

Adattárház

↳ MI-hoz



Tartalom

01 → Bevezetés

02 → Az adat-
architektúra
jelenlegi állapota

03 → Az adattó
tárház
meghatározása

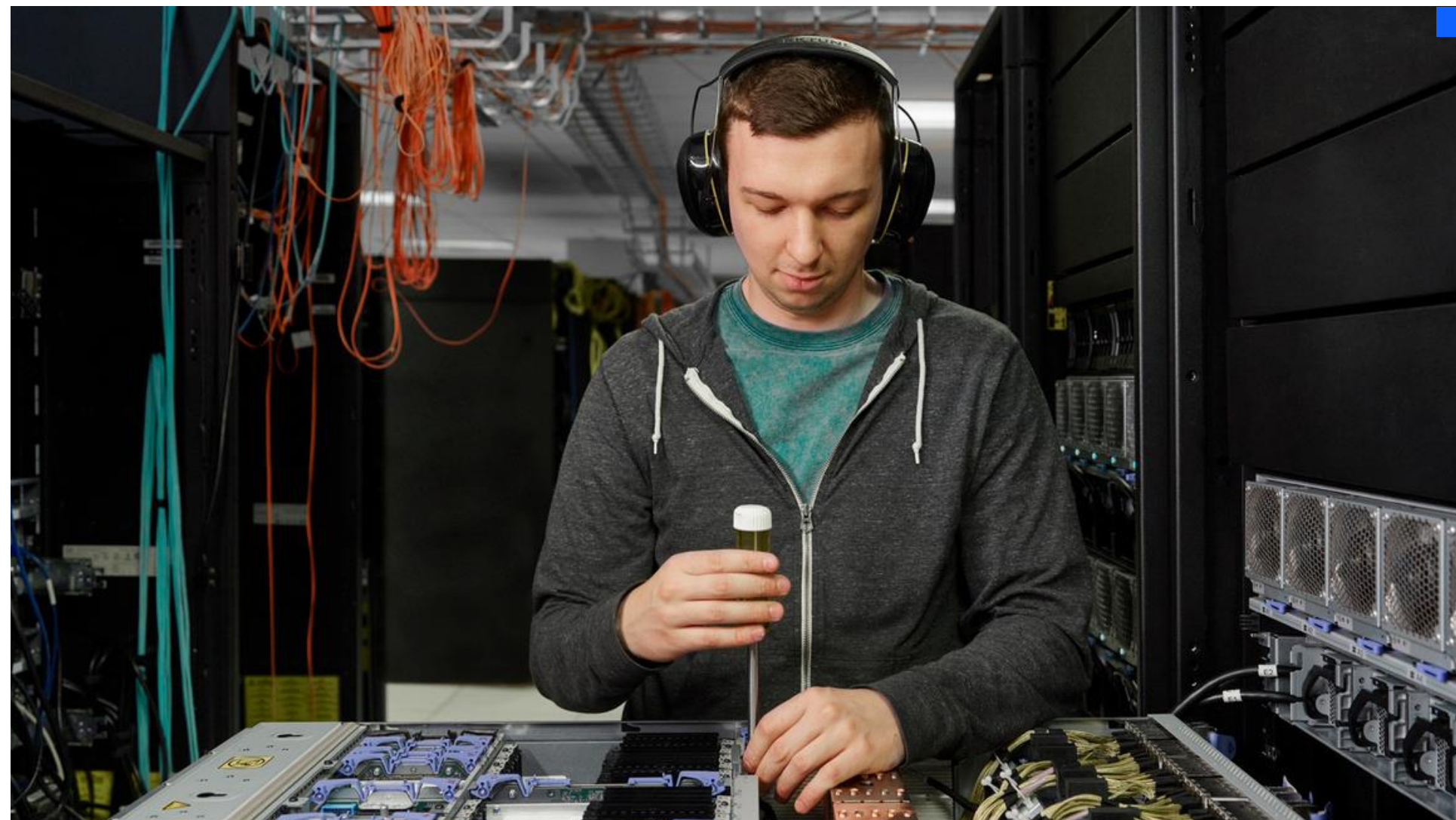
04 → Architektúra
elemek

05 → Költség-
optimalizálási
lehetőségek

06 → Analitikai és
adattudományi
fejlesztések

07 → IBM
watsonx.data

08 → A további
lépések



Ez az írás azokat az adat- és analitikai vezetőket szólítja meg, akik szeretnék csökkenteni a költségeket, könnyebbé tenni az adatokhoz való hozzáférést, és automatizálni az egységes irányítást a mesterséges intelligencia (MI) hatékony skálázása érdekében.

Az adatok minden vállalkozás szívében helyezkednek el. Az alkalmazások működését biztosítják, előrejelző elemzéseket tesznek lehetővé, és javítják a vásárlói és munkavállalói élményeket. Az adatok teljes kihasználása azonban gyakran elmarad, mivel a tárolásuk és az analitikai, illetve mesterséges intelligencia célú hozzáférésük nehézkes.

Ha Ön is monolitikus adattárakat használ, több adatbázissal és adattavakkal, helyben és a felhőben, nincs egyedül; a szervezetek 82%-a küzd az adatsilók okozta korlátokkal.¹ A helyzet csak romlik: az IDC előrejelzése szerint a tárolt adatok mennyisége 2025-re várhatóan 250%-kal nő.²

Az adattavak elméletileg megoldást kínáltak ezekre a problémákra: csak helyezze el az adatait egy központi helyen, és dolgozza fel őket. Azonban nem olyan egyszerű frissíteni az adattavakat, megfelelően katalogizálni az adatokat, vagy biztosítani a megfelelő irányítást.

Az ezekhez szükséges készségek specifikusak, ritkák és költségesek. Ennek eredményeként az adattavak kiépítése és fenntartása költségesnek bizonyult. Az adattárházak ugyanakkor magas teljesítményt nyújtanak terabájtnyi strukturált adat feldolgozásához. Az adattárházak is drágává válhatnak, különösen az új és változó munkaterhelések esetében. A legtöbb szervezet olyan ökoszisztémákban futtatja analitikai és MI-feladatait, amelyek bonyolultak és nem költséghatékonyak. Itt az ideje a változtatásnak.

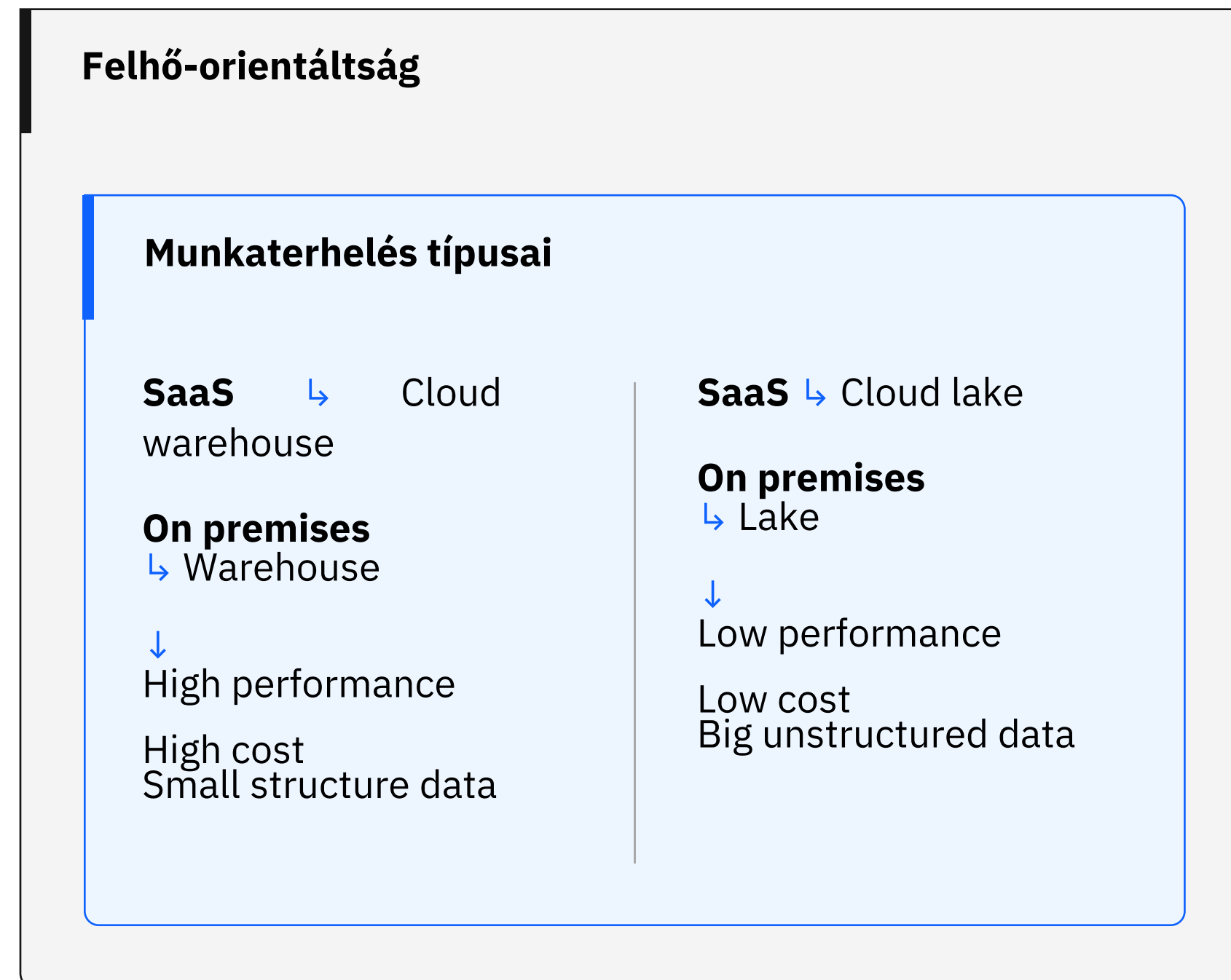
↑250%

A tárol adatok
mennyisége várhatóan
250%-kal nő 2025-re

02

Az adatarchitektúra jelenlegi állapota

A vállalati architektúrában ma már gyakori a helyhez kötött és a felhőalapú tárházak és a testreszabott adattavak kombinációja. Valószínűleg Ön is úgy találja, hogy a költségek, a silókban tárolt adatok és az adatkezelés közötti zsonglőrködés állandó kihívást jelent.



A Data Lakehouse egy
feltörekvő
paradigmaváltás amely,
megváltoztatja hogyan
tárják fel a vállalatok az
új felismeréseket.³

03

Az *data lakehouse* meghatározása

- Válasszon egy olyan tárház megoldást, amely modern adat-alapot biztosít az MI skálázásához

A *data lakehouse* egy feltörekvő architektúra, amely az adattárház teljesítményével és szerkezetével egy *data lake* rugalmasságát kínálja. A legtöbb lakehouse-megoldás nagy teljesítményű lekérdező motort kínál az alacsony költségű tárolás mellett metaadat-irányítási réteggel együtt. Az intelligens metaadatrétegek megkönnyítik a felhasználók számára a strukturálatlan adatok, például a videó és a hang, valamint a félig strukturált adatok kategorizálását és osztályozását, például XML, JSON és e-mailek.

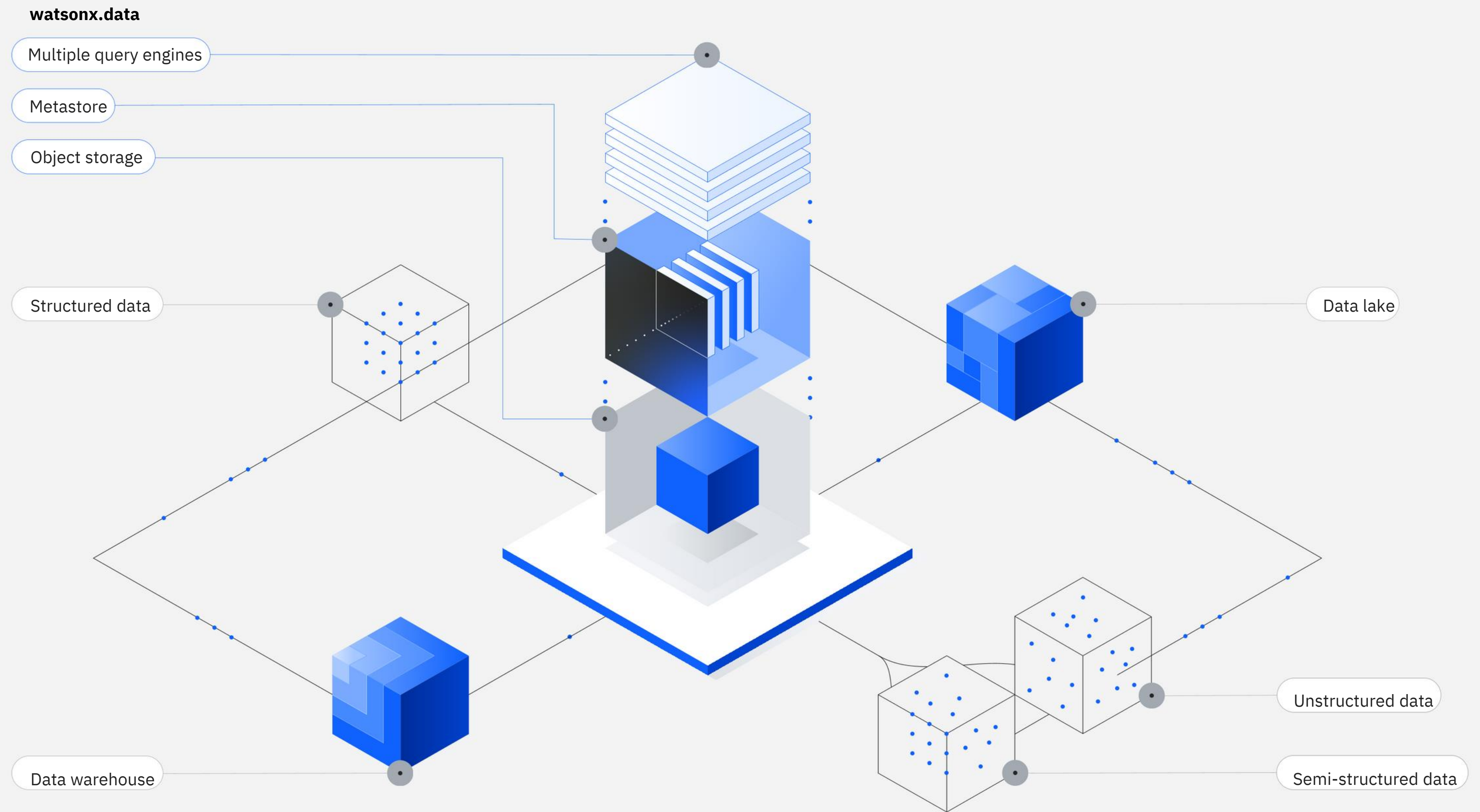
A legjobb *data lakehouse* olyan nyílt forráskódú technológiákat kínál majd, amelyek csökkentik az adatok megkettőzését és egyszerűsítik az összetett ETL-folyamatokat.

Legyen tudatában annak, hogy egyes első generációs *lakehouse* -oknak kulcsfontosságú korlátai vannak, amelyek korlátozzák a költség és összetettség kihívásainak kezelését.

Például az üzleti intelligencia vagy a gépi tanulás (ML) terhelésekre tervezett egyetlen lekérdezőmotor hatástalan lehet, ha más típusú terheléshez használják.

Az IBM adat- és MI-csapata úgy véli, hogy minden munkaterhelés egyedi, és a legmegfelelőbb környezetben kell optimalizálni, amely minimalizálja a költségeket és maximalizálja a teljesítményt. Válasszon egy olyan lakehouse-t, amely optimális teljesítményt nyújt a jobb döntéshozatal érdekében, valamint megadja azt a rugalmasságot, amely szükséges minden típusú adatból származó érték felszabadításához.

Ábra 1. Hogyan lehet a legjobban skálázni és felgyorsítani az MI hatását



04

Architektúra elemek

Infrastruktúra Ez az a hely, ahova a *lakehouse* telepítve lesz - teljes körűen kezelve bármilyen felhő- vagy helyben lévő környezetben.

Tároló Ez a réteg az, ahol az adat fizikailag tárolva van, fájlok formájában, és nyílt adatformátumokban, mint például az Apache Parquet és az Avro. A nyílt adatformátumok fájl-specifikációk és protokollok, amelyeket az open-source közösség számára elérhetővé tettek, hogy bárki felhasználhassa és fejleszthesse őket.

Nyílt táblázatformátumok A nyílt táblázatformátumok, mint például az Apache Iceberg, segítenek a struktúra biztosításában, valamint a megbízhatóság és az SQL egyszerűségének biztosításában nagy adathalmazok esetén. Ezek a formátumok lehetővé teszik, hogy különböző motorok egyszerre férjenek hozzá ugyanahhoz az adathoz, ami segít elkerülni a szállítói kötöttségeket. Olyan módon az adatokat több eszköz és adattár között, például az adattárházában; egyetlen példányú adat segít csökkenteni az adat-replikációt és megszüntetni az adat-szilók kialakulását.



Irányítás:

A metaadatokat is nyílt táblázatformátumokkal tárolják; ezek szolgálnak a fájlformátumok meghatározására bármely eszköz számára, amely képes olvasni vagy írni nyílt adatformátumokat.

Technikai metaadatszolgáltatás Ez a komponens szükséges ahhoz, hogy megértsük, milyen adatok állnak rendelkezésre a tárolási rétegben. A lekérdezési motor számára szükségesek a metaadatok az adatokhoz és táblákhoz, hogy biztosítsa a teljes származási információkat, és tudja, hol találhatóak, hogyan néznek ki, és hogyan lehet azokat olvasni.

Adatkatalógusok Ez a komponens segíti a felhasználókat abban, hogy megtalálják a feladathoz szükséges megfelelő adatokat, és szemantikai információkat biztosít a szabályzatokhoz és szabályokhoz. Várhatóan üzleti metaadatokat, például üzleti terminológiákat és címkéket fog tárolni a keresés és az adatvédelem érdekében.

Szabályzati motor Ez a komponens lehetővé teszi a felhasználók számára, hogy adatvédelmi szabályzatokat határozzanak meg, és lehetővé teszi a motor számára, hogy végrehajtsa ezeket a szabályzatokat. Annak érdekében, hogy egy skálázható irányítási keretrendszer hozzunk létre, a szabályzati motort gyakran a technikai metaadatszolgáltatással és az adatkatalógussal együtt telepítik.

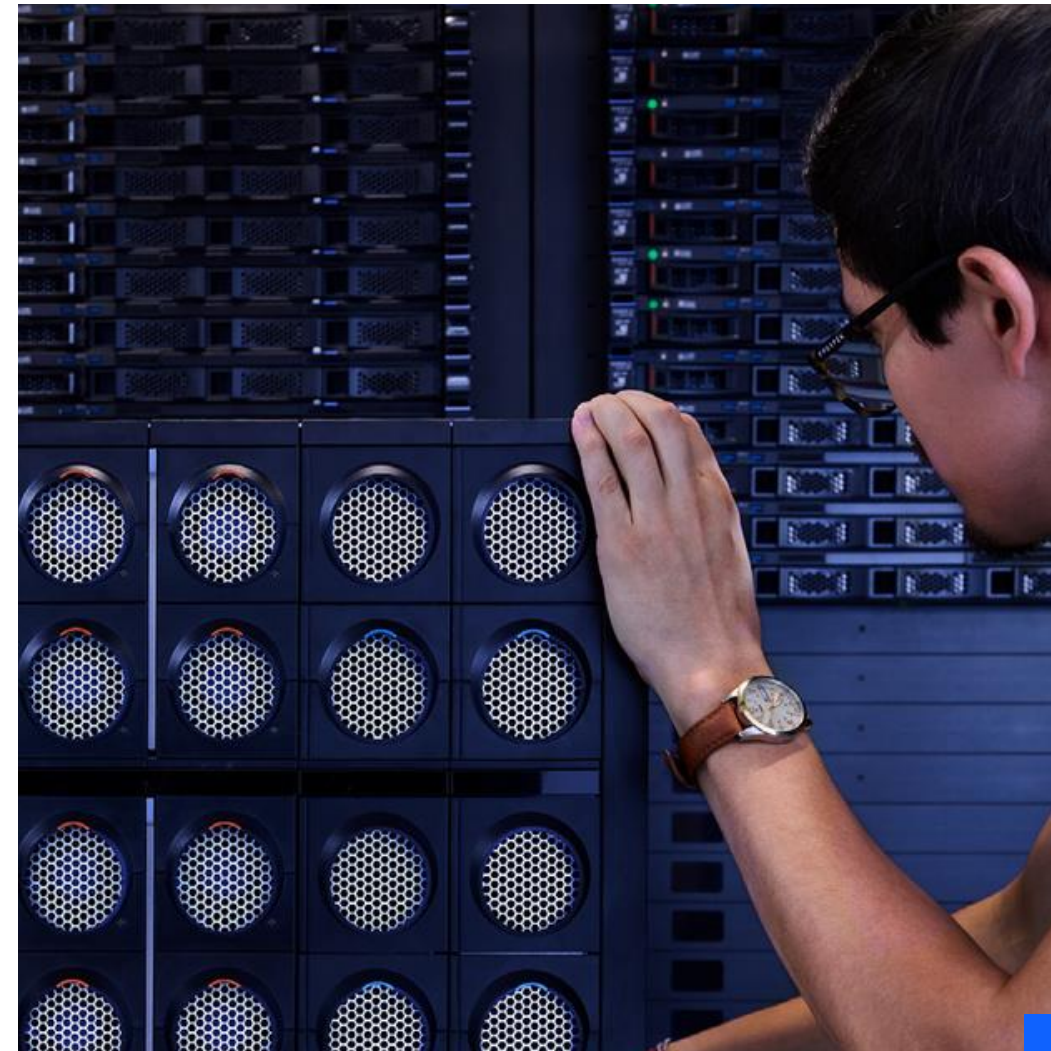
Lekérdezési motor: Ez a komponens az open data lakehouse központi eleme. A lekérdezési motor, amely lehet nyílt forráskódú vagy zárt forráskódú, hozzáfér az adatokhoz nyílt táblázatformátumban, és gyakran a számítási komponensként ismert. A lekérdezési motorok általában két típusba sorolhatók: SQL-alapú lekérdezési motor, mint például a nyílt forráskódú Presto, vagy a nyílt forráskódú Apache Spark motor vagy annak megfelelője. Egy nyílt lakehouse architektúrában a lekérdezési motor teljesen moduláris, ami azt jelenti, hogy a motor dinamikusan skálázható a terhelési igények és a párhuzamosság kielégítésére. A lekérdezési motorok bármilyen katalógushoz és tárolóhoz is csatlakozhatnak.

↓50%

Mostantól lehetőség van gyorsabb, megbízhatóbb betekintést elérni, miközben a felére csökkentheti az adattárház költségeit.⁴

05

Költségoptimalizálási lehetőségek



Ha a szervezete már rendelkezik helyszíni big data megvalósításokkal, a *lakehouse* olcsóbb alternatívát kínál az adatok objektumtárolón nyílt formátumokban történő tárolására. Csökkentheti az analitikai költségeket, egyszerűsítheti a rendszert és felgyorsíthatja az értékteremtést.

Ha már van meglévő adattárház-megvalósítása, a *lakehouse* megközelítés hatalmasan skálázható, alacsonyabb költségű alternatívát jelenthet a nagy elemzési munkaterhelésekhez, amelyek kevésbé érzékenyek a szolgáltatási szintű megállapodásokra (SLA-kra).

Az adatraktárak gyakran drágák és zártak – de egy *lakehouse* segítségével jelentősen csökkenthetők a tárolási és számítási költségek. Optimalizálhatja az adatraktári munkaterheléseket olyan célorientált motorok használatával, amelyek a munkaterhelési követelményein alapulnak. A *lakehouse* nyílt természete megszabadítja Önt a zárt adat-raktártechnológiától, ami kevesebb függőséget jelent az egyes gyártóktól, valamint csökkenti az IT-infrastruktúra fenntartási költségeit.

Az IBM watsonx.data egy nyílt, hibrid és szabályozott adattár, amely minden adat-, elemzési és MI-munkaterhelésre optimalizált.

Analitikai és adattudományi fejlesztések

“Abba az irányba haladunk, hogy a *data lakehouse* a legjobb gyakorlattá váljon.”³
Adam Ronthal, alelnök,
Gartner

A tulajdonosi adatformátumok és a magas tárolási költségek korlátozzák az AI- és ML-modellek együttműködését és bevezetését egy adattárház-környezetben; az adattavak alacsony teljesítményű adattudományi feladatokkal küzdenek. A technológiák elkülönítése olyan infrastrukturális kihívásokhoz vezetett, amelyek mellett biztonsági és felügyeleti problémák is felmerülnek, mivel az AI- és ML-modellek fejlesztéséhez szükséges adatok másolása és mozgatása is kockázatokkal jár.

A data lakehouse nagyszerű megoldás arra, hogy segítséget nyújtson kollégáinak, akik éhesek a szervezet adatállományaiban rejlő betekintésekre. Ha komolyan gondolja, hogy üzleti értéket szeretne kinyerni az egyre növekvő adatmennyiségből, mindenképpen érdemes megfontolni a lakehouse stratégiát.

Adam Ronthal, a Gartner alelnöke és elemzője, azt mondja, hogy „Abba az irányba haladunk, ahol a data lakehouse lesz a bevált gyakorlat.” A legjobb megközelítés egy nyitott, együttműködő és szabályozott környezetet fog kínálni az adattudományi feladatok teljes körű kezeléséhez.

Vizsgáljuk meg az IBM® watsonx.data™-t – a nyitott, hibrid és szabályozott adattárolót, amelyet minden adat-, analitikai- és AI-feladat optimalizálására terveztek.

Skálázza az MI-feladatokat az összes adatára, bárhol is legyenek. A Watsonx.data egy nyitott, hibrid és szabályozott adattároló, amelyet minden adat-, analitikai- és MI-feladatra optimalizáltak, és amely a *data lakehouse* architektúrára épül (1. ábra).

Férjen hozzá az összes adatához, és maximalizálja a munkaterhelés lefedettségét a hibrid felhőkörnyezetekben. Számítson a teljesen felügyelt szolgáltatások zökkenőmentes bevezetésére bármely felhőben vagy helyszíni környezetben. Érjen el bármilyen adatforrást, bárhol is található, egyetlen belépési ponton keresztül, és kombinálja azokat nyílt adatformátumok segítségével. Integrálja meglévő környezetébe nyílt forráskódú és nyílt szabványok használatával, valamint biztosítson interoperabilitást az IBM és harmadik fél szolgáltatásaival.

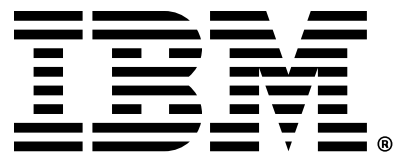
Gyorsítsa fel a megbízható betekintések elérését. Kezdjen gyorsan a beépített irányítási és automatizálási funkciókkal; erősítse a vállalati megfelelést és biztonságot az egész

ökoszisztémán átívelő egységes irányítással. Az átlátható felhasználói élmény és a kattintással vezérelhető konzol segíti csapatait az adatok beolvasásában, elérésében, átalakításában és a feladatok futtatásában. Figyelje meg, milyen gyorsan elfogadják majd a műszerfalat, amely megkönnyíti számukra a költségmegtakarítást és a friss, megbízható betekintések szállítását. **Csökkentse** adatközpontjának költségeit akár 50%-kal a munkaterhelés optimalizálásával, több lekérdező motor és tárolási szint használatával. Optimalizálja a költséges adatközponti munkaterheléseket olyan célzott motorokkal, amelyek automatikusan felfelé és lefelé skálázhatók. Csökkentse a költségeket az adatismétlések megszüntetésével, amikor alacsony költségű objektumtárolást használ; nyerjen több értéket a nem hatékony adat-tavakban lévő adatokból.

08 További lépések

Használja ki az IBM csapat adatkezelési és optimalizálási tudását, amelyet évtizedek során fejlesztettek ki a világ legnagyobb kihívást jelentő adatfeldolgozási munkaterheléseinek kezelésével. Nézze meg, milyen gyorsan teremthet értéket a watsonx.data használatával.





1. Why Unstructured Data is the Future of Data Management, Venturebeat, July 2021.
2. Worldwide IDC Global DataSphere Forecast, 2022-2026, IDC, May 2022.
3. The rise of the data lakehouse: A new era of data value, CIO Magazine, 18 August 2022
4. When comparing published 2023 list prices normalized for VPC hours of IBM watsonx.data to several major cloud data warehouse vendors. Savings may vary depending on configurations, workloads and vendors.

További információért lépjen kapcsolatban az IBM Business Partnerével:

VIR Zrt.
info@virzrt.hu

© Copyright IBM Corporation 2023 IBM Corporation
New Orchard Road Armonk, NY 10504 Produced in the
United States of America May 2023 IBM, the IBM logo,
and watsonx.data are trademarks or registered
trademarks of International Business Machines
Corporation, in the United States and/or other
countries. Other product and service names might be
trademarks of IBM or other companies. A current list of
IBM trademarks is available on ibm.com/trademark. It
is the user's responsibility to evaluate and verify the
operation of any other products or programs with IBM
products and programs. The performance data and
client examples cited are presented for illustrative
purposes only. Actual performance results may vary
depending on specific configurations and operating
conditions. THE INFORMATION IN THIS DOCUMENT IS
PROVIDED "AS IS" WITHOUT ANY WARRANTY,
EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY
WARRANTIES OF MERCHANTABILITY, FITNESS FOR A
PARTICULAR PURPOSE AND ANY WARRANTY OR
CONDITION OF NON-INFRINGEMENT. IBM products
are warranted according to the terms and conditions of
the agreements under which they are provided.

Statement of Good Security Practices: No IT system or
product should be considered completely secure, and
no single product, service or security measure can be
completely effective in preventing improper use or
access. IBM does not warrant that any systems,
products or services are immune from, or will make
your enterprise immune from, the malicious or illegal
conduct of any party. The client is responsible for
ensuring compliance with all applicable laws and
regulations. IBM does not provide legal advice nor
represent or warrant that its services or products will
ensure that the client is compliant with any law or
regulation.