

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

DIMENZIONÁLNY MODEL AKADEMICKÉHO
REPORTOVACIEHO SYSTÉMU
BAKALÁRSKA PRÁCA

2016
MARTIN PRIBULA

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

DIMENZIONÁLNY MODEL AKADEMICKÉHO
REPORTOVACIEHO SYSTÉMU

BAKALÁRSKA PRÁCA

Študijný program: Informatika
Študijný odbor: 2508 Informatika
Školiace pracovisko: Katedra informatiky
Školiteľ: Mgr. András Varga

Bratislava, 2016
Martin Pribula



Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Martin Pribula
Študijný program: informatika (Jednoduché štúdium, bakalársky I. st., denná forma)
Študijný odbor: informatika
Typ záverečnej práce: bakalárska
Jazyk záverečnej práce: slovenský
Sekundárny jazyk: anglický

Názov: Dimenzionálny model Akademického Reportovacieho Systému
Dimensional Model for Academic Reporting System

Cieľ: Cieľom bakalárskej práce je vytvoriť logický model pre dátový sklad „Akademický Reportovací Systém“. Daný systém tvorí sada dobre definovaných star schém, dátových kociek a agregáčnych tabuliek podporujúce preddefinované reporty. Primárne a cudzie kľúče zachovávajú dátovú integritu a dobre definované indexy podporujú efektivitu daných reportov. Fyzický model vyrobený z logického modelu má byť implementovaný na databázovom systéme používanom vo svete biznisu.

Literatúra: Christopher Adamson - The Complete Reference: Star Schema
Ralph Kimball, Margy Ross - The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling

Kľúčové slová: Star schéma, Logický model, Analytický Systém, Dimenzionálne a faktové tabuľky, Reporting, Dátový sklad, Databáza

Vedúci: Mgr. András Varga
Katedra: FMFI.KI - Katedra informatiky
Vedúci katedry: doc. RNDr. Daniel Olejár, PhD.
Dátum zadania: 24.10.2015

Dátum schválenia: 27.10.2015

doc. RNDr. Daniel Olejár, PhD.
garant študijného programu

.....
študent

.....
vedúci práce

PodĎakovanie: Chcel by som sa poĎakovať môjmu školiteľovi Andrásovi Vargovi za jeho ochotu a odbornosť v radách ktoré mi dal aj za jeho trpezlivé usmerňovanie pri tvorbe práce.

Abstrakt

Práca pojednáva o procese návrhu logického modelu v rámci databázového skladu. Podrobne sa teoreticky aj prakticky venuje štyrom základným krokom jeho tvorby. Hlavným cieľom práce je vytvoriť takýto model pre dátový sklad akademického reportovacieho systému. Model obsahuje sadu hviezdicových schém, ktoré sú navrhnuté na základe dopredu zvolených biznis procesov z akademickej pôdy. Z logického modelu je vygenerovaná fyzická databáza implementovaná v Oracle serveri, ktorá je pomocou dobre definovaných indexov, primárnych a cudzích kľúčov optimalizovaná na rýchly beh dotazov a dátovú integritu.

Kľúčové slová: Hviezdicová schéma, Logický model, Dimenzionálne a faktové tabuľky, Databáza, Dátový sklad

Abstract

This bachelor thesis focuses on design of logical model of data warehouse. It describes, practically and theoretically, four basic steps of creating such model. Its main goal is to create logical model for data warehouse of academic reporting system. Model contains set of star schemas. These are design based on chosen buissiness processes taking place on academic grounds. Physical implementation of logical model is created on Oracle database server. This implementation is optimized by well designed indexes, primary and foreign keys to preform fast queries and maintain data integrity.

Keywords: Star schema, Logical model, Dimensional and fact tables, Database, Datawarehouse

Obsah

Úvod	1
1 Dimenzionálne modelovanie	2
1.1 Dimenzionálne modelovanie	2
1.2 Hviezdicová schéma	2
1.2.1 Faktová tabuľka	2
1.2.2 Dimenzné tabuľky	4
1.3 Architektúry dátových skladov	6
1.3.1 Kimballov dimenzionálny dátový sklad	6
1.3.2 Iné architektúry	6
2 Logický model	8
2.1 Postup návrh dimenzionálneho modelu	8
2.2 Požiadavky na analitický systém	9
2.3 Biznis procesy	10
2.4 Definované dimenzie v reportovacom systéme	13
2.4.1 Tabuľka zhodnosti	16
2.5 Hviezdicové schémy	16
2.5.1 Zápis študentov do ročníka	17
2.5.2 Zápis predmetov	20
2.5.3 Hodnotenie predmetov	21
2.5.4 Záverečné práce	23
2.5.5 Štátne skúšky	25
2.5.6 Využitie miestností	28
2.5.7 Rozvrh	29
2.5.8 Technické zariadenia	30
2.5.9 Prihlášky	31
2.5.10 Publikácie	32
3 Fyzický model	35
3.1 Implementácia fyzického modelu	35

<i>OBSAH</i>	vii
3.2 Indexy	36
Záver	38
Dodatok	39

Zoznam obrázkov

1.1	Hviezdicová schéma	3
1.2	Kimballov dimenzionálny dátový skald, [1]	7
2.1	Dimenzná tabuľka - Class	15
2.2	Tabuľka zhodnosti	17
2.3	Príklad dotazu zo schémy zápisu študentov do ročníka	18
2.4	Snehová vločka - Zápis študentov do ročníka	19
2.5	Hviezdicová schéma - Zapisovanie predmetov	21
2.6	Hviezdicová schéma - Hodnotenie predmetov	22
2.7	Hviezdicová schéma - Hodnotenie predmetov pre jednotlivé fakulty	23
2.8	Hviezdicová schéma - Záverečné práce	25
2.9	Hviezdicová schéma - Štátne skúšky	26
2.10	Hviezdicová schéma - Obhajoba záverečnej práce	27
2.11	Hviezdicová schéma - Písomná štátna skúška	28
2.12	Hviezdicová schéma - Využitie miestností	30
2.13	Hviezdicová schéma - Rozvrh	31
2.14	Dotaz - Prihlášky na vysokú školu.	33
2.15	Hviezdicová schéma - Prihlášky na vysokú školu.	34
2.16	Hviezdicová schéma - Publikácie	34

Úvod

Dimenzionálny model je výsledným produktom procesu ktorý nazývame dimenzionálne modelovanie. Jeho účelom je vybudovať dátový sklad v ktorom sa budú hromadiť údaje nami zvolených procesov. Úlohou zhromaždených dát je ich distribuovanie ku koncovému používateľovi. Údaje v sklade nemajú byť používané ako operačné dáta nad ktorými vykonávame transakcie. Dáta majú byť jednoducho interpretovateľné a rýchlo prístupné. Musia poskytovať jednoznačné merateľné fakty, ktorým priradíme dostatočujúcu šírku detailov. Dáta nám poskytujú pohľad na rôzne časti zaznamenávaného prostredia, no nemali by byť od seba navzájom izolované. Preto je dôležité ich navrhnuť tak aby sme ich vedeli na niektorých detailoch interpretovať vzájomne medzi sebou. Vytváranie takéhoto modelu obsahuje štyri kroky. Prvým je výber procesov, ktorých dáta nás zaujímajú, popis a definovanie detailov týchto procesov. V druhom kroku sa zdefinuje granularita dát ktoré budeme uskladňovať. Tretí krok definuje kontexty zvolených procesov pričom berie do úvahy zvolenú granularitu z druhého kroku. Posledným krokom je výber metrických údajov, ktoré sú dôležité pre daný proces.

V tejto práci sa pokúsime navrhnúť takýto dimenzionálny model na akademickej pôde. V prvej kapitole sa pozrieme na problematiku a zdefinujeme pojmy ktoré budeme používať. Ukážeme typy tabuliek používané pri modelovaní, ako spolu súvisia a ktoré typy dát patria do ktorých tabuliek. V druhej kapitole vyberieme niekoľko procesov z akademického prostredia, určíme do akej úrovne detailu ich chceme zaznamenávať, vyberieme správne opisné dáta, aby samotné numerické dáta zachytené v dátovom sklade dávali význam. Nakoniec tento logický model naimplementujeme na skutočnej fyzickej databáze, vytvorenej na Oracle serveri. Pre túto implementáciu zdefinujeme a vytvoríme indexy tak, aby podporovali rýchlosť prístupu k informáciám v dátovom sklade.

Kapitola 1

Dimenzionálne modelovanie

1.1 Dimenzionálne modelovanie

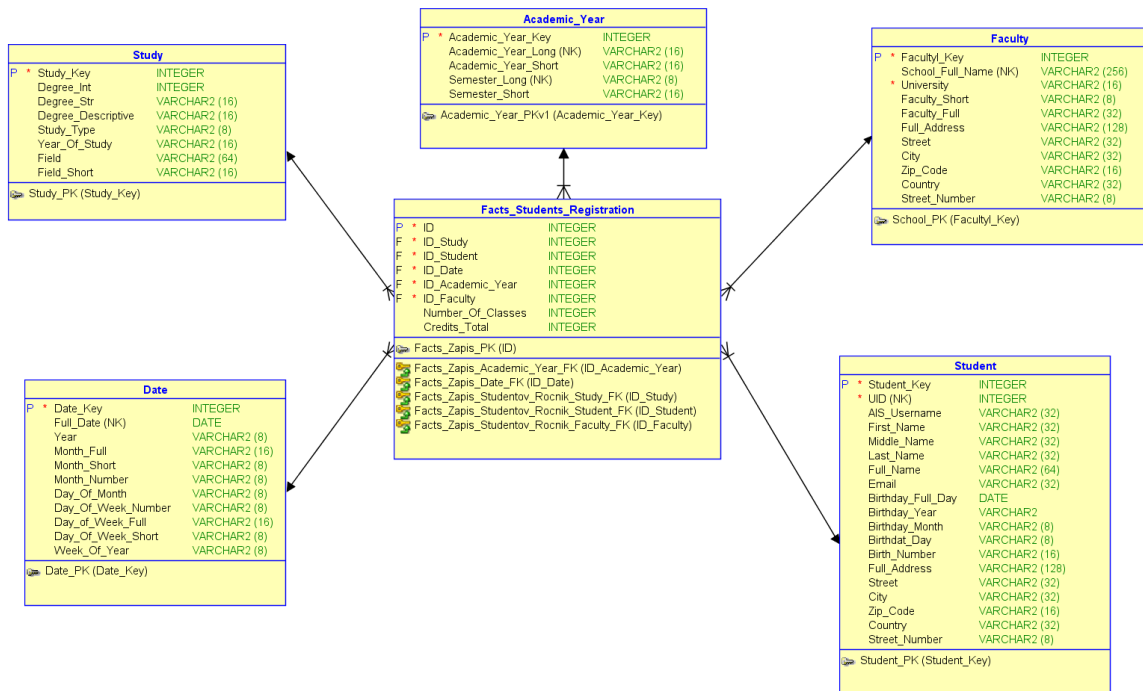
Dimenzionálne modelovanie je dizajn optimalizovaný pre analytický systém. Úlohou analytického systému je vyhodnocovať výsledky biznis procesu. Je odlišný od bežnejších a rozšírenejších transakčných procesov nazývaných aj operačné systémy. Ich úlohou je vykonávať biznis proces a jeho bežné transakcie a úkony. Analytický systém môže byť implementovaný ako relačná databáza alebo multidimenzionálna kocka. Pozeráme sa na údaje z rôznych uhlov a pýtame sa, aký dopad mali jednotlivé podniknuté akcie na celkový prospech biznisu. Tieto údaje vyhodnocujeme a na ich základe upravujeme jednotlivé časti biznis procesu k lepším ziskom. Na rozdiel od operačného systému sa v analytickom systéme nezameriavame na transakcie (vkladanie, aktualizovanie, mazanie) nad dátami, ale robíme nad nimi dotazy a zachováваме historické údaje podľa toho, ako sa mení operačný systém.

1.2 Hviezdicová schéma

Dimenzionálny dizajn v relačnej databáze sa nazýva hviezdicová schéma. Táto schéma obsahuje dimenzné tabuľky a faktovú tabuľku. Faktová tabuľka pozostáva z aditívnych faktov a náhradných kľúčov, ktoré slúžia na prepojenie s každou dimenznou tabuľkou. Dimenzné tabuľky obsahujú stĺpce, ktoré často nazývame dimenzie. Sú spájané do tabuliek podľa logických súvislostí. Meno tejto schémy pochádza z vizuálnej interpretácie, v ktorej keď nakreslíme faktovú tabuľku a po krajoch k nej pripojíme dimenzie, pripomína hviezdu. Vid' obrázok 1.1.

1.2.1 Faktová tabuľka

Centrom hviezdicovej schémy je faktová tabuľka, ktorá sa skladá z kľúčových meraní - faktov, ktoré popisujú proces. Ďalej obsahuje cudzie kľúče, ktoré nám posky-



Obr. 1.1: Hviezdicová schéma

tujú širší kontext pre každé meranie. Na obrázku 1.1 môžeme vidieť faktovú tabuľku `Facts_Students_Registration`, ktorá obsahuje napríklad fakty: `Number_Of_Classes` a `Credits_Total`. Taktiež sa tam nachádzajú cudzie kľúče, ktoré sú označené pred názvom stĺpca skratkou (F).

Je dôležité, aby tabuľka obsahovala všetky biznis definujúce fakty, aj keď sa nám zdá, že môžu byť nepodstatné. Faktová tabuľka nemá veľa stĺpcov. Skôr rastie do hĺbky s veľkým množstvom pribúdajúcich riadkov. Fakty sú aditívne alebo semi-aditívne údaje, zväčša implementované ako integer alebo floating-point čísla. Z toho vyplýva, že napriek veľkej hĺbke faktovej tabuľky je jej implementácia relatívne skromná na diskový priestor, keďže dáta sú prevažne numerické.

Neaditívne fakty

Niektoré fakty však môžu byť neaditívne a nedajú sa teda sčítavať a vytvárať reporty podľa našich predstáv. Sú to napríklad percentuálne alebo pomerové údaje. Riešením problému je neskladovať neaditívne fakty v tabuľke, ale rozbiť ich do viacerých komponentov. Rozbité komponenty už budú aditívne a teda sú skladovateľné. Potom môžeme pôvodný fakt vypočítať priamo v dotaze alebo nami pridanom procese.

Grain

Úroveň detailu na ktorom sú zaznamenané jednotlivé riadky vo faktovej tabuľke nazývame grain. Definovanie tejto vlastnosti je dôležité, aby nevznikali rôzne interpretácie dát. Grain je definovaný ako čo najmenší možný a teda zachytáva údaje v čo najväčšom detaile. Atomické údaje môžu byť pre prípadný report agregované, no ak sú zaznamenané už agregované, nedajú sa rozbiť do podrobnejších detailov. V ďalšom pokračovaní práce budeme grain nazývať aj granularitou tabuľky.

Z obrázku 1.1 pomerne ľahko vidieť, že grain tabuľky `Facts_Students_Registration` je jeden atomický zápis jedného študenta do ročníka a nie počet študentov v ročníku alebo v nejakom štúdiom programe a pod.

Riedkosť

Ďalšou dôležitou vlastnosťou faktovej tabuľky je jej riedkosť. Tabuľka neobsahuje všetky možné kombinácie dimenzií, pretože sú v nej len riadky, ktoré boli zaznamenané skutočnými biznis aktivitami. Preto bývajú faktové tabuľky riedke.

1.2.2 Dimenzné tabuľky

Dimenzné tabuľky obsahujú stĺpce, ktoré poskytujú bohatý kontext pre fakty. Bez tohto kontextu by boli dáta nezmyselné. Údaje by neposkytovali informáciu o časovom úseku v ktorom boli namerané, v akom geografickom kontexte sme ich získali a pod. Na rozdiel od faktovej tabuľky dimenzie sú široké a plytké. Obsahujú veľa stĺpcov a len málo riadkov, ktoré pribúdajú sporadicky. Stĺpce sú textového formátu a sú pripravené rovno na reportovanie. Tabuľky nie sú implementované v tretej normálnej forme, nakoľko jednotlivé reporty si môžu vyžadovať rôzny formát toho istého údaju, alebo chceme filtrovať len pomocou čiastočného údaju (Na obrázku 1.1 v tabuľke `Student` budú redundancie, ktoré nie sú povolené v 3NF, definíciu 3NF nájdeme napríklad u doc. Dr. Tomáša Plachetku [3]. Správnosť 3NF narušia stĺpce `First_Name`, `Middle_Name`, `Last_Name` a `Full_Name` alebo spôsob uloženia adresy, ktorý nespĺňa podmienku, že všetky atribúty vieme určiť z kľúča). Rozumnejšie je však obetovať miesto na disku kvôli rýchlejšiemu vykonávaniu dotazu. Kvôli tomuto dôvodu sú skoro všetky informácie uložené v textovej forme v dimenziách a sú dopredu vypočítané, aby sme ušetrili dotazovací čas.

Kľúče

Jednotlivé riadky v dimenzných tabuľkách sú určené náhradným kľúčom (Na obrázku 1.1 a v celej práci sú vyznačené názvom poľa s príponou `*_Key`), ktorý sa líši od toho

používaného v operačnom systéme, a teda je špeciálny pre analytický systém. Obyčajne je typu integer a je vytváraný pri spracovávaní a nahrávaní dát. Každý záznam navyše obsahuje prirodzený kľúč (Na obrázku 1.1 a v celej práci sú vyznačené skratkou (NK)), ktorý je zdedený z operačného systému a môže niesť relevantnú informáciu pre koncového používateľa. Prirodzené kľúče môžu byť často tvorené viacej ako jedným stĺpcom, aj preto sa vždy definuje náhradný kľúč. Jeho ďalšou výhodou je možnosť úpravy dát riadku v jednom alebo viacerých stĺpcoch a ponechanie si aj starého, aj nového záznamu z dôvodu udržiavania histórie. Toto je možné, lebo existuje náhradný kľúč a teda prirodzený kľúč zostane rovnaký. Dobrými nápadmi sa zdajú aj viac stĺpcové náhradné kľúče. Avšak tieto iba komplikujú a zneprehľadňujú výsledné joiny v dotazoch.

Spájanie dimenzií

Dimenzie sú spájané do dimenzných tabuliek, ktoré reprezentujú hlavné kategórie daného biznisu. Spájajú sa tie, ktoré spolu kontextovo súvisia. Často je ťažké rozhodnúť, ktoré dimenzie patria spolu. Niektoré dokonca nemusia patriť nikam a takýmto vytvárame špeciálnu odpadovú dimenznú tabuľku, kde jednotlivé stĺpce spolu nesúvisia.

Pomaly sa meniace dimenzie

Nahrané dáta v dimenznej tabuľke sa stále môžu meniť v operačnom systéme. Nazývajú sa pomalé lebo oproti faktovým tabuľkám zmeny (riadky tabuľky) pribúdajú rádovo pomalšie. Dimenzie sa môžu meniť rôznymi spôsobmi. Pre potreby tejto práce uvedieme dva z nich: buď nastala chyba a treba ju opraviť, alebo sa dáta zmenili v reálnom svete a treba ich obnoviť. Hviezdicová schéma preto pozná Zmenu typu jedna a Zmenu typu dva.

Zmena typu jedna sa prejaví v dimenzii ako nahradenie pôvodnej hodnoty. Používa sa najmä pre opravu chýb. Toto má za následok zahodenie starej hodnoty a preto aj analytický systém bude ukazovať, že hodnota bola stále rovnaká. Dopad tejto zmeny môžeme vidieť v dvoch rovnakých reportoch z rôzneho dátumu s nezmyselne rôznymi dátami. Napríklad tá istá osoba má dva rôzne dátumy narodenia. Pre zložitejšie dáta sú odchýlky väčšie. Preto by sa mali zmeny typu jeden robiť čo najmenej a treba varovať užívateľov, že takéto situácie nastávajú.

Zmena typu dva má za úlohu zmeniť aktuálny stav dát, pričom zachováva ich históriu a teda pridá nový riadok. Toto je možné vďaka náhradnému kľúču. Táto zmena môže byť tiež mäťúca, keď sa v jednom reporte môžu zjavne vyskytovať duplicitné dáta, preto je nutné používať distinct. Príkladom takejto zmeny je presťahovanie. Potom máme v

tabuľke dva záznamy pre jednu osobu s rôznymi adresami. Takéto zachovávanie histórie je žiadúce.

Pri implementovaní dimenzionálneho modelu treba preto dbať na to, aby každému atribútu bola priradená vhodná pomalá zmena. Zmena typu dva je bezpečnejšia, a teda odporúčanejšia.

1.3 Architektúry dátových skladov

Dátový sklad je systém, ktorý uchováva súčasné a historické dáta nahraté z operačných systémov. Dáta pochádzajú z jedného alebo viacerých zdrojov. Systém údaje analyzuje a reportuje.

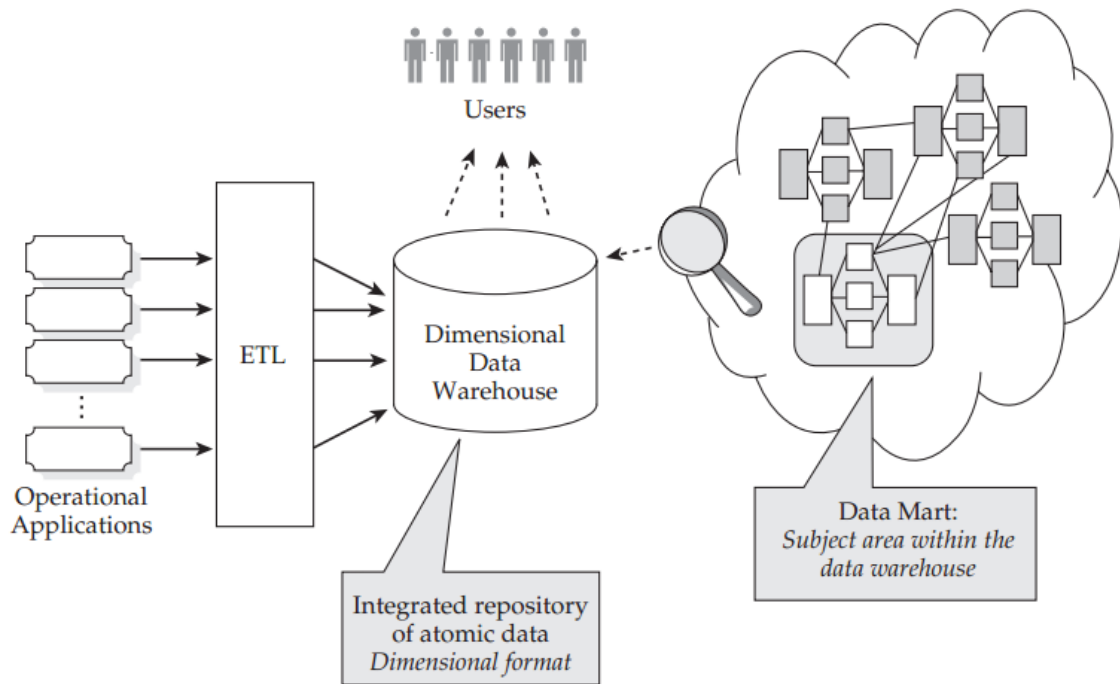
V nasledujúcej časti je uvedený príklad implementácie dátového skladu, ktorú sme sa aj my rozhodli použiť v našej práci. Využíva dimenzionálny dizajn a hviezdicové schémy popísané vyššie.

1.3.1 Kimballov dimenzionálny dátový sklad

Na obrázku 1.2 môžeme vidieť zjednodušenú schému dimenzionálneho dátového skladu. Na ľavo sa nachádzajú dáta operačných procesov v najrôznejších formátoch, ktoré sú spracované ETL procesmi (ETL - „extract, transform, load“). Tieto dáta nahrajú do korporátneho dátového skladu. ETL procesy sú spúšťané periodicky a načítajú dáta z viacerých zdrojov do dátového skladu. Tu sú informácie uložené v jednotnom atomic-kom formáte. Sklad je implementovaný podľa princípov dátového modelovania a teda pozostáva z jednej alebo viacerých databáz obsahujúcich hviezdicové schémy a kocky, ktoré môžu byť priamo prístupné pre koncového používateľa a analytické systémy. V tomto prípade sú data marty len logickou štruktúrou, avšak je možné implementovať ďalšiu vrstvu a to fyzické data marty, kde sú dimenzie agregované alebo preskupované, aby boli vhodnejšie potrebám reportov, dotazov a pod. Pre naše riešenie sme sa rozhodli používať práve túto architektúru.

1.3.2 Iné architektúry

Existujú aj ďalšie architektúry. Podobná Kimballovmu skladu je Immonová korporátne informačná továreň, ktorá sa líši v niekoľkých implementačných detailoch a prístupoch. Oveľa jednoduchší je systém takzvaných samostatných data martov, ktoré sú rýchlejšie a lacnejšie na realizáciu avšak sú vhodné len ako riešenie jednotlivých biznis procesov. Ďalšie informácie ohľadom týchto architektúr je možné nájsť v citovanej literatúre [1], [2].



Obr. 1.2: Kimballov dimenzionálny dátový sklad, [1]

Všetky tri architektúry si našli uplatnenie vo svete biznisu. Kimballová a Immonová sú zamerané širokospektrálnejšie. Naproti tomu samostatné data marty sa sústreďujú len na jednu oblasť. Všetky tri však implementujú dimenzionálny dizajn, každá na inej úrovni, no všetky zachovávajú jeho techniky a princípy.

Kapitola 2

Logický model

V tejto kapitole najskôr ukážeme, akým spôsobom sa správne navrhujú dimenzionálne modely a aký postup sme používali pri návrhu toho nášho. Ďalej popíšeme požiadavky na náš reportovací systém a ako sme tieto požiadavky rozanalizovali. Taktiež popíšeme aké dimenzie sme identifikovali ako potrebné použiť pre naše hviezdnicové schémy, ako sme ich zoskupili do dimenzných tabuliek a aké pomalé zmeny sme pre jednotlivé dimenzie navrhli. Nakoniec sa budeme venovať jednotlivým hviezdnicovým schémam, ich návrhovým detailom ako napríklad navrhnutý grain a charakteristickým vlastnostiam.

2.1 Postup návrh dimenzionálneho modelu

Pri návrhu dimenzionálneho modelu treba brať do úvahy požiadavky ktoré máme na samotný biznis ako aj realitu dát ktoré nám operačné systémy ponúkajú. Návrhu reportovacieho systému pre nejaký biznis musí predchádzať dôkladná a precízna analýza obidvoch faktorov. Pri každom návrhu sa treba riadiť týmito štyrmi kľúčovými rozhodnutiami: Výber biznis procesu, určenie grainu schémy, identifikácia dimenzie a zdefinovanie metrických faktov. Riešenia týchto oblastí sú určené potrebami biznisu a reálnymi dátami, ktoré dostávame z operačných systémov. Na základe týchto štyroch aspektov musí dizajnér alebo tím pomenovať stĺpce a tabuľky, doménové hodnoty a pravidlá. Pričom treba dbať aby navrhovaný systém zodpovedal realite procesu ktorý sa odohráva.

Biznis procesy sú operačné aktivity ktoré sa v našom prípade dejú na akademickej pôde. K takýmto napríklad patrí hodnotenie predmetov, zápis študentov do ročníka, vytváranie rozvrhu, vyhodnocovanie obsadenosti miestností a mnohé iné. Procesy vytvárajú alebo zaznamenávajú údaje, ktoré sa zachytávajú do faktov vo faktových tabuľkách za účelom monitorovania a vylepšovania danej aktivity. Tieto sa sústreďujú na jeden proces. Výber procesu je dôležitý pretože určuje špecifický návrh a umožňuje navrhnúť grain, dimenzie a fakty. Každý proces zodpovedá jednému riadku v dátovom sklade.

Definovanie grainu je ústredným krokom v dimenziálnom návrhu. Grain ustáli čo presne jeden riadok vo faktovej tabuľke reprezentuje. Určenie grainu sa stáva zaväzujúcim rozhodnutím pre celý dizajn. Grain musí byť zvolený pred faktami a dimenziami, pretože musí byť zvážený pri ich výbere a teda ich do istej miery určuje. Táto stálosť zaručuje jednotvárnosť na celom návrhu, ktorý je kritický pre výkon a ľahké používanie systému. Grain sa navrhuje atomický - najmenší možný ako sa to dá. Toto podporuje aj nepredvídateľné a často neočakávané požiadavky na reportovací systém. Každá faktová tabuľka musí mať vlastný grain a nemôžu sa miešať dve tabuľky s rôznymi grainami. Dimenzie nám poskytujú kontextové údaje o tom „kto? kedy? ako? a kde?“ niečo urobil v biznis procese. Dimenzné tabuľky obsahujú opisné údaje používané reportovacím systémom na filtrovanie a grupovanie údajov. Podľa určeného grainu faktovej tabuľky môžu byť identifikované všetky možné dimenzie. Ak je to možné dimenzia spojená s riadkom faktovej tabuľky by mala byť jednodnotová. Dimenzie sú často nazývané aj dušou dátového skladu pretože obsahujú vstupné body a opisné označenia ktoré reportovaciemu systému umožňujú voľnosť pre analýzu. Čas a úsilie venované výberu a návrhu dimenzií sa často na prvý pohľad zdajú až neprimerane dlhé, avšak sú nosným pilierom reportovacieho systému a preto nemôžu byť zanedbané.

Prostredníctvom faktov vieme merať výsledky z vybratého biznis procesu. Fakty sú skoro vždy numerické. Jeden fakt má vždy vzťah jedna ku jednej k meranej udalosti tak ako je to popísane grainom tabuľky. Takto faktová tabuľka zodpovedá fyzickej pozorovanej udalosti a nie požiadavkám niektorého konkrétneho reportu. Vo faktovej tabuľke je dovolené skladovať len údaje s rovnakým grainom. Napríklad, nemôžeme uložiť do jednej tabuľky hodnotenie jedného predmetu študenta a súčasne aj jeho vážený priemer v danom ročníku.

Takto popisuje návrh dimenzionálneho modelu Ralph Kimball vo svojej knihe o dátových skladoch [2]. Týmto postupom navrhovania dimenzionálneho modelu sme sa riadili aj pri vytváraní nášho akademického reportovacieho systému. A teda sme vždy vybrali jeden biznis proces, určili jeho grain a navrhli pre neho dimenzie a fakty, pričom sme brali do úvahy všetky zásady spomínané v tejto časti.

2.2 Požiadavky na analitický systém

Dôležitým krokom pri implementácii dátového skladu je dobre definovať a pochopiť požiadavky na systém ktorý budeme vytvárať, preto v nasledujúcej časti popíšeme, čo sa od nášho dátového skladu očakáva a aké sú požiadavky na jeho funkcionality. Systém musí byť schopný zaznamenávať nasledujúce procesy s popísanými vlastnosťami odohrávajúcimi sa na akademickej pôde.

Zápis študentov do ročníka, kde môžeme vyberať záznamy podľa študenta, odboru na

ktorý sa zapísal, fakulty, univerzity či dátumu zápisu. Taktiež budeme schopný rýchlo nájsť študentov zvolenej fakulty.

Zápis jednotlivých predmetov v danom semestri, ktoré vieme vyhodnocovať podľa rovnakých kritérií ako zápis študentov a navyše aj podľa samotného študenta. Taktiež vieme vyhodnocovať prerekvizity zapísaného predmetu.

Hodnotenie jednotlivých zapísaných predmetov študenta, ktoré sa dajú vyberať podľa odboru, fakulty, univerzity, dátumu skúšky alebo semestra. Zároveň zaznamenáva podrobné ohodnotenie študenta tak, aby sa dal vypočítať vážený priemer. Chceme vedieť porovnávať dáta vygenerované hodnotením predmetov a zápisom predmetov.

Tvorba rozvrhov v danom semestri. Teda prostredníctvom systému sa musia dať filtrovať záznamy podľa miestností v ktorých sa vyučovanie odohráva, podľa predmetov ktoré sa vyučujú a dátumu kedy prebiehajú.

Využitie jednotlivých miestností na univerzite alebo fakulte v čase. Budeme mať k dispozícii údaje o tom kedy a kým je miestnosť používaná a zároveň aj jej kapacitné využitie.

Prihlasovanie na jednotlivé štúdiá univerzity v danom akademickom roku, ktoré sa dajú rozlišovať podľa fakulty odboru, ale aj podľa predchádzajúceho vzdelania a školy odkiaľ uchádzači prichádzajú.

Publikačná činnosť študentov, kde sa dá rozoznávať článok v odbornom časopise od študentskej vedeckej odbornej činnosti a taktiež je zaznamenané ohodnotenie študenta. Filtrovanie podľa študenta, jeho štúdiá, dátumu publikácie a časopisu bude umožnené. Sledovanie životného cyklu bakalárskych, diplomových a dizertačných prác. Systém bude vedieť zachytiť časové intervaly pre jednotlivé statusy (vypísaná práca, priradená študentovi, odovzdaná, ohodnotená a obhájená) v ktorých sa práca nachádzala.

Časové využívanie technických zariadení v budovách univerzity pre jednotlivé semestre. Príklady zariadení sú počítačové miestnosti, prístroje v laboratóriu a pod. Pričom systém bude vedieť rozlíšiť typy týchto zariadení.

Priebeh a vyhodnotenie štátnej skúšky pre každý predmet ktorého sa študent zúčastní, pričom sa dajú rozlíšiť jednotlivé atribúty rôznych typov skúšky ako je obhajoba, písomná, alebo ústna skúška.

2.3 Biznis procesy

V tejto časti si bližšie popíšeme biznis procesy, ktoré sme definovali a z ktorých budeme zbierať dáta na vyhodnocovanie. Pozrieme sa na to čo by mali zachytávať, na aké udalosti na univerzite sa sústredia a ktoré dáta nimi chceme merať.

Pre začatie štúdiá je nutné, aby sa každý študent zapísal na štúdiijný program, ktorý chce v danom roku navštevovať. Proces zápisu študentov do ročníka zachytáva aktivitu

pre každého študenta ktorý sa v danom akademickom roku zapíše na nejaké štúdium - pod štúdiom rozumieme štúdijný program ponúkaný určitou katedrou, ktorý je definovaný aj ročníkom do ktorého sa študent zapisuje. Ide o zaznamenanie samotného zápisu kde zachytíme koľko si študent zapísal predmetov a koľko mu tieto predmety potenciálne prinesú kreditov. Nesústredíme sa na konkrétne predmety, ale na to kto a kam sa zapísal v ktorý školský rok.

Podprocesom zápisu do ročníka by sme mohli nazvať zápis predmetov. Tento zaznamenáva konkrétnejšiu informáciu. Každý zapísaný predmet študenta je osobitne poznačený do dátového skladu. Pamätáme si kto si predmet zapísal, v aký deň, v ktorom semestri akademického roku, či predmet potrebuje na zapísanie nejaké prerekvizity, kto predmet vyučuje a vyučoval, koľko sa za predmet na danom štúdiu dá dosiahnuť kreditov. Tento proces je tiež definovaný tým ktorá katedra daný predmet ponúka, a ktorá ho ponúkala v minulosti. Či je predmet povinný, povinne voliteľný, alebo voliteľný určuje štúdijný program na ktorom sa predmet zapisuje a preto je nutné tieto typy rozpoznávať.

Každý zapísaný predmet musí byť na konci semestra v ktorom sa odohrával aj ohodnotený. Hodnotenie predmetov je teda aktivita ktorá každému zapísanému predmetu študenta v danom akademickom roku priradí vyhodnotenie a teda známku A, B, C, D, E alebo FX, z ktorých sme potom schopný študentovi vypočítať vážený priemer. Známkou je študent schopný dosiahnuť skúškou alebo hodnotením, podľa typu ukončenia predmetu. Rôzne spôsoby ukončenia je potrebné vedieť rozpoznať. Tiež je nutné aby osoba ktorá prezerá hodnotenia predmetov mala prístup len k takým záznamom, ktoré sa týkajú jej zverenej fakulty respektíve univerzity, nakoľko sú to citlivé údaje.

Každý študent ukončuje štúdium záverečnou prácou. Proces analýzy týchto prác podrobne sleduje ich prechody medzi jednotlivými stavmi. Stavby ktoré zaznamenáva sú nasledovné: práca začína vo vypísanom stave a teda jej téma a vedúci sú známi, ale ešte nie je priradená študentovi. Následne sa je schválená respektíve neschválená garantom štúdia a priradená konkrétnemu študentovi. Odovzdaná práca v systéme čaká na svoje ohodnotenie od vedúceho a oponenta. Keď práca dostane známky je v ohodnotenom stave a prejde buď do stavu obhájená, alebo neobhájená. Pre všetky tieto stavy je potrebné zachytiť čas, ktorý v nich strávil.

Niektoré univerzity a fakulty pri ukončení štúdia navyše vyžadujú aj štátne skúšky. Tieto sa skladajú z viacerých častí a rôznych typov skúšok. Skúšky môžu byť z rôznych predmetov. Tieto rozdeľujeme na dva typy ústne a písomne. Tretím typom štátnej skúšky je obhajoba záverečnej práce. Tieto tri majú spoločné vlastnosti, no v mnohom

sa aj líšia. Preto je nutné aby sme tieto vlastnosti vedeli všetky zaznamenať a rozlíšiť ktoré patria ku ktorej skúške. Zaznamenávať budeme výsledky štátnych skúšok, komisiu ktorá bola prítomna, predmet štátnej skúšky. Tieto vlastnosti sú spoločné. Podrobnejšie sa budeme venovať osobitným vlastnostiam v samostatnej sekcii 2.5.5.

Aktivita zaznamenávaná pred samotným štúdiom na univerzite je podávanie prihlášok na vysokú školu. Pri tomto procese zachytávame samotné podanie prihlášky a informácie z nej prístupné. Tieto obsahujú údaje o potenciálnom študentovi, aké najvyššie vzdelanie dosiahol a na aký štúdiálny program sa hlási. Po samotnom podaní prihlášky prebieha prijímacie konanie, ktoré môže mať rôzny priebeh a výsledky. Zaujímá nás či uchádzač o štúdium bol prijatý a či potreboval absolvovať aj testy. Ak uchádzač absolvoval testy, zaujíma nás aké výsledky dosiahol.

Študenti a iní pracovníci na univerzite sa zúčastňujú aj publikačnej činnosti. Túto aktivitu taktiež definujeme ako proces ktorý budeme sledovať a vyhodnocovať. Sústreďovať sa budeme na jednotlivé publikácie študentov a to ako boli kreditovo ohodnotené. Rozlišujeme článok v časopise, študentskú vedeckú odbornú činnosť a prípadné iné typy publikácií.

Aktivity na vysokej škole sa odohrávajú najmä v budovách univerzity. Vyhodnocovanie percentuálneho využitia jednotlivých budov, sektorov a miestností môže byť kľúčové, aby sme vedeli čo sa kde deje a ktoré objekty by sme mohli viacej zaťažiť respektíve, odľahčiť ich využitie. Preto budeme sledovať spodnú úroveň objektov a to jednotlivé miestnosti a na koľko percent sú v čase využité, ale aj koľko ľudí sa v nich nachádza podľa rozvrhu. Sledovať budeme aj kabinety, kancelárie, sklady a rôzne iné typy miestností.

Jednotlivé fakulty potrebujú na svoje fungovanie a robenie výskumu rôzne zariadenia, ktoré sa môžu veľmi líšiť podľa zamerania univerzity alebo fakulty. Sledovanie využitia takýchto prístrojov a zariadení môže byť vitálne pre rozhodovanie jednotlivých katedier o neskoršom vyradení, alebo prípadnom zaobstaraní alebo preorganizovaní týchto prostriedkov. Preto budeme zachytávať kto používa tieto zariadenia a koľko času sú využité percentuálne a absolútne. Pod zariadeniami v tomto procese budeme rozumieť väčšie vybavenie ako počítačové miestnosti, roboty, geodetické merače a iné. Nebudeme sa zameriavať na malé položky typu skúmavka, pipeta.

Na začiatku každého semestra je nutné priradiť jednotlivé vyučované predmety do miestností a rozdeliť ich v čase. Vzniká rozvrh na základe ktorého potom vieme sledovať jednotlivé miestnosti v čase, alebo sa vieme pozrieť ktorý predmet sa kedy a kde koná.

Proces tvorby a analýzy rozvrhu budeme sledovať a dáta z neho ukladať k neskoršej analýze.

2.4 Definované dimenzie v reportovacom systéme

Ako už bolo spomenuté výber vhodných dimenzií je jednou z najdôležitejších častí vytvárania reportovacieho systému. Nasledujúce dimenzie sme identifikovali ako potrebné pre definovanie kontextu k jednotlivým faktom biznis procesov, ktoré sme vybrali pre náš akademický reportovací systém.

Dimenzná tabuľka *Student* nám poskytuje súvislosti medzi dátami a jednotlivými študentami. Do tejto tabuľky sme zoskupili nasledovné dimenzie: Meno, priezvisko a prípadné stredné meno, ktoré sú uložené v samostatných poliach ale sú aj spojené pre rýchlejší prístup k informáciám. Všetky tieto dimenzie majú navrhnutý druhý typ pomalej zmeny nakoľko meno študenta sa môže meniť bez toho aby to bola oprava chyby (napr. svadba). Adresa študenta, ktorá je taktiež uložená ako celá adresa aj jednotlivé časti - krajina, mesto, ulica, číslo domu, poštové číslo. Pre tieto dimenzie sme navrhli pomalú zmenu druhého typu nakoľko zmena adresy je bežná vec. Dátum narodenia, ktorý obsahuje deň, mesiac, rok a rodné číslo. Pomalá zmena týchto dimenzií je typu jedna, lebo zmena dátumu narodenia je len opravením faktickej chyby. Poslednými dimenziami v tejto tabuľke sú údaje prebraté zo systému AIS a to univerzitné identifikačné číslo, ktoré je zároveň aj prirodzeným kľúčom prebratým z operačného systému a užívateľské meno. Obidve dimenzie majú navrhnutú zmenu prvého typu.

Ďalšou dimenziou je názov univerzity a fakulty, ktoré sú uložené každá v samostatnom poli aj spoločne ako jedno pole. Dimenzia fakulty má takisto uloženú aj svoju skratku. Typy zmien implementované na týchto poliach sú typu jedna, nakoľko neočakávame zmenu názvu fakulty alebo univerzity. K týmto kontextom sme pridali ešte adresu sídla fakulty, ktorá je štruktúrou a typom zmeny rovnaká ako v predchádzajúcej tabuľke a zoskupili sme ich do dimenznej tabuľky *Faculty*. Táto popisuje fyzický kontext univerzity a fakulty na ktorej sa určený biznis proces odohráva. Prirodzeným kľúčom v tejto tabuľke je už spomínaný celý názov fakulty aj s univerzitou pod ktorú patrí.

Study je tabuľka ktorá nám poskytuje abstraktné súvislosti štúdia, viažúce sa na daný fakt. Obsahuje celý názov štúdiijného programu ako aj jeho skratku a kód z AIS-u, stupeň štúdia v troch podobách (slovný, číselný, skratka), ročník v ktorom sa študent nachádza a typ štúdia, čo znamená či je štúdium denné alebo externé. Všetky zmeny dimenzií sú prvého typu, lebo nepredpokladáme zmeny názvu programu ani stupňa štúdia počas jedného roka, keďže na začiatku nového školského roka budeme štúdium definované novým rokom. Prirodzeným kľúčom je kód programu spoločne s rokom štúdia.

Dimenzia dátumu je najbežnejšia a najčastejšie používaná, keďže chceme zaznamenávať kedy sa fakty udiali. Naša dimenzná tabuľka dátumu obsahuje vyjadrenie dňa v mesiaci, dňa v týždni, mesiaca v roku, ako aj celý dátum pohromade, ktorý je zároveň primárnym kľúčom tabuľky *Date*. Zmena prvého typu je prirodzená, lebo záznam dátumu sa nemôže meniť.

Podobná dimenzia vyjadruje v ktorom akademickom roku sa proces udial, ktorá tvorí tabuľku *Academic_Year*. Táto tabuľka podľa potreby obsahuje aj možnosť zaznamenať semester. Prirodzeným kľúčom je celé vyjadrenie akademického roka, a v tabuľke so semestrom aj semester. Taktiež je implementovaná zmena prvého typu.


Pre biznis procesy tvorby rozvrhu, využívania miestností a využívania technických zariadení potrebujeme dimenzie ktoré nám určia, kde daný proces zaznamenávame. Na to slúži dimenzná tabuľka *Room*, ktorá v sebe združuje dimenzie, kód miestnosti, ktorý je zároveň aj prirodzeným kľúčom, poschodie na ktorom sa nachádza, sekciu budovy a samotnú budovu pre ktorú ukladáme aj jej adresu. Pre dimenziu sekcie a kódu miestnosti sme navrhli zmenu druhého typu, ak by sa zmenil názov sekcie a s tým aj všetkých kódov aby sme to boli schopný zachytiť. Pre ostatné sme implementovali zmenu druhého typu.

Dimenzie názov publikácie, jej typ, a meno učiteľa ktorý je školiteľom sú zhromaždené do tabuľky *Publication*. Tieto dimenzie využívame pri zaznamenávaní jednotlivých publikácií. Všetky majú implementovanú zmenu druhého typu, pretože sa môže v priebehu zmeniť názov, typ, alebo školiteľ. Publikácie môžu byť uverejňované v časopisoch. Údaje o časopisoch nám zaznamenávajú kontexty: názov časopisu, jazyk v ktorom vychádza, jeho typ a krajina v ktorej vychádza. Všetky majú navrhnuté zmeny druhého typu nakoľko časopis môže zmeniť meno, jazyk, alebo typ.

Pri viacerých procesoch potrebujeme vedieť detaily o vyučovanom, zapisovanom alebo hodnotenom predmete. Preto máme nasledovné dimenzie: meno predmetu, kód z AIS-u, jazyk v ktorom je predmet vyučovaný, počet hodín do týždňa, semester v ktorom sa dá zapísať, ktorá katedra ho poskytuje, kto predmet vedie, a kedy bol predmet otvorený prvý a posledný krát. Tieto sú združené v dimenznej tabuľke *Class*. Táto vo svojich rôznych variáciách niektoré dimenzie neobsahuje, lebo nie sú vitálne pre konkrétne hviezdicové schémy. Dimenzia kód predmetu z AIS-u má implementovanú pomalú zmenu prvého typu. Katedra, ktorá predmet vyučuje má implementovanú zmenu tretieho typu. Ako možno vidieť na obrázku 2.1 tabuľka obsahuje pole *Department_Current*, ktoré hovorí o katedre ktorá predmet vyučuje v súčasnosti a pole *Department_Previous*, ktoré hovorí o predchádzajúcej katedre ktorá predmet poskytovala predtým. Tento typ pomalej zmeny sme implementovali pretože sa chceme pozeráť aj na dáta z minulosti a to konkrétne ktorá fakulta predmet poskytovala. Rovnakú zmenu je vidieť z obrázku 2.1 aj na dimenzii učiteľa. Ostatné dimenzie implementujú pomalú zmenu typu dva. Predmet môže vyžadovať na svoje zapísanie nejakú prerekvizitu a teda iný predmet.

Preto nám vznikne hierarchia na tejto dimenzii, ktorá môže mať viacej úrovní, nakoľko predmety na seba môžu teoreticky nadväzovať ľubovoľne dlho. Riešeniu tohto hierarchického problému sa venujeme v sekcii 2.5.2

V procese zaznamenávajúcom využívanie technických zariadení potrebujeme vedieť aj

Class		
P *	Class_Key	INTEGER
	Name	VARCHAR2 (64)
	Code (NK)	VARCHAR2 (32)
	Number_Of_Classes_Per_Week	VARCHAR2 (32)
	Language	VARCHAR2 (32)
	Type_Of_Evaluation	VARCHAR2 (32)
	Semester	VARCHAR2 (64 CHAR)
	Department_Current	VARCHAR2 (64 CHAR)
	Department_Previous	VARCHAR2 (64 CHAR)
	Teacher_Full_Current	VARCHAR2 (64 CHAR)
	Teacher_First_Name_Current	VARCHAR2 (32 CHAR)
	Teacher_Last_Name_Current	VARCHAR2 (32 CHAR)
	Teacher_Full_Previous	VARCHAR2 (64 CHAR)
	Teacher_First_Name_Previous	VARCHAR2 (32 CHAR)
	Teacher_Last_Name_Previous	VARCHAR2 (32 CHAR)
 Class_PK (Class_Key)		

Obr. 2.1: Dimenzná tabuľka - Class

ich detaily, ktoré nám poskytnú dimenzie: meno zariadenia, typ zariadenia a kľúč z AIS-u, ktorý je aj prirodzeným kľúčom. Meno a typ implementujú zmenu druhého typu. Všetky sú v spoločnej dimenznej tabuľke *Equipment*. Pri tomto procese chceme tiež vedieť kto dané zariadenie používal a teda zachytávame dimenzie: meno používateľa, univerzitné identifikačné číslo, používateľské meno z AIS-u a email, pre prípadné kontaktovanie osoby. Meno a email sú premenlivé údaje preto sme navrhli zmenu druhého typu. Ostatné dimenzie majú zmenu prvého typu nakoľko neočakávame ich prirodzenú zmenu. Tieto sú v spoločnej tabuľke *User*, kde je prirodzeným kľúčom univerzitné identifikačné číslo.

Súvislosti pre jednotlivé diplomové, bakalárske a dizertačné práce získavame pomocou názvu práce, typu (bakalárska/diplomová/dizertačná/), mena školiteľa a katedry ktorá

prácu vypísala. Zmeny v názve práce sa môžu vyskytnúť, avšak nie je potrebné ich špeciálne zaznamenávať a preto sme pre všetky stĺpce navrhli pomalú zmenu typu jedna. Dimenzie sú zhromaždené v tabuľke *Thesis*, ktorej prirodzeným kľúčom je názov práce. Taktiež zachytávame stav práce ktorý je v samostatnej *Status* tabuľke a teda jej kľúčom je názov stavu.

Detaily procesu štátnych skúšok zachytávajú dimenzie: kód predmetu štátnej skúšky z AIS-u, názov predmetu, komisia, a typ skúšky. Tieto sú zoskupené do centrálnej dimenznej tabuľky *Examintaion*, kde je prirodzeným kľúčom kód z AIS-u. Podľa typu skúšky potom definujeme špecifické dimenzie pre jednotlivé typy a spájame ich do špecifických dimenzných tabuliek, ktoré obsahujú aj dimenzie z centrálnej dimenznej tabuľky. Zdefinovali sme tri špecifické dimenzné tabuľky pre ústne skúšky, písomné skúšky a obhajoby prác. Ústna skúška obsahuje vlastné dimenzie: vyťahnutú tému a skúšajúceho. Písomná skúška obsahuje navyše oproti centrálnym dimenziám čas na vypracovanie, počet otázok a opravujúceho. Obhajoba práce zahŕňa navyše školiteľa, oponenta, známku od školiteľa, známku od oponenta, čas prezentácie, názov práce, a kód práce z AIS-u, ktorý je aj prirodzeným kľúčom.

2.4.1 Tabuľka zhodnosti

Rôzne hviezdicové schémy môžu používať tie isté dimenzie. Takýmto dimenziám hovoríme, že sú zhodné v našom modeli. Pre uvedenie si podobnosti, alebo naopak odlišnosti štruktúry dát ktoré v sklade uchováваме je dobré si všimnúť, ktoré dimenzie sa nám zhodujú naprieč hviezdicovými schémami. Potom uvidíme celý reportovací potenciál nášho systému, alebo naopak jeho rozkúskovanosť. Uvidíme, ako sa dáta dajú krížovo prepájať naprieč schémami pomocou týchto dimenzií. Dobrou a relatívne kompaktnou ilustráciou tejto vlastnosti systému je takzvaná tabuľka zhodnosti (z anglického *conformance matrix*). Na obrázku 2.2 môžeme vidieť, že každý riadok je jedna z hlavných faktových tabuliek nášho systému. Stĺpce zase predstavujú jednotlivé dimenzie ktoré sme definovali. Políčko v ktorom je zaznačené „x“ znamená, že faktová tabuľka v riadku má v sebe cudzí kľúč na príslušnú dimenziu v stĺpci. Z obrázku jasne vidieť ktoré dimenzie sú najviac zhodné naprieč systémom a ktoré sú unikátne. Na základe tejto tabuľky usudzujeme, že náš systém je navrhnutý dobre, lebo má relatívne veľkú zhodnosť.

2.5 Hviezdicové schémy

Nasledujú podrobné opisy jednotlivých hviezdicových schém, kde vysvetlíme prečo sú navrhnuté tak ako sú, popíšeme ich jednotlivé časti a vysvetlíme, ako sme vyriešili

Faktová tabuľka/Dimenzia	Fakulta	Dátum	Akademický rok	Štúdium	Študent	Semester	Predmet	Miest	Publik	Vybab	Práca	Skúška	Časop	Použív
Facts_Students_Registration	x	x	x	x	x	x								
Facts_Registered_Class	x	x	x	x	x	x	x							
Facts_Class_Evaluation	x	x	x	x	x		x							
Facts_Thesis	x	x	x	x	x						x			
Facts_Degree_Examination	x	x	x	x	x							x		
Facts_Room_Utilization	x	x	x			x	x	x						
Facts_Schedule	x	x	x			x	x	x						
Facts_Technical_Equipment	x	x	x			x		x		x				x
Facts_Application	x	x	x	x	x									
Facts_Publications	x	x	x	x	x	x			x				x	

Obr. 2.2: Tabuľka zhodnosti

niektoré problémy a vzťahy v jednotlivých schémach. Každá časť bude obsahovať jeden alebo viacej obrázkov skutočného logického návrhu danej schémy. Navrhnutý logický model je priložený ako príloha. Pre čitateľa práce v tlačenej podobe je odporúčané otvoriť si projekt logického návrhu z prílohy, aby videl detaily obrázkov.

2.5.1 Zápis študentov do ročníka

Biznis proces popisujúci zápis študentov do ročníka je logicky znázornený v samostatnej hviezdicovej schéme na obrázku 1.1. Faktová tabuľka Facts_Students_Registration obsahuje svoj primárny kľúč - stĺpec ID. Cudzie kľúče na odkazujúce sa na dimenzné tabuľky sú znázornené veľkým písmenom F a hviezdíčkom. Toto platí pre všetky nasledujúce faktové tabuľky. Samotné merané fakty sú posledné dva stĺpce tabuľky a to počet zapísaných predmetov a celkový počet zapísaných kreditov v aktuálnom ročníku prislúchajúce k sebe navzájom uvedenom poradí. Grain faktovej tabuľky sme sa snažili zvoliť čo najjemnejší a teda keď sa podrobne pozrieme aké dimenzie sa pripájajú na fakty je zrejmé že ukladáme záznamy pre jedného študenta zapísaného na jeden štúdijný program v jednom roku. To znamená že ak sa študent zapíše na dva rôzne štúdijné programy sme schopný ich rozlíšiť vďaka atomicite grainu ktorý sme navrhli. Dimenzie ktoré sme sa rozhodli pripojiť na faktovú tabuľku sú: fakulta, štúdijný program, dátum, akademický rok, a študent. Tieto nám umožňujú filtrovať prípadné reporty podľa jednotlivých študentov, alebo miest odkiaľ študenti pochádzajú. Na uložené fakty sa tiež môžeme pozerieť z pohľadu jednotlivých fakúlt, univerzít alebo štúdijných odborov na ktoré sa študenti zapísali alebo do ktorého ročníka sa zapísali. Dimenzia dátumu a akademického roku nám umožní filtrovať zápisy z konkrétneho časového intervalu. Všetky spomínané filtrovania pomocou dimenzií nám umožňujú aj agregovanie rovnakým spôsobom. Uvedieme príklad detailného reportu dát z tejto hviezdicovej schémy, aby sme ukázali aké rôzne spôsoby pohľadu na dáta nám umožňujú zvolené dimenzie. Chceme celé meno, univerzitné identifikačné číslo, email a dátum zápisu všetkých zapísaných študentov v akademickom roku 2012/2013, ktorý sa zapísali do prvého ročníka štúdia na Prírodovedeckej fakulte na štúdijný odbor molekulárna biológia a nemajú zapísaný minimálny počet kreditov alebo majú príliš veľa kreditov. A teda dotaz na náš systém by vyzeral ako na obrázku 2.3. Dotaz na schému bude veľmi rýchli

pretože používame už vopred pripravené stĺpce a teda nemusíme žiadnymi funkciami spájať dáta. Môže sa zdať, že JOIN operácii je veľa, ale ako uvádzame v sekcii 3.2 kapitoly o fyzickej implementácii, indexy navrhnuté na cudzích kľúčoch faktovej tabuľky a primárnom kľúči každej dimenznej tabuľky umožnia ORACLE plánovaču vykonávania dotazov použiť najrýchlejší možný spôsob JOIN operácie podľa pravidiel ako sú popísane v dokumentácii [4]. Týmto sme znázornili reportovací potenciál a rýchlosť tejto schémy. Ukázali sme prečo je tento návrh dobrý a prečo sú zvolené dimenzie dobre vybrané a potrebné.

```

SELECT s.Full_Name AS "Celé meno",
       s."UID (NK)" AS "Univerzitné identifikačné číslo",
       s.Email AS "Email",
       d."Full_Date (NK)" AS "Dátum zápisu"
FROM Facts_Students_Registration f
LEFT JOIN Faculty u
ON f.ID_Faculty = u.Faculty_Key
LEFT JOIN Student s
ON f.ID_Student = s.Student_Key
LEFT JOIN "Date" d
ON f.ID_Date = d.Date_Key
LEFT JOIN Study st
ON f.ID_Study = st.Study_Key
LEFT JOIN Academic_Year
ON f.ID_Academic_Year = a.Academic_Year_Key
WHERE u.Facult_Full = "Prírodovedecká fakulta Univerzity Komenského"
AND st.Field = "Molekulárna biológia"
AND a."Academic_Year_Long (NK)" = "2012/2013"
AND st.Year_Of_Study = "prvý"
AND (f.Credits_Total < 40
OR f.Credits_Total > 90)

```

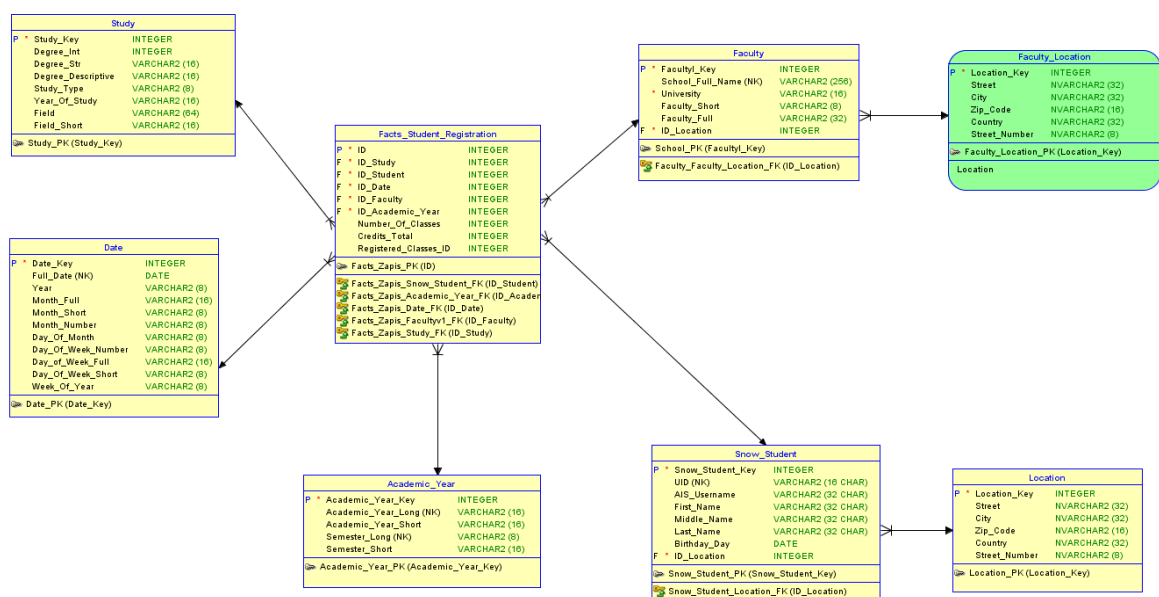
Obr. 2.3: Príklad dotazu zo schémy zápisu študentov do ročníka

Snowflaking

Ako sme uviedli v prvej kapitole v sekcii 1.2.2 o dimenzných tabuľkách hviezdicové schémy nie sú navrhované v tretej normálnej forme. Pre tento biznis proces sme sa rozhodli navrhnuť normalizovanú formu dimenzných tabuliek, aby sme ukázali ako vyzerať a ako sa navrhuje takzvaná snowflake alebo snehová vločka z hviezdicovej schémy. Na obrázku 2.4 je snehová vločka vytvorená z hviezdicovej schémy zápisu študentov do ročníka 1.1. Dimenzná tabuľka *Student* je rozložená na dve, a to tak že sme vyrobili novú tabuľku *Location*, ktorá je napojená pomocou cudzieho kľúča *ID_Location*. Rovnaká úprava je aj na tabuľke *Faculty*, táto je však pripojená na View *Faculty_Location*, aby sme zdôraznili, že dimenzia adresy v tejto snehovej vločke zohráva dve úlohy: adresa študenta a adresa fakulty. Takáto dimenzia sa nazýva aj role-playing dimenzia. Tieto úpravy však nespĺňajú definíciu 3NF, ktorú uvádza doc. Dr. Tomáš Plachetka

vo svojej prezentácii ku predmetu o databázach [3]. Normalizácia ktorú sme použili totižto nespĺňa definíciu prvej normálnej formy, ktorá je podmienkou pre 3NF a teda obsahuje zložený atribút *Full_Name*, ktorý je zložený z troch atribútov *First_Name*, *Middle_Name*, *Last_Name* a tým porušuje 1NF. Takéto stĺpce sme si naďalej ponechali kvôli pohodlnému vytváraniu reportov, aby sme ich nemuseli spájať pri behu dotazov a tým zrýchlili ich beh.

Grain vytvorenej snehovej vločky ako aj logický návrh a pripojenie ostatných dimenzných tabuliek na faktovú tabuľku zostal rovnaký. Pri tvorbe snehovej vločky sa totižto mení iba normalizácia dimenzií, ktorá nemôže nijako narušiť logický návrh vopred danej štruktúry tabuliek.



Obr. 2.4: Snehová vločka - Zápis študentov do ročníka

Predvypočítavanie agregovaných dát

Ak vieme že niektoré dáta bude musieť reportovací systém pravidelne agregovať, pretože budú často žiadané, tak je výhodou si ich už dopredu pripraviť. Zagregovať si dáta podľa najviac používanej dimenzie a takto si ich uložiť do samostatnej hviezdicovej schémy.

V našom dátovom sklade sme identifikovali, že častou požiadavkou na systém bude vybrať študentov zgrupovaných podľa štúdiijného programu na ktorý sa zapísali. Vytvorili sme teda novú schému, ktorá vychádza zo schémy o zapísaných študentoch, ale v tomto prípade už neobsahuje dimenznú tabuľku *Student*, pretože táto sa agregovaním stratila. Takisto sa stratila tabuľka ktorá hovorila o dátume jednotlivých zápisov pretože keď ich zgrupujeme už nevieme rozlíšiť tento detail kvôli jeho menšej granularite.

Ostatné dimenzie sa zachovali, pretože stále má zmysel hovoriť o tom ktorá fakulta dané štúdium ponúka alebo z ktorého akademického roku údaje prezeráme.

2.5.2 Zázpis predmetov

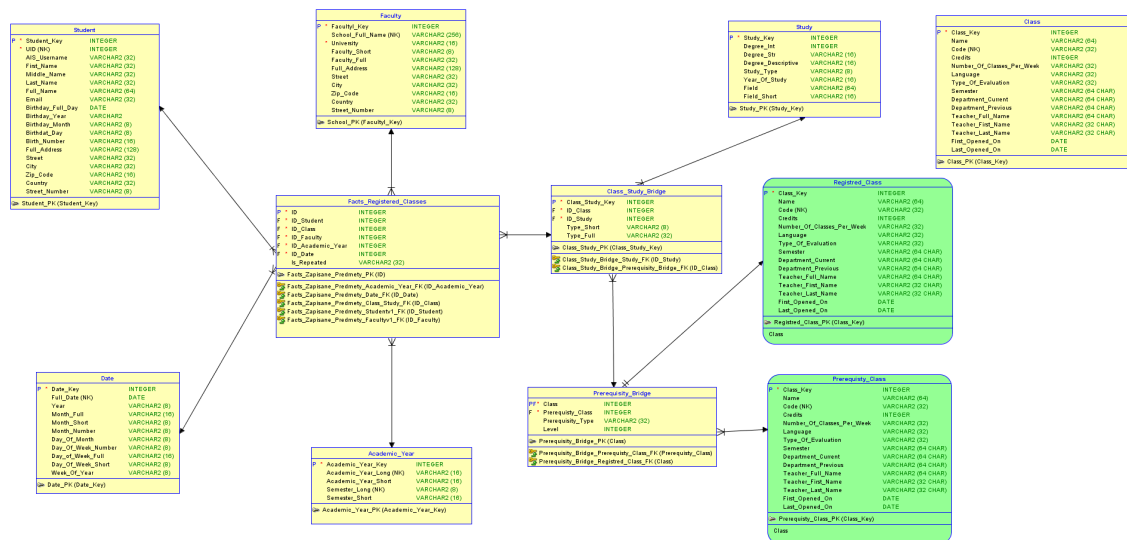
Hviezdicová schéma ktorá opisuje biznis proces zápisu jednotlivých predmetov je jedna z najkomplexnejších schém čo sa týka vzťahov medzi faktovou tabuľkou a dimenziami, ale hlavne dimenziami navzájom medzi sebou. Tieto vzťahy sa nám podarilo namodelovať do návrhu a zachytiť tak realitu procesu v dátach.

Grain faktovej tabuľky je zápis jednej inštancie predmetu. To znamená, že pre každý predmet je zaznamenaný každý osobitný zápis študentom v danom akademickom roku. Faktová tabuľka obsahuje cudzie kľúče odkazujúce na dimenzné tabuľky a navyše ešte zachytáva degenerovanú dimenziu, ktorá hovorí o tom či je daný predmet zapísaný prvý alebo druhý krát pre daného študenta v jeho štúdiu. Degenerovaná dimenzia je taká, ktorá sa logicky nehodila do žiadnej dimenznej tabuľky a preto sme ju uložili vo faktovej tabuľke. Pri zápise predmetu potrebujeme vedieť rozlíšiť jeho typ v danom štúdiu. Jeden predmet môže byť pre jeden štúdijný program povinný a pre druhý iba voliteľný. Tento vzťah medzi dvoma dimenziami sme vyriešili takzvaným bridgom alebo mostíkom. Na obrázku 2.5 je tabuľka *Class_Study_Bridge*, ktorá rieši vzťah predmetu a štúdia v ktorom sa predmet aktuálne zapisuje. Ako je možné vidieť, tak dimenzie predmetu a štúdia v tomto prípade nenapájame priamo na cudzí kľúč do faktovej tabuľky, ale miesto nich do nej pripojíme spomínanú mostíkovú tabuľku. Na mostíkovú tabuľku potom pripojíme samotné dimenzie a na základe ich kombinácie v nej vyplníme polia *Type_Full*, *Type_Short*, ktoré hovoria o vyššie zmienenom type predmetu pre dané štúdium. V našom prípade nenapájame do mostíka priamo dimenziu predmetu ale ďalší mostík, ktorého význam si vysvetlíme v ďalšom odstavci. Týmto spôsobom sme jednoducho vyriešili komplikovaný many-to-many vzťah medzi predmetom a štúdiom.

Predmety majú medzi sebou navzájom ďalší vzťah. Niektoré predmety na to aby mohli byť zapísané potrebujú mať predtým zapísané a absolvované iné predmety. Tieto nazývame prerekvizity. Prerekvizitami nám vzniká hierarchický vzťah medzi predmetmi. Hierarchický vzťah môže mať niekoľko úrovní. Predmet môže mať nula, jednu, a teoreticky až neobmedzene veľa prerekvizít. Takýto hierarchický vzťah vyriešime hierarchickým mostíkom. Na obrázku 2.5 sa tento mostík nachádza pod tabuľkou *Class_Study_Bridge* a volá sa *Prerequisite_Bridge*. Obsahuje odkaz na zapisovaný predmet a odkaz na prerekvizitu tohto predmetu. Mostík obsahuje všetky prerekvizity zapisovaného predmetu a zároveň aj úroveň prerekvizity, ktorú si pamätáme v stĺpci *Level*. Každý predmet má minimálne uloženú jednu prerekvizitu na úrovni nula a to samého seba. Takýmto spôsobom sme vyriešili hierarchiu predmetov a teda vieme si

pozrieť pre daný zapísaný predmet aj jeho prerekvizity. Pri výbere prerekvizít predmetu z databázy budeme musieť vybrať také záznamy, ktoré majú v poli mostíkovej tabuľky *Class* predmet ku ktorému pozeráme prerekvizity a v poli *Level* je číslo väčšie ako nula. Takto si vieme prezrieť všetky prerekvizity daného predmetu.

Ostatné dimenzné tabuľky, ktoré sme ešte nespomínali v schéme zapisovania predmetov sú pripojené priamo na faktovú tabuľku. Ako aj ostatné procesy týkajúce sa študentov aj tento potrebuje na plnohodnotné vyjadrenie svojich kontextov dimenzné tabuľky študenta, fakulty a už spomínaného štúdiijného odboru. Dimenzia dátumu je taktiež prítomná a opisuje dátum zapísania predmetu do zápisného listu študenta. Dimenzia akademického roka nám pomáha jednoduchšie filtrovať a agregovať záznamy pre jednotlivé semestre alebo akademické roky. Spomínaná dimenzia predmetu je v tejto schéme rozdelená do dvoch úloh. Prvou úlohou je zapisovaný predmet a druhou je prerekvizita k tomuto predmetu. Z obrázku 2.5 je vidieť že sme vytvorili dva View-y *Registered_Class* a *Prerequisite_Class*, aby sme ukázali úlohy ktoré spĺňa táto dimenzia.



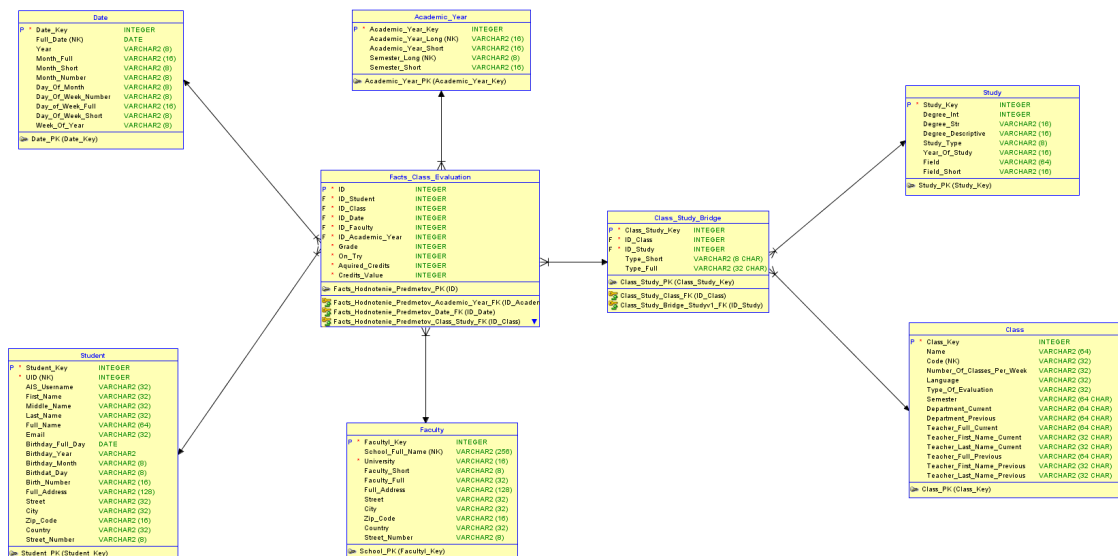
Obr. 2.5: Hviezdicová schéma - Zapisovanie predmetov

2.5.3 Hodnotenie predmetov

Schéma opisujúca proces hodnotenia je z veľkej časti zhodná so schémou o zapisovaní predmetov 2.5.2. Prvý dôvod tejto podobnosti týchto dvoch schém je čiastočná zhoda biznis procesov podľa ktorých sú namodelované aj v reálnom svete. Druhým dôvodom pre ich podobnosť bola požiadavka na reporovací systém, aby vedel vyhodnocovať údaje medzi týmito dvoma procesmi. Pre splnenie tejto požiadavky museli byť schémy vymodelované s určitou zhodou.

Faktová tabuľka obsahuje svoj primárny kľúč, a taktiež cudzie kľúče, ktoré sú odkazmi

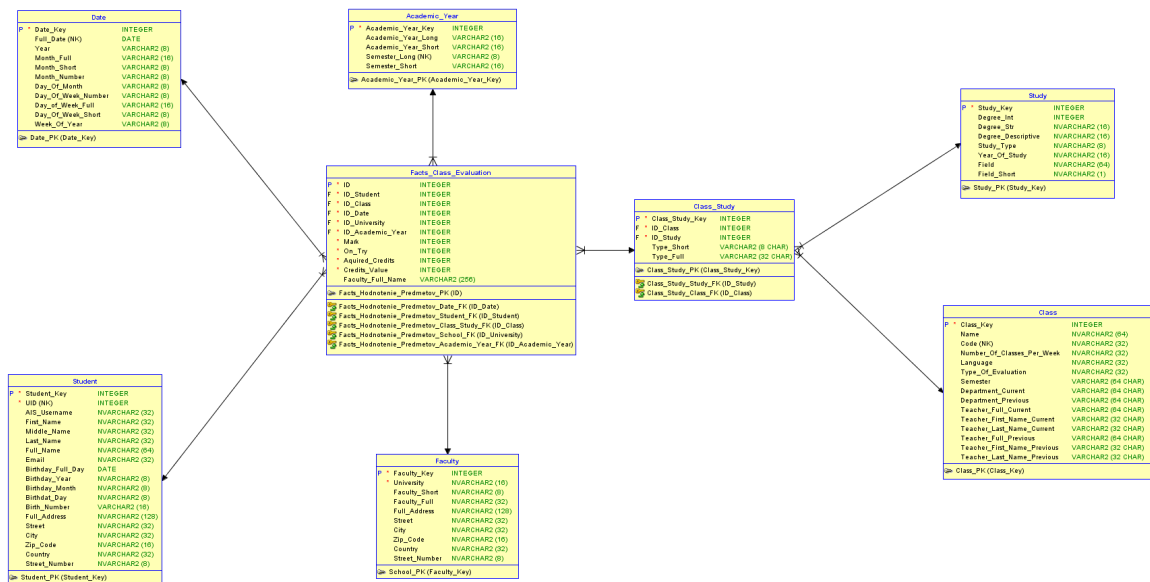
na jednotlivé pripojené dimenzie. Na spodku tabuľky *Facts_Class_Evaluation* z obrázka 2.6 sú štyri atribúty faktov. Stĺpec *Grade* nesie informáciu o pridelenej známke za skúšku. Stĺpec *On_Try* hovorí o tom na ktorí pokus študent skúšku absolvoval. Stĺpec *Aquired_Credits* je hodnota kreditov, ktoré študent za prípadne absolvovaný predmet získal. Grain je určený ako hodnotenie jedného predmetu pre jedného študenta v konkrétnom semestri. Dimenzie v tejto schéme sú skoro úplne rovnaké ako v schéme o zápise predmetov. Vidíme teda tabuľky *Student*, *Date*, *Academic_Year*, *Faculty*, *Class_Study_Bridge*, *Study* a *Class*. Riešenie many-to-many vzťahu medzi štúdiom a predmetom je rovnaké ako pri hviezdicovej schéme o zapísaných predmetoch. Rozdiel je v tom že teraz už nepotrebujeme sledovať hierarchický vzťah dimenzie predmetov a preto na mostík medzi predmetom a štúdiom môžeme napojiť obe dimenzie priamo, tak ako to vidno v pravej časti obrázka 2.6. Požiadavka na vzájomné porovnávanie dát už spomínaných procesov nás viac menej prinútila zadefinovať rovnaké dimenzie pre tieto schémy. Zodpovedajúce si dimenzie (preložené z anglického conformed dimensions) sú nasledovné *Student*, *Date*, *Academic_Year*, *Faculty*, *Class_Study_Bridge*, *Study*. To znamená, že vo všetkých týchto tabuľkách sú rovnaké stĺpce ako v jednej tak aj druhej hviezdicovej schéme. To nám potom umožňuje plniť požiadavky takých reportov ktoré by chceli porovnávať údaje z oboch týchto procesov. Pri takomto porovnávaní sme obmedzení na filterovanie a agregovanie, iba podľa týchto zhodujúcich sa atribútov. V našom prípade sú to všetky spomínané tabuľky a z tabuľky *Class* sú to všetky atribúty okrem predchádzajúceho učiteľa a prvého a posledného otvorenia predmetu. Tieto dva v podobných reportoch nemôžeme selektovať ani podľa nich filtrovať či agregovať. Takémuto krížovému dotazovaniu dát sa hovorí drilling across.



Obr. 2.6: Hviezdicová schéma - Hodnotenie predmetov

Derivovaná schéma

V popise tohoto biznis procesu v sekcii 2.3 sme spomínali možnosť obmedzenia dát z hodnotenia predmetov iba na jednotlivé fakulty. Aby sme dosiahli takúto filtráciu dát vytvoríme úplne novú hviezdicovú schému, tak ako ju vidíme na obrázku 2.7. Pri porovnaní z obrázkom 2.6 nájdeme iba jednu zmenu a to vo faktovej tabuľke. V novej schéme sa ako posledný stĺpec faktovej tabuľky nachádza *Faculty_Full_Name*, ktorý explicitne vyfiltruje záznamy podľa fakulty a poskytne používateľovi iba tie záznamy ku ktorým ma oprávnenie pristupovať. Takto vytvorená schéma sa nazýva derivovaná schéma, pretože sme ju vytvorili jednoduchým obmedzením z pôvodnej hviezdicovej schémy. Máme teda vyriešenú aj požiadavku aby systém umožnil obmedzený prístup k dátam o hodnotení predmetov kvôli ochrane údajov.



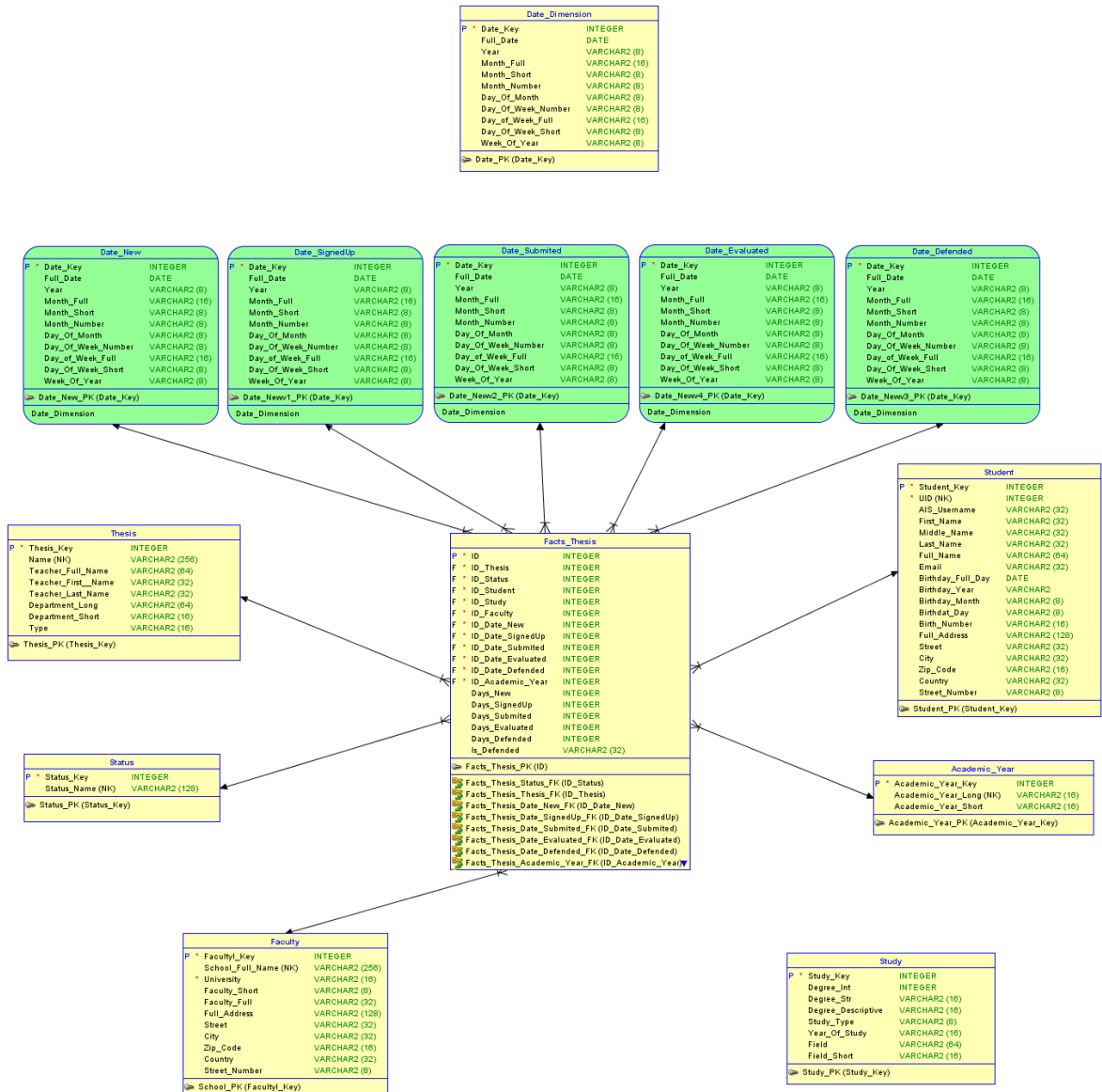
Obr. 2.7: Hviezdicová schéma - Hodnotenie predmetov pre jednotlivé fakulty

2.5.4 Záverečné práce

Úlohou hviezdicovej schémy, ktorá zaznamenáva údaje o záverečných prácach nieje iba zaznamenať dáta samotnej práce, ale hlavne jej prechod medzi stavmi v čase. Je teda navrhnutá z ohľadom na túto požiadavku. Grain faktovej tabuľky je jedna atomická záverečná práca. Požiadavka na zaznamenanie stavov by nás mohla lákať ku ešte jemnejšiemu grainu a to jeden záznam pre každý stav každej zaznamenatej práce, avšak za chvíľu si ukážeme, že toto nieje potrebné a samotná práca je atomická. Faktová tabuľka schémy obsahuje viacej stĺpcov ako ostatné faktové tabuľky nášho systému. Dôvodom sú práve spomínané zachytené stavy práce. Okrem klasických cudzích kľúčov odkazujúcich sa na dimenzie študenta, akademického roku, fakulty, štúdiijného

odboru, stavu práce a samotnej práce obsahuje faktová tabuľka aj päť odkazov na dátumovú dimenziu. Týchto odkazov je päť preto, lebo pri každej zmene stavu miesto toho, aby sme vytvorili nový riadok v tabuľke, tak vyplníme pole príslušného dátumu na kľúč hodnoty v ktorom sa zmena stavu udiala. Možných stavov práce je päť: nová vypísaná práca, schválená práca priradená študentovi, odovzdaná práca, ohodnotená práca, obhájená/neobhájená práca. Keď na začiatku do tabuľky vložíme údaj o novej práci, má vo faktovej tabuľke iba jeden odkaz na dátumovú dimenziu a to na dimenziu *Date_New*, alebo toľko záznamov koľko nám pri nahrávaní dát poskytne operačný systém. Ostatné odkazy na dátumové dimenzie majú hodnotu null. Tieto sa vyplnia vždy keď systém dostane notifikáciu o zmene stavu. Dátum začiatku stavu je teda určený v odkaze na záznam v príslušnej dimenzii a dátum konca toho istého stavu sa zhoduje zo začiatkom ďalšieho stavu v poradí, tak ako sme ich vyššie vymenovali. Okrem samotných dátumov si faktová tabuľka pamätá aj dĺžku trvania práce v každom stave. Tieto dĺžky nám umožňujú rýchly a pohodlný prístup k jednotlivým časovým intervalom presne tak, ako sme to požadovali. Takto navrhnutú faktovú tabuľku nazývame aj Accumulating Snapshot Fact Table (slovenský preklad nieje dostatočne vystihujúci - viac o takýchto tabuľkách sa dá dočítať v knihe Christophera Adamsona [1]) pretože nahromaďuje údaje o jednom zázname, tým že nám o ňom postupne poskytuje viac a viac údajov. Navyše faktová tabuľka obsahuje jednu degenerovanú dimenziu a to fakt či bola práca obhájená alebo nie.

Keď sa zaujímate o entity akými sú záverečné práce tak nám nasledujúce kontexty uložené v dimenzných tabuľkách poskytujú veľmi rozumné a žiadané atribúty podľa ktorých budeme často v reportoch agregovať a filtrovať: fakulta študenta, štúdiálny odbor študenta, meno vedúceho, typ práce a teda či je bakalárska, magisterská alebo dizertačná, akademický rok v ktorom sa odohrávali jej zmeny, aktuálny status práce, meno práce, katedra pod ktorou je práca uvádzaná, samotné údaje o študentovi. Na obrázku 2.8 sú zelenou farbou vyznačené Views *Date_New*, *Date_SignedUp*, *Date_Submitted*, *Date_Evaluated*, *Date_Defended*, ktoré sú vygenerované z dimenznej tabuľky *Date*, ktorá nieje v tomto prípade pripojená na faktovú tabuľku. Tieto Views sú zadefinované úplne rovnako ako tabuľka *Date*, ale každý sa volá inak podľa príslušajúceho stavu. Samotná tabuľka *Date* je takto rozdelená aby sme ukázali to ako jedna dimenzia zastáva rôzne úlohy v tej istej hviezdicovej schéme. Dimenzia *Status* uchováva iba meno aktuálneho stavu čím je tento stav aj definovaný. Odkaz vo faktovej tabuľke na túto dimenziu sa zmení vždy pri zmene stavu práce a obnovení nového dátumu. Toto nám uľahčuje samotný prístup k tomuto údaju. Miesto toho aby sme pozerali ktorý je posledný nenulový odkaz na dimenziu dátumu a podľa toho rozhodli o stave, radšej si ho uložíme dopredu a skrátime dotazovací čas.



Obr. 2.8: Hviezdicová schéma - Záverečné práce

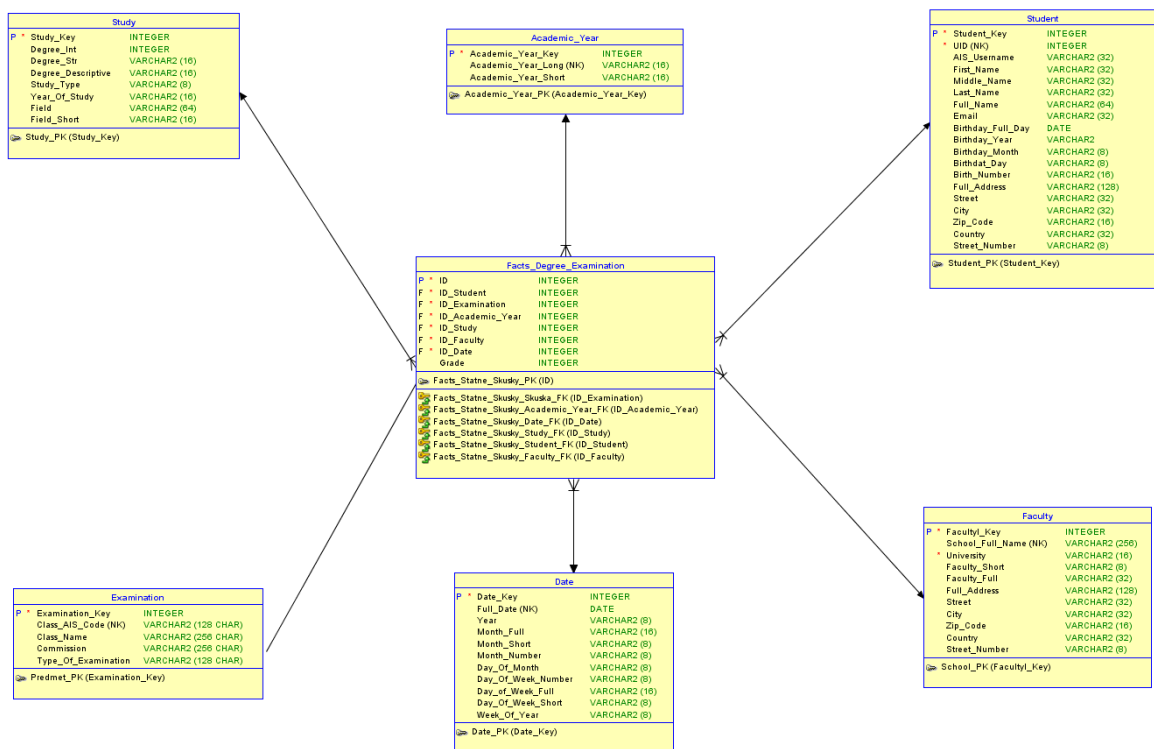
2.5.5 Štátne skúšky

Logický model ktorý zachytáva priebeh štátnych skúšok ja obsiahnutý v štyroch hviezdicových schémach. Ako sme spomínali pri popise tohoto biznis procesu a popise jednotlivých dimenzií reportovacieho systému, poznáme tri typy štátnych skúšok ktoré majú niektoré spoločné a niktore odlišné atribúty. Na základe týchto typov sme vytvorili spomenuté štyri schémy.

Všeobecná schéma

Prvá schéma je všeobecná schéma (z anglického core star) pre štátnu skúšku, ktorá nemá bližšie špecifikovaný žiadny typ. Táto obsahuje spoločné ústredné dimenzie, ktoré sú spoločné pre všetky typy štátnych skúšok a všeobecne sa nazýva ústredná schéma.

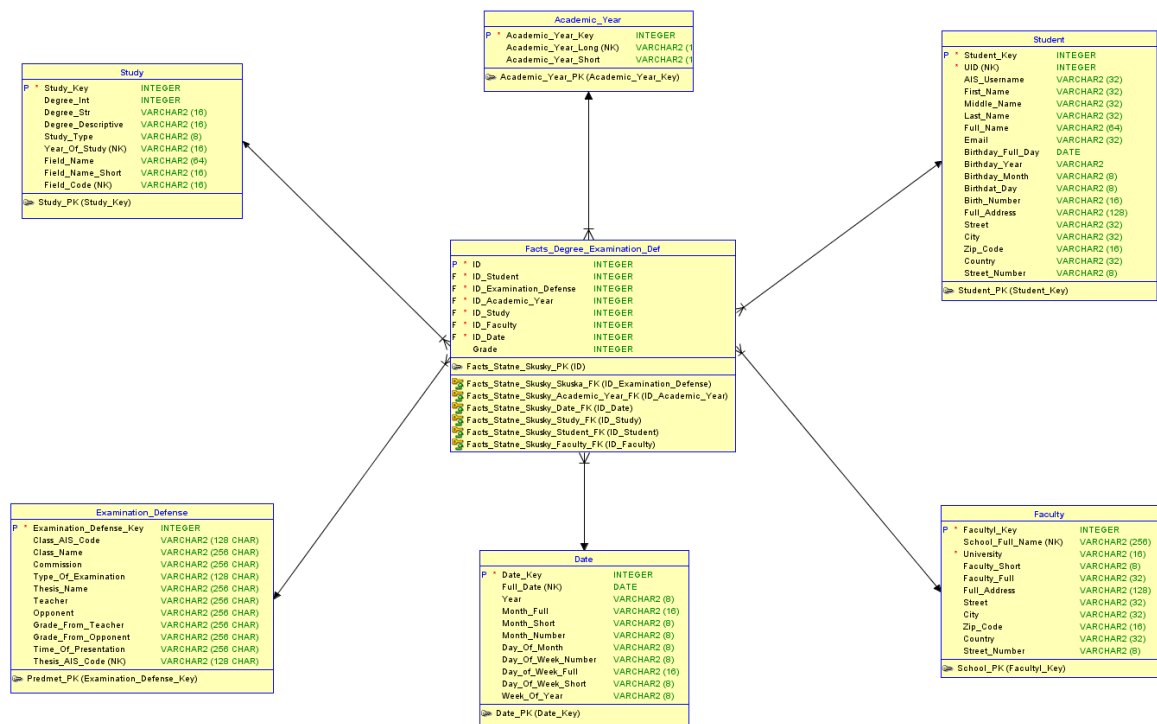
Podrobnejšie si opíšeme jej komponenty, ako ich vidíme na obrázku 2.9. Faktová tabuľka obsahuje štandardný primárny kľúč a cudzie kľúče, ktoré na ňu napájajú a dimenzné tabuľky. Posledným stĺpcom v tejto tabuľke je známka ktorú študent dostal. Grain schémy je určený ako jedna štátna skúška, ktorú absolvoval študent v nejakom čase. Definovaný je teda tak aby sme ho nevedeli zjemniť, ale zároveň vieme agregovať na väčšie skupiny podľa pripojených dimenzií. Dimenzné tabuľky ktoré sú pripojené k faktovej tabuľke sú rovnaké ako v predchádzajúcich schémach, pretože aj v tomto procese sú dôležité detaily týkajúce sa študenta, ktorý je skúšaný a tomu zodpovedá dimenzná tabuľka *Student*. Rovnako chceme vedieť aké štúdium je skúškou ukončované a preto schéma obsahuje tabuľku *Study*, ktorá definuje detaily nutné pre identifikovanie štúdia. Časový kontext udávajú dimenzie *Date* a *Academic_Year*. Všetky tieto dimenzné tabuľky sa zhodujú s tabuľkami v predchádzajúcich sekciách. Dimenzná tabuľka *Examination* obsahuje detaily o samotnej štátnej skúške. Polia *Class_AIS_CODE(NK)*, *Class_Name*, *Commission* a *Type_Of_Examination*, nám nesú údaje o: kóde predmetu štátnej skúšky ktorý pochádza z AIS-u, názve predmetu štátnej skúšky, komisii prítomnej pri štátnej skúške a type štátnej skúšky v poradí v akom sú uvedené. Toto sú všetky komponenty všeobecnej schémy, ktoré sú spoločné aj pre nasledujúce špecifické schémy (z anglického custom stars).



Obr. 2.9: Hviezdicová schéma - Štátne skúšky

Obhajoba záverečnej práce

Hviezda ktorá popisuje špecifický typ štátnej skúšky je priamo vyrobená zo všeobecnej schémy štátnice. Granularita faktovej tabuľky a jej polia sú úplne zhodné so všeobecnou schémou. Vzťahy ku dimenziám štúdia, fakulty, študenta, dátumu a akademickému roku sú tiež úplne rovnaké ako aj samotné dimenzné tabuľky, ktoré tieto vzťahy s faktovou tabuľkou vytvárajú. Rozdiel nastáva v dimenzii skúšky, ktorý môžeme vidieť na obrázku 2.10. Dimenzná tabuľka *Examination_Defense* obsahuje oproti všeobecnej tabuľke štátnice navyše atribúty: meno obhajovanej práce - *Thesis_Name*, školiteľa - *Teacher*, oponenta - *Opponent*, známku od učiteľa a oponenta - *Grade_From_Teacher*, *Grade_From_Opponent*, čas trvania obhajoby - *Time_Of_Presentation* a kód práce z AIS-u - *Thesis_AIS_Code* (NK). Tieto všetky dimenzie sú špecifické pre obhajobu práce.



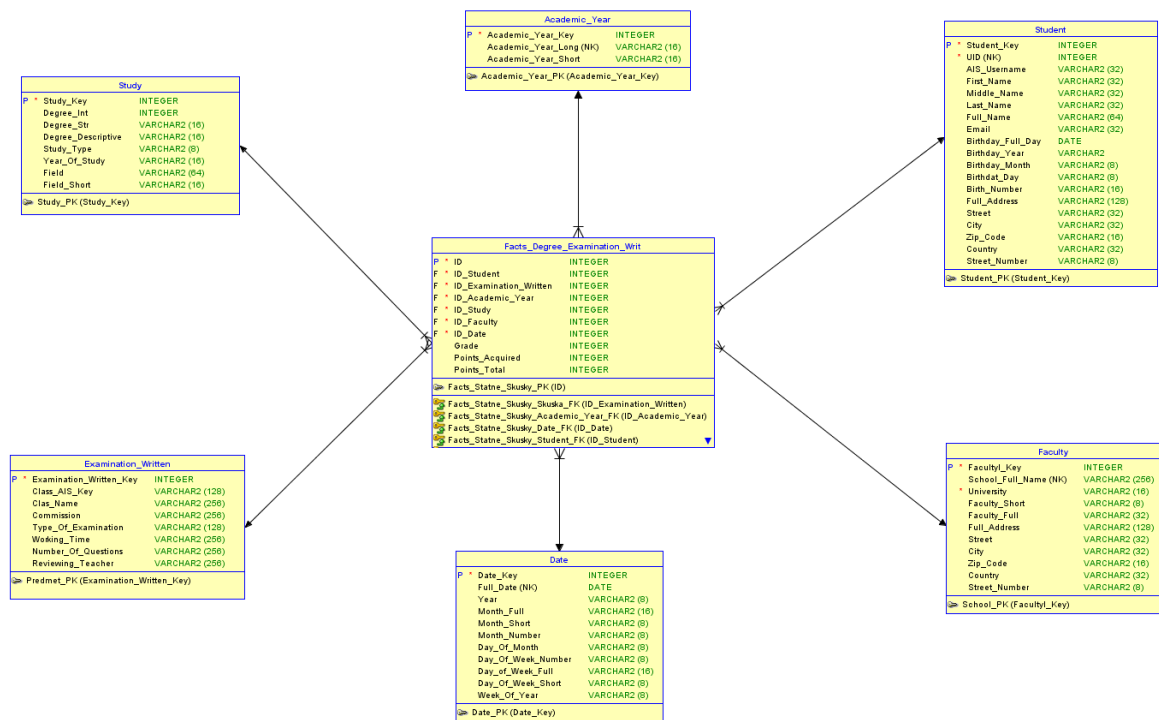
Obr. 2.10: Hviezdicová schéma - Obhajoba záverečnej práce

Ústna skúška

Druhou špecifickou schémou je hviezdička pre ústnu skúšku. Rovnako ako predchádzajúca schéma sa zhoduje v celej faktovej tabuľke a všetkých dimenzných tabuľkách okrem tabuľky *Examination_Verbal*. Táto obsahuje oproti všeobecnej dimenznej tabuľke atribúty téma ústnej skúšky - *Topic* a skúšajúceho učiteľa *Examining_Teacher*.

Písomná skúška

Posledná špecifická hviezdica sa líši aj vo faktovej tabuľke. Oproti všeobecnej obsahuje navyše dva stĺpce ktoré hovoria o tom koľko bodov študent mohol dosiahnuť z písomky a koľko ich skutočne získal - *Points_Total*, *Points_Acquired*. Ako predchádzajúce typy skúšok aj písomná definuje oproti všeobecnej navyše tri dimenzie, ktoré sú pre ňu charakteristické. V tabuľke *Examination_Written* možno vidieť na obrázku 2.11 stĺpce *Working_Time*, *Number_Of_Questions* a *Reviewing_Teacher*, ktoré popisujú koľko má študent času na vypracovanie písomky, koľko otázok písomka obsahuje a kto ju opravoval.



Obr. 2.11: Hviezdicová schéma - Písomná štátna skúška

Pomocou všeobecnej schémy a špeciálnych schém sa nám podarilo namodelovať biznis proces štátnej skúšky. Schémy sú definované tak, aby sme vedeli rozlíšiť medzi jednotlivými typmi štátnych skúšok a zároveň vedeli reportovať dáta generované týmto biznis procesom s možnosťou interpretovať štátnu skúšku ako všeobecnú bez ohľadu na jej skutočný typ podľa potreby.

2.5.6 Využitie miestností

Hviezdica na obrázku 2.12 popisuje využívanosť miestností v čase ako aj jej kapacitnú vyťaženosť. Grain faktovej tabuľky *Facts_Room_Utilization* sme určili ako jeden nepretržitý záznam udalosti v miestnosti. Pod nepretržitým záznamom rozumieme uda-

losť, ktorá sa v miestnosti odohráva bez možnosti, aby ju prerušila nejaká iná udalosť. Udalosti môžu byť rôzne, či už sú to predmety v posluchárňach, pracovná doba zamestnanca v jeho kancelárii alebo stála obsadenosť skladu. Jednotlivé typy udalostí nerozlišujeme a zaznamenávame iba či je miestnosť obsadená, nakoľko pri vyhodnocovaní nás nezaujíma akým spôsobom je miestnosť obsadená, ale ako dlho je obsadená. Tiež sa budeme zaujímať ako dobre je miestnosť využitá kapacitou. Tieto údaje sme schopný ukladať v atribútoch tabuľky *Start_Time*, *End_Time*, a *Duration*, ktoré sa starajú o časové využitie miestnosti. O kapacitné využitie sa starajú atribúty *Capacity_Of_Room* a *Number_Of_Persons* z ktorých sa dá jednoduchým spôsobom vypočítať aj percento tohoto údaju.

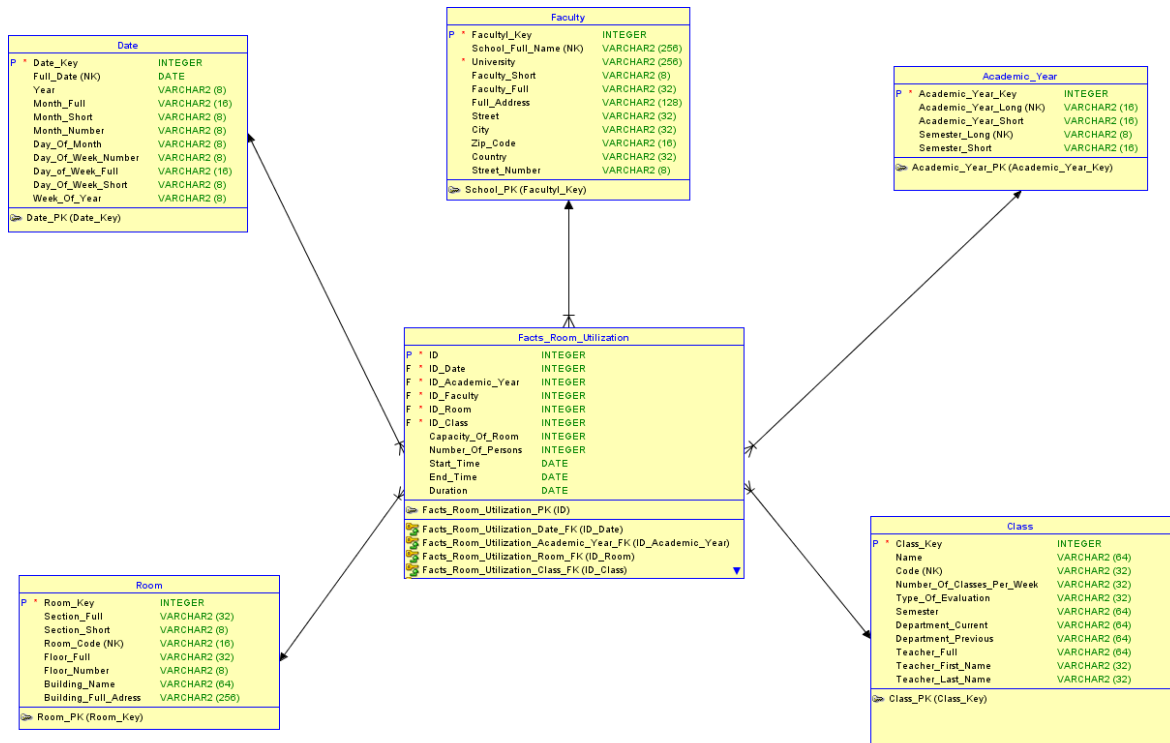
Room je najdôležitejšou dimenznou tabuľkou tejto schémy pretože v nej sú uložené údaje o miestnosti v ktorej sa zaznamenaná udalosť stala. Atribúty tejto tabuľky tvoria hierarchiu, ktorá umožňuje dobrý spôsob filtrovania a agregovania. Obsahuje totiž stĺpce o budove, sekcii budovy, poschodí a samotnej miestnosti. Podľa týchto atribútov potom vieme na údaje pozeráť v nami požadovanej úrovni detailu. Tabuľka *Class* nám navyše poskytuje detailnejšie informácie aký predmet sa odohráva v miestnosti. Odkaz z faktovej tabuľky do tejto dimenzie môže byť aj prázdny resp. neexistujúci a to v prípade, že sa nejedná o udalosť vyučovania predmetu. Detaily o fakulte na ktorej je záznam uskutočnený sú uložené v tabuľke *Faculty*. Dáta o dni udalosti ukladáme do dimenznej tabuľky *Date*. Údaj o semestri a akademickom roku v ktorom je udalosť zaznamenaná sa uloží v tabuľke *Academic_Year*.

Schéma je navrhnutá tak aby sa dali jednoducho vytvárať reporty časového a kapacitného využitia miestností na fakulte, pričom máme možnosť ľahko prejsť na vyššiu úroveň granularity pomocou dobre navrhnutej dimenznej tabuľky *Room* a filtrovať údaje na základe spomínaných dimenzií.

2.5.7 Rozvrh

Malá schéma ktorá popisuje rozvrhy má vo svojej faktovej tabuľke *Facts_Schedule* cudzie kľúče na päť dimenzných tabuliek. Okrem nich obsahuje ďalšie dva stĺpce ktoré poskytujú časové fakty o začiatočnom a konečnom čase predmetu - *Start_Time* a *End_Time*. Grain tabuľky je zadaný ako jeden vyučovaný predmet v miestnosti počas určeného času. V rovnakom čase môžu byť na rôznych miestach vyučované viaceré rôzne predmety.

Class je tabuľka, ktorá obsahuje dimenzie popisujúce predmet, ktorý sa v danú hodinu vyučuje. Obsahuje informácie o názve a kóde predmetu, počtu hodín vyučovaných v jednom týždni, katedre ktorá predmet vedie a konkrétnom vedúcom predmetu. Lokácia záznamu v schéme je udržiavaná v tabuľke *Room*, ktorá popisuje miestnosť v ktorej sa vyučovanie odohráva. Na obrázku 2.13 si môžeme všimnúť, že poskytuje informácie

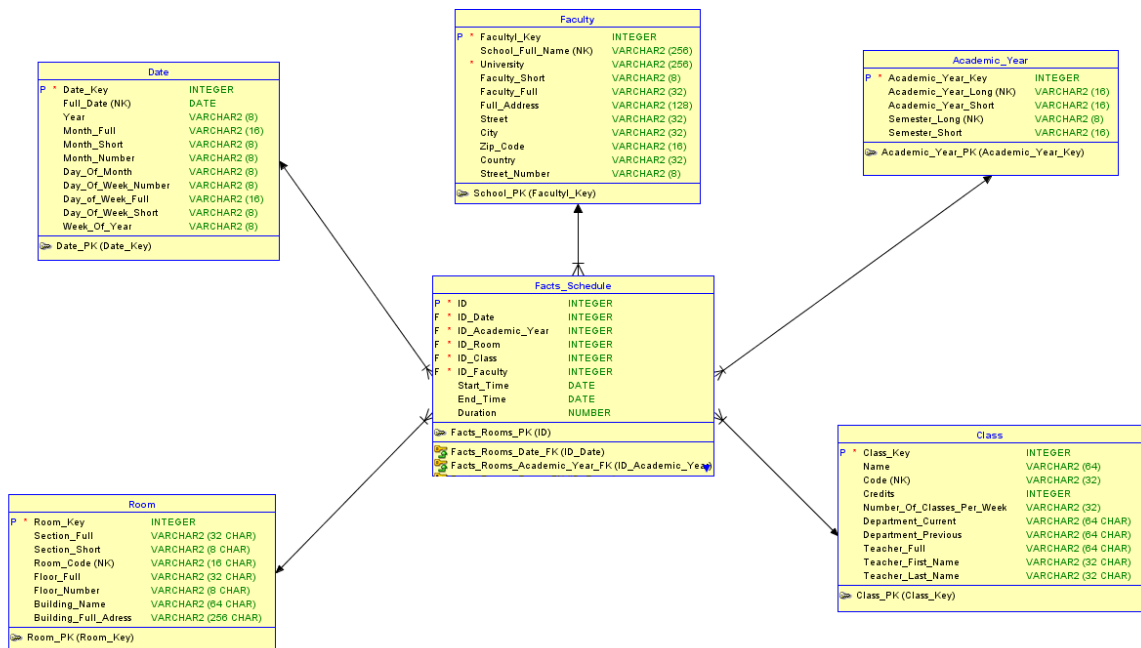


Obr. 2.12: Hviezdicová schéma - Využitie miestností

na úrovni budovy, sekcie budovy, poschodia a konkrétnej miestnosti. Tieto nám potom umožňujú filtrovanie a agregovanie hierarchiu, rovnakú ako v predchádzajúcej schéme. Schéma tiež obsahuje dimenzné tabuľky určujúce na ktorej fakulte je daný predmet vyučovaný a časové údaje o akademickom roku a semestri v ktorom sa záznam odohráva.

2.5.8 Technické zariadenia

Schéma navrhnutá pre meranie využitia jednotlivých technických zariadení na fakulte je štruktúrou veľmi podobná hviezdici navrhnutej v sekcii o využívaní miestností 2.5.6. Faktová tabuľka *Facts_Technical_Equipment* má podobnú štruktúru. Je to spôsobené tým, že v oboch prípadoch meriame čas používania nami zadaného objektu. Obsahuje teda rovnaké atribúty na meranie času: *Start_Time*, *End_Time*, a *Duration*. Tieto však zachytávajú inú udalosť a to používanie konkrétneho technického zariadenia používateľom. Toto určuje aj grain tabuľky: jedno nepretržité použitie zariadenia nejakým človekom. Dimenzia ktorá uchováva údaje o zariadení, ktoré je používané sa volá *Equipment*. Obsahuje názov a typ zariadenia rovnako ako jeho kľúč z AIS-U. V tabuľke *Room* ukladáme údaje o miestnosti v ktorej sa používané zariadenie nachádza. Tabuľka obsahuje rovnaké stĺpce ako v predchádzajúcich dvoch schémach. *User* je dôležitou dimenznou tabuľkou, ktorá obsahuje údaje o používateľovi zariadenia. Obsa-



Obr. 2.13: Hviezdicová schéma - Rozvrh

huje stĺpce *UID (NK)*, *AIS_Username* - údaje o používateľovi z AIS-u, *First_Name*, *Middle_Name*, *Last_Name Full_Name* - atribúty mena a *Email* - emailová adresa. Konkrétny deň používania zariadenia je uložený v dimenzii *Date*. Dáta o fakulte ktorej zariadenie patrí sú uložené v dimenznej tabuľke *Faculty*, ktorá má rovnakú štruktúru atribútov ako vo všetkých predchádzajúcich schémach.

2.5.9 Prihlášky

V nasledujúcej hviezdicovej schéme sme sa snažili zachytiť proces prvého kroku, ktorý treba vykonať, aby sa študent mohol dostať na univerzitu. Aktivita ktorú popisuje je podávanie prihlášok na vysokú školu. Mnohé dáta ktoré táto hviezdica zachytáva sú štruktúrou prebraté z tých, ktoré sa nachádzajú v prihláške. Faktová tabuľka *Facts_Application*, ktorá sa nachádza v centre obrázku 2.15, obsahuje sedem cudzích kľúčov ukazujúcich na dimenzné tabuľky, ktoré neskôr detailnejšie popíšeme. V jej spondej časti sa nachádzajú fakty tejto schémy. *First_Decision*, ktorý hovorí či bol študent prijatý na štúdium, alebo potrebuje ďalšie preskúšanie. Stĺpce *Acquired_Points*, *Percent*, *Total_Possible_Points* sú v prípade, že bol študent prijatý bez ďalšieho skúšania prázdne, respektíve obsahuje hodnotu null. Ak študent potreboval ďalšie preskúšanie tieto fakty hovoria o počte dosiahnutých bodov z prijímacích skúšok v absolútnej hodnote, percentách a celkovom počte možných bodov. Posledným faktom je *Final_Decision*, ktorý hovorí o tom či bol uchádzač prijatý alebo nie. Ak bolo prvé rozhodnutie o prijatí kladné bez potreby ďalšej skúšky, musí byť kladné aj konečné rozhodnutie. Grain fak-

tovej tabuľky je definovaný jednou prihláškou študenta na jeden štúdijný program v danom roku. Rovnaký človek môže podať aj viacej prihlášok v tom istom roku. V tomto prípade bude hviezdicová schéma obsahovať toľko záznamov pre rovnakého uchádzača koľko prihlášok podal

Z prihlášky sme prebrali dáta o uchádzačovi ktoré sme považovali za potrebné a umiestnili sme ich do dimenznej tabuľky *Potential_Student*. Konkrétne obsahuje rodné číslo, ktoré je aj prirodzeným kľúčom, meno uchádzača a adresu trvalého bydliska. Ďalej sme z prihlášky vzali údaje o najvyššom dosiahnutom vzdelaní študenta a zoskupili ich do tabuľky *Highest_Acquired_Education*. Dimenzie v tejto tabuľke nesú informáciu o type školy na ktorej bolo vzdelanie dosiahnuté, stupni vzdelania a v akom odbore bolo získané. Údaje uložené v tabuľke *Previous_School* nám poskytujú bohaté informácie o tom z akých škôl na jednotlivé odbory a fakulty prichádza najviac ľudí. Taktiež dáta v tejto tabuľke môžeme interpretovať z geografického hľadiska. Tabuľky *Desired_Study a Faculty* sme už viac krát v našom dizajne použili, no tentokrát majú iný kontextuálny význam ako zvyčajne. Určujú iba štúdium na fakulte a univerzite na ktoré sa uchádzač hlási a teda nie je jeho poslucháčom ako vo väčšine prípadov. Atribúty týchto tabuliek sú však rovnaké. Časové dimenzie sú zaznamenávané v tabuľkách *Date a Academic_Year*, ktoré zachovávajú údaje o dátume podania prihlášky a v ktorom akademickom roku sa tak udialo.

V tejto schéme si ukážeme aj príklad reportu ktorý využíva agregáciu, aby sme videli použitie definovaných dimenzií aj v praxi. Budeme sa zaujímať o počty prihlášok na Fakultu matematiky fyziky a informatiky, pričom nás bude zaujímať z ktorej strednej školy bolo najviac uchádzačov o štúdium v celej histórii nášho dátového skladu. Do taz na našu schému by potom vyzeral ako na obrázku 2.14. Ako je ľahko vidieť školy sme grupovali podľa ich prirodzeného kľúča z tabuľky *Previous_School*. V prípade tejto schémy a hviezdice opisujúcej zapisovanie študentov do ročníka by sme vedeli robiť zaujímavé drilling across reporty, ktoré by vyhodnocovali koľko percent celkových alebo prijatých uchádzačov sa aj zapíše na štúdium. Zodpovedajúce si dimenzie na ktorých by sme vedeli takéto reporty spájať by zahŕňali dimenzie uchádzača a zapísaného študenta, fakulty, štúdia a akademického roku.

2.5.10 Publikácie

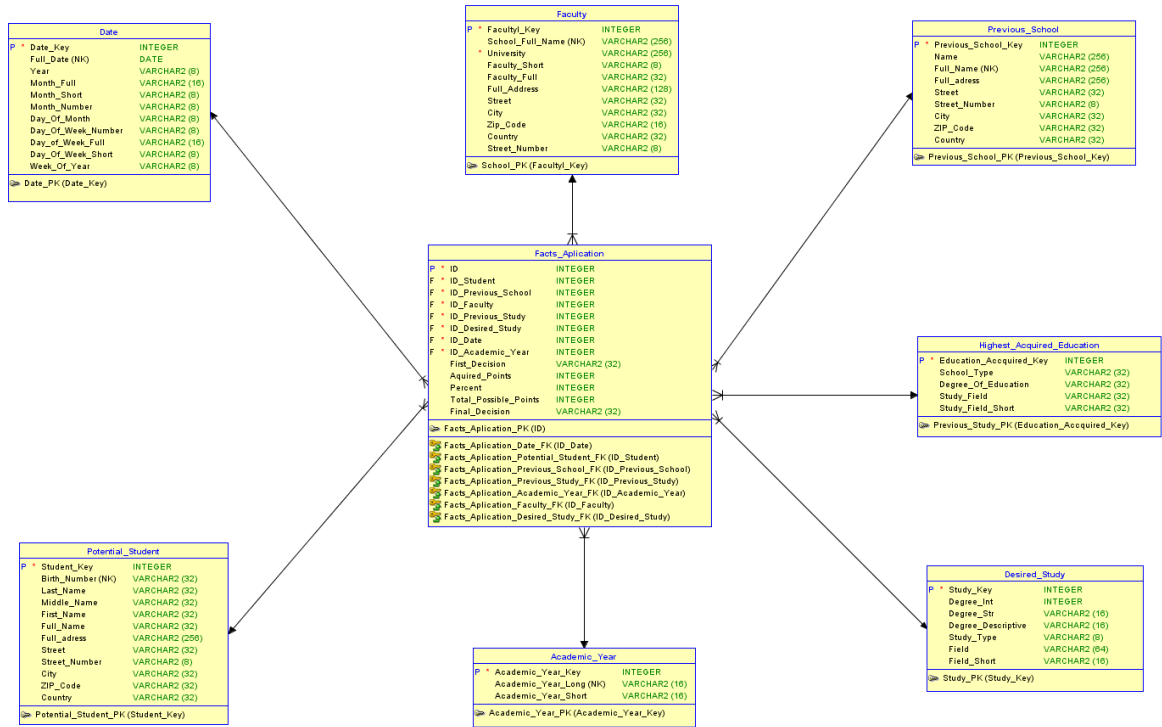
Pre záznamy o publikačnej činnosti študentov sme vytvorili schému s ústrednou faktovou tabuľkou *Facts_Publications* a siedmimi dimenznými tabuľkami, tak ako to môžeme vidieť na obrázku 2.16. Vo faktovej tabuľke si okrem cudzích kľúčov odkazujúcich sa na jednotlivé dimenzie budeme ukladať v poli *Credits_Acquired* aj počet kreditov ktoré študent dosiahol za publikovanie práce. Grain definovaný pre túto tabuľku je na úrovni jednej konkrétnej práce, ktorú študent publikoval či už v časopise alebo ako

```
SELECT p."Full_Name (NK)",
       COUNT(p."Full_Name (NK)") AS "Počet"
FROM Facts_Application f
LEFT JOIN Highest_Acquired_Education h
ON f.ID_Previous_Study = h.Education_Acquired_Key
LEFT JOIN Previous_School p
ON f.ID_Previous_School = p.Previous_School_Key
WHERE h.School_Type = 'stredná škola'
GROUP BY p."Full_Name (NK)"
ORDER BY COUNT(p."Full_Name (NK)") DESC
```

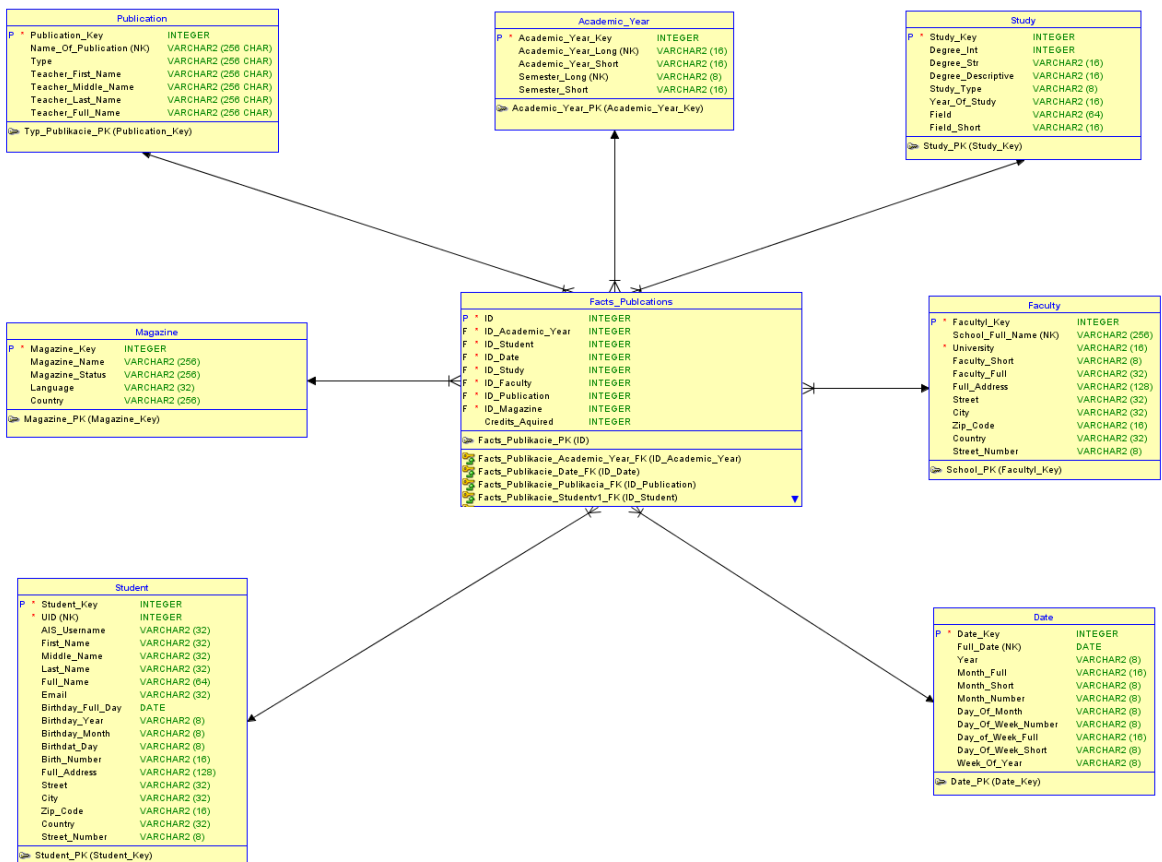
Obr. 2.14: Dotaz - Prihlášky na vysokú školu.

ŠVOČ. Ak je autorov článku viacej budeme mať o tomto konkrétnom článku toľko záznamov koľko má autorov. V tomto prípade budú cudzie kľúče *ID_Student* rôzne, ale odkazy na dimenzie publikácie rovnaké.

Dimenzné tabuľky *Study* a *Faculty* sú zaradené aj do tejto schémy. Sú pripojené na faktovú tabuľku pomocou svojho primárneho kľúča. Dôvodom ich prítomnosti sú details, ktoré nám ohľadom publikácie vedia poskytnúť: štúdium na ktorom je študent odovzdávajúci publikáciu zapísaný, v ktorom sa nachádza ročníku, ktorá fakulta takéto štúdium poskytuje. Zároveň nám umožňujú obmedzovať podľa týchto dimenzií vyberané publikácie. Aby sme samotné publikácie vedeli od seba rozlíšiť vytvorili sme dimenzú tabuľku *Publication*, ktorá nesie podrobnejšie informácie o samotnej publikácii. V zadaných atribútoch tejto tabuľky ukladáme názov publikácie, jej typ a meno vedúceho. Ďalšími dôležitými informáciami sú details ohľadne prípadného časopisu, kde bola práca publikovaná. Tieto sú zhromaždené v tabuľke *Magazine*, ktorá sa na obrázku 2.16 nachádza v ľavej časti. Ako môžeme vidieť obsahuje polia ktoré hovoria o názve časopisu, type, jazyku v ktorom časopis vychádza a krajne pôvodu. Všeobecnejšie dimenzné tabuľky ktoré sme už niekoľkokrát videli v logickom návrhu nášho systému a zaradili sme ich aj do tejto schémy sú *Student*, *Date*, a *Academic_Year*. Tieto nám poskytujú details ohľadom študenta ktorý článok vypracoval, kedy sa tak udialo a aj pohodlnejšie a predvypočítané údaje o tom v ktorom akademickom roku. Návrh tejto schémy je pomerne priamočiary a bezproblémový. V zadanom biznis procese nie sú žiadne zložité vzťahy a závislosti, ktoré treba riešiť. Hlavné body pri modelovaní tejto hviezdice boli určenie dobrej granularita faktovej tabuľky a správny výber dimenzií, ktoré sme popísali.



Obr. 2.15: Hviezdicová schéma - Príhlášky na vysokú školu.



Obr. 2.16: Hviezdicová schéma - Publikácie

Kapitola 3

Fyzický model

V tejto kapitole sa budeme venovať opísaniu technológií ktoré sme si vybrali a použili pre náš dátový sklad. Ďalej ukážeme detaily fyzických objektov, ktoré sme implementovali na základe logického modelu navrhnutého v predchádzajúcej kapitole. Popíšeme implementáciu indexov jednotlivých tabuliek a dôvody prečo sme si zvolili práve tieto indexy.

3.1 Implementácia fyzického modelu

Fyzický model reportovacieho systému sme sa rozhodli implementovať na technológiách od spoločnosti Oracle. Použili sme predinštalovaný databázový server Oracle 12c vo Virtual Boxe [5]. Pre jednotlivé entity z logického modelu sme vytvorili fyzické objekty pričom sme zachovali všetky závislosti. Pre každú star schému sme vymysleli dvoj alebo troj písmenovú predponu, ktorá vo fyzickej databáze identifikuje všetky jej objekty. Pod takýmito objektami rozumieme tabuľky, views, indexy a constrain-ty. Potom sme vytvorili faktovú tabuľku a dimenzné tabuľky tak ako boli zadané v logickom modeli so všetkými závislosťami medzi nimi, ktoré sme definovali v logickom modeli. Primárne kľúče tabuliek sa takisto zhodujú s návrhom. Takisto sme vygenerovali cudzie kľúče medzi tabuľkami ktoré spolu súvisia podľa logického modelu. Názvy entít sa zmenili pre jednotlivé schémy podľa dopredu vybranej identifikačnej predpony. Uvádzame zoznam týchto prefixov, aby sa dalo novému pozorovateľovi ľahšie orientovať vo vytvorených objektoch.

- Zápis študentov do ročníka - SR
- Agregovaná schéma zápisu študentov - AS
- Zápis predmetov - CR
- Hodnotenie predmetov - CE

- Záverečné práce - TH
- Štátne skúšky - DE
- Obhajoba záverečnej prác - DED
- Ústna skúška - DEV
- Písomná skúška - DEW
- Využitie miestností - RU
- Rozvrh - SCH
- Technické zariadenia - TE
- Prihlášky - APP
- Publikácie - PUB

Pri implementácii sme vynechali snehovú vložku zápisu študentov do ročníka, ktorú sme modelovali iba ako ukážku takéhoto návrhu. Pre schému štátnych skúšok sme vytvorili všetky tabuľky všeobecnej schémy, avšak pri implementácii špecifických schém sme vytvorili každej z nich iba faktovú tabuľku a špecifickú dimenziu skúšky (ústnej, písomnej a obhajoby). Následne sme v každej faktovej tabuľke špeciálnej schémy vygenerovali cudzie kľúče na príslušnú špeciálnu dimenznú tabuľku a ostatné kľúče sme pripojili na všeobecné dimenzie, ktoré boli zhodné pre každú špeciálnu schému. Taktiež sme vytvorili dva príkladné reporty uvedené v kapitole 2. Implementovali sme ich ako dva separátne view, pričom sme rešpektovali konvenciu tvorby mien uvedenú vyššie.

3.2 Indexy

Dôležitou časťou implementácie fyzických objektov bolo zdefinovať a vytvoriť správne indexy, aby sme zabezpečili čo najrýchlejší beh dotazov a teda čo najlepší chod celého systému. Moderné databázové serveri, medzi nimi aj Oracle, používajú takzvaný Query Planner, ktorý rozhoduje o tom, akým spôsobom sa bude daný dotaz alebo sekcia príkazov na serveri vykonávať. Jednou z hlavných úloh tohto plánovača je vybrať správne algoritmy JOIN operácií. Aby sme zabezpečili najlepší výber algoritmu pre naše dotazy vyrobili sme indexy na všetkých poliach na ktorých predpokladáme JOIN operácie. Tieto operácie predpokladáme na všetkých primárnych kľúčoch a preto majú primárne kľúče vyrobené indexy. Rovnako predpokladáme JOIN operácie na všetkých cudzích kľúčoch v tabuľkách. V hviezdicových schémach sa väčšinou všetky cudzie kľúče nachádzajú vo faktovej tabuľke. V našom prípade sa nachádzajú ešte v mostíkových tabuľkách. Preto sme vytvorili indexy aj na všetkých cudzích kľúčoch.

Ďalšími poliami na ktorých treba zväžiť vytvorenie indexov sú tie na ktorých predpokladáme veľa filtrovania alebo zoradovania. Pre jednotlivé dimenzné tabuľky ukážeme na ktorých poliach sme navrhli indexovanie, lebo na nich predpokladáme a zároveň aj odporúčame vykonávať bežné dotazovacie operácie. V dimenzných tabuľkách ktoré implmentujú návrh dimenzie študenta sme vytvorili okrem indexu na primárnom kľúči dva indexy. Prvý na poli Last_Name a druhý na poli Birth_Number, pretože predpokladáme, že študenti budú vyhľadávaný najmä podľa týchto atribútov. Triedenie študentov dáva takisto najväčší zmysel podľa priezviska. V dimenzných tabuľkách vytvorených na základe dimenzie fakulty sme vytvorili takisto dva indexy okrem toho na primárnom kľúči. Na poli Faculty_Full, ktoré je celým menom fakulty a School_Full_Name (NK), ktoré obsahuje meno univerzity a fakulty spolu. Rovnako ako pri študentovi, predpokladáme na týchto poliach filtrovanie a zoradovanie. Pre tabuľky implementujúce návrh dimenzie Class bol vytvorený index na poli Code(NK), čo je unikátny kód predmetu. Pre tabuľky miestností je index vytvorený na poli Room_Code, ktoré je skratkou sekcie budovy, poschodia a čísla dverí na poschodí a teda je unikátne. Preto aj predpokladáme, že sa bude podľa neho filtrovať. Pre tabuľku záverečnej práce bol index navrhnutý na pole s názvom - Name(NK). Pre tabuľku publikácie bol implementovaný jeden index na poli Name_Of_Publication(NK), kde predpokladáme filtrovanie podľa mena. V dimenznej tabuľke technického vybavenia bol vytvorený jeden index na poli Name, kde takisto predpokladáme filtrovanie na mene. V tabuľkách štátnych skúšok bol v každom jednom type skúšky implementovaný jeden index na poli ktoré obsahuje kód predmetu štátnej skúšky.

Toto sú navrhnuté a implementované indexy vo fyzickom modeli. Prílišná tvorba indexov často spôsobujú spomaľovanie INSERT, DELETE a UPDATE operácií na poliach s indexami. Náš systém je reportovací a na rozdiel od operačného používa prevažne dotazovacie príkazy a preto nemá implementácia indexov značné nevýhody.

V prílohe ku práci je možné nájsť aj skripty opisanej databázy (meno súboru je Skripty - Fyzicky Model.sql). Tieto skripty sa dajú spustiť na Oracle Servri verzie 11 a 12.

Záver

Hlavným cieľom práce bolo vytvoriť logický model pre dátový sklad reportovacieho systému. V prvej kapitole sme opísali problematiku a zadefinovali pojmy, ktoré boli nutné na pochopenie ako dimenzionálne modelovanie funguje a čo to vlastne je. V druhej kapitole sme opísali štyri jednoduché kroky ako takýto model vytvoriť. Potom sme aj prakticky ukázali, ako sa tieto pravidlá aplikujú v dizajne nášho logického modelu. Identifikovali a zadefinovali sme biznis procesy, ktoré by sme chceli vymodelovať. Vytvorili sme logický model zadefinovaných procesov na základe dobrých praktík definovania grainu, dimenzií, faktov a vzťahov medzi nimi. Pôvodným cieľom bolo aj vytvorenie nerelačnej databázy. Nepodarilo sa nám však vymodelovať multidimenzionálnu kocku z dôvodu časovej tiesne a neskúsenosti s takýmito objektami. Nakoniec sme na základe vytvoreného modelu naimplementovali fyzickú databázu, kde sme vytvorili pre všetky logické objekty ich fyzické implementácie. Vo fyzickom modeli sme si dali záležať na dobre navrhnutých indexoch a rýchlosti dotazov. Na navrhnutej databáze je možné ďalej pracovať a vybudovať kompletný dátový sklad. Toto sa dá uskutočniť implementáciou ETL procesov, ktoré by sklad zásobovali dátami z operačných systémov a vytvorením či už logických alebo fyzických data martov, ktoré by poskytovali multidimenzionálny pôžitok z dát koncovým používateľom.

Literatúra

- [1] CHRISTOPHER ADAMSON, 2010, Star Schema: The Complete Reference, 510 s., Pearson International edition, 978-0071744324
- [2] RALPH KIMBALL - MARGY ROSS, 2013, The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, 600s., 978-1-118-53080-1
- [3] PLACHETKA T., 2015, Prezentácia ku predmetu Úvod do databázových systémov, [online] Dostupné na internete: <http://www.dcs.fmph.uniba.sk/plachetk/TEACHING/DB2015/db2015_8.pdf>
- [4] Oracle Release 8.1.5, Dokumentácia Oracle databázového servera - optimalizácia join operácie, [online] Dostupné na internete: <https://docs.oracle.com/cd/F49540_01/DOC/server.815/a67781/c20c_joi.htm#2437>
- [5] Oracle Server 12c vo Virtual Box-e, [online] Dostupné na internete: <<http://www.oracle.com/technetwork/middleware/oedq/downloads/edq-vm-download-2424092.html>>

Dodatok

K práci prikladám aj USB kľúč, ktorý obsahuje súbory so zdrojovými kódmi pre vytvorenie databázy popísanej v kapitole 3 o fyzickom modeli. Na kľúči sa nachádza aj projekt logického modelu reportovacieho systému opísaného kapitole 2. Tento projekt je možné otvoriť pomocou programu Oracle Data Modeler, ktorého inštaláčne súbory sú súčasťou kľúča.