

Hibrid adattárházak felhasználási lehetőségei

2020. November 10.

Előadó: Erdey Levente

Adattárház régen és most

Data Lake – Hadoop alapokon

Big Data felhasználás

Integráció

Építkezés

Megéri?

Lehetőségek

Új idők, új kihívások – Adattárház régen és most

"A data warehouse is a subject oriented, integrated, nonvolatile, and time variant collection of data in support of management's decisions.,, - Bill Inmon

- Tematikus
- Integrált
- Tartós
- Idősoros

Új elvárások

- Gyorsabb fejlesztési ciklus
- Félig vagy nem strukturált adatok feldolgozása
- Nagy mennyiségű adat befogadása, feldolgozása
- Ad Hoc adat betöltések
- Data Discovery
- Költségek csökkentése
- „Nekünk is kell egy Big Data!”



Data Lake – Hadoop alapokon

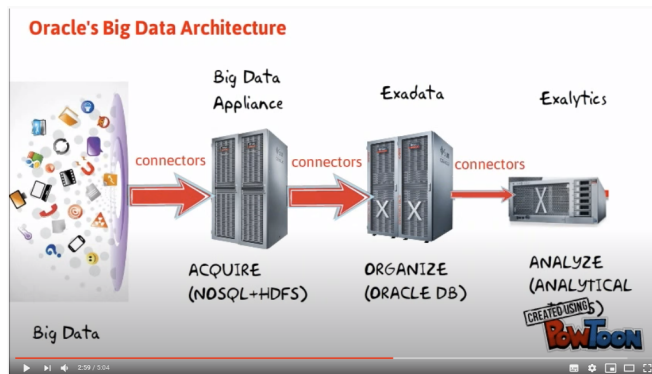
What have the Big Data ever done for us

Új elvárások

- Gyorsabb fejlesztési ciklus
- Félig vagy nem strukturált adatok feldolgozása
- Nagy mennyiségű adat befogadása, feldolgozása
- Ad Hoc adat betöltések
- Data Discovery
- Költségek csökkentése
- „Nekünk is kell egy Big Data!”

Big Data megoldások

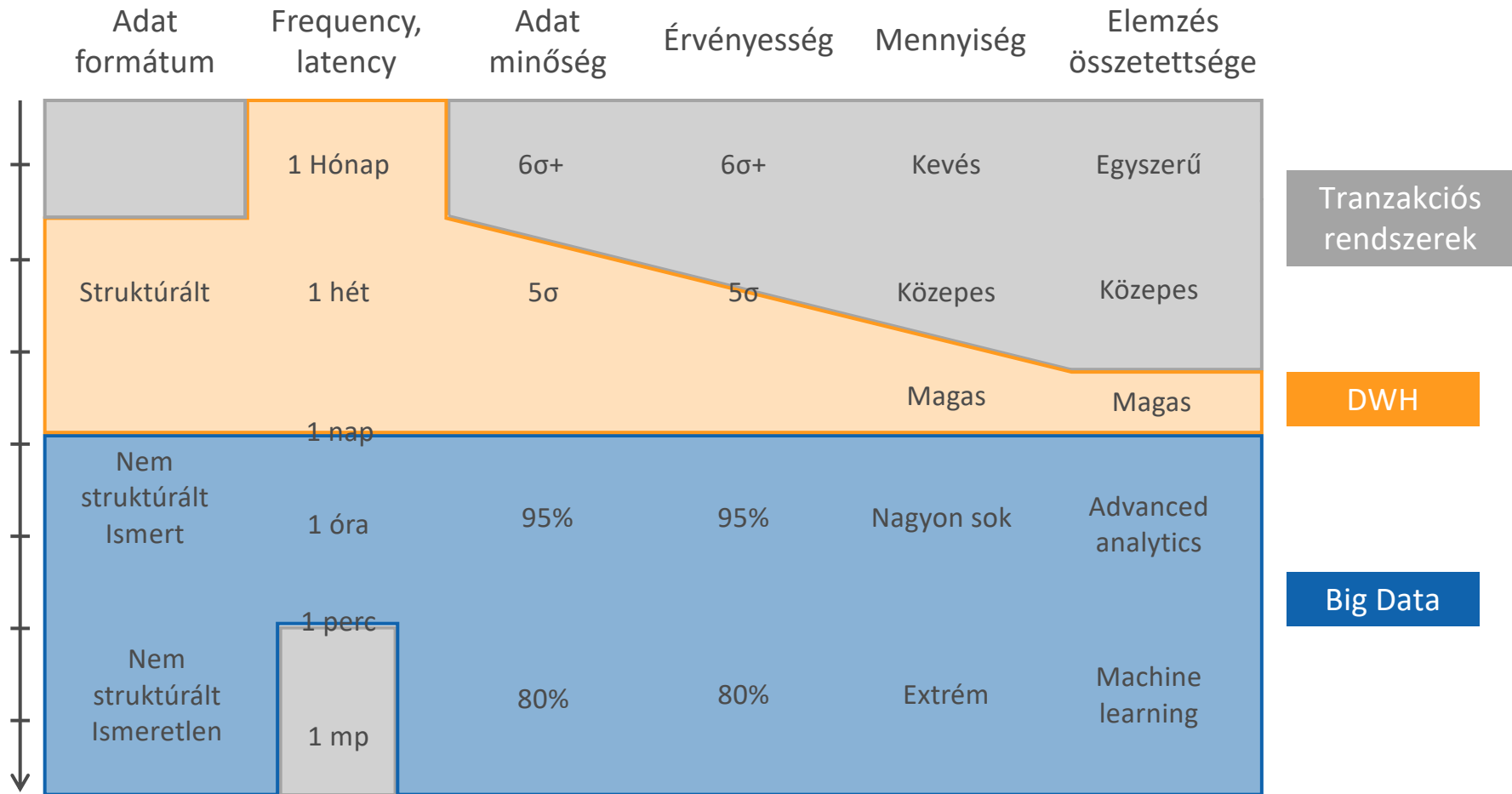
- ➔ Ez kisebb részben technológiai kérdés
- ➔ Több megoldás a félig vagy nem strukturált adatok feldolgozására
- ➔ Egyszerű befogadás, és elosztott tárolás és feldolgozást biztosított
- ➔ Egyszerű befogadás
- ➔ Csak támogatja: nagy mennyiségű adattal is gyors, rugalmas
- ➔ Ennek mindig utána kell számolni, nem egyértelmű
- ➔ Erre a tökéletes megoldás!



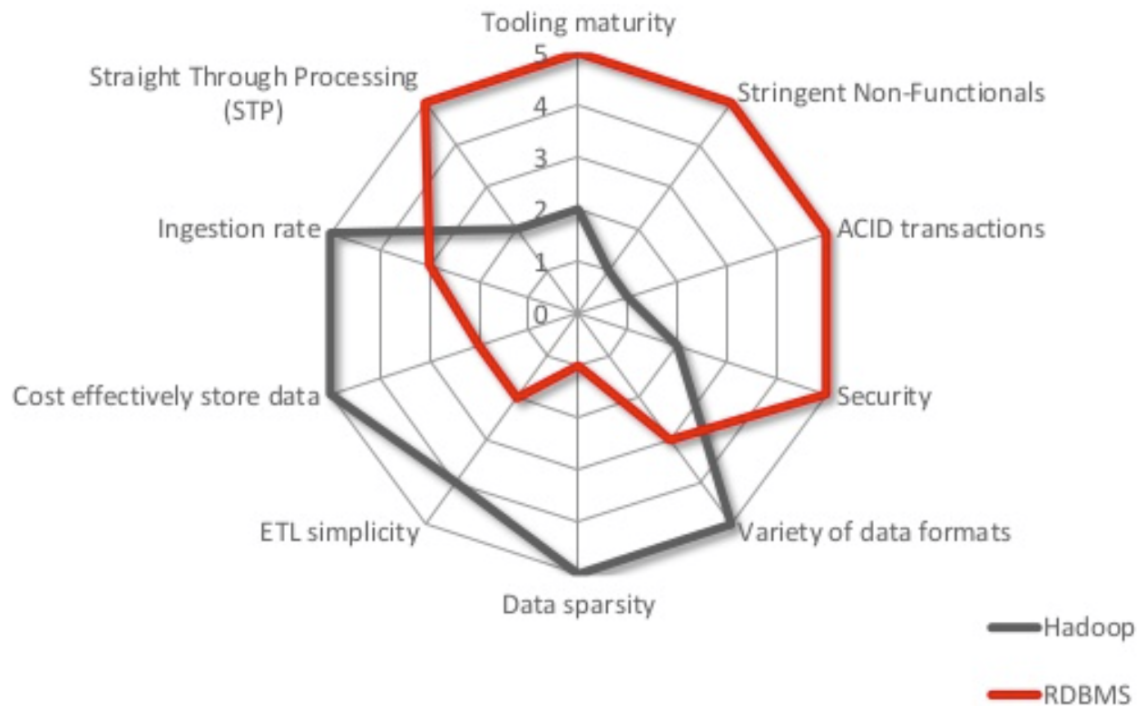
Ne maradjon ki

- IT Biztonsági integráció
- Fejlett adattárház integráció
- Egyszerű üzemeltethetőség
- Magas rendelkezésre állás
- Könnyű felhőbe tenni

Big Data felhasználás



Big Data felhasználás



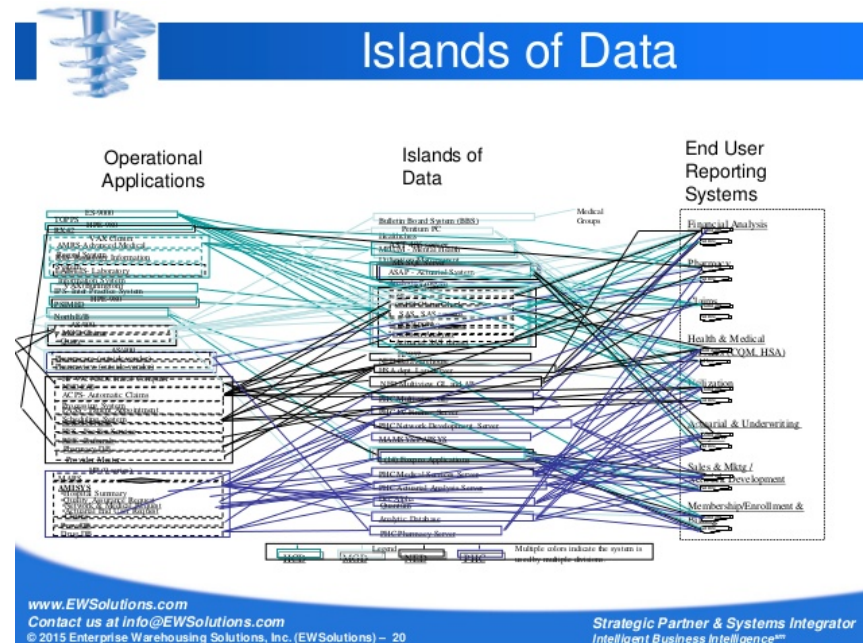
- Mindegyik megoldás másban erős
- A kettő együtt ütőképes
- Szükséges az integráció

Integráció

Stratégia nélkül nem fog menni

Mit kell tartalmaznia a Stratégiának - Minimum

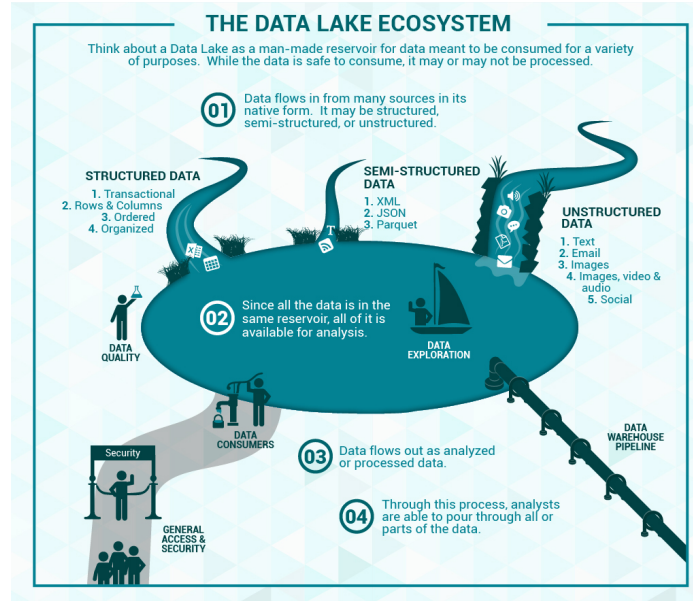
- Integrációs tervezés
- Adatbefogadás és feldolgozás
- DWH / Data Lake határvonalak, és azok átlépése
- ETL helye, funkciója
- TCO költségek tervezése
- Adatvagyon felmérése, és kihasználási lehetőségek
- Üzemeltetés tervezés és oktatás
- BI szakértők oktatása



Építkezés

Greenfield

- Egyszerűbb
- Átláthatóbb
- Drágább
- DWH integrációval egybekötve
- A DWH projektek sikeressége alacsony



- Helyes scope meghatározása
- Stratégia

Brownfield

- Integrációs kihívások
- Csodavárás
- Nem megfelelő lépésekben történő átállás veszélyes
- DataFlow management

Megéri?

Előnyök

- Olcsóbb DWH réteg is elegendő
- Store everything – tényleg kell ez?
- Olcsó számítási és tárolási kapacitás
- Lehet nagyon egyszerű



Hátrányok

- Plusz egy rendszer (üzemeletetés, bonyolultság)
- Azért vannak hardware/licenz költségek
- Jogosultság kezelés bonyolult

Mégis egyre több helyen alkalmazzák, mert az adat érték, és tárolni akarják, feldolgozni.
A technológiában van lehetőség, és folyamatosan fejlődik

Lehetőségek



IGEEEEEN

- Kezelhetetlen adatmennyiséget
- Folyamatosan növekedés ígérete
- Jelentős adatvagyon van / lesz struktúrátlan formában
- Igény egyszeri nagy betöltésekre
- Streaming adatforrások megjelenése
- Távoli adatok összekapcsolása AI-al
- AI

Építsük DWH-t Data Lake-re?



Nem kellene erőltetni

- Pár TB adat évente, és hasonló növekedési tendencia marad
- Csak strukturált adatok tárolása
- Továbbra is (micro) batch jellegű töltésekkel kell számolni
- Jelenlegi riporting igények már nem változnak



Kérdések

T · · **Systems** ·

Let's power
higher performance

Köszönöm a figyelmet!