

S3 & Minio: Cloud Object Storage

Modul 346, BBZW

Patrick Bucher

Speicher kann nach folgenden Kriterien kategorisiert werden:

- Zugriffsgeschwindigkeit
- Kapazität
- Kosten

Als Faustregel gilt:

- **schneller** Speicher ist **teuer/klein**
- **langsamer** Speicher ist **günstig/gross**

Grobe Angaben, Stand 2022

Speichertyp	Zugriffszeit	Kapazität	Preis
CPU-Register	< 1 ns	64 kB	> 1 CHF pro kB
CPU-Cache (L1, L2, ...)	5-10 ns	8-64 MB	> 1 CHF pro MB
RAM	15-20 ns	16-256 GB	~ 4 CHF pro GB
SSD	25-100 μ s	0.25-4 TB	~ 0.1 CHF pro GB
HDD	5-10 ms	1-20 TB	~ 30 CHF pro TB
Tape	10-100 s	1-30 TB	~ 10 CHF pro TB

Einsatzgebiete von Speichertypen

Speichertyp	Einsatzgebiet
CPU-Register	Daten für aktuell laufende Instruktionen
CPU-Cache	Daten für häufig verwendete Instruktionen
RAM	Daten für derzeit laufende Programme (Datenstrukturen)
SSD	Daten für häufigen Zugriff (Dokumente, Programme)
HDD	Daten für seltenen Zugriff (Datenablage, Online-Backup)
Tape	Daten für Langzeitarchivierung (Offline-Backup)

S3: (Amazon) Simple Storage Service

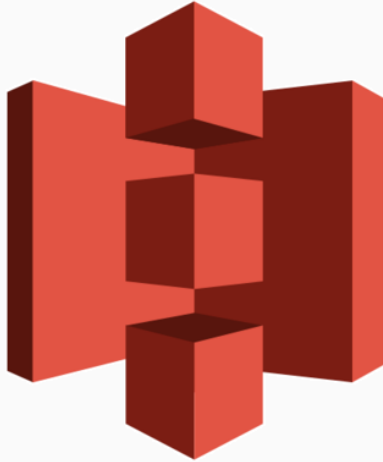


Abbildung 1: Amazon Simple Storage Service (S3)

Amazon Simple Storage Service (S3)

- Object Storage mit Web-Schnittstelle
 - “Object”: unstrukturierte Daten (BLOB: Binary Large Object, “Dateien”)
- Organisationsprinzip: flache Hierarchie
 - **Bucket** (Eimer) als Ordner
 - **Object** (Objekt) als Datei (max. 5 TB)
 - grobe Zugriffsberechtigung auf Stufe Bucket möglich
- Einhängen von Buckets als Dateisystem möglich (FUSE: Filesystem in Userspace)

S3-Speicherklassen

Ähnlich zur Hardware-Speicherhierarchie bietet S3 **Speicherklassen**:

Speicherklasse	Zweck
Standard	Häufig verwendete Daten
Standard-IA	Selten verwendete Daten (IA: Infrequent Access)
One Zone-IA	Selten verwendete Daten in einer bestimmten <i>Zone</i>
Intelligent-Tiering	Automatische Ablage in kosteneffizienter Klasse
Outposts	Ausserhalb von Amazon-Infrastruktur gehosted
Glacier Instant Retrieval	Seltene, aber schnelle Zugriffe
Glacier Flexible Retrieval	Seltene, unterschiedlich schnelle Zugriffe (3 Stufen)
Glacier Deep Archive	Seltene, langsame (dafür günstige) Zugriffe

Eine *Zone (Availability Zone)* bezeichnet die geografische Lage des Speichers.

Amazon organisiert die Cloud-Infrastruktur in **Regionen** und **Zonen**.

- **Region**

- geografischer Ort
- verfügt über eine Reihe von Rechenzentren
- von anderen Regionen getrennt

- **(Availability) Zone**

- geografischer Ort *innerhalb einer Region*
- erstreckt sich über ein oder mehrere Rechenzentren
- von anderen Zonen getrennt (Ausfallsicherheit)

- **Virtual Private Cloud:** Verbund aus gewählten Zonen (optional mit Outpost)

Die Datenverarbeitung und -speicherung kann auf bestimmte Regionen eingeschränkt werden!

Regionen und Zonen (II)

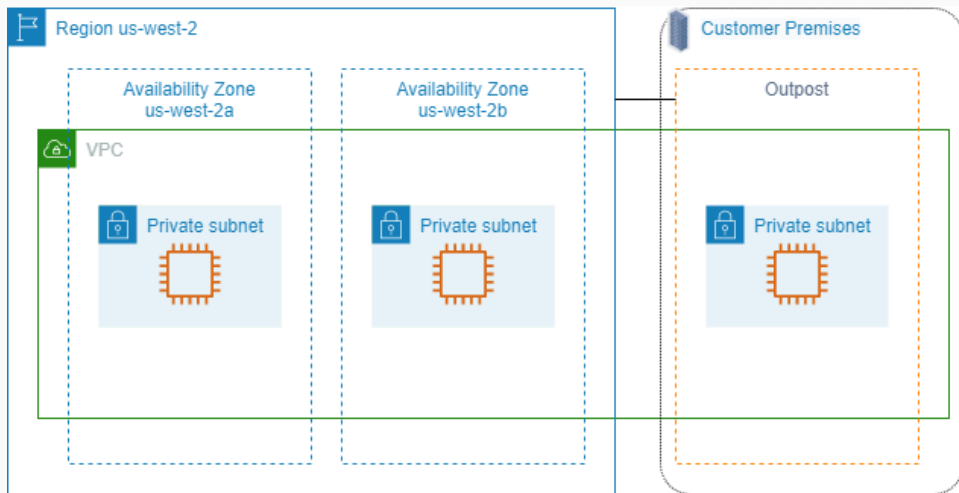


Abbildung 2: Regionen, Zonen, Outpost, VPC

Amazon bietet u.a. die folgenden Regionen an:

- Nordamerika:
 - us-east-1: Northern Virginia, us-east-2: Ohio
 - us-west-1: Northern California, us-west-2: Oregon
 - ca-central-1: Canada
- Europa:
 - eu-central-1: Frankfurt, eu-central-2: Zürich
 - eu-west-1: Irland, eu-west-2: London, eu-west-3: Paris
 - eu-south-1: Mailand, eu-south-2: Spanien
 - eu-north-1: Stockholm
- Afrika, Asien, Mittlerer Osten, Südamerika

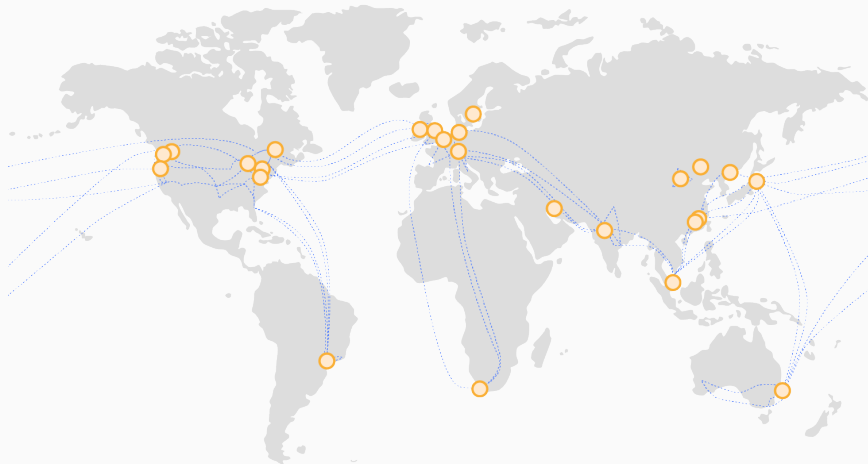


Abbildung 3: Die globale Amazon-Cloud-Infrastruktur



Abbildung 4: Minio

- MinIO ist kompatibel zu Amazon S3, kann aber selber gehostet werden.
 - Kubernetes, OpenShift, Cloud (Amazon, Azure, Google), Linux, Windows usw.
- In Go implementiert
 - Server: [minio](#)
 - Client: [mc](#)

s3cmd

- Kommandozeilenwerkzeug zum Zugriff auf S3-Speicher

s3fs

- Kommandozeilenwerkzeug zum Einbinden von S3-Speicher als FUSE-Dateisystem
 - FUSE: Filesystem in Userspace

- [Cloud Object Storage – Amazon S3](#)
 - [S3: Features](#)
 - [S3: Preismodell](#)
 - [S3: Preisrechner](#)
 - [AWS Free Tier](#)
- [Minio](#)
 - [MinIO-Server](#)
 - [MinIO-Client](#)
 - [Dokumentation \(Linux\)](#)
 - [Dokumentation \(Windows\)](#)
- [s3cmd](#)
- [s3fs](#)