**ADATTÁRHÁZAK ÉS ÜZLETI INTELLIGENCIA**

*FÉLÉVES FELADAT DOKUMENTÁCIÓ*

*NBA statisztikák adattárháza*

Készítő:

**Kasza Barnabás**

Intézmény:

Óbudai Egyetem – Neumann János Informatikai Kar

Szak:

Mérnökinformatikus BSc

Tagozat:

Nappali

Félév:

2023-2024 ősz

Tartalomjegyzék

***Fedlap*** ..................................................................................................................... 1

[Bevezetés 2](#_Toc153103885)

[A projekt témája 2](#_Toc153103886)

[A projekt alapját adó adatbázis 2](#_Toc153103887)

[Adatforrás 3](#_Toc153103888)

[Modellezés 3](#_Toc153103889)

[Use-Case meghatározás 3](#_Toc153103890)

[Use-Case tételesen 3](#_Toc153103891)

[DWH 3](#_Toc153103892)

[STAGE 3](#_Toc153103893)

[DM 4](#_Toc153103894)

[ETL 5](#_Toc153103895)

[EXTRACT 5](#_Toc153103896)

[TRANSFORM 6](#_Toc153103897)

[LOAD 7](#_Toc153103898)

[Power BI 8](#_Toc153103899)

[Riport 8](#_Toc153103900)

[Kódok 12](#_Toc153103901)

[STAGE 12](#_Toc153103902)

[DW & VIEWS 14](#_Toc153103903)

[DM 17](#_Toc153103904)

# Bevezetés

## A projekt témája

A projekthez nem a Projektmunka I. tárgy keretein belül elvégzett munkámat fogom felhasználni, hanem egy teljesen új adatbázist, ami az NBA statisztikai adatokból áll.

## A projekt alapját adó adatbázis

A projekt alapjához a Kaggle (kaggle.com) oldalán található „NBA\_games”

(<https://www.kaggle.com/datasets/nathanlauga/nba-games> ) nevű adatkészletet fogom felhasználni és kibővíteni olyan módon, hogy megfeleljen a projektmunka követelményeinek.

## Adatforrás

Az adatbázis az NBA összes mérkőzésének adatait tartalmazza 2004-től 2020 decemberéig. Az adatok között megtalálható a mérkőzés dátuma, ideje, helyszíne, a csapatok nevei, a végeredmény és a statisztikák. A screenshot of a computer

Description automatically generated

*Adatbázis modell*

Ahhoz, hogy dimenzió és tény táblákat tudjak képezni a meglévő objektumokból, nem volt szükség módosításra, mivel a táblák alapértelmezetten tartalmaznak referenciát más táblákra.

# Modellezés

## Use-Case meghatározás

## Use-Case tételesen

• Total\_points\_by\_player

• AVG\_points\_by\_team

• Top\_10\_player\_scores

• Top\_10\_player\_avg\_score

• Team\_point\_by\_year

* Teams\_avg\_point\_home\_and\_away

# DWH

## STAGE

A **STAGE** rétegben történik az egységesítés, ez azt jelenti, hogy közös adattípust kell meghatározni a táblák és mezők között, ezzel segítve a betöltés felépítését.

A **STAGE** réteg táblái megegyeznek a forrásadatbázis tábláival, viszont két különbség van köztük.

* A mezők mind szöveges adattípusúak, ezzel elérve az egységesítést, és könnyítve vele a betöltést
* Elsődleges / másodlagos kulcsok nem szerepelnek, ahogy megszorítások sem, így egyszerűsítve az ETL folyamatot

A screenshot of a computer

Description automatically generated

## DM

A **DM** réteg tartalmazza a HST réteg tábláiból képzett dimenzió és tény táblákat, illetve az ezekből felépülő csillagsémát. A screenshot of a computer

Description automatically generated

Az alábbi dimenzió és tény táblák találhatóak meg a DM rétegben:

* PlayerDim: A játékosokat leíró attribútumok pl. név, csapat
* GameDim: A meccseket leíró attribútumok pl. id, dátum
* TeamDim: Csapatot leíró attribútumok pl. becenév, város
* TeamStatsDim: A csapatok statisztikáit leíró attribútumok pl. conferencia(east,west)
* NBAStatsFact: Minden meccs statisztikai adatai, illetve a dimenzió táblákra mutató kulcs értékek

# ETL

## EXTRACT

Az Extract folyamat a ETL ciklus első lépése, amelynek során a nyers adatokat a forrásrendszerekből a STAGE rétegbe másoljuk. A folyamat során a következő lépéseket hajtjuk végre:

1. Adatok beolvasása

Az első lépés az, hogy a nyers adatokat a forrásrendszerekből beolvassuk egy ideiglenes tárolási területre. Ezt a folyamatot általában egy adatbázis-kezelő rendszer (DBMS) vagy egy adatbetöltő eszköz végzi.

1. Adatok tisztítása

A beolvasott adatokat ezután tisztítani kell, hogy eltávolítsuk a hibás vagy hiányos adatokat. Ezt a folyamatot általában egy adattisztító eszköz végzi.

1. Adatok feltöltése

A tisztított adatokat ezután a STAGE rétegben található táblákba kell feltölteni. Ezt a folyamatot általában a DBMS végzi.

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

*STAGE folyamat*

A screenshot of a computer

Description automatically generated

*STAGE táblák feltöltése*

*Egy Dataflow folyamaton kersztül feltöltöm az összes táblát a STAGE rétegben.Minden adat varchar-ként kerül be a STAGE rétegbe.*

## TRANSFORM

A Transform folyamat felelős az adatok vissza konvertálásáért. Itt beállítjuk a típusokat majd betöltjük a transform rétegbe.

A screenshot of a computer

Description automatically generated

*Transform folyamat*

## LOAD

A Load folyamat felelős a csillagsémában található dimenzió és tény táblák feltöltéséért. Itt történik a historizálás is. SCD (Slowly Chanding Dimensions) használatával valósítottam meg.

*Load folyamat* A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

*Dimenzió táblák és tény tábla feltöltése*

A screenshot of a computer

Description automatically generated

# Power BI

## Riport

* Total\_points\_by\_player

A screenshot of a computer

Description automatically generated

* AVG\_points\_by\_team

A graph with colorful lines and numbers

Description automatically generated

* Top\_10\_player\_scores

A graph of numbers and names

Description automatically generated

* Top\_10\_player\_avg\_score

A blue and white bar graph

Description automatically generated

* Team\_point\_by\_yearA graph of different colored lines

  Description automatically generated

* Teams\_avg\_point\_home\_and\_away A blue and white striped background

  Description automatically generated

# Kódok

## STAGE

1. CREATE TABLE [Games](
2. [GAME\_DATE\_EST] varchar(256),
3. [GAME\_ID] varchar(256),
4. [GAME\_STATUS\_TEXT] varchar(256),
5. [HOME\_TEAM\_ID] varchar(256),
6. [VISITOR\_TEAM\_ID] varchar(256),
7. [SEASON] varchar(256),
8. [TEAM\_ID\_home] varchar(256),
9. [PTS\_home] varchar(256),
10. [FG\_PCT\_home] varchar(256),
11. [FT\_PCT\_home] varchar(256),
12. [FG3\_PCT\_home] varchar(256),
13. [AST\_home] varchar(256),
14. [REB\_home] varchar(256),
15. [TEAM\_ID\_away] varchar(256),
16. [PTS\_away] varchar(256),
17. [FG\_PCT\_away] varchar(256),
18. [FT\_PCT\_away] varchar(256),
19. [FG3\_PCT\_away] varchar(256),
20. [AST\_away] varchar(256),
21. [REB\_away] varchar(256),
22. [HOME\_TEAM\_WINS] varchar(256)
23. )
24. CREATE TABLE [Games\_deatils] (
25. [GAME\_ID] varchar(256),
26. [TEAM\_ID] varchar(256),
27. [TEAM\_ABBREVIATION] varchar(256),
28. [TEAM\_CITY] varchar(256),
29. [PLAYER\_ID] varchar(256),
30. [PLAYER\_NAME] varchar(256),
31. [NICKNAME] varchar(256),
32. [START\_POSITION] varchar(256),
33. [COMMENT] varchar(256),
34. [MIN] varchar(256),
35. [FGM] varchar(256),
36. [FGA] varchar(256),
37. [FG\_PCT] varchar(256),
38. [FG3M] varchar(256),
39. [FG3A] varchar(256),
40. [FG3\_PCT] varchar(256),
41. [FTM] varchar(256),
42. [FTA] varchar(256),
43. [FT\_PCT] varchar(256),
44. [OREB] varchar(256),
45. [DREB] varchar(256),
46. [REB] varchar(256),
47. [AST] varchar(256),
48. [STL] varchar(256),
49. [BLK] varchar(256),
50. [TO] varchar(256),
51. [PF] varchar(256),
52. [PTS] varchar(256),
53. [PLUS\_MINUS] varchar(256)
54. )
55. CREATE TABLE [Players] (
56. [PLAYER\_NAME] varchar(256),
57. [TEAM\_ID] varchar(256),
58. [PLAYER\_ID] varchar(256),
59. [SEASON] varchar(256)
60. )
61. CREATE TABLE [Ranking] (
62. [TEAM\_ID] varchar(256),
63. [LEAGUE\_ID] varchar(256),
64. [SEASON\_ID] varchar(256),
65. [STANDINGSDATE] varchar(256),
66. [CONFERENCE] varchar(256),
67. [TEAM] varchar(256),
68. [G] varchar(256),
69. [W] varchar(256),
70. [L] varchar(256),
71. [W\_PCT] varchar(256),
72. [HOME\_RECORD] varchar(256),
73. [ROAD\_RECORD] varchar(256),
74. [RETURNTOPLAY] varchar(256)
75. )
76. CREATE TABLE [Team] (
77. [LEAGUE\_ID] varchar(256),
78. [TEAM\_ID] varchar(256),
79. [MIN\_YEAR] varchar(256),
80. [MAX\_YEAR] varchar(256),
81. [ABBREVIATION] varchar(256),
82. [NICKNAME] varchar(256),
83. [YEARFOUNDED] varchar(256),
84. [CITY] varchar(256),
85. [ARENA] varchar(256),
86. [ARENACAPACITY] varchar(256),
87. [OWNER] varchar(256),
88. [GENERALMANAGER] varchar(256),
89. [HEADCOACH] varchar(256),
90. [DLEAGUEAFFILIATION] varchar(256))

## DW & VIEWS

1. , CREATE TABLE Games (
2. GAME\_DATE\_EST DATE,
3. GAME\_ID INT PRIMARY KEY,
4. GAME\_STATUS\_TEXT VARCHAR(255),
5. HOME\_TEAM\_ID INT,
6. VISITOR\_TEAM\_ID INT,
7. SEASON INT,
8. TEAM\_ID\_home INT,
9. PTS\_home INT,
10. FG\_PCT\_home FLOAT,
11. FT\_PCT\_home FLOAT,
12. FG3\_PCT\_home FLOAT,
13. AST\_home INT,
14. REB\_home INT,
15. TEAM\_ID\_away INT,
16. PTS\_away INT,
17. FG\_PCT\_away FLOAT,
18. FT\_PCT\_away FLOAT,
19. FG3\_PCT\_away FLOAT,
20. AST\_away INT,
21. REB\_away INT,
22. HOME\_TEAM\_WINS INT,
23. );
24. CREATE TABLE Game\_Details (
25. GAME\_ID INT,
26. TEAM\_ID INT,
27. TEAM\_ABBREVIATION VARCHAR(10),
28. TEAM\_CITY VARCHAR(50),
29. PLAYER\_ID INT,
30. PLAYER\_NAME VARCHAR(50),
31. NICKNAME VARCHAR(50),
32. START\_POSITION VARCHAR(10),
33. COMMENT VARCHAR(255),
34. MIN INT,
35. FGM INT,
36. FGA INT,
37. FG\_PCT FLOAT,
38. FG3M INT,
39. FG3A INT,
40. FG3\_PCT FLOAT,
41. FTM INT,
42. FTA INT,
43. FT\_PCT FLOAT,
44. OREB INT,
45. DREB INT,
46. REB INT,
47. AST INT,
48. STL INT,
49. BLK INT,
50. TO INT,
51. PF INT,
52. PTS INT,
53. PLUS\_MINUS INT,
54. );
55. CREATE TABLE Rankings (
56. TEAM\_ID INT,
57. LEAGUE\_ID VARCHAR(255),
58. SEASON\_ID VARCHAR(255),
59. STANDINGSDATE VARCHAR(255),
60. CONFERENCE VARCHAR(255),
61. TEAM VARCHAR(255),
62. G INT,
63. W INT,
64. L INT,
65. W\_PCT FLOAT,
66. );
67. CREATE TABLE Team (
68. [LEAGUE\_ID] INT,
69. [TEAM\_ID] INT,
70. [MIN\_YEAR] INT,
71. [MAX\_YEAR] INT,
72. [ABBREVIATION] varchar(256),
73. [NICKNAME] varchar(256),
74. [YEARFOUNDED] INT,
75. [CITY] varchar(256),
76. [ARENA] varchar(256),
77. [ARENACAPACITY] INT,
78. [OWNER] varchar(256),
79. [GENERALMANAGER] varchar(256),
80. [HEADCOACH] varchar(256),
81. [DLEAGUEAFFILIATION] varchar(256),
82. );
83. CREATE TABLE Players (
84. PLAYER\_NAME VARCHAR(255),
85. TEAM\_ID INT,
86. PLAYER\_ID INT,
87. SEASON INT,);
88. CREATE VIEW TeamStatsDimView AS
89. SELECT
90. TEAM\_ID,
91. LEAGUE\_ID,
92. SEASON\_ID,
93. STANDINGSDATE,
94. CONFERENCE,
95. TEAM,
96. G,
97. W,
98. L,
99. W\_PCT
100. FROM
101. Rankings;
102. CREATE VIEW GameDimView AS
103. SELECT
104. GAME\_ID,
105. GAME\_DATE\_EST,
106. HOME\_TEAM\_ID,
107. VISITOR\_TEAM\_ID,
108. SEASON,
109. TEAM\_ID\_home,
110. TEAM\_ID\_away
111. FROM
112. Games;
114. CREATE VIEW TeamDimView AS
115. SELECT
116. TEAM\_ID,
117. ABBREVIATION,
118. NICKNAME,
119. YEARFOUNDED,
120. CITY,
121. ARENA,
122. ARENACAPACITY,
123. OWNER,
124. GENERALMANAGER,
125. HEADCOACH,
126. DLEAGUEAFFILIATION
127. FROM
128. Team;
129. CREATE VIEW PlayerDimView AS
130. SELECT
131. PLAYER\_ID,
132. PLAYER\_NAME,
133. SEASON,
134. LEAGUE\_ID
135. FROM
136. Players p
137. Inner join Rankings r on p.TEAM\_ID = r.TEAM\_ID;
139. CREATE VIEW NBAStatsFactView AS
140. SELECT
141. PLAYER\_ID,
142. TEAM\_ID,
143. TEAM\_ID\_home,
144. TEAM\_ID\_away,
145. HOME\_TEAM\_WINS,
146. GAME\_DATE\_EST,
147. GAME\_STATUS\_TEXT,
148. PTS\_home,
149. FG\_PCT\_home,
150. FT\_PCT\_home,
151. FG3\_PCT\_home,
152. AST\_home,
153. REB\_home,
154. PTS\_away,
155. FG\_PCT\_away,
156. FT\_PCT\_away,
157. FG3\_PCT\_away,
158. AST\_away,
159. REB\_away,
160. MIN,
161. FGM,
162. FGA,
163. FG\_PCT,
164. FG3M,
165. FG3A,
166. FG3\_PCT,
167. FTM,
168. FTA,
169. FT\_PCT,
170. OREB,
171. DREB,
172. REB,
173. AST,
174. STL,
175. BLK,
176. TO\_,
177. PF,
178. PTS,
179. PLUS\_MINUS
180. FROM
181. Game\_Details gd
182. Inner join Games g on gd.GAME\_ID = g.GAME\_ID;

## DM

1. CREATE TABLE NBAStatsFact (
2. PLAYER\_ID INT,
3. TEAM\_ID INT,
4. GAME\_ID INT,
5. TEAM\_ID\_home INT,
6. TEAM\_ID\_away INT,
7. HOME\_TEAM\_WINS INT,
8. GAME\_DATE\_EST DATE,
9. GAME\_STATUS\_TEXT VARCHAR(255),
10. PTS\_home INT,
11. FG\_PCT\_home FLOAT,
12. FT\_PCT\_home FLOAT,
13. FG3\_PCT\_home FLOAT,
14. AST\_home INT,
15. REB\_home INT,
16. PTS\_away INT,
17. FG\_PCT\_away FLOAT,
18. FT\_PCT\_away FLOAT,
19. FG3\_PCT\_away FLOAT,
20. AST\_away INT,
21. REB\_away INT,
22. MIN INT,
23. FGM INT,
24. FGA INT,
25. FG\_PCT FLOAT,
26. FG3M INT,
27. FG3A INT,
28. FG3\_PCT FLOAT,
29. FTM INT,
30. FTA INT,
31. FT\_PCT FLOAT,
32. OREB INT,
33. DREB INT,
34. REB INT,
35. AST INT,
36. STL INT,
37. BLK INT,
38. TO\_ INT,
39. PF INT,
40. PTS INT,
41. PLUS\_MINUS INT,
42. PRIMARY KEY (PLAYER\_ID, TEAM\_ID, GAME\_ID, TEAM\_ID\_home, TEAM\_ID\_away),
43. FOREIGN KEY (PLAYER\_ID) REFERENCES PlayerDim(PLAYER\_ID),
44. FOREIGN KEY (TEAM\_ID) REFERENCES TeamDim(TEAM\_ID),
45. FOREIGN KEY (GAME\_ID) REFERENCES GameDim(GAME\_ID),
46. FOREIGN KEY (TEAM\_ID\_home) REFERENCES TeamDim(TEAM\_ID),
47. FOREIGN KEY (TEAM\_ID\_away) REFERENCES TeamDim(TEAM\_ID));
48. CREATE TABLE PlayerDim (
49. PLAYER\_ID INT PRIMARY KEY,
50. PLAYER\_NAME VARCHAR(255),
51. TEAM\_ID INT,
52. SEASON INT,
53. LEAGUE\_ID INT,
54. FOREIGN KEY (TEAM\_ID) REFERENCES TeamDim(TEAM\_ID));
55. CREATE TABLE TeamDim (
56. TEAM\_ID INT PRIMARY KEY,
57. ABBREVIATION VARCHAR(256),
58. NICKNAME VARCHAR(256),
59. YEARFOUNDED INT,
60. CITY VARCHAR(256),
61. ARENA VARCHAR(256),
62. ARENACAPACITY INT,
63. OWNER VARCHAR(256),
64. GENERALMANAGER VARCHAR(256),
65. HEADCOACH VARCHAR(256),
66. DLEAGUEAFFILIATION VARCHAR(256));
67. CREATE TABLE TeamStatsDim (
68. TEAM\_ID INT PRIMARY KEY,
69. LEAGUE\_ID INT,
70. SEASON\_ID INT,
71. STANDINGSDATE DATE,
72. CONFERENCE VARCHAR(255),
73. TEAM VARCHAR(255),
74. G INT,
75. W INT,
76. L INT,
77. W\_PCT FLOAT);
78. CREATE TABLE GameDim (
79. GAME\_ID INT PRIMARY KEY,
80. GAME\_DATE\_EST DATE,
81. HOME\_TEAM\_ID INT,
82. VISITOR\_TEAM\_ID INT,
83. SEASON INT,
84. TEAM\_ID\_home INT,
85. TEAM\_ID\_away INT,
86. FOREIGN KEY (HOME\_TEAM\_ID) REFERENCES TeamDim(TEAM\_ID),
87. FOREIGN KEY (VISITOR\_TEAM\_ID) REFERENCES TeamDim(TEAM\_ID),
88. FOREIGN KEY (TEAM\_ID\_home) REFERENCES TeamDim(TEAM\_ID),
89. FOREIGN KEY (TEAM\_ID\_away) REFERENCES TeamDim(TEAM\_ID)s);