



# Adattárház és Big Data

Definíciók, különbségek, gyakorlati hasznosítás

Bánki Zsolt

[banki.zsolt@virzrt.hu](mailto:banki.zsolt@virzrt.hu)



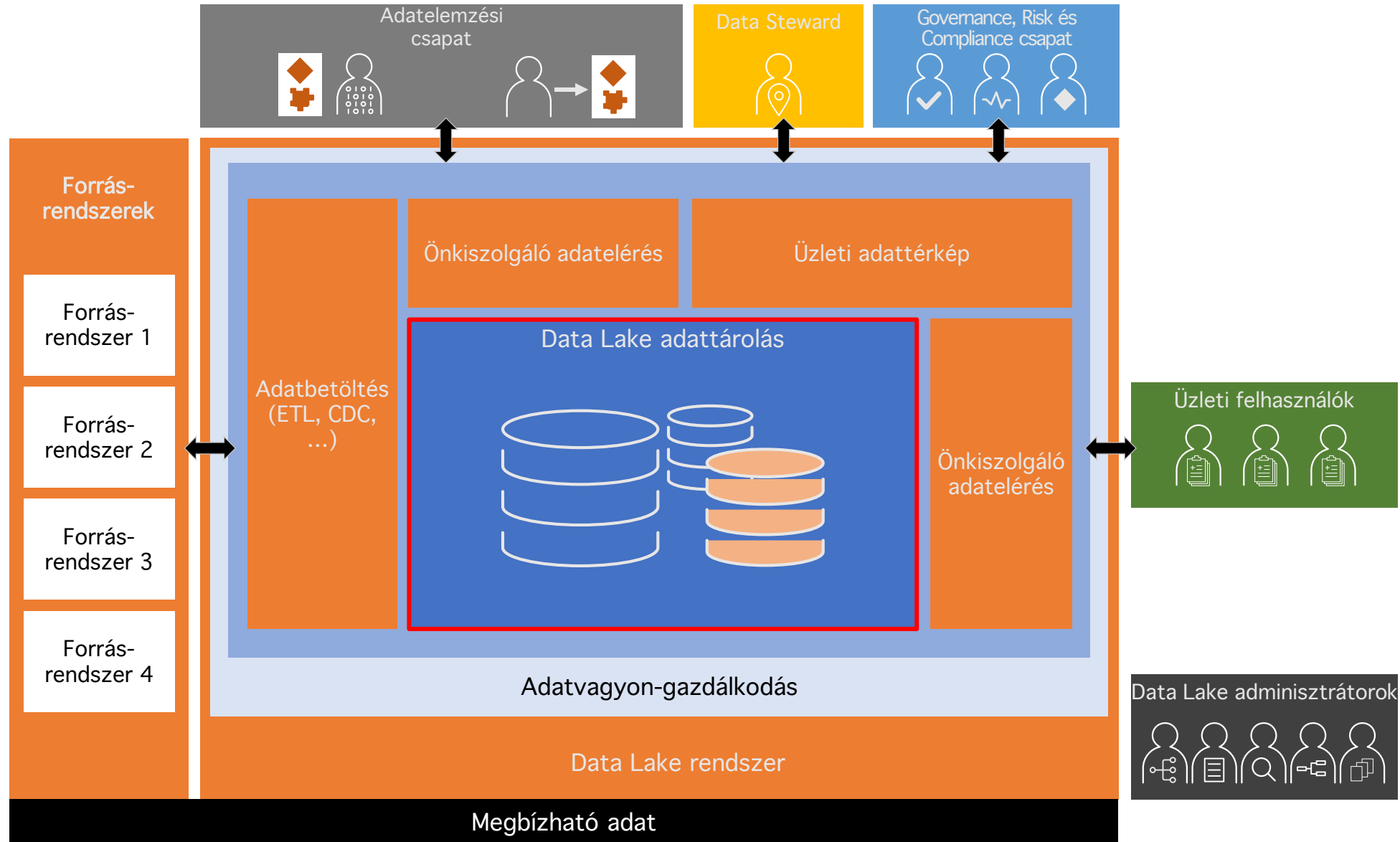
# Adattárház

- Az adattárház rendszerek lehetővé teszik a vezetői igényekre **optimalizált** sok adatot **összevontan** kezelő elemzői / vezetői lekérdezéseket és a **személyre szabott gyors riportolást**.
- Az összvállalati szintű adatkezelést a központi adattárház a kapcsolódó **adatpiacok integrált rendszerével** együtt látja el, melyek egy-egy szervezeti egység speciális információs igényeinek megfelelően optimalizáltak.
- Amikor az adatokat adattárházba töltjük be, az **ELT**) rendszer elolvassa a nyers adatokat, kinyeri az információkat, **átalakítja** (például az SQL táblák formátumainak való megfeleléshez), majd továbbítja azokat
  - az adattárházakban soha nem fogjuk az eredeti, nyers adatok egy példányát megtalálni, csupán a tartalom egy „szabványos másolatát”.
  - Információ kérésekor (például egy üzleti alkalmazás lekérése esetén) mindig az **előzetesen formázott és előre strukturált adatok** töltődnek be.

# Big Data

- Az adattárházban is „Big Datát” tárolunk – sok adatot. Nem a méret számít, hanem a tárolás és feldolgozás módja különbözteti meg az adattárházat a „Data Lake”-től.
- Új fogalom „Big Data” helyett: „Data Lake” (adat-tó)
  - Rendezetlen felszín alatti ökoszisztéma
- A Data Lake rendszer elve az, hogy az összes adatot adatfájlok formájában tárolja a merevlemezeken (nem feltétlenül rendezett SQL adatbázisokban).
- Amikor egy adattavat töltünk fel, a belső adatkezelő csak nyers adat formátumot másol a lemezünkre. Ezen a ponton nem történik adatkivonás vagy -módosítás. Az adatok igény szerint (on-demand) – valamilyen ügyfélalkalmazáson keresztül – kerülnek majd feldolgozásra (például indexelésre, formázásra vagy átalakításra).
- Data Lake adatmenedzsment rendszere a Hadoop.

# Data Lake adattárolás - Hibrid Adatmenedzsment



# Hibrid Adatmenedzsment - Adattó és Adattárház

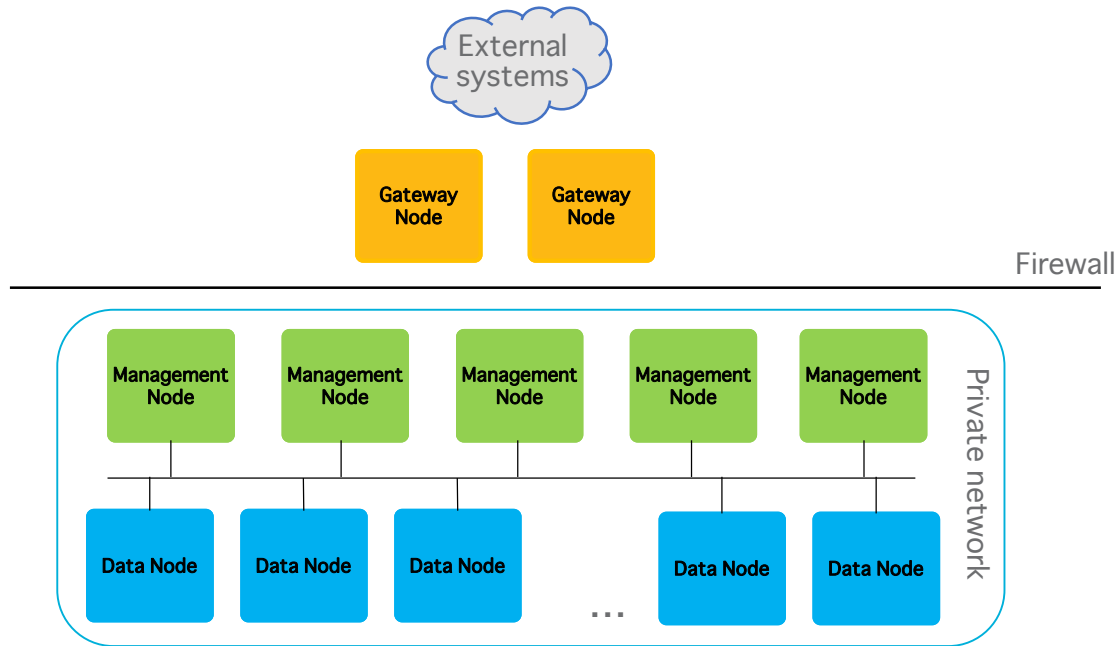
	Adattó - Hadoop	Adattárház
Adattárolás	Strukturált, részben és nem strukturált és akár nyers adatok.	Strukturált, előfeldolgozott és üzletileg legkritikusabb adatok.
Esettanulmányok	IoT analitika, adattárház offloading, adatelemzési homokozó (adatfelderítés, adatbányászat)	Üzleti elemzések, riportok komplex BI elemzés
Felhasználók	Adatbányász, tapasztalt üzleti felhasználók	Üzleti felhasználók riportokon keresztül
Elemzési interfész	SQL, Java, Spark (Scala, Python, R), stb.	SQL
Tárolókapacitás	Olcsó, HDD alapú tárolók és Object Storage	Gyors, flash alapú tárolás
Elemzés válaszidő	IoT analitika: közel valós idejű Komplex elemzés: több perc, óra	Pár másodperc
Adatbiztonság	Kiforróban, sok szolgáltatás menedzselése szükséges	Kiforrott

..és tovább:  
adattárház  
modernizáció,  
kiegészítés

# Adattárház kikönnyítés, tárkapacitás csökkentés

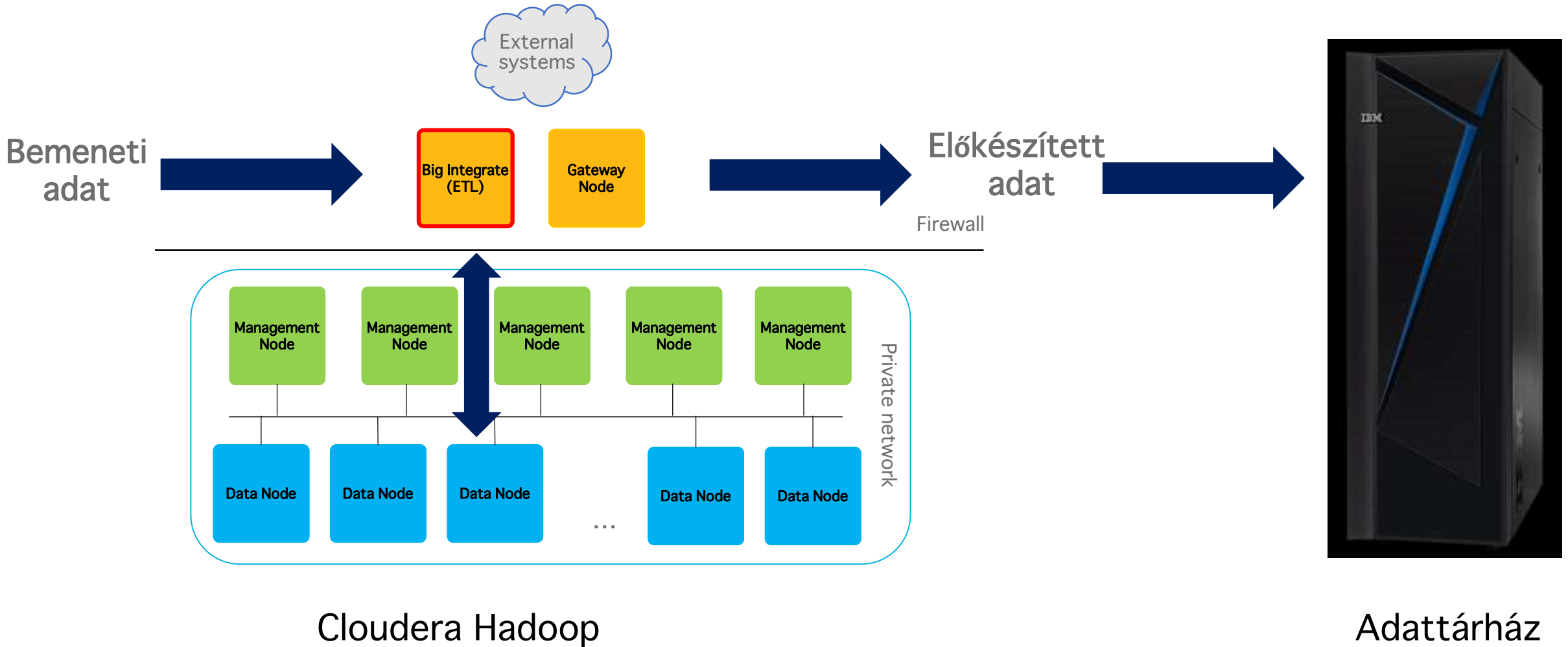


Abstraction layer – Common SQL



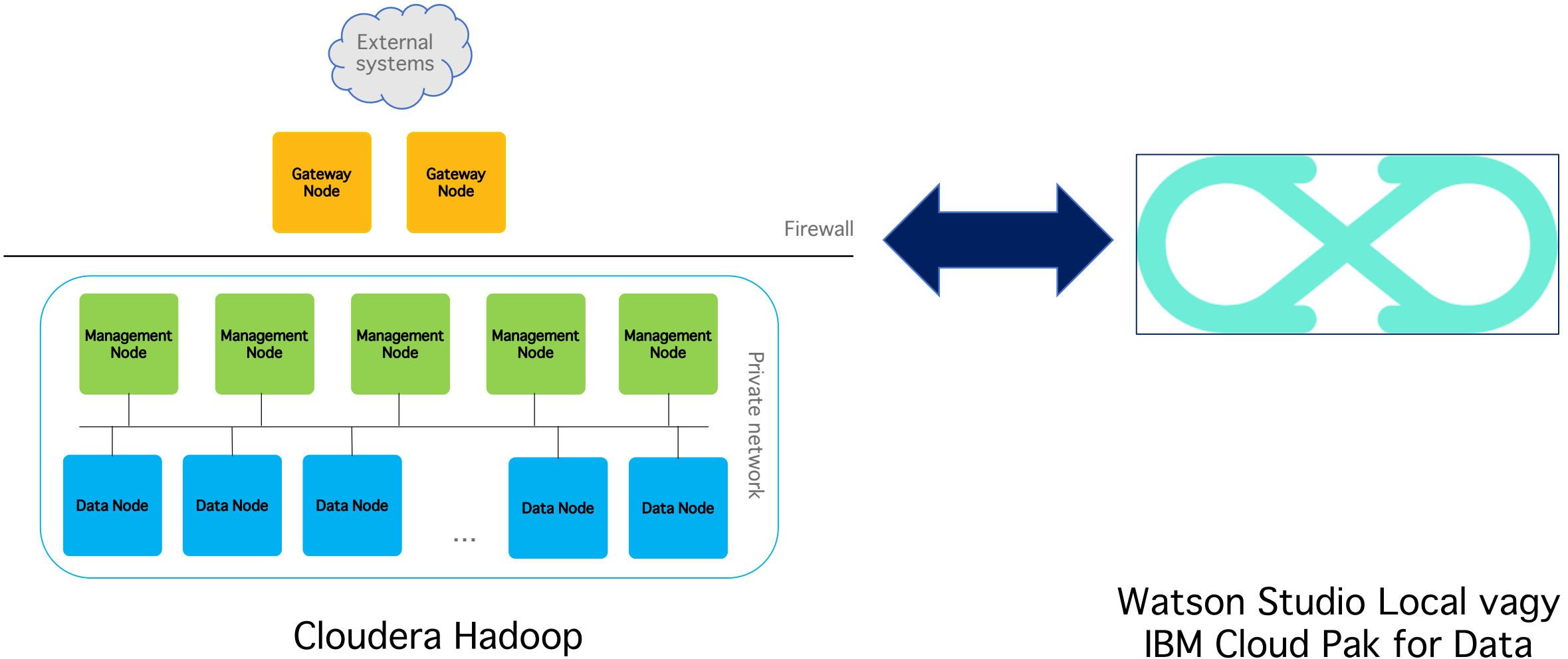
Adattárház

# Adattárház kikönnyítés, ETL betöltés gyorsítás





# Hadoop, mint adatelemzési homokozó



# Köszönöm a figyelmüket!

Bánki Zsolt

[banki.zsolt@virzrt.hu](mailto:banki.zsolt@virzrt.hu)

