AWS Verweis Architekturen

Amazon EC2

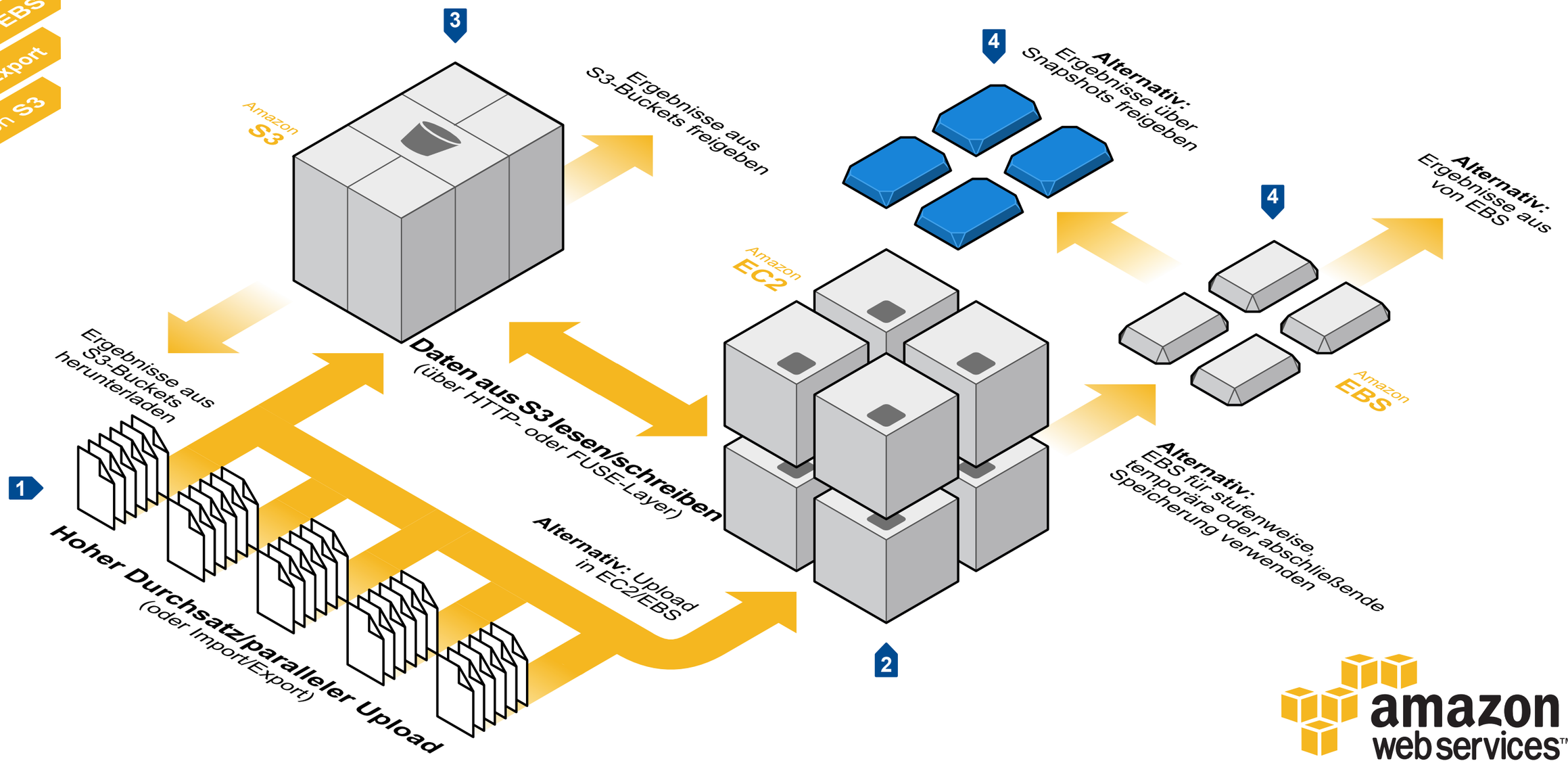
Amazon EBS

AWS Import/Export

Amazon S3

LARGE SCALE COMPUTING & GROßE DATENSÄTZE

Amazon Web Services wird häufig in Bereichen mit hohen Computingaufwand, wie wissenschaftlichem Rechnen, Simulation oder Forschungsprojekten eingesetzt. Diese Szenarios beinhalten riesige Datensätze aus wissenschaftlichen Geräten, Messgeräten oder anderen Rechenprojekten. Nachdem diese Datensätze erfasst sind, müssen sie durch mit hohen Rechenleistungen analysiert werden, um aussagekräftige Ergebnisdatensätze zu erhalten. Im Idealfall sind die Ergebnisse bereits bei der Erfassung der Daten verfügbar. Oftmals werden diese Ergebnisse später einem größeren Publikum veröffentlicht.



System Übersicht

1 Um große Datensätze in AWS hochzuladen, ist es wichtig, die verfügbare Bandbreite effektiv zu nutzen. Hierzu können Sie Daten parallel über mehrere Clients in **Amazon Simple Storage Service (S3)** hochladen. Jeder dieser Clients nutzt Multithreading, um simultane Uploads oder stufenweise Uploads für zur weiteren Parallelisierung zu ermöglichen. TCP-Einstellungen wie die Fensterskalierung und die selektive Bestätigung können zur weiteren Optimierung des Durchsatzes angepasst werden. Durch effektive Optimierungen sind Uploads von mehreren Terabytes pro Tag möglich. Eine weitere Alternative für große Datensätze ist **Amazon Import/ Export.** Hierbei wird das Versenden von Speichergeräten an AWS und das Speichern der entsprechenden Inhalte in **Amazon S3** oder **Amazon EBS** unterstützt.

2 Die parallele Verarbeitung umfangreicher Aufträge ist kritisch und existierenden parallele Anwendungen können in der Regel auf mehreren Instanzen der **Amazon Elastic Compute Cloud (EC2)** ausgeführt werden. Eine parallele Anwendung setzt manchmal einen großen Arbeitsspeicher voraus, so dass alle Knoten effizient lesen und schreiben können. S3 kann als ein solcher Arbeitsspeicher eingesetzt werden – entweder direkt über HTTP oder über eine FUSE-Layer (z. B s3fs oder SubCloud) wenn die Anwendung ein POSIX-Dateisystem benötigt.

3 Sobald der Auftrag ausgeführt und die Ergebnisdaten in **Amazon S3**, **Amazon EC2,** Instanzen heruntergefahren werden, und die Ergebnisdaten heruntergeladen werden. Die Ausgabedaten können für andere Nutzer freigegeben werden – entweder durch die Erteilung von Leseberechtigungen an ausgewählte Nutzer oder an alle Nutzer durch den Einsatz temporärer URLs.

4 Anstatt **Amazon S3** können Sie **Amazon EBS** verwenden, um den Eingangsdatensatz bereitzustellen, als temporärer Speicherort zu fungieren und/oder den Ausgangsdatensatz zu erfassen. Während des Uploads sind zudem parallele Upload-Streams und eine TCP-Optimierung anwendbar. Darüber hinaus können bei Uploads, die UDP verwenden, höhere Geschwindigkeiten erwartet werden. Der Ergebnisdatensatz kann auf EBS-Datenträgern gespeichert werden. Hier können dann auch Snapshots der Datenträger für die Freigabe der Daten erstellt werden.