

Unbegrenzte Möglichkeiten: Big Data und Business Intelligence in der Cloud

Constantin Gonzalez
Amazon Web Services Germany GmbH
München

Schlüsselworte: Cloud, AWS, Amazon Web Services, Big Data, Analytics, Business Intelligence, BI, Hadoop. Database, Storage.

Einleitung

Wissenschaftlicher, Entwickler und viele andere Technologen aus zahlreichen Branchen nutzen Amazon Web Services zum Erfüllen der Herausforderungen der immer größer werdenden Menge, Vielfältigkeit und Geschwindigkeit digitaler Informationen. Amazon Web Services stellen eine umfassende Palette von Cloud Computing-Ressourcen zur Vereinfachung des Big Data-Managements, Senkung der Kosten, Erzielung eines Wettbewerbsvorteils und Beschleunigung der Innovation dar.

Big Data mühelos in die Cloud verlagern

Es scheint, als würden heutzutage überall Daten erzeugt – von Ihren Kunden in sozialen Netzwerken bis zu den Instances, in denen Ihre Webanwendungen ausgeführt werden. AWS erleichtert die Bereitstellung der Speicher-, Rechen- und Datenbankservices, die Sie benötigen, um diese Daten in nützliche Informationen für Ihr Unternehmen umzuwandeln. AWS bietet außerdem Datenübertragungsservices, mit denen Big Data schnell in die bzw. aus der Cloud verschoben werden können: AWS Direct Connect und AWS Import/Export. Darüber hinaus fallen für in AWS eingehenden Datenverkehr keine Kosten an.

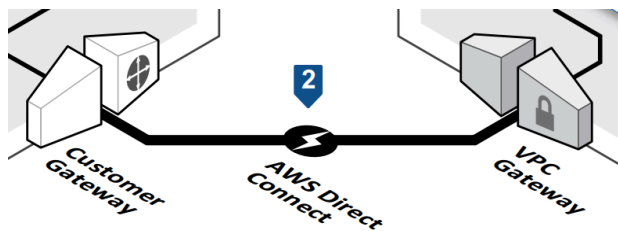


Abb. 1: AWS Direct Connect

Speicher für das Internet, Speicher für EC2.

Unabhängig davon, ob Sie pharmazeutische Daten für die Analyse, Finanzdaten für Berechnungen und Preisfindung oder Multimediadateien wie Fotos und Videos speichern, ist Amazon Simple Storage Service (S3) der ideale Ort für eine beständige Speicherung von Originalinhalten. Mit einer Beständigkeit von 99,99999999 % und ohne einzelne Fehlerstelle ist Amazon S3 Ihr elementarer Speicher für Big Data-Objekte.

Amazon Elastic Compute Cloud (EC2) bietet virtuelle Maschinen als Service. Amazon Elastic Block Store (EBS) bietet Festplatten als persistenten Speicher für diese virtuellen Maschinen. Amazon EBS-Volumes liefern die einheitliche Leistung, die benötigt wird, um Big Data-Verarbeitungslasten wie Ihre eigenen relationalen oder NoSQL-Datenbanken, Unternehmensanwendungen und leistungsstarken verteilten Netzwerkdateisysteme mit kurzer Latenz zu bewältigen.

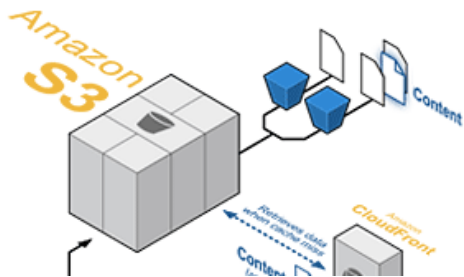


Abb. 2: Amazon Simple Storage Service (S3)

SSD-Laufwerke, zu Ihren Diensten.

NoSQL-Datenspeicher profitieren enorm von der Geschwindigkeit von SSD-Laufwerken (Solid State Drives). DynamoDB nutzt diese standardmäßig. Doch wenn Sie mit Alternativen aus AWS Marketplace arbeiten, z. B. Cassandra oder MongoDB, können Sie mit der Instance-Klasse "High I/O" Ihren Zugriff auf Terabytes von auf Solid State-Laufwerken gespeicherten Daten bedarfsabhängig beschleunigen.

Skalieren Sie Ihre NoSQL-Datenbank ohne Einschränkungen... oder Probleme.

Wenn Sie eine NoSQL-Datenbank ohne den damit verbundenen Betriebsaufwand benötigen, hört Ihre Suche bei Amazon DynamoDB auf. DynamoDB ist ein schneller vollständig verwalteter NoSQL-Datenbankservice, mit dessen Hilfe beliebig große Datenmengen einfach und wirtschaftlich gespeichert und abgerufen sowie sämtliche Datenverkehrsansforderungen erfüllt werden können. Amazon DynamoDB bietet einen garantierten Durchsatz und dank seiner Latenz im einstelligen Millisekundenbereich eignet sich der Service besonders für Online-Spiele, digitale Werbung, mobile Anwendungen und viele andere Big Data-Anwendungen.

Big Data ist nicht gleich NoSQL; auch relationale Datenbanken sind groß.

Die Big Data-Innovation geht über NoSQL hinaus, denn es geht hauptsächlich darum, je nach Ihren Geschäftsanforderungen die für Ihre Daten geeignete Technologie anzubieten. Relationale Datenbanken bieten eine schnelle, berechenbare und einheitliche Leistung und sind für Transaktionsverarbeitungslasten, z. B. Verkaufsstellen und Finanzgeschichte, optimiert. Relationale Datenbanken spielen in umfangreichen Big Data-Architekturen eine ergänzende Rolle zu NoSQL-Datenbanken.

Mit Amazon RDS ist das Einrichten, Betreiben und Skalieren einer relationalen Datenbank in der Cloud ganz einfach. Dieser Service stellt kosteneffiziente und individuell anpassbare Kapazität zur Verfügung und erledigt gleichzeitig zeitraubende Verwaltungsaufgaben für Datenbanken, sodass Sie sich stärker auf Ihre Anwendungen und Ihr Geschäft konzentrieren können.

Oracle und Amazon Web Services.

Amazon Web Services (AWS) und Oracle haben zusammengearbeitet, um Kunden bequeme Optionen zur Bereitstellung von Unternehmensanwendungen in der Cloud zu bieten. Kunden können Unternehmenslösungen mit Oracle-Anwendungen nicht nur erstellen und mithilfe von Datenbank- und Middleware-Software von Oracle in der AWS-Cloud ausführen, sondern sogar gesamte Unternehmenssoftware-Stacks von Oracle in Amazon EC2 starten.

Einfache Analysen für jedermann.

Amazon Elastic MapReduce (EMR) stellt das leistungsstarke Hadoop-Framework in Amazon EC2 als benutzerfreundlichen Managed Service zur Verfügung. Mit Amazon EMR können Sie sich auf Ihre Abfragen vom Typ "Map/Reduce" konzentrieren. Nutzen Sie die breite Palette von Hadoop-Tools, während die Bereitstellung auf einer hochskalierbaren, sicheren Infrastrukturplattform erfolgt.

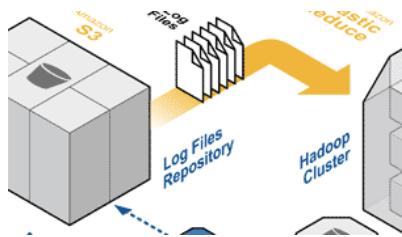


Abb. 3: Amazon Elastic Mapreduce

Beispiel für das Starten eines Amazon Elastic MapReduce-Streaming-Jobs:

```
./elastic-mapreduce --create --stream --ami-version 3.0.3 \ --instance-type  
m1.large --arg "-files" --arg  
"s3://elasticmapreduce/samples/wordcount/wordSplitter.py" \ --input  
s3n://elasticmapreduce/samples/wordcount/input --mapper wordSplitter.py --  
reducer aggregate \ --output s3n://myawsbucket/output/2014-01-16
```

Supercomputing zum selbst festgelegten Preis.

Wie schnell käme Ihr Projekt mit weiteren 1000 virtuellen Maschinen voran? Und wie mit 10 000? Der in Amazon Elastic MapReduce integrierte Amazon Spot Market lässt Sie Ihren eigenen Preis für die benötigten Datenverarbeitungsressourcen auswählen. Das bedeutet, dass Sie Ihr eigenes Kosten- und Leistungsverhältnis bestimmen und bei Bedarf mehr Spot-Instances hinzuziehen oder die Kosten signifikant senken können.

Data Warehousing in Petabyte-Dimensionen in Minutenschnelle.

Amazon Redshift stellt ein schnelles, vollständig verwaltetes Data Warehouse im Petabyte-Bereich für unter 1000 USD pro Terabyte im Jahr zur Verfügung. Amazon Redshift bietet eine schnelle Abfrage- und E/A-Leistung für Datenmengen praktisch beliebiger Größe mithilfe einer spaltenbasierten Speichertechnologie und parallelen Verteilung von Abfragen auf mehrere Knoten. Sie können in Minutenschnelle ein vollständig verwaltetes Data Warehouse mit automatischen Backups und integrierter Verschlüsselung bereitstellen. Vorhandene Business Intelligence-Tools lassen sich problemlos einbinden.

Online-Speicher für Software in der Cloud.

AWS Marketplace ist ein Online-Shop, in dem Entwickler und IT-Experten Software für die Ausführung in der AWS-Cloud einfach finden und einsetzen können. Sie können nun Software im AWS Marketplace für die Erfassung, Analyse und Zusammenarbeit finden, wenn Sie zur Unterstützung von Big Data-Projekten mit großen Datenmengen arbeiten. AWS Marketplace bietet Software zur Unterstützung einer Palette von Big Data-Lösungen einschließlich NoSQL, Hadoop, Cassandra und Business Intelligence (BI).

Streaming von Big Data und Analysen in Echtzeit.

Amazon Kinesis ist ein verwalteter Service für die Echtzeitverarbeitung gestreamter Big Data. Amazon Kinesis unterstützt Datendurchsätze von Megabytes bis zu Gigabytes an Daten pro Sekunde und kann nahtlos skaliert werden, um Daten-Streams aus Hunderten bis Tausenden unterschiedlichen Quellen verarbeiten zu können. Da dieser Service Ihnen hohe Verfügbarkeit und Beständigkeit bei niedrigen Kosten gewährleistet, können Sie sich jetzt auf die Auswertung Ihrer Daten konzentrieren und dadurch schneller und günstiger bessere Entscheidungen treffen.

Selten genutzte Daten zu extrem niedrigen Kosten archivieren.

Mit Amazon Glacier können Sie die Verwaltungsaufgaben für den Betrieb und die Skalierung von Archivierungsspeichern AWS überlassen und die Daten auf sehr einfache Weise für lange Zeit (Jahre oder Jahrzehnte) aufbewahren. Amazon Glacier ist ein überaus kostengünstiger Speicher für die Archivierung. Die Preise beginnen bei 0,01 USD pro GB pro Monat. Es fallen keine vorherigen Investitionskosten an und alle laufenden Betriebskosten sind im Preis inbegriffen.

Erste Schritte

Begeben Sie sich auf eine geführte Tour mit praxisnahen Daten und Beispielen durch Hadoop und Amazon Elastic MapReduce, um zu erfahren, wie Sie Ihre Daten besser nutzen können.

Erfahren Sie mehr darüber, wie Sie mit Amazon Redshift, einem schnellen, vollständig verwalteten Data Warehouse im Petabyte-Bereich, Informationen über Ihre Anwendungen, Ihre Kunden und Ihr Unternehmen gezielter nutzen können.

Mehr Informationen dazu finden Sie unter:

<http://aws.amazon.com/de/big-data/>,

<http://aws.amazon.com/de/big-data/getting-started/>,

<http://aws.amazon.com/de/oracle/>

Kontaktadresse:

Constantin Gonzalez

Amazon Web Services Germany GmbH

c/o Marcel-Breuer-Str. 12

D-80807 München

Telefon: +49 (0) 89-35 80 31 03 9

E-Mail glez@amazon.de

Internet: aws.amazon.com