

# Analytics External Data API Developer Guide

Salesforce, Spring '24





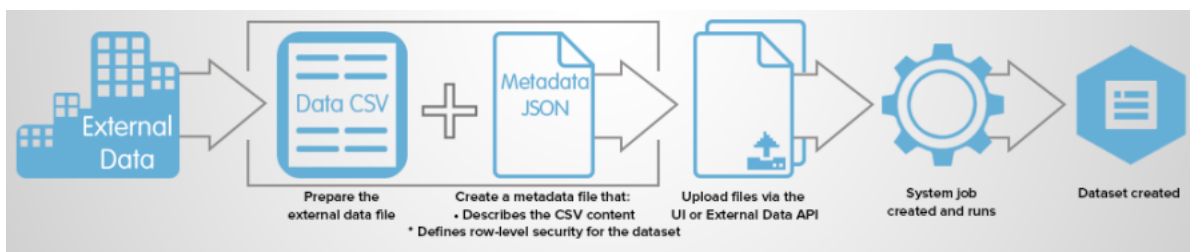
# CONTENTS

- [External Data](#) ..... 1
  - Load External Data into a Dataset ..... 1
  - External Data API Reference ..... 4
- [Release Notes](#) ..... 13



# EXTERNAL DATA

You can integrate external data into CRM Analytics to make the data available for queries from lenses and designer.



The External Data API enables you to upload external data files to CRM Analytics. The External Data API can upload .csv files, and you can optionally specify the structure of your data by defining metadata in JSON format.

The External Data API is available in API version 31 and later.

The high-level steps for uploading external data by using the API are:

1. Prepare your data in CSV format, and then create a metadata file to specify the structure of the data.
2. Connect programmatically to your Salesforce organization.
3. Configure the upload by inserting a row into the `InsightsExternalData` object, and then set input values such as the name of the dataset, the format of the data, and the operation to perform on the data.
4. Split your data into 10-MB chunks, and then upload the chunks to `InsightsExternalDataPart` objects.
5. Start the upload by updating the `Action` field in the `InsightsExternalData` object.
6. Monitor the `InsightsExternalData` object for status updates, and then verify that the file upload was successful.

## Load External Data into a Dataset

### External Data API Limits

When working with the External Data API, consider the following limits, in addition to the general CRM Analytics limits.

The following limits apply to all supported editions.

Value	Limit
Maximum file size per external data uploads	40 GB
Maximum file size for all external data uploads in a rolling 24-hour period	50 GB
Maximum number of external data jobs per dataset that can be run in a rolling 24-hour period	50
Maximum number of characters in a field	32,000


Value	Limit
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

## Prepare Data Files

To upload external data from .csv files into a dataset, first prepare your data files.

External data can be loaded into a dataset by preparing two files.


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

 **Note:** Providing a metadata file is recommended.

The data and metadata files are used to populate a dataset with the external data. For detailed information about formatting CSV data and JSON metadata, see the [Analytics Extended Metadata \(XMD\) Developer Guide](#).

## Connect to Salesforce

After preparing your data files, the next step in loading external data into CRM Analytics is to connect to your Salesforce organization by using standard Salesforce APIs.

 **Note:** The following examples use SOAP API, but you can use any of the Salesforce APIs, such as REST API or Apex. The examples assume that you're using the Web Services Connector.

To load external data into CRM Analytics, first connect to your Salesforce organization. Use the PartnerConnection object to log in to your organization, as shown in the following example. You need to supply a username, password, and endpoint.

```
ConnectorConfig config = new ConnectorConfig();
config.setUsername(username);
config.setPassword(password);
config.setAuthEndpoint(endpoint);

PartnerConnection partnerConnection = new PartnerConnection(config);
```

For more information about the Web Services Connector (WSC), see [Introduction to the Web Services Connector](#). For more information about user authentication, see [Security and the API](#) in the *SOAP API Developer Guide*.

## Configure the Upload

Configure the external data upload by inserting a row into the InsightsExternalData object and setting configuration values.


After establishing a connection with Salesforce, insert a row into the InsightsExternalData object to configure and control the upload. The InsightsExternalData object provides a "header" that contains information about the upload, such as the name of the dataset, the format of the data, and the operation to perform on the data. You can also provide the metadata file. The following example inserts a row into the InsightsExternalData object and sets configuration values.

```
SObject sobj = new SObject();
soobj.setType("InsightsExternalData");
soobj.setField("Format", "Csv");
soobj.setField("EdgemartAlias", DatasetName);
```

```
sobj.setField("MetadataJson",metadataJson);
sobj.setField("Operation","Overwrite");
sobj.setField("Action","None");

SaveResult[] results = partnerConnection.create(new SObject[] { sobj });

for(SaveResult sv:results)
    if(sv.isSuccess())
        parentID = sv.getId();
```

 **Note:** The WSC converts the metadata.json file to a Base64-encoded string, but if you're using REST API, you need to make this conversion yourself.

For detailed information about the `InsightsExternalData` object, see [The InsightsExternalData Object](#).

## Add the Data

When uploading external data files, you can use the `InsightsExternalDataPart` object to load the data in smaller chunks.

After inserting a row into the `InsightsExternalData` (header) object, split your data into 10-MB chunks and upload the chunks to `InsightsExternalDataPart` objects. You associate the part objects with the header object by setting the `InsightsExternalDataId` field on the part objects to the ID of the header object. The part objects contain the bytes of data and must be assigned part numbers in a contiguous sequence, starting with 1.

Ensure that the chunks of data are smaller than 10 MB. If the data is compressed, it must be compressed first and then split into 10-MB chunks. Only the gzip format is supported.

The following example splits a file into 10-MB chunks, and then uploads the chunks to `InsightsExternalDataPart` objects.

```
List<File> fileParts = chunkBinary(dataFile); //Split the file

for(int i = 0;i<fileParts.size();i++)
{
    SObject sobj = new SObject();
    sobj.setType("InsightsExternalDataPart");
    sobj.setField("DataFile", FileUtils.readFileToByteArray(fileParts.get(i)));
    sobj.setField("InsightsExternalDataId", parentID);
    obj.setField("PartNumber",i+1); //Part numbers should start at 1
    SaveResult[] results = partnerConnection.create(new SObject[] { sobj });
    for(SaveResult sv:results)
        if(sv.isSuccess())
            rowId = sv.getId();
}
```

For detailed information about the `InsightsExternalDataPart` object, see [InsightsExternalDataPart](#).

## Manage the Upload

After you've created a header and uploaded the data parts by using the `InsightsExternalData` and `InsightsExternalDataPart` objects, update the `Action` field on the header object to `Process` to start processing the data.

The following example sets the `Action` field and updates the row in the `InsightsExternalData` object.

```
SObject sobj = new SObject();
sobj.setType("InsightsExternalData");
```

```
sobj.setField("Action","Process");
sobj.setId(parentID); // This is the rowID from the previous example.
SaveResult[] results = partnerConnection.update(new SObject[] { sobj });

for(SaveResult sv:results)
    if(sv.isSuccess())
        rowId = sv.getId();
```

When the `Action` field is set to `Process`, a dataflow job is created and marked active. You can monitor the `Status` field of the header object to determine when the file upload is completed. After the `Action` field is updated to request processing, no user edits are allowed on the objects.

## Append the Data

After you upload data into a dataset, you can append more data.

To append data, update the `Action` field to `Load` and the `Operation` field to `Append`.

This example sets the `Action` and `Operation` fields and appends data to the `InsightsExternalData` object. Use the `Mode` field with a value of `Incremental` to enable faster uploads.

```
SObject sobj = new SObject();
sobj.setType("InsightsExternalData");
sobj.setField("InsightsExternalDataId", parentID);
sobj.setField("Action","Load");
sobj.setField("Operation","Append");
sobj.setField("Mode","Incremental");

SaveResult[] results = partnerConnection.update(new SObject[] { sobj });

for(SaveResult sv:results)
    if(sv.isSuccess())
        parentID = sv.getId();
```

If `Mode` isn't specified or is set to `None`, the append is a bulk upload that processes data slower.

## External Data API Reference

---

### The InsightsExternalData Object

With the `InsightsExternalData` object, you can configure and control external data uploads. You can use it to provide metadata, trigger the start of the upload process, check status, and request cancellation and cleanup.

The `InsightsExternalData` object is used with the `InsightsExternalDataPart` object, which holds the parts of the data to be uploaded. Together, they provide a programmatic way to upload a large file in parts and trigger a dataflow into a dataset. The first step is to insert a row into the `InsightsExternalData` object. Data parts are then uploaded to `InsightsExternalDataPart` objects. The `Mode` field is used to enable faster uploads. The `Action` field of the `InsightsExternalData` object is updated to start processing and request cancellations. After the `Action` field is updated to request processing, no user edits are allowed on the objects, except to request cancellation.



**Note:** The standard system fields (`CreatedById`, `CreatedDate`, `LastModifiedById`, `LastModifiedDate`, and `SystemModstamp`) are documented in [System Fields](#) in the *Salesforce Object Reference*.

The `InsightsExternalData` object is available in API version 31 and later.





## Supported Calls

`create()`, `delete()`, `describeObjects()`, `query()`, `retrieve()`, `update()`, `upsert()`

## Fields



Field	Details
Action	<p><b>Type</b> Picklist</p> <p><b>Properties</b> Create, Filter, Group, Sort, Update</p> <p><b>Description</b> The action to perform on this data. Picklist values are:</p> <p><b>Abort</b> <b>Reserved for future use.</b> The user no longer wants to upload the data and is requesting that the system stop processing, if possible.</p> <p><b>Delete</b> <b>Reserved for future use.</b> The user wants to remove uploaded data parts as soon as possible. Implies that an Abort status is queued.</p> <p><b>None</b> The user hasn't completed the data upload. This value is the default when the object is created.</p> <p><b>Process</b> The user completed the data upload and is requesting that the system process the data.</p>
CompressedMetadataLength	<p><b>Type</b> Int</p> <p><b>Properties</b> Create, Filter, Group, Nillable, Sort, Update</p> <p><b>Description</b> The length of the compressed metadata .json file. This field is overwritten when data is uploaded. This system field isn't editable.</p>
Dataflow	<p><b>Type</b> String</p> <p><b>Properties</b> Create, Filter, Group, Nillable, Sort, Update</p> <p><b>Description</b> For dataflows that were created in API version 34.0 and later. The unique ID of the dataflow that was used to create the dataset. You can use this field to get the status of the dataflow. This system field isn't editable.</p>

Field	Details
Description	<p><b>Type</b> String</p> <p><b>Properties</b> Create, Filter, Nillable, Sort, Update</p> <p><b>Description</b> The description of the dataset that is only used when creating the dataset.</p>
EdgemartAlias	<p><b>Type</b> String</p> <p><b>Properties</b> Create, Filter, Group, Sort, Update</p> <p><b>Description</b> The alias of a dataset, which must be unique across an organization. The alias must follow the same guidelines as other field names, except that they can't end with "__c". Can be up to 80 characters. For more information, see <i>Field Names</i> in the <a href="#">CRM Analytics External Data Format Developer Guide</a>.</p>
EdgemartContainer	<p><b>Type</b> String</p> <p><b>Properties</b> Create, Filter, Group, Nillable, Sort, Update</p> <p><b>Description</b> The name of the app that contains the dataset.</p> <ul style="list-style-type: none"> <li>• If the name is omitted when you're creating a dataset, the name of the user's private app is used.</li> <li>• If the name is omitted for an existing dataset, the system resolves the app name.</li> <li>• If the name is specified for an existing dataset, the name is required to match the name of the current app that contains the dataset.</li> </ul> <p> <b>Note:</b> Use the developer name or the ID of the app for the name. To get the developer name or ID, run this query:</p> <pre>SELECT Id, DeveloperName, Name, AccessType, CreatedDate, Type FROM Folder where Type = 'Insights'</pre> <p>For example, the display label of an app is Analytics Cloud Public Datasets, but the developer name is AnalyticsCloudPublicDatasets.</p>
EdgemartLabel	<p><b>Type</b> String</p> <p><b>Properties</b> Create, Filter, Group, Nillable, Sort, Update</p>

Field	Details
	<b>Description</b> The display name for the dataset. Can be up to 255 characters.
FileName	<b>Type</b> String <b>Properties</b> Create, Filter, Group, Nillable, Sort, Update <b>Description</b> Identifier of the external data file, such as the file name. A unique value isn't required. It can contain only alphanumeric characters and underscores. It must begin with a letter, not include spaces, not end with an underscore, and not contain two consecutive underscores. The maximum file name is 255 characters. <div>  <b>Note:</b> CRM Analytics doesn't populate this field. You can manually update it via the External Data API. </div>
Format	<b>Type</b> Picklist <b>Properties</b> Create, Filter, Group, Sort, Update <b>Description</b> The format of the uploaded data. Picklist values are: <div> <b>Csv</b>  The data is in CSV format. </div> <div> <b>Binary</b>  Reserved for Salesforce internal use. </div>
isDependentOnLastUpload	<b>Type</b> Boolean <b>Properties</b> Create, Defaulted on create, Filter, Group, Sort, Update <b>Description</b> <b>Reserved for future use.</b> When <code>false</code> , indicates that this upload depends on the previous upload to the same dataset name.
isIndependentParts	<b>Type</b> Boolean <b>Properties</b> Create, Defaulted on create, Filter, Group, Sort, Update

Field	Details
	<p><b>Description</b></p> <p><b>Reserved for future use.</b> When <code>true</code>, indicates that file parts were divided on row boundaries and can be processed independently of each other. The default is <code>false</code>.</p>
LicenseType	<p><b>Type</b></p> <p>picklist</p> <p><b>Properties</b></p> <p>Create, Defaulted on create, Filter, Group, Nillable, Restricted picklist, Sort, Update</p> <p><b>Description</b></p> <p>The license type of the external data file. Possible values are:</p> <ul style="list-style-type: none"> <li>• <code>Aqs</code> (Analytics Query Service)</li> <li>• <code>Cdp</code> (Data Cloud)</li> <li>• <code>DataPipelineQuery</code> (Data Pipeline Query)</li> <li>• <code>EinsteinAnalytics</code> (CRM Analytics)</li> <li>• <code>MulesoftDataPath</code> (Mulesoft DataPath)</li> <li>• <code>IntelligentApps</code> (Intelligent Apps)</li> <li>• <code>Sonic</code> (Salesforce Data Pipelines)</li> </ul> <p>The default value is <code>EinsteinAnalytics</code>.</p>
MetaDataLength	<p><b>Type</b></p> <p>Int</p> <p><b>Properties</b></p> <p>Create, Filter, Group, Nillable, Sort, Update</p> <p><b>Description</b></p> <p>The length of the metadata .json file. This field is overwritten when data is uploaded. This system field isn't editable.</p>
MetadataJson	<p><b>Type</b></p> <p>Blob (Base64-encoded string)</p> <p><b>Properties</b></p> <p>Create, Nillable, Update</p> <p><b>Description</b></p> <p>Metadata in JSON format, which describes the structure of the uploaded file.</p>
Mode	<p><b>Type</b></p> <p>picklist</p> <p><b>Properties</b></p> <p>Create, Filter, Group, Nillable, Restricted picklist, Sort, Update</p> <p><b>Description</b></p> <p>The upload mode for the data. Possible values are:</p>

Field	Details
	<ul style="list-style-type: none"> <li>Incremental</li> <li>None</li> </ul> <p>The default value is None.</p>
NotificationEmail	<p><b>Type</b> String</p> <p><b>Properties</b> Create, Filter, Group, Nillable, Sort, Update</p> <p><b>Description</b> The email address to send notifications to. Can be up to 255 characters and can contain only one email address. Defaults to the current user's email address.</p>
NotificationSent	<p><b>Type</b> Picklist</p> <p><b>Properties</b> Create, Filter, Group, Nillable, Sort, Update</p> <p><b>Description</b> Indicates when to send notifications about the upload. Picklist values are:</p> <p><b>Always</b> Always send notifications.</p> <p><b>Never</b> Never send notifications.</p> <p><b>Failures</b> Send notifications if the upload process failed.</p> <p><b>Warnings</b> Send notifications if warnings occurred during the upload.</p> <p><b>Success</b> Send notifications if the upload is successful.</p> <p><b>SuccessFailures</b> Send notifications if the upload is successful or if the process failed.</p> <p><b>WarningsFailures</b> Send notifications if warnings or failures occurred during the upload.</p> <p><b>WarningsSuccess</b> Send notifications if the upload is successful or if warnings occurred during the upload.</p> <p>The default value is 'WarningsFailures'.</p>
Operation	<p><b>Type</b> Picklist</p> <p><b>Properties</b> Create, Filter, Group, Sort, Update</p>

Field	Details
	<p><b>Description</b></p> <p>Indicates which operation to use when you're loading data into the dataset. Picklist values are:</p> <p><b>Append</b></p> <p>Append all data to the dataset. Creates a dataset if it doesn't exist.</p> <p> <b>Note:</b> If the dataset or rows contain a unique identifier, the append operation isn't allowed.</p> <p><b>Delete</b></p> <p>Delete the rows from the dataset. The rows to delete must contain one (and only one) field with a unique identifier.</p> <p><b>Overwrite</b></p> <p>Create a dataset with the given data, and replace the dataset if it exists.</p> <p><b>Upsert</b></p> <p>Insert or update rows in the dataset. Creates a dataset if it doesn't exist. The rows to upsert must contain one (and only one) field with a unique identifier. For more information about unique identifiers, see <i>isUniqueld</i> in the <a href="#">CRM Analytics External Data Format Developer Guide</a>.</p> <p> <b>Note:</b> A metadata JSON file is required for the append, upsert, and delete operations. The data and metadata for the append and upsert operations must match the dataset on which the operation is happening. (All columns, dimensions, and measures must match exactly.) The metadata for the delete operation must be a subset of the dataset columns.</p>
Status	<p><b>Type</b></p> <p>Picklist</p> <p><b>Properties</b></p> <p>Create, Filter, Group, Sort, Update</p> <p><b>Description</b></p> <p>The status of this data upload. The initial value is null. Picklist values are:</p> <p><b>Completed</b></p> <p>The data upload job was completed successfully. Data parts are retained for 7 days after completion.</p> <p><b>CompletedWithWarnings</b></p> <p>The data upload job completed, but contains warnings. Data parts are retained for 7 days after completion.</p> <p><b>Failed</b></p> <p>The data upload job failed. Data parts are retained for 7 days after failure.</p> <p><b>InProgress</b></p> <p>The data upload job is in progress.</p> <p><b>New</b></p> <p>The data upload job has been created.</p>

Field	Details
	<p><b>NotProcessed</b> The data upload job was aborted on user request. Data parts have been removed.</p> <p><b>Queued</b> The data upload job has been scheduled. This system field isn't editable. The default value is <code>New</code>.</p>
StatusMessage	<p><b>Type</b> String</p> <p><b>Properties</b> Create, Nillable, Update</p> <p><b>Description</b> The reason for the file upload failed or has warnings. This system field isn't editable.</p>
SubmittedDate	<p><b>Type</b> String</p> <p><b>Properties</b> Create, Filter, Nillable, Sort, Update</p> <p><b>Description</b> The time when the upload was submitted or set to Process. This system field isn't editable.</p>
Target	<p><b>Type</b> picklist</p> <p><b>Properties</b> Create, Filter, Group, Nillable, Restricted picklist, Sort, Update</p> <p><b>Description</b> The target for the external data. Valid values are <code>Dataset</code>.</p>

## The InsightsExternalDataPart Object

The InsightsExternalDataPart object enables you to upload an external data file that has been split into parts.

The InsightsExternalDataPart object works with the InsightsExternalData object. After you insert a row into the InsightsExternalData object, you can create part objects to split up your data into parts. If your initial data file is larger than 10 MB, split your file into parts that are smaller than 10 MB.



**Note:** The standard system fields (`CreatedById`, `CreatedDate`, `LastModifiedById`, `LastModifiedDate`, and `SystemModstamp`) are documented in [System Fields](#) in the *Salesforce Object Reference*.

The InsightsExternalDataPart object is available in API version 31 and later.

## Supported Calls

`create()`, `delete()`, `describeObjects()`, `query()`, `retrieve()`, `update()`, `upsert()`

## Fields

Field	Details
CompressedDataLength	<p><b>Type</b> Int</p> <p><b>Properties</b> Create, Filter, Group, Nillable, Sort, Update</p> <p><b>Description</b> The length of the compressed data. This field is overwritten when data is uploaded.</p>
DataFile	<p><b>Type</b> Blob (Base64-encoded string)</p> <p><b>Properties</b> Create, Nillable, Update</p> <p><b>Description</b> The data bytes. Parts are required to be smaller than 10 MB. For data greater than 10 MB, compress the file and then split it into parts. Only the gzip format is supported.</p>
DataLength	<p><b>Type</b> Int</p> <p><b>Properties</b> Create, Filter, Group, Nillable, Sort, Update</p> <p><b>Description</b> The length of the data. This field is overwritten when data is uploaded.</p>
InsightsExternalDataId	<p><b>Type</b> String</p> <p><b>Properties</b> Create, Filter, Group, Sort</p> <p><b>Description</b> The ID of the InsightsExternalData object that this part belongs to.</p>
PartNumber	<p><b>Type</b> Int</p> <p><b>Properties</b> Create, Filter, Group, Sort, Update</p> <p><b>Description</b> The part number. Part numbers are required to be in a contiguous sequence, starting with 1. (For example, 1, 2, 3, etc.)</p>



# RELEASE NOTES

Use the Salesforce Release Notes to learn about the most recent updates and changes to the Analytics External Data API.

For a list of all current developer changes, including Analytics External Data API, see [CRM Analytics](#) in the Salesforce Release Notes.



**Note:** If the Analytics Development section in the Salesforce Release Notes isn't present, there aren't any updates for that release.