



Taloflow's Instruction Manual on Gathering Data for Cost Analysis of Different Cloud Object Providers

Last updated: September 26 2023

Object Storage Assumptions

These instructions will help you provide the key assumptions we need to perform an analysis of Object Storage costs for your use case.

Number of Objects

Please enter your best estimate for the number of objects stored. You can get this information from your cloud provider's console or via script.

AWS

Get the number of objects from the AWS S3 Lens console:

1. Access the [S3 Lens console](#).
1. Click on one of the available (there should be a default) S3 Lens dashboard in the table.
2. Once open, under **Overview**, you will see a total object count.
3. Convert the object count to an integer (e.g.: 850K ➡ 850,000) and enter into the number of objects field in the Taloflow form.

To access S3 Storage Lens dashboards, you must use an IAM user and not a root account. Your AWS administrator must update your IAM permissions to allow for the `s3:ListStorageLensConfigurations`

There are other ways to get the number of S3 objects covered in [this post](#).

Google Cloud Platform

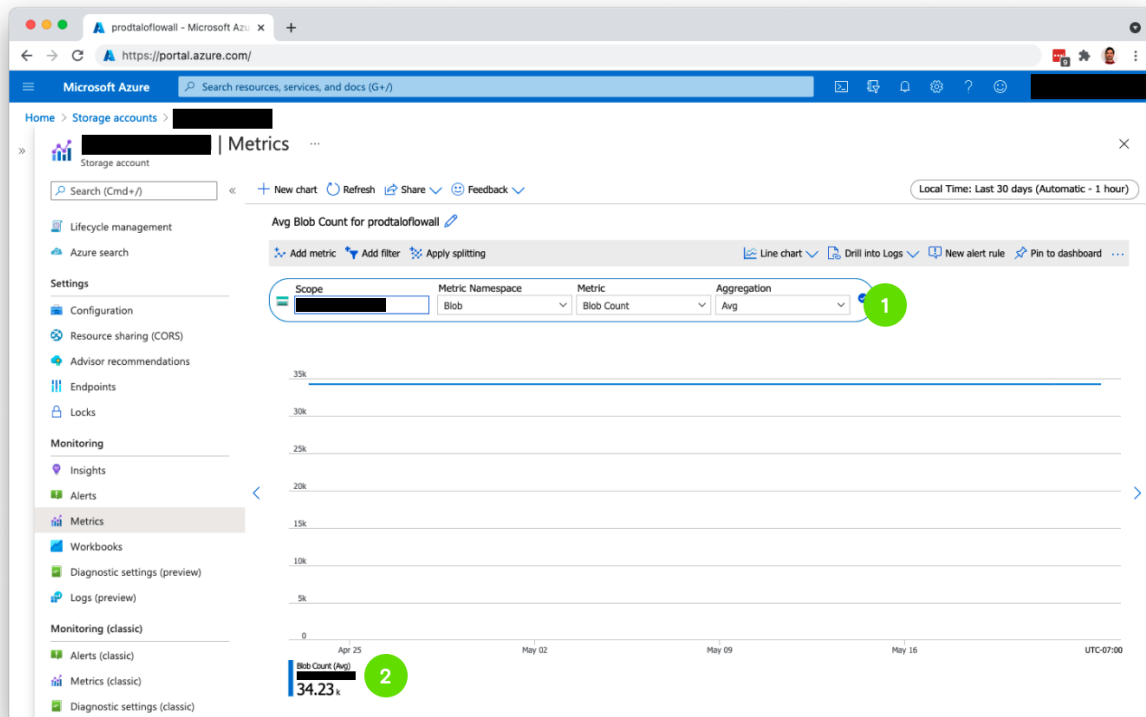
Get the number of objects by running one of the scripts below in your terminal or in Google Cloud Shell:

- `gsutil du | wc -l` will list every object (i.e.: file) in the default project
- `gsutil du -p <PROJECT_ID> | wc -l` will list the objects for a specific project (replace `<PROJECT_ID>` with the relevant project ID).

If you have more than one project to include in the analysis, please run `gsutil du -p <PROJECT_ID> | wc -l` for each project and sum the results.

Microsoft Azure

1. Log into the [Azure Portal](#)
2. Select **Storage Accounts** under the services list
3. Select a storage account
4. In the left panel, under the Monitoring group, click on **Metrics**
5. Set up the **Scope** to the storage account, **Metric Namespace** to `Blob`, and **Metric** to `Blob count`, and Aggregation to `Avg` (See #1 in the screenshot below).
6. The **Blob Count** (i.e.: object count) will be at a ticker at the bottom of the chart (See #2 in the screenshot below).
7. Convert the object count to an integer (e.g.: 850K $\Rightarrow$ 850,000) and enter into the number of objects field in the Taloflow form.



Blob count is equal to the number of blob objects in the storage account.

Azure currently does not allow for multiple storage accounts in a single metric view, so you will need to repeat the steps above for each storage account and total the Blob counts. (Azure is working on this)

Data Transfer Estimates

Taloflow already captures the **INTER**-Regional GB Transfer IN and OUT from the uploaded cost reports. Do not include that data transfer when estimating the **INTRA**-Regional GB Transfer IN and OUT.

Intra-Regional GB Transfer IN (Monthly)

Entering data in this field is optional but may produce a more accurate analysis because the storage migration could result in some additional ongoing transfer costs between services on your cloud provider of origin and the new storage provider.

Please provide the GB monthly volume of data transfer that occurs when other services (like a virtual machine) read or write data to a storage bucket in the same region (i.e.: **intra**-region). This back and forth is not captured in your current usage reports because it is intra-regional and therefore it is free of charge.

One way to estimate the amount of such traffic is to take your average file size in storage and multiply it by the number of times your services likely read or write data to object storage.

Intra-Regional GB Transfer OUT (Monthly)

Please provide your best estimate of the GB monthly volume of data transfer from storage containers or buckets into applications in the same region (a.k.a: intra-region/VPC traffic). Turning on VPC flow logging might help one get a more accurate estimate.

- [VPC Flow Logs on AWS](#)
- [Flow Logging on Microsoft Azure](#)
- [Using VPC Flow Logs on Google Cloud Platform](#)

Cloud Usage History

For some reports, uploading usage history from cloud providers is required. This doc covers how to obtain that information from the major cloud providers, like AWS, Azure, and Google Cloud.

Sharing cloud usage history

Sharing cloud usage history helps us create an accurate representation of the cost-benefits of migrating to a new service (e.g.: AWS S3 ➡ Azure Blob Storage) or adopting a new product (e.g.: forecast APM cost). Once generated, you can download your usage history from a cloud provider and submit to Taloflow via our secure upload service.

Cost and usage reports only contain billing-related information.

AWS

Your AWS usage history is available in AWS Cost and Usage Reports (sometimes abbreviated as CURs).

You must have [AWS Billing and Cost Management Console Access](#) to perform the following steps.

Generate AWS Cost and Usage Reports

To start generating AWS Cost and Usage Reports, please follow these steps:

1. Sign in to the [AWS Billing and Cost Management console](#).
2. On the navigation pane, choose **Cost & Usage Reports**.
3. Click **Create Report**. Give your report a name. For example, `general-cost-report`. Make sure that both `Include resource IDs` and `Data refresh` settings are selected.
4. Under delivery options, select the S3 bucket where your reports currently reside and click **Verify** OR create a new S3 bucket if you do not currently have one for the report. If you get a prompt to add a default bucket policy, accept it.
5. We recommend you create a path prefix in the next field (e.g.: `main`). Make sure that `Hourly`, `GZIP` and `Create New Report Version` are selected. Click **Next**, then click **Review & Complete**.

AWS should generate your first Cost and Usage Reports in 8-24 hours time.

Download AWS Cost and Usage Reports

1. Sign in to the [AWS Billing and Cost Management console](#)

2. On the navigation pane, choose **Cost & Usage Reports**.
3. A list of your Cost and Usage Reports will be available. Click on the relevant one.
4. Click-through the S3 bucket folders until you select the latest report (ideally full month), and download to `csv.gz` file.

If it's available, it's much better to have a full-month report from the last month (not the current/ongoing month) so we can capture more variance. The report path will indicate the period like in this example: `20210501-20210601/`

Google Cloud Platform

To export billing data from Google Cloud Platform (GCP) you have to create a billing report table in BigQuery.

You must have [Billing enabled](#) on your project to perform the following steps.

Generate the Billing Report in BigQuery

1. Set up Cloud Billing data export to BigQuery by following [these instructions](#).
2. Run the following query in [BigQuery](#) after replacing `<PROJECT>`, `<DATASET>`, `<BILLING_ACCOUNT_ID>` and `<INVOICE_MONTH>` with the name of the relevant project, the relevant BigQuery dataset you created in the previous step, the [Billing Account ID](#), and the last month in the following format `yyyyMM` (e.g.: `202105`):

This query will extract the relevant information for **Cloud Storage**.

```
SELECT
billing_account_id, service.id AS service_id,
service.description AS service_description, sku.id AS sku_id,
sku.description AS sku_description, usage_start_time,
usage_end_time, project.id AS project_id, location.location,
location.country, location.region, location.zone, cost, currency,
currency_conversion_rate, usage.amount AS usage, usage.unit,
usage.amount_in_pricing_units AS usage_in_pricing_units,
usage.pricing_unit, TO_JSON_STRING(credits) AS credits_array,
```

```
IFNULL((SELECT SUM(CAST(c.amount * 1000000 as int64)) FROM UNNEST(credits)
c), 0) / 1000000 AS total_credits,

invoice.month, cost_type, adjustment_info.description AS
adjustment_info_description,

adjustment_info.mode AS adjustment_info_mode, adjustment_info.type AS
adjustment_info_type

FROM `<PROJECT>.<DATASET>.gcp_billing_export_v1_<BILLING_ACCOUNT_ID>`

WHERE service.description = "Cloud Storage" AND invoice.month =
'<INVOICE_MONTH>' ORDER BY invoice.month, usage_start_time,
usage_end_time;
```

Download the Billing Report from BigQuery

1. If there are *less than* 16k records from the query, then you can simply click **Save Results**, then select `CSV` (local file) to download the report.
2. If there are *more than* 16k records from the query, then you'll have to perform these additional steps:
 - a. Click on **Query History**
 - b. Click on the query you just ran
 - c. Click on the **Temporary Table** link (Destination Table) and click export to GCS
 - d. Chose `GZIP` compression
 - e. For file location, browse and select a GCS bucket where you want to save the file
 - f. For the file name it would be better to use a wildcard in case more than one file needs be created, so you can put something like `report-*.csv.gz` (Read more about this [here](#))
 - g. When the export finishes you can download the files from your GCS bucket directly.

If there is more than 1 report file, you need to put all the files into a `.zip` or `.tar.gz` file before uploading to Taloflow.

Microsoft Azure

If you are on Microsoft Azure startup credits and can only access your billing data from the Microsoft Azure Sponsorships portal, **you cannot get an analysis performed by Taloflow**. This is because this limited report does not have critical data or report columns necessary for cost analysis, including: `bandwidth`, `chargeType`, `unitOfMeasure` and `productName`.

Microsoft Azure has many different account types and some of these have their own specific ways of accessing the billing report for export. The following two sets of linked instructions work in the vast majority of cases. However, if these steps are not relevant to your account type, please contact us at help@taloflow.ai and we'll provide different instructions.

1. [Download from Cost Management + Billing console](#)
2. [Billing Export Method](#)

If you're using the Billing Export method, please use **Actual Cost** and a **Daily export of last-month costs** (or month-to-date costs if last-month is unavailable).