**Korszerű Adatbázisok**

**Beadandó feladat**

**Kovács Levente**

**Y2XWHJ**

**2025.**

# InfluxDB Egy időalapú nem-relációs adatbázis

Az adatbázis-kezelő rendszerek világában az utóbbi években egyre nagyobb figyelmet kaptak a nem-relációs (NoSQL) megoldások. Különösen azok, amelyek valamilyen speciális adattípust céloznak, mint például az időbélyegekkel ellátott (time-series) adatok tárolását és kezelését. Az InfluxDB egy ilyen, időalapú, nem-relációs adatbázis-kezelő rendszer, amelyet kifejezetten a nagy mennyiségű, gyorsan keletkező időbélyegzett adatok hatékony kezelésére fejlesztettek ki. E beszámoló célja az InfluxDB adatmodelljének, gyakorlati használati eseteinek, valamint az interfészek és adatkezelési lehetőségeinek bemutatása.

## Az InfluxDB adatmodellje

Az InfluxDB nem relációs módon, hanem úgynevezett mérési pontokat (measurements) tárol, amelyek egy adott esemény vagy érték időben történő rögzítését jelentik. Az adatmodell központi eleme az időbélyeg (timestamp), amely minden rögzített adat kötelező eleme, hiszen az InfluxDB lényegében idővonalon értelmezi az adatokat. Emellett három fő komponens határozza meg az adatstruktúrát:

* **Measurement** – Gyakorlatilag egy logikai táblának felel meg, például: "hőmérséklet", "CPU\_kihasználtság", "forgalom", stb.
* **Tags** – Indexelt kulcs-érték párok, amelyek jellemzőként írják le az adatforrást (pl. város=Budapest, érzékelő=szenzor1). A tag-ek alapján történő lekérdezés gyors, mivel ezek automatikusan indexelve vannak.
* **Fields** – Tényleges mérési értékek, amelyek nem indexeltek (pl. hőmérséklet=21.3, sebesség=55.7). Egy measurementhez több field is tartozhat.
* **Timestamp** – Az időpont, amikor az adott adat keletkezett. E nélkül nincs érvényes adatpont ebben az adatbázis-kezelő rendszerben.

Az InfluxDB tehát kvázi sorozatokat (series) kezel, amelyek a measurement, tag-set és field kombinációjából jönnek létre. Ez lehetővé teszi, hogy ugyanazon measurement alatt több ezer különböző szenzor vagy eszköz adatai külön sorozatként legyenek kezelve, mégis egységes struktúrában.

A séma szabadabb, mint egy relációs adatbázis esetén: nem szükséges előre deklarálni a mezőket, és egy measurement sorozat különböző field-ekkel is rendelkezhet időben eltérően. Ugyanakkor fontos figyelembe venni a performanciát befolyásoló sématervezési ajánlásokat – például hogy a tag-ek száma ne nőjön túl gyorsan (cardinality problémák).

## Használati esetek és gyakorlati alkalmazás

Az InfluxDB-t tipikusan olyan helyzetekben alkalmazzák, ahol az adatok gyors ütemben, gyakran másodpercenként többször is keletkeznek, és hosszú időn keresztül kell őket tárolni és elemezni.

1. **IoT (Internet of Things) és szenzoradatok:** Talán a leggyakoribb felhasználás. Akár több ezer érzékelő is küldhet adatot másodpercenként, például hőmérsékletet, páratartalmat, rezgést vagy GPS-koordinátát. Az InfluxDB kiválóan kezeli az ilyen típusú tömeges adatáramlást.
2. **Rendszerfigyelés (monitoring):** Szerverszolgáltatások, hálózati forgalom, CPU-használat, memóriakihasználtság, stb. rendszeres mérésére használható. Gyakran a Telegraf ügynökkel együtt telepítik, ami képes automatikusan gyűjteni rendszeradatokat és küldeni azokat az InfluxDB felé.
3. **Pénzügyi idősorok:** Árfolyamok, tranzakciók, kriptovaluták értékének követése. Mivel a pénzügyi adatoknak is jellemzője a gyors változás és időbélyegzettség, az InfluxDB jól használható ezen a területen is.
4. **Energia- és környezetgazdálkodás:** Villamosenergia-fogyasztás, napelemtermelés, szélenergia-mérések – ezek az adatok szintén tipikusan időalapúak, és gyakran nagy időbeli felbontást igényelnek.
5. **Agrárinformatika és meteorológia:** Talajnedvesség, hőmérséklet, csapadék, NDVI-index – ezek mind olyan mérési sorozatok, amelyek időben változnak és elemzést igényelnek.

A fenti példák közös jellemzője az adatok időbeli jellege, a gyakori frissülés, és az, hogy gyakran szükség van visszamenőleges elemzésre vagy előrejelzések készítésére.

## Az InfluxDB kezelőfelületei és interfészei

Az InfluxDB használatához többféle felület is rendelkezésre áll, a fejlesztők és rendszergazdák igényeitől függően:

1. **InfluxQL:** A korábbi verziók SQL-szerű lekérdezőnyelve, amely egyszerűbb, gyors lekérdezésekre ideális. Például:

SELECT mean("hőmérséklet") FROM "időjárás" WHERE "város"='Szeged' AND time > now() - 7d GROUP BY time(1h);

1. **Flux:** Az újabb, rugalmasabb lekérdezőnyelv, amely programozhatóbb, adattranszformációkra és összetettebb aggregálásokra is alkalmas. Példa:

from(bucket: "mérések")

|> range(start: -1d)

|> filter(fn: (r) => r.\_measurement == "hőmérséklet" and  
 r.város == "Debrecen")

|> mean()

1. **Grafikus felület (GUI):** Az InfluxDB beépített webes felületet biztosít, ahol lehetőség van vizualizációk (grafikonok) létrehozására, adatfolyamok monitorozására, dashboard-ok építésére.
2. **CLI:** Parancssoros eszköz, amellyel lehetőség van bucket-ek létrehozására, adatfeltöltésre, lekérdezések futtatására.
3. **REST API:** Az InfluxDB HTTP-alapú API-kat kínál, így külső programból (pl. Python, JavaScript) is könnyen integrálható.
4. **Telegraf:** Egy különálló adatgyűjtő ügynök, amely több száz input plugint támogat.

A rendszer konfigurálható úgynevezett retention policy-kkel is, amelyek meghatározzák, hogy egyes adatok meddig kerüljenek tárolásra.

## Az InfluxDB összegzése

Ez egy rendkívül hatékony, időalapú nem-relációs adatbázis-kezelő rendszer, amely jól illik a mai adatintenzív környezetbe, különösen az IoT, monitoring és idősoros analitika területein. Rugalmassága, skálázhatósága és a fejlett interfészek miatt könnyen integrálható más rendszerekbe is. Az adatmodellje egyszerű, de jól átgondolt: különlegessége az időbélyegek, a mért értékek és a leíró metaadatok megfelelően strukturált kombinációja.