

Shopee S3 User Manual, Forms & FAQs

Introduction

Important notes



PICs

Please contact new S3 PIC for all requests/issues

New PIC: xukai@shopee.com (Active PIC from Dec 2021)

Old PICs: duy.nguyen@shopee.com, hongrui.com@shopee.com, ducmanh.le@shopee.com (Stopped supporting S3 products from Dec 2021)

Standard usage

Most user requests will fall to this type:

- Small data size (less than **100GB**)
- Low QPS (less than **500**)
- Small number of files (less than **10,000** files)
-

New application workflow Space CMDB (preferred way): [\[UI\] S3 User Guide 用户指南 - V1.4.2](#)

Special request

For special S3 usage like:

- Huge amount of data (more than 5TB/project)
- Need to expose S3 service to public for special cases
- Special compliance/regulation requirements (sensitive user data, need to comply with Government law)
- High network traffic
- Require High/Very High QPS
- Require Low latency
- Need special whitelisting
-

If you have these kind of **special requirements** for S3, can check this [S3 Special Request Form](#)

If you are still not sure if your data is big enough to be listed as special request, can consul me: [Duy Nguyen](#)

Or open a ticket here: <https://jira.shopee.io/servicedesk/customer/portal/4/create/206>

Which cluster should I use?

There are many clusters to choose from, depend on your requirements

What is Tier?: [S3 Tiering & Usecases](#)

| Domain | Version | Tier | IDC | Can access from | Usage | SLOs | Price | Notes |
|--|----------|------|---------|---|--|------|--|-------|
| ia.s3.sto.shopee.io (Default one) | Nautilus | IA | SG (MH) | <ul style="list-style-type: none">• Shopee IDCs• Shopee Office /VPNs | General purpose storage for any type of data <ul style="list-style-type: none">• Avg Latency• High Capacity | TBD | Reference price: \$0.18/GB (The prices here are al for reference, we wont charge you anything) | |

| | | | | | | | | |
|------------------------------|----------|---------|------------|--|--|-----|-----------------------------------|--|
| s3. shopeemobile. com | Luminous | Fast | SG (MH) | <ul style="list-style-type: none"> Shopee IDCs Shopee Office /VPNs | Reserved for high demand applications which requires <ul style="list-style-type: none"> Very low latency Low capacity | TBD | Reference price: \$1.2/GB | |
| smap.s3.sto. shopee.io | Nautilus | IA | SG (MH) | <ul style="list-style-type: none"> Shopee IDCs | Dedicated cluster for SMAP team | TBD | Reference price: \$0.18/GB | |
| deep.s3.sto. shopee.io | Nautilus | IA | SG (MH) | <ul style="list-style-type: none"> Shopee IDCs | Dedicated cluster for DEEP team | TBD | Reference price: \$0.18/GB | |
| di.s3.sto.shopee. io | Nautilus | IA | SG (MH) | <ul style="list-style-type: none"> Shopee IDCs | Dedicated cluster for DI/DE team | TBD | Reference price: \$0.18/GB | |
| archive.s3.sto. shopee.io | Nautilus | Archive | SG2 | <ul style="list-style-type: none"> Shopee IDCs Shopee Office /VPNs | For storing long-term archive (3++ years) For off-site backup/archive, in case our main clusters in MH are all down... | TBD | Reference price: \$0.12/GB | |
| | | | | | | | | |

What are the **download** and **upload** speed of each clusters?

Last updated: 9 June 2021

| Shared Clusters | Download speed (same IDC) / MBps | Download 1 GB in / Seconds (3s.f) | Download speed (cross IDC) / MBps | Notes |
|---------------------------|----------------------------------|-------------------------------------|-----------------------------------|-------|
| s3hot | 700 | 1.46 | TBD | |
| s3ia | 300 | 3.41 | TBD | |
| s3archive | 300 | 3.41 | TBD | |
| Dedicated Clusters | | | | |
| s3smap | 350 | 2.93 | TBD | |
| s3deep | 340 | 3.01 | TBD | |
| s3di | 320 | 3.20 | TBD | |

Reference: [S3 User Upload/Download Speeds](#)

Last updated: 9 June 2021

| Shared Clusters | Upload speed (same IDC) / MBps | Download 1 GB in / Seconds (3s.f) | Upload speed (cross IDC) / MBps | Notes |
|---------------------------|--------------------------------|-------------------------------------|---------------------------------|-------|
| s3hot | 560 | 1.83 | TBD | |
| s3ia | 670 | 1.53 | TBD | |
| s3archive | 330 | 3.10 | TBD | |
| Dedicated Clusters | | | | |
| s3smap | 600 | 1.55 | TBD | |
| s3deep | 630 | 1.63 | TBD | |
| s3di | 600 | 1.55 | TBD | |

Reference: [S3 User Upload/Download Speeds](#)

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
 - 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](#)

Introduction

Standard usage

Most user requests will fall to this type:

- Small data size (less than **100GB**)
- Low QPS (less than **500**)
- Small number of files (less than **10,000** files)
-

New application workflow Space CMDB (preferred way): [\[UI\] S3 User Guide 用户指南 - V1.4.2](#)

Special request

For special S3 usage like:

- Huge amount of data (more than 5TB/project)
- Need to expose S3 service to public for special cases
- Special compliance/regulation requirements (sensitive user data, need to comply with Government law)
- High network traffic
- Require High/Very High QPS
- Require Low latency
- Need special whitelisting
-

If you have these kind of **special requirements** for S3, can check this [S3 Special Request Form](#)

If you are still not sure if your data is big enough to be listed as special request, can consul me: [Duy Nguyen](#)

Or open a ticket here: <https://jira.shopee.io/servicedesk/customer/portal/4/create/206>

Which cluster should I use?

There are many clusters to choose from, depend on your requirements

What is Tier?: [S3 Tiering & Usecases](#)

| Domain | Version | Tier | IDC | Can access from | Usage | SL0s | Price | Notes |
|--|--------------------------|------|---------|--|---|------|--|-------|
| ia.s3.sto.shopee.io (Default one) | Nautilus | IA | SG (MH) | <ul style="list-style-type: none"> • Shopee IDCs • Shopee Office /VPNs | General purpose storage for any type of data <ul style="list-style-type: none"> • Avg Latency • High Capacity | TBD | Reference price: \$0.18/GB (The prices here are al for reference, we wont charge you anything) | |
| s3.shopeemobile.com | Luminous | Fast | SG (MH) | <ul style="list-style-type: none"> • Shopee IDCs • Shopee Office /VPNs | Reserved for high demand applications w hich requires <ul style="list-style-type: none"> • Very low latency • Low capacity | TBD | Reference price: \$1.2/GB | |

| | | | | | | | | |
|--------------------------|----------|---------|---------|--|---|-----|-----------------------------------|--|
| smap.s3.sto.shopee.io | Nautilus | IA | SG (MH) | <ul style="list-style-type: none"> Shopee IDCs | Dedicated cluster for SMAP team | TBD | Reference price: \$0.18/GB | |
| deep.s3.sto.shopee.io | Nautilus | IA | SG (MH) | <ul style="list-style-type: none"> Shopee IDCs | Dedicated cluster for DEEP team | TBD | Reference price: \$0.18/GB | |
| di.s3.sto.shopee.io | Nautilus | IA | SG (MH) | <ul style="list-style-type: none"> Shopee IDCs | Dedicated cluster for DI/DE team | TBD | Reference price: \$0.18/GB | |
| archive.s3.sto.shopee.io | Nautilus | Archive | SG2 | <ul style="list-style-type: none"> Shopee IDCs Shopee Office /VPNs | For storing long-term archive (3++ years) For off-site backup/archive, in case our main clusters in MH are all down... | TBD | Reference price: \$0.12/GB | |
| | | | | | | | | |

What are the **download** and **upload** speed of each clusters?

Last updated: 9 June 2021

| Shared Clusters | Download speed (same IDC) / MBps | Download 1 GB in / Seconds (3s.f) | Download speed (cross IDC) / MBps | Notes |
|---------------------------|----------------------------------|-------------------------------------|-----------------------------------|-------|
| s3hot | 700 | 1.46 | TBD | |
| s3ia | 300 | 3.41 | TBD | |
| s3archive | 300 | 3.41 | TBD | |
| Dedicated Clusters | | | | |
| s3smap | 350 | 2.93 | TBD | |
| s3deep | 340 | 3.01 | TBD | |
| s3di | 320 | 3.20 | TBD | |

Reference: [S3 User Upload/Download Speeds](#)

Last updated: 9 June 2021

| Shared Clusters | Upload speed (same IDC) / MBps | Download 1 GB in / Seconds (3s.f) | Upload speed (cross IDC) / MBps | Notes |
|---------------------------|--------------------------------|-------------------------------------|---------------------------------|-------|
| s3hot | 560 | 1.83 | TBD | |
| s3ia | 670 | 1.53 | TBD | |
| s3archive | 330 | 3.10 | TBD | |
| Dedicated Clusters | | | | |
| s3smap | 600 | 1.55 | TBD | |
| s3deep | 630 | 1.63 | TBD | |
| s3di | 600 | 1.55 | TBD | |

Reference: [S3 User Upload/Download Speeds](#)

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
 - 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 don't have such a thing... If you delete/overwrite your files....then your data are gone...
 - **No**. We don't 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](#)