Managed DB Services @ AWS

Tóbi Tamás, Technology Head @ TC2

NAPIREND

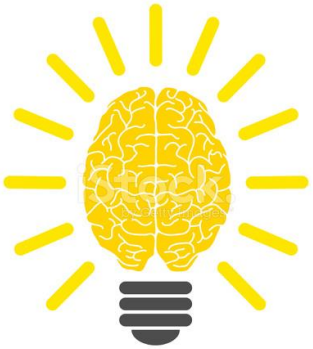
- SQL service management, scaling?
- Problémás, költséges és bonyolult on-premise megoldások
- Amazon RDS, Redshift, Redshift-Spectrum, Athena, DynamoDB, Neptune

On-premise megoldások nehézségei

- Host operációs rendszer management: patching, update, security,
- SQL service patching, update, security, maintenance,
- Hardware maintenance, update, purchasing, replacing, handling failures, storage, latency,
- Scalability? Availability? – általában bonyolult megoldások (Master-Save, Multi-Master replication, shared-nothing, sharding, synchronization, load-balancing…)
- Licensing,
- Backup & Restore,
- Monitoring & alerting

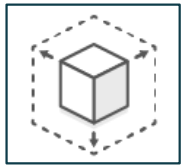
Miért AWS?

- **Egyszerű adminisztrálni**: az AWS gondoskodik a host OS és DB service update-ről, licensing-ről, időzítésről,
- **Egyszerűen skálázható:** push-button (vertikális), Aurora autoscaling (horizontális) skálázás, storage scaling
- **99.95% rendelkezésre állás**, automatizált backup, point-in-time recovery, self-healing,
- **Multi-AZ**, master-slave, “push button” read-replica, (cross-region)
- **Olcsó**: stop/start funkció, pay-as-you-go, BYOL (Bring Your Own License)



AMAZON RELATIONAL DATABASE SERVICE (RDS)

Easy to Administer: host OS és SQL service management-et az AWS veszi át (security updates, patching, etc**..), License-included**, AWS WEB console, CLI, SDK…



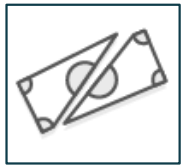
Highly Scalable: Amazon Aurora **serverless & autoscaling**, read-replicas, storage "push button" scaling

Available and Durable: Multi-AZ SLA min. 99.95% , automated backup, automated host-replacement, self-healing

Fast: push-button (vertical) scaling, autoscaling (horizontal), SSD backed, multi-AZ, read-replicas

Up to: 3.9TB memory, 25GBps bandwidth, 128 CPU core, storage: SSD 16TB, 40k IOPS ***

Secure: Network isolation (VPC), storage encryption (KMS), resource level permissions (IAM)



Inexpensive: **pas-as-you go**, license-included & BYOL options, reserved instances, stop-start, hourly rates

Amazon RDS engines

- Commercial:

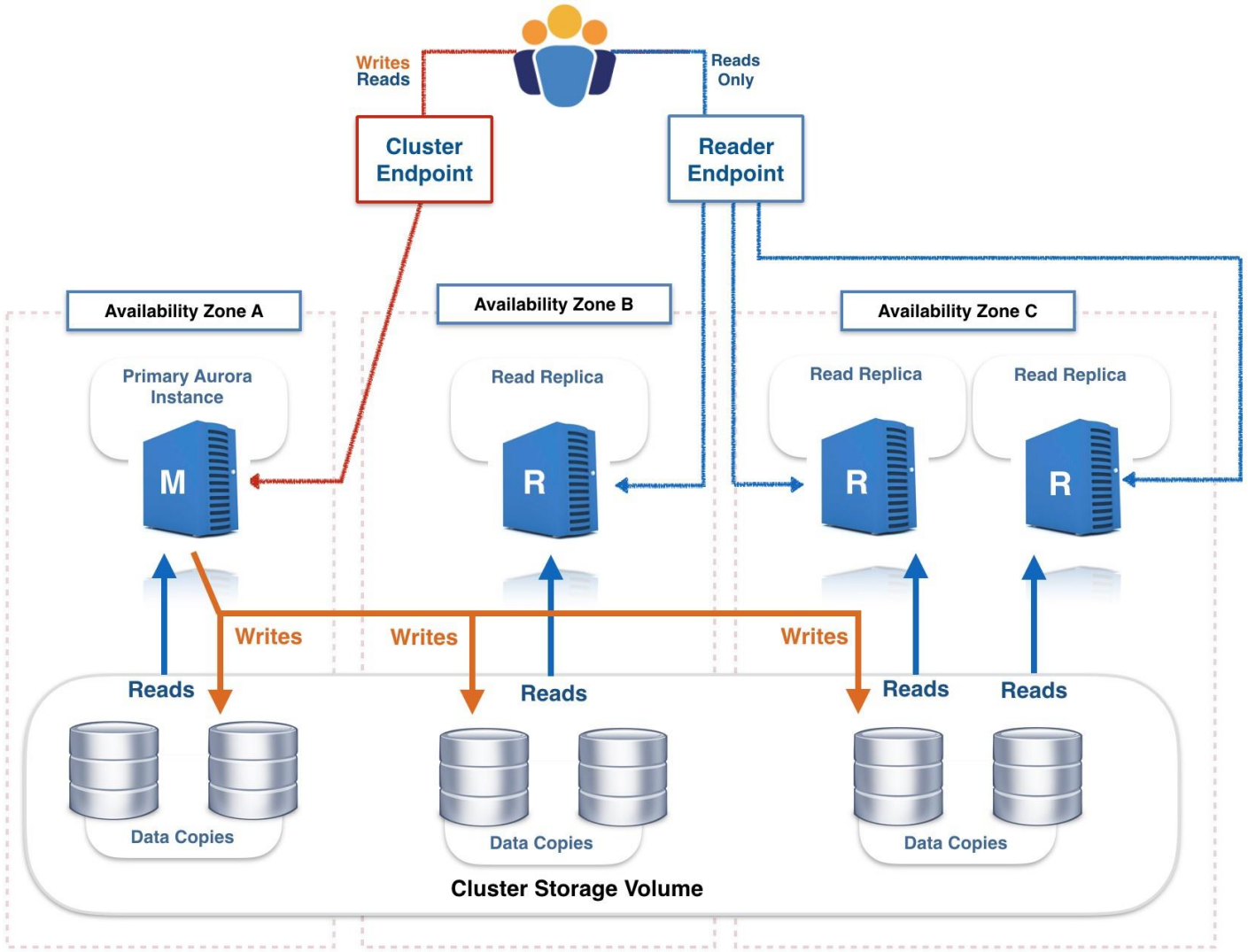
- Open source:

- Cloud native:

Amazon Aurora

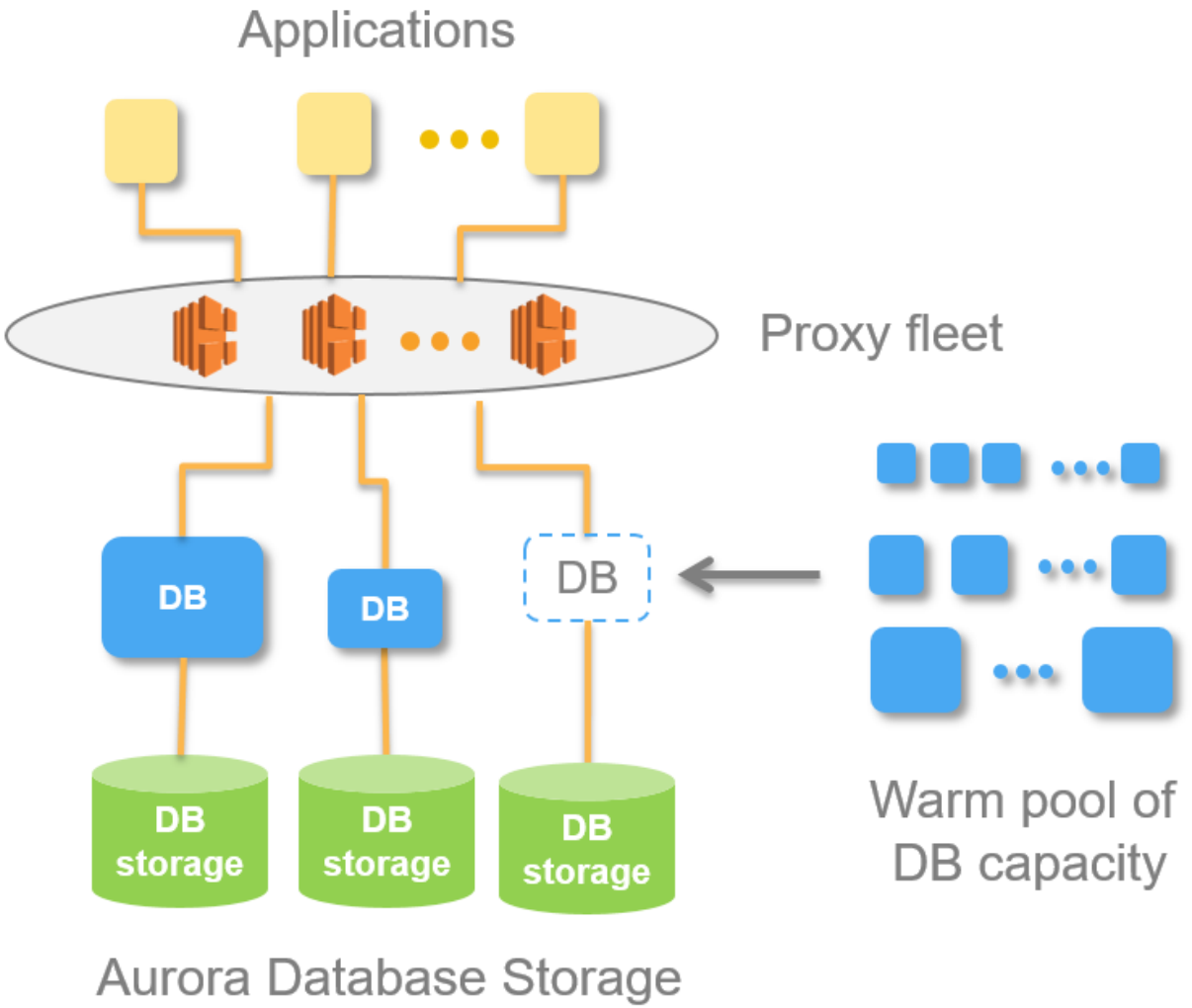
- cloud native, scalable, distributed RDMS engine developed by AWS
- MySQL (5.6, 5.7) és PostgreSQL (9.6, 10.4) interface
- **Min. 99.99% SLA, failover < 30 sec!**
- 15db read-replica
- Cross-region read-replica
- **AutoScaling: CPU vagy connection based**
- Automatic storage allocation, 64TB maximum

Amazon Aurora - Architektúra



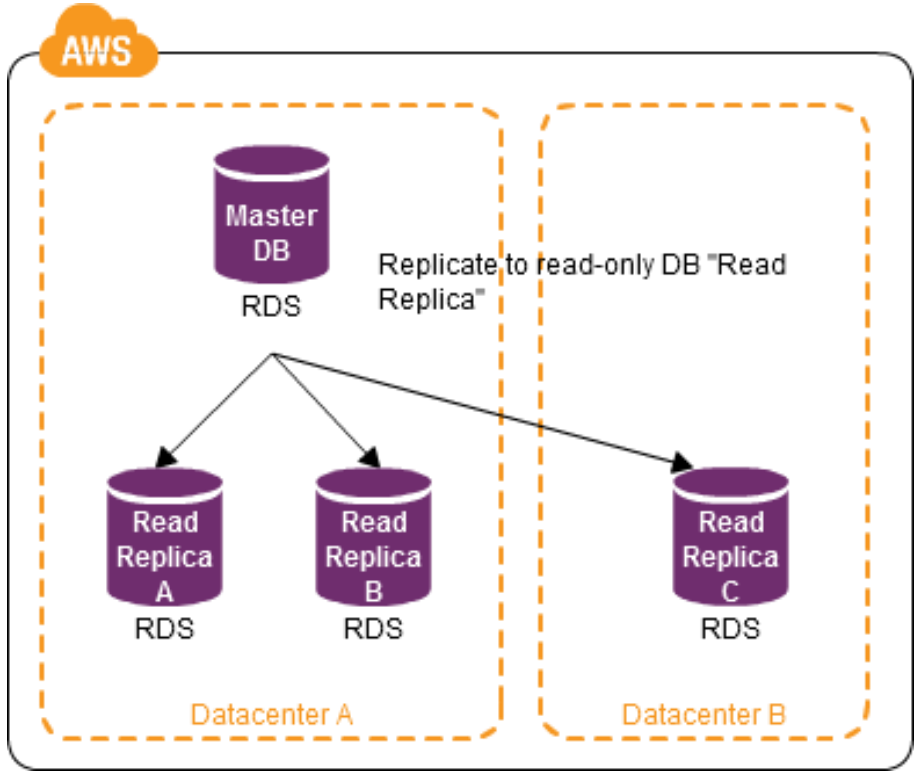
Amazon Aurora - Serverless

- On-demand, autoscaling configuration for Amazon Aurora
- Specify min. and max. values of capacity units (like autoscaling, but not instances)
- Aurora Capacity Units (ACU), **$0.06 per ACU** Hour
- MySQL 5.6 only
- Access available from AWS VPC only
- Automatically start-up, shut down, and scale up or down: **capacity based**

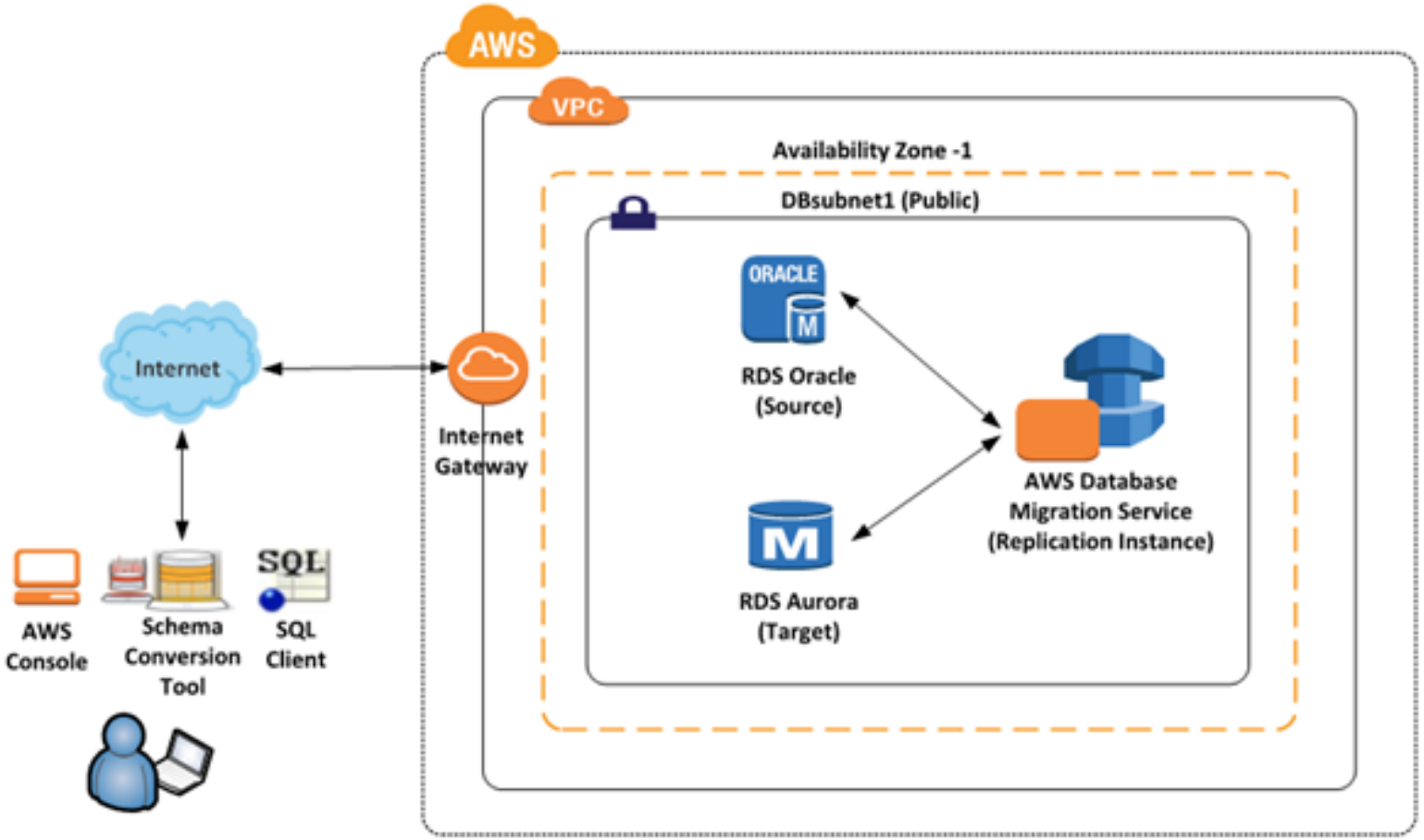


Read-replica

- AWS RDS: “egy kattintással indítható” read-replica
- **Ideális: reporting és statisztikák (ha az alkalmazás támogatja_RO végpontok / skálázás)**
- SQL server, Oracle vagy más adatforrás: DMS
- Read-replica előnyei:
  - Asszinkron – nem terheli a source tranzakciókat
  - few seconds lag (Aurora: ms range)
  - Pay-as-you-go
  - Cross-region
  - Egyszerűen indítható, leállítható
- Hátrány: a source DB tulajdonságait örökli



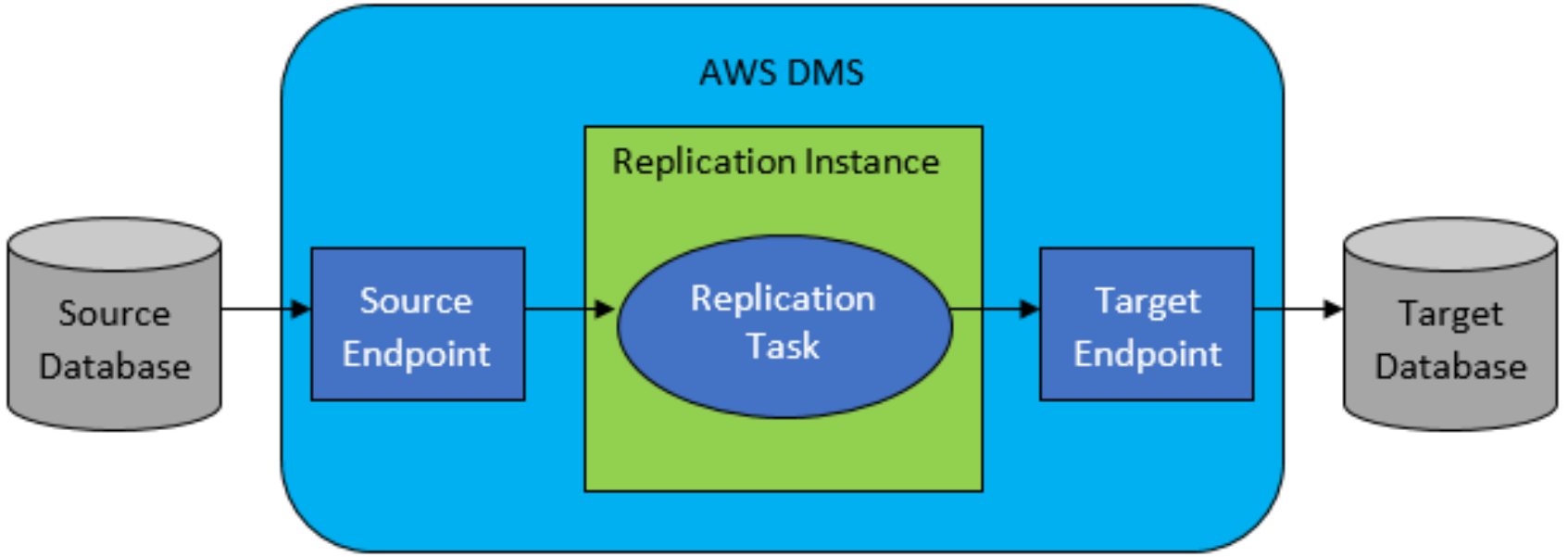
Amazon Aurora – DB migrálás



- MySQL: native mysqldump or DMS
- Schema Conversion Tool (SCT),
- Database Migration Service (DMS)
- Database Migration best practice, Oracle to Aurora PostgreSQL, playbook

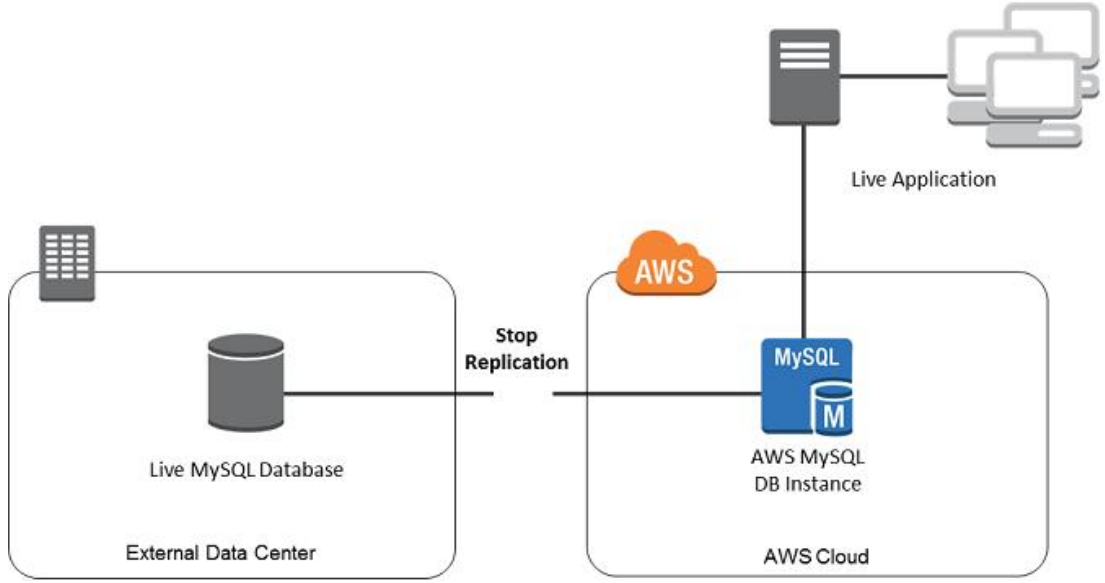
Amazon Database Migration Service (DMS)

- A scalable cloud service for RDBMS, Redshift and NoSQL continuous data migration and synchronization



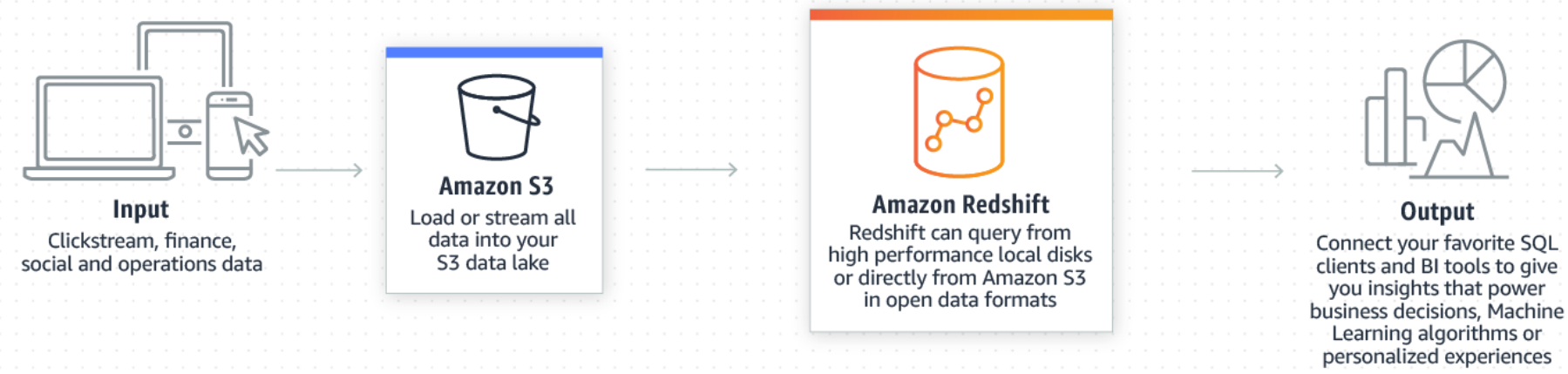
Amazon Database Migration Service (DMS)

- **Sources**: ORACLE, Azure SQL, MS-SQL Server, PostgreSQL, MySQL, SAP ASE, MongoDB, IBM DB2, AWS-S3
- **Targets**: ORACLE, MS-SQL Server, PostgreSQL, MySQL, Redshift, SAP ASE, DynamoDB, AWS-S3
- One-time setup, continuous sync
- Schema conversion support with SCT
- Scale up and down migration resources
- Automatic failover



Amazon Redshift

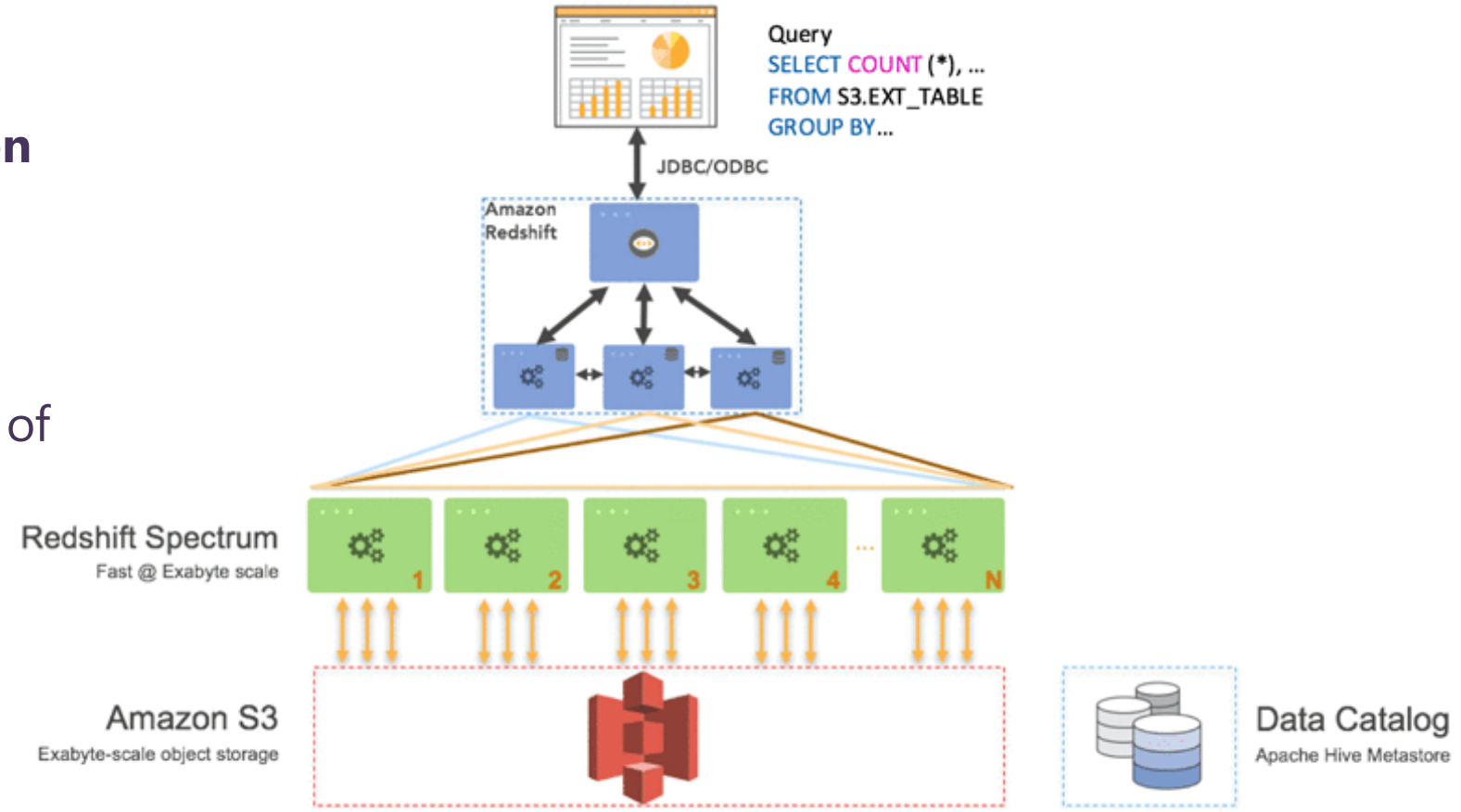
- Fast, scalable SQL data warehouse (interface: PostrgeSQL / JDBC)
- **Petabyte scale, massively parallel query execution,** columnar storage
- **Machine learning:** Amazon Redshift uses machine learning to deliver high throughout based on your workloads
- **Data lake**: query open file formats you already use, such as **Avro, CSV, Grok, JSON, ORC, Parquet, and more, directly in S3**
- **Tableau / PowerBI support**



Amazon Redshift Spectrum

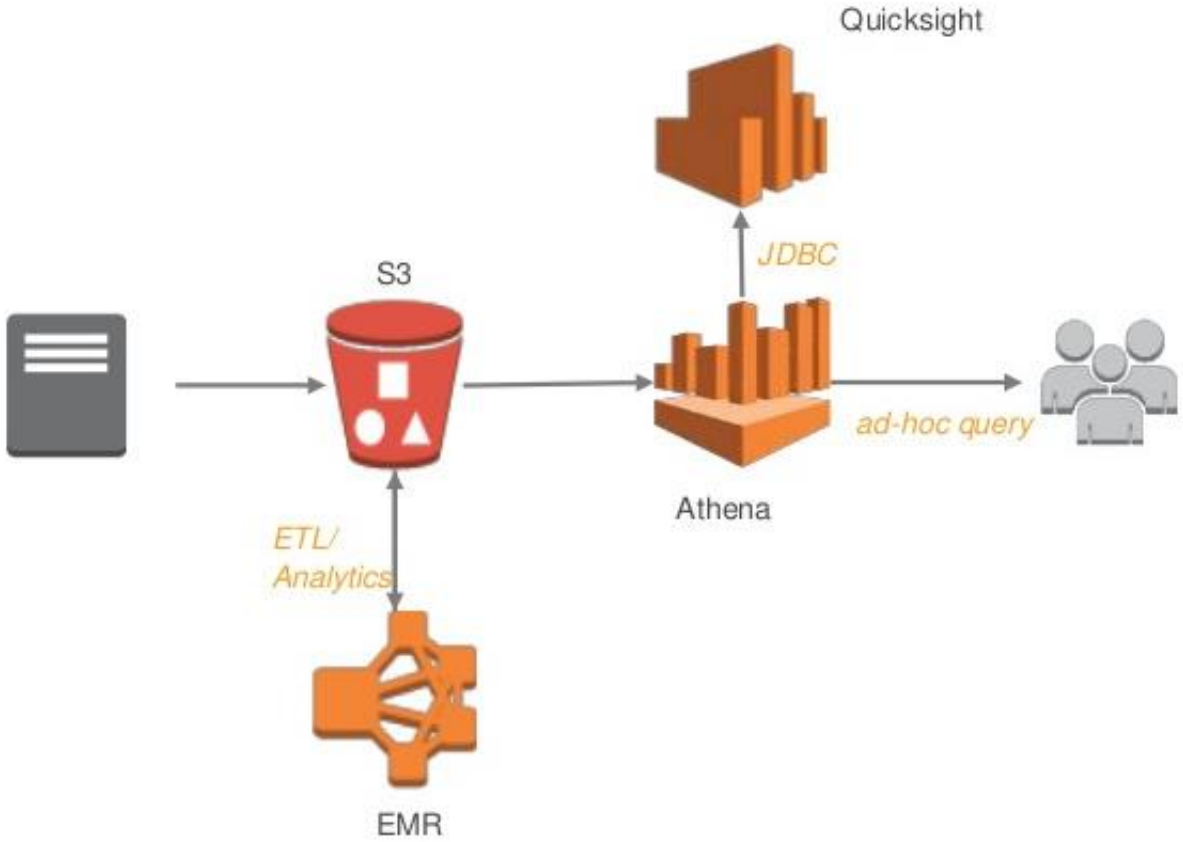
amazon REDSHIFT

- **Serverless** solution: but ***AWS Redshift is necessary*, query on** S3 structured data files, Apache Hive / AWS Glue data catalog
- Pay-as-you-go: $5 per terabyte of data scanned by SQL query
- **Multi-exabyte scale!**
- **Table definitions:** Apache Hive Metastore / AWS Glue

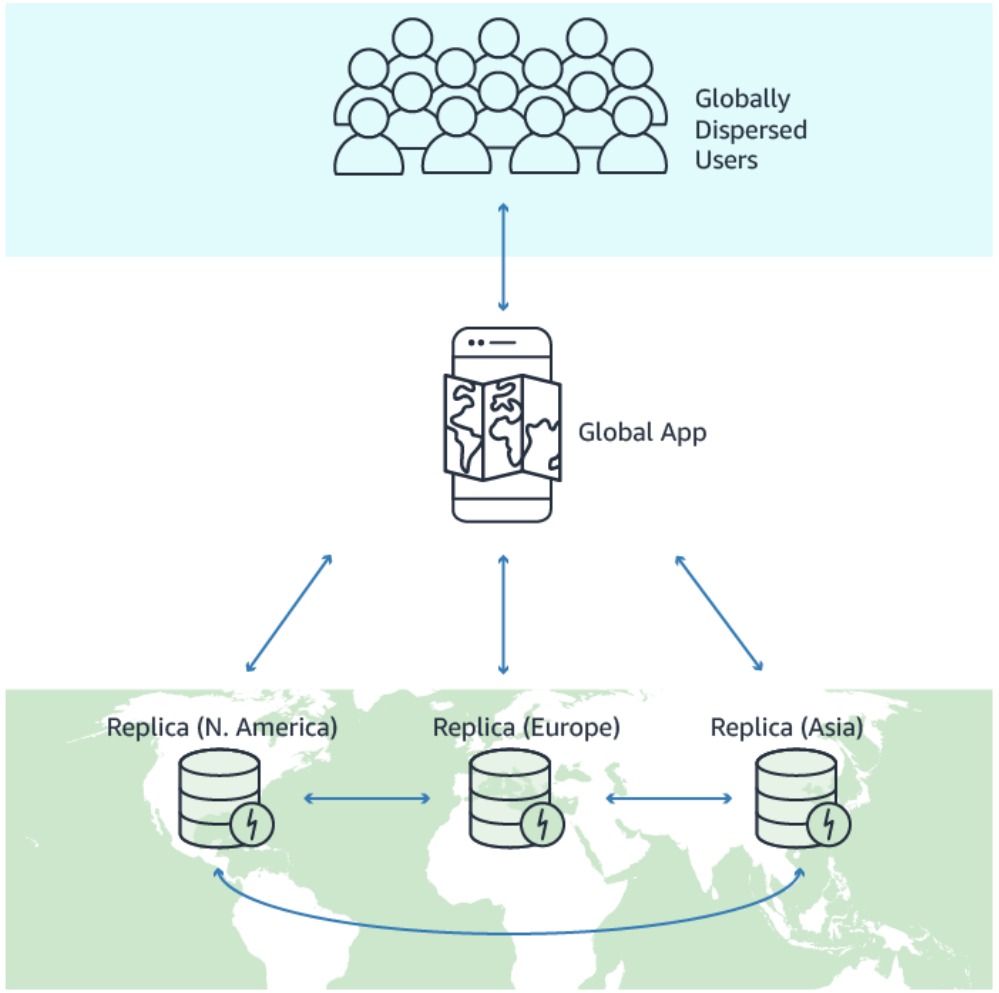


Amazon Athena

- **Serverless** solution: standard SQL query straight on S3 storage, AWS Glue data-catalog, ***Redshift is not necessary!***
- **Exabyte scale, zero spin-up time**
- **Compressed data support**
- **BI tools:** Tableau, JDBC, API…
- **Pricing:** $5 per TB of data scanned



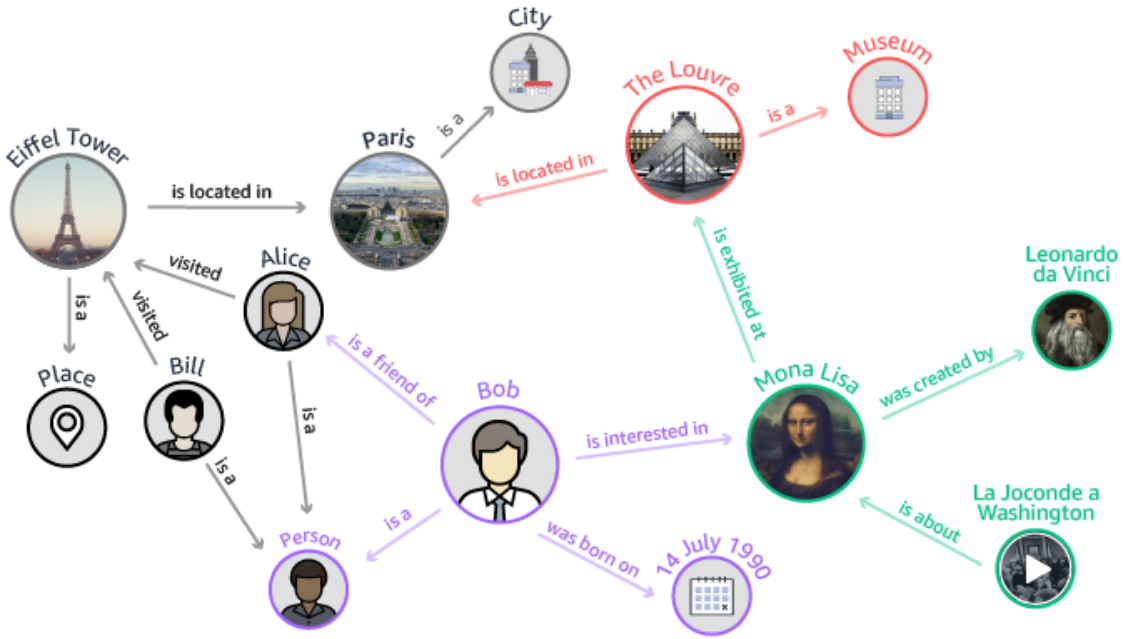
Amazon DynamoDB: NoSQL



- **Serverless** solution: fully managed, auto-scaled, distributed, schema-less NoSQL database
- **Virtually unlimited throughput and storage.**
- **Pricing**: data stored + throughput
- Single-digit millisecond responsiveness (**DAX**: microsecond responsiveness)
- **SLA:** minimum **99.999%!** *(less than 5minute / year outage)*
- **Multi-Region, Multi-Master tables**

Amazon Neptune: GraphDB

- Kakukktojás: “graph database” szolgáltatás (nem serverless és NEM SQL)
- **Open Graph APIs:** Apache TinkerPop Gremlin, Gremlin and SPARQL, RDF model support
- Amazon Aurora –hoz hasonló: magas rendelkezésre állás, elosztottság, read-replica, ACID, stb…
- **Előny**: nem szükséges komplikált SQL join-okat írni, a reláció közvetlenül lekérdezhető és tárolható
- Adat betöltése közvetlenül S3-ból: property graph, CSV, RDF, Turles, stb…
- **Up to 100,000s** of queries per second



ÖSSZEFOGLALÁS

- Strukturált SQL és NoSQL adattárolás AWS felhőn?
- A megoldások tárháza sokrétű és sokszínű.
- Egyszerű, olcsó, hatékony, megbízható, skálázható, igényekre formálható.