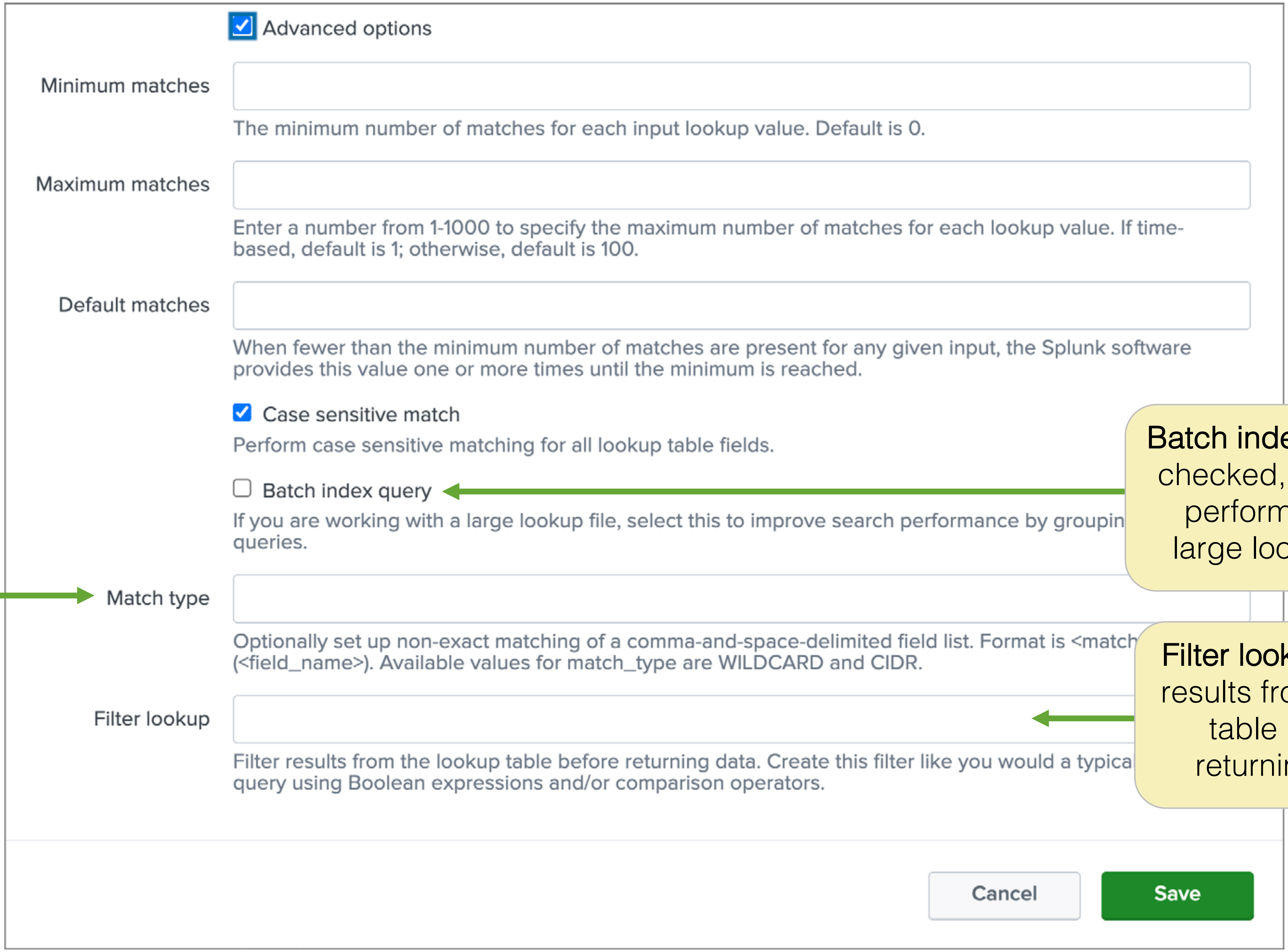
Minimum matches: min # of matches for each input lookup value

Maximum matches: max # of matches for each lookup value

Default matches: value to output when fewer than the minimum matches are returned for a given input; defaults to an empty string

Case sensitive match: if unchecked, case insensitive matching is performed for all fields in lookup table

Apply Advanced Lookup Options (cont.)



☑ Advanced options

Minimum matches

The minimum number of matches for each input lookup value. Default is 0.

Maximum matches

Enter a number from 1-1000 to specify the maximum number of matches for each lookup value. If time-based, default is 1; otherwise, default is 100.

Default matches

When fewer than the minimum number of matches are present for any given input, the Splunk software provides this value one or more times until the minimum is reached.

☑ Case sensitive match

Perform case sensitive matching for all lookup table fields.

☐ Batch index query

If you are working with a large lookup file, select this to improve search performance by groupin queries.

Match type

Optionally set up non-exact matching of a comma-and-space-delimited field list. Format is <match (<field_name>). Available values for match_type are WILDCARD and CIDR.

Filter lookup

Filter results from the lookup table before returning data. Create this filter like you would a typica query using Boolean expressions and/or comparison operators.

Match type: allows for non-exact matching using the match types WILDCARD and CIDR

Batch index query: if checked, improves performance of large lookup files

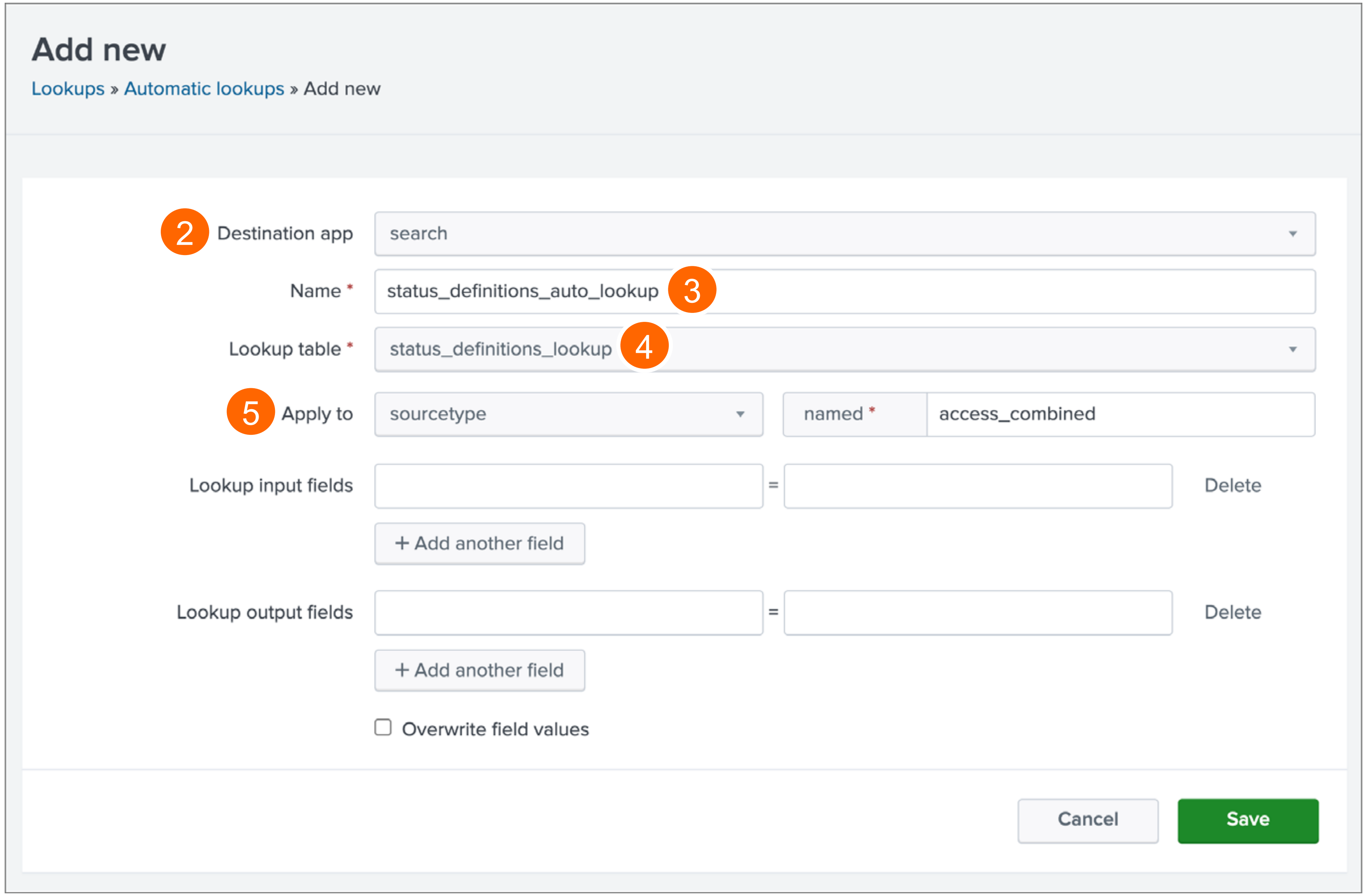
Filter lookup: filters results from lookup table before returning data

| 12 August 2022

Create an Automatic Lookup

Settings > Lookups > Automatic lookups > 1 New Automatic Lookup

1. Click **New Automatic Lookup**
2. Select a **Destination app**
3. **Name** the automatic lookup
4. Select the **Lookup table** definition
5. Select host, source, or sourcetype to apply to the lookup and specify the name



Add new

Lookups » Automatic lookups » Add new

2 Destination app: search

Name *: status_definitions_auto_lookup 3

Lookup table *: status_definitions_lookup 4

5 Apply to: sourcetype named *: access_combined

Lookup input fields: = Delete

+ Add another field

Lookup output fields: = Delete

+ Add another field

□ Overwrite field values

Cancel Save

Create an Automatic Lookup (cont.)

Settings > Lookups > Automatic lookups > **New Automatic Lookup**

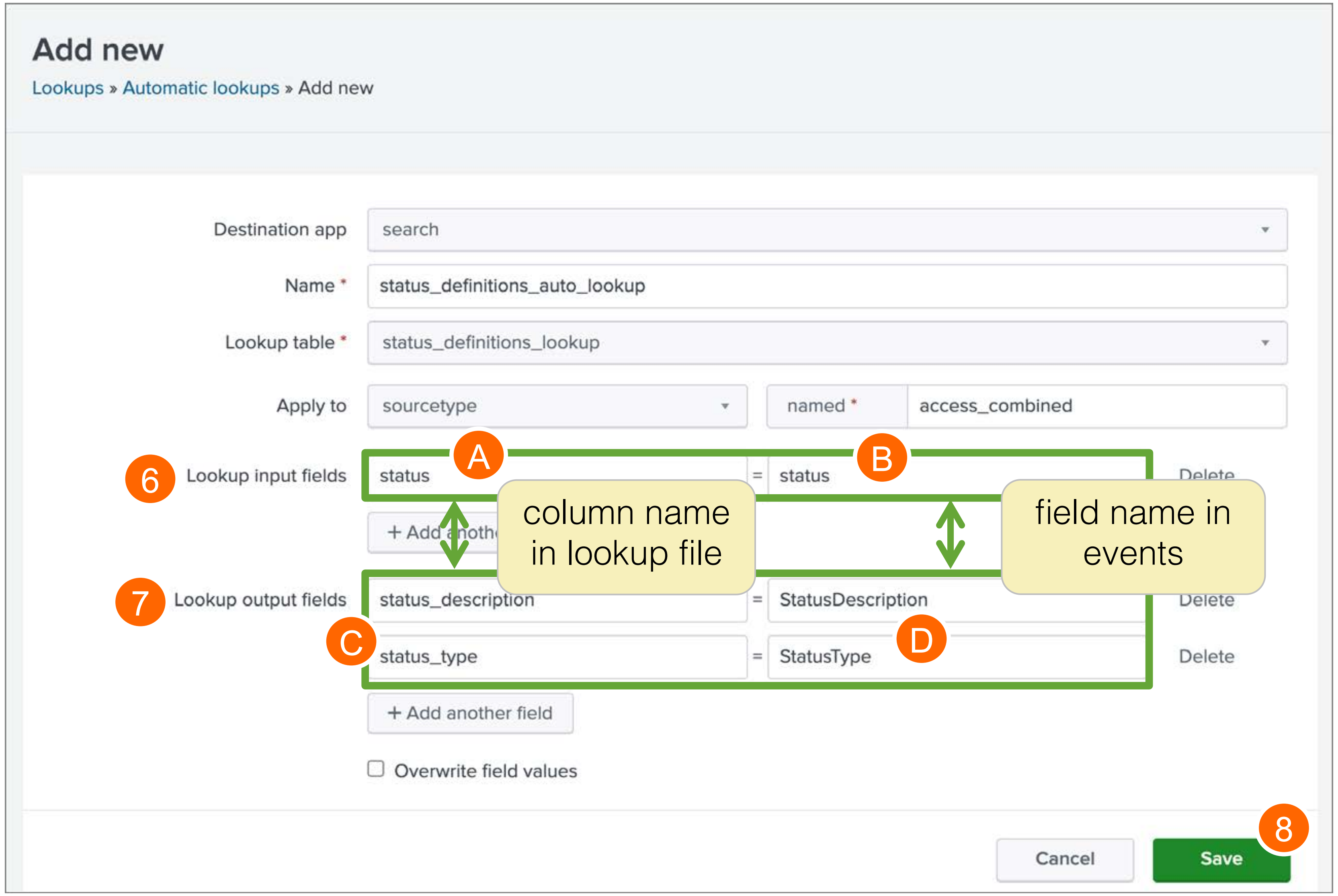
6 Define the **Lookup input fields**: the field(s) that exist in your events that you are relating to the lookup table

- A Column name in CSV
- B Field name in events

7 Define **Lookup output fields**

- C Field name in lookup table
- D Name you want displayed in results, otherwise column name from CSV is inherited

8 Save



Add new

Lookups » Automatic lookups » Add new

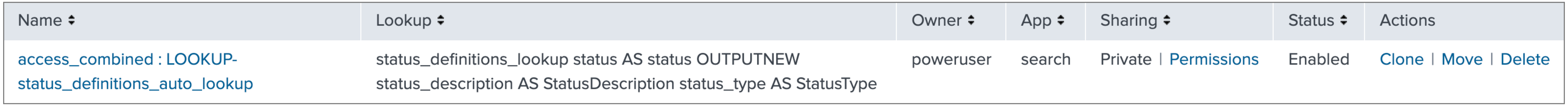
Destination app	search
Name *	status_definitions_auto_lookup
Lookup table *	status_definitions_lookup
Apply to	sourcetype — named * access_combined
6 Lookup input fields	status = status (Delete)
	+ Add another field
7 Lookup output fields	status_description = StatusDescription (Delete)
	status_type = StatusType (Delete)
	+ Add another field

☐ Overwrite field values

Cancel | Save

Create an Automatic Lookup (cont.)

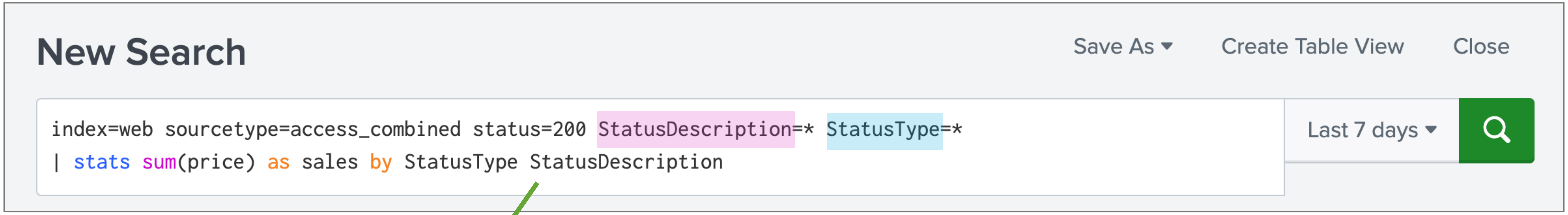
- Lookup is now listed in **Lookups > Automatic lookups**
- Automatic lookups are applied to all searches at search time



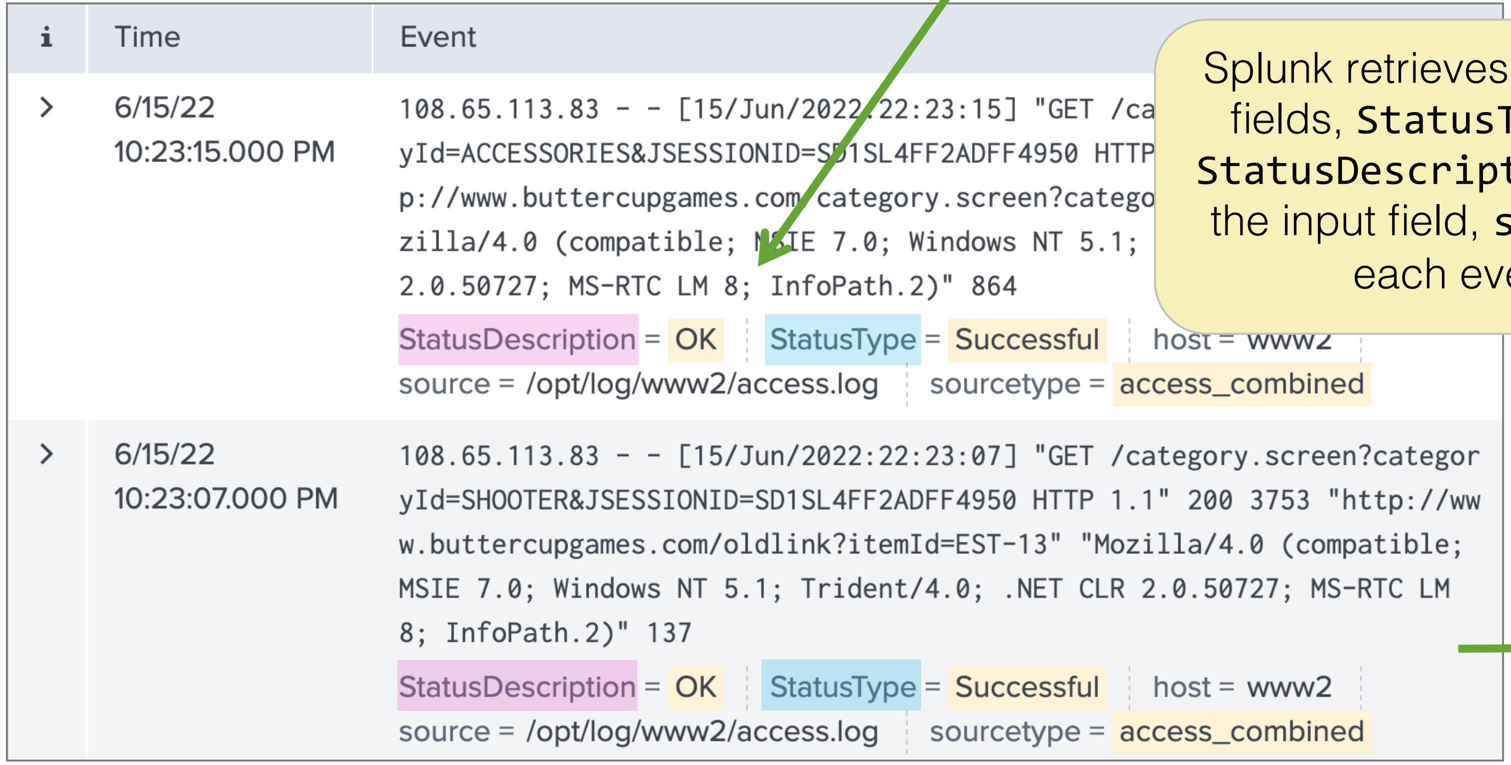
Name	Lookup	Owner	App	Sharing	Status	Actions
access_combined : LOOKUP-status_definitions_auto_lookup	status_definitions_lookup status AS status OUTPUTNEW status_description AS StatusDescription status_type AS StatusType	poweruser	search	Private | Permissions	Enabled	Clone | Move | Delete

Use an Automatic Lookup in Search

To use an automatic lookup, specify the output fields in your search



```
index=web sourcetype=access_combined status=200 StatusDescription=* StatusType=*
| stats sum(price) as sales by StatusType StatusDescription
```

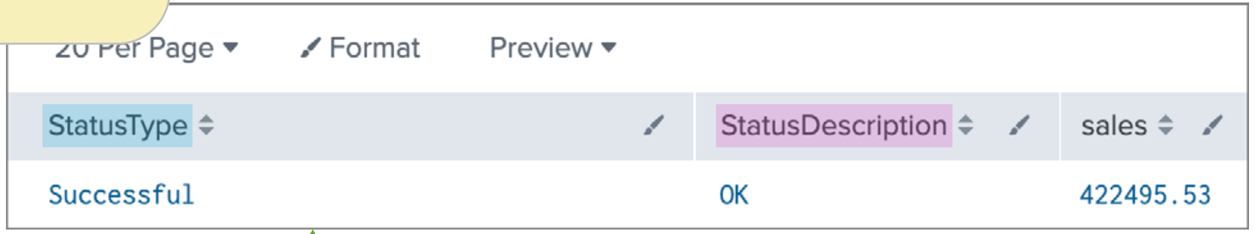


i	Time	Event

>	6/15/22
	10:23:15.000 PM
		108.65.113.83 - - [15/Jun/2022:22:23:15] "GET /ca
		yId=ACCESSORIES&JSESSIONID=SD1SL4FF2ADFF4950 HTTP
		p://www.buttercupgames.com/category.screen?catego
		zilla/4.0 (compatible; MSIE 7.0; Windows NT 5.1;
		2.0.50727; MS-RTC LM 8; InfoPath.2)" 864
		StatusDescription = OK | StatusType = Successful | host = www2
		source = /opt/log/www2/access.log | sourcetype = access_combined

>	6/15/22
	10:23:07.000 PM
		108.65.113.83 - - [15/Jun/2022:22:23:07] "GET /category.screen?categoryId=SHOOTER&JSESSIONID=SD1SL4FF2ADFF4950 HTTP 1.1" 200 3753 "http://www.buttercupgames.com/oldlink?itemId=EST-13" "Mozilla/4.0 (compatible; MSIE 7.0; Windows NT 5.1; Trident/4.0; .NET CLR 2.0.50727; MS-RTC LM 8; InfoPath.2)" 137
		StatusDescription = OK | StatusType = Successful | host = www2
		source = /opt/log/www2/access.log | sourcetype = access_combined

Splunk retrieves the output fields, **`StatusType`** and **`StatusDescription`**, using the input field, **`status`**, for each event



StatusType	StatusDescription	sales
Successful	OK	422495.53

Results are returned!

Create a Lookup Lab Exercise

Time: 20-25 minutes

Tasks:

- Add lookup table files to your search environment
- Create a lookup definitions
- Create an automatic lookup
- Verify your automatic lookup is working in search

Geospatial Lookups

Topic Objectives

- Describe the use of geospatial lookups
- Examine `KML/KMZ` geospatial lookup files
- Add a geospatial lookup file
- Define a geospatial lookup

Geospatial Lookups

- Matches region names in your events to region names in lookup and outputs fields with corresponding geographic feature info
- Location coordinate ranges are provided by geographic feature collections: `.KML` and `.KMZ` files
- Geospatial lookups can be invoked in searches to generate choropleth map visualizations
- Splunk ships with two geospatial lookup files:
  - `geo_us_states`
  - `geo_countries`

Denmark
Latvia
Moscow
Ireland
U.K.
Netherlands
Berlin
Amsterdam
London
Belarus
Poland
Germany
Kiev
Ukraine
Paris
Vienna
France
Switzerland
Austria
Hungary
Geneva
Romania
Italy
Georgia
Portugal
Madrid
Rome
Istanbul
Greece
Spain
Turkey
Athens
Tunisia

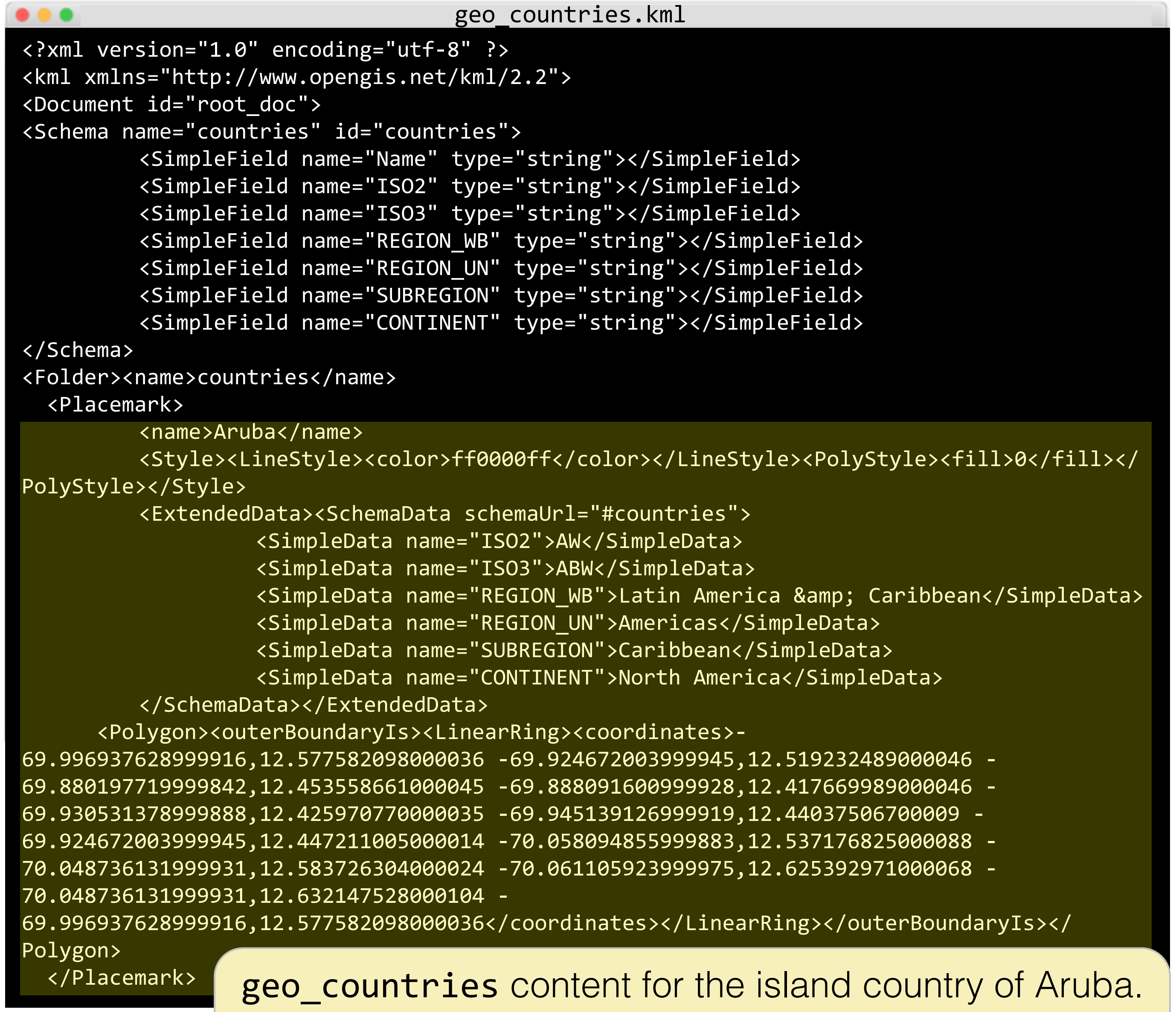
.KMZ

geo_countries

Choropleth map visualization

Geospatial Lookup Files

- Provides geographic feature information used to define a geospatial lookup
  - KML: a type of XML file
  - KMZ: a zipped KML file
- Rely on polygons which are closed shapes that start and end at the same coordinate
- Many are available online or can be created from scratch using software such as Google Earth



geo_countries.kml

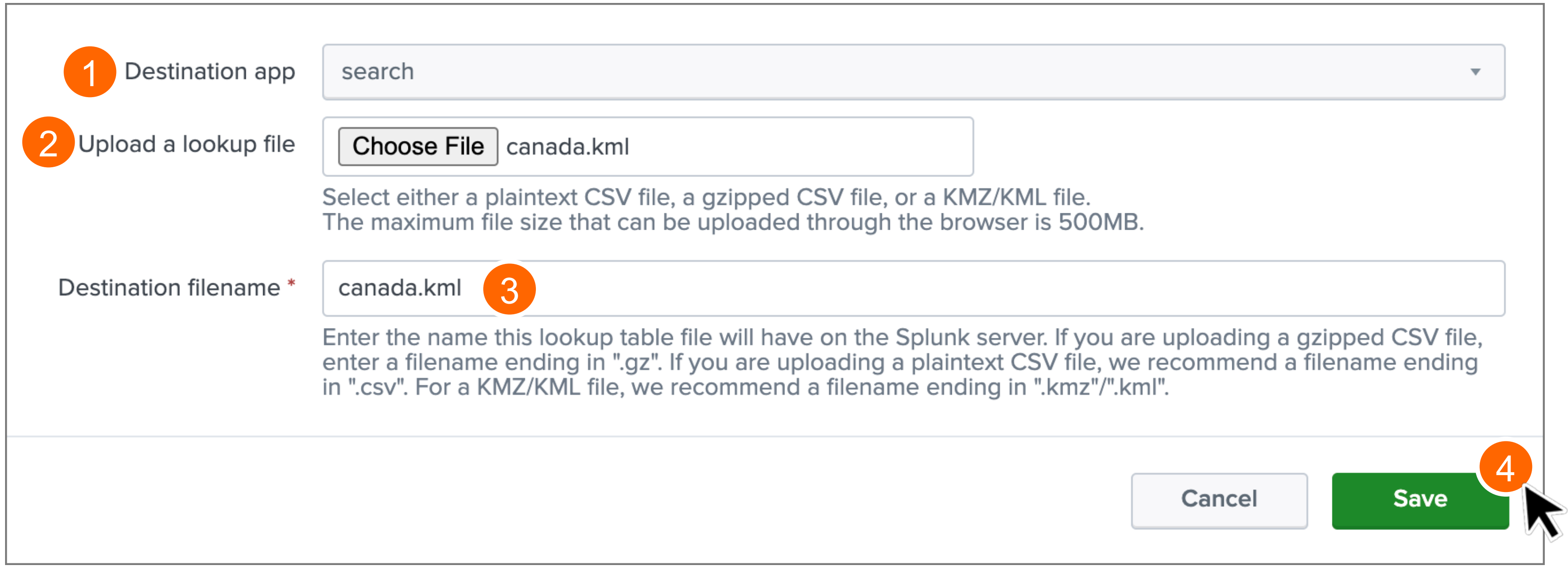
```
<?xml version="1.0" encoding="utf-8" ?>
<kml xmlns="http://www.opengis.net/kml/2.2">
<Document id="root_doc">
<Schema name="countries" id="countries">
        <SimpleField name="Name" type="string"></SimpleField>
        <SimpleField name="ISO2" type="string"></SimpleField>
        <SimpleField name="ISO3" type="string"></SimpleField>
        <SimpleField name="REGION_WB" type="string"></SimpleField>
        <SimpleField name="REGION_UN" type="string"></SimpleField>
        <SimpleField name="SUBREGION" type="string"></SimpleField>
        <SimpleField name="CONTINENT" type="string"></SimpleField>
</Schema>
<Folder><name>countries</name>
  <Placemark>
        <name>Aruba</name>
        <Style><LineStyle><color>ff0000ff</color></LineStyle><PolyStyle><fill>0</fill></
PolyStyle></Style>
        <ExtendedData><SchemaData schemaUrl="#countries">
                <SimpleData name="ISO2">AW</SimpleData>
                <SimpleData name="ISO3">ABW</SimpleData>
                <SimpleData name="REGION_WB">Latin America &amp; Caribbean</SimpleData>
                <SimpleData name="REGION_UN">Americas</SimpleData>
                <SimpleData name="SUBREGION">Caribbean</SimpleData>
                <SimpleData name="CONTINENT">North America</SimpleData>
        </SchemaData></ExtendedData>
     <Polygon><outerBoundaryIs><LinearRing><coordinates>-
69.996937628999916,12.577582098000036 -69.924672003999945,12.519232489000046 -
69.880197719999842,12.453558661000045 -69.888091600999928,12.417669989000046 -
69.930531378999888,12.425970770000035 -69.945139126999919,12.44037506700009 -
69.924672003999945,12.447211005000014 -70.058094855999883,12.537176825000088 -
70.048736131999931,12.583726304000024 -70.061105923999975,12.625392971000068 -
70.048736131999931,12.632147528000104 -
69.996937628999916,12.577582098000036</coordinates></LinearRing></outerBoundaryIs></
Polygon>
  </Placemark>
```

geo_countries content for the island country of Aruba. The **Polygon** tag (highlighted) contains the coordinates Splunk uses to define its choropleth map data.

Add a Geospatial Lookup Table File

Settings > Lookups > Lookup table files > **New Lookup Table File**

1. Select a **Destination app**
2. Browse and select the `.kmz` or `.kml` file to use for the lookup table
3. Enter a name for the lookup file
4. Save



1 Destination app: search

2 Upload a lookup file: Choose File canada.kml

Select either a plaintext CSV file, a gzipped CSV file, or a KMZ/KML file.
The maximum file size that can be uploaded through the browser is 500MB.

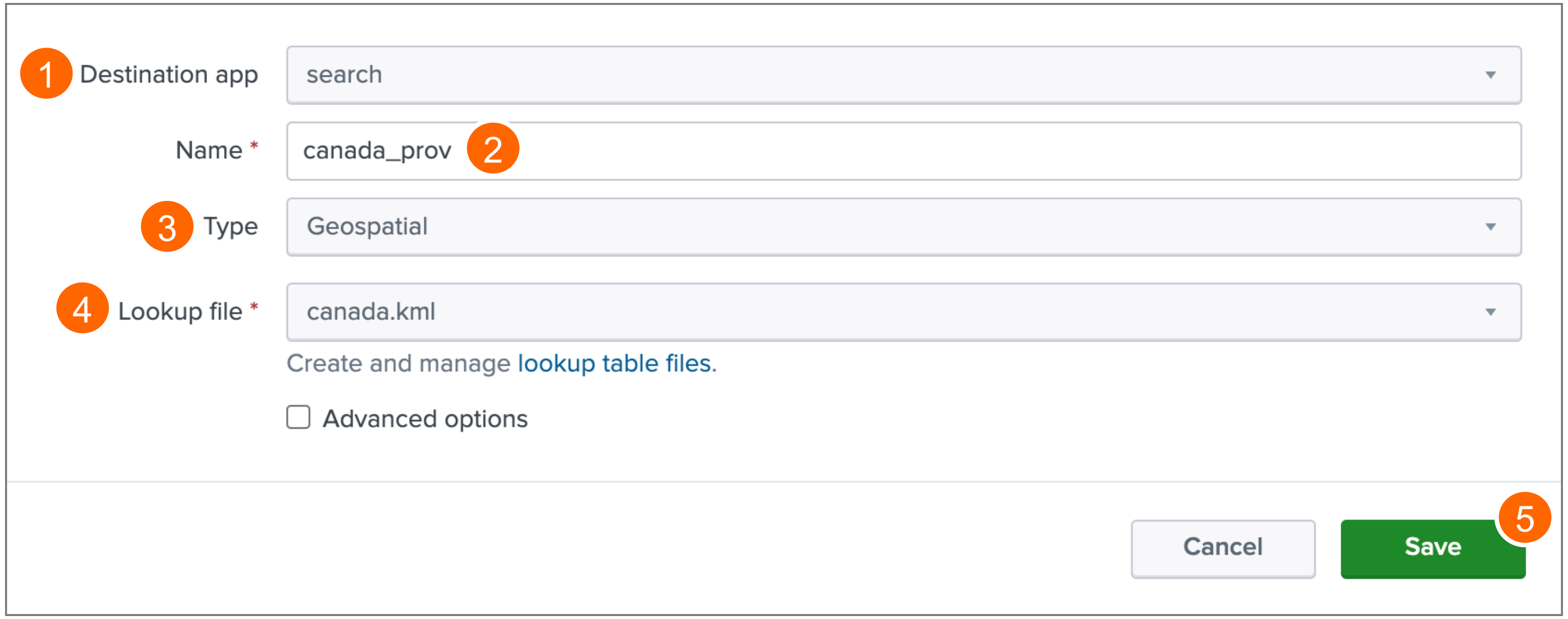
Destination filename *: canada.kml 3

Enter the name this lookup table file will have on the Splunk server. If you are uploading a gzipped CSV file, enter a filename ending in ".gz". If you are uploading a plaintext CSV file, we recommend a filename ending in ".csv". For a KMZ/KML file, we recommend a filename ending in ".kmz"/".kml".

Define a Geospatial Lookup

Settings > Lookups > Lookup definitions > **New Lookup Definition**

1. Select a **Destination app**
2. **Name** the lookup definition
3. Change **Type** to **Geospatial**
4. Select the **Lookup file** from the drop-down list
5. Save



1 Destination app: search

Name *: canada_prov 2

3 Type: Geospatial

4 Lookup file *: canada.kml

Create and manage lookup table files.

□ Advanced options

External Lookups

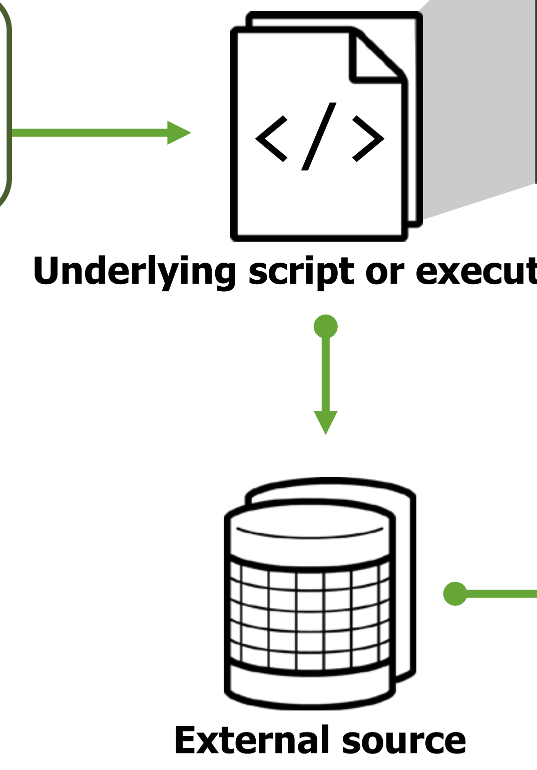
Topic Objectives

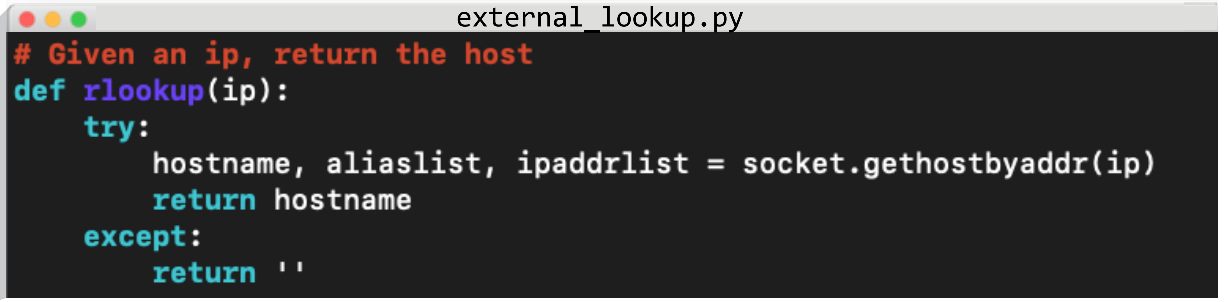
- Define the use of external lookups
- Examine an `external_lookup.py` lookup script
- Configure external lookups

What are External Lookups

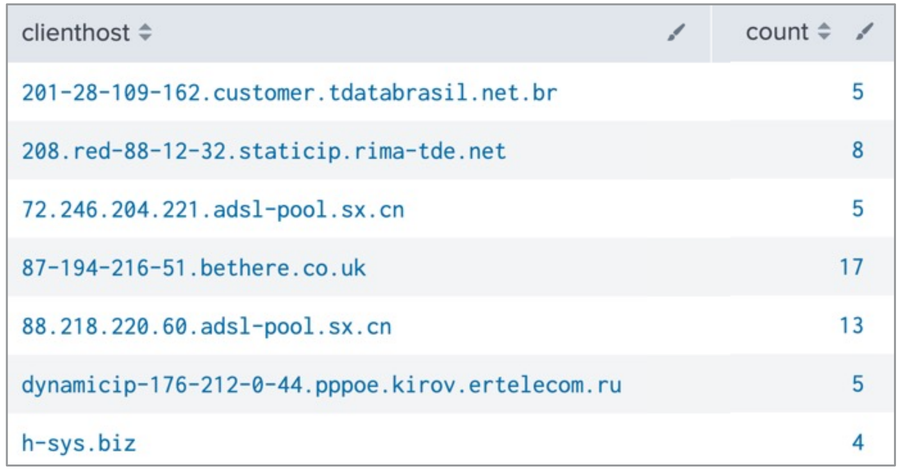
- External lookups use scripts or executables to populate events with field values from an external source
- Often referred to as **scripted lookups**

```
index=web sourcetype=access_combined
| lookup dnslookup clientip
| stats count by clienthost
```




external_lookup.py

```
# Given an ip, return the host
def rlookup(ip):
    try:
        hostname, aliaslist, ipaddrlist = socket.gethostbyaddr(ip)
        return hostname
    except:
        return ''
```



clienthost	count
201-28-109-162.customer.tdatabrasil.net.br	5
208.red-88-12-32.staticip.rima-tde.net	8
72.246.204.221.adsl-pool.sx.cn	5
87-194-216-51.bethere.co.uk	17
88.218.220.60.adsl-pool.sx.cn	13
dynamicip-176-212-0-44.pppoe.kirov.ertelecom.ru	5
h-sys.biz	4

Manage an External Lookup Script

- Must be a Python script or binary executable
- Must be added to your Splunk deployment in either:
  - `$SPLUNK_HOME/etc/searchscripts`
  - `$SPLUNK_HOME/etc/apps/<app_name>/bin`

`external_lookup.py`

- Splunk ships with a sample script **`external_lookup.py`** in **`$SPLUNK_HOME/etc/system/bin`**

To use the sample script:

1. Move **`external_lookup.py`** script to appropriate directory
2. Create **`dnslookup`** definition as shown in next slide
3. Invoke the lookup using either:

```
...| lookup dnslookup clienthost
```

```
...| lookup dnslookup clientip
```

- Splunk passes values for **`clienthost`** into script and script returns **`clientip`** (or vice versa)
- Returned values are used to populate **`clientip`** or **`clienthost`** in the results

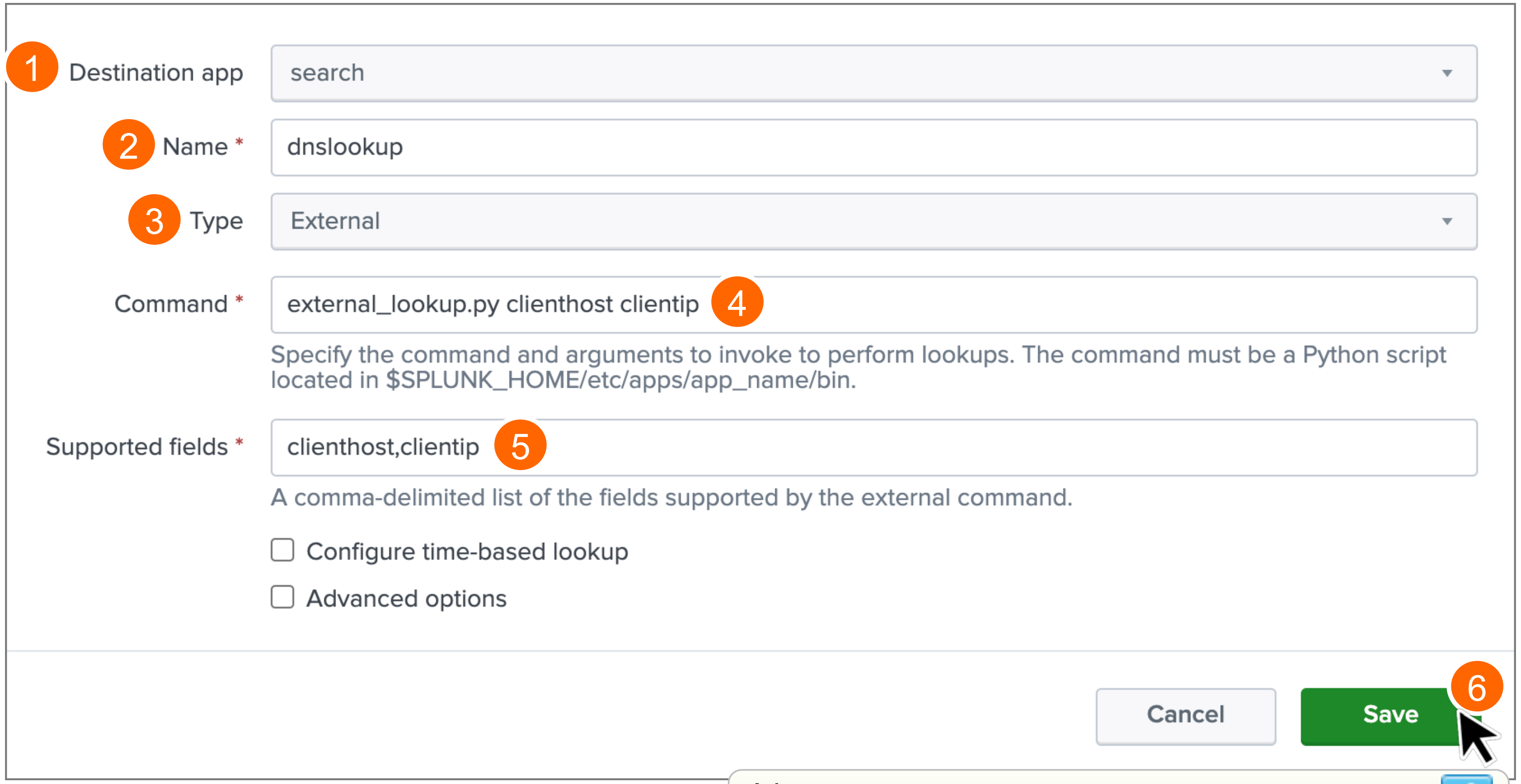
Note

The first step has already been completed in the lab environment.

Configure an External Lookup

Settings > Lookups > Lookup definitions > **New Lookup Definition**

1. Select **Destination app**
2. **Name** the lookup definition
3. Change **Type** to **External**
4. Enter script name and arguments passed to script
5. List all fields supported by the lookup
6. Save



1 Destination app: search

2 Name *: dnslookup

3 Type: External

Command *: external_lookup.py clienthost clientip 4

Specify the command and arguments to invoke to perform lookups. The command must be a Python script located in $SPLUNK_HOME/etc/apps/app_name/bin.

Supported fields *: clienthost,clientip 5

A comma-delimited list of the fields supported by the external command.

□ Configure time-based lookup

□ Advanced options

Note

The arguments passed to the script are the field headers from the input/output CSV files.

Geospatial & External Lookups Lab Exercise

Time: 10 minutes

Tasks:

- Upload and define a geospatial lookup and verify its contents in search
- Define an external lookup and use it in search

KV Store Lookups

Topic Objectives

- Define the use of KV Store lookups
- Identify the steps to set up a KV Store lookup
- Examine the KV Store lookups `collections.conf` file
- Create a KV Store lookup definition
- Identify the options for populating a KV store lookup
- Compare file-based CSV lookups to KV Store lookups

Use KV Store Lookups

- Instead of matching against values in a CSV file, you can also match against values in a **KV Store** (key value store)
- Use for large lookup tables or ones that are updated often
- KV Store saves and retrieves data in **collections** of key-value pairs
  - Similar to database tables in which each record has a unique key
  - Provides multiuser access locking so that multiple users can not edit the same record at the same time

Steps to Set up a KV Store Lookup

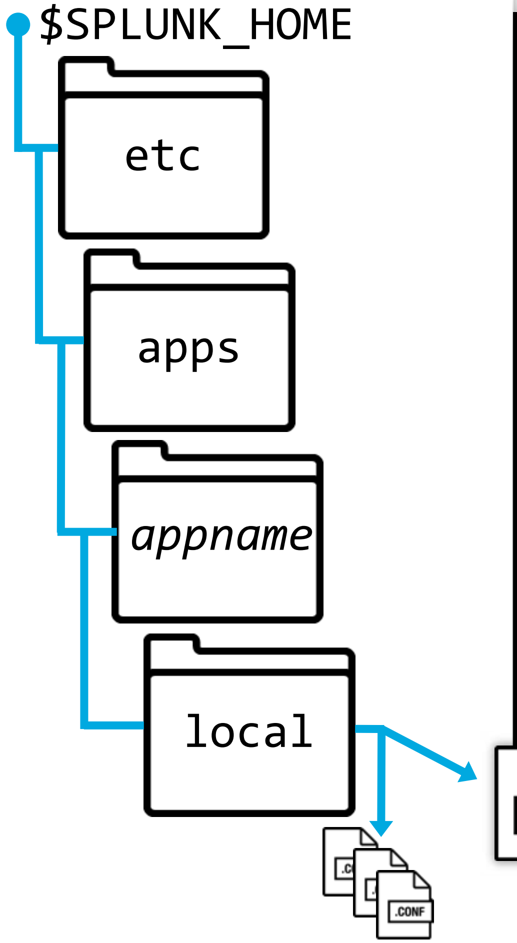
1. Add configuration stanzas to `collections.conf` (admin only)
2. Create KV Store definition
3. Populate the KV Store lookup with data using:
   - `outputlookup` command (admin and power user ability)
   - REST API (admin ability)
   - A front-end form (not discussed in this course)

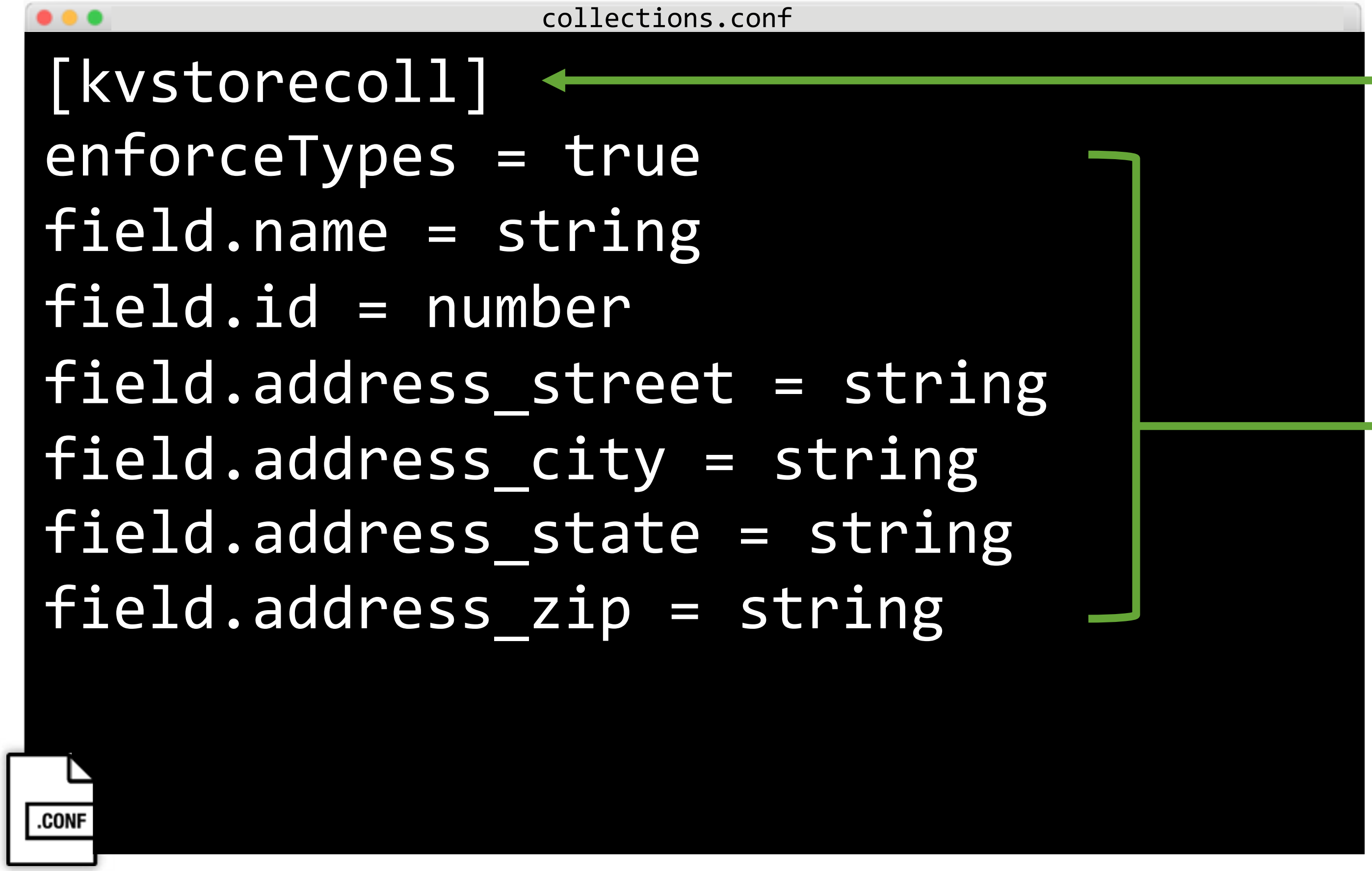
Note

Once defined, the admin can share the KV Store collection with other apps and users

Examine the KV Store: `collections.conf`

An admin must add a stanza for each KV Store in the `collections.conf` file before a definition can be created




collections.conf

```
[kvstorecoll]
enforceTypes = true
field.name = string
field.id = number
field.address_street = string
field.address_city = string
field.address_state = string
field.address_zip = string
```

A `[collection_name]` is required

Admins must define the data types for each field and have the option to enforce those data types

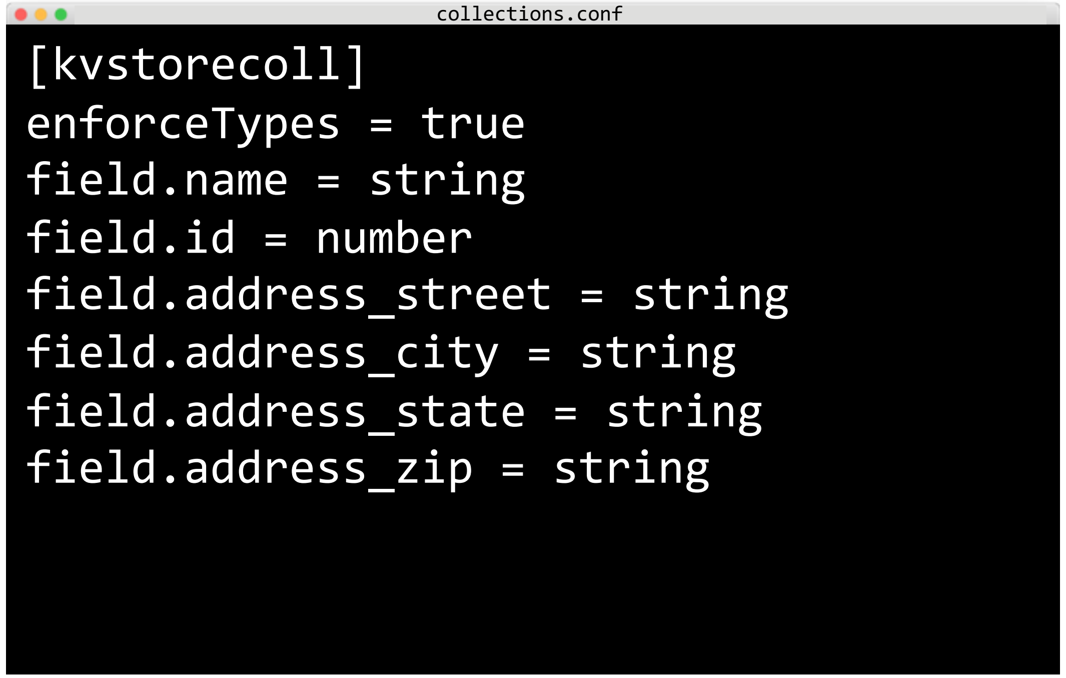
Data type options include:
`number`, `bool`, `string`, `time`

Note

When data type values are enforced, any invalid value added to a collection causes record insertion to fail.

Set Up a KV Store: `collections.conf` (cont.)

- Enforcing data types is useful if you want to:
  - Guarantee a field is always treated as a specific data type
  - Improve the collection’s accelerations (beyond the scope of this course)
- For example, an admin would create the following configuration stanza to enforce the data types of this JSON record



collections.conf

```
[kvstorecoll]
enforceTypes = true
field.name = string
field.id = number
field.address_street = string
field.address_city = string
field.address_state = string
field.address_zip = string
```

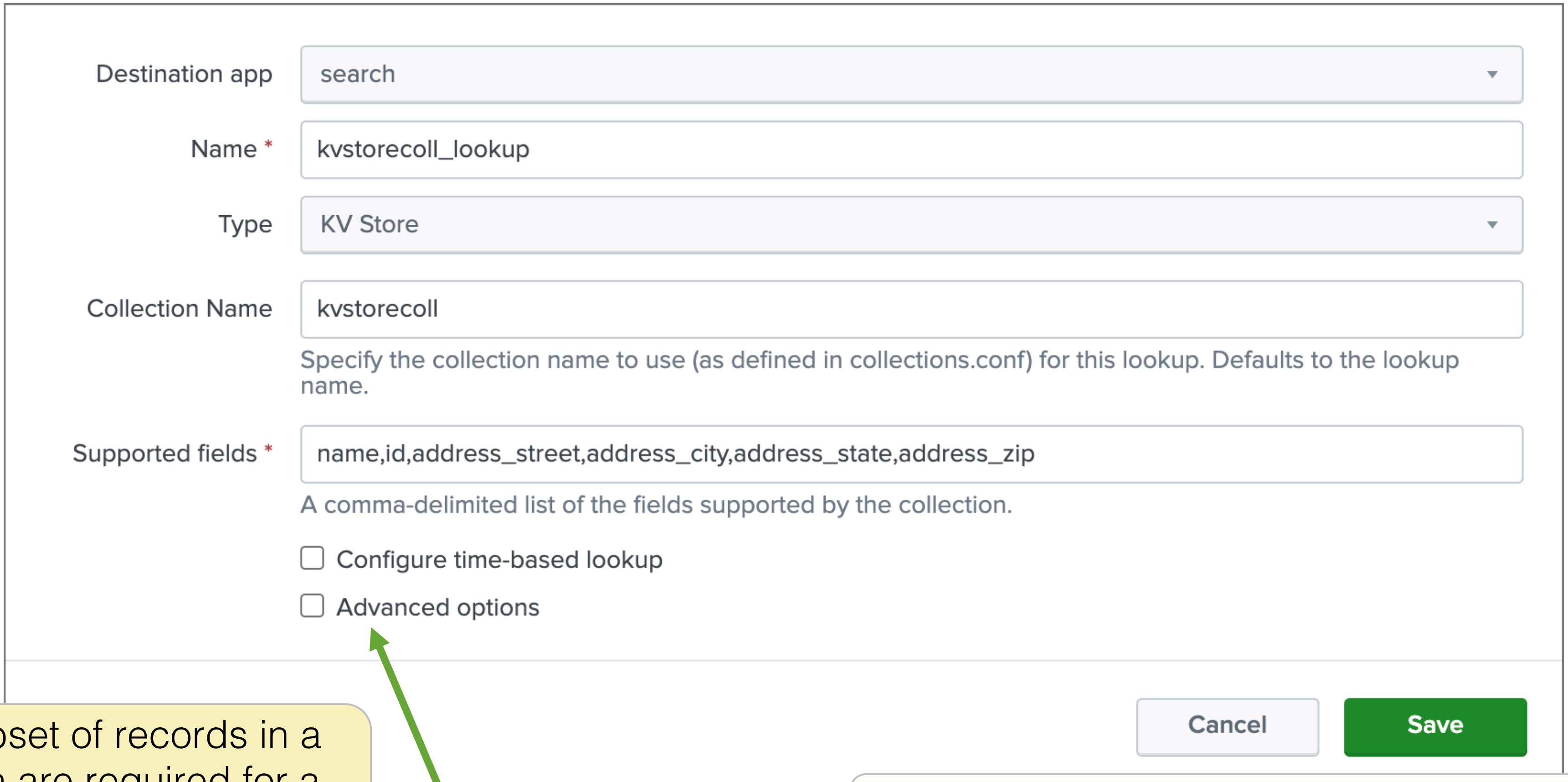
```
{
      "name" : "Splunk Seattle",
      "id" : 123,
      "address" :
      {
            "street" : "1730 Minor Avenue",
            "city" : "Seattle",
            "state" : "WA",
            "zip" : "98101"
      }
}
```

Sample JSON data

Set Up a KV Store: Create a Definition

Settings > Lookups > Lookup definitions > **New Lookup Definition**

1. Choose Destination app
2. Enter Name that will be used in the search string
3. Change to “KV Store”
4. Enter Collection Name as defined in `collections.conf`
5. List all fields supported by the lookup
6. Save



Destination app: search ▾

Name *: kvstorecoll_lookup

Type: KV Store ▾

Collection Name: kvstorecoll

Specify the collection name to use (as defined in collections.conf) for this lookup. Defaults to the lookup name.

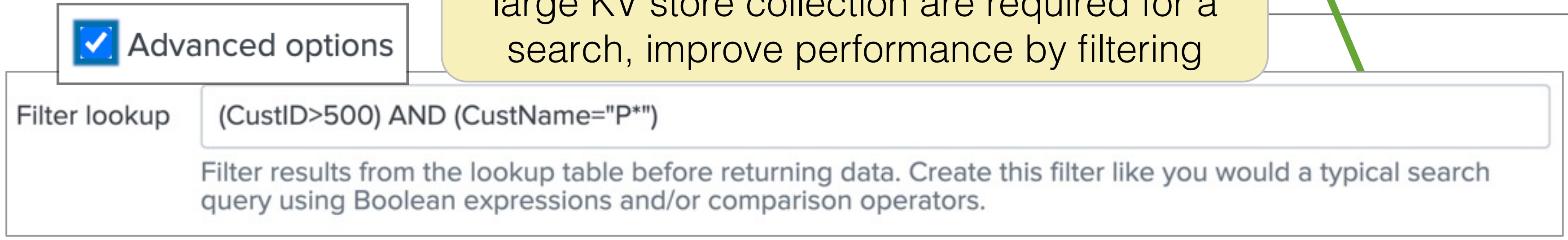
Supported fields *: name,id,address_street,address_city,address_state,address_zip

A comma-delimited list of the fields supported by the collection.

☐ Configure time-based lookup

☐ Advanced options

If only results from a subset of records in a large KV store collection are required for a search, improve performance by filtering



☑ Advanced options

Filter lookup: (CustID>500) AND (CustName="P*")

Filter results from the lookup table before returning data. Create this filter like you would a typical search query using Boolean expressions and/or comparison operators.

Note

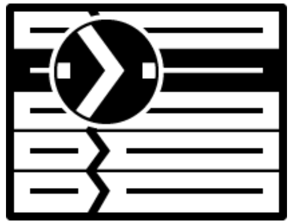
Each collection must have at least two Supported fields. One of these fields must match the values of a field in your event data.

Set Up a KV Store: Populating

Option 1: Use `outputlookup` to write search results into a specific KV Store collection

```
...
|outputlookup kvstorecoll_lookup
```

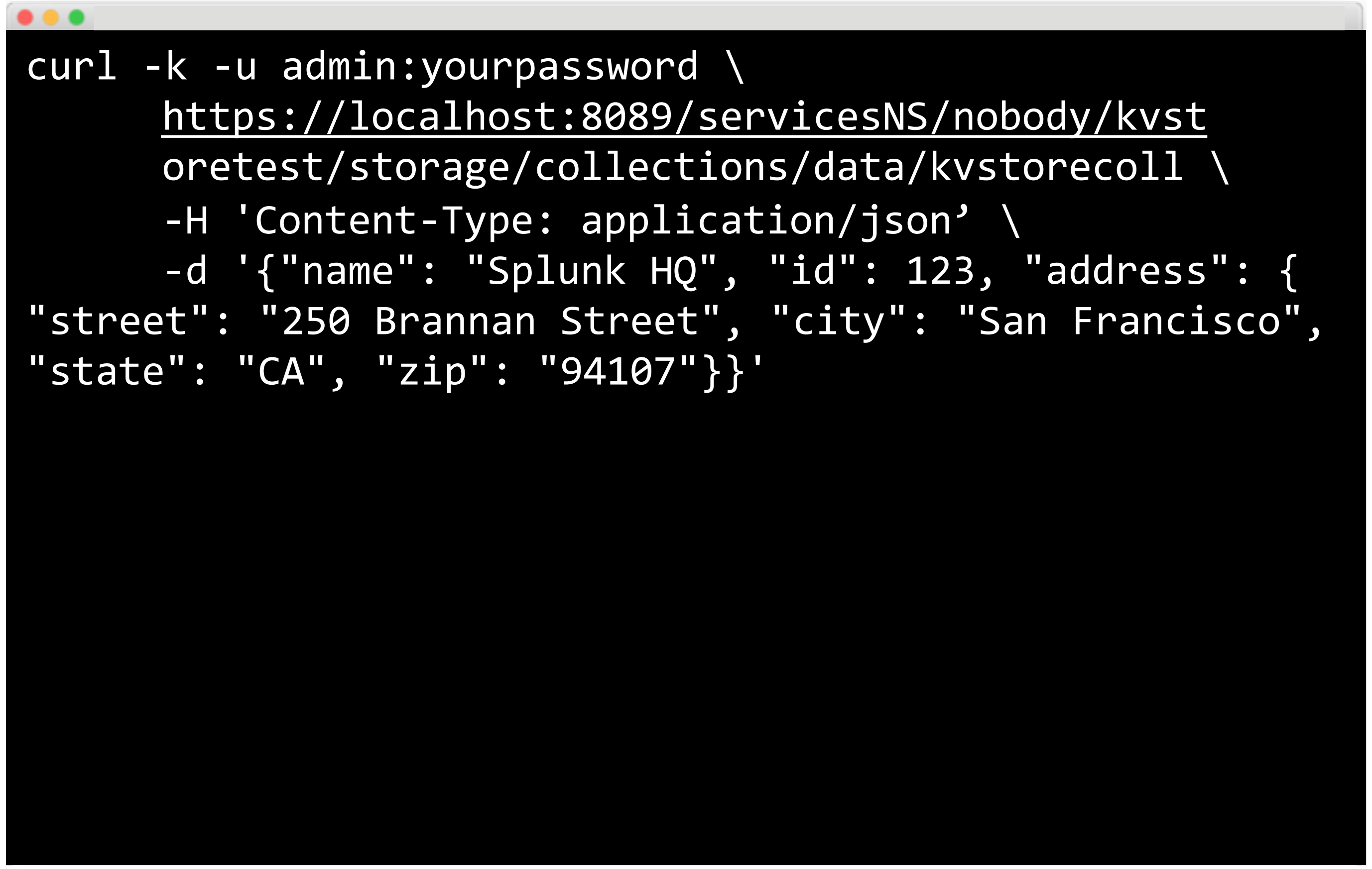
…sends results to…



Note

You must have access to the collection in order to write results using the **`outputlookup`** command.

Option 2: Use Splunk REST API



```
curl -k -u admin:yourpassword \
      https://localhost:8089/servicesNS/nobody/kvst
      oretest/storage/collections/data/kvstorecoll \
      -H 'Content-Type: application/json’ \
      -d '{"name": "Splunk HQ", "id": 123, "address": {
"street": "250 Brannan Street", "city": "San Francisco",
"state": "CA", "zip": "94107"}}'
```

CSV Files vs KV Store

	File-based (CSV)	KV Store
Allows for per-record insertion and editing		✓
Suitable for frequent updating		✓
Allows for data type enforcement		✓
Allows for field accelerations		✓
Provides REST API access to the data collection		✓
Require a full rewrite of the file to edit values	✓	
Supports case-sensitive field lookups	✓	✓
Supports case-insensitive field lookups	✓	
Uploading lookup file is not mandatory		✓
Allows for multiuser access locking		✓
Matches against values in a KV Store		✓
Matches against values in a .csv file	✓	

Best Practices for Lookups

Best Practices for Lookups

splunk>

Best Practice

- Order fields in lookup tables so that 'key' field is first (leftmost), followed by other values
- After uploading, validate lookup in search by using:

```
| inputlookup <lookup>
```

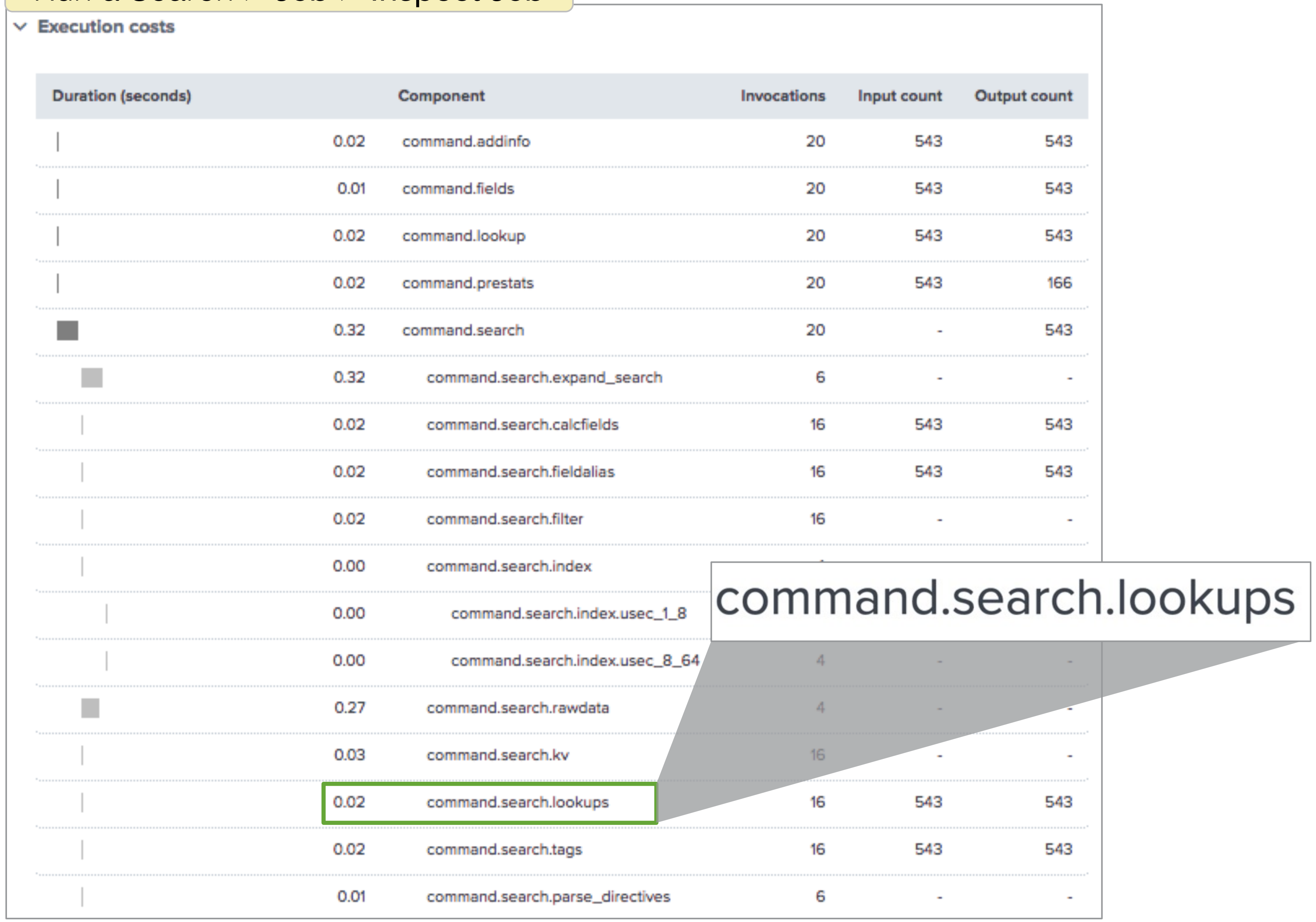
- For commonly used fields, make lookups automatic
- Use gzipped CSV files or KV Store for large lookups
- Keep your lookups fresh and relevant:
  - Do you really need the lookup table to contain a year's worth of data or is one week enough?
  - Maintain the lookup table and delete older data if not needed

key	value
…	…
…	…
…	…
…	…

Best Practices for Lookups (cont.)

Best Practice

Run a Search > Job > Inspect Job



Execution costs

Duration (seconds)	Component	Invocations	Input count	Output count
0.02	command.addinfo	20	543	543
0.01	command.fields	20	543	543
0.02	command.lookup	20	543	543
0.02	command.prestats	20	543	166
0.32	command.search	20	-	543
0.32	command.search.expand_search	6	-	-
0.02	command.search.calcfields	16	543	543
0.02	command.search.fieldalias	16	543	543
0.02	command.search.filter	16	-	-
0.00	command.search.index			
0.00	command.search.index.usec_1_8			
0.00	command.search.index.usec_8_64	4	-	-
0.27	command.search.rawdata	4	-	-
0.03	command.search.kv	16	-	-
0.02	command.search.lookups	16	543	543
0.02	command.search.tags	16	543	543
0.01	command.search.parse_directives	6	-	-

command.search.lookups

- `command.search.lookups` in job inspector will show how long lookups took to execute
- If there is latency, see if there is one or many lookups being invoked against large files/tables

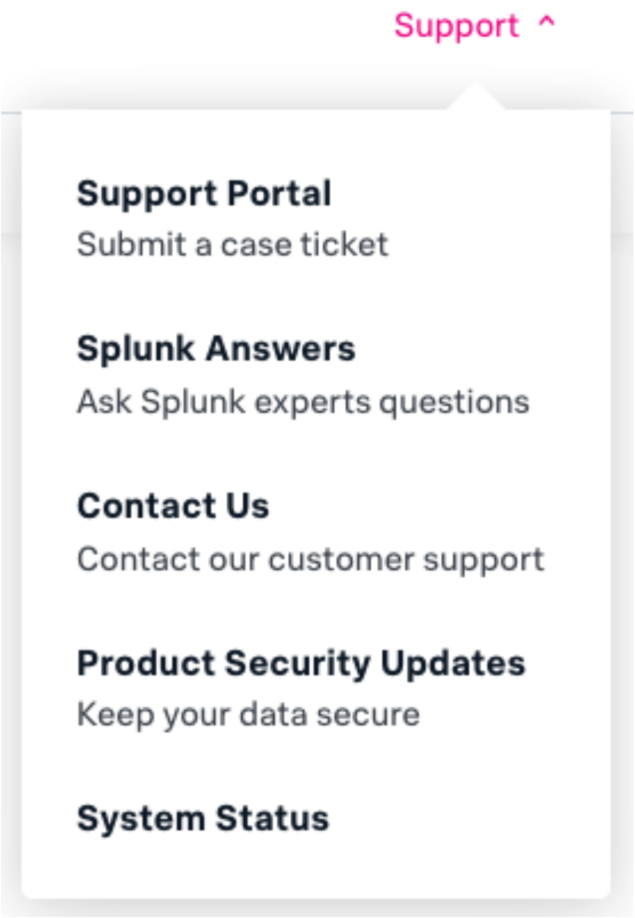
Wrap-up Slides

Community

- Splunk Community Portal
  community.splunk.com
  - Answers
  - Discussions
  - Splunk Trust
  - User Groups
  - Ideas
- Splunk Blogs
  splunk.com/blog/
- Splunk Apps
  splunkbase.com
- Splunk Dev Google Group
  groups.google.com/forum/#!forum/splunkdev
- Splunk Docs on Twitter
  twitter.com/splunkdocs
- Splunk Dev on Twitter
  twitter.com/splunkdev
- Splunk Live!
  splunklive.splunk.com
- .conf
  conf.splunk.com

Support Programs

- Web
  - Documentation: dev.splunk.com and docs.splunk.com
  - Wiki: wiki.splunk.com
- Splunk Lantern
  Guidance from Splunk experts
  - lantern.splunk.com
- Global Support
  Support for critical issues, a dedicated resource to manage your account – 24 x 7 x 365
  - Web: splunk.com/index.php/submit_issue
- Enterprise, Cloud, ITSI, Security Support
  - Web: splunk.com/en_us/about-splunk/contact-us.html#tabs/customersupport
  - Phone: (855) SPLUNK-S or (855) 775-8657



Support ^

Support Portal
Submit a case ticket

Splunk Answers
Ask Splunk experts questions

Contact Us
Contact our customer support

Product Security Updates
Keep your data secure

System Status

Learning Paths (cont.)

Knowledge Manager - Recommended Courses

Free eLearning courses are in blue and courses with an * are present in both learning paths.

- What is Splunk *
- Introduction to Splunk *
- Using Fields *
- Introduction to Knowledge Objects
- Creating Knowledge Objects
- Creating Field Extractions
- Enriching Data with Lookups
- Data Models
- Introduction to Dashboards
- Dynamic Dashboards
- Using Choropleth
- Search Optimization *

Learning Paths

Search Expert - Recommended Courses

Free eLearning courses are in blue and courses with an * are present in both learning paths.

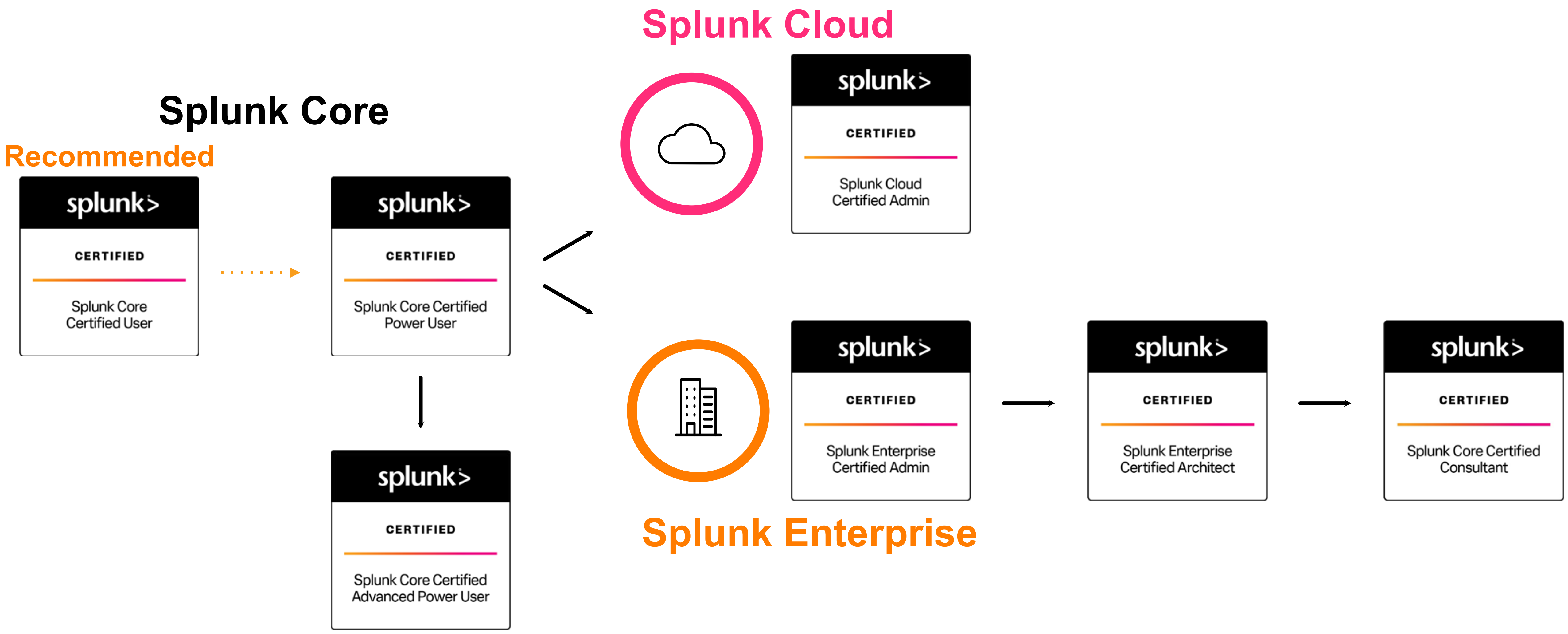
- What is Splunk *
- Introduction to Splunk *
- Using Fields *
- Scheduling Reports and Alerts
- Visualizations
- Statistical Processing
- Working with Time
- Comparing Values
- Result Modification
- Leveraging Lookups and Subsearches
- Correlation Analysis
- Search Under the Hood
- Multivalue Fields
- Search Optimization *

Splunk Certification

Offerings & Requirements

Splunk Core and Beyond

Regardless of which Splunk product you use, it all starts with Splunk Core



Splunk Core

Recommended

splunk> CERTIFIED Splunk Core Certified User

splunk> CERTIFIED Splunk Core Certified Power User

splunk> CERTIFIED Splunk Core Certified Advanced Power User

Splunk Cloud

splunk> CERTIFIED Splunk Cloud Certified Admin

Splunk Enterprise

splunk> CERTIFIED Splunk Enterprise Certified Admin

splunk> CERTIFIED Splunk Enterprise Certified Architect

splunk> CERTIFIED Splunk Core Certified Consultant

| 12 August 2022

Splunk Core Certified User

This entry-level certification demonstrates an individual's basic ability to navigate and use Splunk software

Prerequisite Certification(s):

- None

Prerequisite Course(s):

- None

Splunk Core Certified User Exam

Time to study! We suggest candidates looking to prepare for this exam complete Fundamentals 1 **or** the following courses:

- What is Splunk?
- Intro to Splunk
- Using Fields
- Scheduling Reports and Alerts
- Visualizations
- Statistical Processing
- Working with Time
- Leveraging Lookups and Subsearches
- Search Optimization
- Enriching Data with Lookups
- Data Models

See here for registration assistance.

Congratulations! You are a...

splunk>

CERTIFIED

Splunk Core
Certified User

Recommended Next Step

- Splunk Core Certified Power User

Splunk Core Certified Advanced Power User

This certification demonstrates an individual's ability to generate complex searches, reports, and dashboards with Splunk’s core software to get the most out of their data

Prerequisite Certification(s):

- Splunk Core Certified Power User

Prerequisite Course(s):

- None

Splunk Core Certified Advanced Power User Exam

Time to study! We suggest candidates looking to prepare for this exam complete Fundamentals 3, Creating Dashboards, and Advanced Searching & Reporting **or** the following courses:

- Using Fields
- Working with Time
- Comparing Values
- Result Modification
- Leveraging Lookups and Subsearches
- Correlation Analysis
- Search Under the Hood
- Multivalue Fields
- Search Optimization
- Creating Field Extractions
- Enriching Data with Lookups
- Data Models
- Using Choropleth
- Introduction to Dashboards
- Dynamic Dashboards

See here for registration assistance.

Congratulations! You are a...

splunk>

CERTIFIED

Splunk Core Certified
Advanced Power User

Recommended Next Steps

- Splunk Enterprise Certified Admin
- Splunk Cloud Certified Admin

Thank You