Define initial data for a new table dataset

When you create a new table dataset with the Table Editor you start by defining initial data. You have three options for initial data.

- **An index and source type combination** - You can populate your new dataset with events associated with a combination of indexes and source types.
- **An existing dataset** - Your dataset can get its initial data from a dataset that already exists. The dataset can be a table dataset, a data model dataset, a CSV lookup table, or a CSV lookup definition.
- **A search** - You can base your dataset on the results of any search string, as long as it does not include transforming commands.

If you use Splunk Analytics for Hadoop and want to create a dataset based on data from a virtual index, you must get your initial data either from a search that references the virtual index or from an existing dataset that already has the virtual index data.

Identify an index and source type combination for initial data

1. In the Search & Reporting app, open the Datasets listing page.
2. Click **Create New Table Dataset** to go to the initial data setup screen of the Table Editor.
3. Select **Indexes & Source Types**.
4. Choose an index that you want to use for initial data. If you do not want to select a specific index, select **All indexes**.
5. Select a source type that you want to use for initial data. If you do not want to select a specific source type, select **All source types**.

   If you select both **All indexes** and **All source types**, you risk creating an overly broad dataset that contains all of the events indexed by your Splunk implementation (with the exception of events in `_internal` and other internal indexes, which you must specify by name). In general you should avoid creating overly broad datasets. The datasets feature is designed for creating narrow views of data.

   A preview of your dataset appears. Rows are events, columns are fields, and cells are field values.

6. (Optional) Click **Add an index and one or more source types**... to create a dataset that pulls data from more than one index and source type combination.
7. Select existing fields that you want to see in your dataset. Click **OK** when you are done.

   Hover over a listed field to see field statistics, such as the percentage of events in the dataset that have the field, and the top values for the field.

8. (Optional) If you are not seeing a field choice that you are expecting, add the missing field.
   - At the bottom of the field list, click **Add a missing existing field**.
   - Enter the field and click **Add**.
   - Select the added field.

9. Use the dataset preview pane to verify that this is the initial data that you want. If you do not find the existing fields or field values that you were expecting you can remove this selection and select another one.
10. (Optional) If you are not sure whether the index and source type combination you have chosen contains the events you are looking for, change the **Sample** setting at the top of the preview pane to see random events from the dataset or select a new sample.
11. When you are satisfied that your index, source type, and field selections provide the correct initial data for your dataset, click **Done** to move on to the Table Dataset Editor.
**Use an existing dataset for initial data**

The **Datasets** tab lets you select an existing dataset for your initial data. You can select any dataset that you can otherwise see on the Datasets listing page, including data model datasets, lookup tables, and lookup definitions.

When you create a dataset that uses an existing dataset for initial data, you can choose between cloning and extending the existing dataset.

**Prerequisites**

- See About datasets.
- See Dataset extension.

**Steps**

1. In the Search & Reporting app, open the Datasets listing page.
2. Click **Create New Table Dataset** to go to the initial data setup screen of the Table Editor.
3. Select **Existing Datasets**.
4. Select either **Clone** or **Extend**.

<table>
<thead>
<tr>
<th>Selection</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clone</td>
<td>Creates an identical copy of the original dataset. Only table datasets can be cloned.</td>
</tr>
<tr>
<td>Extend</td>
<td>Creates a dataset that is extended from an existing dataset. Changes made to the original dataset propagate down to the extended dataset. All dataset types can be extended.</td>
</tr>
</tbody>
</table>

5. If you are working with a lookup table file, select the fields that you want to use in your table.

   The fields you select are the only fields that will make up your dataset, along with _raw and _time, which are required. You can hover over a field to see field statistics, such as the percentage of events in the dataset that have the field, and the top values for the field.

   Table datasets, data model datasets, and lookup definitions have fixed fields. When you create a new dataset by cloning or extending a dataset with fixed fields, you do not get to choose which of those fields you want to start with in your dataset.

6. (Optional) If you are not seeing a field choice that you are expecting, add the missing field.
   - At the bottom of the field list, click **Add a missing existing field**.
   - Enter the field and click **Add**.
   - Select the added field.

7. Use the dataset preview pane to verify that this is the initial data that you want. If you do not find the existing fields or field values that you were expecting you can select a different dataset.

8. (Optional) If you are not sure whether the dataset you have chosen contains the events you are looking for, change the **Sample** setting at the top of the preview pane to see random events from the dataset or select a new sample.

9. When you are satisfied that your dataset selection provides the correct initial data for your dataset, click **Done** to move on to the Table Dataset Editor.

**Provide a search string for initial data**

There are four methods that you can follow to derive the search string for initial data. Once you provide the search string, the other initial data setup steps are the same.
The search string you provide must identify the fields that its search commands operate on. For example, a search that only includes commands like sendemail, highlight, or delete will be invalid because those commands do not require that you identify the fields that they operate upon.

**Provide the full search string in the Table Editor**

1. In the Search & Reporting app, open the Datasets listing page.
2. Click **Create New Table Dataset** to go to the initial data setup screen of the Table Editor.
3. Click **Search (Advanced)**.
4. Provide a search string that returns data you want in your table.
5. Continue the initial data definition process by following the steps in Preview your dataset and select its starting fields.

**Use a search string you have designed in the Search view**

1. In the Search view, design a search that returns events that you want in your table.
2. Click **New Table** to use the search as the initial data for a new table dataset. The Table Editor opens with **Search (Advanced)** selected and the search string you designed in the search field.
3. (Optional) Add additional SPL until you have a search that returns results that you would like to use in a dataset.
4. Continue the initial data definition process by following the steps in Preview your dataset and select its starting fields.

**Start with a search string that includes your index and source type selections**

1. In the Search & Reporting app, open the Datasets listing page.
2. Click **Create New Table Dataset** to go to the initial data setup screen of the Table Editor.
3. Select **Indexes & Source Types**.
4. Choose an index that you want to use for initial data. If you do not want to select a specific index, select **All indexes**.
5. Choose a source type that you want to use for initial data.
6. (Optional) Click **Add an index and one or more source types...** to create a dataset that pulls data from more than one index and source type combination.
7. Click **Search (Advanced)**. The search field populates with the index and source type combination that you have selected.
8. (Optional) Add additional SPL until you have a search that returns results that you would like to use in a dataset.
9. Continue the initial data definition process by following the steps in Preview your dataset and select its starting fields.

**Start with a search string that extends an existing dataset**

This method creates a dataset that is extended from an existing dataset. Changes made to the original dataset propagate down to the extended dataset. All dataset types can be extended.

1. In the Search & Reporting app, open the Datasets listing page.
2. Click **Create New Table Dataset** to go to the initial data setup screen of the Table Editor.
3. Select **Existing Datasets**.
4. Select **Extend**.
5. Select a dataset that you want to use for initial data. If you have a significant number of datasets to choose from, click the magnifying glass to search for the dataset you want.
6. Click **Search (Advanced)**. The search field populates with SPL referencing your selected dataset.
7. (Optional) Select the fields you would like to see in your dataset. You can select fields whether or not the original dataset type has fixed fields.
8. (Optional) Add additional SPL until you have a search that returns results that you would like to use in a dataset.
9. Continue the initial data definition process by following the steps in Preview your dataset and select its starting fields.

**Preview your dataset and select its starting fields**

When you begin this task, you must have first used one of the previous four tasks to define the search string for your initial data.

1. After you define the search string for your initial data, press the **Enter** key on your keyboard or click the magnifying glass icon to run the search.

   A preview of your dataset appears. Rows are events, columns are fields, and cells are field values. Update the search and run it again until you are satisfied with the results.

2. Select existing fields that you want to see in your dataset. Click **OK** when you are done.

   Hover over a listed field to see field statistics, such as the percentage of events in the dataset that have the field, and the top values for the field.

3. (Optional) If you are not seeing a field choice that you are expecting, add the missing field.
   ♦ At the bottom of the field list, click **Add a missing existing field**.
   ♦ Enter the field and click **Add**.
   ♦ Select the added field.

4. Use the dataset preview pane to verify that this is the initial data that you want. If you do not find the existing fields or field values that you were expecting you can modify the search.

5. (Optional) If you are not sure whether your search string will return the events you are looking for, change the **Sample** setting at the top of the preview pane to see random events from the dataset or select a new sample.

6. When you are satisfied that your index, source type, and field selections provide the correct initial data for your dataset, click **Done** to move on to the Table Dataset Editor.