Data Models

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 and identify uses of data models and datasets
- Create a data model
- Use a data model in Pivot
- Describe data model acceleration options
- Define `tsidx` files and their role in data model acceleration

Course Outline

- Introduce Data Model Datasets
- Design Data Models
- Create a Pivot
- Accelerate Data Models

Introduce
Data Model Datasets

Topic Objectives

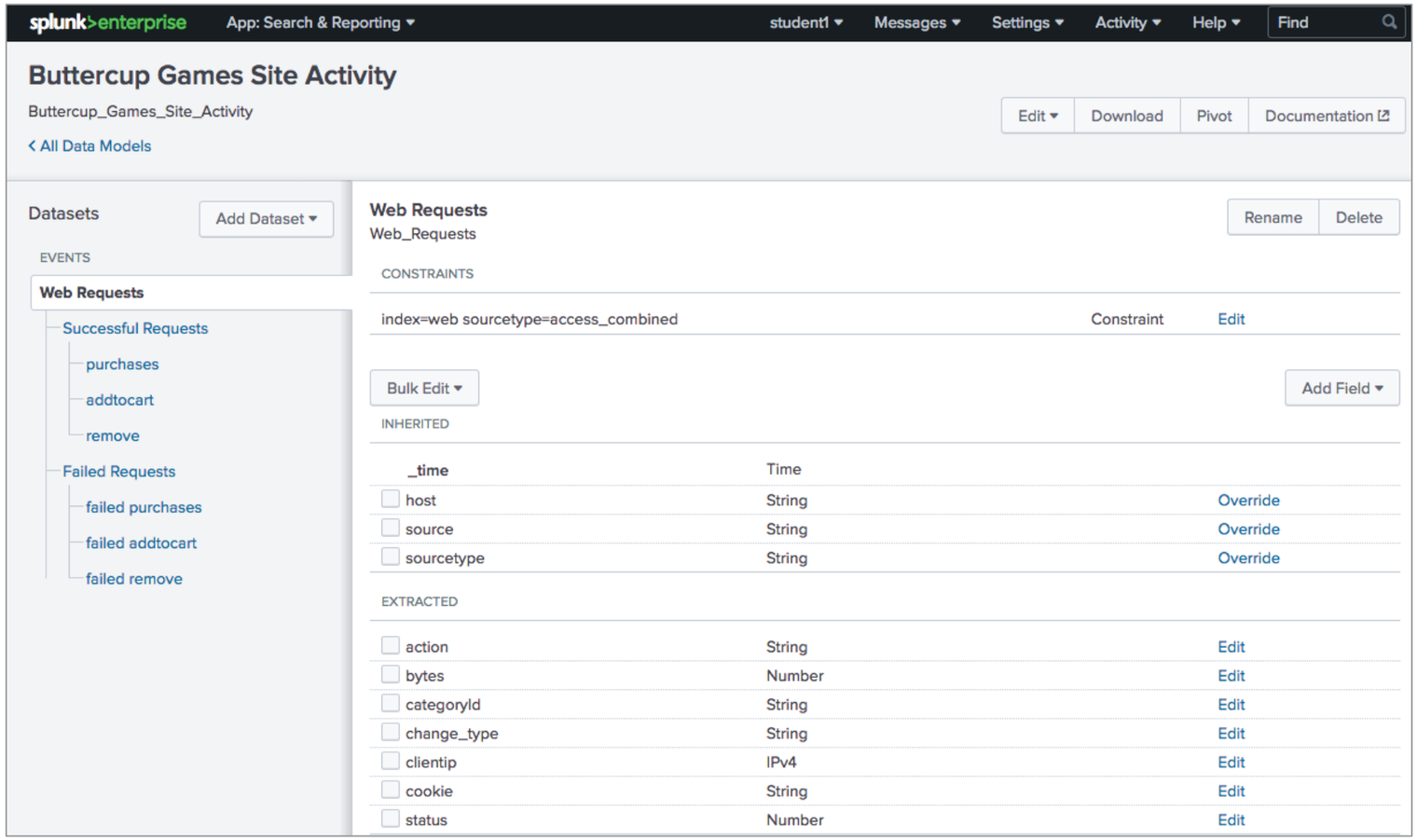
- Define data models
- Add event, search, and transaction datasets to data models
- Identify event object hierarchy and constraints
- Add fields based on **`eval`** expressions to transaction datasets

Data Models

- A data model is a knowledge object that applies an information structure to raw data, making it easier to use
- Each data model is designed by the user to represent a specific category of event data
- The Pivot interface can use data models to generate reports and dashboard panels
- Data models can be accelerated for faster performance

Overview of Data Model Datasets

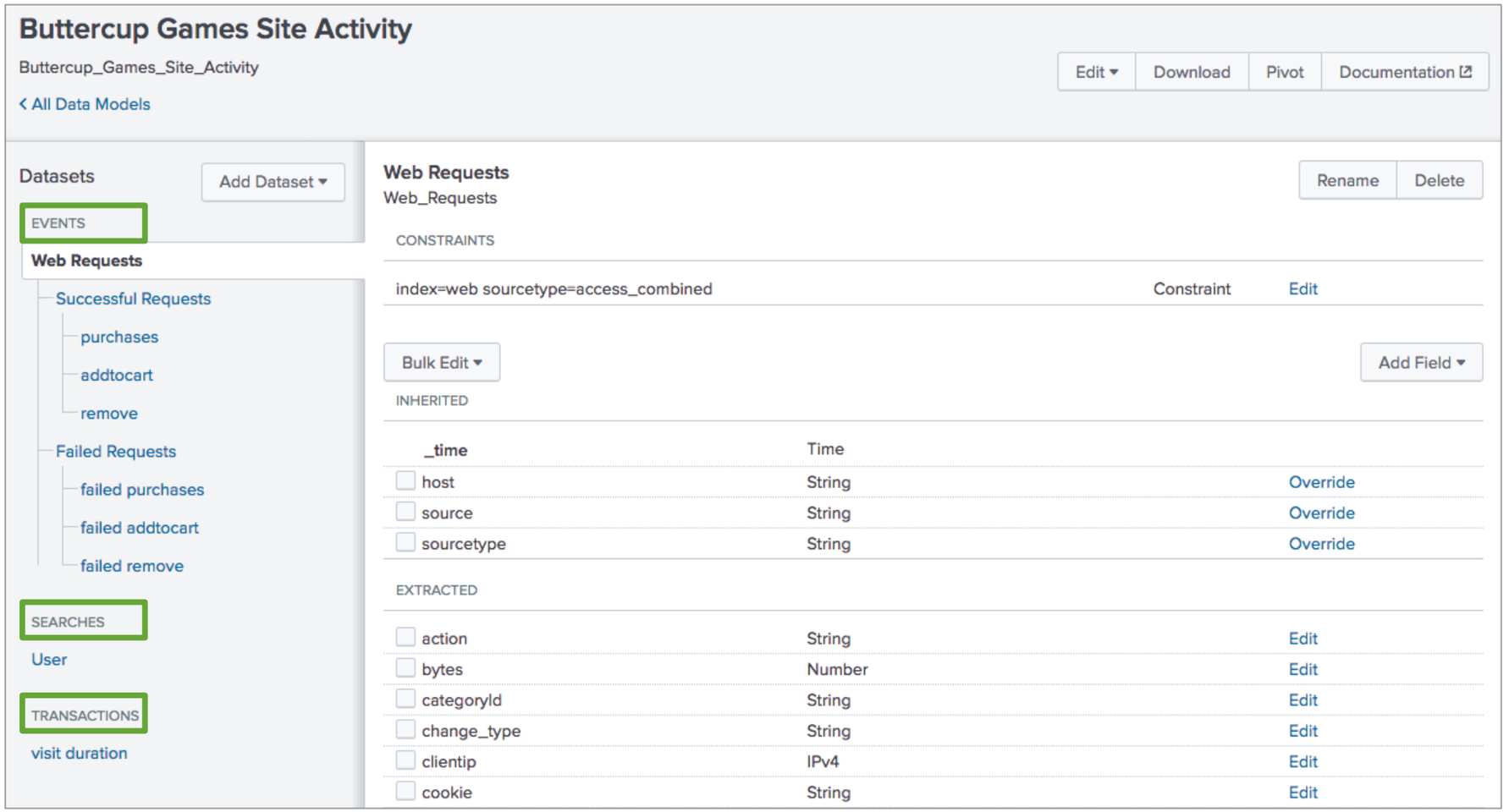
- Data models are hierarchically structured datasets containing searches and fields
- Each event, search, or transaction is saved as a separate dataset



Data Model Dataset Types

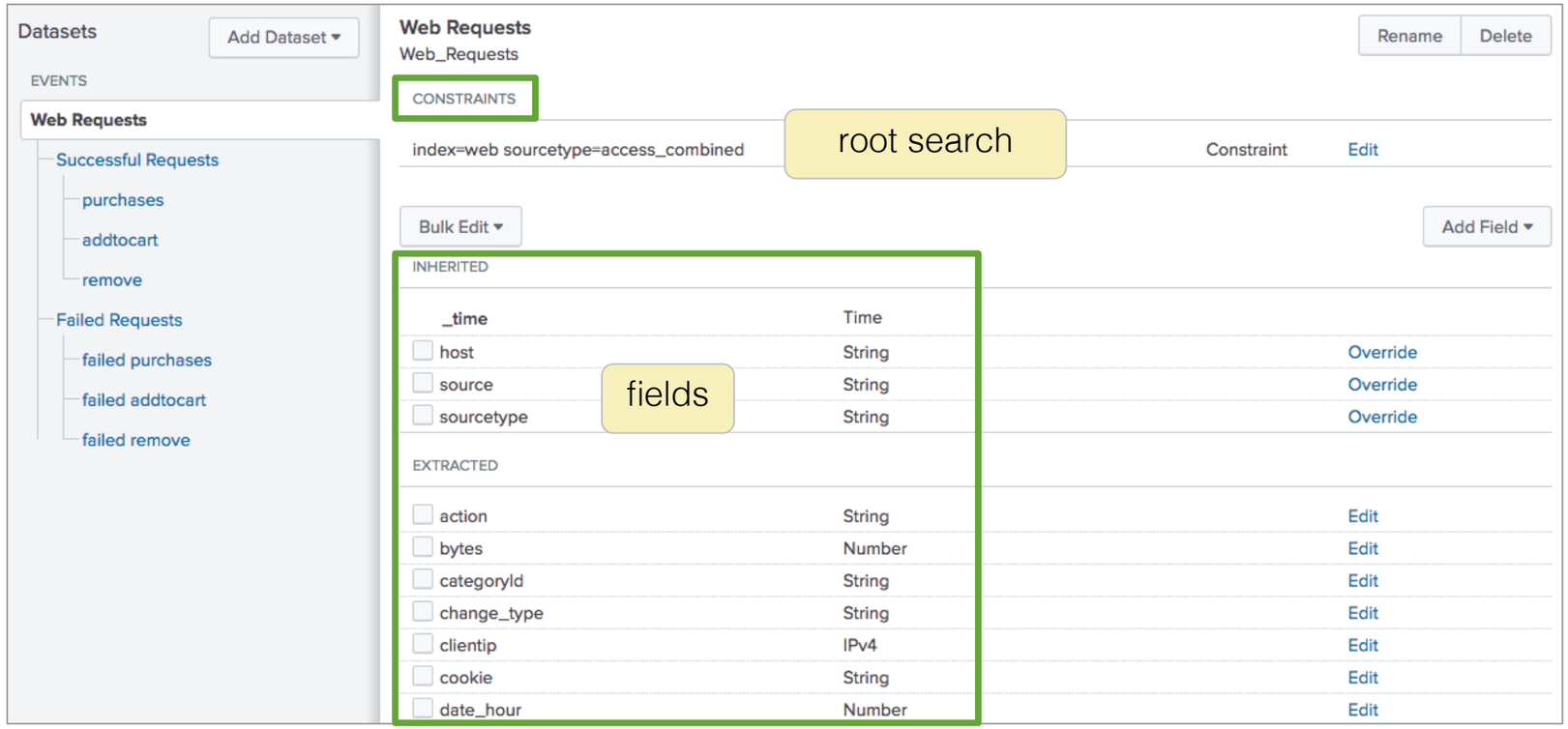
Data Models can contain 3 types of datasets:

- Events
- Searches
- Transactions

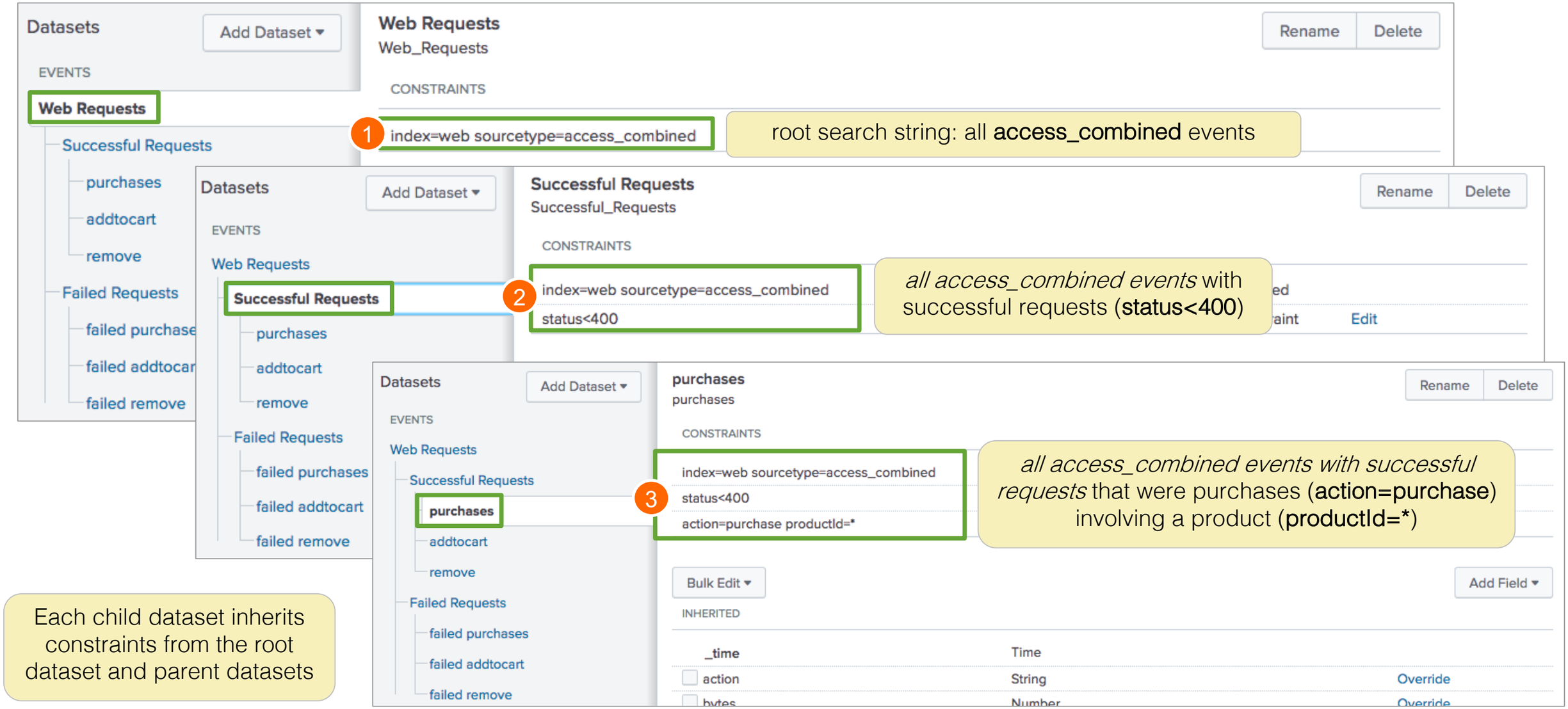


Data Model Event Datasets

- All event datasets (root and children) contain:
  - **Constraints**: essentially the search broken down into a hierarchy
  - **Fields**: associated with the events



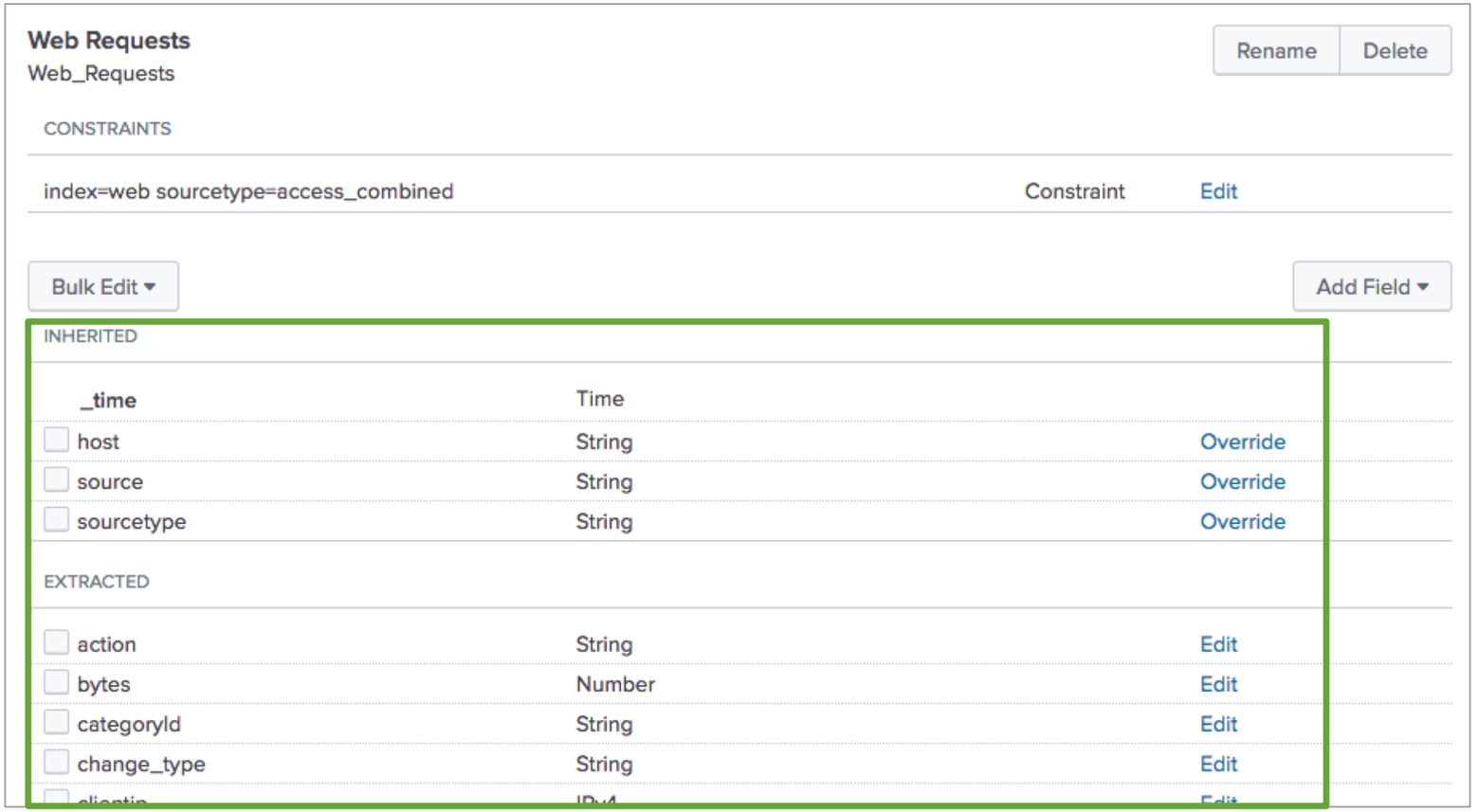
Event Object Hierarchy and Constraints



Each child dataset inherits constraints from the root dataset and parent datasets

Dataset Fields

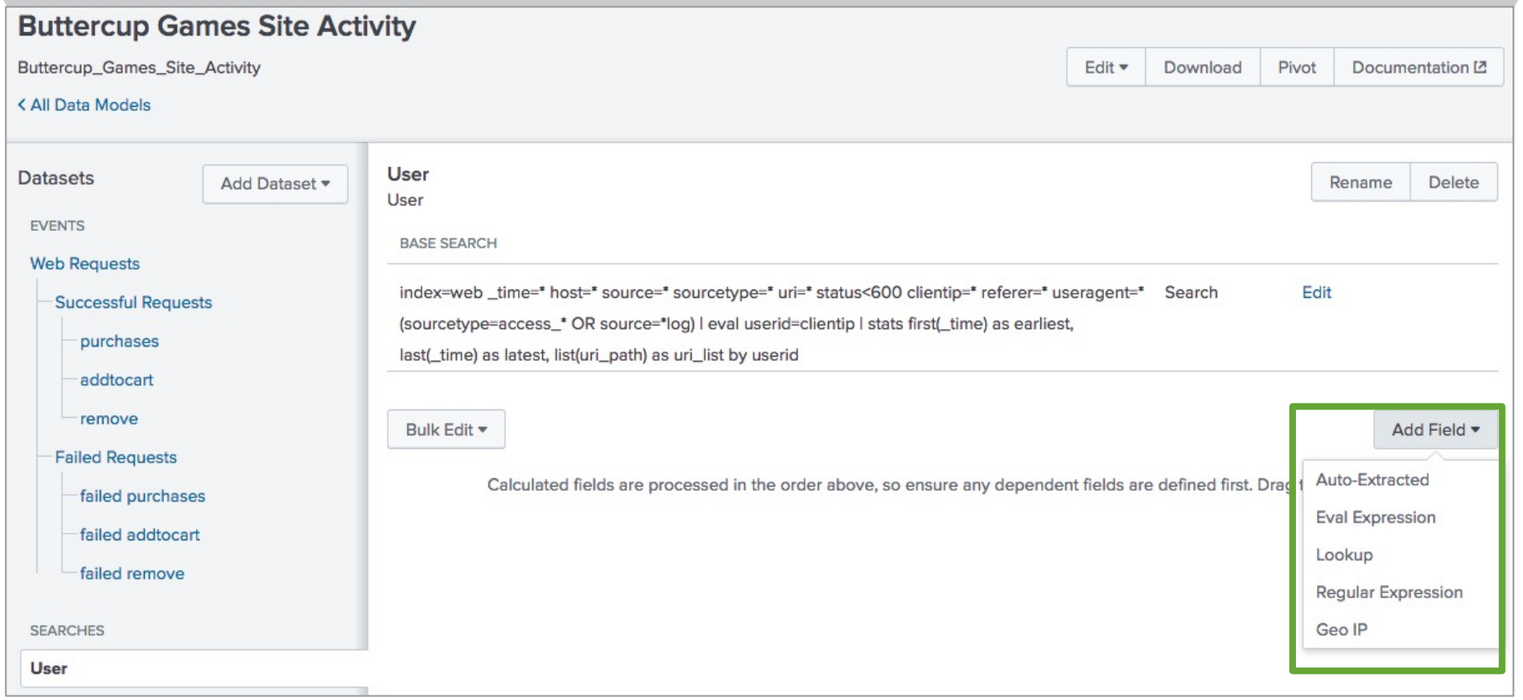
- Select the fields you want to include in the dataset
- Like constraints, fields are inherited from parent datasets



Data Model Root Search Datasets

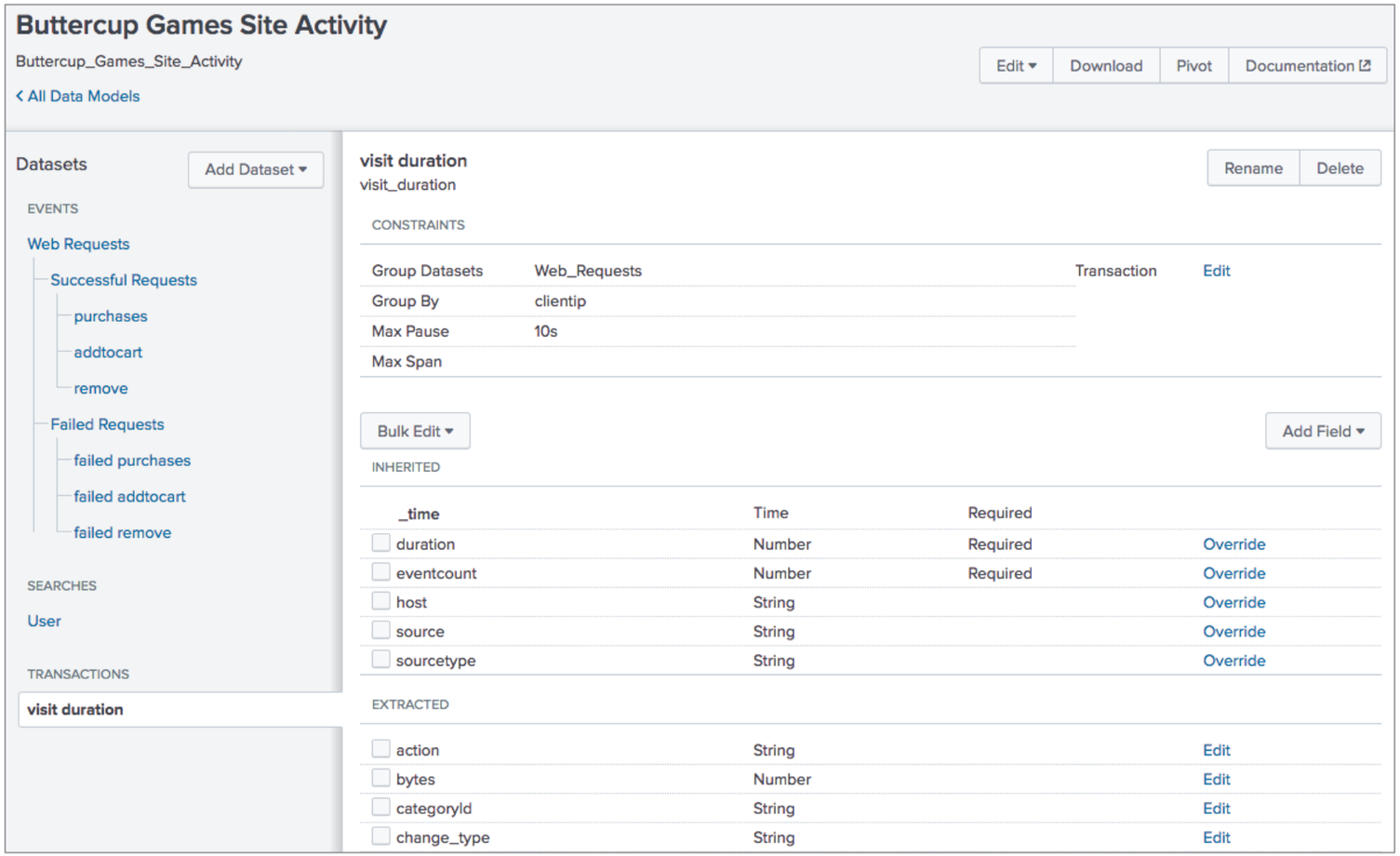
- Based on an arbitrary search to define the dataset it represents
- Include a transforming command if you want to include field(s) that aggregate over the entire dataset
- Use the **Add Field** button to add more fields

```
index=web _time=* host=* source=* sourcetype=* uri=* status<600 clientip=* referer=* useragent=* (sourcetype=access_* OR source=*log)
| eval userid=clientip
| stats first(_time) as earliest, last(_time) as latest, list(uri_path) as uri_list by userid
```



Data Model Root Transaction Datasets

- Based on a transaction
- Uses fields that have already been added to the data model using either event or search datasets



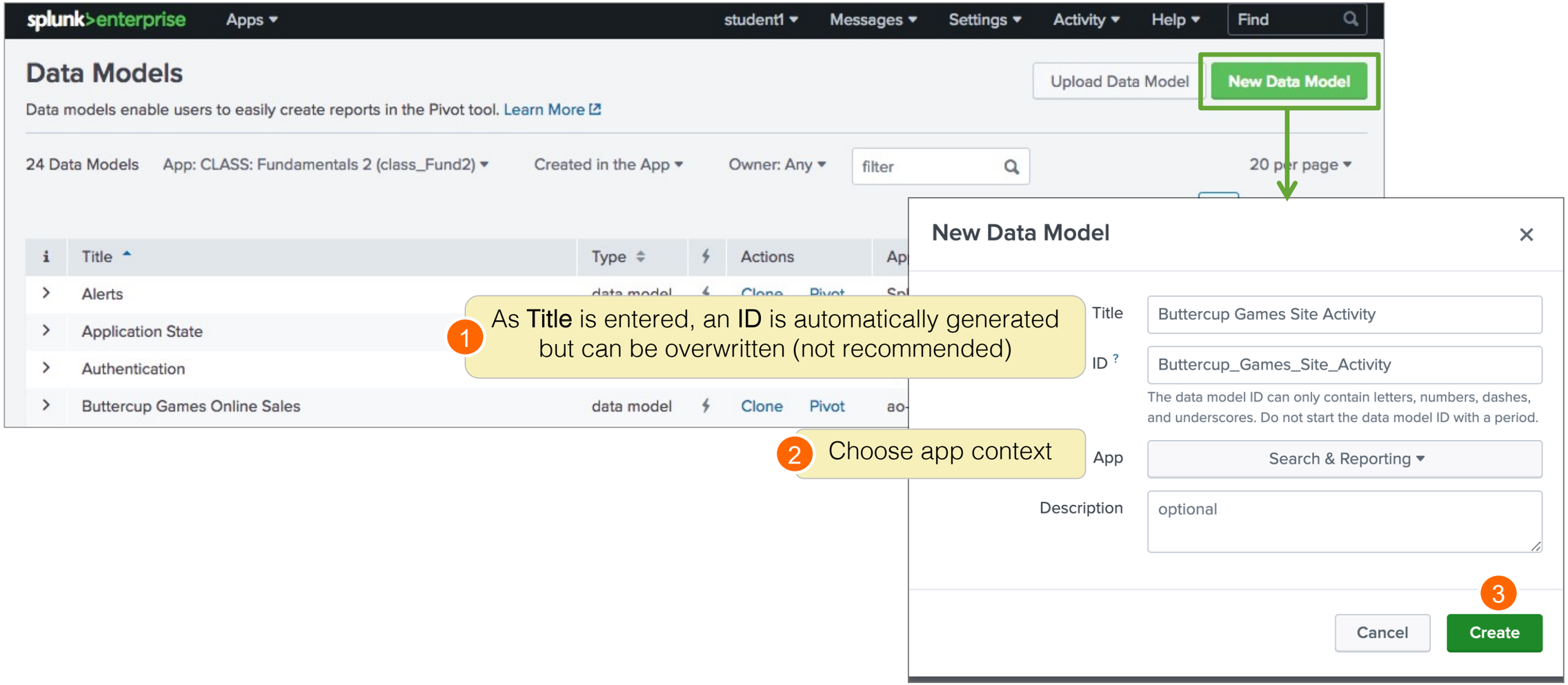
Design Data Models

Topic Objectives

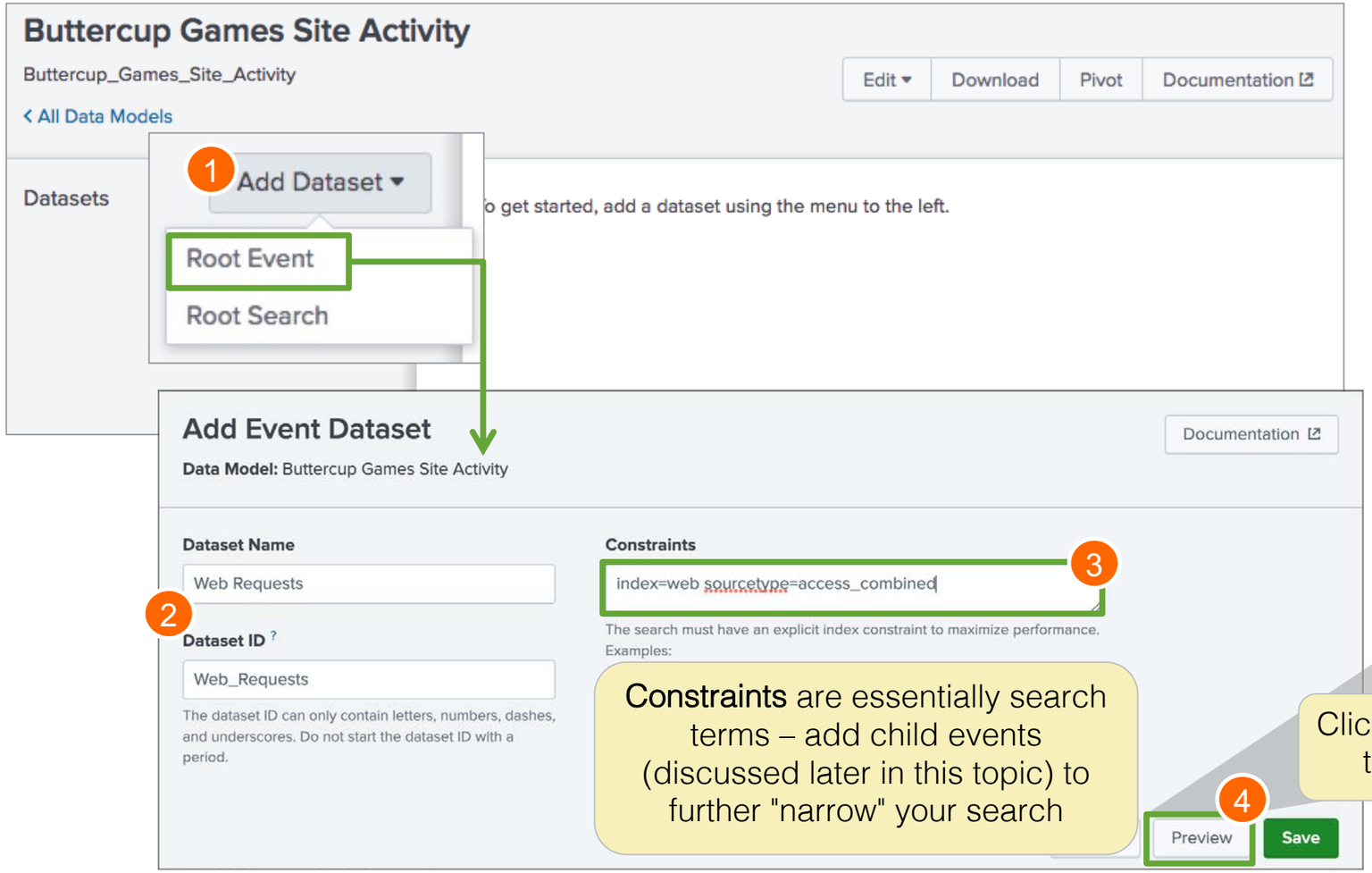
- Create a data model
- Add root event dataset to a data model
- Add fields to data models
- Add child datasets to a data model
- Test a data model
- Define permissions for a data model
- Upload/download data models for backup and sharing

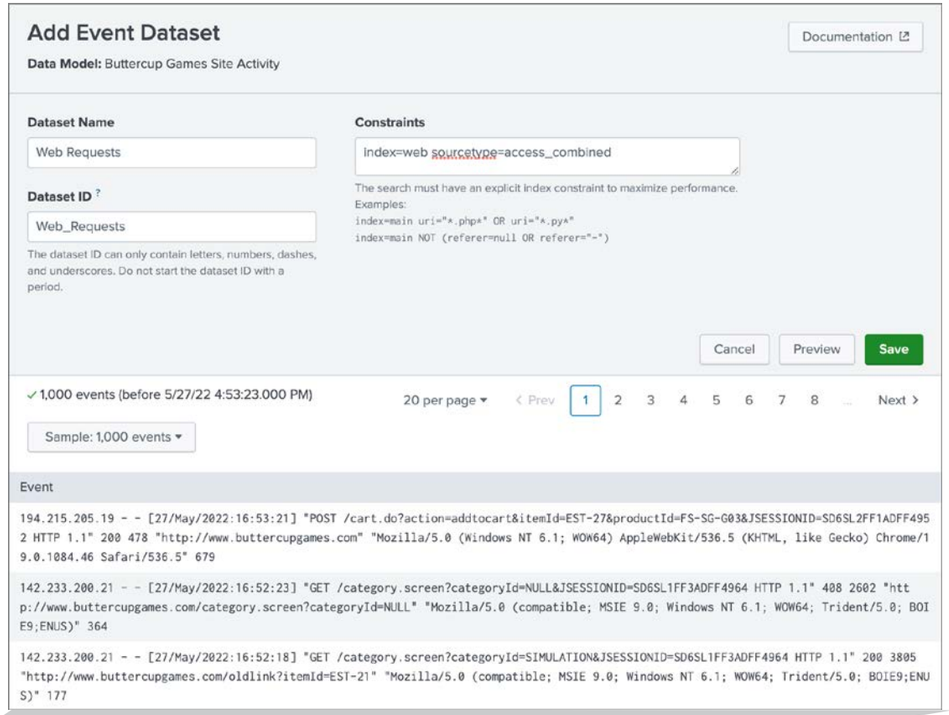
Create a Data Model

Settings > Data models



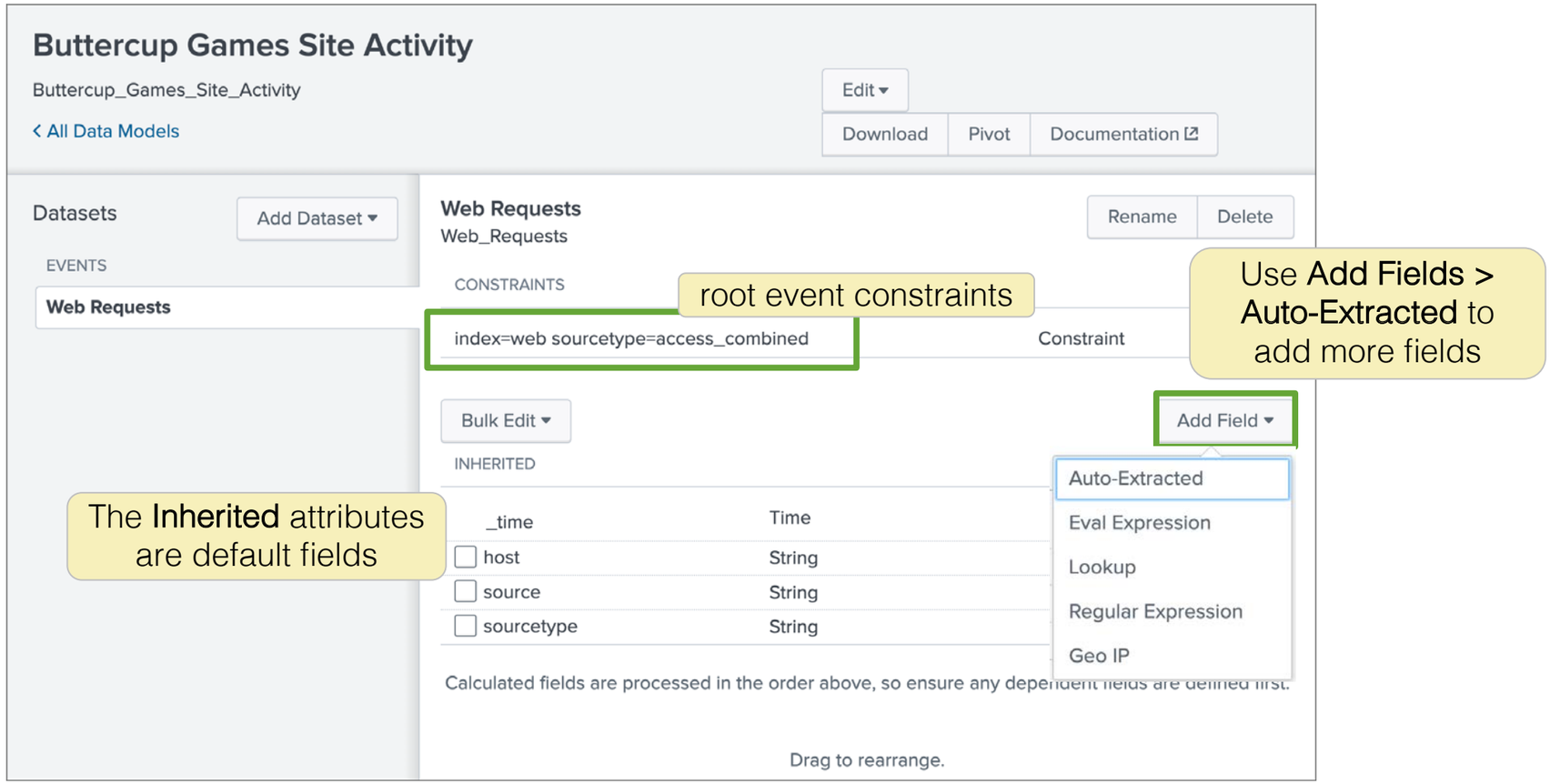
Add a Root Event





Add a Root Event (cont.)

In this example, the root event of this data model represents all web requests

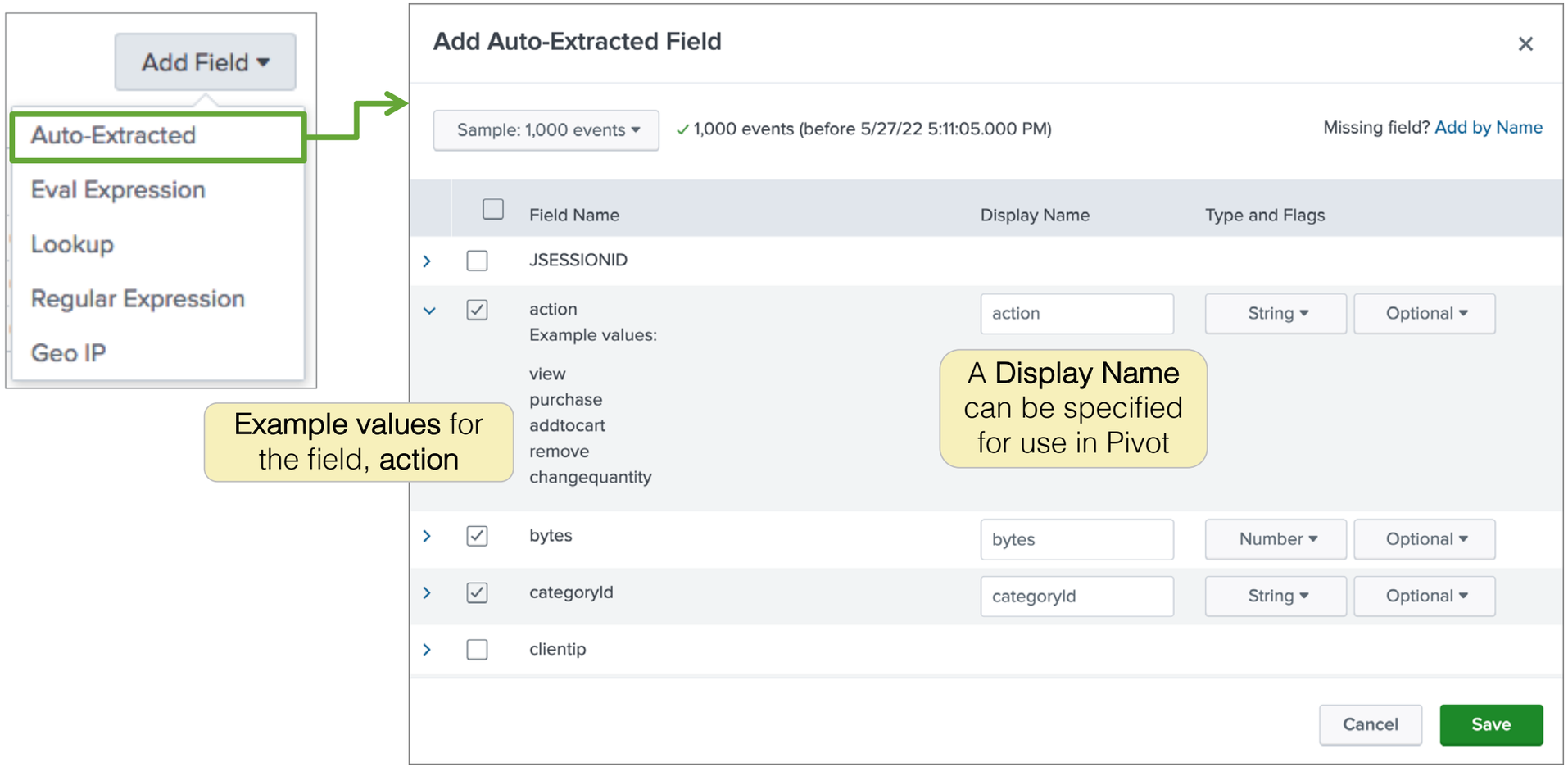


Add Fields

- **Auto-Extracted:** default fields or manually extracted fields
- **Eval Expression:** new field based on an expression you define
- **Lookup:** leverage an existing lookup table
- **Regular Expression:** extract a new field based on regex
- **Geo IP:** geographical fields such as latitude/longitude or country

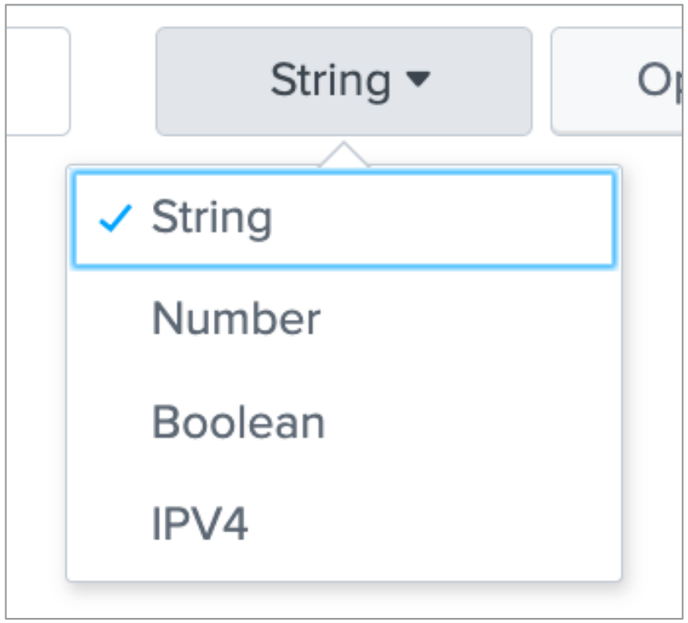
Add Fields: Auto-Extracted

Check auto-extracted fields to add them to the dataset



Field Types

- **String**: Recognized as alphanumeric
- **Number**: Recognized as numeric
- **Boolean**: Recognized as true/false or 1/0
- **IPV4**: Recognized as IP addresses

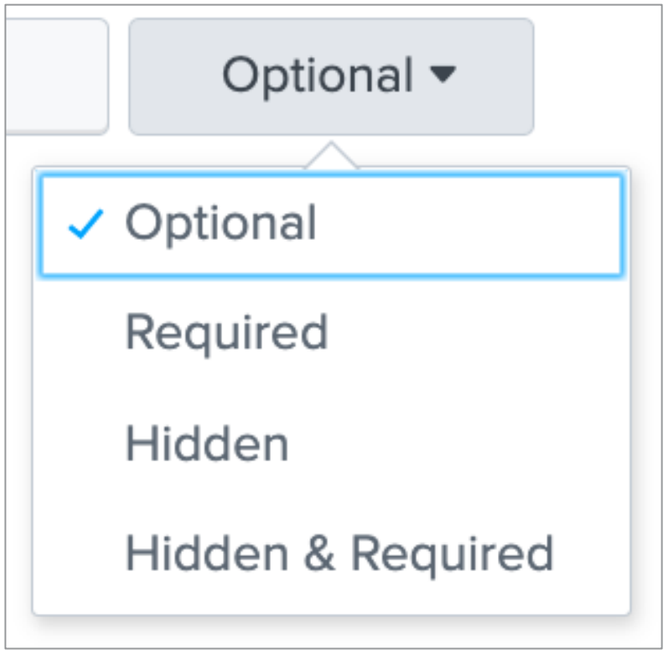


Note

At least one IPV4 attribute type must be present in the data model in order to add a Geo IP attribute.

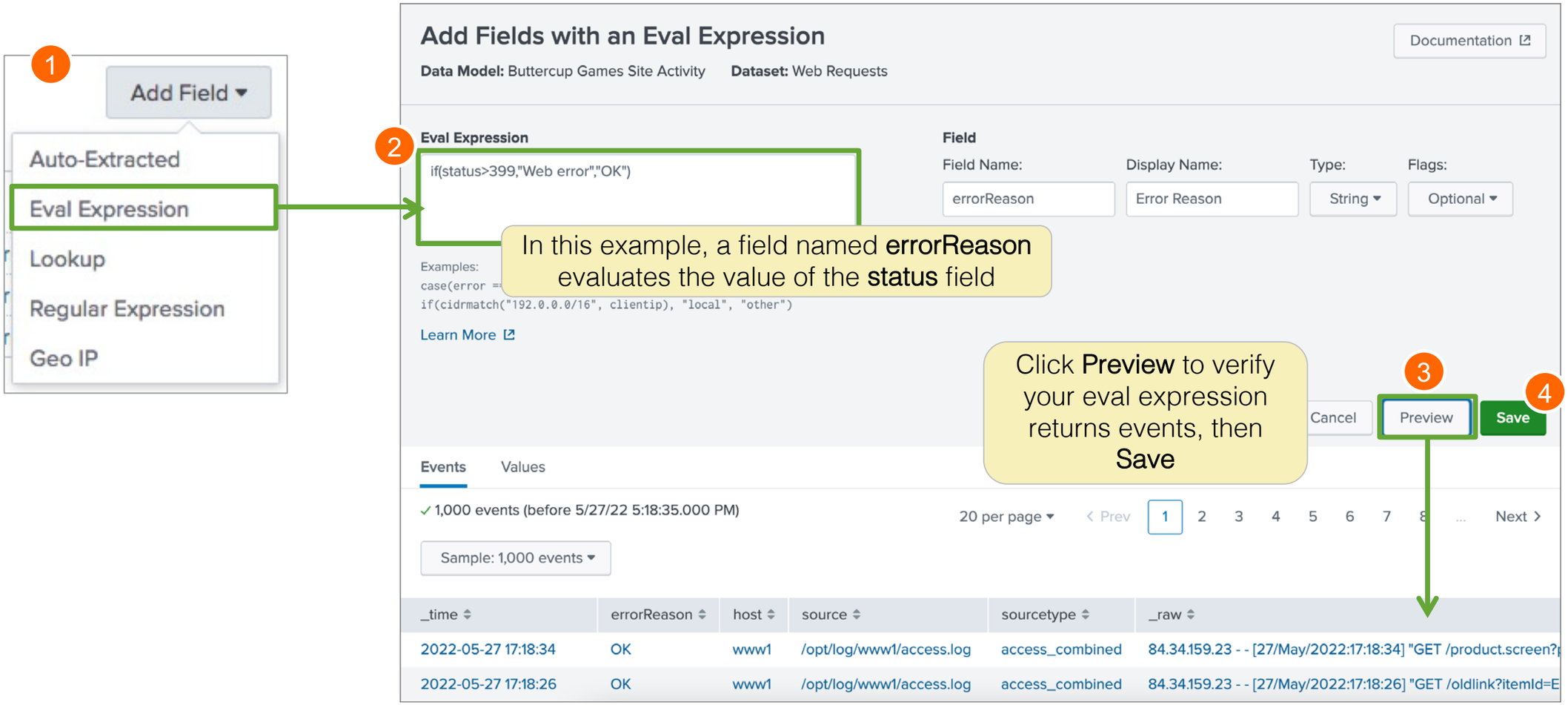
Field Flags

- **Optional**: Field doesn't have to appear in every event
- **Required**: Only events that contain this field are returned in Pivot
- **Hidden**: Field is not displayed to Pivot users when they select the dataset in Pivot
- **Hidden & Required**: Only events that contain this field are returned, and the fields are hidden from use in Pivot



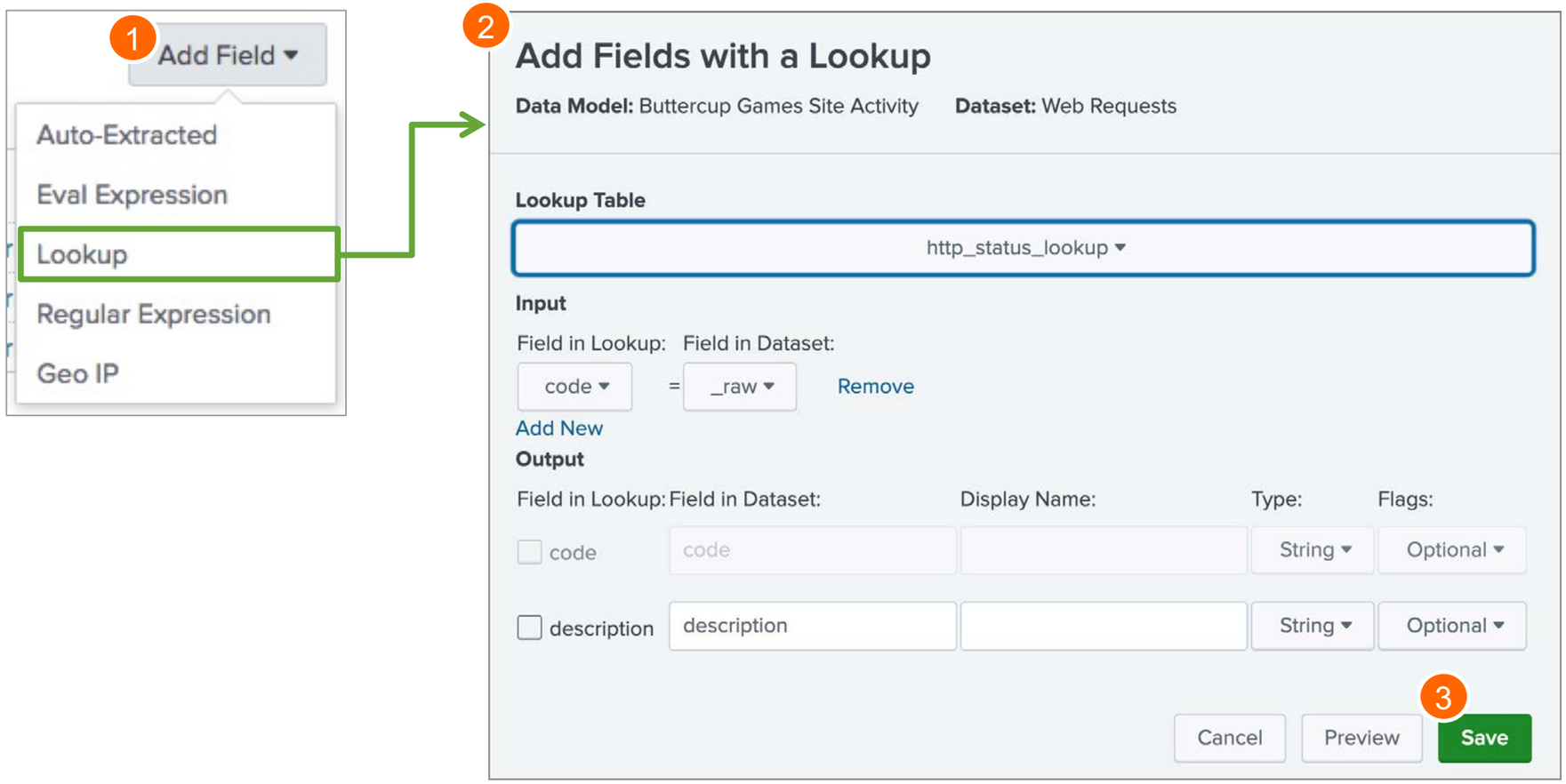
Add Fields: Eval Expressions

Define a new field using an eval expression



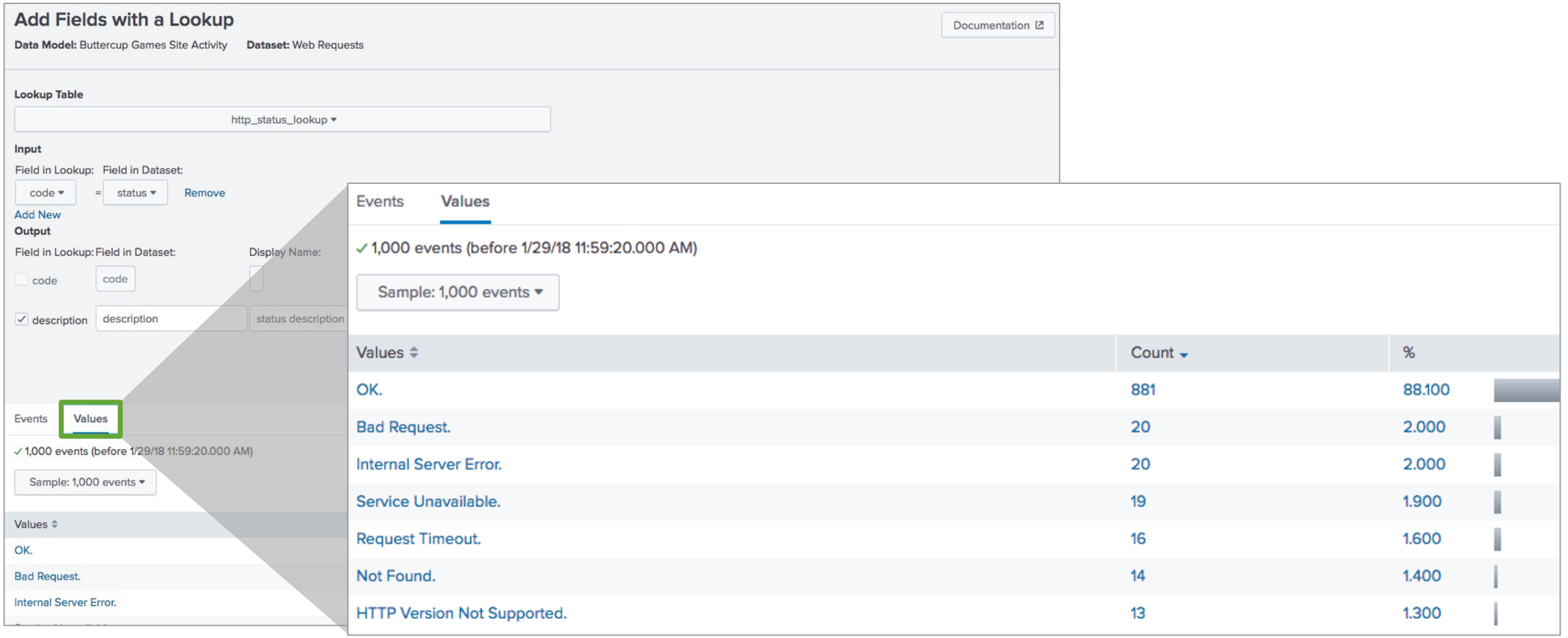
Add Fields: Lookups

- Leverage an existing lookup definition to add fields to the dataset
- Configuration is similar to setting up an automatic lookup



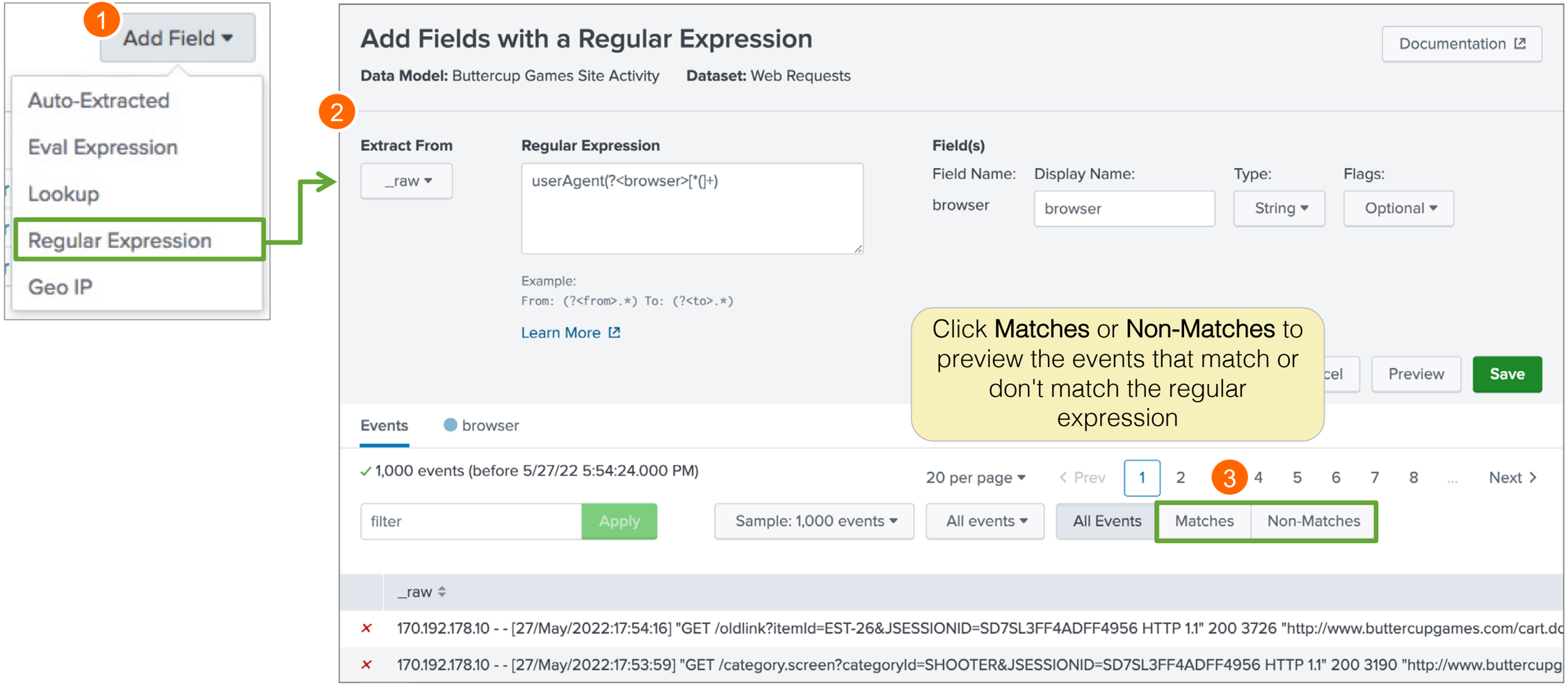
Add Fields: Lookups (cont.)

Click **Preview** and use the events and values tabs to verify your results



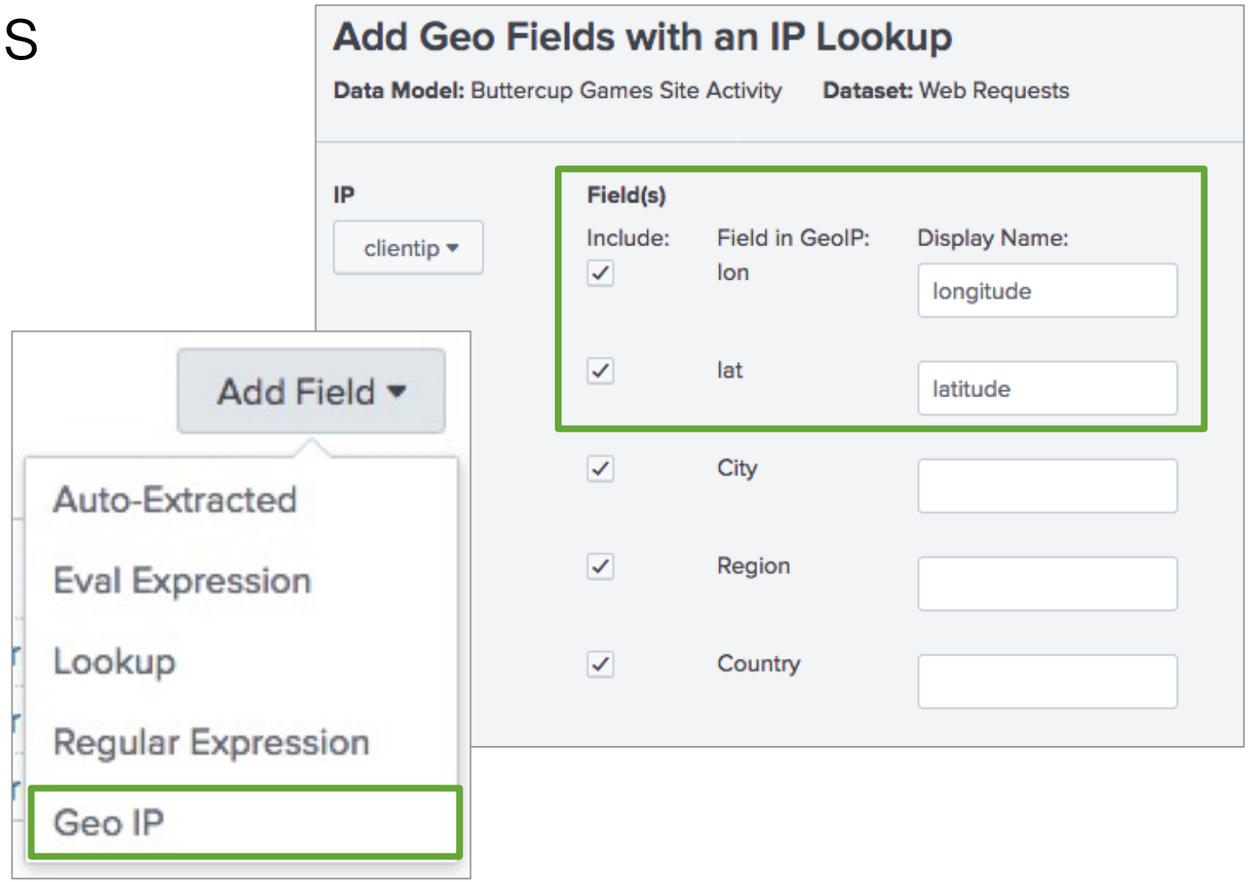
Add Fields: Regular Expression

You can define a new field using a regular expression



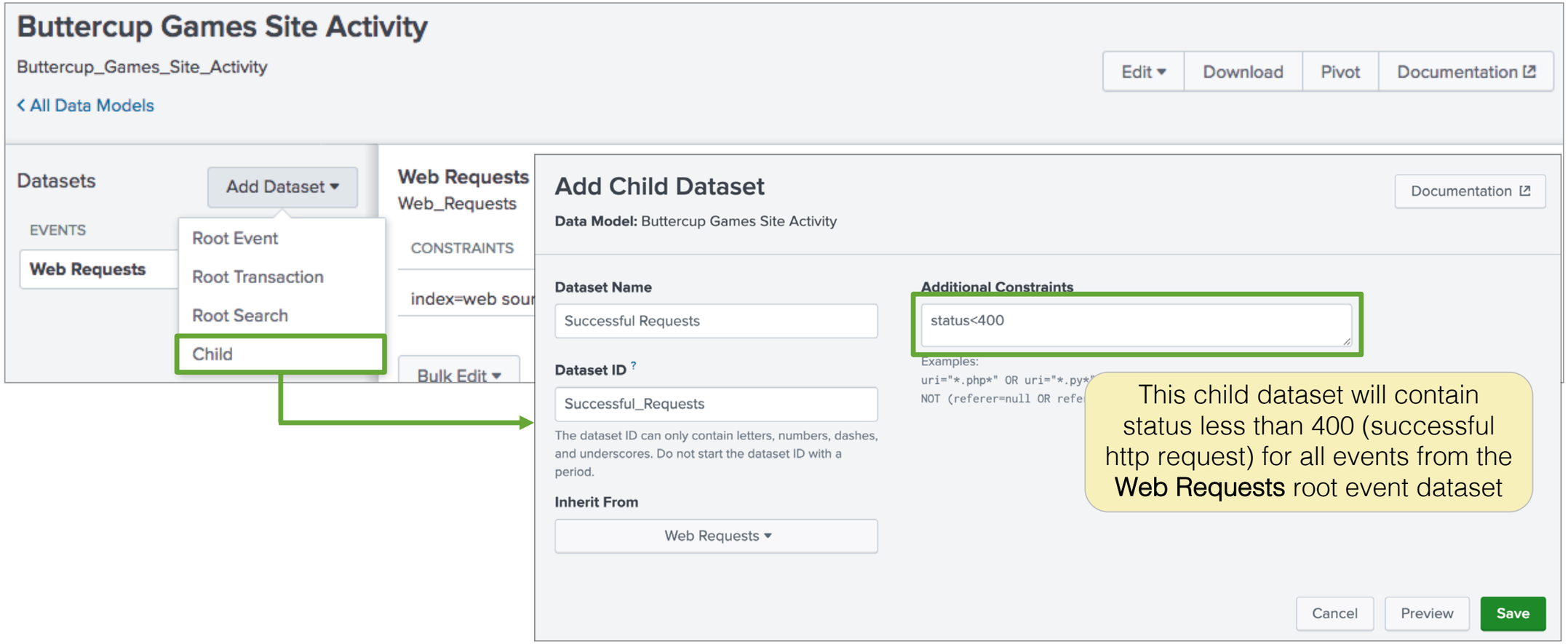
Add Fields: GeoIP

- Requires latitude/longitude fields
- Requires at least one IP field configured as an IPv4 field type
- Select the field that contains the mapping to `lat`/`lon`
- Identify the `lat`/`lon` and geo fields in the data



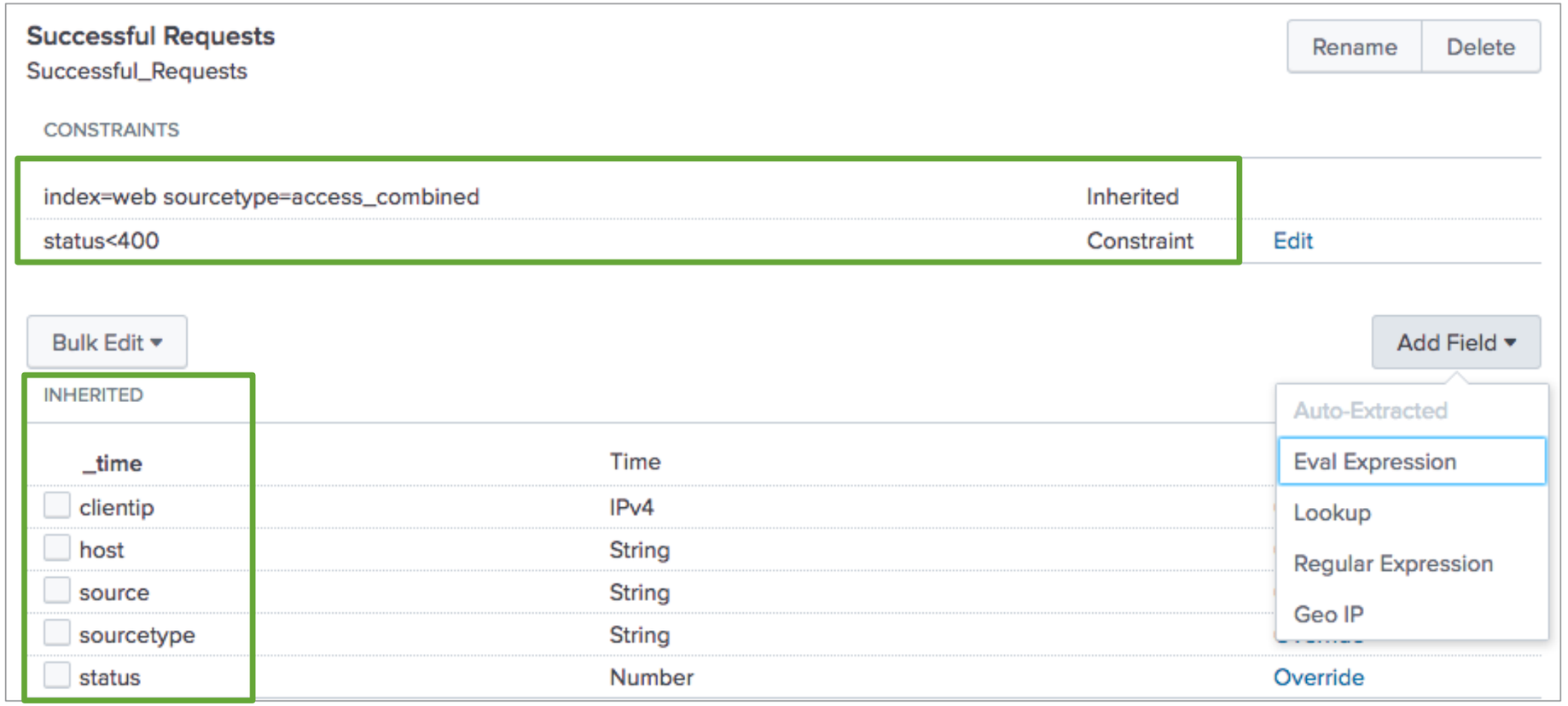
Add Child Datasets

New child datasets inherit the parent's constraints and **Additional Constraints** can be added

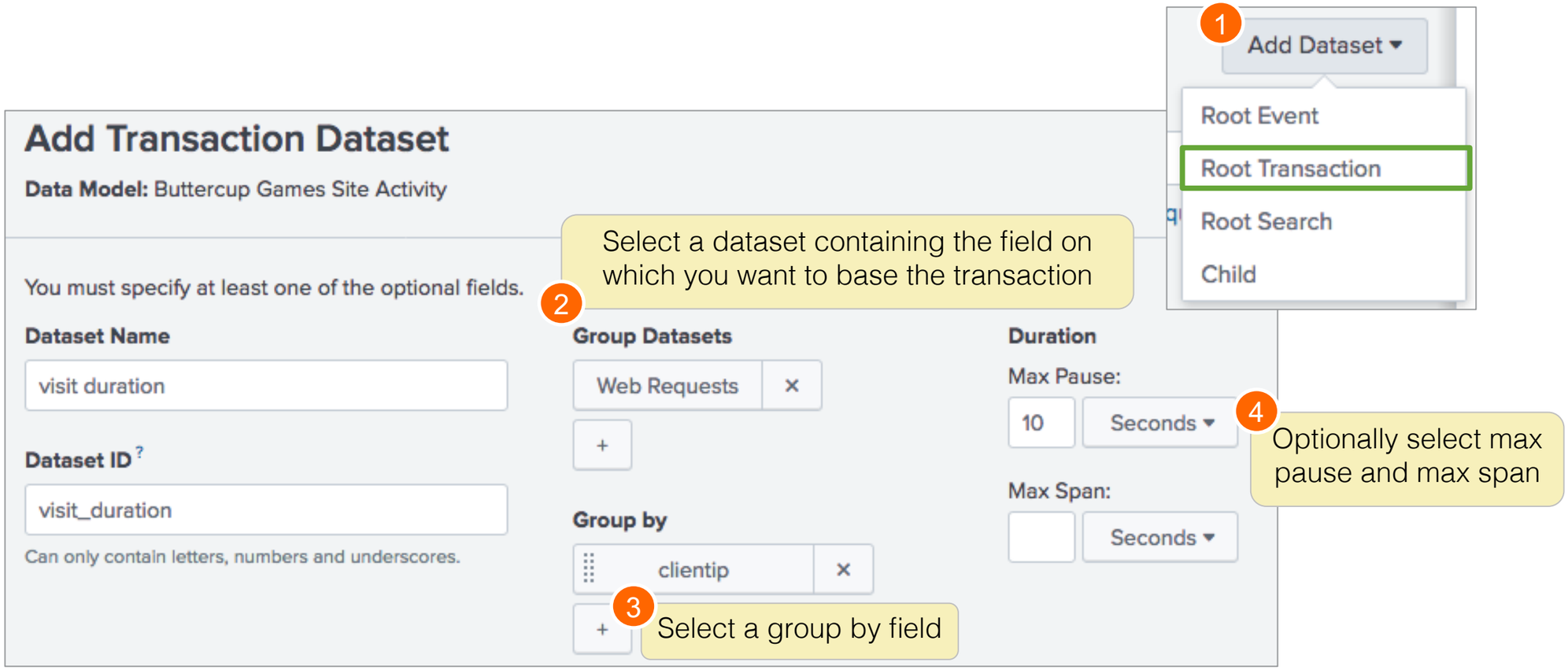


Add Child Datasets: Fields

- Child datasets inherit all fields from the parent events
- Inherited fields can be overridden and more fields can be added



Add a Transaction Dataset

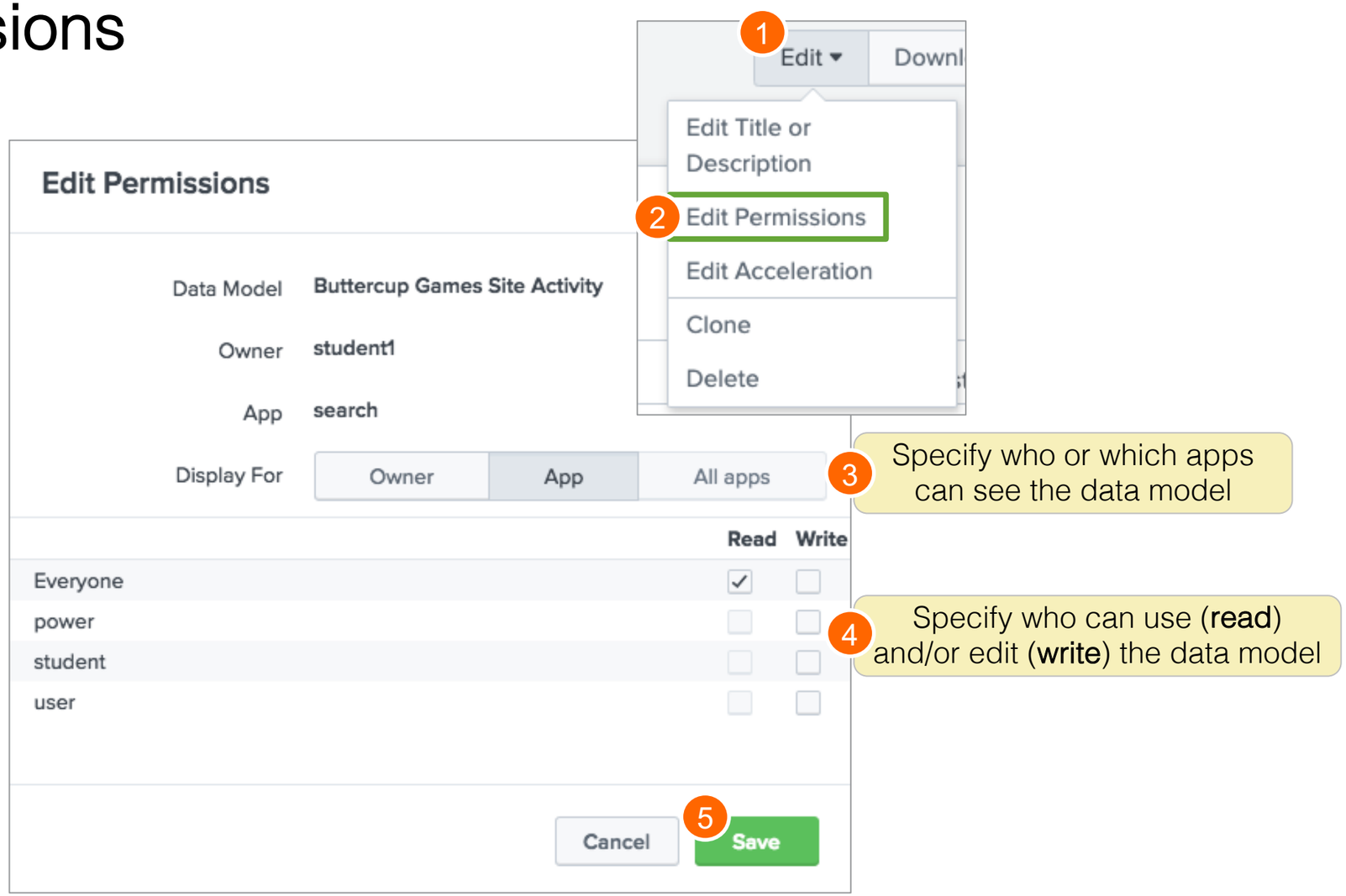


Transaction Dataset Considerations

- An event or search dataset must exist before adding a transaction dataset
- Auto-extracted fields must be added to the root event or root search dataset to be available in the transaction dataset
- Transaction datasets cannot benefit from persistent data model acceleration

Set Permissions

Edit > Edit Permissions

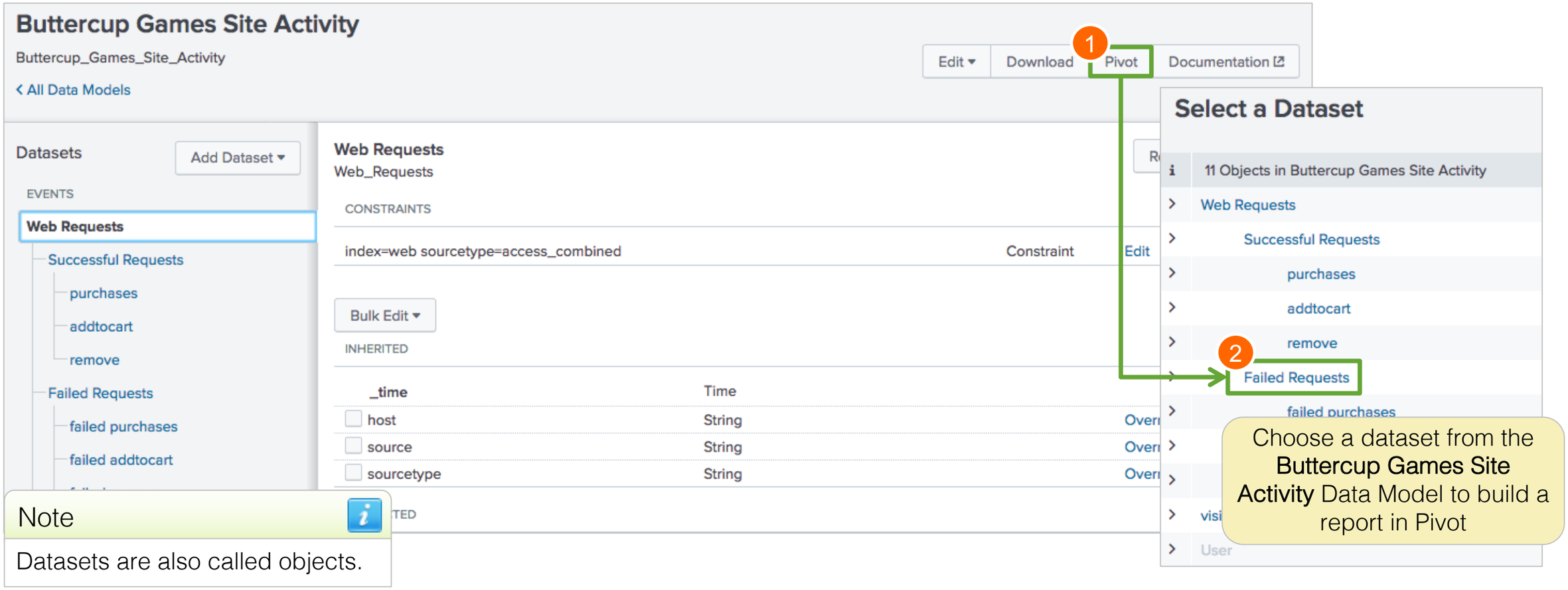


Note

Once the data model is displayed for a particular app, the admin can make changes to the data model, regardless of the permissions set by the owner.

Test the Data Model

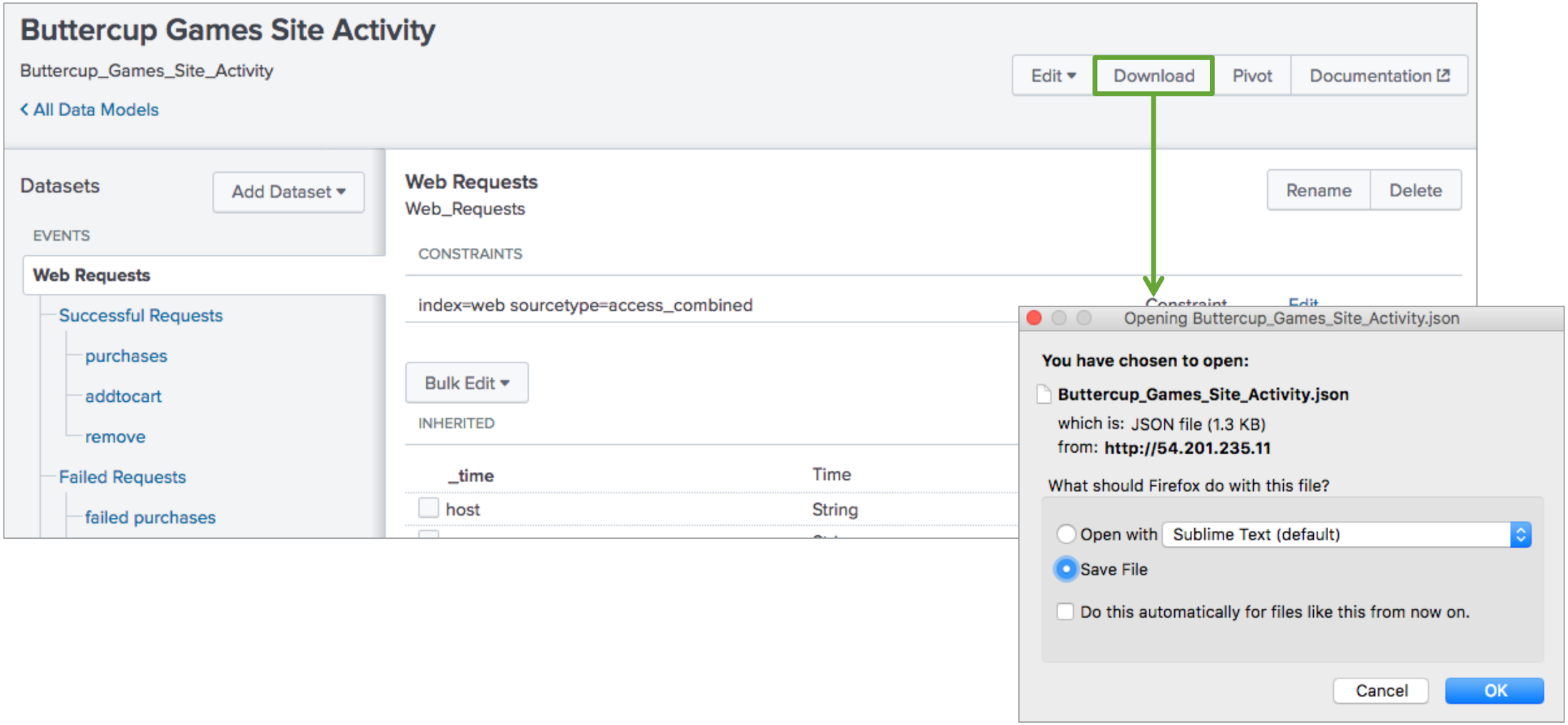
After completing your data model, it is recommended to test your datasets by building a pivot (next topic)



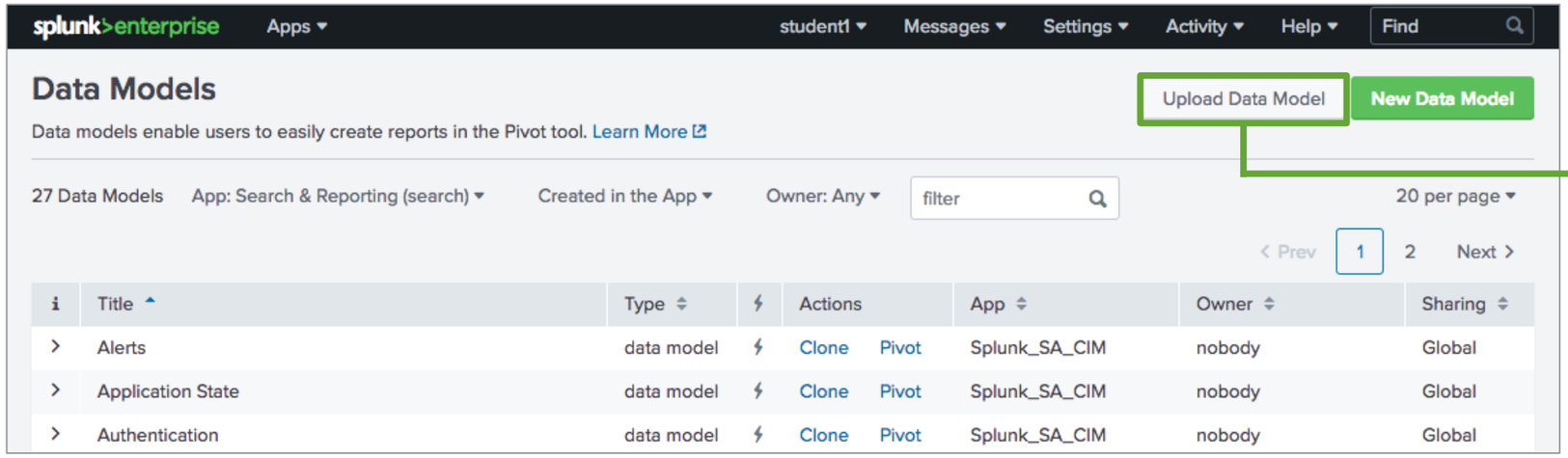
Download and Upload Data Models

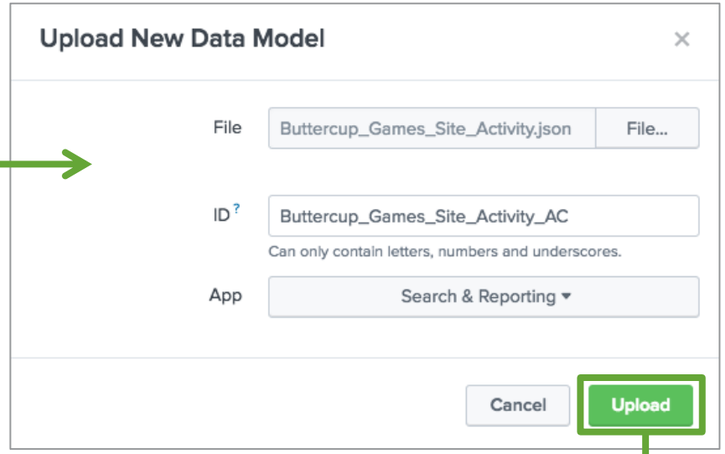
- The Splunk Web interface allows for data models to be downloaded or uploaded
- Benefits of downloading/uploading:
  - Back up important data models
  - Collaborate with other Splunk users to create/modify/test data models
  - Move data models from a test environment to production instance

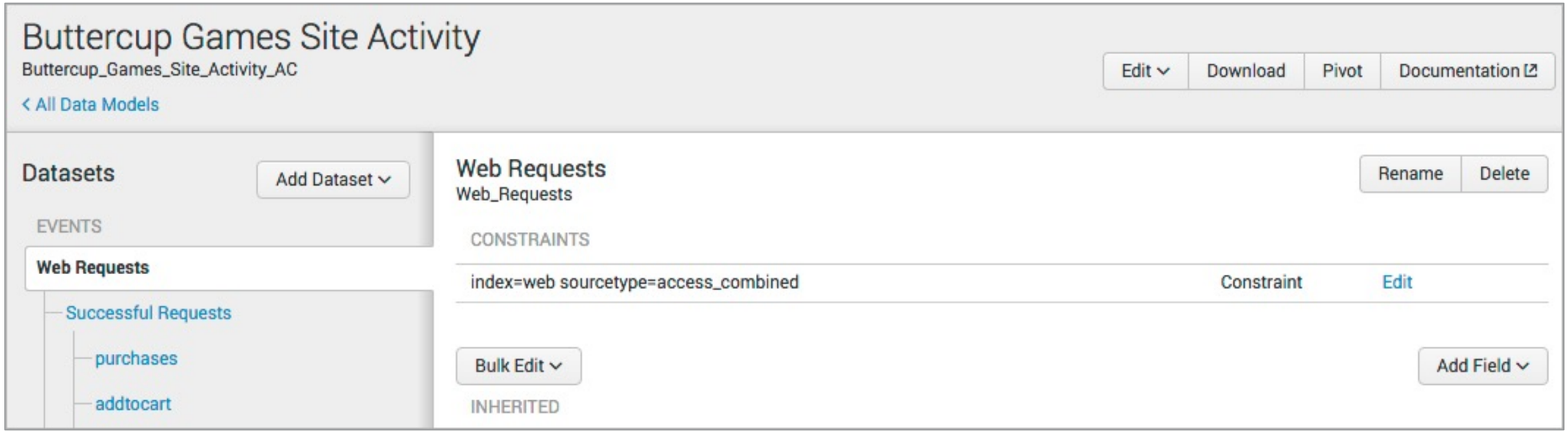
Download a Data Model



Upload a Data Model







Design Data Models Lab Exercise

Time: 10 minutes

Tasks:

- Create a data model and add a root event dataset
- Add auto-extracted fields to the root event dataset and modify their display names
- Add child event datasets to the root event dataset

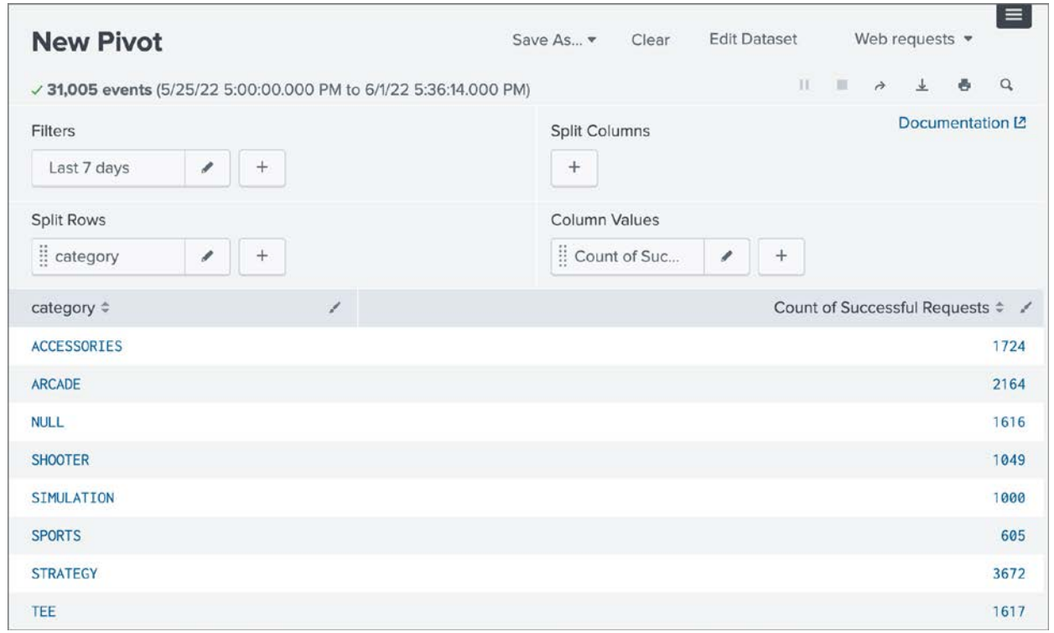
Create a Pivot

Topic Objectives

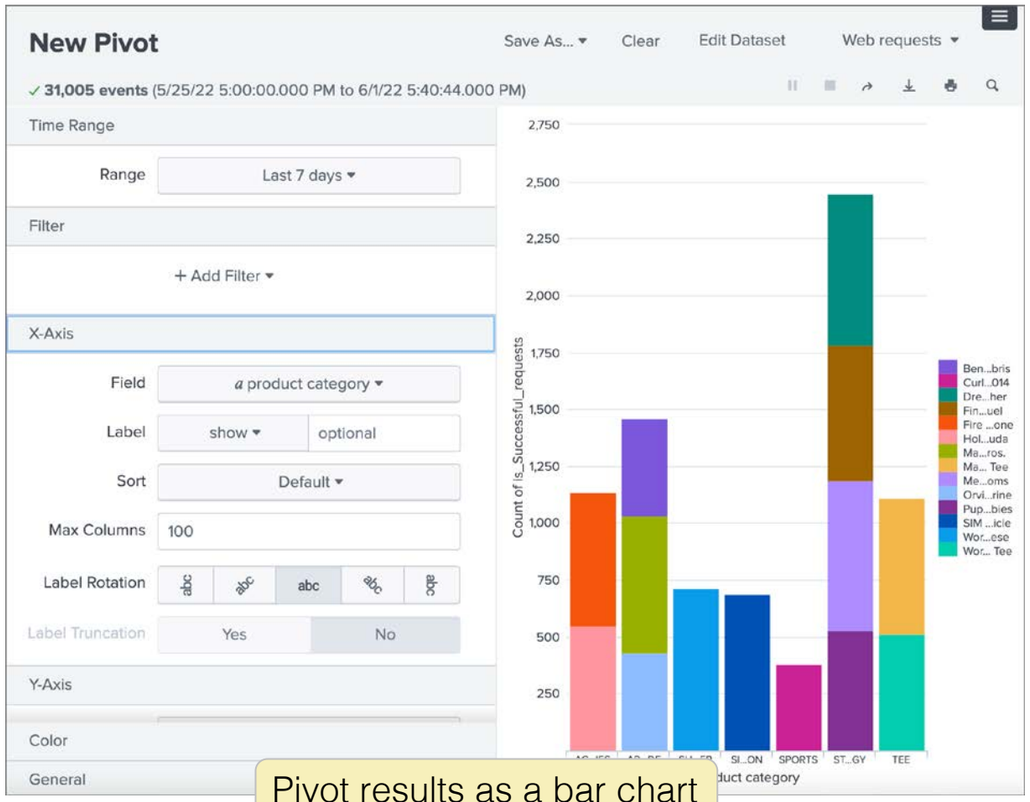
- Identify benefits of using Pivot
- Create and configure a Pivot
- Visualize a Pivot
- Save a Pivot
- Use Instant Pivot
- Access underlying search for Pivot

What is Pivot?

- Alternate method to access data **without** using search language
- Quick way to design visualizations of data using Splunk Web
- Requires use of datasets



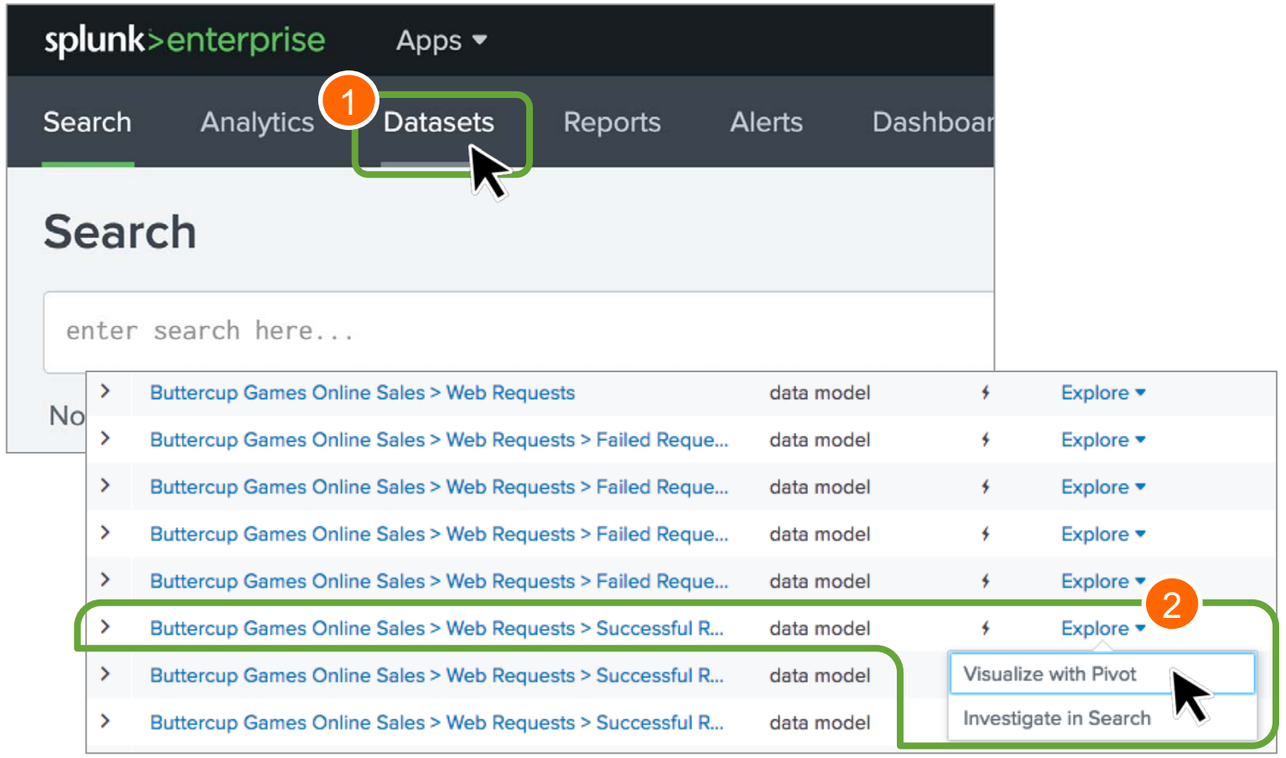
Pivot results as a table



Pivot results as a bar chart

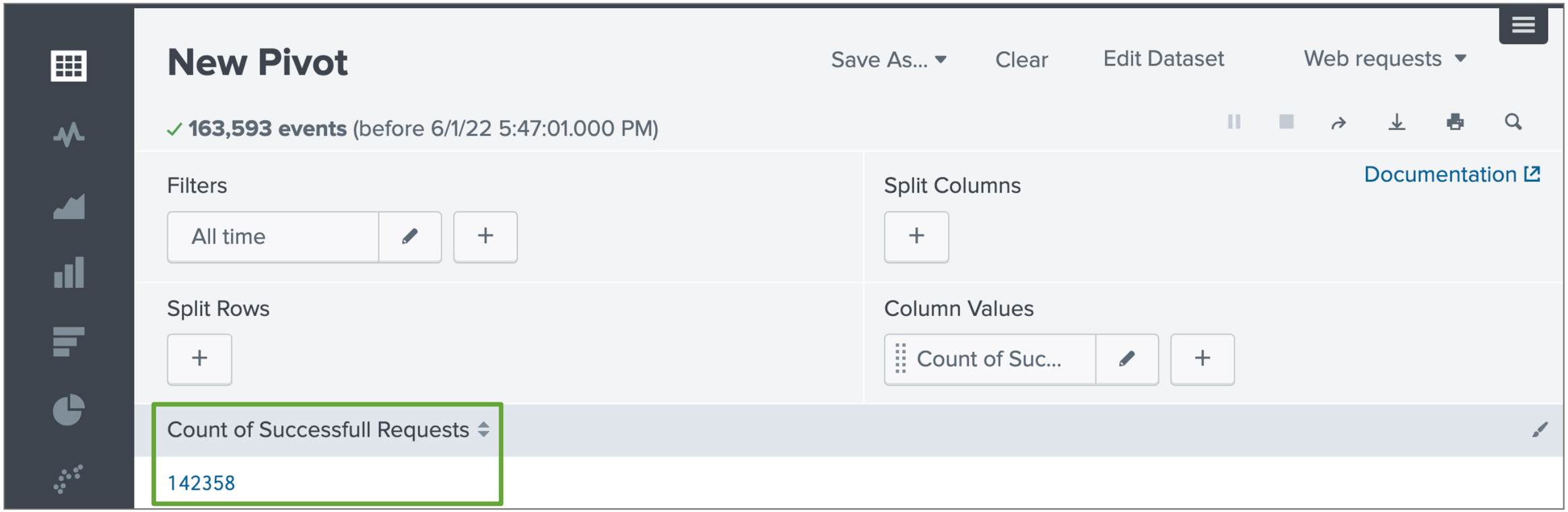
Select a Dataset for Pivot

1. Select **Datasets** tab from the app navigation bar
   - Datasets represent specific data from the data model
   - Data models are composed of multiple datasets
2. Choose a dataset, then click **Explore > Visualize with Pivot**



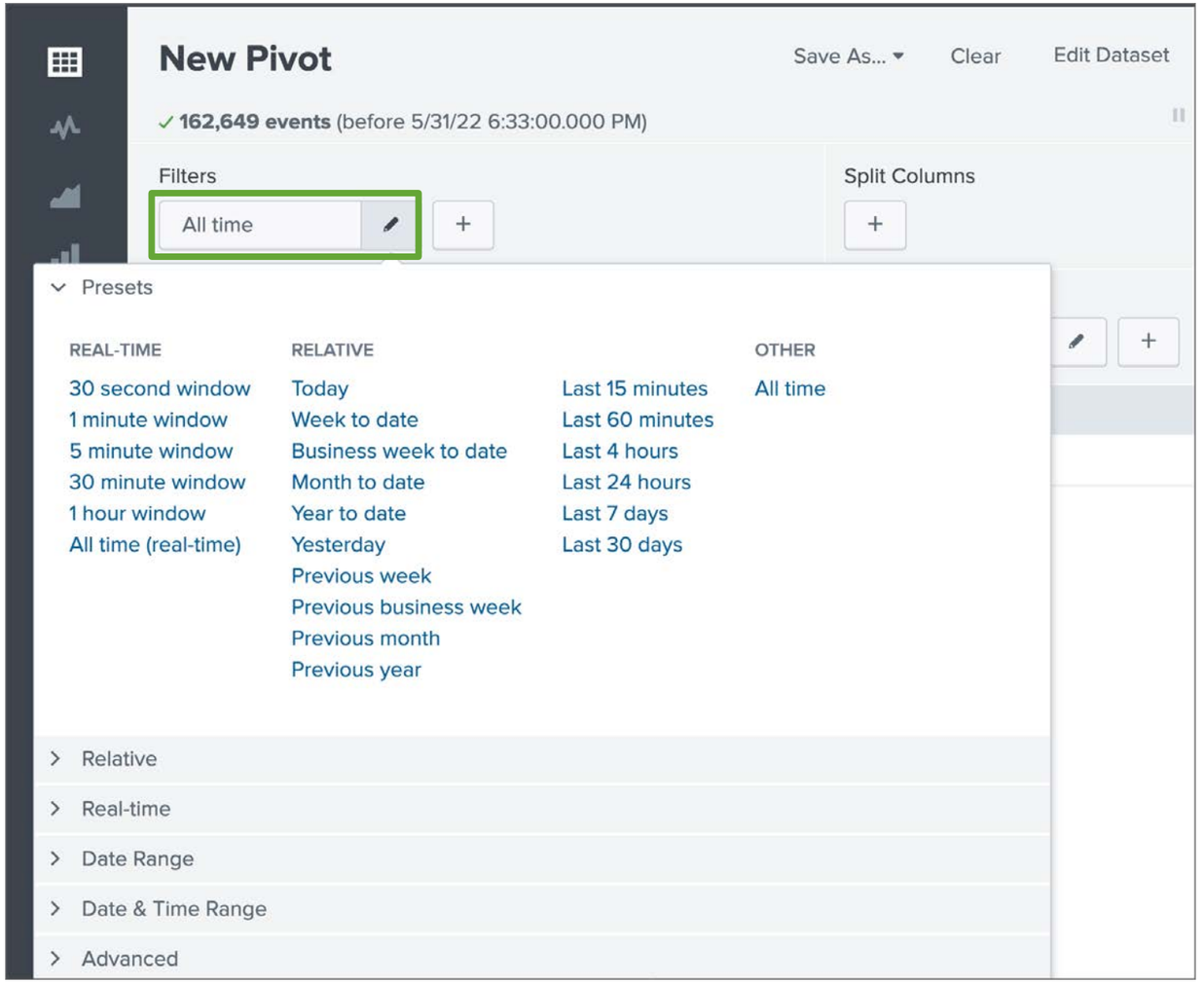
Open in Pivot

The Pivot automatically populates with a count of events for the selected dataset



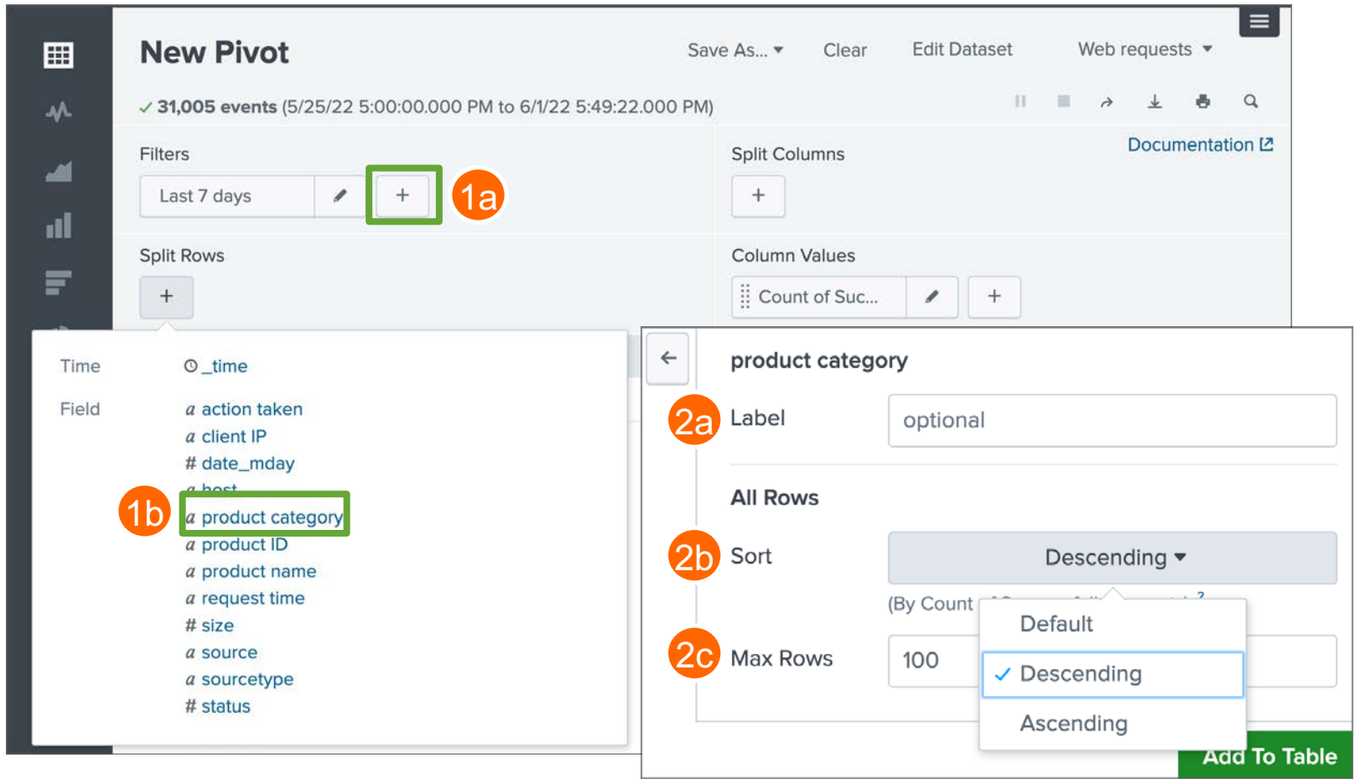
Select a Time Range

- Defaults to returning results over **All time**
- Click the  to select the desired time range
- The pivot runs immediately upon selecting the new time range

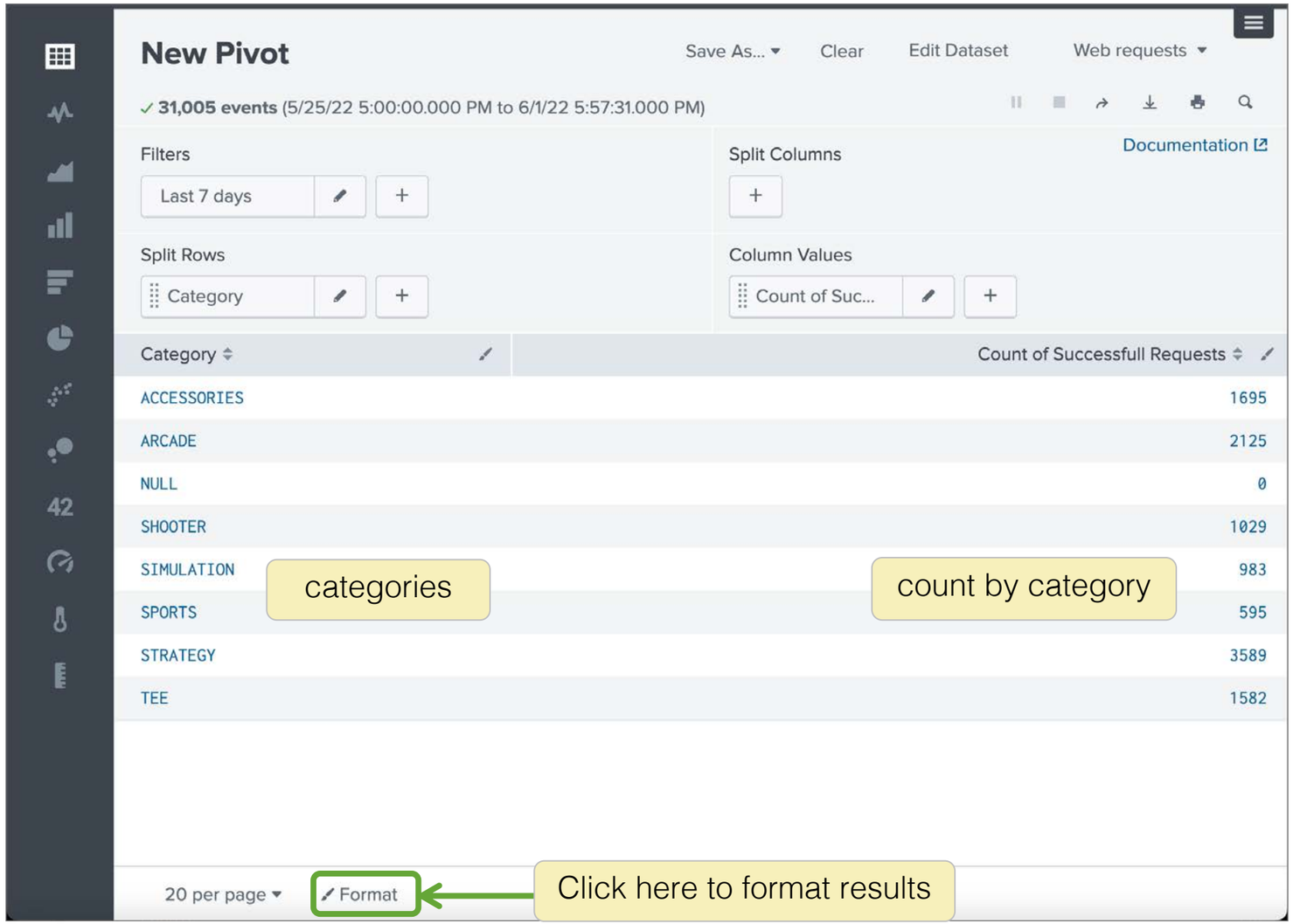


Split Rows

1. Click + under **Split Rows** for a list of available fields to split by
2. Choose a field
   - Enter a field **Label** (optional)
   - Choose a **Sort** order
   - Define maximum number of rows to display

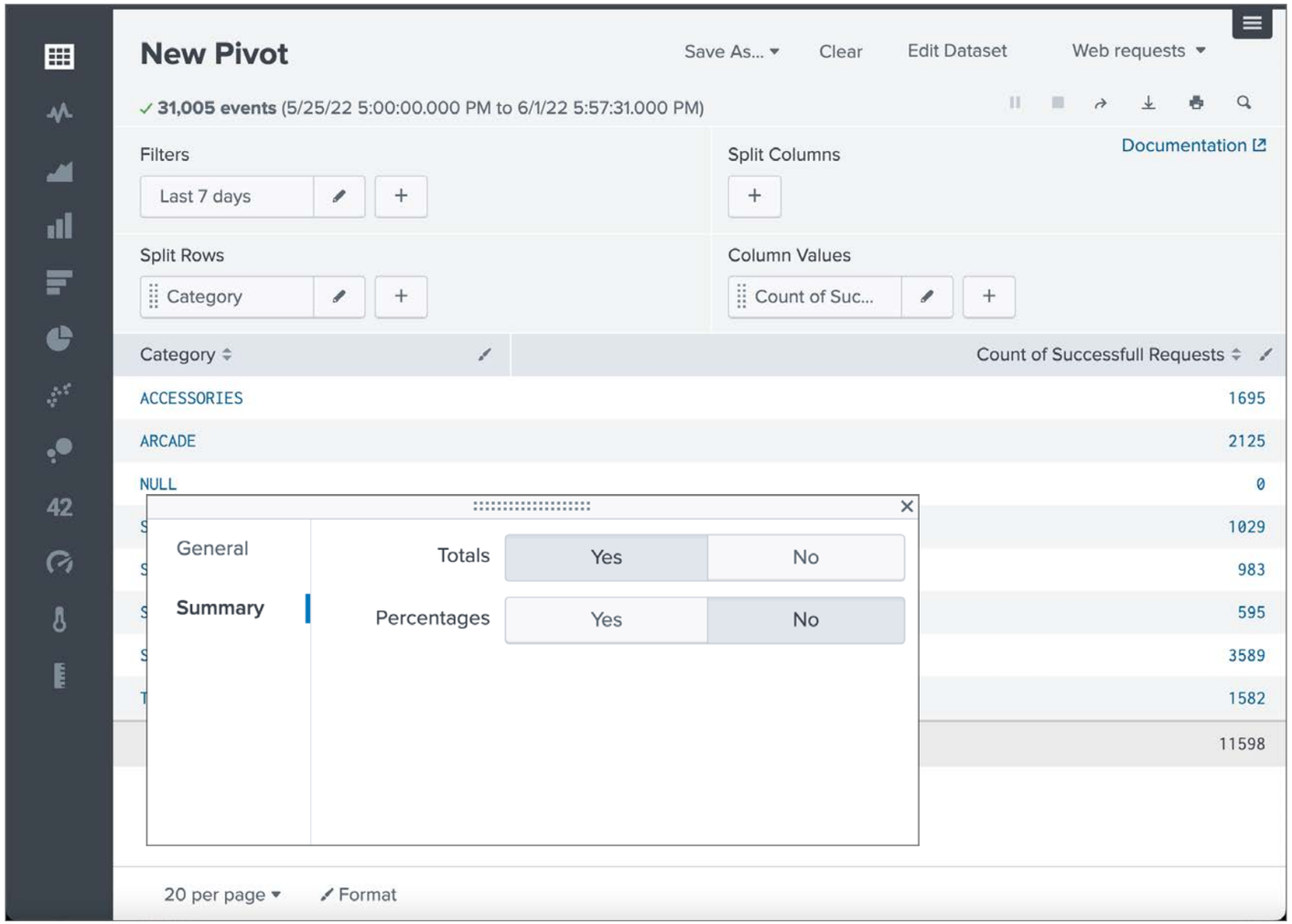


Split Rows: Results

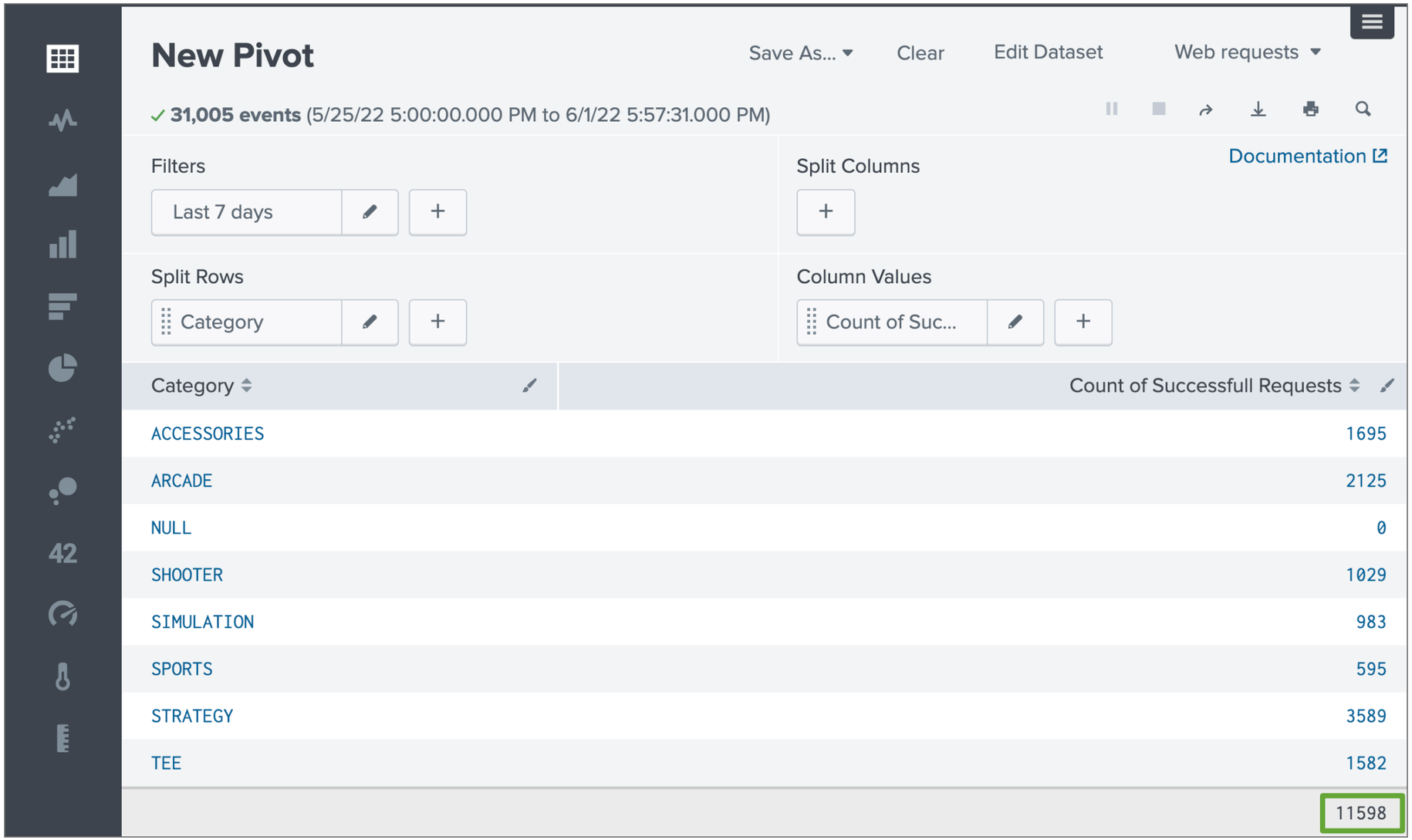


Format the Results

- Various formatting options are available

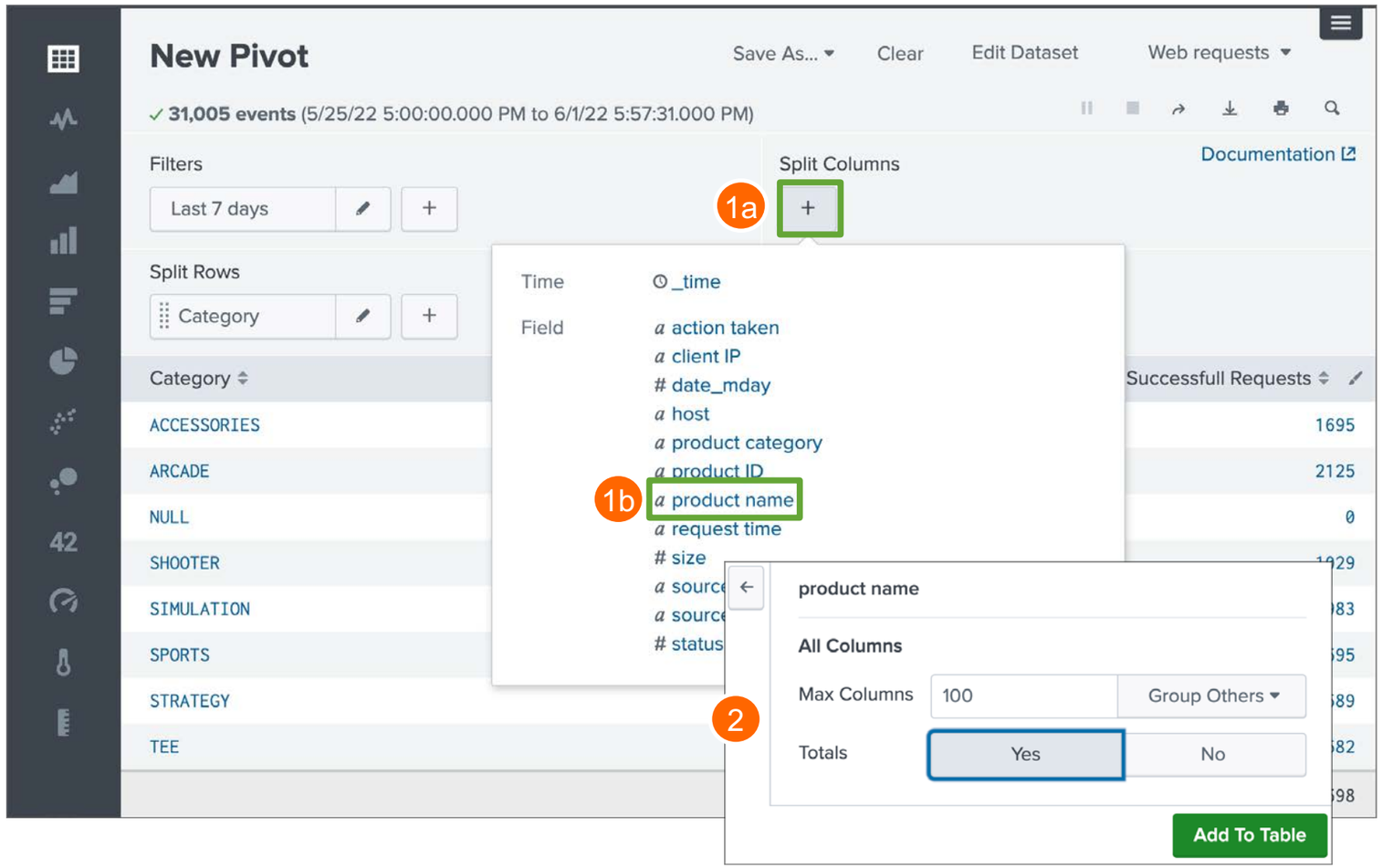


Formatted Results

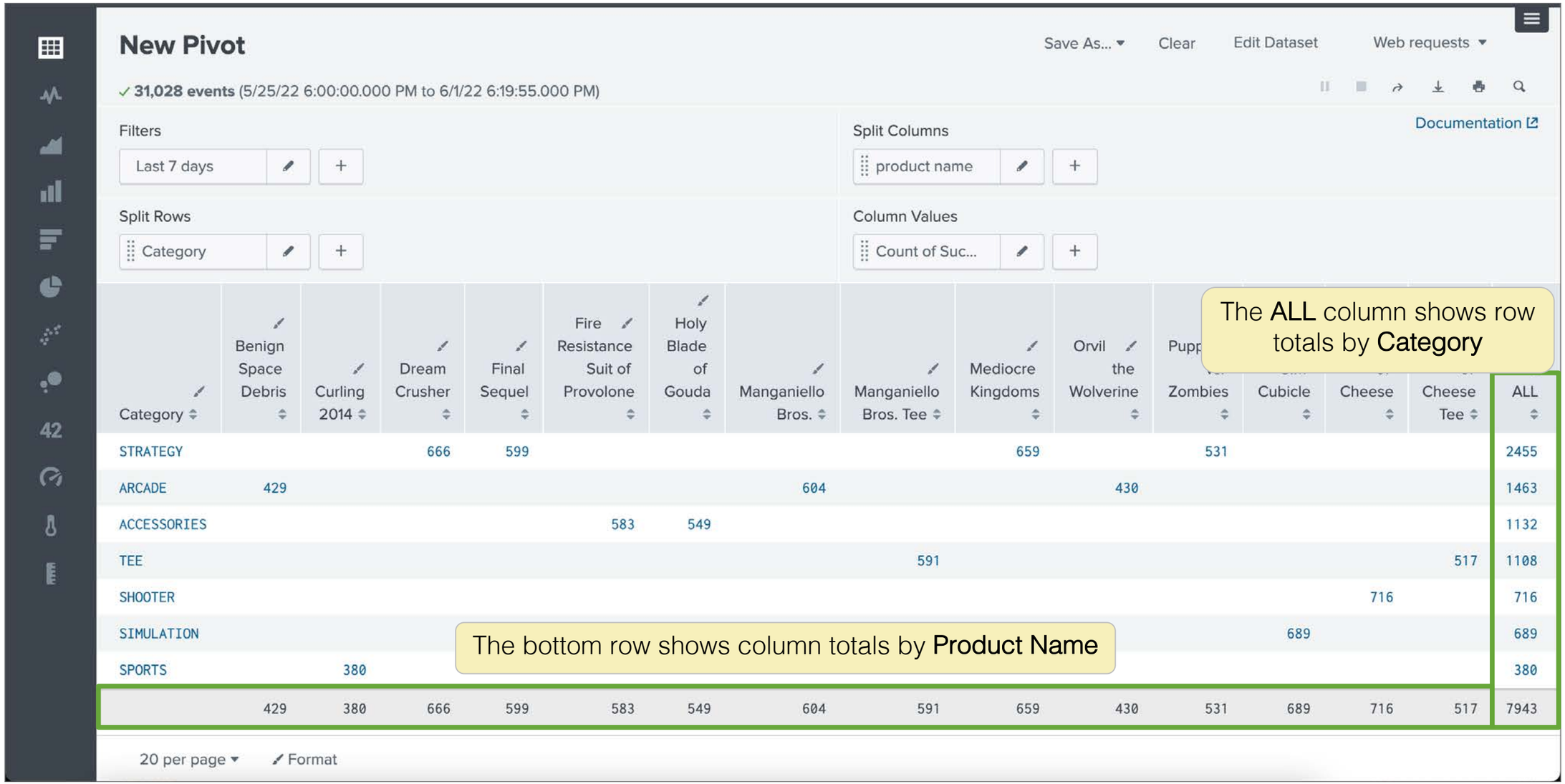


Split Columns

1. Click + under **Split Columns** and select a field to split by
2. Select the desired formatting options



Split Columns: Results



Category	Benign Space Debris	Curling 2014	Dream Crusher	Final Sequel	Fire Resistance Suit of Provolone	Holy Blade of Gouda	Manganiello Bros.	Manganiello Bros. Tee	Mediocre Kingdoms	Orvil the Wolverine	Pupp... Zombies	Cubicle	Cheese	Cheese Tee	ALL
STRATEGY			666	599					659		531				2455
ARCADE	429						604			430					1463
ACCESSORIES					583	549									1132
TEE								591						517	1108
SHOOTER													716		716
SIMULATION												689			689
SPORTS		380													380
	429	380	666	599	583	549	604	591	659	430	531	689	716	517	7943

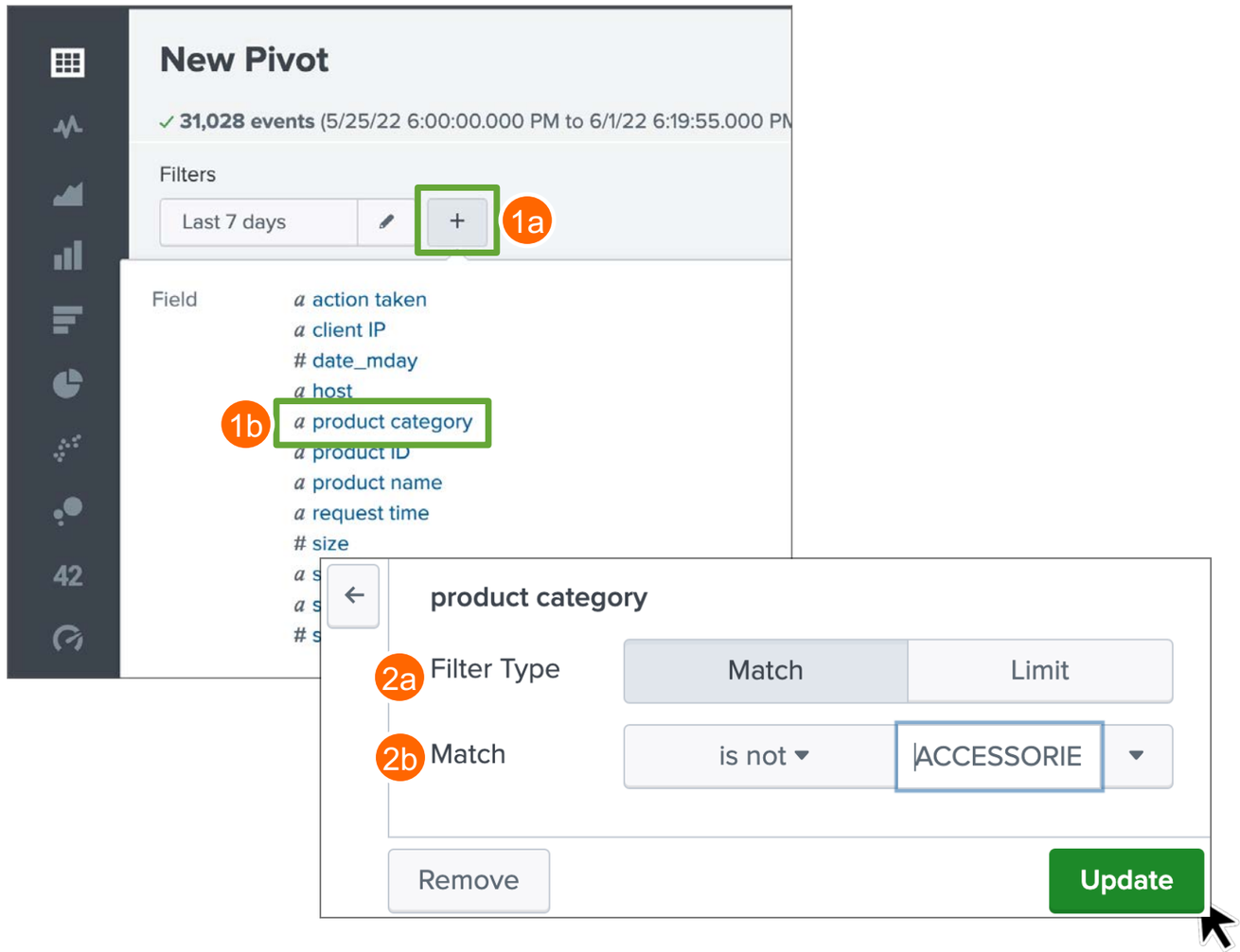
Add Additional Filters

- Refine a pivot by filtering on field-value pairs

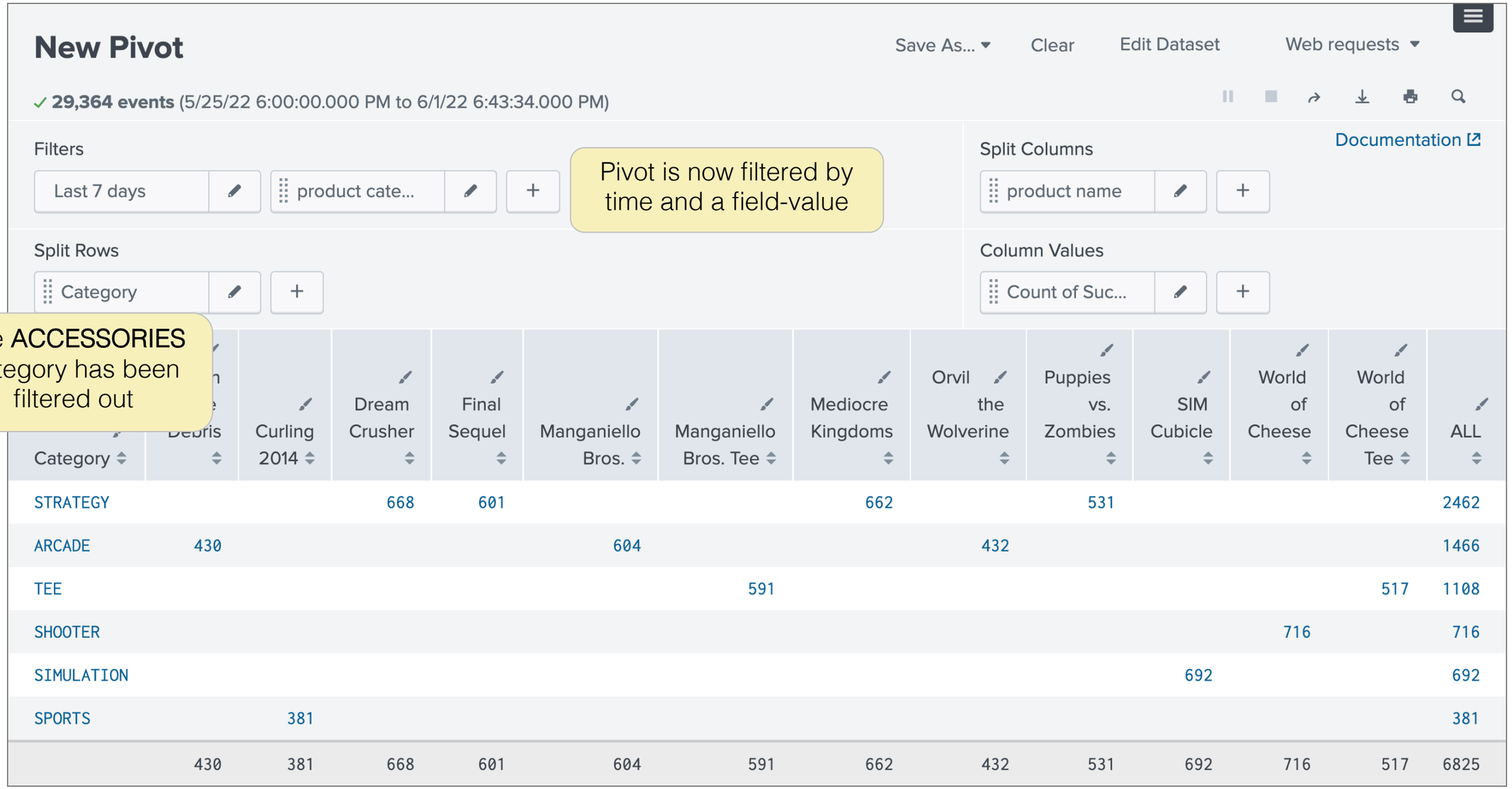
1. Click + under **Filters** and choose a **Field**

2a. Set the **Filter Type**

2b. Configure **Match** filter and value constraints



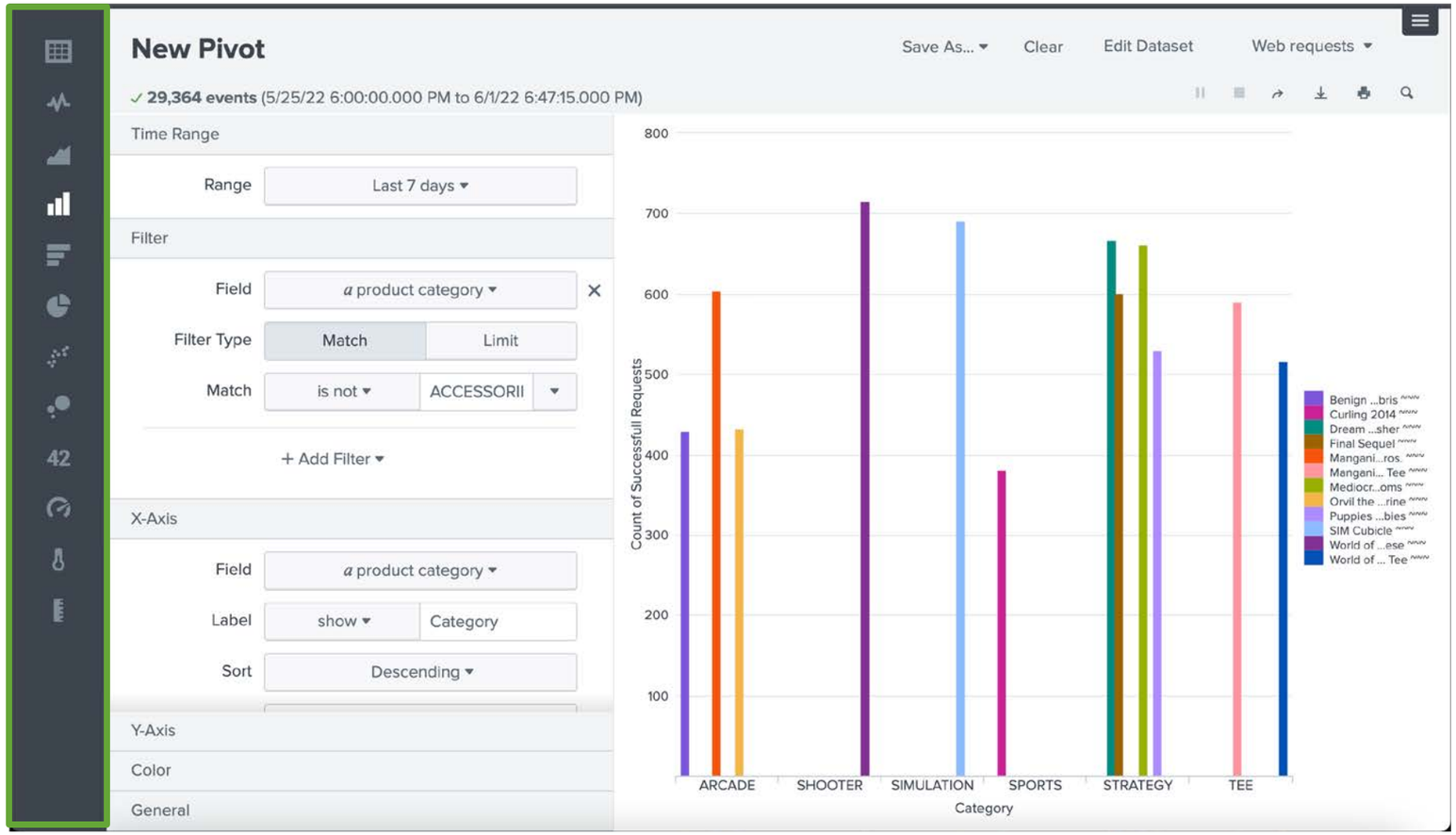
Filtered Pivot: Results



Category	Debris	Curling 2014	Dream Crusher	Final Sequel	Manganiello Bros.	Manganiello Bros. Tee	Mediocre Kingdoms	Orvil the Wolverine	Puppies vs. Zombies	SIM Cubicle	World of Cheese	World of Cheese Tee	ALL
STRATEGY			668	601			662		531				2462
ARCADE	430				604			432					1466
TEE						591						517	1108
SHOOTER											716		716
SIMULATION										692			692
SPORTS		381											381
	430	381	668	601	604	591	662	432	531	692	716	517	6825

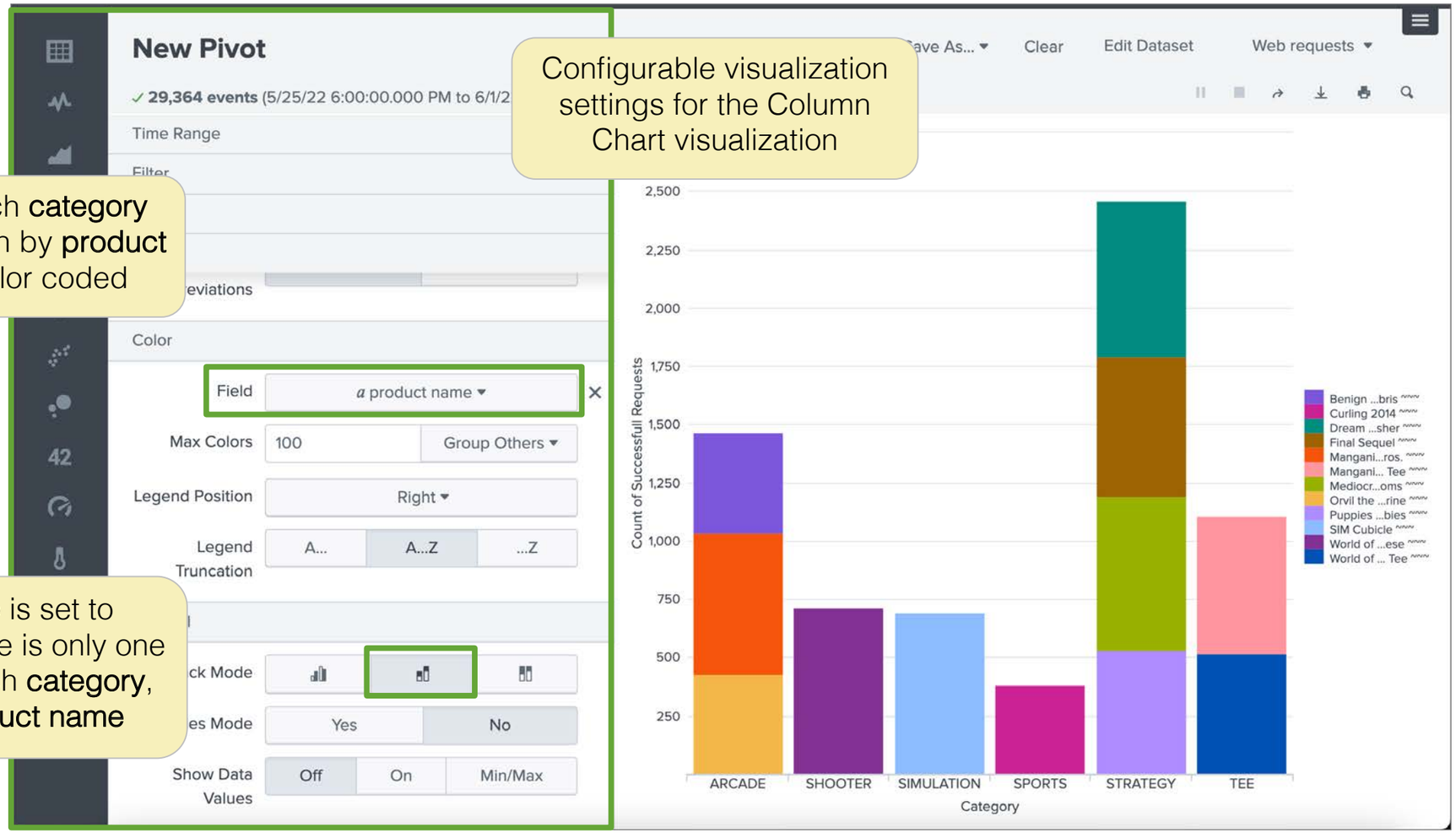
Pivot Visualizations

The Pivot Document Action Bar provides various options for visualizing Pivot results: Line Chart, Area Chart, Column Chart, etc.



Modify Visualization Settings

Each visualization offers different, configurable settings



Configurable visualization settings for the Column Chart visualization

Results for each **category** are broken down by **product name** and color coded

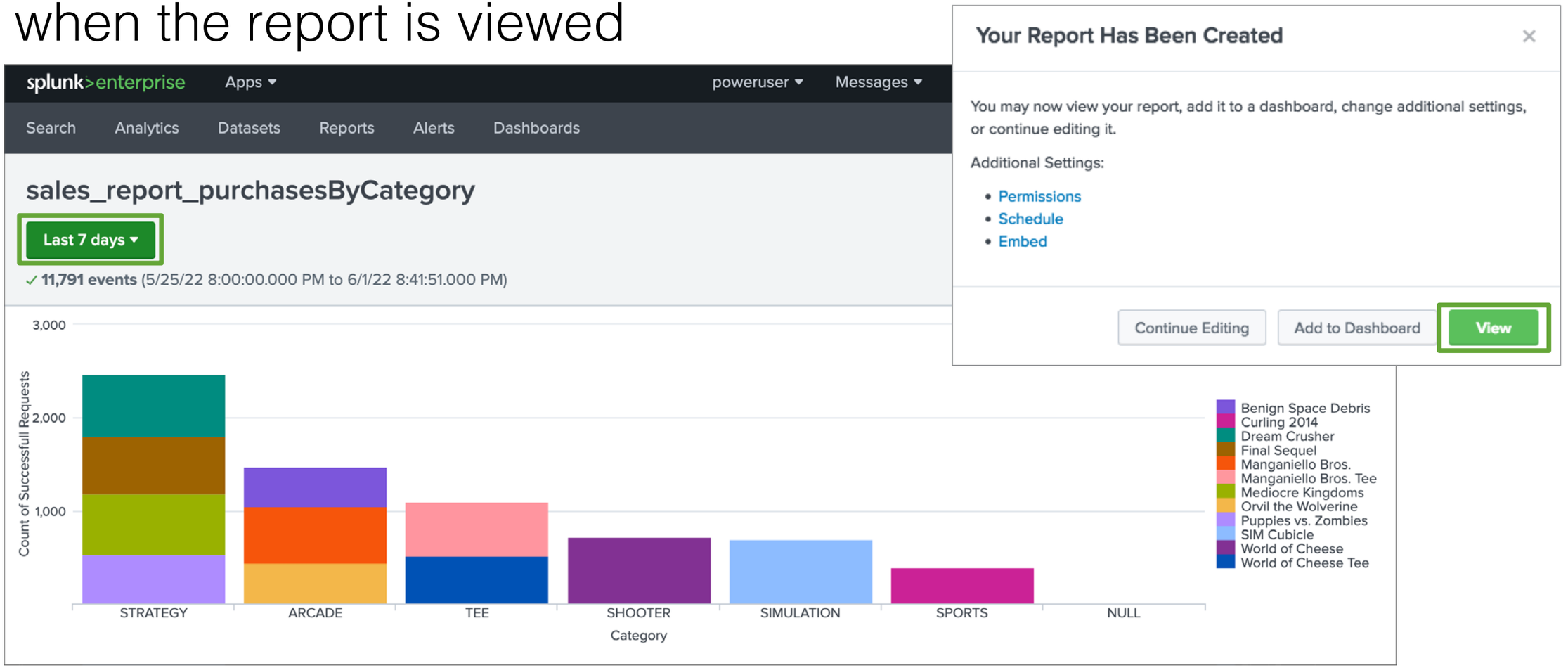
Stack Mode is set to **stacked** so there is only one column for each **category**, split by **product name**

Save a Pivot

Pivots can be saved as reports or as part of a dashboard panel

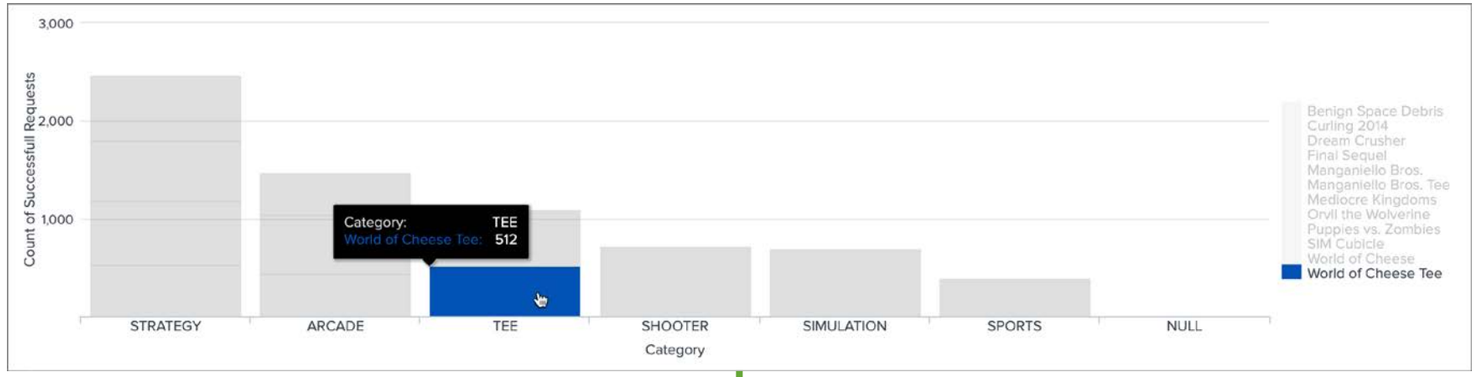
View a Pivot Report

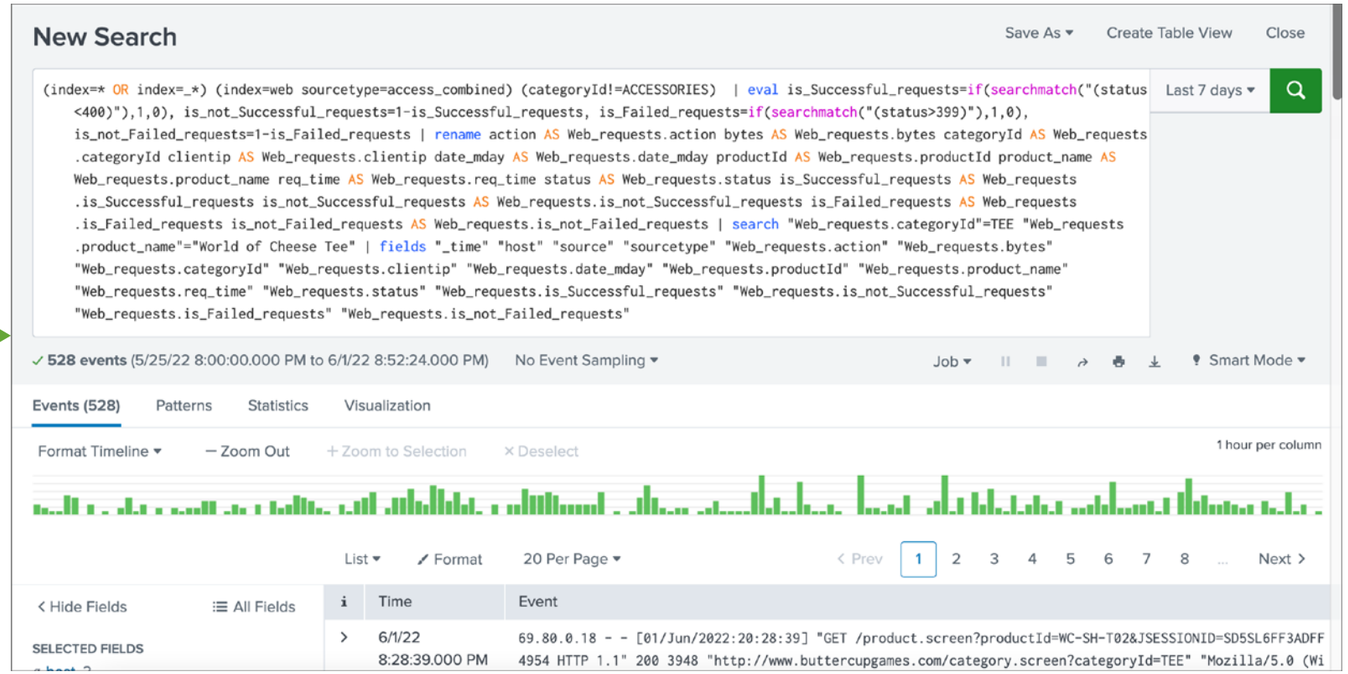
If a Time Range Picker was added to the saved report, it displays when the report is viewed



Interact with a Pivot Report

- Mouse over an object to reveal its details
- When you click on part of the report, a drilldown occurs that shows the underlying search details





Note

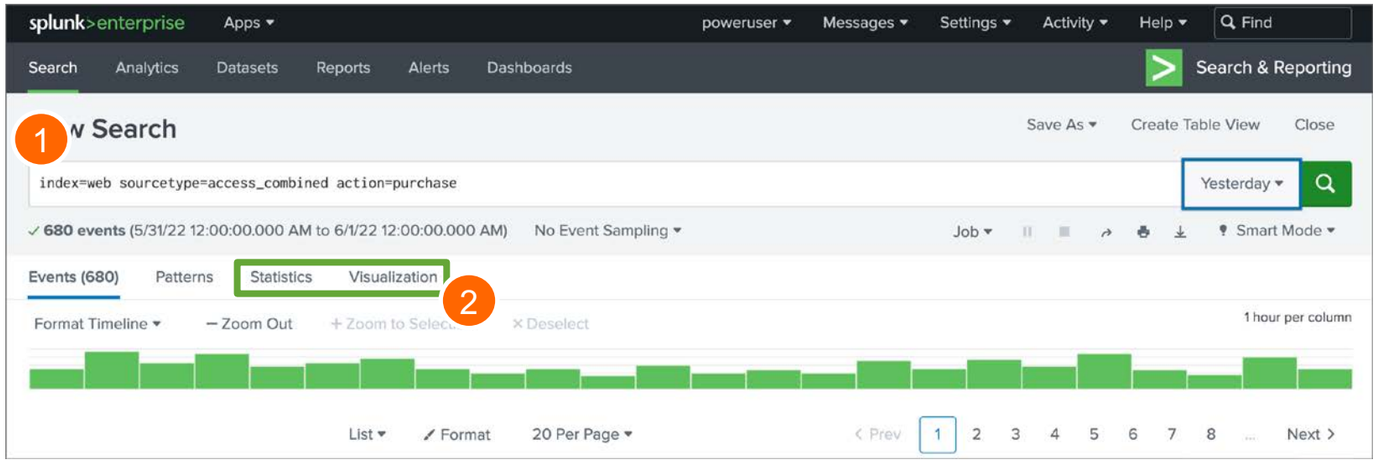
The search generated by drilldown may be more detailed than your original search. However, it produces the same results.

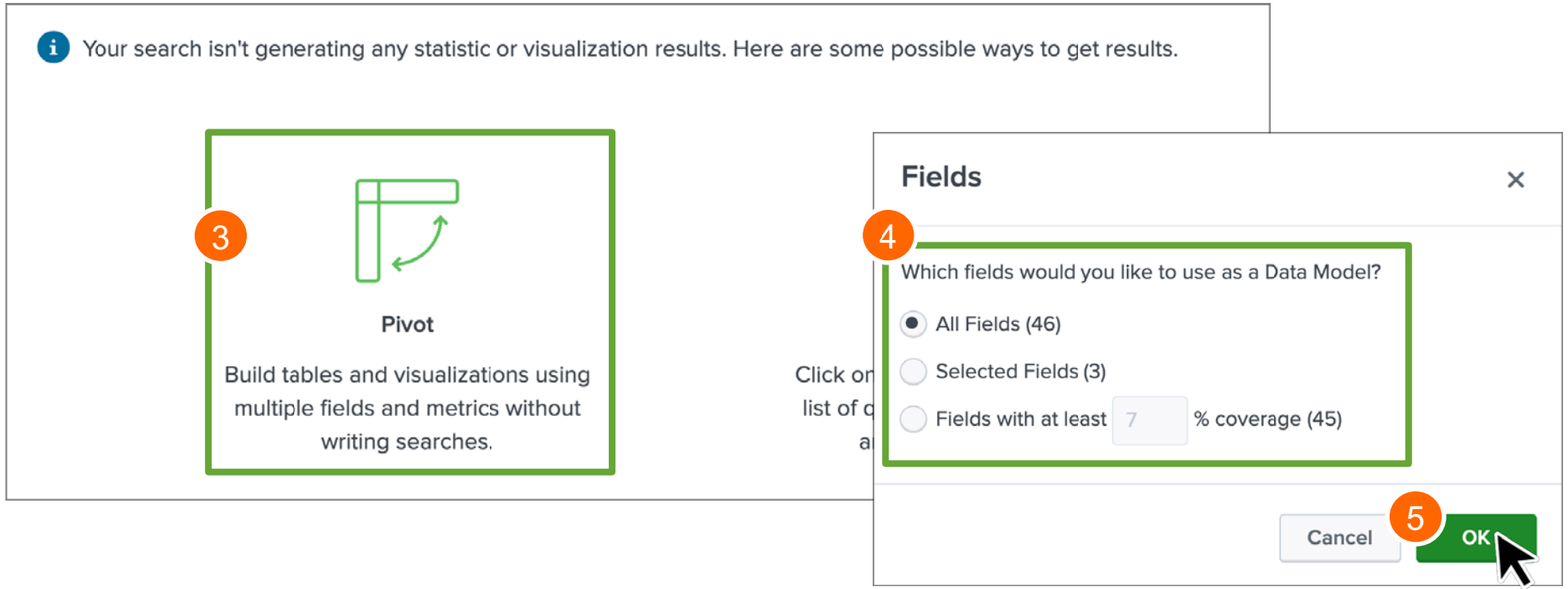
Instant Pivot Overview

- Instant Pivot allows you to utilize the Pivot tool without a preexisting data model
- Instant Pivot creates an underlying data model utilizing the search criteria entered during the initial search

Access Instant Pivot

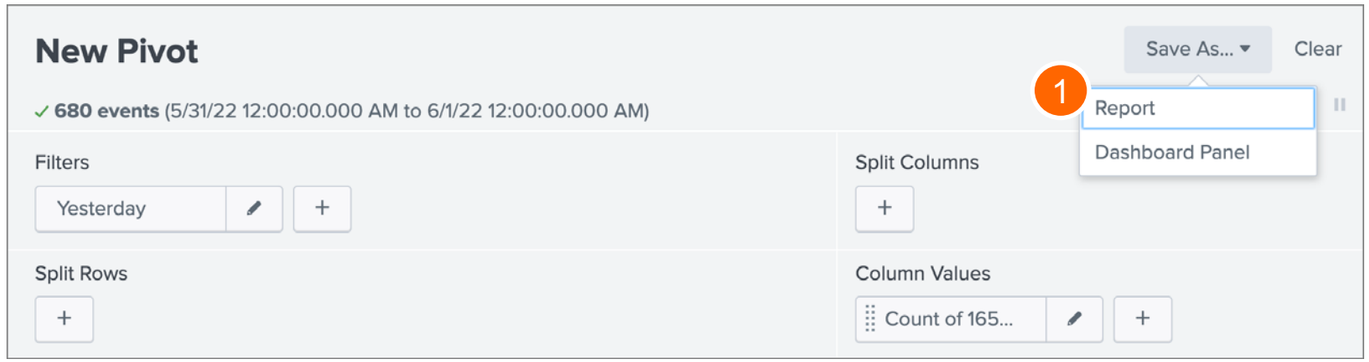
1. Execute a non-transforming search
2. Choose **Statistics** or **Visualization** tab
3. Click **Pivot** icon
4. Select the fields to be included in the data model object
5. Click OK to create the Pivot

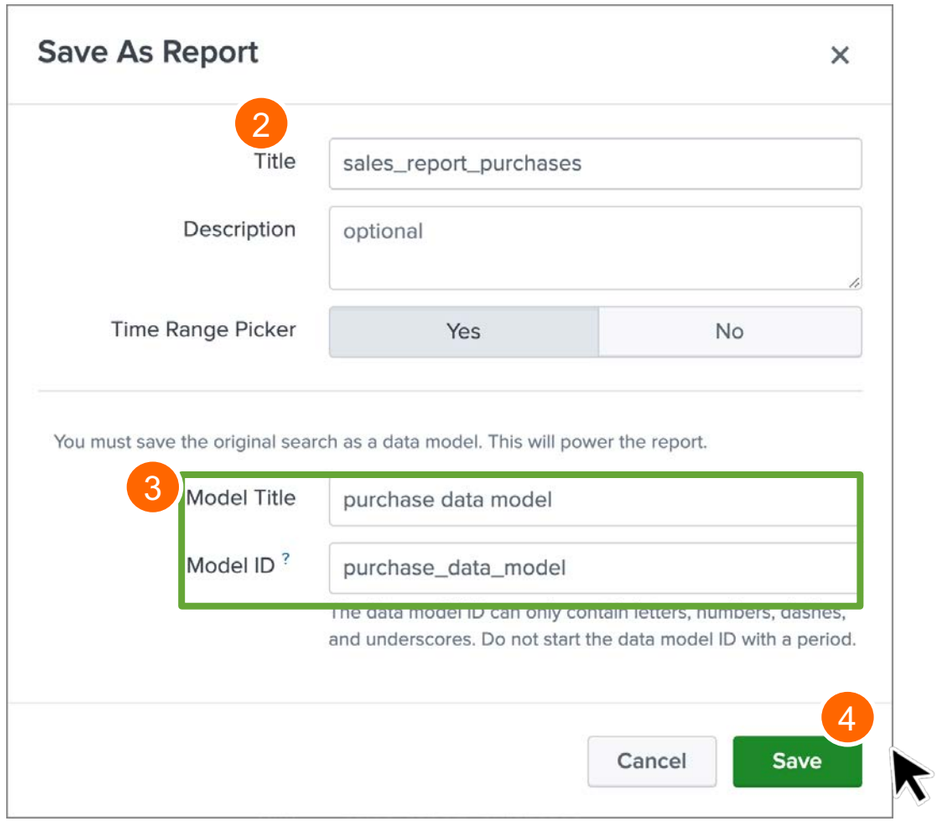




Save an Instant Pivot as a Report

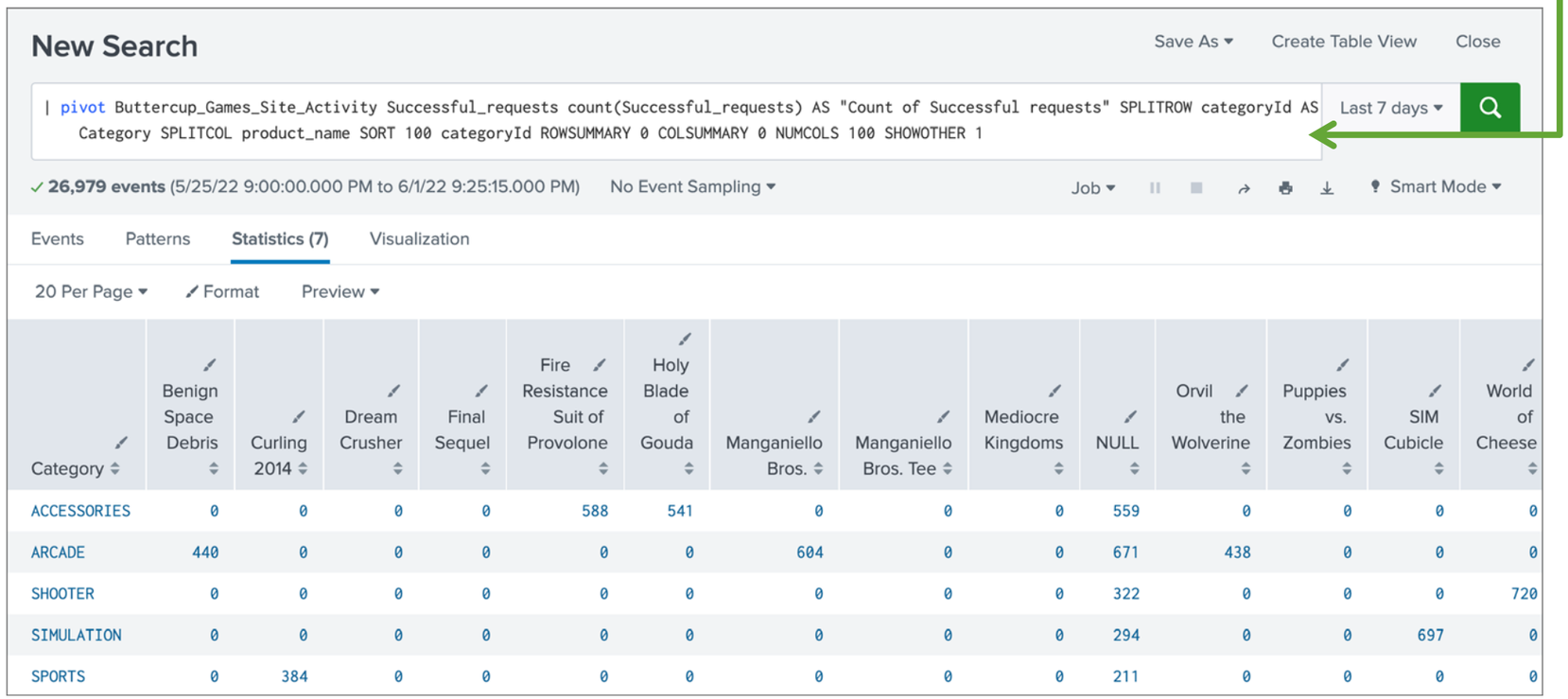
- A **Model Title** is required when saving an Instant Pivot as a report
- The **Model ID** is automatically generated based on the **Model Title**





View Underlying Search

View the underlying search for an active Pivot by clicking 



Category	Benign Space Debris	Curling 2014	Dream Crusher	Final Sequel	Fire Resistance Suit of Provolone	Holy Blade of Gouda	Manganiello Bros.	Manganiello Bros. Tee	Mediocre Kingdoms	NULL	Orvil the Wolverine	Puppies vs. Zombies	SIM Cubicle	World of Cheese
ACCESSORIES	0	0	0	0	588	541	0	0	0	559	0	0	0	0
ARCADE	440	0	0	0	0	0	604	0	0	671	438	0	0	0
SHOOTER	0	0	0	0	0	0	0	0	0	322	0	0	0	720
SIMULATION	0	0	0	0	0	0	0	0	0	294	0	0	697	0
SPORTS	0	384	0	0	0	0	0	0	0	211	0	0	0	0

View Underlying Search (cont.)

```
| pivot Buttercup_Games_Online_Sales successful_request count(successful_request) AS "Count of Successful Requests" SPLITROW categoryId AS category SPLITCOL
    product_name FILTER categoryId isNot ACCESSORIES TOP 100 count(successful_request) ROWSUMMARY 0 COLSUMMARY 0 NUMCOLS 100 SHOWOTHER 1
```

Data model name

```
| pivot Buttercup_Games_Online_Sales
```

Object (dataset) name

```
successful_request
count(successful_request) AS "Count of
Successful Requests"
```

Split row field (or attribute)

```
SPLITROW categoryId AS category
```

Split column field (or attribute) and filter field/value pair

```
SPLITCOL product_name FILTER categoryId
isNOT ACCESSORIES
```

Create a Pivot Lab Exercise

Time: 15 minutes

Tasks:

- Test your data model by creating a simple Pivot
- Add a field to your data model that uses an eval expression
- Add fields from a lookup to your data model
- Verify that new fields work by creating Pivot reports
- Create a visualization in the Pivot editor
- Add Pivot reports to a dashboard

Accelerate Data Models

Topic Objectives

- Define ad-hoc and persistent data model acceleration
- Accelerate a data model
  - Describe the role of `tsidx` files in persistent data model acceleration
  - Review considerations about data model acceleration

Data Model Acceleration

- Generates summaries to speed pivot and report completion times
- Takes the form of inverted time-series index (**`tsidx`**) files that have been optimized for speed
- Consists of two forms:
  - Ad hoc data model acceleration
  - Persistent data model acceleration

Ad Hoc Data Model Acceleration

- An ad hoc acceleration summary is built anytime a user accesses a data model dataset in Pivot
- Storing summaries in dispatch directories on the search head allows for acceleration of all 3 root dataset types: event, search, and transaction (and their children)

Ad Hoc Data Model Acceleration (cont.)

- Summaries are created for each user currently accessing a data model dataset in Pivot which increases search head load
- The initial acceleration summary is built over all time and as the report is fine-tuned in the Pivot editor, the performance improves
- Ad hoc acceleration is temporary and only exists for the duration of the Pivot session
  - Reports or dashboard panels made in a pivot session do not benefit from ad hoc acceleration when ran outside of the Pivot session

Persistent Data Model Acceleration

- Persistent data model acceleration builds dedicated summaries in indexes and exists as long as the data model exists
- Once accelerated, Splunk maintains the dedicated summaries
- Reports and dashboard panels generated from persistently accelerated data models complete more quickly
- Summaries can be used by Pivot, **`datamodel`**, and **`tstats`**
- Multiple users can access the summary at the same time

Persistent Data Model Acceleration (cont.)

- Admin permissions or the **`accelerate_datamodel`** capability is required to accelerate a data model
- Private data models can't be accelerated
- Accelerated data models can't be edited
- When accelerating a data model, only the following datasets are accelerated:
  - Root event datasets
  - Root search datasets that only include streaming commands
  - Child datasets of these qualifying root datasets

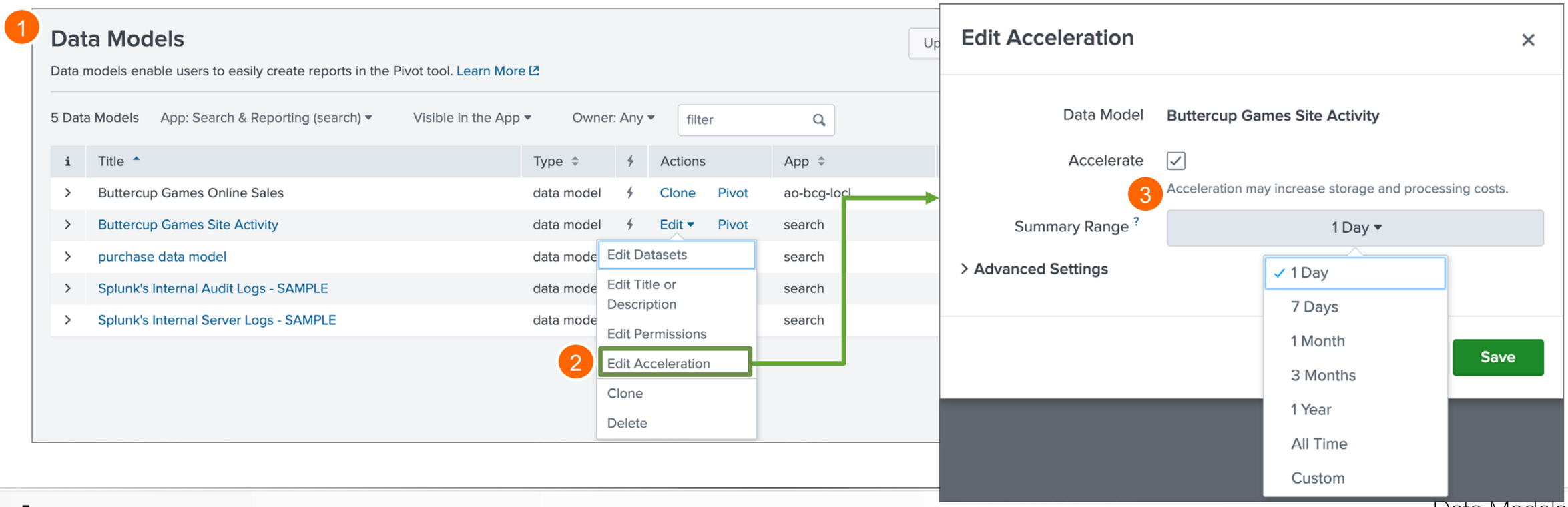
Compare Data Model Accelerations

Ad Hoc	Persistent
The acceleration is built every time the Pivot editor is accessed	Explicitly defined before using
Exists only for the duration of user's Pivot editor session	Exists as long as data model exists
Runs over all time (i.e., can't be scoped to specific time range)	Can be scoped to specific time ranges
Has no restrictions while in the Pivot editor	Has some restrictions
Reports run without any acceleration	Reports run faster and perform better overall

The rest of this topic discusses **persistent** data model acceleration only

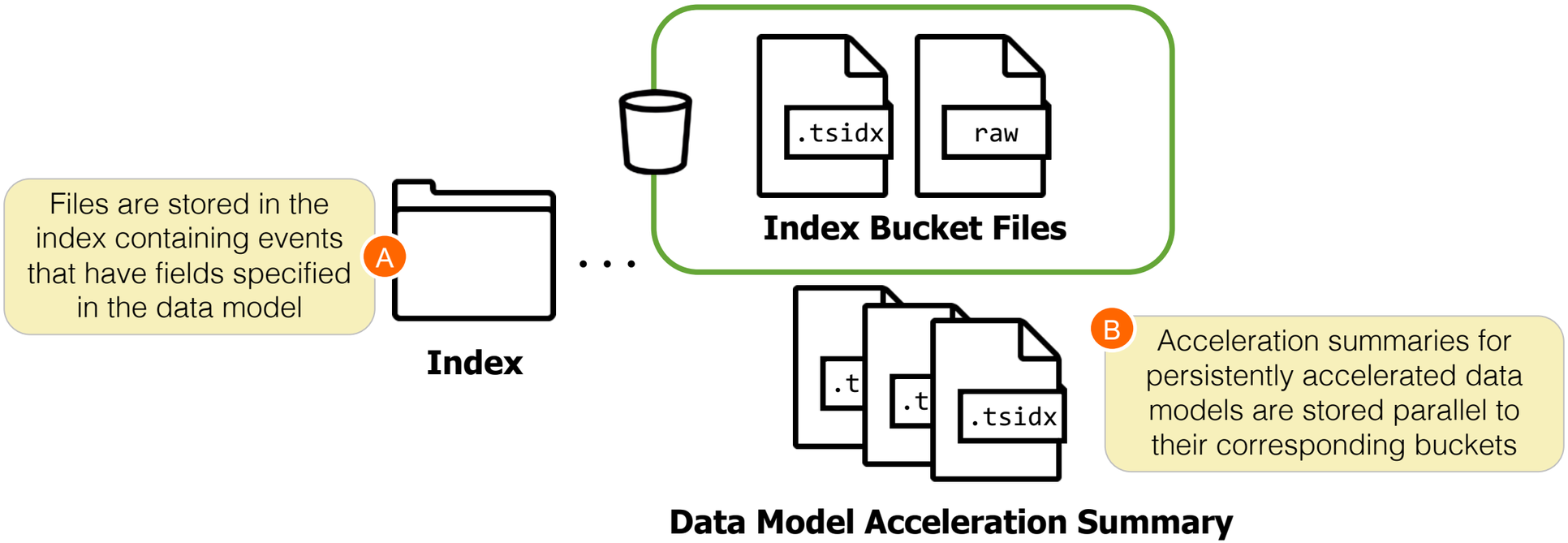
Accelerate a Data Model

1. Click **Settings > Data Models**
2. Select a data model and click **Edit > Edit Acceleration**
3. Click the **Accelerate** check box and choose a **Summary Range**



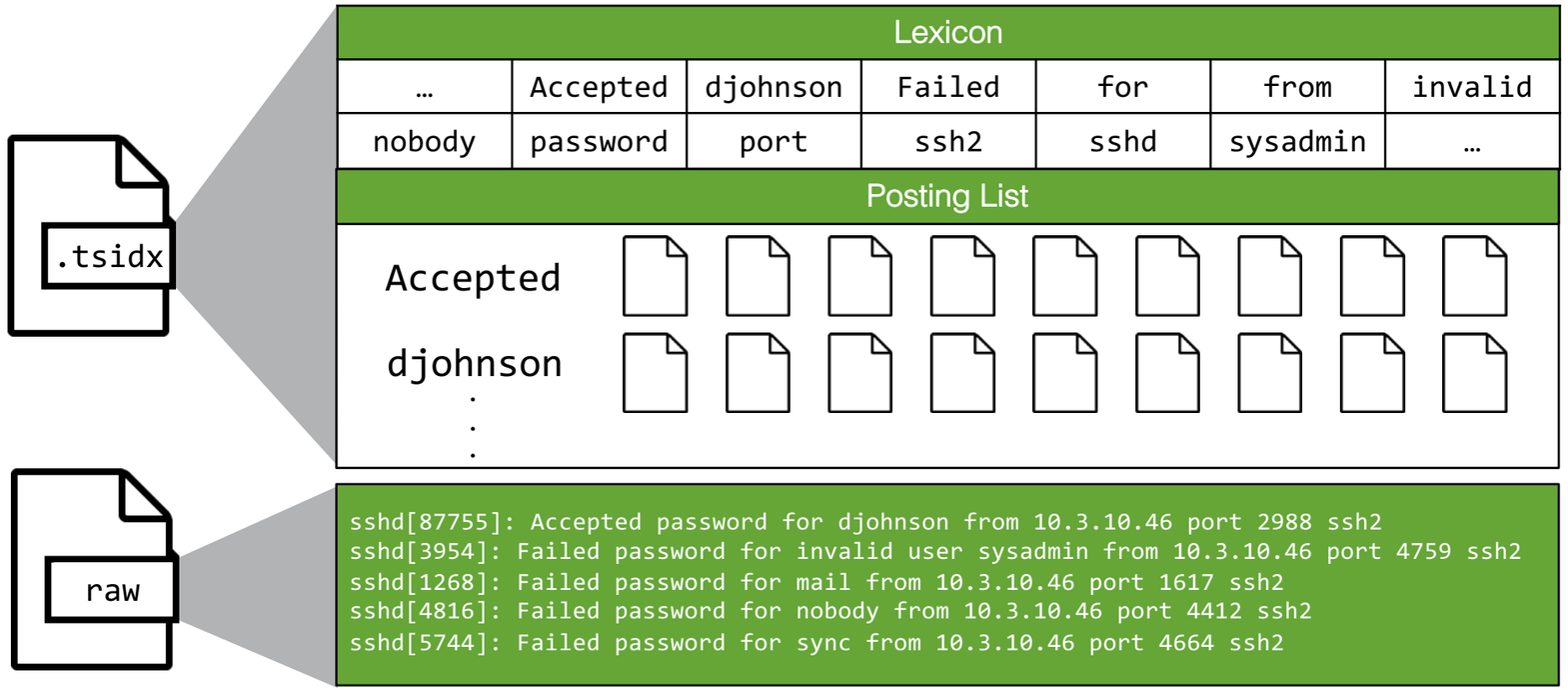
After Accelerating a Data Model

Splunk builds an acceleration summary for the specified summary range in the form of inverted time-series index (`tsidx`) files



Time Series Index (`tsidx`) Files

- Exist inside buckets alongside raw data files
- Consist of a lexicon and a posting list and the indexed `field::value` combinations (`host`, `source`, and `sourcetype`)



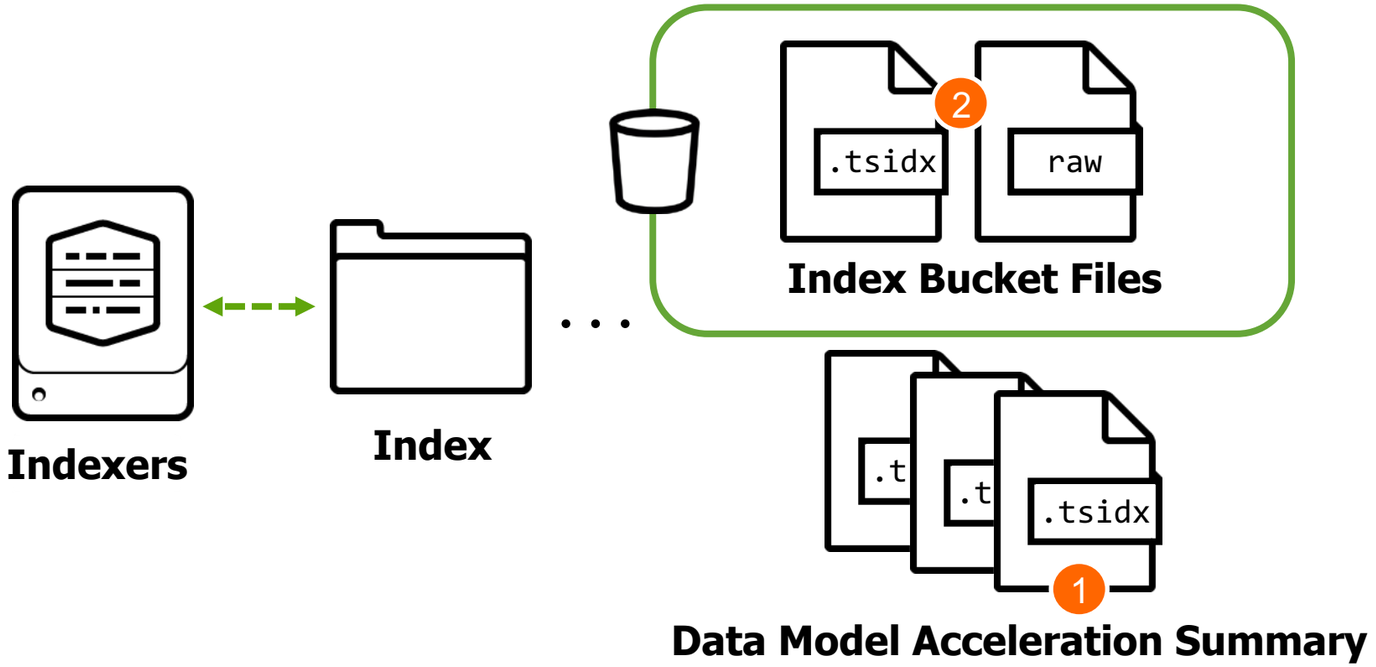
The **lexicon** is an alpha-numerically ordered list of terms found in the data at index time

The **posting list** is an array of pointers that match each term to events in the raw data files

Splunk uses the pointers to search just the events that match the terms, making the search much more efficient

Search With Data Model Acceleration

1. Indexer retrieves information about the data model that has been stored on disk in the `tsidx` files that make up the acceleration summary
2. Indexer pulls additional events from bucket files if search is outside data summary range



Accelerated Data Model Considerations

- The acceleration summary always contains a store of data that at least meets the summary range (may slightly exceed)
- Splunk updates **`tsidx`** files every 5 minutes and removes outdated summary data every 30 minutes
- Accelerated data model summaries can be accessed through:
  - Pivot editor
  - Searches using **`pivot`**, **`tstats`** or **`datamodel`** (outside the scope of this course)

Accelerate Data Models Lab Exercise

Time: 20 minutes

Tasks:

- Accelerate a data model
- Explore your data model using search commands

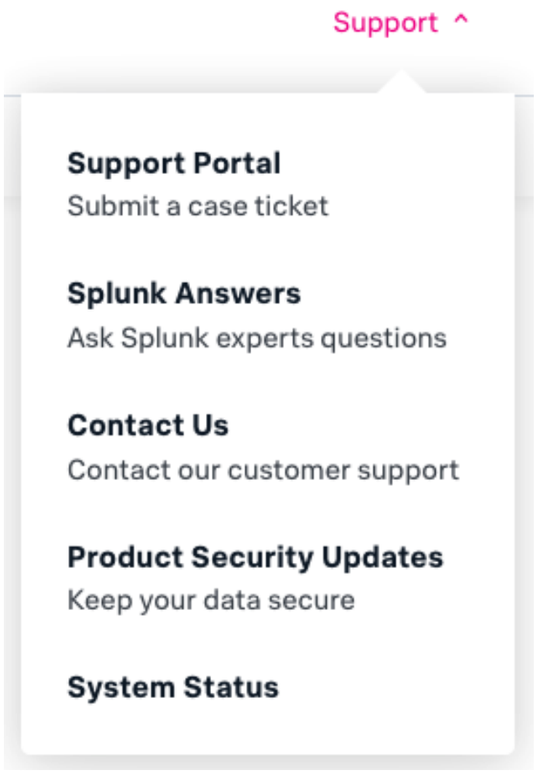
Wrap-up Slides

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
- **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



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.

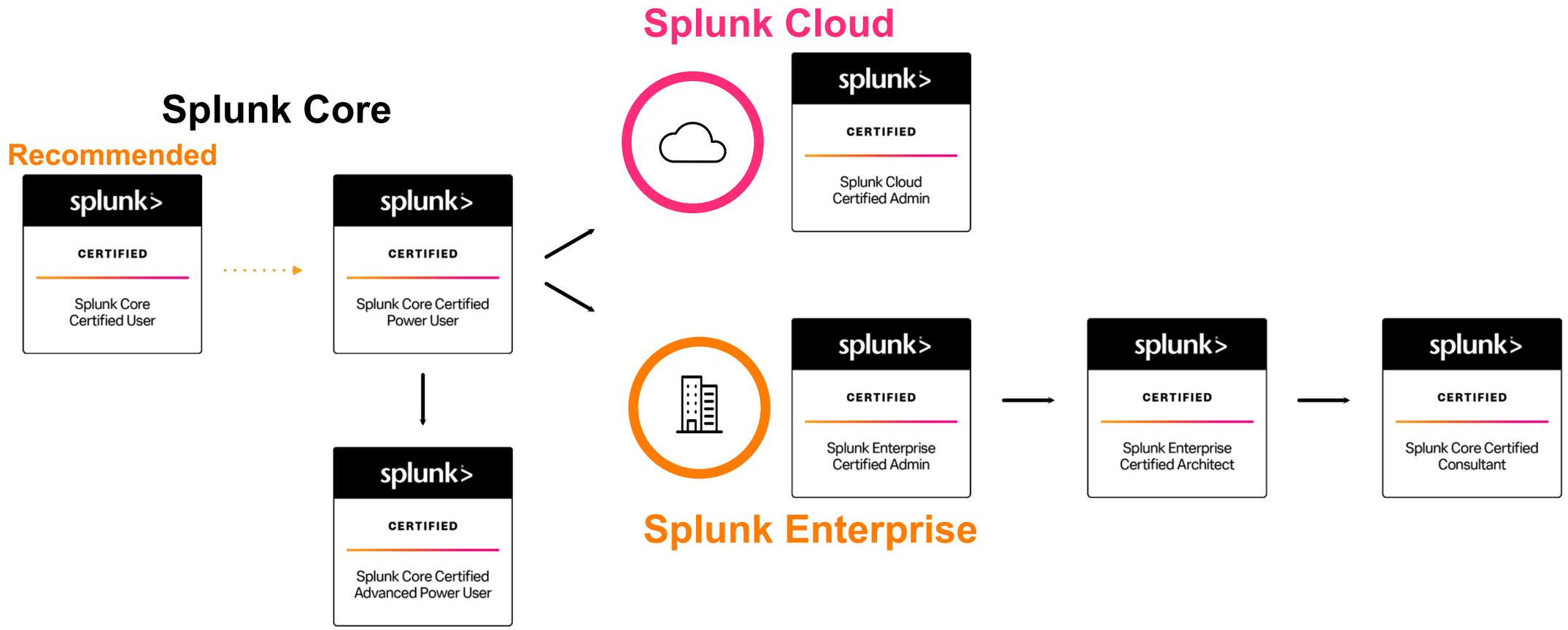
- What is Splunk *
- Introduction to Splunk *
- Using Fields *
- Scheduling Reports and Alerts
- Visualizations
- Statistical Processing
- Working with Time
- Comparing Values
- Result Modification
- Leveraging Lookups and Subsearches
- Correlation Analysis
- Search Under the Hood
- Multivalue Fields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

Recommended Next Step

- Splunk Core Certified Power User

Splunk Core Certified Power User

This entry-level certification demonstrates an individual's foundational competence of Splunk's core software

Prerequisite Certification(s):

- None

Prerequisite Course(s):

- None

Splunk Core Certified Power User Exam

Time to study! We suggest candidates looking to prepare for this exam complete Fundamentals 2 **or** the following courses:

- Visualizations
- Statistical Processing
- Working with Time
- Comparing Values
- Result Modification
- Correlation Analysis
- Search Under the Hood
- Introduction to Knowledge Objects
- Creating Knowledge Objects
- Creating Field Extractions
- Data Models
- Using Choropleth

See here for registration assistance.

Congratulations! You are a...

Recommended Next Steps

- Splunk Core Certified Advanced Power User
- Splunk Enterprise Certified Admin
- Splunk Cloud Certified Admin

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...

Recommended Next Steps

- Splunk Enterprise Certified Admin
- Splunk Cloud Certified Admin

Thank You