
Wave Analytics External Data Format Reference

Salesforce, Winter '17



CONTENTS


- External Data Metadata Overview 1
- External Data Metadata Format Reference 4

EXTERNAL DATA METADATA OVERVIEW

To upload external data into a Wave Analytics dataset, make sure that you have prepared your data and metadata files.

You can load external data into a dataset by preparing two files.

- A data file, which contains the external data, in comma-separated value (CSV) format
- An optional metadata file, which describes the structure of the data file in JSON format

 **Note:** Providing a metadata file is recommended. Otherwise, every field is treated as text.

The data and metadata files are used to populate a dataset with the external data.

CSV Example

The following CSV example contains data that conforms to the .json metadata file that's described next.

```
Name,Amount,CloseDate
opportunityA,100.99,6/30/2014
opportunityB,99.01,1/31/2012
```

The first row in the CSV file lists the field names for your dataset. Each subsequent row corresponds to a record of data. A record consists of a series of fields delimited by commas. For information on creating valid field names, see [Field Names](#).

JSON Example

The following JSON example represents a SalesData object with three fields: Name, Amount, and CloseDate. The example corresponds to the preceding CSV example.

```
{
  "fileFormat": {
    "charsetName": "UTF-8",
    "fieldsEnclosedBy": "\"\"",
    "fieldsDelimitedBy": ",",
    "numberOfLinesToIgnore": 1
  },
  "objects": [
    {
      "connector": "AcmeCSVConnector",
      "description": "",
      "fullyQualifiedName": "SalesData",
      "label": "Sales Data",
      "name": "SalesData",
      "fields": [
        {
          "description": "",
          "fullyQualifiedName": "SalesData.Name",
          "label": "Account Name",
          "name": "Name",

```

```

        "isSystemField": false,
        "isUniqueId": false,
        "isMultiValue": false,
        "type": "Text"
    },
    {
        "description": "",
        "fullyQualified_name": "SalesData.Amount",
        "label": "Opportunity Amount",
        "name": "Amount",
        "isSystemField": false,
        "defaultValue": "0",
        "isUniqueId": false,
        "type": "Numeric",
        "precision": 10,
        "scale": 2,
        "format": "$#,##0.00"
    },
    {
        "description": "",
        "fullyQualified_name": "SalesData.CloseDate",
        "label": "Opportunity Close Date",
        "name": "CloseDate",
        "isSystemField": false,
        "isUniqueId": false,
        "type": "Date",
        "format": "MM/dd/yyyy",
        "fiscalMonthOffset": 0
    }
]
}
]
}

```

CSV Format

The External Data API uses a strict format for field values to optimize processing for large sets of data. Note the following when generating .csv files.

- The delimiter for field values in a row must be a comma.
- If a field value contains a comma, a new line, or a double quote, the field value must be contained within double quotes: for example, "Director of Operations, Western Region".
- If a field value contains a double quote, escape the double quote by preceding it with another double quote: for example, "This is the ""gold"" standard".
- Field values aren't trimmed. A space before or after a delimiting comma is included in the field value. A space before or after a double quote generates an error for the row. For example, John, Smith is valid. John, Smith is valid, but the second value is " Smith". "John", "Smith" isn't valid.
- Numeric values in the CSV file can't contain any formatting (such as currency symbols or grouping separators). For example, \$1,000.00 is not a valid numeric value; the correct value is 1000.00.
- The maximum numeric value is 36,028,797,018,963,967 and the minimum is -36,028,797,018,963,968.

- Dates must conform to specific formats, and they must match the formats exactly. For more information, see [Date Formats](#).
- At least one column in the CSV file must contain dimension values.
- If column headers are specified, the number of column headers must equal the number of columns in each record.

For more information about field names, see [Field Name Restrictions](#).

External Data Limits

The following limits apply to all supported editions. Contact salesforce.com to extend the limits.

External Data Limits

Limit	Value
Maximum file size per external data upload	If you use the External Data API: 40 GB. If you use the user interface: 512 MB.
Maximum file size for all external data uploads in a rolling 24-hour period	50 GB
Maximum number of external data files that can be uploaded in a rolling 24-hour period	50
Maximum number of characters in a field	32,000
Maximum number of fields in a record	5,000 (including up to 1,000 date fields)
Maximum number of characters for all fields in a record	400,000

EXTERNAL DATA METADATA FORMAT REFERENCE

The metadata describes the structure of external data files. The metadata file is in JSON format. The .json file consists of 3 main sections: file format, object information, and field information. Include all required fields when you create a record, but you can leave out optional fields.

The File Format Section

The file format section of the metadata file specifies information about the format of the data file, including the character set and delimiter character.

Field Name	Type	Required?	Description
charsetName	String	No	The character set of the .csv file. If this field is included, it must be set to UTF-8. Example: "charsetName": "UTF-8"
fieldsEnclosedBy	String	No	The character that can be used to enclose fields in the .csv file. Only double quotes are supported. If a double quote is used within a field, escape it by preceding it with another double quote. If this field is included, it must be set to "\". Example: "fieldsEnclosedBy": "\""
fieldsDelimitedBy	String	No	The character that separates field values in the .csv file. If this field is included, it must be set to ",". Example: "fieldsDelimitedBy": ","
linesTerminatedBy	String	No	<i>Deprecated. Do not use.</i>
numberOfLinesToIgnore	Number	No	The number of lines for the parser to ignore. (Allows you to specify a header.) <ul style="list-style-type: none">• When the .csv file doesn't have a header, set to 0.• When the .csv file has a header, set to 1. Example: "numberOfLinesToIgnore": 1

The Objects Section

The objects section of the metadata file specifies information about the top-level database object, including object-level security information, display name, and API name.

 **Note:** The metadata file can contain only 1 object definition.


Field Name	Type	Required?	Description
rowLevelSecurityFilter	String	No	<p>The predicate that's used to apply row-level security on the dataset.</p> <p>When entering the predicate in the metadata file JSON, escape double quotes around string values.</p> <p>Example:</p> <pre>"rowLevelSecurityFilter": "'OwnerId' == \">\$User.Id\ '"</pre> <p>For more information about creating the predicate, see the <i>Security Implementation Guide for Wave Analytics</i>.</p>
connector	String	Yes	<p>The string that uniquely identifies the client application.</p> <p>Example:</p> <pre>"connector": "AcmeCSVConnector"</pre>
description	String	No	<p>The description of the object. Must be less than 1,000 characters.</p> <p>Example:</p> <pre>"description": "The SalesData object tracks basic sales data."</pre>
fullyQualifiedName	String	Yes	<p>The full path that uniquely identifies the record. Must be less than 1,000 characters.</p> <p>Example:</p> <pre>"fullyQualifiedName": "CRM.SalesData"</pre> <p>For information on creating valid field names, see Field Names.</p>
label	String	Yes	<p>The display name for the object. Can be up to 40 characters.</p> <p>Example:</p> <pre>"label": "Sales Data"</pre>
name	String	Yes	<p>The unique API name for the object. Can be up to 255 characters.</p> <p>Example:</p> <pre>"name": "SalesData"</pre> <p>For information on creating valid field names, see Field Names.</p>


Field Name	Type	Required?	Description
fields	Array	Yes	The array of fields for this object.


The Fields Section

The fields section of the metadata file specifies information about each field in the record, including data type and formatting information.

 **Note:** The fields must be in the same order as the CSV columns are in.

Field Name	Type	Required?	Description
fullyQualifiedName	String	Yes	<p>The full path that uniquely identifies the field (object.field). Must be less than 1,000 characters.</p> <p>Example:</p> <pre>"fullyQualifiedName": "SalesData.Amount"</pre> <p>For information on creating valid field names, see Field Names.</p>
label	String	Yes	<p>The display name for the field. Can be up to 255 characters.</p> <p>Example:</p> <pre>"label": "Opportunity Amount"</pre>
name	String	Yes	<p>The unique API name for the field. Can be up to 255 characters.</p> <p>Example:</p> <pre>"name": "Amount"</pre> <p> Note: With Summer '15 Plus, we recommend that field names in datasets use no more than 40 characters. Long field names increase the likelihood of exceeding character limits when you augment datasets, because names are appended. Field names longer than 40 characters are currently supported, but we recommend shortening them to avoid issues.</p> <p>For information on creating valid field names, see Field Names.</p>
description	String	No	<p>The description of the field. Must be less than 1,000 characters.</p> <p>Example:</p> <pre>"description": "The Amount field contains the opportunity amount."</pre>
isSystemField	Boolean	No	<p>Indicates whether this field is a system field to be excluded from query results.</p> <p>Example:</p> <pre>"isSystemField": false</pre>

Field Name	Type	Required?	Description
type	String	Yes	<p>The type of the field. Can be Text, Numeric, or Date.</p> <p>Example:</p> <pre>"type": "Numeric"</pre>
defaultValue	String	No	<p>The default value of the field, if any. All numeric types require a default value.</p>
isUniqueId	Boolean	No	<p>Indicates whether this field is the primary key for the record. This field is required for incremental extract. Only 1 field can be set to be the unique ID.</p> <p> Note: Only text fields can be unique IDs. Numeric, date, and multi-value fields can't be unique IDs.</p> <p>Example:</p> <pre>"isUniqueId": false</pre>
isMultiValue	Boolean	No	<p>For text fields only. Indicates whether the field has multiple values. Applies only to Text fields.</p> <p>Example:</p> <pre>"isMultiValue": false</pre>
multiValueSeparator	String	No	<p>For text fields only. The character that separates multiple values. The default is ",".</p> <ul style="list-style-type: none"> • If <code>isMultiValue</code> equals true, specify a value. • If <code>isMultiValue</code> equals false, this field can be set to null. <p>Example:</p> <pre>"multiValueSeparator": ";"</pre>
format	String	Yes (for Date values only)	<p>The format of the numeric or date value. See also Numeric Formats and Date Formats.</p> <p>Example:</p> <pre>"format": "\$#,##0.00" (Numeric)</pre> <pre>"format": "dd-MM-yy HH:mm:ss" (Date)</pre>
precision	Number	Yes (for Numeric values)	<p>The maximum number of digits in a numeric value, or the length of a text value.</p> <p>For numeric values: Includes all numbers to the left and to the right of the decimal point (but excludes the decimal point character). Value can be up to 18.</p> <p>For text values: Value defaults to 255 characters, but can be up to 32,000 characters.</p> <p>Example:</p>

Field Name	Type	Required?	Description
			<code>"precision": 10</code>
scale	Number	Yes (for Numeric values)	<p>The number of digits to the right of the decimal point in a numeric value. Must be less than the precision value.</p> <p>Example:</p> <p><code>"scale": 2</code></p>
canTruncateValue	Boolean	No	<p>Indicates whether to truncate a value when the value exceeds the precision. The default is true.</p> <ul style="list-style-type: none"> • If true, truncates the value. • If false, the row is rejected. <p>Example:</p> <p><code>"canTruncateValue": true</code></p>
decimalSeparator	String	No	<p>For numeric fields only. The character that separates digits in a decimal number. Can be used to handle international number formats where the decimal separator is ",". The default is ".".</p> <p>Example:</p> <p><code>"decimalSeparator": ",",</code></p>
fiscalMonthOffset	Number	No	<p>For date fields only. The difference, in months, between the fiscal year and the calendar year. For example, if the fiscal year starts in January, the offset is 0. If the fiscal year starts in October, the offset is 9.</p> <p>Example:</p> <p><code>"fiscalMonthOffset": 9</code></p> <p> Note: This attribute also controls whether Wave Analytics generates fiscal date fields. To generate fiscal date fields, set <code>fiscalMonthOffset</code> to a value other than 0.</p> <p>See also Date Handling in Datasets.</p>
isYearEndFiscalYear	Boolean	No	<p>For date fields only. Indicates whether the fiscal year is the year in which the fiscal year ends or begins. Because the fiscal year can start in one calendar year and end in another, specify which year to use for the fiscal year.</p> <ul style="list-style-type: none"> • If true, then the fiscal year is the year in which the fiscal year ends. The default is true. • If false, then the fiscal year is the year in which the fiscal year begins. <p>Example:</p> <p><code>"isYearEndFiscalYear": true</code></p> <p>This field is relevant only when <code>fiscalMonthOffset</code> is greater than 0.</p>

Field Name	Type	Required?	Description
			See also Date Handling in Datasets .
<code>firstDayOfWeek</code>	Number	No	<p>For date fields only. The first day of the week for the calendar year and, if applicable, fiscal year. Use 0 to set the first day to be Sunday, 1 to set the first day to be Monday, and so on. Use -1 to set the first day to be January 1. The default is -1.</p> <p>Example:</p> <pre>"firstDayOfWeek": 0</pre> <p>See also Date Handling in Datasets.</p>
<code>isSkipped</code>	Boolean	No	<p>Indicates whether to skip the field when the data is uploaded.</p> <p>Example:</p> <pre>"isSkipped": true</pre>

Field Name Restrictions


Field names in the .csv file and the metadata file:

- Can contain only alphanumeric and underscore characters
- Must begin with a letter
- Can't end with an underscore
- Can't contain 2 consecutive underscore characters, except when ending with “__c” (case-sensitive)
- Must be unique across all fields of the object

Numeric Formats

An example of a typical numeric value is \$1,000,000.99, which is represented as \$#,##0.00. You are required to specify the precision and scale of the number. The format is specified by using the following symbols:

Symbol	Meaning
0	One digit
#	Zero or 1 digit
.	This symbol is the default decimal separator. Use the <code>decimalSeparator</code> field to set the decimal separator to a different symbol.
-	Minus sign
,	Grouping separator
\$	Currency sign

 **Note:** The format for numeric values isn't used in data ingestion. It is used only to specify how numeric values are formatted when displayed in the UI. For data ingestion, numeric values can't contain any formatting (such as currency symbols or grouping separators). For example, \$1,000.00 is not a valid numeric value; the correct value is 1000.00.

Date Formats

For Date fields, specify the format of the date by using one of the following formats. Dates must match the format exactly and can't have any extra text. For example, if the date format is "MM-dd-yyyy hh:mm:ss" and the value is "12-31-2015 12:00:00.0000", the upload fails because the value has extra milliseconds.

 **Note:** The timestamp portion of each date format is optional.

Format	Sample Value
yyyy-MM-dd'T'HH:mm:ss.SSS'Z'	2014-04-29T16:53:34.000Z
yy-MM-dd'T'HH:mm:ss.SSS'Z'	14-04-29T16:53:34.000Z
yyyy-MM-dd'T'HH:mm:ss'Z'	2014-04-29T16:53:34Z
yy-MM-dd'T'HH:mm:ss'Z'	14-04-29T16:53:34Z
yyyy-MM-dd HH:mm:ss	2014-06-03 11:31:45
yy-MM-dd HH:mm:ss	14-06-03 11:31:45
dd.MM.yyyy HH:mm:ss	03.06.2014 11:31:45
dd.MM.yy HH:mm:ss	03.06.14 11:31:45
dd/MM/yyyy HH:mm:ss	03/06/2014 11:31:45
dd/MM/yy HH:mm:ss	03/06/14 11:31:45
dd/MM/yyyy hh:mm:ss a	03/06/2014 11:31:45 AM
dd/MM/yy hh:mm:ss a	03/06/14 11:31:45 AM
dd-MM-yyyy HH:mm:ss	03-06-2014 11:31:45
dd-MM-yy HH:mm:ss	03-06-14 11:31:45
dd-MM-yyyy hh:mm:ss a	03-06-2014 11:31:45 AM
dd-MM-yy hh:mm:ss a	03-06-14 11:31:45 AM
MM/dd/yyyy hh:mm:ss a	06/03/2014 11:31:45 AM
MM/dd/yy hh:mm:ss a	06/03/14 11:31:45 AM
MM-dd-yyyy hh:mm:ss a	06-03-2014 11:31:45 AM
MM-dd-yy hh:mm:ss a	06-03-14 11:31:45 AM
HH:mm:ss dd/MM/yyyy	11:31:45 03/06/2014
HH:mm:ss dd/MM/yy	11:31:45 03/06/14

 **Note:** Only the preceding formats are supported.

These formats use the following symbols:

Symbol	Meaning
yyyy or yy	Four-digit year (yyyy) or two-digit year (yy)
MM	Two-digit month (01–12)
dd	Two-digit day (01–31)
'T'	A separator that indicates that time of day follows
HH	Two-digit hour (00–23)
mm	Two-digit minute (00–59)
ss	Two-digit seconds (00–59)
SSS	Optional three-digit milliseconds (000–999)
'Z'	The reference UTC time zone