Bagó Boglárka (SI2JMR), Gábor Krisztián(), Gergő (), Kike (), Hegedüs Alexandra(XM6YBD)

GDE Gazdaságinformatika Bsc

aDATBÁZISOK   
PROJEKTMUNKA  
 DOKUMENTÁCIÓ

íVKEMENCE ADATAINAK FELDOLGOZÁSA,

2024. szeptember 28.

Tartalom

[I. Bevezetés 3](#_Toc179375975)

[1.1 Csapat bemutatása 3](#_Toc179375976)

[1.2 Szerepkörök magyarázata 3](#_Toc179375977)

[2. Lépések 4](#_Toc179375978)

[3.Téma ismertetése 5](#_Toc179375979)

[3.1 Témakör 5](#_Toc179375980)

[3.2 Projekt adatai 6](#_Toc179375981)

[3.3 Projekt célja 6](#_Toc179375982)

[4. Funkciók ismertetése 7](#_Toc179375983)

[4.1 Általános műveletek,input 7](#_Toc179375984)

[4.2 Eredmény prezentálása 8](#_Toc179375985)

[II.A munkafolyamat részletei 9](#_Toc179375986)

[1.Adatfeldolgozás excellel és szabadszemmel 9](#_Toc179375987)

[2.Relációs adatbázis fontossága a projektnél 10](#_Toc179375988)

[3.Adatbázis megtervezése ER diagrammal 10](#_Toc179375989)

[3.1 Definíció 10](#_Toc179375990)

[3.2 Az aktuális entitások és attribútumok meghatározásának lépései 11](#_Toc179375991)

[4.Adatok feltöltése, ellenőrzése 13](#_Toc179375992)

[Az első egyeztetés során megállapítottuk,hogy hibás adatokat nem szabad és nem lehet feltölteni az adatbázis kezelőrendszerbe, ezért először tisztítani fogjuk az adatokat és normalizálni.Felvetődütt gondolatok: 13](#_Toc179375993)

[pythonnal , az adatokátfutni nincs -e type hiba,kimaradás..stbk az adatok és az adattípusok helyességé 13](#_Toc179375994)

[5.Adatbázis létrehozása 13](#_Toc179375995)

[6.Adatbázis létrehozása 13](#_Toc179375996)

[7.Lekérdezések 14](#_Toc179375997)

[8. Allekérdezések, UNION 14](#_Toc179375998)

[9. Teljesítmény optimalizálás 14](#_Toc179375999)

[10. Adattisztítás. Az adatok hibával terheltek 14](#_Toc179376000)

[11. Milyen tudásuk alapján tudták azonosítani a hibákat? 15](#_Toc179376001)

[12. Az adatok egy ipari adatgyűjtő program írja az adatbázisba. 15](#_Toc179376002)

[13. Biztonsági mentésre milyen stratégiát választanának? 15](#_Toc179376003)

[III.Összegzés 16](#_Toc179376004)

[1.Alfejezet Címsor 2 és 3 16](#_Toc179376005)

[Mellékletek 16](#_Toc179376006)

# I. Bevezetés

## 1.1 Csapat bemutatása

|  |  |  |  |
| --- | --- | --- | --- |
| Szerepkör | Név | NeptunKód | Szak |
|  | Gábor Krisztián |  | Mérnökinfo |
|  | Gergő |  | Mérnökinfo |
|  | Kike |  | Mérnökinfo |
|  | Bagó Boglárka | SI2JMR | GazdInfo |
|  | Hegedüs Alexandra | XM6YBD | GazdInfo |

## 1.2 Szerepkörök magyarázata

1. **Projektvezető (Project Manager)**

A projektvezető felelős az egész projekt szervezéséért és koordinálásáért. Feladata, hogy irányítsa a csapat munkáját, figyelemmel kísérje a projekt haladását, és biztosítsa, hogy mindenki időben elvégezze a feladatát. Emellett ő tartja a kapcsolatot az oktatókkal vagy a tanácsadókkal, és gondoskodik arról, hogy a projekt megfeleljen az előírt követelményeknek. Az ő feladata továbbá a határidők betartása, az erőforrások allokálása, és a felmerülő akadályok megoldása.

2. **Adat- és üzleti elemző (Data and Business Analyst)**

Az adat- és üzleti elemző feladata az ívkemence működési adatai és üzleti szempontú követelményeinek elemzése. Ő vizsgálja meg, hogyan gyűjthetők össze az adatok, miként dolgozhatók fel, és hogyan használhatók fel ezek az információk a projekt céljainak eléréséhez. Elemzi az adatokat és meghatározza, hogy milyen mutatók fontosak a projekt sikeres teljesítéséhez. Az üzleti elemző felel továbbá az adatok értelmezéséért, valamint a jelentések és vizualizációk elkészítéséért, hogy a csapat többi tagja megfelelően láthassa az eredményeket.

3. **Vezető fejlesztő (Lead Developer)**

A vezető fejlesztő felelős az adatok feldolgozását végző rendszerek és szoftverek fejlesztéséért. Feladata az ívkemence adatainak kezeléséhez szükséges algoritmusok és kód megírása, illetve a technikai megvalósítás irányítása. Ő biztosítja, hogy az adatfeldolgozó szoftverek hatékonyan működjenek, és képesek legyenek a nagy mennyiségű adat kezelésére. Szorosan együttműködik az elemzővel és a projekt többi tagjával, hogy a technikai megoldások összhangban legyenek a projekt céljaival.

4. **Vezető tesztelő (Lead Tester)**

A vezető tesztelő felelős a fejlesztett rendszerek minőségének biztosításáért. Ő végzi a szoftverek tesztelését annak érdekében, hogy azok megfelelően működjenek, és minden adatfeldolgozási funkció hibátlan legyen. A tesztelési fázis során különböző forgatókönyvek alapján teszteli az ívkemence adatok feldolgozását, és ellenőrzi, hogy az adatok pontosan és helyesen jelennek meg a rendszerben. A felmerülő hibákat jelentésekkel dokumentálja, és javaslatokat tesz azok kijavítására.

5. **Adatmérnök (Data Engineer)**

Az adatmérnök feladata az ívkemence adatainak begyűjtése, tárolása és feldolgozása. Ő tervezi meg az adatok áramlását biztosító technikai infrastruktúrát, és felelős a különböző adatbázisok létrehozásáért, karbantartásáért, valamint azok integrációjáért. Az adatmérnök biztosítja, hogy az ívkemencéből származó nyers adatok hatékonyan feldolgozhatók legyenek, és hozzáférhetők a csapat többi tagja számára elemzés céljából. Szoros együttműködésben dolgozik a vezető fejlesztővel, hogy a fejlesztett rendszerek képesek legyenek kezelni a nagyméretű és komplex adathalmazokat. Ő gondoskodik arról is, hogy az adatok tiszták, konzisztens formátumúak és megfelelően strukturáltak legyenek az elemzési célok érdekében.

## 2. Lépések

A projekt lépései és a hozzájuk tartozó feladatok ütemtervvel való egybehangolása során minden szakaszt alaposan áttekintettünk. Amennyiben eltérések voltak a 1-es verzióhoz képest, azokat szekvenciálisan megjelöltük és korrigáltuk, ahol szükséges volt.

|  |  |  |  |
| --- | --- | --- | --- |
| Verzió | Dátum | Szerző | Tevékenység |
| 1. | 2024.10.08 | Mindenki | **1.alkalom**: Megbeszélés, alapok lefektetése:   * Elemzés Excellel, * Kiindulópontok keresés, * Értelmezés |
| 1. | 2024.10.08 | HA | Dokumentáció írása , főbb pontok hozzáadása, ER Diagram 1.verzió |
| 1. | 2024.10.09 | VG | Adattisztítás, normalizálás |
| 1. | 2024.10.09 | HA | Projekt lépéseinek tervezése, megbeszélés részleteinek hozzáadása a dokumentumhoz |
| 2. | 2024.10.09 | HA | ER Diagram 2.verzió |
| 1. | 2024. 10. 11 | BB | Diagramok és kiértékelésük |
| 1. | 2024.10.14 | HA | Github porject inicializálása  Alap használati útmutató  Python DB táblák készítése |
| 1 | 2024.10.14 | BB,VG, HA | Github szinkronizálás, összekapcsolás PyCharmmal |
| 1. | 2024.10.16 | VG, HA | Py adatisztítás rész bővítése, Diagram elemzés átnézése |

## 3.Téma ismertetése

### 3.1 Témakör

Az elektromos ívkemence olyan kemence, amely elektromos ív segítségével fűti az anyagot. Az ipari ívkemencék mérete a körülbelül egy tonnás kapacitású, kis egységektől a másodlagos acélgyártáshoz használt 400 tonnás egységig terjednek. Az Electric Arc Kemence (EAF) egy acélkészítő kemence, amelyben acélhulladékot fűtik és olvad fel a kemenceelektródák és a fémfürdő között. A hűtőpanelek (hűtési elemek) az ívkemencékben olyan hűtőrendszerek, amelyek a kemence falait és egyéb forró részeit hűtik, hogy megakadályozzák a túlmelegedést és a szerkezeti károsodásokat. A panelekben csövekben áramoltatják a vizet, célja a kemence élettartamának növelése.

Egy ilyen kemence ára függ a nagyságától, kapacitásától, teljesítményétől…stb , de egészen magas, akár millió euros árat is elérhet. A minőség és a biztonság szempontjából is kulcsfontosságú a megfelelő szabályzás, ezért van szükség az adatok folyamatos értelmezésére , kontextusba helyezésére és vizsgálatára.

**Adagok**: Újra olvasztott acélhulladék.

**Hűtőpanelek**:A hűtést végző vízzel feltöltött segédrendszer.

1. ábra cím: ELEKTROMOS ÍVKEMENCE

  
Forrás megjelölés: Facebook, Fémipari Szakember- Steel Worker csoport

### 3.2 Projekt adatai

Projekt neve: Ívkemence adatainak kezelése

Projekt kezdete: 2024.10.08

Projekt elvárt befejezési ideje:2024.11.17

Projekt tényleges befejezési ideje:

Kommunikációhoz használt felületek,keretrendszerek,programok: GitHub, PyCharm, Discord

Kapott adatok: Adagok.csv, Hűtőpanelek.csv

### 3.3 Projekt célja

Az adatok feldolgozása Python és SQL használatával, CSV formátumú fájlokból. Adatok létrehozása, lekérdezése, módosítása és törlése. Ezen folyamatok során kiemelt jelentőséget kap az adattisztítás, az adatok integritásának biztosítása, valamint az anomáliák és hibák detektálása és javítása, továbbá a trendek és mintázatok disszonanciáinak feltárása.(két egyforma folyamat nincsen de jellegében kevés eltérést mutasson)

A folyamat főbb lépései a következők:

**Adatbevitel és betöltés**: Az adatok importálása CSV fájlokból Python segítségével, mely során a kapott struktúrák SQL adatbázisba kerülnek. Ez a lépés magában foglalhatja az adatok előfeldolgozását, például formátumkonverziókat, hiányzó értékek kezelését, valamint típuskonzisztenciák ellenőrzését.

**Adatlekérdezés és manipuláció**: SQL segítségével specifikus lekérdezések futtatása az adatbázisban, hogy előre meghatározott kritériumok alapján kinyerhessük a releváns információkat. Az adatok módosítása (szerkesztése) során az adatintegritás fenntartása kritikus fontosságú, míg az adatok törlése gondos elemzés után történik, hogy elkerüljük a hasznos információk elvesztését.

**Anomáliák és hibák kiszűrése**: Az adatok elemzése során matematikai és statisztikai módszerekkel azonosíthatók az anomáliák és hibák. Ez magában foglalja a hibás vagy szokatlan adatok detektálását, például statisztikailag szignifikáns eltérések feltárását, amelyek hibás adatgyűjtésre, mérési hibákra, vagy szabálytalanságokra utalhatnak.

**Trend disszonanciák elemzése**: A kapott adatok trendjeinek vizsgálata, különös tekintettel a hosszú távú mintázatokra és azok időbeli alakulására. A disszonanciák feltárása során figyelembe vesszük a megszokott trendektől való eltéréseket, melyek mögöttes okainak statisztikai módszerekkel történő azonosítása hozzájárul a döntéshozatali folyamatokhoz. A trend disszonanciák előfordulása utalhat váratlan külső tényezőkre, rendszerszintű hibákra vagy nem szándékolt változásokra az adatgyűjtési folyamatban.

Az egész folyamat célja az adatok konzisztenciájának és pontosságának biztosítása, amely nélkülözhetetlen az eredmények tudományos értelmezéséhez, valamint az adatok megbízható felhasználásához döntéstámogató rendszerekben.

## 4. Funkciók ismertetése

### 4.1 Általános műveletek,input

CRUD adatbevitel,szerkesztés,törlés,lekérdezés

### 4.2 Eredmény prezentálása

**Adatok és diagramok kiértékelése**

A képen szöveg, képernyőkép, Betűtípus, sor látható

Automatikusan generált leírás

A képen szöveg, képernyőkép, Betűtípus, sor látható

Automatikusan generált leírás

A képeken látható vonaldiagramok egy ívkemence hűtőpaneljeinek hőmérsékleti adatait mutatják az idő függvényében. X tengelyen tehát az mérési idő látható másodperc pontossággal, az Y tengelyen pedig a mért hőmérséklet.

**Az adatok és diagram alapján történő következtetések:**

* A hőmérsékletadatok jellemzően a 30°C és 50°C közötti tartományban mozognak, ami normális működésre utal a panelek többségénél.
* Az adatok folyamatosak, nincsenek jelentős mérési szünetek, és a hűtőpanelek működésében általánosan stabilitás mutatkozik.
* Bizonyos esetekben kiszúrtunk ingadozásokat és esetleges szélsőértékeket, ahol a hőmérséklet váratlanul megugrik vagy csökken.

**Panelek hőmérsékletének részletezése:**

* **Panel 5, 9, 13:**
  + Ezek a panelek stabilan a 30-50°C közötti tartományban mozognak, néhány kisebb hőmérséklet-ingadozással.
  + Az ingadozások általában a kemence működésének természetes változásaihoz köthetők, például a be- és kikapcsolási ciklusokhoz, valamint az anyagfeldolgozási szakaszokhoz. A folyamatok jellegükben megegyezőek
  + Nincs jelentős kiugrás vagy 0-ra való visszaesés.
* **Panel 1,6, 11:**
  + Itt figyelhetők meg jelentősebb hőmérsékleti kiugrások, akár 150-200°C körüli lokális csúcsok is. Ez eltér a többi paneltől, amelyek értékei stabilabbak.
  + Lehetséges okok: Ezeknek a paneleknek a mérésénél időszakos zavarok vannak. Valószínűleg a hibás adatok a szenzorok hibájából adódnak, főleg hogy a forráspont feletti érétkek nem reálisak.
* **Panel 1,3, 10, 14:**
  + Ezeknél a panelek működése során szintén láthatók kisebb ingadozások, azonban ezek mérsékeltek. A hőmérséklet ritkán megy 40-50°C fölé, és viszonylag konstans marad.
  + Az ingadozások oka lehet például a hűtőfolyadék áramlásának változása vagy a kemence hőmérsékleti ciklusai okozta változás, ez lehet reális adat, de meg kell nézni, hogy miből adódik az eltérés.

**Hőmérsékleti eltérések okainak lehetséges magyarázata:**

* **Hűtési rendszer működése:** Bizonyos panelek esetében a hőmérséklet rövid időre jelentősen megnövekszik(60-70 fok körüli értékre), ami utalhat a hűtési rendszerben fellépő átmeneti problémákra. Ezek lehetnek a hűtőfolyadék áramlásának megszakadásai, eltömődések vagy szivárgások.
* **Kemence belső működése:** Azoknál a paneleknél, ahol magasabb hőmérsékleti csúcsokat tapasztalunk (pl. Panel 6 és 11), feltételezhető, hogy közelebbi kapcsolatban állnak a kemence aktívabb zónáival. Ezek a területek hőtermelőbbek, így a panelek is magasabb hőmérsékleteknek vannak kitéve.
* **Hőmérséklet-szenzorok érzékenysége:** A jelentős hőmérsékleti kiugrásokat okozhatja a hőmérséklet-szenzorok átmeneti hibája is, bár a legtöbb adat a megfelelő működésre utal.

# II.A munkafolyamat részletei

## 1.Adatfeldolgozás excellel és szabadszemmel



Az adatok első vizsgálatakor a következő megállapításokat tettük:

* Dátumok el vannak csúszva a két file között (nem egyszerre kezdődik az idősor)
* Címek hibás karaktereket tartalmaznak
* Szükség van PK létrehozására, és egyértelmű hozzárendelésre
* Különböző nagyságú időintervallumok vannak megadva
* Nincsenek mértékegységek
* Adagközi idő hiányzik
* Idő alapján párosíthatóak az adatok
* 17- es adagnál eltérés látható

jellegre hasonlóak a görbék? akkor ugyanaz

kétszer ugyanazt az eredményt nem lehet kapni

lehet a kiindulási anyagban szennyeződés

meghibásodás

* a hűtőpanelek csv egy azonos időpontban a hűtőrendszer különböző pontjain mért értékek, ebből kifolyólag ha ugrásszerű változást látunk az hiba
* 07.19.10:45 percnél kiugró hőm. adatsor = berendezés hiba
* 07.18 18:25-től anomália valószínűleg lehűlés és csak rámondták ,hogy adag
* 27-es adag végén van egy kiugró hőmérséklet , csúcs az 1-es panelnél

## 2.Relációs adatbázis fontossága a projektnél

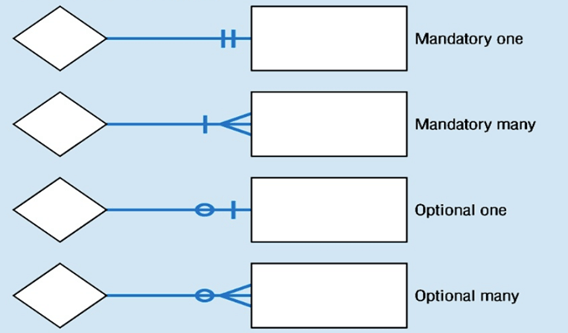
Az adatbázisunkban az adatokat táblákban tároljuk, amelyeket SQL nyelvvel kezelünk. Ez biztosítja az adatok konzisztenciáját és integritását, ezzel megbízható, pontos és biztonságos adatkezelést kínálnak.

A szerkezet felépítéséből fakad, hogy relációs adatbázist használunk: mivel a relációs adatbázis modelljében az adatok táblázatos formában kerülnek tárolásra, ahol minden tábla sorai különböző rekordokat, míg oszlopai a rekordokhoz tartozó attribútumokat reprezentálják A kapott adataink már a kezdetektől ilyen formát követnek, azonban az adatok szerkezetének továbbra is meg kell felelnie a normalizációs követelményeknek. A normalizációs folyamat biztosítja, hogy az adatbázisunk minimális redundanciát tartalmazzon, ezáltal hatékonyabb tárolást és adatkezelést eredményez. Ezen túlmenően segít elkerülni az anomáliákat (pl. beszúrási, törlési és módosítási anomáliák), amelyek az adatok konzisztenciáját veszélyeztethetnék.

## 3.Adatbázis megtervezése ER diagrammal

### 3.1 Definíció

ER diagram Az Entity Relationship Diagram rövidítése, más néven ERD egy diagram, amely az adatbázisban tárolt entitáshalmazok kapcsolatát jeleníti meg, segít egy jól megtervezett adatbázis létrehozásában. Az ER diagramok három alapfogalom alapján készülnek:

* entitások: hely,személy,objektum,esemény,fogalom valós dolog, kell hogy legyenek attribútumai és elsődleges kulcsa
* gyenge entitás: dupla gyémánt kapcsolja össze az entitással, nincs elsődleges kulcsa, gyenge entitáskészletben ez az erős entitáskészlet elsődleges kulcsának és részkulcsának kombinációja.
* attribútumok: entitás tulajdonsága
* kapcsolatok: ami összekapcsolja az az entitásokat, attribútumokat, gyakran azonosíthatjuk igékkel pl: tamás részt vesz a kémia órán

ER Diagram fő összetevői és szimbólumai:

|  |  |
| --- | --- |
| Jelölés | Magyarázat |
| téglalap | entitás típus |
| dupla téglalap | gyenge entitás(nincs elsődleges kulcsa) |
| ellipszis | attribútum |
| gyémánt | kapcsolat típus |
| sor | attribútumokat és entitásokat kapcsol össze |
| elsődleges kulcs | attribútumok aláhúzottak |
| dupla ellipszis | többértékű attribútum |

### 3.2 Az aktuális entitások és attribútumok meghatározásának lépései

#### 3.2.1 Entitás azanosítás:

adagok tábla: adagok

panelek tábla: panelek 1- 15-ig

#### 3.2.2 Kapcsolat azonozítás:

idő alapján kapcsolhatóak az adagok a panelekkel, a paneleknek függősége van az adagolási ciklusokra (ha egy adagot hozzáadunk megváltozik a hűtőpanel által mért hőmérséklet

#### 3.2.3 Kardinalitás azonosítása:

Egy adaghoz több hűtőpanel hőm. kapcsolódik 1:N

Egy panelhez egy adag tartozik 1:1

#### 3.2.4 Attribútumok azonosítása:

adagok:

* adagszám – elsődleges kulcs
* kezdeti dátum
* kezdeti idő
* vég dátum
* vég idő
* adagok közötti idő
* adag idő

panelek:

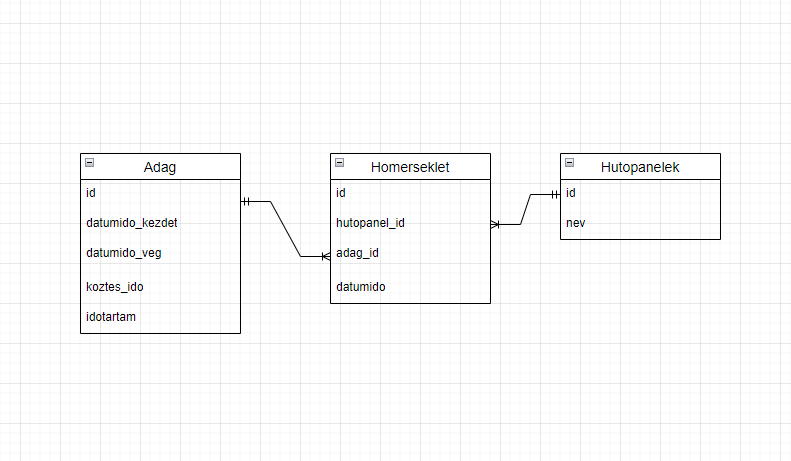
* adagszám – idegen kulcs – időintervallum egyeztetés után
* hőmérséklet ideje
* hőmérséklet mértéke

#### 3.2.5 ER Diagram létrhozása

2. ábra ER Diagram 1.verzió



3. ábra ER Diagram 2.verzió



## 4.Adatok feltöltése, ellenőrzése

## Az első egyeztetés során megállapítottuk,hogy hibás adatokat nem szabad és nem lehet feltölteni az adatbázis kezelőrendszerbe, ezért először tisztítani fogjuk az adatokat és normalizálni.Felvetődütt gondolatok:

* dátum és idő egybevonása a könnyebb kezelhetőségért, átláthatóságért kevesebb hibalehetőség(DATETIME mindenhol)
* adagok számának kinevezése elsődleges kulccsá (id), amit idegen kulcsként a hűtőpanelek tábla is megkap(adag\_id)
* Hűtőpanelek file is kapjon egy azonosítót (panel\_id)
* Hómérsékletek (homerseklet\_id)
* egy időpont többször egymás után szerepel , alig eltérő adatokkal, érdemes ezt redukálni-? nem, azért mert nem egyformák, és nem tudjuk milyen a vízáramoltatás ( 100 l/sec vagy 10 l/sec így nem tudjuk megítélni mennyit változtat 0,2 fok eltérés, mekkora a súlya)
* sor címek hibás karaktereket tartalmaznak : UTF-8 encodeinggal is hibás karaktereket jelenít meg, valószínűleg más encodeingban lett elmentve (pl.:Windows-1252)

## 5.Adatbázis létrehozása

metódus indoklása

## 6.Adatbázis létrehozása

6.1 DCL és TCL: SQLite-n belül nincs jogosultságkezelés. a. Írják le, hogy melyik résztvevőnek milyen jogosultságra lett volna szüksége, a munkájához.

6.2 b. Vizsgálják meg, hogy van-e, és ha van, akkor milyen esetben tranzakciókezelésre.

6.3 1 tranzakciót tegyenek bele akkor is, ha nem feltétlen szükséges csak a példa kedvéért.

## 7.Lekérdezések

7.1 Végezzenek el lekérdezéseket, melyek összecsatolnak táblákat (JOIN).

7.2. Használjanak aggregációs függvényeket.

7.3. Írják le, hogy miért végezték el ezeket a műveleteket, és milyen következtetésekre jutottak.

## 8. Allekérdezések, UNION

8.1 Szükség esetén készítsenek egymásba ágyazott lekérdezéseket

8.2 A részeredményeiket INSERT és SELECT segítségével tegyék ki külön táblákba

8.3. Vizsgálják meg, hogy mit lehetne kimutatni, vagy ellenőrizni az UNION utasításokkal

## 9. Teljesítmény optimalizálás

9.1 Melyik mezőre milyen kulcsot, indexet alkalmaznának, és miért?

9.2 . Milyen módszert javasolnának az adatműveletek gyorsítására, ha megkapnának 10 évnyi adatot és azt kellene feldolgozni?

## 10. Adattisztítás. Az adatok hibával terheltek

10.1 Milyen módszerrel, illetve logika alapján kerestek, szűrtek hibás adatokat?

10.2. Milyen módszerrel kezelték azokat?

## 11. Milyen tudásuk alapján tudták azonosítani a hibákat?

11.1 Mi alapján tekintették a hibás adatokat hibás adatnak? pl. Fizika órán tanulták, otthon tapasztalták, valamilyen logika alapján stb. b.

11.2 Adatbevitelhez milyen érvényesítési szabályokat, milyen kényszereket javasolnának

## 12. Az adatok egy ipari adatgyűjtő program írja az adatbázisba.

Az adatok rögzítése és feldolgozása során milyen szerepköröket hoznának létre, és azokat milyen jogosultsággal látnák el?

## 13. Biztonsági mentésre milyen stratégiát választanának?

# III.Összegzés

## 1.Alfejezet Címsor 2 és 3

# Mellékletek