Ceph/S3 FAQs

FAQs

- **Q1**: What is Shopee S3?
  - Shopee S3 is similar to Amazon S3, but not the same as Amazon S3.
  - We build Shopee S3 using Ceph, and it only implement a *subset of S3 features* compared to AWS S3
  - So in some rare cases, Ceph S3 does not support some features like AWS S3

- **Q2**: Can each team have a dedicated account/space in S3?
  - Yes

- **Q3**: Can S3 be accessed from different IDCs in different countries?
  - Yes. But our S3 is based in SG idc, so you may expect some latency if querying it from other countries
  - By default, all VPN users + Shopee offices can access S3
  - Some cases we need to update our whitelist to allow access to S3 service

- **Q4**: What is the range of file size recommended to store in S3?
  - S3 is suitable for storing and sharing `small files.` Ideally less than `100MB/file`

- **Q5**: But S3 can store large files as well?
  - Yes. But we don't really recommend it
  - You should split your data into many smaller objects, then you can read/write in parallel

- **Q6**: Is there any limitation for the number of files in each space/bucket?
  - For number of files in each bucket
    - `No limitation,` but expect `high latency` if you use LIST command in a bucket with `more than 100k objects`
  - For number of buckets per account
    - `Yes`, the limit is `1000 buckets/account`

- **Q7**: Is S3 open to all teams to apply space and quota?
  - Yes, of course :D

- **Q8**: Should I store sensitive data like user personal info, credit card info on S3?
  - `NO!`
  - Sensitive data like :
    - Personal user info
    - Credit card info
    - Data need to follow Government Compliances/Laws/Standards
    - Other data depend on how you define *sensitive data*
  - Should not be stored on S3, unless you implement the encryption part yourself.
  - But in general, we `don't recommend` to store `sensitive data` like that on S3, `use it at your own risks!`

- **Q9**: Should I use **HTTP** or **HTTPS** when accessing to Shopee S3?
    - Should use **HTTPS**.
    - However, we do support **HTTP** as well

- **Q10**: I just accidentally delete/overwrite my files....do you have any backups or some way to get historical data or version control in S3?
    - **No**. Sadly, we dont have such a thing... If you delete/overwrite your files....then your data are gone...
    - **No**. We dont have version control or anything similar to it. If you want version control, can use GitLab instead

- **Q11**: If our data is large, should we split it to multiple 100MB files?
    - **Yes.** S3 protocol only supports up to 5GB for a **single** large file in a **single** PUT operation without using Multipart Upload API. Moreover, the larger the single object filesize, the slower the PUT operation.
    - Therefore, for large data objects (>100MB), we recommend using S3 Multipart Upload API, to optimize for upload speeds and reduce timeouts. S3 Multipart Upload API allows up to 5TB filesize uploads.
    - More information can be found here https://aws.amazon.com/fr/blogs/aws/amazon-s3-multipart-upload/
    - Some clients such as Minio and AWS SDK supports higher level PUT operations that automatically invoke multipart upload when object filesize is detected to be large.

- **Q12**: I cannot use my minio cli to list bucket with many files, what should I do?
    - Can refer to this guide as workaround: S3 Bucket Listing Issue