Spike in exported records from Salesforce cloud

You might need to look for a sudden, high-volume increase in exported records when doing the following:

- [Protecting a Salesforce cloud deployment](#)

## Prerequisites

In order to execute this procedure in your environment, the following data, services, or apps are required:

- Splunk Enterprise or Splunk Cloud Platform
- CRM, ERP, and other business application data
- Splunk Add-on for Salesforce

## Example

A sudden, high-volume increase in exported records can indicate unauthorized, non-compliant, and potentially malicious behavior. Because so many people in your organization have access to Salesforce, this is an activity you want to monitor for regularly.

To optimize the search shown below, you should specify an index and a time range.

1. Populate the lookup_sfdc_usernames lookup provided by the Salesforce Add-on with live values from your site.
2. Run the following search:

```bash
ROWS_PROCESSED>0 EVENT_TYPE=API OR EVENT_TYPE=BulkAPI OR EVENT_TYPE=RestAPI |
lookup lookup_sfdc_usernames USER_ID |
bucket _time span=1d |
stats sum(ROWS_PROCESSED) AS rows BY _time Username |
stats count AS num_data_samples max(eval(if(_time >= relative_time(maxtime, ",-1d@d"), 'rows',null))) AS rows |
avg(eval(if(_time<relative_time(maxtime,"-1d@d"),'rows',null))) AS avg stdev(eval(if(_time<relative_ |
time(maxtime,"-1d@d"),'rows',null))) AS stdev BY Username |
where 'rows' > upperBound AND num_data_samples >=7
```

## Search explanation

The table provides an explanation of what each part of this search achieves. You can adjust this query based on the specifics of your environment.

---

The information provided in Splunk Lantern is intended for informational and educational purposes only. All information is provided in good faith, however, Splunk disclaims any and all representations and warranties, express and implied, regarding the information provided, including without limitation any warranties and representations regarding the completeness, adequacy or accuracy of the information. You agree to take full responsibility for the results arising from the use of the information provided.
### Splunk Search

- `ROWS_PROCESSED>0` | Filter for event types with at least one processed row.
- `EVENT_TYPE=API OR EVENT_TYPE=BulkAPI OR EVENT_TYPE=RestAPI` | Filter for API-related events.
- `|lookup lookup_sfdc_usernames USER_ID` | Convert the SFDC USER_ID into a friendly username via a lookup.
- `|bucket _time span=1d` | Group events based on _time, effectively flattening the actual _time value to the same day.
- `|stats sum(ROWS_PROCESSED) AS rows BY _time USER_NAME` | Count and aggregate per user, per day.
- `|stats count AS num_data_samples max(eval(if(_time >= relative_time(maxtime, "-1d@d"), 'rows',null))) AS rows` | Calculate the mean, standard deviation, and most recent value.
- `avg(eval(if(_time<relative_time(maxtime,"-1d@d"),'rows',null))) AS avg` | 
- `stdev(eval(if(_time<relative_time(maxtime,"-1d@d"),'rows',null))) AS stdev BY USER_NAME` | 
- `|eval lowerBound=(avg-stdev*2), upperBound=(avg+stdev*2)` | Calculate the bounds as a multiple of the standard deviation.
- `|where 'rows' > upperBound AND num_data_samples >=7` | Display events that have a frequency of occurrence above the calculated upperBound and seven or more data samples.

### Result

While there are no traditional false positives in this search, there will be a lot of noise. Every time this search runs, it will accurately measure a spike in the number of records monitored.
How you handle these alerts depends on where you set the standard deviation. If you set a low standard deviation (2 or 3), you are likely to get a lot of events that are useful only for contextual information. If you set a high standard deviation (6 or 10), the amount of noise can be reduced enough to send an alert directly to analysts.

For most environments, these searches can be run once a day, often overnight, without a lag. If you want to run this search more frequently, or if this search is too slow for your environment, use a summary index that first aggregates the data.

When this search returns values, initiate your incident response process and identify the user demonstrating this behavior. Capture the time of the event, the user's role, and number of records downloaded. If possible, determine the system used to download this data and its location. Contact the user and their manager to determine if the download is authorized, and then document that it was authorized and by whom. If you cannot find authorization, the user credentials may have been used by another party and additional investigation is warranted.

**GDPR Relevance**: Detecting and proving that individuals within the organization are not abusing or misusing legitimate access to assets that store and process personal data is an industry best practice and can be considered an effective security control, as required by Article 32. This is applicable to processing personal data from the controller and needs to also be addressed if contractors or sub-processors from third countries or international organizations access and transfer personal data (Article 15).