



Adattárház technológiák az AWS-ben

Stadler Gellért

2024.09.10.



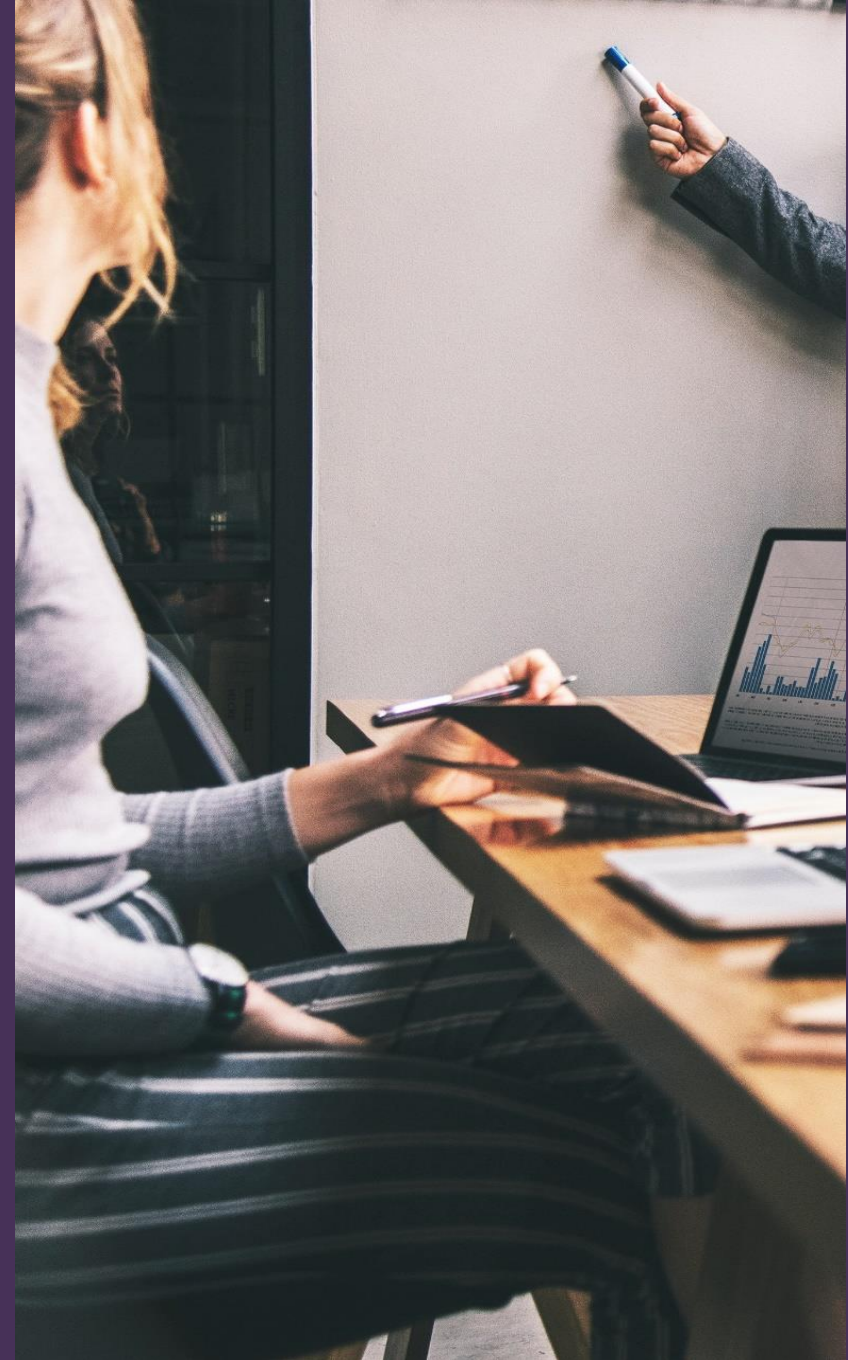
Tartalom

- AWS Modern Adatstratégia
- AWS Referencia Architektúra
- AWS Redshift



Miért van rá szükség?

A hagyományos, on-premise architektúra egyre kevésbé képes költséghatékonyan kielégíteni a gyorsuló ütemben keletkező üzleti igényeket. A vállalatok adat és információ iránti éhsége jelentős költségnövekedéssel jár adattárolási és feldolgozási funkcióknál egyaránt. Egy költség-hatékonyabb, **felhasználás alapon árazódó, rugalmasan skálázható** felhő infrastruktúra bevezetésével a vállalatok olyan adat stratégiát alakíthatnak ki, amely lehetővé tesz számukra releváns üzleti információk előállítását a korábbinál gyorsabban és olcsóbban.



Számszerűsíthető előnyök

Infrastruktúra költségek csökkenése

- Beruházási költség
- Amortizáció
- Üzemeltetési költségek
- Leállási idő költsége

Skálázódási és integrációs költségek csökkenése

- Azonnali skálázódás (lefelé is!)
- Csúcsidőszakok kezelése
- Várakozási idő nélkül

Hatékonyság növelés „beépített” AWS funkciókkal

- Integrációs adapterek
- Automatizáló eszközök
- Algoritmusok és modellek

Egyéb előnyök

Data Governance és Compliance

- Sarbanes-Oxley
- FedRamp
- ...

Adatbiztonság

- AWS security szolgáltatások
- Alert és Log kezelés

Alkalmazotti elégedettség

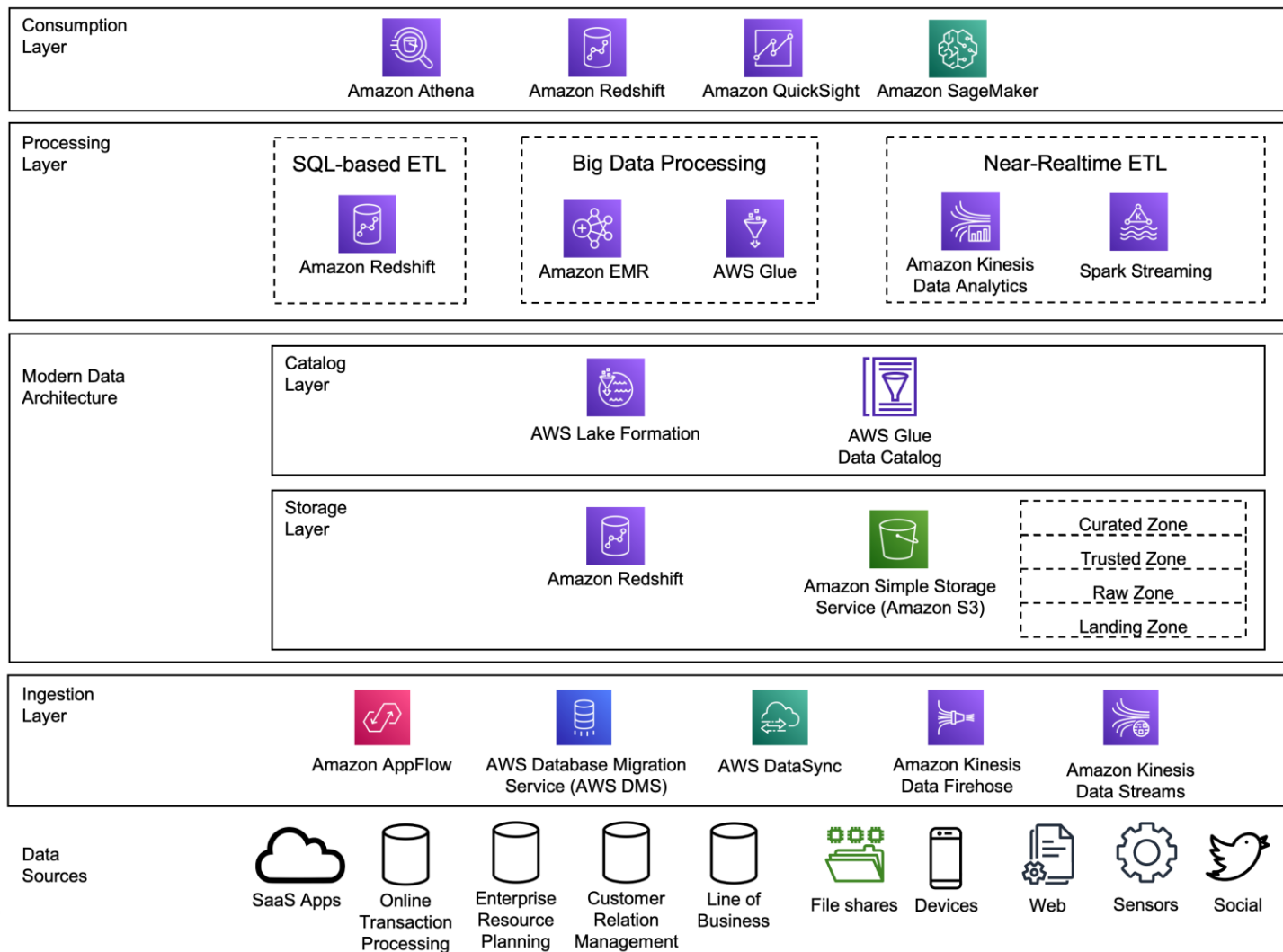
- Gyorsabb, hatékonyabb munkavégzés
- Új képességek elsajátítása
- Karrier út

Új technológiák adoptálása

- Kész háttérrel érkező új technológiák, zöldmezős fejlesztés helyett
- Alacsony bekerülési költséggel
- „Cloud First”



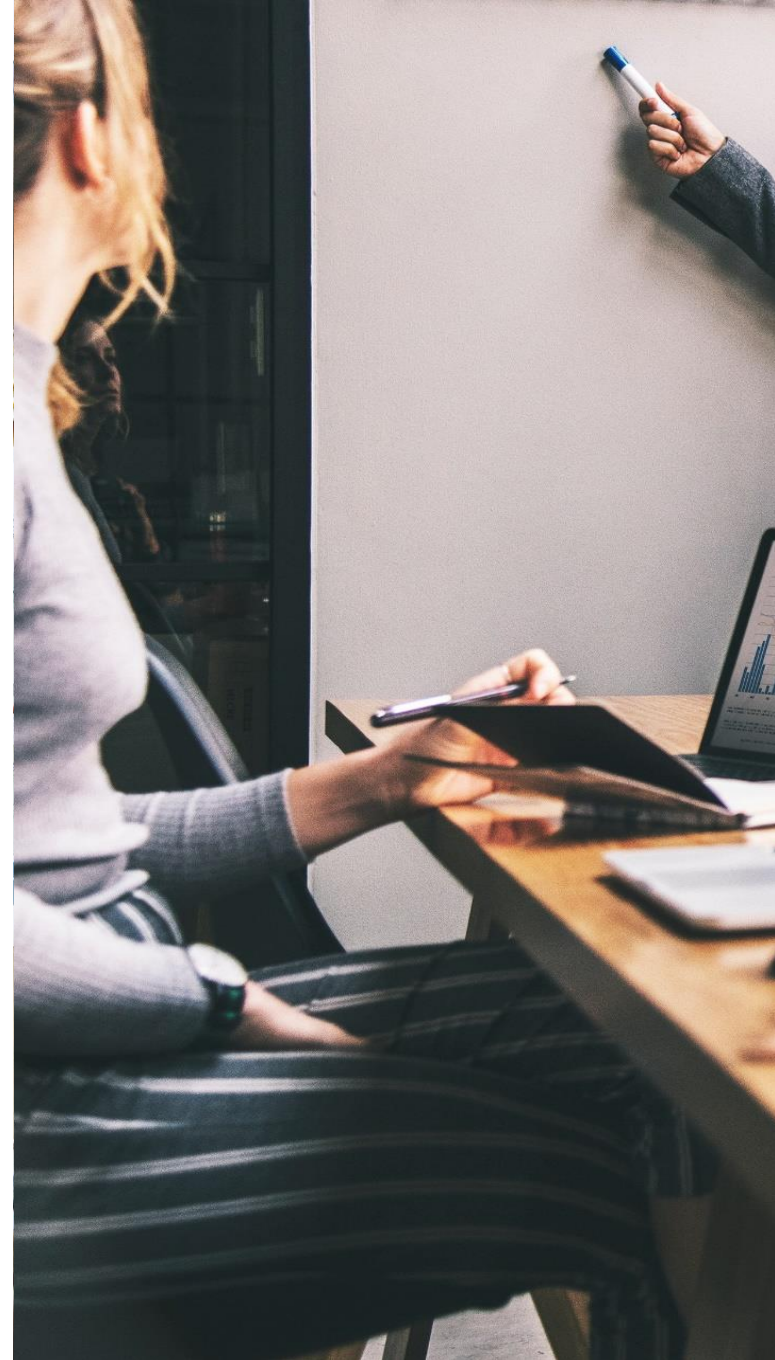
AWS Referencia Architektúra





Mi az AWS Redshift?

- **Menedzselt adattárház** szolgáltatás, amely exabyte-os (10^{18} bytes) méretekre skálázható:
 - Automatikus provisioning
 - Automatikus backup és replikáció
 - Automatikus workload management
 - Monitoring
- Hagyományos relációs adatbázisokhoz képest akár 10x jobb teljesítménnyel:
 - Masszívan párhuzamos architektúra (adattárolás és lekérdezés végrehajtás)
 - Oszlopos tárolás
 - Adattömörítés
- SQL / ODBC / JDBC kompatibilitás
- **„Serverless” adattárház szolgáltatás:** Redshift Serverless



Tipikus Redshift használati esetek

ADATTÁRHÁZ
MODERNIZÁCIÓ

EGYESÍTETT
ADATTÁRHÁZ ÉS
DATA LAKE

PÉNZÜGY /
TŐZSDE

TÁVKÖZLÉS / IoT

CLICKSTREAM /
AD IMPRESSION

LIFE SCIENCE

GAMING /
SOCIAL

Fontosabb Redshift funkciók

Integráció

- PostgreSQL alapú SQL
- Redshift Spectrum
- Zero-ETL
- Streaming ingestion
- Federated Query

Adatkezelés

- Komplex adattípusok (SUPER, VARBTYPE, Geometry, Geography)
- Lambda Functions
- Materializált nézetek

ML & AI

- Model készítés,
 - training,
 - használat
 - SQL nyelven belül
- Külső modellek használatának lehetősége
- Automatic Workload Management

Használati modellek

- **Serverless:** 100% felhasználás alapú árazás, automatikus skálázódás limitekkel
- **Provisioned:** Folyamatosan lekötött kapacitás, automata vagy kézi skálázódással



Versenyképes árazás

Node-ok száma	Tárolási Kapacitás	Havi díj
1 (dc2.large)	160 GB	~210 USD ¹
8 (dc2.large)	1,2 TB	~1700 USD ¹
4 (ra3.xplus)	100 TB	~6000 USD ¹
Serverless 8 RPU	50 TB	~2600 USD ²
Serverless 16 RPU	100 TB	~5200 USD ²

¹Provisioned mode, 90%-os kihasználtsággal számolva

²Napi 12 óra működéssel számolva



Miért a TC2?



Tanúsítványaink lefedik az összes lehetséges területet az AWS palettáján.



Köszönöm a figyelmet!

Stadler Gellért
gellert.stadler@tc2.hu