

Datenbanken: Grundlagen und AWS-Services

Einführung in Datenbanken

- **Definition:** Eine Datenbank ist ein organisiertes System zur Speicherung und Verwaltung von Daten, das einen schnellen Zugriff und effiziente Verwaltung ermöglicht.
- **Funktionen:**
 - Speicherung großer Datenmengen
 - Schneller Zugriff und Abfrage
 - Aktualisierung und Verwaltung von Daten
 - Sicherheit und Schutz sensibler Informationen

Arten von Datenbanken

- **Relationale Datenbanken (SQL):**
 - Organisieren Daten in Tabellen mit Zeilen und Spalten
 - Verwenden von Structured Query Language (SQL) für Abfragen
 - Geeignet für strukturierte Daten und komplexe Abfragen
- **NoSQL-Datenbanken:**
 - Flexiblere Strukturen für unstrukturierte oder semi-strukturierte Daten
 - Verschiedene Typen wie Dokumenten-, Schlüssel-Wert-, Graph- und Spaltenfamilien-Datenbanken
 - Skalierbar und geeignet für große Datenmengen

Datenbankmanagementsysteme (DBMS)

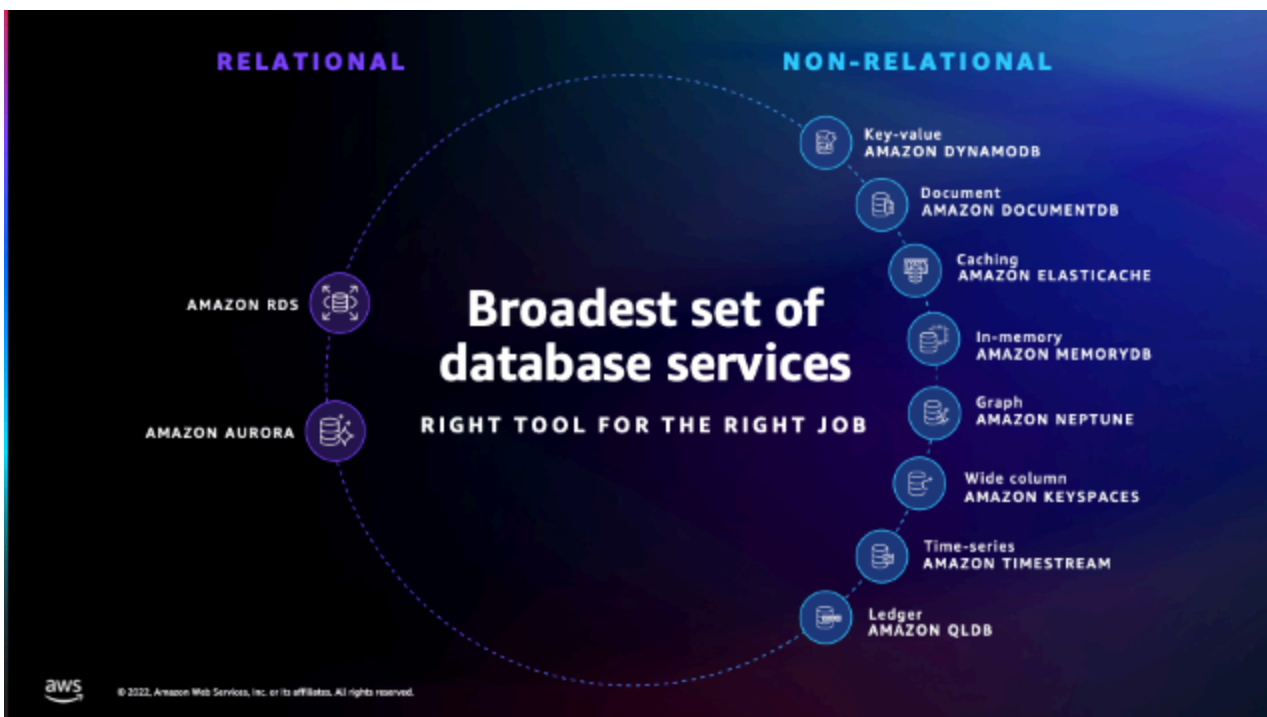
- **Definition:** Software zur Verwaltung von Datenbanken, die Funktionen wie Datenspeicherung, Abfrage, Sicherheit und Integrität bereitstellt.
- **Beispiele:**
 - Relationale DBMS: MySQL, PostgreSQL, Oracle
 - NoSQL DBMS: MongoDB, Cassandra, Amazon DynamoDB

Anwendungsfälle für verschiedene Datenbanktypen

Datenspeicher	Anwendungsfall
Datenbank auf EC2	Volle Kontrolle über Instanz und Datenbank; bevorzugte DB nicht unter RDS verfügbar
Amazon RDS	Traditionelle relationale Datenbank für OLTP; strukturierte Daten; bestehende Anwendungen, die RDBMS erfordern
Amazon Aurora	Hochleistungsfähige relationale Datenbank mit MySQL- und PostgreSQL-Kompatibilität; ideal für anspruchsvolle, geschäftskritische Anwendungen
Amazon DynamoDB	Name/Wert-Paar-Daten; unvorhersehbare Datenstruktur; In-Memory-Leistung mit Persistenz; hoher I/O-Bedarf; dynamische Skalierung erforderlich
Amazon RedShift	Data Warehouse für große Mengen aggregierter Daten; hauptsächlich OLAP-Workloads
Amazon Neptune	Beziehungen zwischen Objekten sind von hohem Wert, z. B. soziale Netzwerke, Wissensgraphen
Amazon ElastiCache	Schneller temporärer Speicher für kleine Datenmengen; hochvolatile Daten (nicht persistent)

Übergang zu AWS-Datenbankservices

Nachdem wir die Grundlagen von Datenbanken und ihre verschiedenen Typen besprochen haben, betrachten wir nun die spezifischen Datenbankservices, die Amazon Web Services (AWS) anbietet.



AWS Datenbankservices

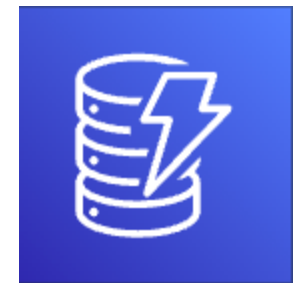
Amazon Relational Database Service (RDS)

- **Beschreibung:** Ein verwalteter Service, der das Einrichten, Betreiben und Skalieren relationaler Datenbanken in der Cloud erleichtert.
- **Unterstützte Datenbank-Engines:**
 - Amazon Aurora
 - PostgreSQL
 - MySQL
 - MariaDB
 - Oracle
 - Microsoft SQL Server
- **Hauptmerkmale:**
 - Automatisierte Backups und Patches
 - Multi-AZ-Bereitstellungen für hohe Verfügbarkeit
 - Read Replicas für verbesserte Leseleistung
 - Skalierbare Rechen- und Speicherkapazität



Amazon DynamoDB

- **Beschreibung:** Ein vollständig verwalteter NoSQL-Datenbankservice mit schneller und vorhersehbarer Leistung sowie nahtloser Skalierbarkeit.
- **Hauptmerkmale:**
 - Schema-loser Aufbau für flexible Datenstrukturen
 - Automatische Skalierung von Durchsatz und Speicher
 - Unterstützung für globale Tabellen für Multi-Region-Replikation
 - Amazon DynamoDB Accelerator (DAX) für In-Memory-Caching



Amazon Aurora

- **Beschreibung:** Amazon Aurora ist ein vollständig verwalteter relationaler Datenbankservice, der mit MySQL und PostgreSQL kompatibel ist und speziell für die Cloud entwickelt wurde.
- **Hauptmerkmale:**
 - Bis zu fünfmal schneller als MySQL und dreimal schneller als PostgreSQL
 - Automatische Skalierung des Speicherplatzes (bis zu 128 TB)
 - Multi-AZ-Bereitstellung mit kontinuierlicher Sicherung und automatischer Wiederherstellung
 - Unterstützt Read Replicas für Lastverteilung und hohe Verfügbarkeit
 - Integration mit AWS-Diensten wie AWS Lambda, Amazon S3 und Amazon Redshift



Amazon RedShift

- **Beschreibung:** Ein schneller, vollständig verwalteter Data-Warehouse-Service, der die Analyse aller Daten mit Standard-SQL und vorhandenen Business-Intelligence-Tools ermöglicht.
- **Hauptmerkmale:**
 - Spaltenbasierte Speicherung für effiziente Abfragen
 - Massiv parallele Verarbeitung für schnelle Leistung
 - Automatische Replikation und kontinuierliche Backups
 - Integration mit AWS-Services wie S3 und EMR



Amazon ElastiCache

- **Beschreibung:** Ein Webservice, der das Einrichten, Betreiben und Skalieren von In-Memory-Caches in der Cloud erleichtert.
- **Unterstützte Engines:**
 - Memcached
 - Redis
- **Hauptmerkmale:**
 - Reduziert Latenz und verbessert Durchsatz für leseintensive Anwendungen
 - Unterstützt komplexe Datenstrukturen und erweiterte Datenverarbeitung
 - Hohe Verfügbarkeit mit Multi-AZ-Unterstützung
 - Skalierbar mit Cluster-Modus für Redis



Amazon EMR

- **Beschreibung:** Ein verwalteter Service, der die Verarbeitung großer Datenmengen mit Open-Source-Tools wie Apache Hadoop, Spark und HBase erleichtert.
- **Hauptmerkmale:**
 - Skalierbare Datenverarbeitung für Big Data-Analysen
 - Integration mit anderen AWS-Services wie S3 und DynamoDB
 - Kosteneffiziente Verarbeitung durch Nutzung von Spot-Instanzen
 - Unterstützung für verschiedene Workloads wie Batch-Verarbeitung, ETL und Data Warehousing



Sicherheitskonzepte für AWS-Datenbankservices

- **Sicherheitsgruppen:**
 - Steuerung des Datenverkehrs durch Zuweisung von Sicherheitsgruppen an Amazon RDS, DynamoDB und andere Datenbanken.
 - Ermöglicht Zugriffsregeln für eingehenden und ausgehenden Datenverkehr, z. B. Port 3306 für MySQL.
- **VPC:**
 - Nutzung von Amazon Virtual Private Cloud (VPC), um Datenbanken in isolierten Netzwerken bereitzustellen.
 - Kombination mit Subnetzen, Internet- und NAT-Gateways, um eine sichere Kommunikation zu ermöglichen.
- **IAM-Berechtigungen:**
 - Verwaltung des Zugriffs auf AWS-Datenbankressourcen durch rollenbasierte Zugriffskontrollen (IAM-Rollen).
 - Dynamische Zugriffskontrollen für API-Aufrufe durch AWS SDKs.

Automatisierte Backups und Wiederherstellung in RDS

- **Automatische Backups:**
 - AWS RDS bietet tägliche Snapshots und Log-Backup für eine einfache Wiederherstellung.
 - Die Aufbewahrungsdauer für Backups kann von 1 bis 35 Tagen konfiguriert werden.
- **Snapshots und Point-in-Time-Wiederherstellung:**
 - Benutzer können manuell Snapshots für spezifische Zeitpunkte erstellen.
 - Point-in-Time-Wiederherstellung ermöglicht eine Wiederherstellung zu einem bestimmten Zeitpunkt, z. B. vor einer Datenmanipulation.
- **Cross-Region-Replikation:**
 - Multi-AZ- und Multi-Region-Backups zur Sicherstellung der Datenverfügbarkeit und -resilienz.

Preisgestaltung und Kostenoptimierung für AWS-Datenbanken

- **On-Demand-Preise:**
 - Kosten basieren auf genutzter Rechen- und Speicherkapazität ohne langfristige Verpflichtungen.
- **Reserved Instances:**
 - Günstigere Preisoptionen durch Vorauszahlung für 1 oder 3 Jahre, ideal für kontinuierliche Arbeitslasten.
- **Speicher- und IOPS-Kosten:**
 - Optimierung durch Auswahl passender Speicheroptionen (z. B. General Purpose SSD vs. Provisioned IOPS SSD).
- **Kostenoptimierung mit ElastiCache:**
 - ElastiCache als Caching-Lösung, um Abfragen zu reduzieren und Latenzzeiten zu senken.

Optimierung von Datenbankleistung und Skalierbarkeit

- **Read Replicas in RDS:**
 - Skalierung durch das Hinzufügen von Read Replicas, um Leseanfragen von der primären Datenbank zu entlasten.
- **DynamoDB Auto Scaling:**
 - DynamoDB passt Kapazitätsgrenzen automatisch an die Datenmenge und den Durchsatzbedarf an.
- **Caching mit ElastiCache:**
 - Verbesserung der Anwendungsleistung durch Zwischenspeichern häufig abgerufener Daten.
- **Partitionierung und Sharding:**
 - Verteilung großer Datensätze auf mehrere Partition

Skalierung bei AWS-Datenbankservices

Amazon RDS

- **Vertikale Skalierung:** Anpassung der Instanzgröße, z. B. von `db.t3.micro` auf `db.m5.large`.
- **Horizontale Skalierung:** Nutzung von Read Replicas und Multi-AZ-Bereitstellungen für höhere Verfügbarkeit und Leseleistung.

Amazon DynamoDB

- **Automatische Skalierung:** DynamoDB passt die Kapazität dynamisch an den Bedarf an.
- **Partitionierung:** Automatische Partitionierung großer Datenmengen zur Leistungssteigerung.
- **Global Tables:** Multi-Region-Replikation für globale Verfügbarkeit und Fehlertoleranz.

Amazon Aurora

- **Automatische Speicher-Skalierung:** Erhöht den Speicherplatz bei Bedarf bis zu 128 TB.
- **Aurora Serverless:** Bedarfsbasierte Skalierung ohne feste Instanzengröße.
- **Read Replicas:** Unterstützung für bis zu 15 Lesereplikate zur Optimierung der Leseleistung.

Snapshots und Backups in AWS-Datenbankservices

Amazon RDS

- **Automatische Backups:** Tägliche Snapshots und Transaktionslogs für eine Point-in-Time-Wiederherstellung.
- **Manuelle Snapshots:** Speichern von Snapshots als Sicherungspunkte für Wartungen.
- **Cross-Region-Backups:** Kopieren von Backups in andere Regionen zur Sicherstellung der Verfügbarkeit.

Amazon DynamoDB

- **On-Demand Backups:** Vollständige Backups ohne Beeinflussung laufender Anwendungen.
- **Point-in-Time-Wiederherstellung (PITR):** Wiederherstellung einer Tabelle zu einem beliebigen Punkt innerhalb der letzten 35 Tage.
- **Global Tables:** Multi-Region-Replikation für automatische, konsistente Datensicherung über mehrere Regionen.

Amazon Aurora

- **Automatische Backups:** Kontinuierliche Snapshots für eine Point-in-Time-Wiederherstellung.
- **Manuelle Snapshots:** Sicherungspunkte für geplante Wartungen.
- **Continuous Backup und Restore:** Kontinuierliche Speicherung in Amazon S3 zur nahtlosen Wiederherstellung ohne Datenverlust.

Zusammenfassung

AWS-Datenbankservices bieten flexible, skalierbare und sichere Optionen für moderne Datenanforderungen:

- **Skalierung:** Bedarfsgesteuerte Anpassungen durch automatische Skalierung, Read Replicas und Multi-AZ-Bereitstellungen.
- **Verfügbarkeit:** Höhere Verfügbarkeit und Fehlertoleranz durch Cross-Region-Replikation und Multi-AZ-Optionen.
- **Datensicherung:** Umfangreiche Backup- und Wiederherstellungsoptionen mit automatischen und manuellen Snapshots sowie Point-in-Time-Wiederherstellung.

Mit diesen Funktionen können Unternehmen ihre Datenbanken effizienter, sicherer und kosteneffektiver verwalten.