

云存储应用技术

第七章：对象存储

丁烨

dingye@dgut.edu.cn

网络空间安全学院

2019-11-14



東莞理工學院
DONGGUAN UNIVERSITY OF TECHNOLOGY

对象存储概述

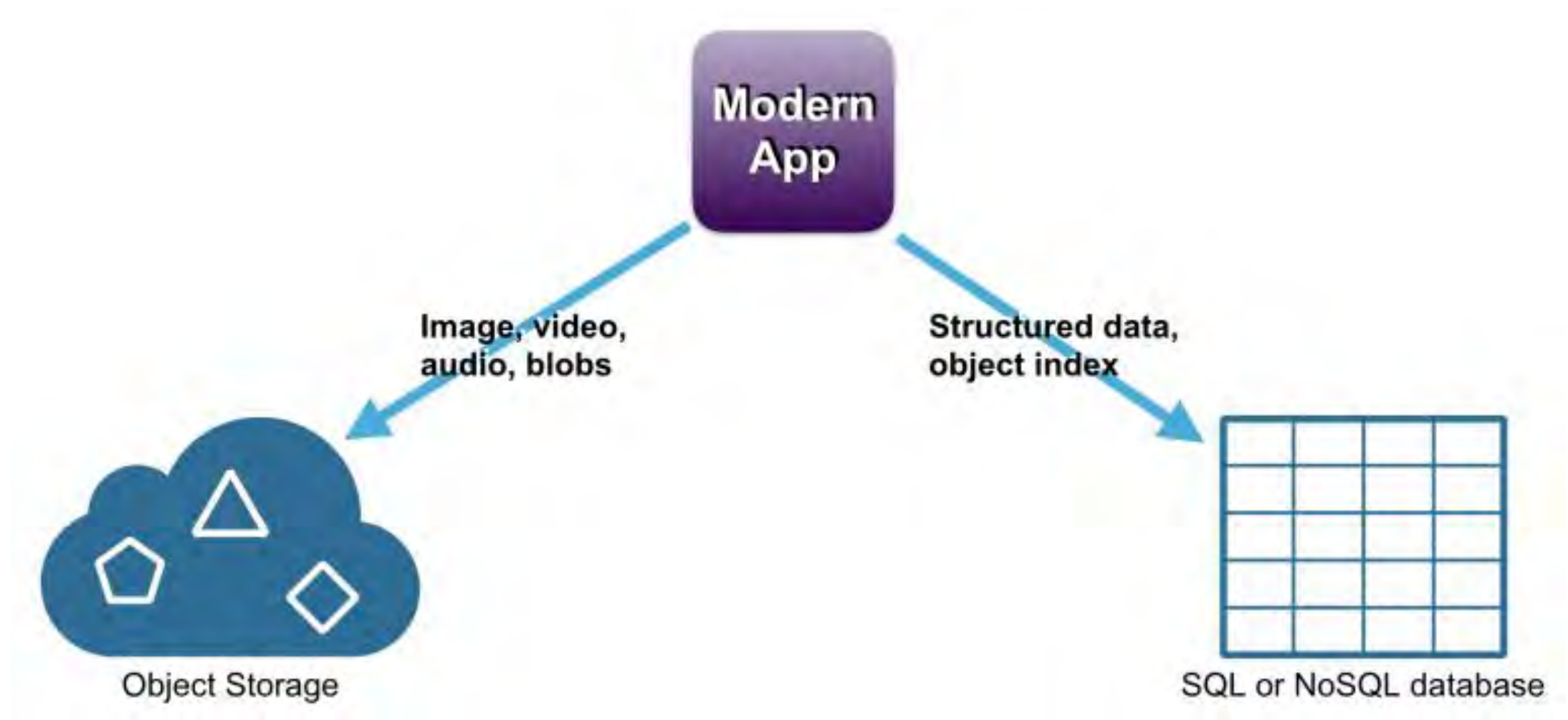
MinIO

OpenIO

Cloudian

其他对象存储平台

- ❖ 互联网应用一般会产生两种不同类型的数据：
- ❖ 结构化数据（Structured Data）
- ❖ 通常是指高度结构化和高度整洁的数据，适合存放于关系型数据库中
- ❖ 非结构化数据（Unstructured Data）
- ❖ 一般是指关系型数据库不宜处理的数据，例如多媒体文件等





- ❖ 对象存储（Object Storage）
- ❖ 一种计算机存储结构，与块存储、文件存储、以及数据库为同一级别
- ❖ 不同于其他的存储结构，对象存储将数据作为“对象（Object）”来处理
- ❖ 一个对象通常包含一份非结构化数据（Data）、对应的元数据（Metadata）、以及该对象的全局独特标识符（Global Unique ID）
- ❖ 不同于其他存储结构，对象存储通常深度依赖于云存储系统

对象存储概述

对象存储的特点

- ❖ 不像文件系统中的文件，对象存储在扁平结构中
- ❖ 对象存储只有对象池，没有文件夹，没有目录，也没有层次体系
- ❖ 你只要提供对象 ID，就可以请求某个对象

- ❖ 对象存储提供了极高的灵活性，因为对象元数据是任意的
- ❖ 元数据并不仅限于存储系统认为很重要的对象（不同于文件系统中的固定元数据）
- ❖ 也可以手动添加任何类型或任何数量的元数据
- ❖ 例如，对象存储可以指定与对象关联起来的应用程序的类型，指定应用程序的重要性，指定赋予对象的数据保护级别，指定是否允许将该对象复制到另一个站点或多个站点，指定何时删除该对象等

对象存储概述

对象存储的特点

- ❖ 类似于分布式云存储对于 POSIX 的文件存储要求，对象存储需要对客户端透明化
- ❖ “你只要提供对象 ID，就可以请求某个对象”
- ❖ 而无需关注对象存储内部复杂的同步和文件处理机制

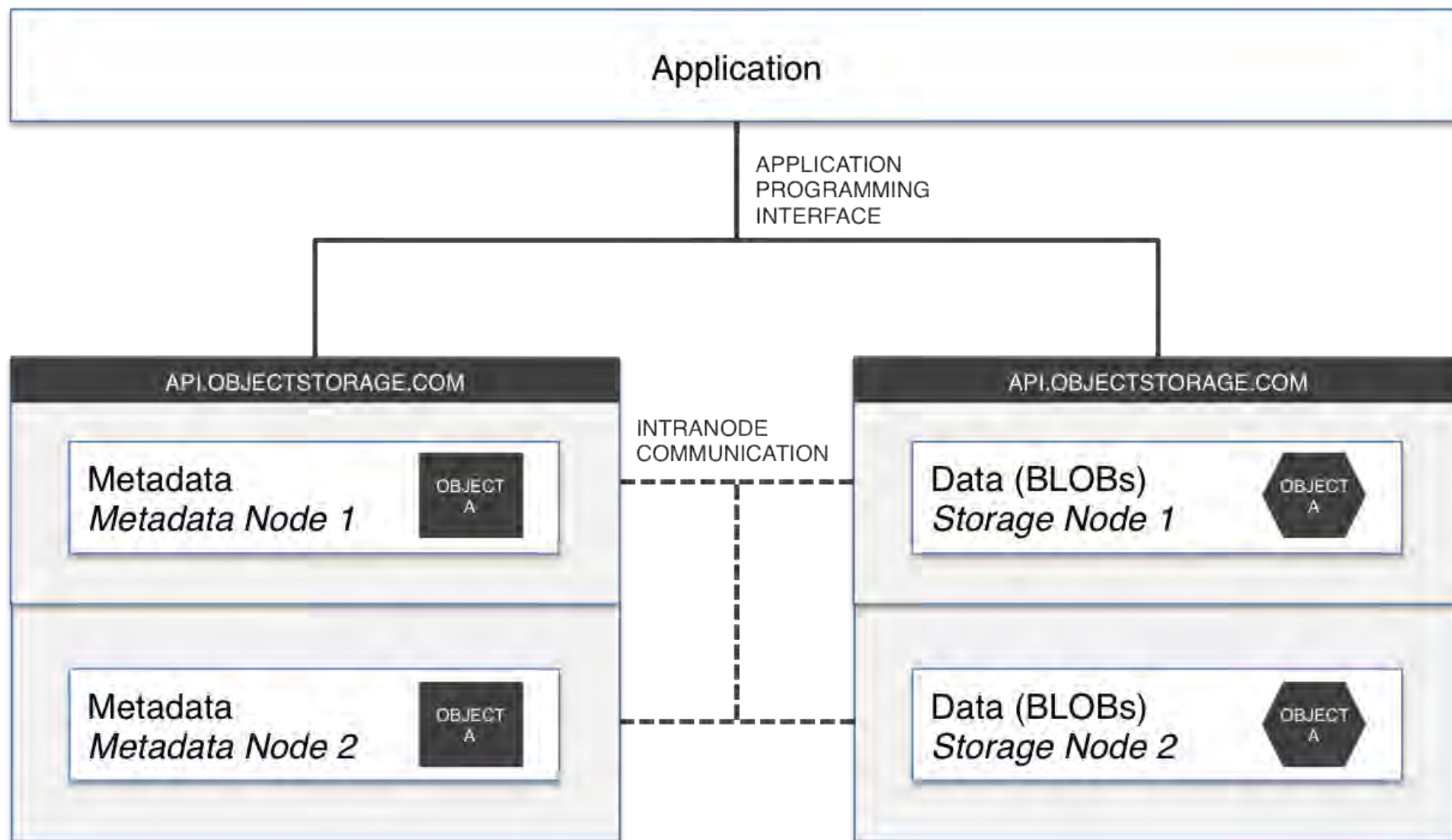
对象存储概述

对象存储的特点



对象存储概述

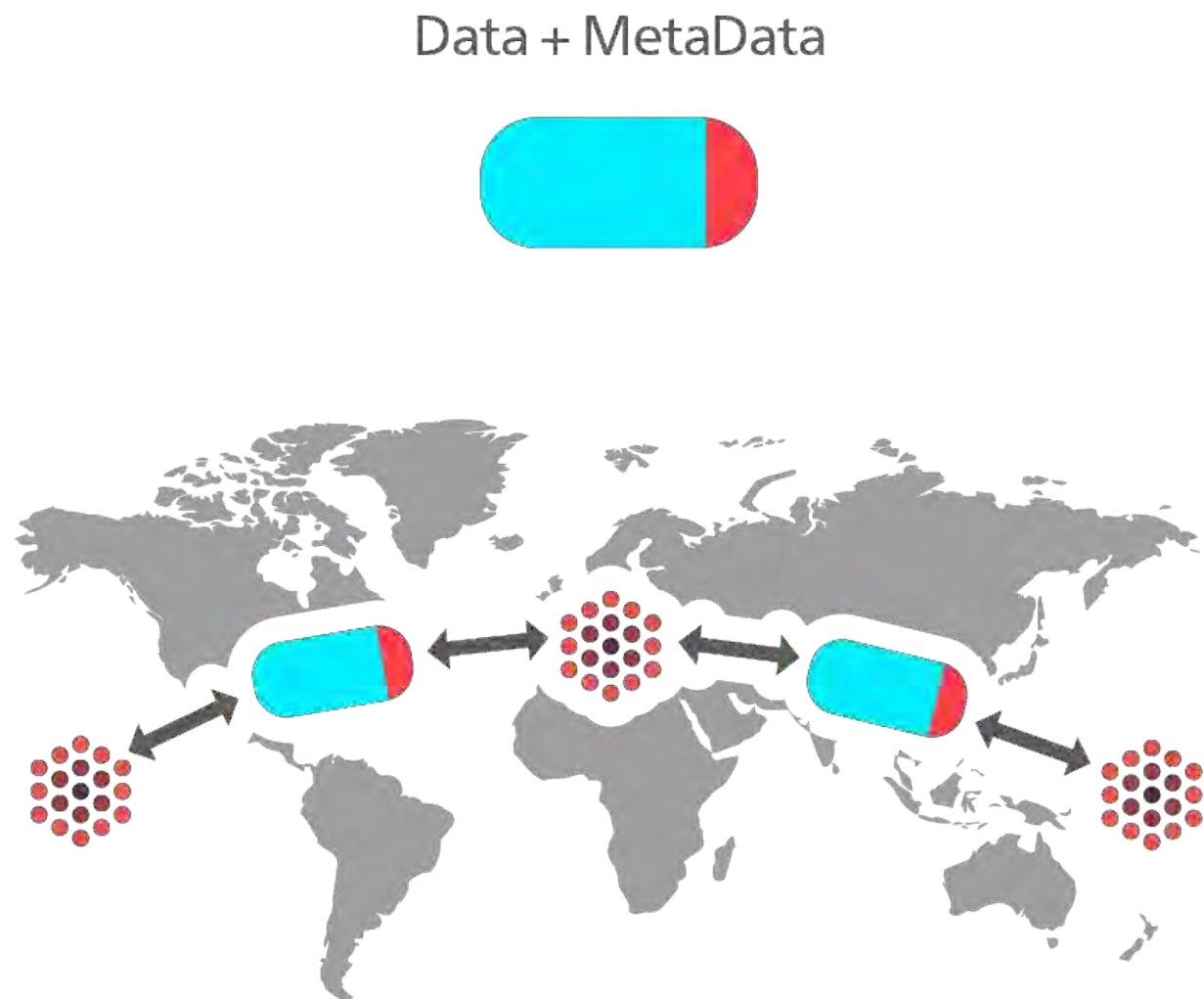
对象存储的特点



对象存储概述

对象存储的特点

- ❖ 对象存储可以部署在复杂的云存储网络中
- ❖ 对象可以是本地的，也可以是远在千里之外的云服务器上
- ❖ 但由于它们是在扁平的地址空间，检索方式一模一样
- ❖ 对象存储系统内部需要实现对象的存储和同步



- ❖ 易于使用 (Ease of Use)
- ❖ 只要提供对象 ID, 就可以请求某个对象
- ❖ 可扩展性 (Scalability)
- ❖ 上传一个对象, 就可以获取对象 ID, 系统会自动处理存储、同步等诸多问题
- ❖ 敏捷性 (Agility)
- ❖ 无需使用复杂的 POSIX 结构, 只需使用对象 ID 获取对象

对象存储概述

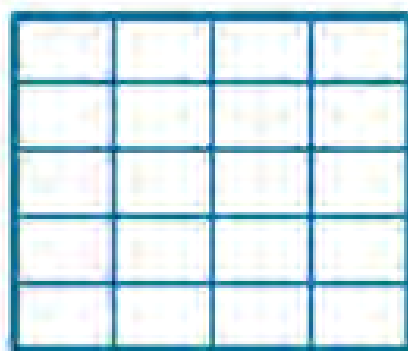
对象存储的特点



File System

C:\folder\music.m4a

sysadmin required for
integrity and scale



Database / Structured Data

```
SELECT * FROM table;  
INSERT INTO table;
```

sysadmin and DBA required
for scale, integrity and
performance



Object Storage

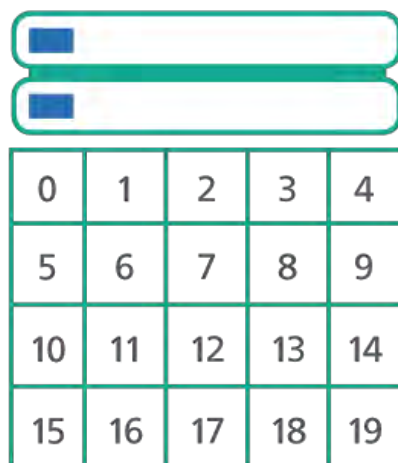
```
GET /object/Kbg1Bn7qepo  
PUT /object/Kbg1Bn7qepo
```

sysadmin not required

对象存储概述

对象存储的特点

Block



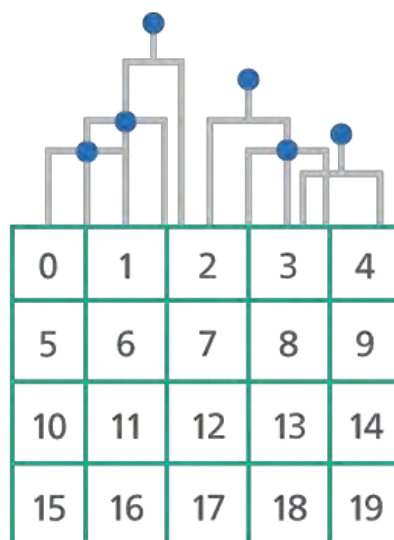
Client Via OS

Fixed Sys Attributes

Transactional Data

Performance

File



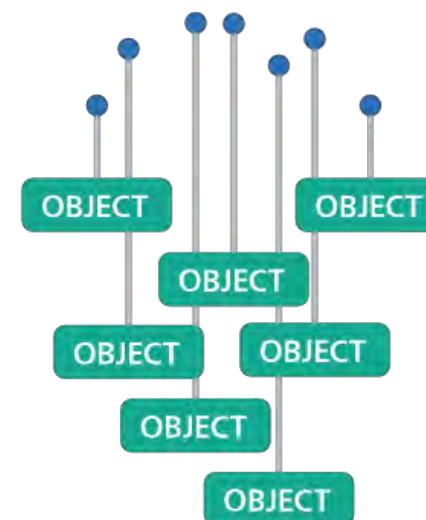
Client Via OS

Fixed Sys Attributes

Shared Changing File

Access, Single Site

Object



Client is App

Custom Metadata

Shared Semi-Static File

Scalable, Multi-Site

- ❖ 对象存储非常适合互联网 App 的开发
- ❖ 对象存储不适合大规模数据分析平台，尤其是海量小数据文件

- ❖ MinIO
 - ❖ OpenIO
 - ❖ Cloudian
 - ❖ OpenStack Swift
- ❖ 大部分主流云计算供应商（如阿里云、AWS、Google Cloud 等）都提供对象存储服务



对象存储概述

MinIO

OpenIO

Cloudian

其他对象存储平台

- ❖ MinIO
- ❖ <https://min.io/>
- ❖ 开源免费，源代码：<https://github.com/minio/minio>
- ❖ MinIO 是一个使用 Go 实现的对象存储系统
- ❖ 作为一套对象存储系统，MinIO 适合保存各类非结构化数据，例如照片、视频、历史记录（log）、数据备份、以及 Docker Container Image 等
- ❖ MinIO 支持单对象最大 5TB



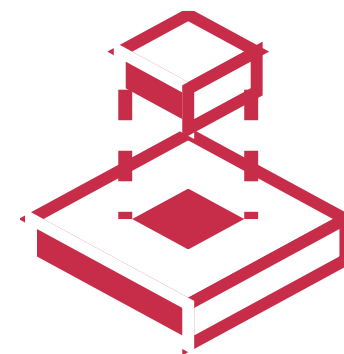
[Product](#) ▾[Docs](#)[Partners](#)[Subscription](#)[Resources](#)[Blog](#)[Download](#)

High Performance Object Storage

Build high performance data infrastructure for machine learning, analytics and application data workloads with MinIO

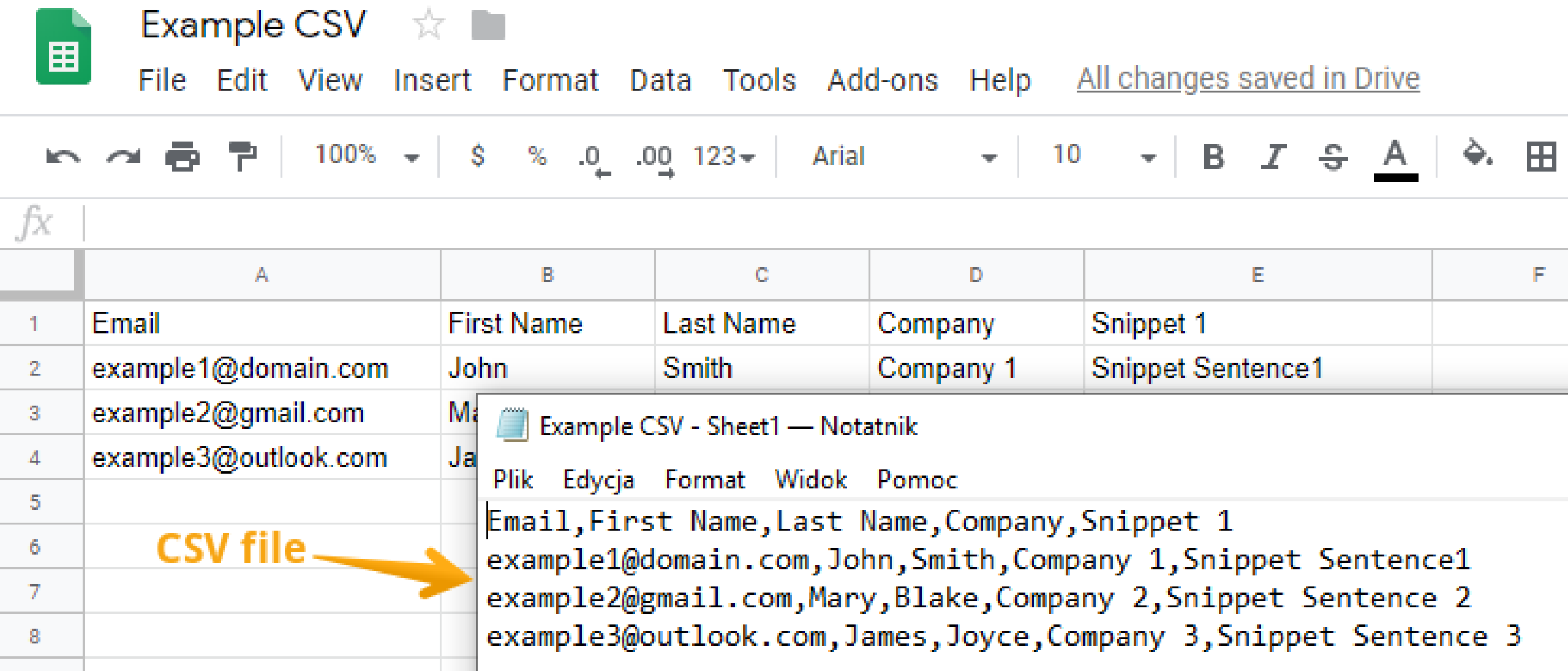


- ❖ 兼容 SQL Select
- ❖ MinIO 高度兼容 Amazon S3, 可以使用类似 Amazon S3 的数据接口
- ❖ Spark、Presto、Hive、Flink 等数据处理平台可以直接使用 SQL Select 访问 MinIO 对象, 而无需下载整个对象到本地
- ❖ MinIO 兼容 Parquet、JSON、CSV 等常见数据格式, 亦支持压缩或加密对象格式



- ❖ 逗号分隔值 (Comma-Separated Values, CSV)
- ❖ CSV 文件以纯文本形式存储表格数据 (数字和文本)
- ❖ 文件由任意数目的记录组成, 记录间以某种换行符分隔
- ❖ 每条记录由字段组成, 字段间的分隔符是其它字符或字符串, 最常见的是逗号或制表符
- ❖ 通常, 所有记录都有完全相同的字段序列

- ❖ CSV 有时也称为字符分隔值, 因为分隔字符也可以不是逗号
- ❖ CSV 列表方式被定义在 FORTRAN 77 (77 代表 1977 年) 中



Example CSV

File Edit View Insert Format Data Tools Add-ons Help [All changes saved in Drive](#)

100% \$ % .0 .00 123 Arial 10 B I S A

	A	B	C	D	E	F
1	Email	First Name	Last Name	Company	Snippet 1	
2	example1@domain.com	John	Smith	Company 1	Snippet Sentence1	
3	example2@gmail.com	Mary	Blake	Company 2	Snippet Sentence 2	
4	example3@outlook.com	James	Joyce	Company 3	Snippet Sentence 3	
5						
6						
7						
8						

CSV file →

Example CSV - Sheet1 — Notatnik

Plik Edycja Format Widok Pomoc

```
Email,First Name,Last Name,Company,Snippet 1
example1@domain.com,John,Smith,Company 1,Snippet Sentence1
example2@gmail.com,Mary,Blake,Company 2,Snippet Sentence 2
example3@outlook.com,James,Joyce,Company 3,Snippet Sentence 3
```

- ❖ CSV 是一种行存储（Row Store）模式
- ❖ 传统的数据库延续了这种存储模式，即行存储
- ❖ 行存储不利于压缩数据，因为一行中的字段出现重复的可能性较小
- ❖ 不同行出现重复的可能性也较小（甚至不允许重复）

- ❖ 当数据量极大，需要使用云存储的时候，不能压缩数据将带来较高的数据存储成本

❖ 列存储 (Column Store)

RowId	EmpId	Lastname	Firstname	Salary
001	10	Smith	Joe	40000
002	12	Jones	Mary	50000
003	11	Johnson	Cathy	44000
004	22	Jones	Bob	55000

❖ **行存储 (Row Store)** 把一行中的数据值串在一起存储起来, 然后再存储下一行的数据, 以此类推:

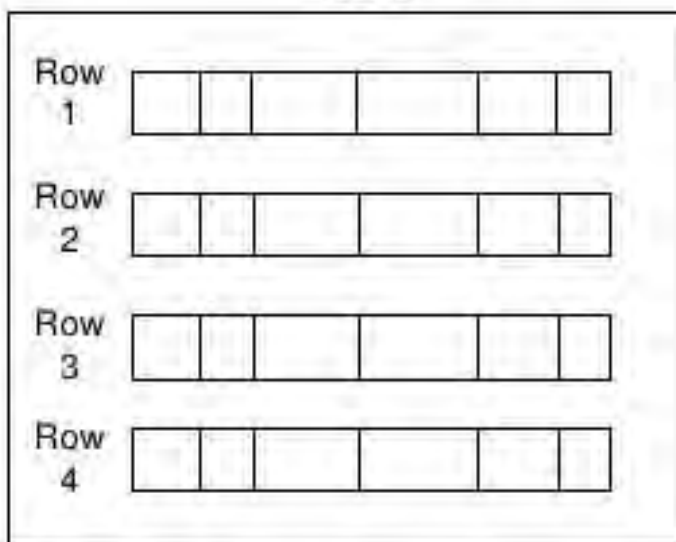
- ❖ 001,10,Smith,Joe,40000
- ❖ 002,12,Jones,Mary,50000
- ❖ 003,11,Johnson,Cathy,44000
- ❖ 004,22,Jones,Bob,55000

❖ **列存储 (Column Store)** 把一列中的数据值串在一起存储起来, 然后再存储下一列的数据, 以此类推:

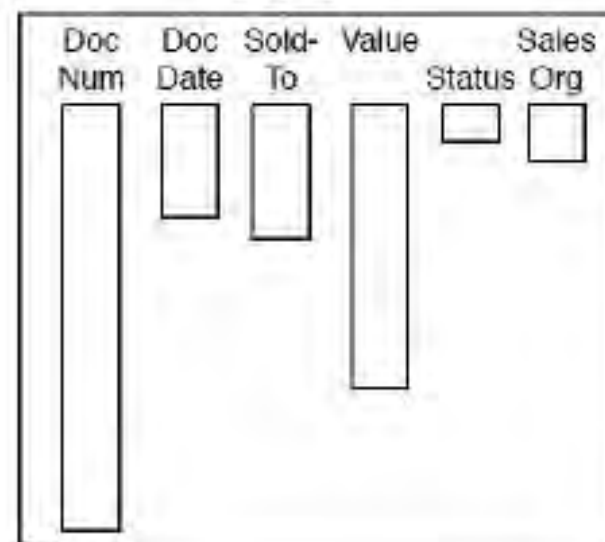
- ❖ 001,002,003,004
- ❖ 10,12,11,22
- ❖ Smith,Jones,Johnson,Jones
- ❖ Joe,Mary,Cathy,Bob
- ❖ 40000,50000,44000,55000

Document Number	Document Date	Sold-To Party	Order Value	Status	Sales Organization	...
95769214	2009-10-01	584	10.24	CLOSED	Germany Frankfurt	...
95769215	2009-10-01	1215	124.35	CLOSED	Germany Berlin	...
95779216	2009-10-21	584	47.11	OPEN	Germany Berlin	...
95779217	2009-10-21	454	21.20	OPEN	Germany Frankfurt	...

Row Store



Column Store



❖ Apache Parquet

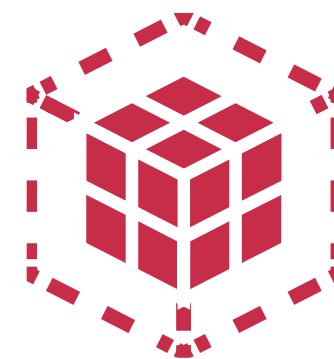
❖ <http://parquet.apache.org/>



- ❖ Apache Parquet 是一个开源免费的列存储格式
- ❖ Apache Parquet 属于 Hadoop 生态链中的一员
- ❖ Apache Parquet 使用了一个高效的列存储压缩算法
- ❖ 不同列可以采用不同的压缩方法
- ❖ 列存储压缩可以指定该列的数据格式，并针对数据格式采用特定的处理方法
- ❖ 查询（Query）一列中的某一格数据无需遍历整列

❖ 数据安全

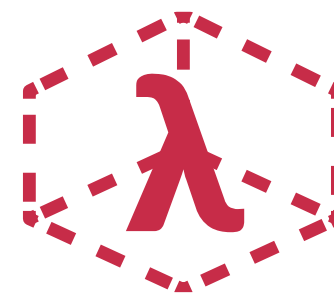
- ❖ MinIO 提供了高效的数据加密算法，确保了数据的：
- ❖ 保密性（Confidentiality）、一致性（Integrity）、真实性（Authenticity）
- ❖ 加密过程安全有效，且只需在原数据中添加很少的额外信息



- ❖ 分布式聚合
- ❖ MinIO 分布式聚合机制允许 MinIO 分布式运行在多个服务器中
- ❖ 同时提供一致的对象存储服务



- ❖ 支持 AWS Lambda 高响应消息队列
- ❖ MinIO 可以通过 AWS Lambda 触发消息通知
- ❖ 兼容 AWS SNS/SQS 事件系统
- ❖ 兼容 Kafka、NATS、AMQP、MQTT 等消息队列
- ❖ 支持 Webhooks 事件响应



- ❖ 用户系统及访问控制
- ❖ MinIO 支持多种先进的用户系统，包括：
- ❖ WSO2、Keycloak、Okta、Ping Identity、Active Directory 等



- ❖ 容灾系统
- ❖ MinIO 的分布式机制允许丢失最多一半的服务器或一半的物理存储，而仍可提供服务
- ❖ 容灾过程通过 x64 和 ARM CPU 的 SIMD 进行加速





18.8K+

GITHUB STARS

277.8M+

DOCKER PULLS

5939

SLACK MEMBERS

466

CONTRIBUTORS

- ❖ Ubuntu 18.04
- ❖ `wget https://dl.min.io/server/minio/release/linux-amd64/minio`
- ❖ 或
- ❖ `wget https://dl.min.io/server/minio/release/linux-ppc64le/minio`

- ❖ `chmod +x minio`
- ❖ `./minio server /data`

❖ macOS Catalina

❖ `brew install minio/stable/minio`

❖ `minio server /data`

❖ 或

❖ `wget https://dl.min.io/server/minio/release/darwin-amd64/minio`

❖ `chmod +x minio`

❖ `./minio server /data`

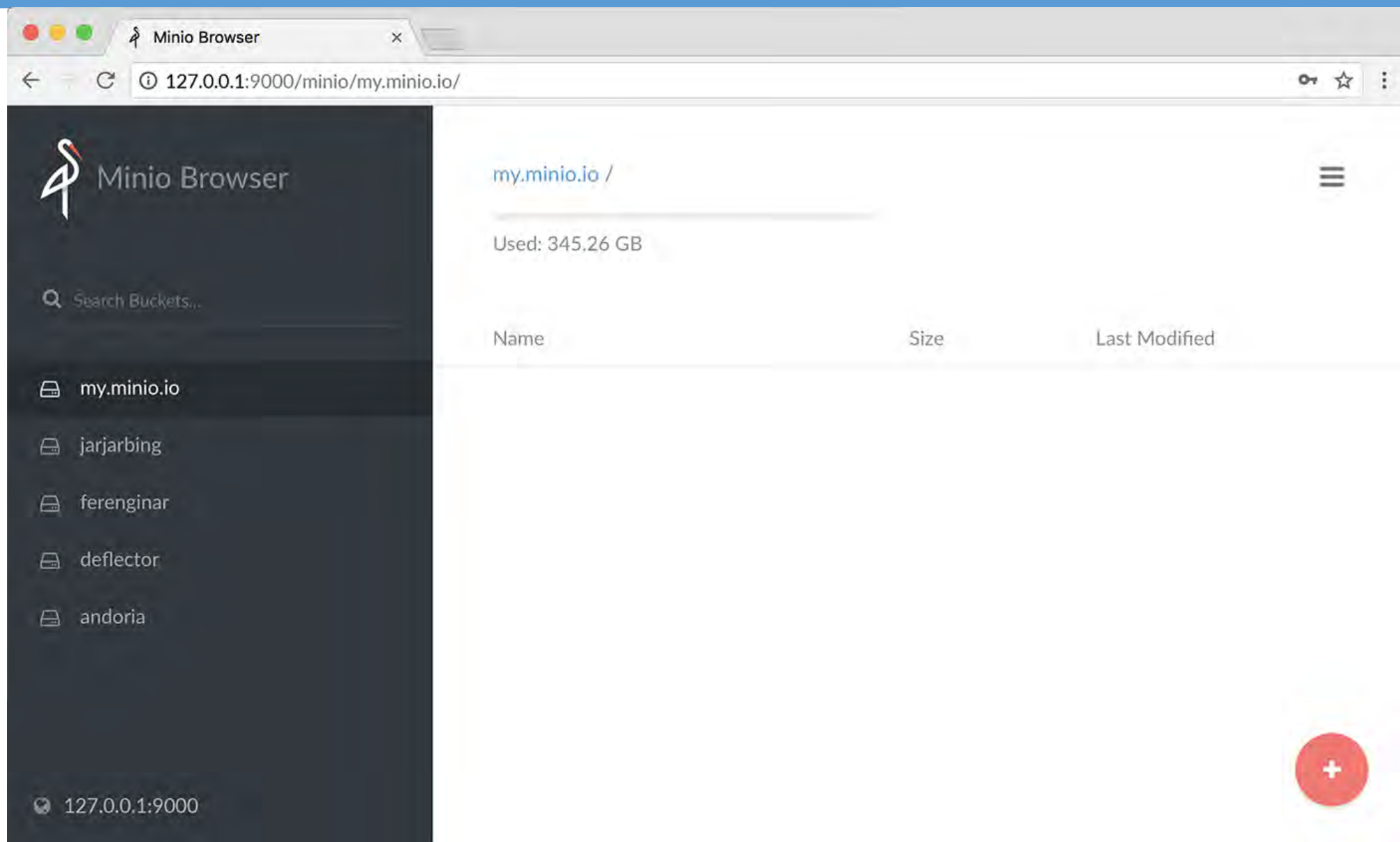
❖ Windows 10

❖ 下载: <https://dl.min.io/server/minio/release/windows-amd64/minio.exe>

❖ `minio.exe server D:\Photos`

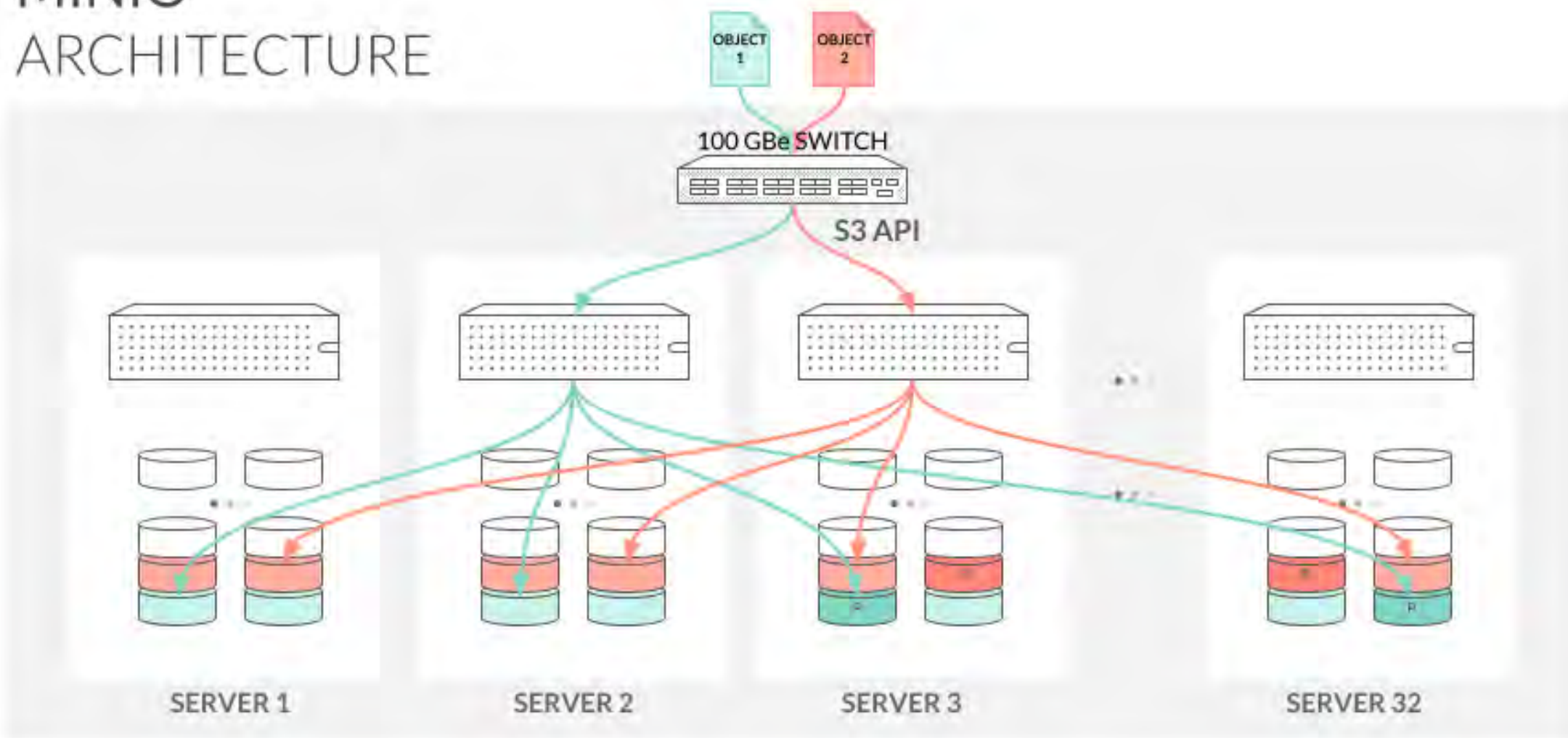
- ❖ MinIO 也提供 Docker Container Image, 用于快速搭建隔离的对象存储系统
- ❖ `docker pull minio/minio`
- ❖ `docker run -p 9000:9000 minio/minio server /data`

- ❖ 安装完毕后，可通过 <http://127.0.0.1:9000> 访问 MinIO 的图形界面



- ❖ 分布式 MinIO 的搭建方法
- ❖ <https://docs.min.io/docs/distributed-minio-quickstart-guide.html>

MINIO ARCHITECTURE



❖ API SDK

- ❖ Python: <https://docs.min.io/docs/python-client-quickstart-guide.html>
- ❖ JavaScript: <https://docs.min.io/docs/javascript-client-quickstart-guide.html>
- ❖ Go: <https://docs.min.io/docs/golang-client-quickstart-guide.html>
- ❖ C#/.NET: <https://docs.min.io/docs/dotnet-client-quickstart-guide.html>

```
var Minio = require('minio')

var minioClient = new Minio.Client({
  endPoint: 'play.min.io',
  port: 9000,
  useSSL: true,
  accessKey: 'Q3AM3UQ867SPQQA43P2F',
  secretKey: 'zuf+tfteSlswRu7BJ86wekitnifILbZam1KYY3TG'
});
```

对象存储概述

MinIO

OpenIO

Cloudian

其他对象存储平台

- ❖ OpenIO
- ❖ <https://www.openio.io/>
- ❖ 部分开源: <https://github.com/open-io>

- ❖ 商业版的 MinIO
- ❖ 安全性、可靠性都好过 MinIO
- ❖ 访问效率是目前知名对象存储项目中最高的
- ❖ OpenIO 融入了**网格计算 (Grid Computing)** 技术



❖ 网格计算 (Grid Computing)

- ❖ 通过利用大量异构计算机（通常为台式机）的未用资源（CPU 周期和磁盘存储），将其作为嵌入在分布式电信基础设施中的一个虚拟的计算机集群，为解决大规模的计算问题提供一个模型
 - ❖ 网格计算包括**共享异构资源**（基于不同的平台，硬件、软件体系结构，以及计算机语言），这些资源位于不同的地理位置，属于一个使用公开标准的网络上的不同的管理域
 - ❖ 网格计算的焦点放在支持跨管理域计算的能力
 - ❖ 网格计算包含了大量的用户管理和授权管理的子模块
-
- ❖ **网格计算与云计算的主要区别：**
 - ❖ 云计算集群一般是同构的，而网格计算集群一般是异构的

[Product](#)[Subscription](#)[Customers](#)[Resources](#)[About](#)[Contact us](#)

High Performance Object Storage for Big Data and AI

Why choose between performance and scalability,
when you can have the best of both?

[Learn more](#)[Subscription plans](#)



Hyper-scalable object storage

Scale seamlessly from Terabytes to Exabytes. Simply add nodes to expand capacity, without rebalancing data, and watch as performance increases linearly.



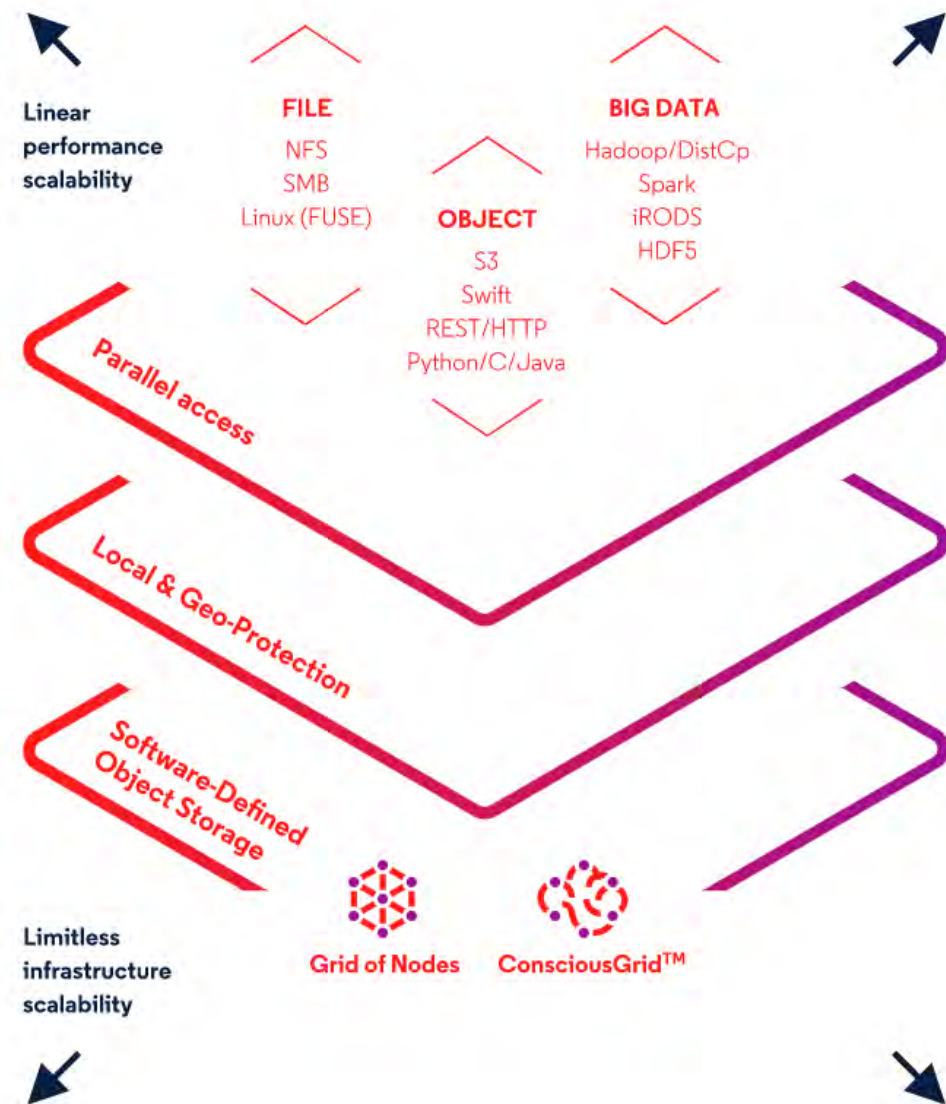
Designed for high performance

Transfer data at 1 Tbps and beyond. Experience consistent high performance, even during scaling operations. Ideal for capacity-intensive and challenging workloads.



Hardware agnostic, lightweight solution

Use servers and storage media that suit your evolving needs. Avoid vendor lock-in. You can combine heterogenous hardware at any time, of different specs, generations, and capacities.



Store your data

- Flexible and instant scaling
- Optimal positioning of data
- Performance optimization
- Smart hardware management

Manage your data

- Fine-grained storage policies for automated tiering
- Object versioning and container snapshots
- On the fly and delayed compression
- WORM mode

Protect your data

- Erasure coding and n-time replication with multiple dynamic storage policies available
- Geo-distributed and stretched cluster topology
- Encryption of data at rest

Access your data

- S3 and Swift APIs
- Native Python, C and Java APIs
- File access through OIO-FS
- Big Data integrations

[See all features](#)

[Technology & requirements](#)

- ❖ 安装和使用指南：
- ❖ <https://docs.openio.io/latest/source/sandbox-guide/quickstart.html>
- ❖ Docker 容器试用版：
- ❖ `docker run -ti --tty openio/sds`

对象存储概述

MinIO

OpenIO

Cloudian

其他对象存储平台

❖ Cloudian

❖ <https://cloudian.com/>

❖ 整合了对象存储的硬件平台



对象存储概述

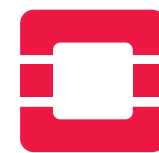
MinIO

OpenIO

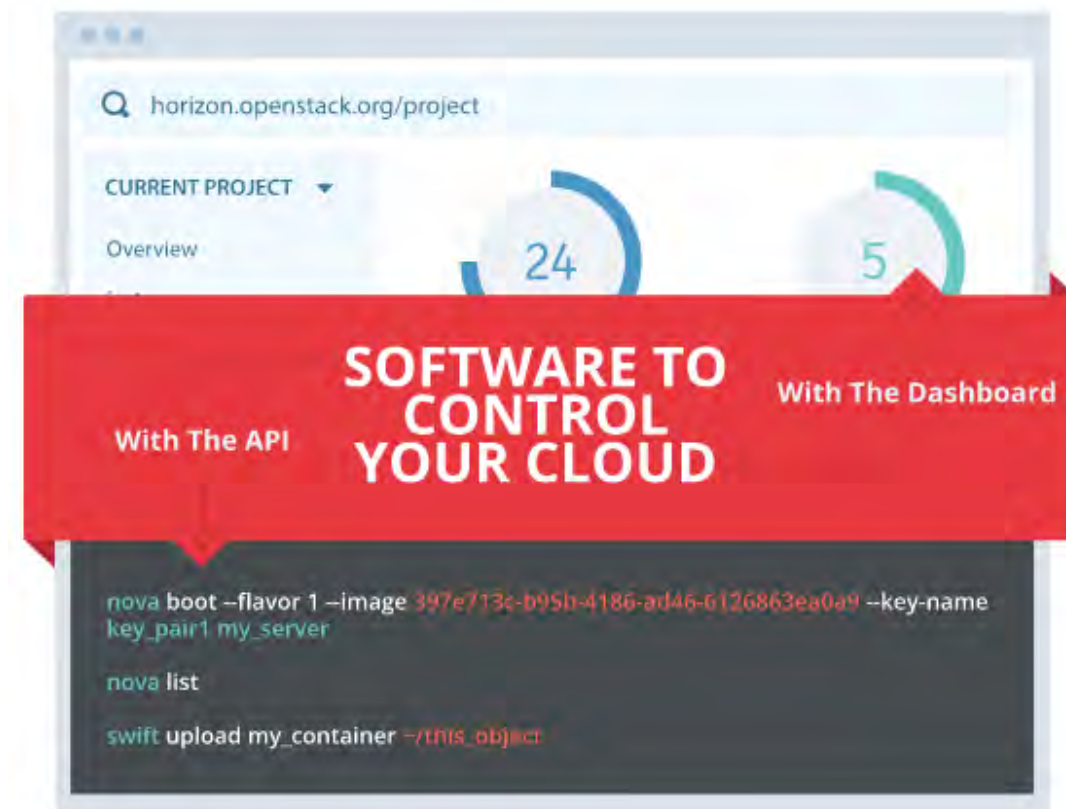
Cloudian

其他对象存储平台

- ❖ OpenStack
- ❖ <https://cn.aliyun.com/product/oss>
- ❖ OpenStack 是美国国家航空航天局（NASA）和 Rackspace 合作研发的云计算软件，以Apache 授权条款授权，并且是自由和开放源代码软件



openstack®



- ❖ OpenStack 组件
- ❖ Nova - 运算
- ❖ Keystone - 身份验证机制
- ❖ Glance - 虚拟机磁盘映像档 (Virtual Machine Image) 发送服务
- ❖ Cinder - Block 数据访问
- ❖ Neutron - 网络管理
- ❖ Swift - 对象存储
- ❖ Horizon - 简易 Web 界面和管理控制台
- ❖ Trove - 数据库管理
- ❖ Sahara - 海量数据运算布署
- ❖ 等等...

- ❖ OpenStack Swift / SwiftStack
- ❖ <https://www.swiftstack.com/>



- ❖ OpenStack Swift（现已改名 SwiftStack）是 OpenStack 其中一个组件

SwiftStack Data Platform

The platform for intelligent data and enables deep learning at ultra-scale

▶ LEARN MORE

❖ 阿里云对象存储 OSS

❖ <https://cn.aliyun.com/product/oss>



- ❖ 海量、安全、低成本、高可靠的云存储服务，提供 99.9999999999% 的数据可靠性
- ❖ 使用 RESTful API 可以在互联网任何位置存储和访问，容量和处理能力弹性扩展，多种存储类型供选择全面优化存储成本

- ❖ 三大云服务平台
- ❖ Google Cloud: <https://cloud.google.com>
- ❖ Amazon Web Services: <https://aws.amazon.com>
- ❖ 阿里云: <https://www.aliyun.com>



Google Cloud



❖ 课外阅读

- ❖ 《云存储技术——分析与实践》，刘洋著，经济管理出版社
- ❖ <http://product.dangdang.com/24247525.html>
- ❖ 《Ahead in the Cloud》，Stephen Orban（GM of AWS）
- ❖ <https://www.amazon.com/Ahead-Cloud-Practices-Navigating-Enterprise/dp/1981924310/>
- ❖ 《Cloud Computing: Concepts, Technology & Architecture》，Thomas Erl
- ❖ <https://www.amazon.com/Cloud-Computing-Concepts-Technology-Architecture/dp/0133387526/>

Thanks!