



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA PODNIKATELSKÁ

FACULTY OF BUSINESS AND MANAGEMENT

ÚSTAV INFORMATIKY

INSTITUTE OF INFORMATICS

NÁVRH A IMPLEMENTACE DATABÁZOVÉ APLIKACE PRO ANALÝZU SPORTOVNÍCH VÝKONŮ

DESIGN AND IMPLEMENTATION OF DATABASE APPLICATION FOR ANALYSIS OF SPORTS
PERFORMANCES

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

Filip Ternovszky

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. Jan Luhan, Ph.D., MSc

BRNO 2018

Zadání bakalářské práce

Ústav:	Ústav informatiky
Student:	Filip Ternovszky
Studijní program:	Systémové inženýrství a informatika
Studijní obor:	Manažerská informatika
Vedoucí práce:	Ing. Jan Luhan, Ph.D., MSc
Akademický rok:	2017/18

Ředitel ústavu Vám v souladu se zákonem č. 111/1998 Sb., o vysokých školách ve znění pozdějších předpisů a se Studijním a zkušebním řádem VUT v Brně zadává bakalářskou práci s názvem:

Návrh a implementace databázové aplikace pro analýzu sportovních výkonů

Charakteristika problematiky úkolu:

Úvod
Cíle práce, metody a postupy zpracování
Teoretická východiska práce
Analýza současného stavu
Vlastní návrhy řešení
Závěr
Seznam použité literatury
Přílohy

Cíle, kterých má být dosaženo:

Vytvořit webovou aplikaci pro interpretaci, správu a následnou analýzu dat fotbalových sportovních výkonů z navržené a vytvořené databáze.

Základní literární prameny:

CONOLLY, T., C. BEGG a R. HOLOWCZAK. Mistrovství - Databáze: Profesionální průvodce tvorbou efektivních databází. Praha: Computer Press, 2009. 584 s. ISBN 978-80-251-2328-7.

CUESTA, H. a J. HUF. Analýza dat v praxi. Brno: Computer Press, 2015. 296 s. ISBN 978-80-2-1-4361-2.

LABERGE, R. Datové sklady - Agilní metody a business intelligence. Praha: Computer Press, 2012. 352 s. ISBN 978-80-251-3729-1.

POKORNÝ, J. a M. VALENTA. Databázové systémy. Praha: České vysoké učení technické v Praze, 2013. 265 s. ISBN 978-80-01-05212-9.

SODOMKA, P. a H. KLČOVÁ. Informační systémy v podnikové praxi. 2. aktual. a rozš. vyd. Brno: Computer Press, 2010. 501 s. ISBN 978-80-251-2878-7.

Termín odevzdání bakalářské práce je stanoven časovým plánem akademického roku 2017/18

V Brně dne 28.2.2018

L. S.

doc. RNDr. Bedřich Půža, CSc.
ředitel

doc. Ing. et Ing. Stanislav Škapa, Ph.D.
děkan

Abstrakt

Zameraním tejto bakalárskej práce je návrh a implementácia databázovej aplikácie pre podporu športových analýz. Význam riešení je demonštrovaný vo futbalovom prostredí a zastáva pozitívny vplyv techniky na šport. Vďaka automatizovanému systému zbierania dát a ich rýchlemu poskytnutiu klientom pomocou portálu, vnáša nový rozmer do fungovania športových klubov. Dôkladná analýza dát reprezentuje progresívne riadenie cesty k úspechu v športe.

Abstract

The main aim of this bachelor's thesis is design and implementation of database application for analysis of sports performances. The importance of the solution is demonstrated in the football environment and holds the positive influence of the technique on sport. Thanks to the automated data collection system and their fast delivery to the clients through the portal, brings a new direction to the operation of sports clubs. Data analysis represents progressive management of the path to success in sport.

Kľúčové slová

dáta, informácie, databázová aplikácia, portál, dátová analýza, report

Key words

data, information, database application, portal, data analysis, report

Bibliografická citace

TERNOVSZKY, F. *Návrh a implementace databázové aplikace pro analýzu sportovních výkonů*. Brno: Vysoké učení technické v Brně, Fakulta podnikatelská, 2018. 66 s.

Vedoucí bakalářské práce Ing. Jan Luhan, Ph.D., MSc.

Čestné prohlášení

Prohlašuji, že předložená bakalářská práce je původní a zpracoval jsem ji samostatně. Prohlašuji, že citace použitých pramenů je úplná, že jsem ve své práci neporušil autorská práva (ve smyslu Zákona č. 121/2000 Sb., o právu autorském a o právech souvisejících s právem autorským).

V Brně dne 15. května 2018

.....

podpis autora

Pod'akovanie

Moja veľká vďaka patrí pánovi Ing. Janovi Luhanovi, Ph.D., MSc. za vedenie mojej bakalárskej práce, a za jeho cenné a odborné rady.

OBSAH

Úvod.....	10
1 CIELE PRÁCE, METÓDY A POSTUPY SPRACOVANIA	12
2 TEORETICKÉ VÝCHODISKÁ PRÁCE	14
2.1 Databázový prístup.....	14
2.1.1 Dáta a informácie.....	14
2.1.2 Databáza.....	15
2.1.3 Rozdelenie databáz	15
2.1.4 Systém riadenia bázy dát	16
2.1.5 Databázová aplikácia	16
2.1.6 Databázový systém	16
2.1.7 Pohľady.....	17
2.1.8 Triggre	18
2.1.9 Replikácie	18
2.2 Dátový a informačný prístup.....	19
2.2.1 Dáta, informácie a znalosti	19
2.2.2 Povaha dát.....	19
2.2.3 Zdroje dát.....	20
2.2.4 Problém.....	21
2.2.5 Príprava dát	21
2.2.6 Prieskum dát	21
2.2.7 Vizualizácia výsledkov	22
2.2.8 Business intelligence.....	22
2.2.9 Dolovanie dát.....	23
2.3 Dátové modelovanie.....	24
2.3.1 Dátový typ a dĺžka	24
2.3.2 Relácia	25
2.3.3 Typy kľúčov.....	25
2.3.4 Cudzie kľúče a referenčná integrita	26
2.3.5 Funkčné závislosti.....	26
2.3.6 Normalizácia	26
2.4 Prístup programovacích jazykov a aplikácií	27
2.4.1 Jazyk SQL.....	27
2.4.2 Jazyk PHP	28
2.4.3 MySQL	29
2.4.4 MSSQL Server.....	29
2.4.5 Power BI	29
2.4.6 Microsoft Azure.....	30
3 ANALÝZA SÚČASNÉHO STAVU.....	31
3.1 Analýza českého trhu s BI.....	31
3.2 Software as a Service	32
3.3 Analýza využitia športových analýz vo svete a v Českej republike	33
3.3.1 Premier League	33
3.3.2 HET Liga	34
3.3.3 SAP Sports One	35
3.4 Metódy automatizovaného zbierania dát v športových stretnutiach	37
3.4.1 Spracovanie projekcie.....	37

3.4.2	Vytvorenie vizualizácie	38
3.5	Zhrnutie súčasného stavu	39
4	VLASTNÉ NÁVRHY RIEŠENÍ.....	40
4.1	Návrh a implementácia platformy.....	40
4.1.1	Definícia entít, atribútov, dátových typov a dátových dĺžok.....	41
4.1.2	Určenie kľúčov a zaistenie referenčnej integrity	42
4.1.3	Užívateľské prispôsobenie.....	42
4.2	Štatistiky a reporting	47
4.2.1	Vypracované reporty.....	48
4.2.2	Komplexný report zo sledovaného obdobia	48
4.2.3	Individuálne porovnanie hráčov	49
4.2.4	Znázornenie vplyvu chyby na rating hráča.....	49
4.2.5	Analýza spoľahlivosti hráčov v osobných súbojoch	50
4.3	Administrácia databázového servera a portálu	52
4.4	Potenciál na trhu.....	53
4.4.1	Vízia.....	53
4.5	Ekonomické zhodnotenie	56
	ZÁVER	60
	ZOZNAM POUŽITÝCH ZDROJOV	61
	ZOZNAM POUŽITÝCH SKRATIEK A SYMBOLOV.....	63
	ZOZNAM OBRÁZKOV	64
	ZOZNAM TABULIEK	65
	ZOZNAM PRÍLOH.....	66

ÚVOD

Posledné desaťročie sa môžeme stretávať s čoraz väčším výskytom súvislostí medzi slovami šport, futbal a biznis. Príliv peňazí do športu z neho robí investičný a marketingový nástroj. Rozmohol sa trend zaobstarávania agentov zastupujúcich športové individuality, čo spolu so záujmami bohatých majiteľov klubov vnáša do obehu množstvo finančného kapitálu.

Existuje množstvo vecí, nad ktorými sa nezamýšľame a neskúmame ako fungujú na pozadí, no ich výsledný produkt berieme ako fakt. Zaznamenávanie a uchovávanie informácií sledujeme už od dávnej histórie. Informácie pomáhali a mali vplyv nie len na obchodníkov, ale aj na ľudí, ktorí sa v tomto odbore profesijne nepohybovali.

Nie je náhoda, že vznik písma ako dorozumievacieho prostriedku, ktorý časom veľmi rýchlo získal prívlastok ako prostriedok zapisovací, mal slúžiť najmä k budovaniu systému evidencie. Triedenie evidovaných údajov v dátach umožňuje rýchle vyhľadávanie a hlavne preto sa dnes so systémom usporiadaných dát stretávame deň čo deň.

Dnes nie ako holé fakty – dáta, ale informácie, sú súčasťou, takmer každého existujúceho podniku a stretávame sa s nimi denne či už pri štúdiu, nákupe a objednávaní tovaru respektíve služieb. Nachádzajúc sa na papieri alebo v počítači ako súbor usporiadaných dát, sa dostávame k pojmu databáza.

Motivácií na vytvorenie databázového systému je niekoľko. Vďaka informáciám má podnik prehľad vo svojej činnosti, klientoch, tovaroch, poskytovaných službách a pomáhajú mu analyzovať a premieňať svoje poznatky k podpore strategického rozhodovania.

Táto práca je rozdelená do troch častí – teoretickej, analytickej a praktickej. Jej snahou je spojiť využitie možností informačných technológií spolu s ľudským faktorom výkonu v športovej činnosti. Výsledný produkt má slúžiť k lepšiemu pochopeniu individuality resp. tímu, opierajúc sa o získané informácie zo zozbieraných dát, čím nachádza miesta,

priestory, k potencionálnemu nie len športovému, ale aj finančnému, marketingovému a reputačnému rastu klubov.

V teoretickej časti sú zhrnuté základné východiská, potrebné k pochopeniu a objasneniu danej problematiky. Hlavné sústredenie je na technickú časť projektu.

Analytická časť sa venuje hlavne momentálnym trendom v oblasti analýzy dát v podnikoch a športových kluboch. Skúma možné alternatívy využitia softvérových prostriedkov a je motiváciou pre finálnu časť tejto bakalárskej práce.

Praktická časť obsahuje vlastné návrhy riešení a demonštruje nadobudnuté poznatky na príkladoch. Vytvorený je komplexný návrh riešenia implementovateľného do regiónu východnej Európy obsahujúci prakticky vytvorenú aplikačnú platformu.

1 CIELE PRÁCE, METÓDY A POSTUPY SPRACOVANIA

Hlavným cieľom tejto bakalárskej práce je navrhnúť a implementovať databázovú aplikáciu, slúžiacu ako zdroj informácií, k analýze športových, futbalových, výkonov. Práca primárne demonštruje poznatky na vlastnom zozbieranom vzorku dát, nadobudnutým pozorovacou metódou a systematickým organizovaním. Hmatateľným výsledkom tejto práce je vypracovanie analytických reportov z reálneho prostredia, ktorých úlohou je poskytovať manažérom, scoutom, agentom alebo fanúšikom, iný ako povrchný pohľad na hráčske a tímové výkony.

Čiastkové ciele práce:

- spracovať teoretické východiská práce slúžiace k vypracovaniu technického riešenia platformy,
- vykonať analýzu súčasného stavu a identifikovať previazanosť medzi podnikovými a športovými softvérovými alternatívami,
- navrhnúť a implementovať časť platformy databázovej aplikácie,
- navrhnúť správu automatizovanej administrácie databázového servera,
- pomocou podporných programov vypracovať analytické reporty,
- ekonomicky zhodnotiť dostupnosť navrhnutého riešenia.

Dosiahnutie cieľov sa opiera o nadobudnuté technické poznatky, vlastné skúsenosti s prostredia futbalu, a nájdenie prepojenia medzi športom a informačnými technológiami. Tieto poznatky sú postupne využité pri tvorení výsledného produktu, počínajúc technickým pozadím až po prezentáciu hlásení.

Pri spracovaní bakalárskej práce bolo použitých niekoľko metód a postupov spracovania. Analýza ako rozbor faktov, vlastností a vzťahov, ktoré je nutné chápať pri riešení problematiky databázových systémov a podporných spolupracujúcich aplikácií. Na opačnú stranu, syntéza docieľuje zovšeobecnenie, čo pri množstve dostupných produktov, funkcií a iných rôznorodostí, uľahčí prístup k vybranej téme.

Metódy abstrakcie a konkretizácie slúžia k vytvoreniu istého predpokladu vlastností a ďalších vzťahov pre definíciu jednotlivých procesov systému. Špecifikuje sa zameranie entít v databázovom systéme, deklaruje sa vymedzenie voči potencionálnym zákazníkom vo vybranom regióne.

Vzniknuté vzťahy a nadobudnuté informácie vo výsledku do hĺbky viac obohatí využitie indukcie a dedukcie. Jedná sa o výskyt javov a dokázanie záverov v praktickej časti práce.

Využitou metódou, ktorá obohacuje a skúma výsledky, z oblasti dolovania dát, bolo k-means clustering. Je nutné sa na jednotlivé metódy a postupy pozerat' komplexne do hĺbky, šírky, ale aj zvonku a zvnútra. Každá z metód napomáha k dôslednejšiemu pochopeniu a dosiahnutiu vzdelania v súvislosti problematiky databázových systémov a športových analýz.

2 TEORETICKÉ VÝCHODISKÁ PRÁCE

V teoretických východiskách sú popísané znalosti a pojmy, ktoré sú potrebné pre preskúmanie a prácu s danou tematikou. Slúžia ako podklady pre návrhy vlastných riešení. Kapitola postupne prechádza databázový, dátový a informačný prístup, dátové modelovanie a prístup programovacích jazykov.

2.1 Databázový prístup

Nasledujúca kapitola popisuje všeobecný pohľad na databázový systém, jeho fungovanie a vzťahy medzi jednotlivými prvkami systému.

2.1.1 Dáta a informácie

Rozdiel medzi dátami a informáciami je veľmi jasný. Dáta chápeme ako nespracované fakty, ktoré majú istú dôležitosť pre jednotlivca alebo organizáciu (2, s. 36).

Dáta, ktoré prešli spracovaním alebo dostali nejakú štruktúru, ktorá im dáva význam, sa nazývajú informácie (2, s. 36).

Pri transformovaní dát na informácie a ich následnom uchovaní je potrebné si položiť niekoľko otázok. Aké informácie chceme zachovať? Prečo ich chceme zachovať? Ako a kde ich chceme uložiť? Na prvý pohľad sa tieto otázky môžu zdať zbytočné, ale pre konečný výsledok stojí za to nad nimi uvažovať (4).

2.1.2 Databáza

K správne mu chápaniu tohto pojmu slúžia tieto definície.

1. Databáza je veľké úložisko dát, ktoré môže byť používané súčasne viacerými klientami. Všetky uložené dáta, ktoré užívatelia požadujú, sú zapracované s minimálnym množstvom duplikácií. Môžeme ju chápať ako zdieľaný zdroj spoločnosti (2, s. 36).
2. Databáza tiež obsahuje popis zozbieraných dát, ktoré sa nazývajú metadáta. Z tohto dôvodu autori definujú databázu ako seba-popisujúcu kolekciu integrovaných záznamov (2, s. 36 – 39).
3. Pri analýze informačných potrieb organizácie sa prejavuje snaha o identifikáciu dôležitých objektov, ktoré je treba reprezentovať v databáze, a logických vzťahov medzi týmito objektami (2, s. 36 – 39).

2.1.3 Rozdelenie databáz

Databázy môžeme rozdeliť na:

- papierové databázy - aj keď ich súčasné využitie je takmer nulové, stále sa jedná o istý typ usporiadaných dát a preto ich radíme do členenia (2, s. 36),
- súborovo orientované databázy - typickým javom pre tieto databázy je, že pre každú tabuľku sa používa jeden samostatný súbor. Sú predchodcami relačných databázových systémov (3),
- relačné databázové systémy - vyznačujú sa lepšou dátovou integritou, za ktorú je zodpovedná samotná databáza. Tá taktiež preberá zodpovednosť za dáta a tie sú potom bezpečnejšie (3),
- objektovo orientované databázy - dáta sa neukladajú do tabuliek, ale ako objekt s vlastnosťami. Ponúka mechanizmy ako sú dedičnosť a zapuzdrenie (3).

Z pohľadu správy ich môžeme rozdeliť na transakčné a analytické. V transakčných sa dáta neustále menia a odpovedajú aktuálnemu stavu diania. Slúžia ako hlavný prostriedok

pre zhromažďovanie a správu dát podnikov po celom svete. Ich označenie je OLTP (online transaction processing) (2, s. 461).

Opačným prípadom sú analytické databázy, v ktorých sa ukladajú staršie a časovo závislé dáta prebraté zvyčajne z transakčných databáz. Analytické databázy plnia iný typ úloh a ich štruktúra je odlišná. Sú určené na zobrazovanie štatistických dát, trendov, a podobne (2, s. 466).

2.1.4 Systém riadenia bázy dát

Je softwarový systém, ktorý slúži na definíciu, vytváranie a údržbu databázy a tiež k nej poskytuje riadený prístup (2, s. 36).

Okrem iného umožňuje užívateľom vkladať, aktualizovať, mazať a vyvolávať dáta z databázy. Pokiaľ existuje miesto, kde sú všetky dáta uschované a popis dát, je možné využiť všeobecnú možnosť dopytovania na tieto dáta pomocou dopytovacieho jazyka ako je napríklad SQL (z angl. Structured Query Language) (2, s. 38 – 39).

2.1.5 Databázová aplikácia

Pod týmto pojmom sa skrýva počítačový program, ktorý komunikuje s databázou vyvolávaním požiadavku pre systém riadenia bazy dát (2, s. 36).

Klienti využívajú väčší počet databázových aplikácií, ktoré používajú k vytváraniu a správe databázy a generovaniu informácií. Tieto programy môžu byť napríklad online aplikáciami s prístupom cez webový portál (2, s. 39).

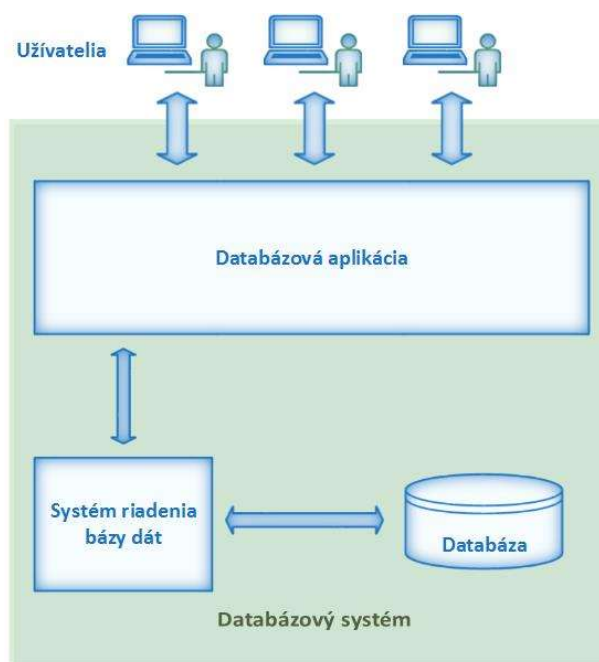
2.1.6 Databázový systém

Do databázy ukladáme objekty z reálneho sveta, ktoré sú identifikované, vo forme tabuliek. Tabuľky sú vzájomne logicky prepojené. Vďaka týmto prepojeniam dokáže

databázová aplikácia, t. j. počítačový program komunikujúci s databázou, vyvolávať príkazy jazyka SQL pre systém riadenia bazy dát (2, s. 36 – 39).

Systém riadenia bazy dát umožňuje definovať a vkladať dáta, triediť ich, vyberať ich, pridávať, aktualizovať, mazať (2, s. 36 – 39).

Z databázy, databázovej aplikácie a systému riadenia bazy dát sa skladá databázový systém (2, s. 36 – 39).



Obrázok č. 1: Databázový prístup užívateľov do databázy pomocou databázovej aplikácie a systému riadenia bazy dát (Vlastné spracovanie podľa 2, s. 40)

2.1.7 Pohľady

Pohľady sú virtuálne tabuľky, poskytujúce každému užívateľovi náhľad na individuálne upravenú časť databázy. Zároveň je jej podmnožinou. Slúžia hlavne k dopytovaniu, ale používajú sa aj v ďalšej užitočnej aplikácií, keď je potrebné z bezpečnostných dôvodov užívateľov „odstrihnúť“ od niektorých dát (2, s. 36 – 39).

Predstavujú silný mechanizmus pre zlepšenie kontroly prístupu k dátam. Ide o tzv. nezávislosť na logických dátach. To znamená, že užívatelia pohľadov si nie sú vedomí prípadných zmien v schémach databázy (5, s. 114).

2.1.8 Triggre

V prípade triggrov, resp. spúšťačov, sa jedná v podstate o uložené procedúry, ktoré sa vykonávajú v prípade určitej, dopredu definovanej udalosti. Existuje nadväznosť na príkazy INSERT, UPDATE a DELETE, pomocou ktorých sa manipuluje s dátami (viac v kapitole 2.2 Jazyk SQL). Využitie spúšťačov prispieva k automatizácii databázových procesov a systému ako celku (2, s. 508).

2.1.9 Replikácie

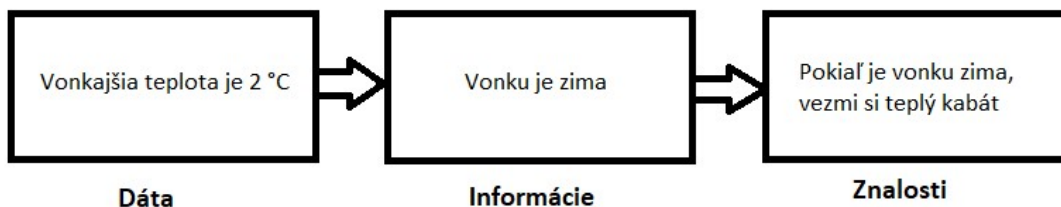
Tvoria distribuovaný databázový systém, v ktorom sa definujú ako proces kopírovania a údržby databázových objektov. Zmeny aplikované na jednom mieste sú zachytené a uložené predtým ako sa distribuujú do vzdialenej lokácie. Ich hlavnou výhodou je dostupnosť údajov v prípade poruchy servera alebo výpadku siete. Ide o vytváranie duplikátov údajov v reálnom čase – požiadavky sa vykonávajú na Master databáze a replikácie na Slave (18, s. 145).

2.2 Dátový a informačný prístup

V predchádzajúcej kapitole bolo popísané, čo sú to dáta a informácie. Tá nasledujúca sa bude venovať tomu, ako k nim pristupovať a chápať ich. Poskytne detailnejší pohľad k problematike transformácie dát na znalosti.

2.2.1 Dáta, informácie a znalosti

Dáta predstavujú fakty o svete. Môžu nimi byť vek alebo teplota, ale sú to v podstate len obyčajné čísla. Informácia sa objaví vtedy, keď s týmito číslami začneme bližšie pracovať a zistíme ich hodnotu a význam, môžeme vďaka nim vytvárať vedomé rozhodnutia. Avšak o znalostiach môžeme hovoriť až vtedy, keď sa naše zozbierané dáta a informácie zmenia na istú sadu pravidiel, ktoré nám pomáhajú sa rozhodovať. Znalosti v praxi zahŕňajú teoretické, ale aj praktické porozumenie danej oblasti. Prediktívnou analýzou, ktorú aplikujeme, dokážeme stimulovať chovanie a nastaviť si presné očakávania do budúcnosti (6, s. 25).



Obrázok č. 2: Prevedenie dát na znalosti (Vlastné spracovanie podľa 6, s. 25)

2.2.2 Povaha dát

Vnímať dáta musíme vždy v množnom čísle. Stretneme sa s nimi vo väčšine situácii v živote, či už v štruktúrovanej alebo neštruktúrovanej podobe. Sú stavebným prvkom všetkých ľudských aktivít. Dáta môžeme vnímať dvomi spôsobmi:

- kategorické – nominálne a rádové,
- číselné – spojité a nespojité (6, s. 26).

Kategorické dáta sú hodnoty a pozorovania, ktoré sa dajú rozdeliť do skupín, kategórií, dvomi spôsobmi. Nominálnou kategorickou premennou môže byť bývanie – vlastné, prenájom. Rádová premenná má už svoje zoradenie, napríklad vek – mladosť, dospelosť, staroba (6, s. 26).

Číselné dáta sú merateľné hodnoty či pozorovania. Nespojité sú tie, ktoré možno spočítať, ale ktoré sa líšia a sú samostatné – počet znakov v texte. Pod spojitými chápeme také, ktoré môžu nadobudnúť akúkoľvek hodnotu v intervale (konečnom aj v nekonečnom), napríklad cena Bitcoinu počas minulého roka (6, s. 26).

2.2.3 Zdroje dát

Základom každej kvalitnej analýzy, či už regresnej alebo štatistickej, je kvalitné zbieranie dát. Nie vždy však platí, že jednoducho a pohodlne získané dáta aj splnia svoj účel a budú užitočné. Žiada sa niekoľko možností a postupov aké dáta k analýze použijeme (7).

Primárne dáta môžeme zbierať z dôvodu nesprávnosti, neúplnosti alebo neaktuálnosti sekundárnych. Výhodou je, že tento súbor dát zodpovedá predmetu skúmania. Na druhú stranu, zber primárnych dát je náročný z ekonomického, časového, ale aj technického pohľadu (7).

Sekundárne dáta prispievajú svojim pohodlnejším spôsobom nadobudnutia. Ich nevýhoda môže spočívať vo fakte, že nie vždy boli získané za rovnakým účelom. Z tohto dôvodu je fáza spracovania náročnejšia ako pri primárnych dátach (7).

Bežným spôsobom ukladania dát sú textové súbory. Tie sa dajú ľahko prevádzať do iných formátov, tie textové môžu byť napríklad:

- CSV,
- TXV,
- XML,
- JSON.

Najpoužívanejší nástroj slúžiaci na dátovú analýzu je Microsoft Excel. Ponúka niekoľko veľmi dobrých funkcií a možnosť využitia Visual Basic for Application. Súbor Excelu vieme jednoducho previesť na jeden z vyššie spomenutých formátov (6, s. 39-40).

Organizovanou zbierkou dát je databáza, ktorá bola bližšie popísaná v kapitole 2.1.2.

2.2.4 Problém

Problém definujeme všeobecnými otázkami. „*Porozumenie predpokladom a cieľom z danej perspektívy je kľúčom k tomu, aby bol projekt dátovej analýzy úspešný.*“ (6, s. 27)

Typy kladených otázok pri dátovej analýze:

- deduktívne,
- prediktívne,
- deskriptívne,
- vysvetľujúce,
- všeobecné,
- korelačné (6, s. 27).

2.2.5 Príprava dát

Príprava dát ma podstatu v zaistení, vyčistení, normalizácii a prevode na optimálnu sadu. Kladie sa dôraz na elimináciu kvalitatívnych odchýlok (6, s. 28).

Kapitola 2.3 sa venuje príprave a modelovaniu dát podrobnejšie.

2.2.6 Prieskum dát

Vo svojej podstate ide o pohľad na dáta v grafickej, respektíve štatistickej podobe. Snažíme sa nájsť obrazce, spojenia a vzťahy. Veľmi blízko súvisí s vizualizáciou, z ktorej je možné vysledovať zmysluplné vzory (6, s. 28).

2.2.7 Vizualizácia výsledkov

Posledný krok v procese analýzy musí zodpovedať niekoľkým otázkam. Odpovedá sa na spôsob vizualizácie výsledkov (tabuľky, grafy atď.) a jej umiestnenie (plagát, PC, smartfón, web apod.). Každá voľba musí závisieť od konkrétnej analýzy. Účelom je nechať vyniknúť nadobudnuté skutočnosti a vzťahy. Predstavuje jednoduchý spôsob náhľadu na dátovú sadu, ktorý nemá len funkciu popisu a skúmania, ale aj podporu rozhodovania (6, s. 29 – 30).

2.2.8 Business intelligence

Výkonnostná vrstva a vrstva užívateľskej prezentácie umožňujú využitie systému dátového skladu, inak povedané výstup informácií, formou analýzy dát. Pri práci s podobným riešením je možné sa zaujímať o niektoré aspekty, napríklad:

- informácie vo výsledných zostavách,
- základný návrh dimenzií a metrík,
- objemy dát,
- indexy,
- dátové oddiely,
- načítanie dát,
- kvalita dát,
- zdroj dát (9, s. 143).

2.2.9 Dolovanie dát

Elementárne tento pojem označuje snahu nájsť v dátach zmysel, ktorý nemožno vidieť „na povrchu“. Úvodom sa musia prijať hypotézy a potom pomocou dopytov na dáta zisťovať komplexné súvislosti. Nehľadajú sa odpovede na počty, objemy predaja apod. Výsledkom sú vzťahy, súvislosti, profilovanie, vedúce k lepšiemu zameraniu a potenciálnemu navýšeniu ziskovosti (9, s. 153).

Algoritmus zhukovania k-means optimalizuje ťažiská jednotlivých zhukov a následne priradí dané prvky k ťažisku. V každom opakovaní sa počíta vzdialenosť prvku od ťažiska a táto hodnota musí byť pre každý zhuk minimálna (14).

2.3 Dátové modelovanie

Holé fakty alebo suché dáta k analýze nestačia. Musia dostať podobu, na ktorej je závislé vyťahovanie informácií a získavanie znalostí. Proces organizácie a štruktúrovania dát nie je vždy jednoduchý – figuruje v ňom stret názorov užívateľa, návrhára a programátora. Ďalšia kapitola priblíži relačné modelovanie.

2.3.1 Dátový typ a dĺžka

Údaje, ktoré o zvolenom objekte chceme sledovať, popisuje dátová štruktúra. Štruktúra je pevne daná, či už údaje vyplníme alebo nie (10).



	Jméno	Bydliště	Povolání	Telefon
	Jan Novák	Brno	Instalater	
	Jiřina Máčková	Praha	Švadlena	5444235621
Ø	Tomáš Vrbník	Zvolen	Montér	

Obrázok č. 3: Dátová štruktúra a obsah vety (10)

Pre zvolený atribút je potrebné vybrať dátový typ a jeho dĺžku. Elementárne rozlišujeme tieto dátové typy:

- textový – meno: Peter,
- číselný – plat: 1 500 €,
- dátumový – dátum narodenia: 1.1.1990,
- objektový – fotka na karte,
- logický – muž: true (10).

Rozhodovanie o dĺžke je naviazané na zvolený atribút. Napríklad pre atribút „Názov futbalového klubu“ musíme uvažovať aj o možnosti zápisu dlhších názvov ako napríklad „Ballspielverein Borussia 09 e.V. Dortmund“ – obsahuje 41 znakov aj s medzerami. Preto môžeme vybrať dátový typ textový, pre počet znakov 60 (10).

2.3.2 Relácia

V úvode je nutné si pojmy vyjasniť. Sledované veci, predmety, činnosti, apod., sa nazývajú entity. Reprezentujú dôležitosť pre užívateľa databáze. Akoby sa z prvého posudku mohlo javiť, pojem relácia značí nejaký druh vzťahu. Toto tvrdenie nie je celkom správne. Reláciu definujeme ako dvojrozmernú tabuľku (v praxi sa tieto pojmy bežne zamieňajú), skladajúcu sa z riadkov a stĺpcov, ktorá musí spĺňať nasledujúce vlastnosti:

1. riadky obsahujú dáta o entite,
2. stĺpce obsahujú dáta o atribútoch entity,
3. bunky v tabuľke uchovávajú jedinú hodnotu,
4. všetky položky v stĺpci sú rovnakého druhu,
5. všetky stĺpce majú jedinečný názov,
6. na poradí stĺpcov a riadkov nezáleží,
7. žiadne dva riadky nesmú obsahovať identické hodnoty (8, s. 79).

2.3.3 Typy kľúčov

Principiálne rozlišujeme kľúče jedinečné a nejedinečné. Ich hlavnou úlohou je identifikovanie každého riadku v relácii, respektíve odkazovanie na ďalší, jednoznačne identifikovateľný, riadok inej relácie. Obsahuje jeden alebo viac atribútov, pričom kľúč obsahujúci dva a viac atribútov sa označuje ako zložený kľúč – rovnako ako kľúče založené na jednom stĺpci (hodnotách atribútu), môžu byť zložené kľúče jedinečné či nejedinečné (8, s. 81).

Z konkrétnejšieho pohľadu, primárny kľúč je zvolený kandidátny kľúč, ktorý identifikuje všetky riadky v relácii. Kandidátny kľúč je taký kľúč, o ktorom sa uvažuje ako o primárnom, ale nebol zvolený ako primárny, aj napriek tomu, že spĺňa vlastnosť identifikácie každého riadku v relácii. Tento vzťah možno prirovnať k voľbám – vždy môže vyhrať len jeden kandidát (8, s. 81).

2.3.4 Cudzie kľúče a referenčná integrita

Pri reprezentovaní vzťahu medzi reláciami, umiestnime hodnotu z jednej relácie do druhej. Používajú sa hodnoty primárneho kľúča. Atribút uchovávajúci tieto hodnoty sa v praxi označuje ako cudzí kľúč. Pravidlo, ktoré je potrebné vo väčšine prípadov zaistiť aby každá hodnota cudzieho kľúča, odpovedala hodnote toho primárneho. Ak je toto pravidlo potvrdené, hovoríme o obmedzení referenčnej integrity (8, s. 81).

2.3.5 Funkčné závislosti

Pomocou funkčných závislostí môžeme navrhnúť schému databázy a predísť problémom s redundanciou, nekonzistenciou, alebo zablokovaním dát pri vkladaní. Je to vzťah vytvorený medzi atribútmi. Najlepším a najčastejším príkladom je vzťah dátumu narodenia a rodného čísla – atribút „dátum narodenia“ je funkčne závislý na atribúte „rodné číslo“ – záznamy s rovnakými rodnými číslami (pozn. pred lomkou; v prípade žien – mesiac + 50) majú rovnaký dátum narodenia (8, s. 91).

2.3.6 Normalizácia

Zámerni vytvorenia normalizovaného dátového modelu sú zamedzenie výskytu duplicitných dát a zaručenie najjednoduchšej formy ich konzistencie. Existuje niekoľko úrovní normalizácie, ale v praxi sa obvykle využívajú tieto tri nasledujúce:

- prvá normálna forma má za úlohu zaistiť, aby mal každý riadok len atomické hodnoty a obsahoval jednoznačný identifikátor,
- druhá normálna forma musí spĺňať požiadavky prvej a navyše oddeliť polia, ktoré sú nezávislé na kľúči, do samostatnej tabuľky,
- tretia normálna forma musí vyhovovať podmienkam prvej a druhej, a okrem toho musí odstrániť závislosti neklúčových polí v tabuľke umiestnením ich do samostatnej tabuľky (9, s. 201).

2.4 Prístup programovacích jazykov a aplikácií

Aby bolo možné efektívne implementovať a využívať vytvorený dátový model, je dôležité mať znalosti jazyka SQL. Správna kombinácia jazyka SQL a PHP môže byť výhodnejšia pre užívateľský prístup cez web. K správe databázy a analýze informácií prispievajú aplikácie typu:

- PHPMyAdmin,
- MySQL,
- MSSQL,
- Power BI,
- Salesforce a pod.

Táto kapitola teoreticky popíše dôležité aspekty prístupu k programovacím jazykom a aplikáciám využiteľných v databázových systémoch.

2.4.1 Jazyk SQL

SQL je najrozšírenejším databázovým jazykom. Jeho počiatky siahajú do roku 1974, kedy prototypová implementácia tohto jazyka v projektoch firmy IBM znamenala vývin a rozvoj sveta databáz (5, s. 57).

2.4.1.1 Hlavné vlastnosti SQL

Hlavnými vlastnosťami jazyka SQL sú:

- relatívne ľahko naučiteľný jazyk,
- neprocedurálny jazyk – je potreba zadať aké informácie požadujeme, nie ako ich získať,
- má voľný formát, čo znamená, že nie sú presne určené miesta obrazovky na písanie kódu,
- štruktúra príkazov je založená na štandardných anglických slovách,
- môže byť realizovaný mnohými užívateľmi (5, s. 57).

Inštitút vzdelávania informačných technológií udáva štyri hlavné dôvody prečo ovládať jazyk SQL. Sú nimi využiteľnosť, komunikácia, nenáročnosť a bezpečnosť. Využívajú ho databázy MySQL, Oracle, Microsoft SQL a ďalšie iné (1).

2.4.1.2 Príkazy

Nasledujúca tabuľka popisuje funkciu niektorých príkazov jazyka SQL.

Tabuľka č. 1: Vybrané príkazy jazyka SQL (Zdroj 5, s. 60 – 118)

PRÍPADY POUŽITIA	PRÍKAZY
DEFINÍCIA DÁT	CREATE TABLE, ALTER TABLE, DROP TABLE, CREATE SCHEMA
MANIPULÁCIA S DÁTAMI	SELECT, INSERT, DELETE, UPDATE
POSTUPNOSŤ SPRACOVANIA V PRÍKAZE SELECT	FROM, WHERE, GROUP BY, HAVING, SELECT, ORDER BY
AGREGAČNÉ FUNKCIE	COUNT, SUM, AVG, MIN, MAX
ZLÚČENIE TABULIEK	INNER JOIN, OUTER JOIN, FULL JOIN, CROSS JOIN
VYTVORENIE POHLADU	CREATE VIEW

2.4.2 Jazyk PHP

Dôležitosť tohto programovacieho jazyka spočíva v jeho špecializácii sa na webové stránky. Obsahuje rozsiahly súbor funkcií v základnej knižnici. Prirodzená podpora databázových systémov a rôznych platforiem ako Linux a Windows má na svedomí, že PHP sa stal štandardom prakticky všade. Tento voľne dostupný skriptovací jazyk mení stránky na dynamické a spracúva udelené príkazy od užívateľa (10).

2.4.3 MySQL

Kľúčovými vlastnosťami databázového servera sú výkon pre správu dát a udržiavanie úložiska. Systém musí rýchlo reagovať na požiadavky užívateľa a efektívne ukladať nazhromaždené údaje. Open source software, momentálne patriaci pod Oracle, MySQL spĺňa až na pár výnimiek kvalitatívne požiadavky na úrovni svojich cenovo náročnejších kolegov. Dokáže spolupracovať s celou škálou operačných systémov, a keďže jeho cena je nulová, jeho využitie na webových portáloch podporujúcich správu databáz je enormné (12).

2.4.4 MSSQL Server

Hlavnou funkciou Microsoft SQL servera je správa relačných databázových systémov. Primárne sa využíva na vkladanie, úpravu, mazanie, záchranu dát a kontrolu prístupov. Disponuje prepracovanými funkciami pre administráciu databázového servera a databázy. Ponúka širokú škálu služieb a mainstreamové a špecializované edície. Zameranie spoločnosti Microsoft na trhu začína od malých aplikácií jedného stroja, až po veľké internetové aplikácie so súbežným použitím (13).

2.4.5 Power BI

Microsoft eviduje prvé korene tejto interaktívnej aplikácie v roku 2009. Užívateľské rozhranie sa dostalo do dnešnej podoby s balíkom Office 365 v roku 2013. Ponúka interaktívne vizualizácie, ktoré si klient samostatne vytvára, a tým vznikajú jedinečné analytické reporty. Možnosťami sú využitia schopností dátového skladu, prípravy dát alebo objavovania dát (11, s. 1 - 4).

2.4.6 Microsoft Azure

„Windows Azure can be anything you want it to be.“ (17, s. 1)

Produkt Microsoft Azure je cloudová platforma poskytujúca širokú škálu služieb. Umožňuje vytvárať, nasadzovať, spravovať riešenia takmer na akýkoľvek účel. Ponúka rovnaké možnosti pre malé lokálne podniky, ako aj pre obrovské podniky s kancelármi po celom svete. Azure môže poskytovať platformu k analýze, dolovaniu dát alebo dokonca vytváraniu aplikácií (17, s.1).

3 ANALÝZA SÚČASNÉHO STAVU

Teoretické znalosti z odvetvia sú veľmi dôležité pre pochopenie a spracovanie vlastných návrhov riešenia. Minimálne rovnakú dôležitosť musíme klásť na analýzu súčasného stavu. Slúži k pochopeniu momentálnych trendov a nájdeniu príležitostí pre nové riešenia. Kapitola postupne prejde od podnikových trendov u databázovo-analytických produktoch, až k trendom v športových databázovo-analytických produktoch demonštrovaných na futbale, ktoré vedú k pochopeniu vzájomného vzťahu medzi nimi.

3.1 Analýza českého trhu s BI

Podobnosť so svetovým trhom s aplikáciami BI môžeme nájsť v niekoľkých aspektoch. Počet ponúk a dostupnosť produktov a technológií prispievajú k jeho neprehľadnosti. Ponuku vieme rozdeliť do troch oblastí:

- BI platformy a obchodné BI riešenia vytvorené na základe platforiem,
- BI nástroje integrované v transakčných systémoch,
- dielčie nástroje a technológie – databázové platformy, ETL nástroje, pokročilé analytické aplikácie a nástroje pre dolovanie dát (15, s. 444).

Český a svetový trh s BI platformami je sústredený výlučne na analytickú funkcionality (SAS Enterprise Intelligence Platform) alebo na úplné podnikové riešenie (SAP Business Suite). Charakteristický predpoklad týchto systémov je nezávislosť na operačnom systéme a databázovej platforme. Výslednú výkonnosť platforiem ovplyvňuje najmä návrh architektúry a dátovej štruktúry (15, s. 445 - 446).

Jediné zjavné spojenie medzi užívateľom a BI platformou je nástroj prístupu k dátam. Ponuka týchto nástrojov je veľmi široká, či už je zameranie na pasívne prehliadanie predom nastavených reportov alebo na funkcionality pre kreatívne vytváranie vlastných. Špičku ľadovca tvoria nástroje pre podporu riadenia výkonnosti podniku. Produkty môžeme rozdeliť do nasledovných oblastí:

- reportingové nástroje – vytváranie zostáv a čerpanie dát z databázy, skladu alebo OLTP systému,
- analytické nástroje – tvorba vlastných analýz s intuitívnym ovládaním bez nutnosti pomoci IT špecialistu,
- CPM nástroje – podpora strategického riadenia (15, s. 446 - 447).

3.2 Software as a Service

Toto riešenie bolo vytvorené ako reakcia na zložitosť, oneskorenia a vysoké náklady tradičného podnikového softvéru. Namiesto toho je implementácia a integrácia samotného riešenia predávaná ako služba od predajcu. Implementácia ale neprebíha priamo u zákazníka, ale môže byť spustená na vzdialených serveroch, ktoré sú podporované predajcami (19).

Spoločnosti Salesforce, NetSuite a RightNow sú príkladom úspešného riešenia pre riadenia vzťahov so zákazníkmi. Tieto programy vyžadujú týždne alebo mesiace na implementáciu, ktorá sa u tradičných riešení pohybuje v priebehu okolo jedného roku. Drasticky znižujú vstupné náklady a nie je nutné kupovať dodatočný hardvér. Zvyčajne sa platí na princípe predplatného (mesačný poplatok) (19).

Riešenia Software as a Service sú viac užívateľsky prívetivé ako tradičné BI programy, ktoré boli navrhnuté pre analytických expertov v rámci organizácie. Zatiaľ čo tradičný prístup umožňuje veľmi komplexnú analýzu, Software as a Service sa často vytvárajú na to, aby slúžili podnikovým používateľom pre prístup k prieskumu dát, zatiaľ čo špecialisti sa môžu sústrediť na zložitejšie otázky (19).

3.3 Analýza využitia športových analýz vo svete a v Českej republike

Spolu s vývojom technológií sa priamo úmerne zvyšuje tlak na vrcholových športovcov podrobených skúmaniu ich výkonnosti. V dnešnej dobe sme schopní pomocou kvalitnej vzorky dát a jej analýzy určiť miesta zlyhania alebo úspechu vo výkone jednotlivca alebo tímu. Hovorí sa, že vďaka správne využitiu tohto trendu, získalo Nemecko v roku 2014 titul svetového šampióna vo futbale. Ďalšia časť tejto bakalárskej práce prinesie pohľad prístup k športovým analýzám vo svete a v Česku.

3.3.1 Premier League

Britská redakcia Sky Sports priniesla v roku 2015 nový spôsob porovnania futbalových hráčov. Prestupové obdobie je vždy plné špekulácií a dohadovania sa, a preto sa rozhodli implementovať dáta z populárnej hry Football Manager do jednotlivých relácií. Prečo práve túto databázu a prečo môže byť táto databáza pre ich činnosť užitočná?

Je to najkomplexnejšia zbierka dát vzťahujúcich sa na profesionálny svet futbalu. 1 300 skautov sleduje a analyzuje viac ako 600 000 hráčov a trénerov po celom svete. Skúma jednotlivo 250 vlastností, či už technických, fyzických, mentálnych alebo charakterových. Hráči sú hodnotení na škále 1-20 pre každú vlastnosť a tieto vlastnosti sú im priradené tímom výskumníkov zodpovedných za klub alebo región.

Databáza ponúka hráčske prednosti a slabiny, rovnako aj predpoveď k potenciálnemu zlepšeniu. Dáta však môžu byť skreslené subjektivitou autora alebo naviazanosťou hráčskych vlastností na výkony v tréningu a nie v zápasovom prostredí.

Dnes majú Sky Sports, a všeobecne všetky blízke strany k Premier League, iný prístup k dátam. Využívajú plne automatizovaný systém na báze popísanej v kapitole 3.3. Marketing anglickej najvyššej súťaže nabral vysoké obrátky. Disponujú fantasy fanúšikovskou ligou, aktivitou na sociálnych sieťach, televízne práva na jeden zápas sa pohybujú na úrovni 10 miliónov libier. Časti tejto finančnej injekcie sa investujú do

skvalitnenia prístupu a ponuky k divákovi nadobudnúť informácie priamo zo štúdia, kde hostia a vedátori pracujú s interaktívnou tabuľou a vytvárajú kreatívne analýzy.

3.3.2 HET Liga

Oproti Premier League, česká najvyššia futbalová súťaž finančne zaostáva aj napriek príchodu nových investorov z Ázie do klubov Sparta Praha a Slávia Praha. České kluby získali vďaka svojej úspešnosti v európskych súťažiach miestenku navyiac, ktorá zaistí účasť ďalšiemu klubu zúčastniť sa kontinentálnych súťaží, čo zapríčiňuje príviv ďalších finančných prostriedkov do národnej súťaže.

Priestor pre zlepšenie a pokrok je mimo ihriska. Komunikácia s fanúšikmi, relácie s podrobnými analýzami, spestrenie sociálnych sietí a lepší prístup k dátam zo zápasov sú miesta, v ktorých by si mala HET Liga brať príklad od bohatšej a atraktívnejšej Premier League.

Portál hetliga.cz sprostredkúva dáta z jednotlivých zápasov, ale pre zložitejšie porovnania a prepočty sú nedostačujúce. U hráčov, ktorí nastúpili na obranných pozíciách zbiera len obranné štatistiky a u útočníkov len útočné, tento fakt zabraňuje komplexnému hodnoteniu hráčov a zisťovaniu skrytých vlastností ako je napríklad workrate alebo sklon k chybám. Priekopníkom v Česku je pražský futbalový klub Slávia, využíva podobné technológie ako kluby v zahraničí – Midtjylland, Bayern Mníchov, Hoffenheim či Arsenal Londýn. Systém slúži k podpore scoutingu a podrobnej dátovej analýze jednotlivých hráčov, hráčskych skupín a tímu ako celku.

3.3.3 SAP Sports One

Spoločnosť SAP, jeden z najväčších softvérových výrobcov na svete, vypustila na trh produkt SAP Sports One so schopnosťami podpory tímového managementu, plánovania tréningu, kontroly hráčskej kondície, skúmania výkonnosti a scoutingu.

Internetová komunikácia s reprezentantom produktu SAP mi pomohla priblížiť vlastnosti produktu a procesy spojené s implementáciou.

3.3.3.1 Popis a vlastnosti produktu

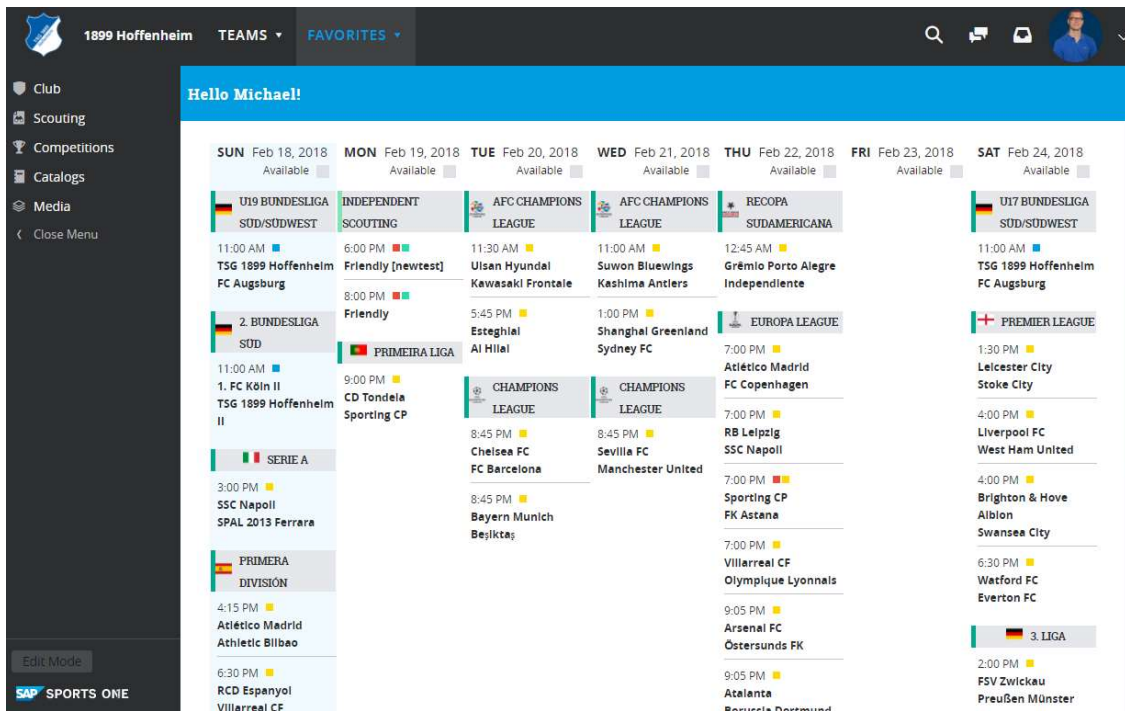
Produkt je dostupný bez ohľadu na krajinu a súťaž, do ktorej má byť zavedený. Po tom, čo zákazník kontaktuje spoločnosť so záujmom o dostupné programové riešenia, nasleduje etapa komunikácie a prieskumu implementačného prostredia. V prípravnej fáze prebiehajú školenia všetkých užívateľov. SAP odporúča zamestnanie interných klubových analytikov, aby sa detailným ovládaním aplikácie nemuseli zaoberať tréneri. Výsledná cena je individuálna, závisí od rozsahu implementácie, počtu zariadení využívajúcich aplikáciu a ďalších technických nutností k chodu tohto cloudového riešenia. Z pravidla ceny začínajú na 499 EUR na 1 rok, za 1 užívateľa, za 1 dielčiu softwarovú časť.

V dostupnej 14 dňovej skúšobnej verzii je možné nahliadnuť do prostredia aplikácie a riešenia UI. UI pôsobí moderne a intuitívne. Idea celého projektu je fantastická, ale nie vo všetkých dielčích častiach dotiahnutá do konca a uchopiteľná. Koordinácia scoutingového tímu je prehľadná, avšak vytváranie tzv. scouting reportov nie je podložené reálnymi dátami, ale stojí na subjektívnom hodnotení scouta zodpovedajúceho za svoj skrátený zoznam scoutovaných hráčov – shortlist.

Efektívne sa pracuje s videoanalýzou zo zápasov. Real-time sledovanie zápasu zbiera všetky hráčske aktivity zo zápasu, sníma pohyb, súdržnosť tímu bez lopty, správanie sa hráčov atď. Po zápase sa dáta uložia do hráčskej karty; tento systém nie je najšťastnejší pre porovnávanie a chyba celkový súhrn hráčskych výkonov z daného zápasu.

3.3.3.2 Zhrnutie

Celkovo produkt pôsobí ucelene a prepojene medzi všetkými časťami systému, ale stále je však v štádiu zlepšovania sa. Pre konkrétne analytické riešenie by bolo vhodnejšie zvoliť viac špecializovanú alternatívu a využiť len niektoré ponúkané časti systému, ako napríklad automatizované zbieranie dát zo zápasov.



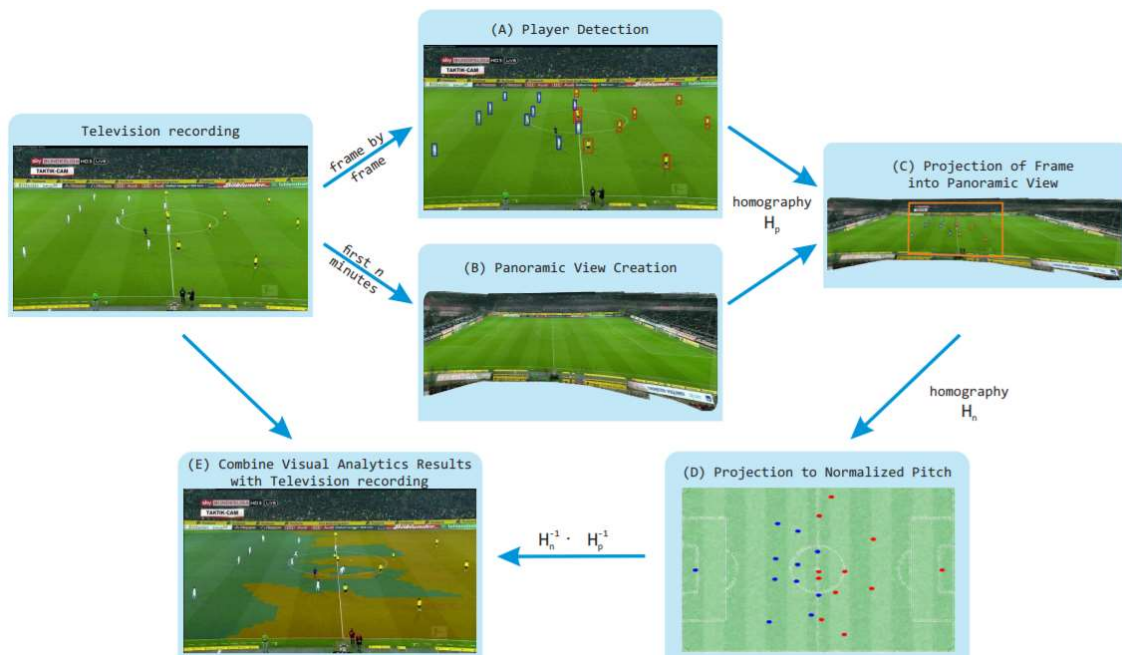
Obrázok č. 4: Ukážka UI zo skúšobnej verzie SAP Sports One

3.4 Metódy automatizovaného zbierania dát v športových stretnutiach

Predstavme si jeden súťažný víkend českej futbalovej ligy. 16 tímov, 8 zápasov po celej republike. Manuálne získavanie dát v takomto prípade je pracné a nepraktické. Zároveň je nutné byť v jeden čas na niekoľkých miestach zároveň. Z hľadiska ľudských zdrojov nám to napovedá, že čím skôr chceme dátami disponovať, musíme niekoho na ich zbieranie zamestnať. Prečo to nechať za úlohu našim počítačom? Dlhodobo nám to môže ušetriť čas, peniaze a zvýšiť kvalitu dátovej zbierky.

3.4.1 Spracovanie projekcie

Využitím televíznych kamier, niekoľkých prepočtov a algoritmov podporujúcich detekciu je možné vytvoriť procesný model, ktorý nám v reálnom čase vytvorí normalizované ihrisko.



Obrázok č. 5: Workflow systému pre vytvorenie projekcie (16)

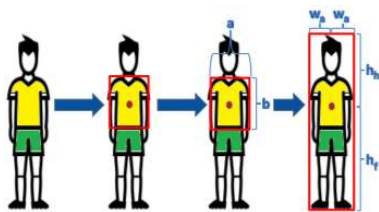
V tímových športoch je nutné vždy detekovať hráčov oboch tímov a loptu. Modelovanie hráčov je dosiahnuté kombináciou detekcie pohybu a farby.

3.4.2 Vytvorenie vizualizácie

K dokonalej projekcii cez televíznu kameru je potrebné určiť si oporné body na ihrisku, podľa ktorých sa systém orientuje. Určia sa vzdialenosti medzi bodmi a na základe niekoľkých kalkulácií vzniká výsledný obraz.

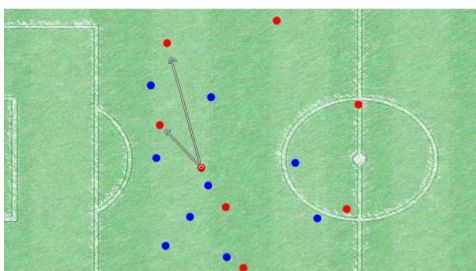


Obrázok č. 6: Prechod z panoramatického pohľadu na normalizovaný (16)



Obrázok č. 7: Ilustrácia detekcie hráča na základe farby dresu (16)

Následne sa získa obraz, ktorý môže vyzerat' takto:



Obrázok č. 8: Normalizovaná projekcia (16)

Metódy počítačového videnia s vizualizáciou pohybu podporujú analytickú činnosť a výpočty slúžiace k vylepšeniu účinnosti poskytnutej analýzy.

3.5 Zhrnutie súčasného stavu

Trh s produktami BI má hlavný rozmer v obchodných a výrobných podnikoch nie len v Českej republike. Uplatnenie analytikov vo firmách je čoraz väčšie a skrýva v sebe aj vysokú atraktivitu. Napriek širokej škále dostupných produktov je na trhu stále veľa príležitostí.

Technické možnosti, ktoré máme možnosť v súčasnosti sledovať, sú pri ich správnom využití epicentrom príležitostí rôzneho druhu. Moderný kamerový systém spojený so softvérom vytvárajúcim projekciu len demonštruje technologický pokrok a implementáciu riešení do odvetví, kde by sme to najmenej čakali. Automatizované riešenia len podtrhujú výdobytky, ktoré momentálne sledujeme.

Do športu sa technika a nové technológie dostávajú o niečo pomalšie. V motoršporte si publikum veľmi rýchlo zvyklo na podsúvané dáta. Na druhú stranu, v športoch ako sú futbal, hokej a tenis to ide ťažšie. Technika nezohráva tak veľkú úlohu ako napríklad pri závodoch F1.

Potreba sledovania dát v podnikoch poriadne vzrástla (SalesForce, Palantir a pod.). Management je nútený sledovať KPI, výkony svojich zamestnancov atď. Ak túto myšlienku pretavíme do športu, otvára sa nám nespočetné množstvo príležitostí, v ktorých hľadáme príčiny úspechu či neúspechu. A aby sme nepodľahli svojim dojmom, musíme sa opierať o zozbierané dáta zo športových udalostí.

V západnej Európe, v bohatších futbalových ligách, momentálne sledujeme pokrok myšlienok analýz v športe. Zvyšok sveta potrebuje cenovo dostupnejšie, komplexnejšie riešenia, veľmi podobné návrhu projektu v tejto bakalárskej práci.

4 VLASTNÉ NÁVRHY RIEŠENÍ

Táto kapitola čerpá z teoretických východísk spracovaných v predchádzajúcich kapitolách. Sústreďí sa na dosiahnutie cieľa práce a prináša nadobudnuté výsledky z analýz.

Prvým krokom pre vlastné návrhy riešení v tejto bakalárskej práci je vymedzenie častí projektu, ktoré táto práca rieši:

1. návrh a implementácia časti platformy – databázový systém, aplikačná resp. užívateľská vrstva – demonštrácia teórie na praktických príkladoch,
2. ukážka výsledkov reportingu za pomoci podporných aplikácií,
3. návrhy administrácie servera,
4. potenciál využitia na trhu,
5. ekonomické zhodnotenie.

Projekt je orientovaný na prácu štatistických a analytických zamestnancov športových klubov, po ktorých je v súčasnej dobe vo svete veľký dopyt.

Pre efektívne spracovanie tohto projektu je nutné dodržať stanovené kroky vo vymedzeniach. Splnenie jednotlivých krokov zaručí komplexný pohľad na stanovený cieľ a jeho dielčie časti. Smerovanie projektu je orientované na hmatateľný, finálny a využiteľný produkt vo forme reportov, ktoré znázorňujú, či už pozitívne alebo negatívne, aspekty v hre. Projekt využíva automatizované zbieranie dát produktu SAP Sports One.

4.1 Návrh a implementácia platformy

Pre vytvorenie zmysluplného databázového systému je potrebné dodržať určité postupy. To však neznamená, že neskoršie úpravy budú mať kritický vplyv na celý projekt. Na zdokumentovaných princípoch pracujú rôzne časti projektu. Návrh platformy je definovaný tak, aby spolupracoval so systémom SAP Sports One a využíval jeho jedinú

dielčiu časť, tou je automatizované zbieranie dát a import do vytvoreného databázového systému.

4.1.1 Definícia entít, atribútov, dátových typov a dátových dĺžok

Rozhodnutie o spôsobe dosiahnutia cieľa tejto bakalárskej práce vzišlo podľa vlastného úsudku a výberu, po ktorom sa dosiahol záver, že analýza športových výkonov sa najlepšie preukazuje na futbale. Samotný databázový systém nie je rozsiahly na vrstevnej úrovni entít a objektov, ale na objem dát a možnosti analýz.

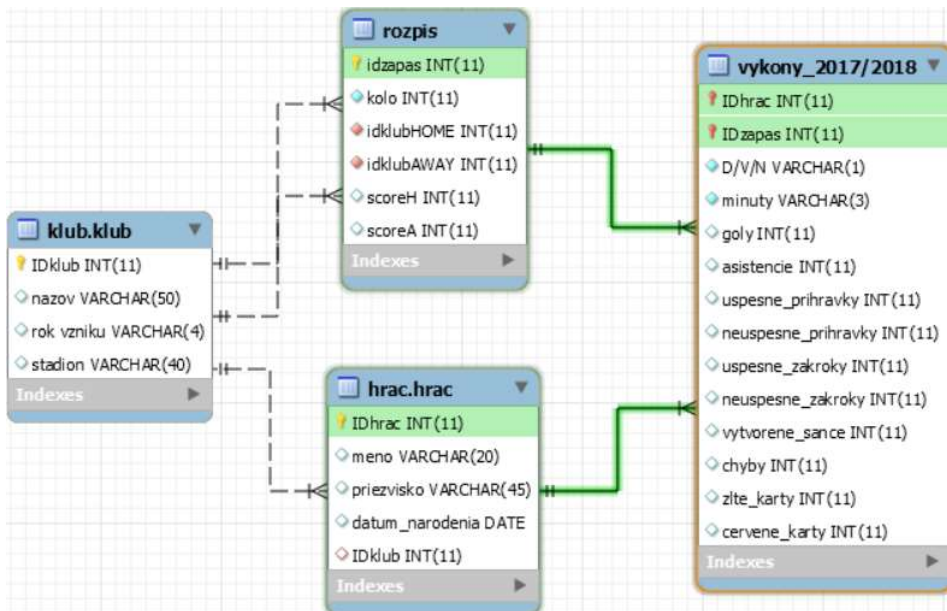
Neskorší obsah dát bol prispôsobený českej HET lige a zvolenými entitami sú: hráč, klub, rozpis, výkon. Nasledujúci obrázok popisuje atribúty jednotlivých entít, ich dátový typ a dĺžku.

klub.klub	rozpis	hrac.hrac	vykony_2017/2018
IDklub INT(11)	idzapas INT(11)	IDhrac INT(11)	IDhrac INT(11)
nazov VARCHAR(50)	kolo INT(11)	meno VARCHAR(20)	IDzapas INT(11)
rok vzniku VARCHAR(4)	idklubHOME INT(11)	priezvisko VARCHAR(45)	D/V/N VARCHAR(1)
stadion VARCHAR(40)	idklubAWAY INT(11)	datum_narodenia DATE	minuty VARCHAR(3)
Indexes	scoreH INT(11)	IDklub INT(11)	goly INT(11)
	scoreA INT(11)	Indexes	asistencie INT(11)
	Indexes		uspesne_prihravky INT(11)
			neuspesne_prihravky INT(11)
			uspesne_zakroky INT(11)
			neuspesne_zakroky INT(11)
			vytvorene_sance INT(11)
			chyby INT(11)
			zlte_karty INT(11)
			cervene_karty INT(11)
			Indexes

Obrázok č. 9: Popis entít, atribútov, dátových typov a dĺžok

4.1.2 Určenie kľúčov a zaistenie referenčnej integrity

Po zvolení entít a ich atribútov bolo nutné určiť primárne, zložené a cudzie kľúče a zaistiť referenčnú integritu. E-R diagram vyzerá nasledovne:



Obrázok č. 10: Diagram znázorňujúci vzťahy medzi reláciami

Žlté značenie pri IDklub, IDhrac a IDzapas vyjadruje primárne kľúče. Červené značenie v relácii vykony_2017/2018 pri atribútoch IDhrac a IDzapas vyjadruje zložený primárny kľúč z dvoch cudzích kľúčov.

4.1.3 Uživateľské prispôsobenie

Súčasťou uživateľského prispôsobenia je vytvorenie pohľadov, vďaka ktorým má užívateľ, z pohľadu správcu, bezpečnejší prístup k dátam. Vytvorené pohľady boli znázornené a implementované pomocou jazykov HTML a PHP. Jazyk PHP umožňuje spoluprácu s SQL databázou a prepojenie môže vyzeráť takto:

```

<?php
$db_host = 'localhost'; // Server Name
$db_user = 'root'; // Username
$db_pass = ''; // Password
$db_name = 'vut_ft_bak'; // Database Name

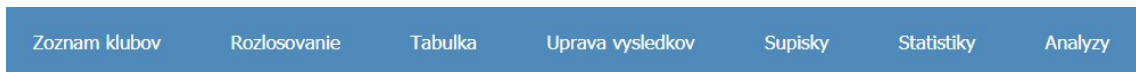
$conn = mysqli_connect($db_host, $db_user, $db_pass, $db_name);
if (!$conn) {
    die ('Failed to connect to MySQL: ' . mysqli_connect_error());
}

```

Obrázok č. 11: Časť kódu jazyka PHP k prepojeniu s databázou SQL

Vzhľadom na fakt, že úvodná platforma funguje na osobnom počítači, bolo pripojenie na vytvorenú databázu aplikované formou localhostu, kde prihlasovacie meno bolo root – bez hesla, a názov databázy bol vut_ft_bak. Tento proces zabezpečuje software s názvom XAMPP. K prípadnému ostrému chodu, je nutné zaistiť lepšie bezpečnostné podmienky, ako napríklad zmena prihlasovacieho mena a hesla.

Vytvorený webový portál má v hlavičke menu, ktoré slúži k plynulému prechodu medzi inými časťami portálu. Ukážka:



Obrázok č. 12: Menu vytvoreného portálu

Každá z kategórií uvedených v menu obsahuje súbor dát, s ktorým môže užívateľ manipulovať a prispôbovať ho podľa svojich predstáv. Avšak, nemôže údaje a záznamy vymazávať. V podstate, na pozadí, ide hlavne o príkazy select využité v rôznych kombináciách s iným praktickým využitím.

V aktuálnej časti práce budú postupne znázornené významné časti portálu s kľúčovými časťami programovacieho kódu.

HET LIGA 2017/2018

Zoznam klubov

NÁZOV	ROK VZNIKU	STADIÓN
SK Slavia Praha	1892	Eden Arena
1. FC Slovácko	1894	Stadion Miroslava Valenty
FC Zbrojovka Brno	1913	Mestsky fotbalovy stadion Srbska
Bohemians 1905	1905	Dolicek
FK Jablonec	1945	Strelnice
FC Slovan Liberec	1899	U Nisy
FK Mlada Boleslav	1902	Mestsky stadion Mlada Boleslav
SK Sigma Olomouc	1919	Andrův stadion
FC Banik Ostrava	1922	Mestsky stadion Vitkovice
FC Viktoria Plzeň	1911	Doosan Arena
AC Sparta Praha	1893	Letna
FK Teplice	1945	AGC Arena Na Stínadlech
FC Fastav Zlín	1919	Letna
FC Vysočina Jihlava	1948	Stadion v Jiráskově ulici
FK Dukla Praha	1948	Stadion Juliska
MFK Karvina	2003	Mestsky stadion Karvina

Obrázok č. 13: Zoznam klubov

```
$sql = 'SELECT *
      FROM `klub.klub`';

$query = mysqli_query($conn, $sql);
<?php
$no    = 1;
$total = 0;
while ($row = mysqli_fetch_array($query))
{
    echo '<tr>
        <td>'.$row['nazov'].'</td>
        <td>'.$row['rok vzniku'].'</td>
        <td>'.$row['stadion'].'</td>
    </tr>';
    $no++;
}
?>
```

Obrázok č. 14: Naplnenie HTML tabuľky údajmi z SQL pomocou PHP



Obrázok č. 15: Výber podmienky z dropdown menu pre select

Rozlosovanie

KOLO	DOMACI	-	-	HOSTE
3	FK Dukla Praha	1	0	FC Fastav Zlin
3	FC Viktoria Plzen	1	0	SK Sigma Olomouc
3	FC Zbrojovka Brno	0	0	Bohemians 1905
3	SK Slavia Praha	2	0	FC Vysocina Jihlava
3	MFK Karvina	1	1	FK Mlada Boleslav
3	FK Teplice	1	1	FK Jablonec
3	FC Slovan Liberec	1	1	AC Sparta Praha
3	1. FC Slovacko	5	2	FC Banik Ostrava

Obrázok č. 16: Výsledok výberu - 3. kolo

Manuálne vloženie a úprava výsledkov jednotlivých kôl funguje za pomoci jquery, ktorá pridá za riadok vytiahnutý z databázy ikony. Pod každou ikonou je definovaná funkcia. Ako bolo na začiatku kapitoly vymedzené, že záznamy nie je možné pomocou portálu vymazať, z tohto dôvodu je funkcia delete zablokovaná.

ID	Kolo	ID Domaci	Domaci	-	-	ID Hoste	Hoste	
1	1	3	FC Zbrojovka Brno	<input type="text" value="1"/>	<input type="text" value="3"/>	9	FC Banik Ostrava	  <input type="button" value="Save"/>

Obrázok č. 17: Úprava pomocou jquery prepojenou s pohľadom a individuálnymi poľami z tabuľky rozpis pre manuálnu úpravu výsledkov

Manuálne alebo automatizované vloženie aktuálnych výsledkov má vplyv na ďalšiu časť aplikačnej vrstvy, a tou je tabuľka. Automaticky sa aktualizuje vďaka príkazu select v PHP časti kódu stránky.

```

Select
  Domaci as Klub,
  count(*) as OZ,
  count(case when scoreH > scoreA then 1 end) as V,
  count(case when scoreH = scoreA then 1 end) as R,
  count(case when scoreA > scoreH then 1 end) as P,
  sum(scoreH) as VG,
  sum(scoreA) as IG,
  sum(scoreH) - sum(scoreA) as GR,
  sum(
    case when scoreH > scoreA then 3 else 0 end
    + case when scoreH = scoreA then 1 else 0 end
  ) as Body
from (
  select Domaci, scoreH, scoreA from Zapasy
  union all
  select Hoste, scoreA, scoreH from Zapasy
) a
where scoreH is not NULL or scoreA is not NULL
group by Domaci
order by Body desc, GR desc

```

Obrázok č. 18: Príkaz select k vytvoreniu aktuálnej tabuľky

Tabuľka výkony bola naplnená vlastnými dátami z dôvodu neúplnosti dát v iných zdrojoch na internete. Tým sa zaručilo sledovanie konkrétnych atribútov a vo výsledkoch nechýbali údaje, resp. boli vytvorené správne kalkulácie v nasledujúcich analýzach. Sledovanie prebehlo jednoduchým spôsobom so zoznamom hráčov a atribútov a značením na hárok.

4.2 Štatistiky a reporting

K vytvoreniu kvalitného reportu z futbalového prostredia nám nestačí iba vedieť hodnotu v kolónke góly pri mene sledovaného. Dôležitým faktorom je vytvoriť si doplnkový atribút s názvom rating, ktorý bude poskytovať komplexné hodnotenie na základe všetkých atribútov, či už negatívnych alebo pozitívnych. Preto, pre účely tohto projektu, bola vytvorená škála vplyvu jednotlivých vlastností na celkový rating hráča.

Tabuľka č. 2: Vplyv výskytu atribútu na celkový rating

<i>atribút</i>	<i>vplyv na rating</i>
<i>gól</i>	+0,7
<i>asistencia</i>	+0,7
<i>úspešná prihrávka</i>	+0,055
<i>neúspešná prihrávka</i>	-0,09
<i>úspešný zákrok</i>	+0,07
<i>neúspešný zákrok</i>	-0,09
<i>vytvorená šanca</i>	+0,4
<i>chyba</i>	-0,45

Vybrané atribúty sú ohodnotené na princípe, ktorý využíva spoločnosť Sports Interactive vo svojich vysoko realistických herných simuláciách Football Manager. Z dôvodov autorských práv som pridal drobný koeficient a zlúčil dielčie atribúty do jednej. Výška vplyvu na rating je založená na výskyte v zápase a prenesenia jednotky atribútu na celkový výsledok v zápase.

Rating začína na hodnote 5.

Sledované obdobie bolo úvodných 5 zápasov HET ligy u družstiev FC Zbrojovka Brno a FC Viktoria Plzeň. Po 5. kole nastala dvojtýždňová reprezentačná prestávka.

Hracske statistiky

Meno	Priezvisko	Klub	Pocet Zapasov	Minuty	Goly	Asistencie	Body	Uspesne prihravky	Uspesnost prihraviek	Uspesne zakroky	Uspesnost zakrokov	Klucove prihravky	Chyby	Zlte karty	Cervene karty	Rating ↓
Patrik	Hrosovsky	FC Viktoria Plzen	4	350	2	1	3	174	89.69 %	7	38.88 %	9	0	0	0	8.24
Tomas	Hajek	FC Viktoria Plzen	5	419	1	0	1	285	93.44 %	25	65.78 %	2	0	2	0	8.19
Lukas	Hejda	FC Viktoria Plzen	5	380	1	0	1	255	93.75 %	24	60.00 %	5	2	0	0	7.90
David	Limbersky	FC Viktoria Plzen	3	212	1	0	1	153	85.47 %	12	66.66 %	3	0	0	0	7.75
Ian	Konic	FC Viktoria	5	411	0	2	2	192	82.05 %	20	40.81 %	14	2	0	0	7.33

Obrázok č. 19: Tabuľkové zobrazenie štatistík hráčov po 5 kole zoradených podľa ratingu

Časť portálu sústreďujúca sa na hráčske štatistiky obsahuje funkcie, ktoré zoradujú dáta podľa zvoleného atribútu.

4.2.1 Vypracované reporty

Všetky reporty sú interaktívne a vypracované zo zozbieraných dát. Na ich vytvorenie boli využité podporné aplikácie ako Power BI či Microsoft Azure. Úlohou analýz a reportov je uľahčiť budúce rozhodovanie a nájsť oporné body, indikátory, pre pozitívny posun.

4.2.2 Komplexný report zo sledovaného obdobia

Príloha číslo 2, resp. komplexný report zobrazuje a porovnáva zozbierané dáta v sledovanom období pre definované kluby. Relevantnosť výsledkov môže potvrdiť fakt, že po 5 zápasoch sa FC Viktória Plzeň nachádzala na 1. mieste tabuľky HET ligy, zatiaľ čo FC Zbrojovka Brno okupovala miesta predurčujúce súboj o záchranu v najvyššej súťaži.

Pri porovnaní priemerného ratingu oboch družstiev, môžeme celosezónne predpokladať, že rating nad 7 je veľmi pozitívny a môže sa predpokladať súboj o najvyššie priečky, a rating pod 6 napovedá, že kluby sa v tabuľke budú nachádzať v okolí spodnej štvrtiny. Obrovským rozdielom medzi oboma tímami bol v priemernom počte úspešných prihrávok (až 46 na hráča) a v počte individuálnych chýb (až 89 celkovo).

4.2.3 Individuálne porovnanie hráčov

Slovenský stredopoliar Patrik Hrošovský podával v úvode sezóny výborné výkony ohodnotené ratingom 8,24, čo mu zaistilo miesto v nominácii Slovenska na kvalifikačné zápasy proti Anglicku a Slovinsku. Rovnako sa objavili špekulácie o jeho možnom prestupe do talianskeho klubu Atalanta Bergamo.

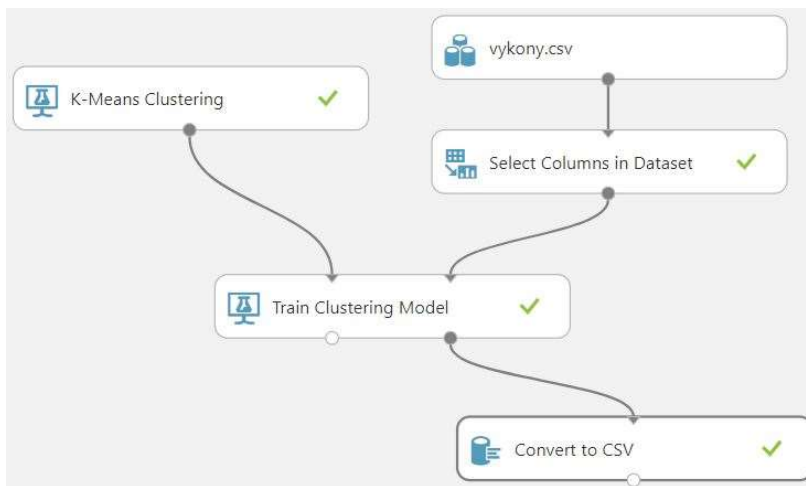
Príloha číslo 3, individuálne sledovanie hráčov, porovnáva hráčske výkony v zápasoch nie na úrovni ratingu, ale na úrovni hodnôt jednotlivých atribútov majúcich vplyv na rating. Tento report pomáha konfrontovať hráčske vlastnosti v rôznych aspektoch hry a nesústreďí sa na celkový výkon hráča podľa ratingu.

4.2.4 Znázornenie vplyvu chyby na rating hráča

V ďalšom reporte vyplývajúcom z analýzy, príloha 4, bol použitý k-means clustering. Aplikoval sa na vzorku dát v tabuľke výkony, kde sa vybrali atribúty, ktoré majú negatívny vplyv na rating. Výsledkom bolo vytvorenie dvoch zhlukov zachytávajúcich nízky alebo vysoký pomer chybovosti hráča v zápase.

Do vzorku boli pridané dva atribúty, prvý určoval vzdialenosť od jedného zhluku a druhý od ďalšieho. K-means clustering sa vyobrazuje na súradnicovej sieti, a teda, po skalárnom súčine týchto dvoch doplnených atribútov a ich následnom spriemerovaní, dostaneme produkt chyby v nami sledovanom intervale.

V reporte z prílohy 4 je vypočítaný percentuálny rating a percento chybovosti z celkového vzorku dát a zobrazení sú len hráči s odohratými 180 minútami a viac. Tento report poukazuje hlavne na to, akým spôsobom hráč dosiahol svoj prípadný vysoký alebo nízky rating. Napríklad: Krmenčík má vysoký produkt chyby, na rozdiel od iných hráčov s vysokým produktom chyby má aj nadpriemerný rating, takže môžeme predpokladať, že kvalitu svojich výkonov zvyšuje atribútmi s vysokým vplyvom na rating – góly a asistencie. Hájek je príkladom dosiahnutia vysokého ratingu bez množstva gólov a asistencií.



Obrázok č. 20: Diagram procesu k-means clustering v aplikácii Microsoft Azure

K-means clustering v kombinácii s ďalšími funkciami reprezentuje riešenie dolovania dát ako súčasti analýzy, ktoré nie je možné vidieť na prvý pohľad a preto posúva výsledný report na vyššiu úroveň.

4.2.5 Analýza spoľahlivosti hráčov v osobných súbojoch

Posledný report, príloha číslo 5, je zameraný na úspešnosť jednotlivca v osobných súbojoch. Ako úspešný osobný súboj sa evidovalo získanie lopty v obrannej pozícii (bez lopty), alebo útočný víťazný súboj (vedenie lopty 1v1). Je zrejmé, že najviac osobných súbojov absolvujú hráči nastupujúci v obrannej línii. Do výsledkov boli zahrnutí hráči, ktorí sa zúčastnili minimálne desiatich vyššie spomínaných situácií. Cieľom bolo zistiť mieru spoľahlivosti hráčov v kontaktných situáciách. Výsledky je možné interpretovať nasledovne.



Obrázok č. 21: Interpretácia výsledkov z reportu v prílohe 5

Osa Y znázorňuje úspešnosť v súbojoch v percentách. Osa X znázorňuje ako často boli hráči súčasťou týchto súbojov. Z toho vyplýva, že sieť môžeme rozdeliť na 4 štvoruholníky:

- ľavý horný – nízky výskyt, vysoká úspešnosť,
- ľavý dolný – nízky výskyt, nízka úspešnosť,
- pravý dolný – vysoký výskyt, nízka úspešnosť,
- pravý horný – vysoký výskyt, vysoká úspešnosť.

Červenou farbou vyznačený štvoruholník značí stav, ktorý by prezentoval hráča s nadpriemernými schopnosťami v osobných súbojoch. Čím bližšie sa bod nachádza červenému indikátoru, tým lepšie výkony daný hráč predvádzal.

4.3 Administrácia databázového servera a portálu

Táto kapitola rieši systémovú bezpečnosť a pravidelnú automatizovanú správu databázy, ktorá ušetrí prácu a čas zamestnancovi.

Inštaláciu každej inštancie databázového systému s prístupom k replikáciám u klienta, ktorý si zakúpi licenciu, vykonáva človek poverený firmou dodávateľa, teda nami. Fyzický prístup do databázy má len vlastník a zvolený administrátor – nemá oprávnenie mazať dáta ani iné objekty databázy. Ostatní odoberatelia majú možnosť len sledovať poskytnuté dáta.

Zálohovanie a aktualizácia databázovej replikácie, na ktorej odoberanie sa môžu pripojiť externí klienti využívajúci tento produkt, sú nastavené na každý utorok o 22:00, to znamená po každom súťažnom víkende vrátane letnej prestávky. Celý tento proces prebieha bez prerušenia chodu databázového servera.

Celý plán údržby a automatických činností sa zobrazuje v sekcii Maintenance Plan alebo SQL Server Agent v podkategórii Jobs. Activity monitor, sleduje výkon a zaťaženie databázového servera a na základe distribuovaných údajov administrátorovi, je možné zablokovať prístup užívateľom spôsobujúcim nadmerné zaťaženie systému.

Prihlasovanie do portálu prebehne po zadaní prihlasovacích údajov, login a heslo. Následne užívateľ musí napísať tzv. one-time password, ktoré sa generuje každú minútu v podpornej aplikácii na prihlasovanie.

Portál je prispôsobený externým užívateľom len na čítanie dát. Administrátor má nastavené oprávnenie manipulácie s dátami aj cez portál. Dôvodom je jednoduchší prístup ku konkrétnym údajom cez pohľady.



Obrázok č. 22: Podporná aplikácia k one-time password

4.4 Potenciál na trhu

Futbal je biznis, IT je biznis, skĺbenie týchto dvoch odvetví vytvorí na trhu priestor obrovských rozmerov najmä v česko-slovenských podmienkach, resp. na celom svete okrem pokročilých líg ako sú v Anglicku, Francúzsku, Nemecku, Španielsku a Taliansku. Česká najvyššia súťaž aktuálne prežíva obrovský príliv peňazí od ázijských investorov, zároveň má výborné meno v rámci európskych klubových súťaží, kde má 1 miestenku do skupinovej fázy Ligy majstrov, 1 miestenku do predkola rovnakej súťaže, ďalšie 2 kluby majú možnosť hrať Európsku ligu a víťaz národného poháru postupuje priamo do skupinovej fázy Európskej ligy. Tieto fakty zaručujú ďalší príliv peňazí na český futbalový trh.

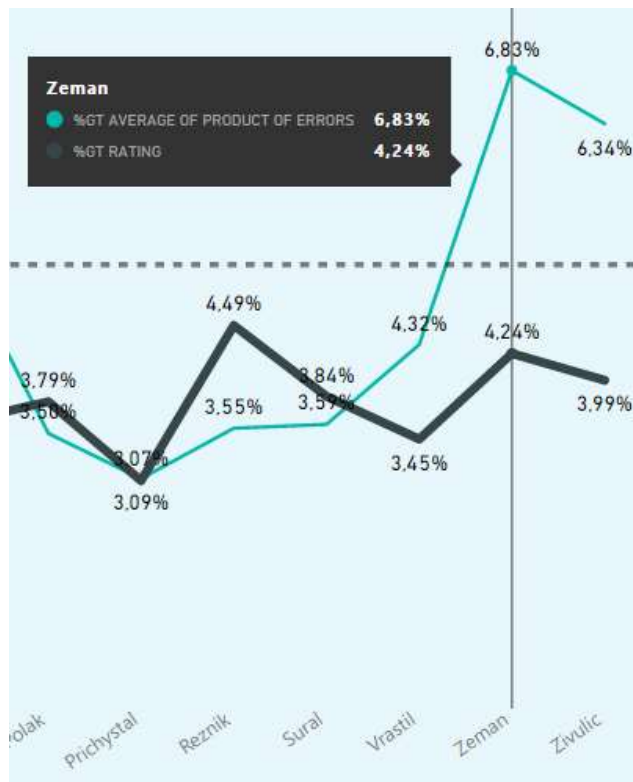
4.4.1 Vízia

Produkt navrhnutý a z časti implementovaný v tejto bakalárskej práci má byť skĺbením automatizovaného systému zbierania dát od SAP Sports One a analyticko-reportovacieho systému vytvoreného v návrhoch riešení tejto práce. Jeho víziou, je slúžiť pri podpore rozhodovania, či už o zostave alebo ohľadom investícií do hráčov, nie len klubom, ale aj trénerom, agentom, scoutom, športovým riaditeľom a fanúšikom.

Modelové situácie:

- pred koncom prestupového obdobia sa klubu dlhodobo zraní hráč. Klub dovtedy nesledoval žiadneho hráča na tento post. Preto siahne po tejto aplikácii a vďaka štatistikám, analýzám a reportom získa prehľad o potencionálnych náhradách veľmi rýchlo. Zároveň nebolo nutné vynaložiť náklady na vycestovania scouta, lístok na zápas a nefinančný náklad – subjektívny názor scouta,
- agent sleduje zápasy v mládežníckych kategóriach, nie vždy je však schopný stihnúť všetky zápasy a preto pomocou tejto aplikácie nájde vhodného hráča, ktorého bude zastupovať a bude ho považovať za svoju dlhodobú investíciu,

- nie všetky kluby disponujú možnosťou sledovať hráčov ďaleko za oceánom, títo hráči však môžu byť talentovanejší a lacnejší ako súčasný kmeňoví v klube. Aby zamestnanec kluby získal prehľad o hráčoch, ktorí môžu byť prestupovým cieľom a neskôr v budúcnosti odídu za podstatne väčšie peniaze, nemusí cestovať fyzicky napr. do Brazílie, ale vďaka juhoamerickému balíčku má prehľad z pohodlia domova,
- pokiaľ majú hráči prístup ku svojim individuálnym dátam, môžu monitorovať svoje výkony, definovať oblasti v ktorých sa môžu zlepšiť a rovnako podkladať údaje pri vyjednávaniach o zmluvách,
- ak športový klub disponuje nízkym rozpočtom, musí sa spoliehať na efektívne správanie sa na prestupovom trhu. Priblížiť sa bližšie k efektivite typu „za málo peňazí veľa muziky,“ je nutné siahnuť po hráčoch, ktorí sú považovaní, že sú svojim vekom za zenitom, že sa po zranení nevrátia do formy alebo dlho nehrali vysokú súťaž. V prípade týchto hráčov sa porovná ich výkonnosť so zvyškom, a na základe dát zistíme, či ich nízka na prestupovom trhu je podhodnotená.
- hlbší pohľad na dáta odhalí nedostatky v individuálnych výkonoch. Rating nemusí mať maximálnu výpovednú hodnotu a preto je nutné sa pozrieť na jednotlivé hodnoty v atribútoch zvlášť. Nasledujúci obrázok poukazuje na pomer výskytu nedostatkov v podobne negatívnych atribútoch na rating:



Obrázok č. 23: Pomer ratingu a produktu chyby z celkového počtu

Opieranie sa o dáta v rozhodujúcich chvíľach môže klubu, hráčovi alebo agentovi vyniesť nemalé finančné prostriedky navyše. Kreativite využitia dát sa medze nekladú.

4.5 Ekonomické zhodnotenie

V ekonomickom zhodnotení musíme uvažovať dvomi smermi. Jeden pohľad tvoria náklady na vytvorenie funkčného prostredia s dôveryhodným technickým a programových požadím, a druhý pohľad vytvára výsledná cena produktu, ktorého súčasťou je užívateľská implementácia a udelenie prístupov pre koncového zákazníka.

Tabuľky zaznamenávajú náklady na technický implementáciu a cenové ponuky pre potenciálnych zákazníkov.

Tabuľka č. 3: Implementačné náklady na technickú realizáciu

Popis produktu	počet	jednotková cena	celková cena
Server 1 : SR590 Xeon Gold 5118 (12C 2.3GHz 16.5MB Cache/105W) 16GB (1x16GB, 2Rx8 RDIMM), O/B, 930-8i, 1x750W, XCC Standard, Tooless Rails	1	€ 3 768,00	€ 3 768,00
Foundation Service - 4Yr Next Business Day Response	1	€ 921,00	€ 921,00
ThinkSystem M.2 with Mirroring Enablement Kit	1	€ 163,00	€ 163,00
1.5m, 10A/100-250V, C13 to IEC 320-C14 Rack Power Cable	2	€ 13,00	€ 26,00
ThinkSystem 32GB TruDDR4 2666 MHz (2Rx4 1.2V) RDIMM	6	€ 838,00	€ 5 028,00
EMEA Hardware Basic Installation Server Standard Business Hours - IP1	1	€ 349,00	€ 349,00
ThinkSystem 430-8e SAS/SATA HBA	2	€ 275,00	€ 550,00
ThinkSystem 750W(230/115V) Platinum Hot-Swap Power Supply	1	€ 188,00	€ 188,00
ThinkSystem XClarity Controller Standard to Advanced Upgrade	1	€ 27,00	€ 27,00
ThinkSystem XClarity Controller Advanced to Enterprise Upgrade	1	€ 18,00	€ 18,00

ThinkSystem M.2 CV3 128GB SATA 6Gbps Non- Hot Swap SSD	2	€	113,00	€	226,00
ThinkSystem 1Gb 2-port RJ45 LOM	1	€	138,00	€	138,00
ThinkSystem Intel I350-T4 PCIe 1Gb 4-Port RJ45 Ethernet Adapter	1	€	338,00	€	338,00
Windows Svr 2016 Standard ROK (16 core) - MultiLang	1	€	639,00	€	639,00
					€ 12 379,00
Storage : Lenovo					
ThinkSystem DS2200 SFF SAS Dual Controller Unit	1	€	5 625,00	€	5 625,00
External MiniSAS HD 8644/MiniSAS HD 8644 2M Cable	4	€	63,00	€	252,00
Foundation Service - 4Yr Next Business Day Response	1	€	2 108,00	€	2 108,00
EMEA Hardware Basic Installation Storage Standard Business Hours - IP1	1	€	349,00	€	349,00
Lenovo Storage 900GB 10K 2.5" SAS HDD	14	€	388,00	€	5 432,00
					€ 13 766,00
Inštaláčn é a implementáčn é práce	1	€	6 090,00	€	6 090,00
					€ 6 090,00
Microsoft SQL Server Enterprise	1	€	11 600,00	€	11 600,00
					€ 11 600,00
SAP Sports One Data Collection	1	€	15 000,00	€	15 000,00
					€ 15 000,00
					€ 58 835,00

Celkové implementačné a inštaláčn é náklady budú minimálne na hranici 58 835,00 €.

Ceny produktov Lenovo vrátane inštaláčn ých prác vznikli po konzultácii so zamestnancom tejto korporácie (Solution Sales Representative). Cena za Microsoft SQL Server Enterprise bola nadobudnutá z oficiálnej stránky Microsoft. Náklady za SAP

Sports One Data Collection sú orientačné po konzultácii so zamestnancom (Product Engineer).

Tabuľka č. 4: Produkty a služby dostupné pre zákazníkov

Popis produktu	počet/jednotka	obdobie	celková cena
prístup do systému	1/účet	1 rok	€ 1 500,00
odoberanie snímkovej replikácie	1/do 1 db	1 rok	€ 4 000,00
regionálny balíček dát	1/región	1 rok	€ 2 000,00
služby reportingu	1/liga	0,25 roka	€ 750,00
konzultácia a vzdelávanie	1 osoba/termín	7 dní	€ 350,00
doplňkové zmluvné podmienky			
% z predaja/obchodu hráča	1/hráč	-	2,5 %
zaobstaranie analytického SW	1/licencia	-	+ 20 % marža
% zo súťažných ziskov	1 súťaž/odmeny	1 rok	4,5 %
extra služby			
kamerový systém a zbieranie dát	1 kamera/obdobie	1 rok	€ 1 000,00
sledovanie mládeže	1 klub/akadémia	1 rok	€ 3 000,00

Produkty podliehajú základnému rozdeleniu na kontinentálne a ďalej na lokálne regióny. Základnú logiku tvorí členenie podľa systému vo svetových korporáciách – EMEA, North America, Asia Pacific a Latin America.

Zákazníci, kluby alebo agenti, ktorí si zakúpia tento balíček produktov získajú väčší prehľad vo výkonoch svojich hráčov. Obohatia svoje vedomosti a názory, ktoré sa nebudú opierať len o dojem, ale reálne dáta. Hráči budú schopní rovnako monitorovať svoju hru a zamerať sa na oblasti, v ktorých sa môžu zdokonaľiť a posunúť sa na vyššiu úroveň. Produkt môže slúžiť ako nástroj kontroly a podpory rozhodovania nie len na ihrisku, ale aj na trhu s hráčmi a zamestnancami.

Modelácia príjmov zo základnej objednávky:

Tabuľka č. 5: Základná objednávka

objednávka	počet	obdobie	celková cena
prístup do systému	1	1 rok	€ 1 500,00
regionálny balíček dát	4	1 rok	€ 8 000,00
služby reportingu	2	1 rok	€ 6 000,00
konzultácia a vzdelávanie	2	7 dní	€ 700,00
			€ 16 200,00

Kalkulácia základnej objednávky ráta len so základným balíčkom produktu. Finálna cena spolu s doplnkovými zmluvnými podmienkami môže podľahnúť vyjednávaniu. Tabuľka vyjadruje záujem len jedného klienta. Len východná Európa, ktorá nedisponuje veľkým pokrytím obdobnými produktami, poskytuje obrovskú príležitosť uživenia sa týmto produktom. V prepočte je nutné uzavrieť zmluvu so 4 klientami, vo výške ilustračnej objednávky za každú zmluvu, aby sme pokryli vstupné náklady na realizáciu.

Pri súčasnom potenciáli na trhu je to veľmi reálna výzva a je možné ju niekoľkonásobne prekonať.

ZÁVER

Cieľom bakalárskej práce bolo spracovať návrh a implementáciu databázovej aplikácie pre analýzu športových výkonov. Správne využitie podobných systémov vo východnej Európe môže strategicky, výkonnostne ale aj finančne povzbudiť fungovanie športových klubov v tomto regióne. Hlavnou víziou navrhnutého produktu je vybaviť kluby kvalitnými informáciami, po ktorých analýze kluby nadobudnú poznatky a zároveň nájdu priestor a možnosti pre zlepšenie.

Práca je rozdelená do troch častí. Prvá, teoretická časť, obsahuje teoretické východiská pre technické pochopenie vybranej problematiky. Rieši prístupy databázové, dátové a informačné, programovacích jazykov a aplikácií a dátového modelovania. Správnym ovládaním týchto prístupov, na základe teoretických východísk, bolo dokázané vytvorenie platformy, teda jedného z cieľov tejto bakalárskej práce.

Druhá časť práce, analýza súčasného stavu, je orientovaná na porovnanie, najprv, dostupných podnikových business intelligence softvérových produktov v Českej republike, a následne plynulo prechádza na produkty a pokrytie trhu športovými produktami slúžiacich k analýze výkonov. Zhŕňa využívanú metódu automatizovaného zbierania dát pomocou zápasovej televíznej kamery. Dôležitosť tejto kapitoly zohráva fakt, že vo finálnom návrhu produktu je použitých niekoľko dielčích častí zo súčasných programových alternatív.

V záverečnej časti je zdokumentovaný praktický výstup projektu a návrh ďalších pokračovaní spolu s ekonomickým ohodnotením. Podstatným výstupom zo zozbieraných dát, ich implementácie do systému a následné využitie, sú vypracované reporty, ktoré skúmajú dáta viac do hĺbky a prinášajú nové poznatky pre riadiace jednotky klubov. Prakticky implementovaná aplikačná platforma, respektíve portál, demonštruje rôzne funkcie, ktoré finálny produkt má spĺňať.

Verím, že implementácia obdobných systémov do reálneho prostredia môže mať výrazný pozitívny vplyv na prácu progresívnych ľudí v športe.

ZOZNAM POUŽITÝCH ZDROJOV

- (1) BARTEK, M. *Štyri hlavné dôvody prečo sa naučiť jazyk sql*. In: *Ivit* [online]. 2016 [cit. 2017-03-06].
- (2) BEGG, C., HOLOWCZAK, R. a T. CONOLLY. *Mistrovství - Databáze : Profesionální průvodce tvorbou efektivních databází*. Praha: Computer Press, 2009. 584 s. ISBN 978-80-251-2328-7.
- (3) BERRINGTON, J. *Databases. Anaesthesia & Intensive Care Medicine* [online]. Elsevier, 1703, 18(3), 155-157 [cit. 2017-03-06]. DOI: 10.1016/j.mpaic.2016.11.016. ISSN 1472-0299.
- (4) FOX, R. *Moving from data to information. OCLC Systems & Services: International Digital Library Perspectives* [online]. Emerald Group Publishing Limited, 2004, 20(3), 96-101 [cit. 2017-03-06]. DOI: 10.1108/10650750410551415. ISSN 1065-075X.
- (5) POKORNÝ, J. a M. VALENTA. *Databázové systémy*. Praha: České vysoké učení technické v Praze, 2013. 265 s. ISBN 978-80-01-05212-9.
- (6) CUESTA, H. a J. HUF. *Analýza dat v praxi*. Brno: Computer Press, 2015. 296 s. ISBN 978-80-251-4361-2.
- (7) MONTGOMERY, D. *Collecting Data. Quality and Reliability Engineering International* [online]. 2016, 32(2), 333-333 [cit. 2018-01-18]. DOI: 10.1002/qre.1974. ISSN 0748-8017.
- (8) KROENKE, D., AUER D. a J. GONER. *Databáze*. Brno: Computer Press, 2015. 496 s. ISBN 978-80-251-4352-0.
- (9) LABERGE, R. a J. GONER. *Datové sklady: agilní metody a business intelligence*. Brno: Computer Press, 2012. 350 s. ISBN 978-80-251-3729-1.
- (10) KOCH, M. a B. NEUWIRTH. *Datové a funkční modelování. Vyd. 4., rozšířené*. Brno: Akademické nakladatelství CERM, 2010. 142 s. ISBN 978-80-214-4125-5.
- (11) FERRARI, A. a M. RUSSO. *Introducing Microsoft Power BI*. Redmond, Washington: Microsoft Corporation, 2016. 407 s. ISBN 978-1-5093-0228-4.
- (12) I-S, STROE. *MySQL databases as part of the Online Business, using a platform based on Linux*. *Database Systems Journal* [online]. 2011, (3), 3-12 [cit. 2018-02-05]. ISSN 2069-3230.

- (13) CORLATAN, C., LAZAR, M., LUCA, V. a O. PETRICICA. *Query Optimization Techniques in Microsoft SQL Server*. Database Systems Journal [online]. 2014, (2), 33-48 [cit. 2018-02-05]. ISSN 2069-3230.
- (14) SARMA, T., VISWANATH, P. a B. REDDY. *Single pass kernel k-means clustering method*. Sadhana [online]. 2013, 38(3), 407-419 [cit. 2018-02-05]. DOI: 10.1007/s12046-013-0143-3. ISSN 0256-2499.
- (15) SODOMKA, P. a H. KLČOVÁ. *Informační systémy v podnikové praxi. 2. aktualiz. a rozš. vyd.* Brno: Computer Press, 2010. 501 s. ISBN 978-80-251-2878-7.
- (16) STEIN, M., JANETZKO, H. a A. LAMPRECHT. *Bring It to the Pitch: Combining Video and Movement Data to Enhance Team Sport Analysis*. Visualization and Computer Graphics, IEEE Transactions on [online]. USA: IEEE, 2018, 24(1), 13-22 [cit. 2018-02-18]. DOI: 10.1109/TVCG.2017.2745181. ISSN 1077-2626.
- (17) TULLOCH, M. *Introducing Windows Azure*. Redmond, Washington: Microsoft Corporation, 2013. 142 s. ISBN: 978-0-7356-8288-7.
- (18) LACKO, E. *Mistrovství v SQL Server 2012*. Brno: Computer Press, 2013. 639 s. ISBN 978-80-251-3773-4.
- (19) CRISTESCU, M. *Traditional Enterprise Business Intelligence Software Compared to Software as a Service Business Intelligence*. Informatică economică [online]. Inforec Association, 2016, 20(1), 39-47 [cit. 2018-04-24]. DOI: 10.12948/issn14531305/20.1.2016.04. ISSN 1453-1305.

ZOZNAM POUŽITÝCH SKRATIEK A SYMBOLOV

OLTP	On-line Transaction Processing
SQL	Structured Query Language
BI	Business Intelligence
ETL	Extract, Transform, Load
CPM	Corporate Performance Management
UI	User Interface
E-R diagram	Entity-Relation diagram
KPI	Key Performance Indicator

ZOZNAM OBRÁZKOV

Obrázok č. 1: Databázový prístup užívateľov do databázy pomocou databázovej aplikácie a systému riadenia bázy dát	17
Obrázok č. 2: Prevedenie dát na znalosti	19
Obrázok č. 3: Dátova štruktúra a obsah vety	24
Obrázok č. 4: Ukážka UI zo skúšobnej verzie SAP Sports One.....	36
Obrázok č. 5: Workflow systému pre vytvorenie projekcie	37
Obrázok č. 6: Prechod z panoramatického pohľadu na normalizovaný	38
Obrázok č. 7: Ilustrácia detekcie hráča na základe farby dresu.....	38
Obrázok č. 8: Normalizovaná projekcia	38
Obrázok č. 9: Popis entít, atribútov, dátových typov a dĺžok.....	41
Obrázok č. 10: Diagram znázorňujúci vzťahy medzi reláciami	42
Obrázok č. 11: Časť kódu jazyka PHP k prepojeniu s databázou SQL.....	43
Obrázok č. 12: Menu vytvoreného portálu	43
Obrázok č. 13: Zoznam klubov.....	44
Obrázok č. 14: Naplnenie HTML tabuľky údajmi z SQL pomocou PHP	44
Obrázok č. 15: Výber podmienky z dropdown menu pre select.....	44
Obrázok č. 16: Výsledok výberu - 3. kolo	45
Obrázok č. 17: Úprava pomocou jquery prepojenou s pohľadom a individuálnymi poľami z tabuľky rozpis pre manuálnu úpravu výsledkov	45
Obrázok č. 18: Príkaz select k vytvoreniu aktuálnej tabuľky.....	46
Obrázok č. 19: Tabuľkové zobrazenie štatistík hráčov po 5 kole zoradených podľa ratingu	48
Obrázok č. 20: Diagram procesu k-means clustering v aplikácii Microsoft Azure.....	50
Obrázok č. 21: Interpretácia výsledkov z reportu v prílohe 5	51
Obrázok č. 22: Podporná aplikácia k one-time password.....	52
Obrázok č. 23: Pomer ratingu a produktu chyby z celkového počtu	55

ZOZNAM TABULIEK

Tabuľka č. 1: Vybrané príkazy jazyka SQL	28
Tabuľka č. 2: Vplyv výskytu atribútu na celkový rating	47
Tabuľka č. 3: Implementačné náklady na technickú realizáciu.....	56
Tabuľka č. 4: Produkty a služby dostupné pre zákazníkov	58
Tabuľka č. 5: Základná objednávka.....	59

ZOZNAM PRÍLOH

Príloha 1: Trojvrstvová architektúra ANSI - SPARC

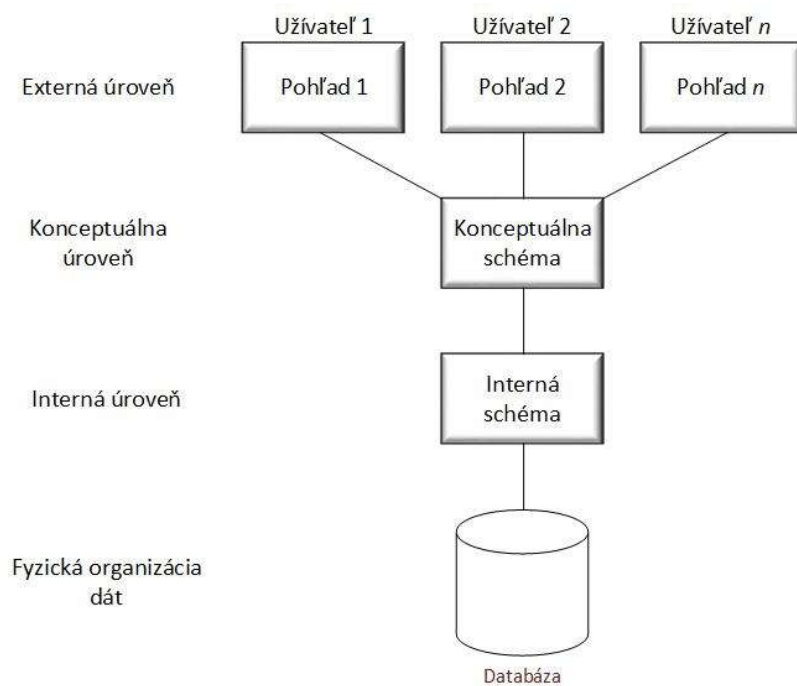
Príloha 2: Komplexný report zo sledovaného obdobia

Príloha 3: Individuálne porovnanie hráčov

Príloha 4: Znázornenie vplyvu chyby na rating hráča

Príloha 5: Analýza spoľahlivosti hráčov v osobných súbojoch

Príloha 1: Trojvrstvová architektúra ANSI – SPARC (Vlastné spracovanie podľa 2, s. 57)



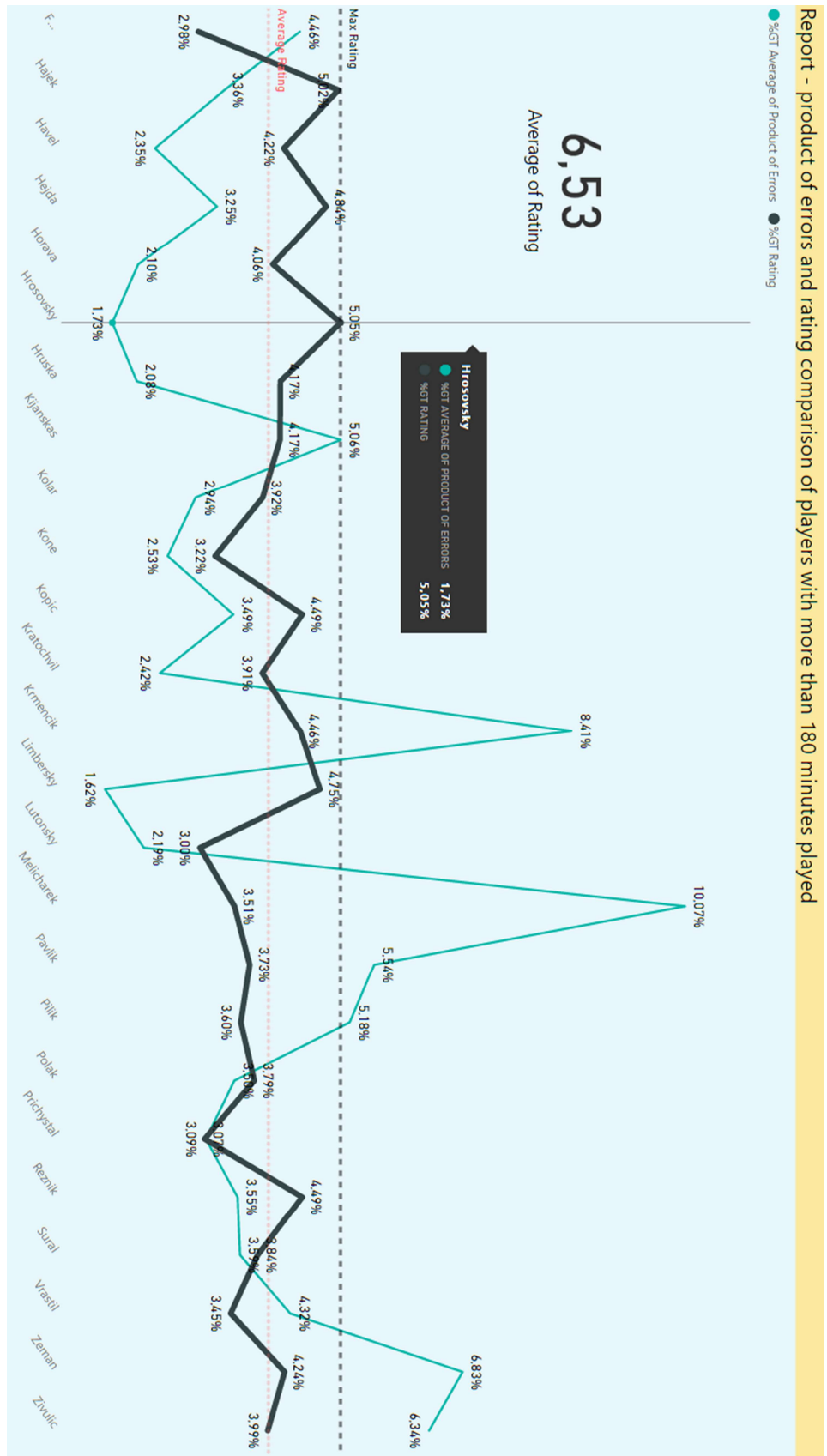
Príloha 2: Komplexný report zo sledovaného obdobia



Príloha 3: Individuálne porovnanie hráčov



Príloha 4: Znáznornenie vplyvu chyby na rating hráča



Príloha 5: Analýza spoľahlivosti hráčov v osobných súbojoch

