

Datová kostka pro analýzy výzkumu a vývoje inovací pro datový sklad ZČU

Dominik Jež¹

1 Úvod

Centrum informatizace a výpočetní techniky poskytuje služby Západočeské univerzitě v Plzni, které jsou spojené s rozvojem a provozem informačních technologií. Cílem práce bylo vytvořit datovou kostku, která bude napojena na stávající systémy pro evidenci projektu na ZČU (Granty a Projekty od firmy ders). Jedno odvětví centra se věnuje tvoření datových skladů, pomocí kterých lze analyzovat data sloužící např. k výkaznictví výročních zpráv, k tvorbě a udržení strategických záměrů univerzity, dalším analýzám či hledání nečekaných korelací.

V dnešní době je k dispozici nesčetné množství dat, které je nutné rozčlenit a uložit do systému, aby informace šly efektivně zpracovávat – filtrace či vizualizace. V práci jsou vybrána vhodná data z CEPu (Centrální Evidence Projektů), která slouží pro naplnění datové kostky. V závěru je demonstrována vizualizace dat nad vytvořenou datovou kostkou.

2 Struktura registru CEP

IS VaVaI představuje informační systém výzkumu, vývoje a inovací, který integruje a shromažďuje data o výzkumu, vývoji a inovacích. Je to jediný oficiální autorizovaný zdroj, který je podporován z veřejných rozpočtů České republiky. Celkem obsahuje čtyři vzájemně propojené oblasti: CEP, CEA (Centrální evidence aktivit VaVaI), RIV (Rejstřík informací o výsledcích) a VES (Evidence veřejných soutěží ve VaVaI). (IS VaVaI (2022))

CEP obsahuje údaje o projektech a poskytovanou účelovou podporu. Jednotlivé údaje jsou uvedeny ve třech fázích: při zahájení, během řešení a po ukončení řešení. Pro upřesnění, CEP obsahuje údaje určující projekt – název, předmět řešení, příjemce, řešitelé, údaje o uzavření smlouvy, kategorie výzkumu, poskytovatel podpory, program a mnohem více. Veškerá data jsou dostupná ke stažení ve formátu JSON po zpřístupnění tokenu. (IS VaVaI (2022))

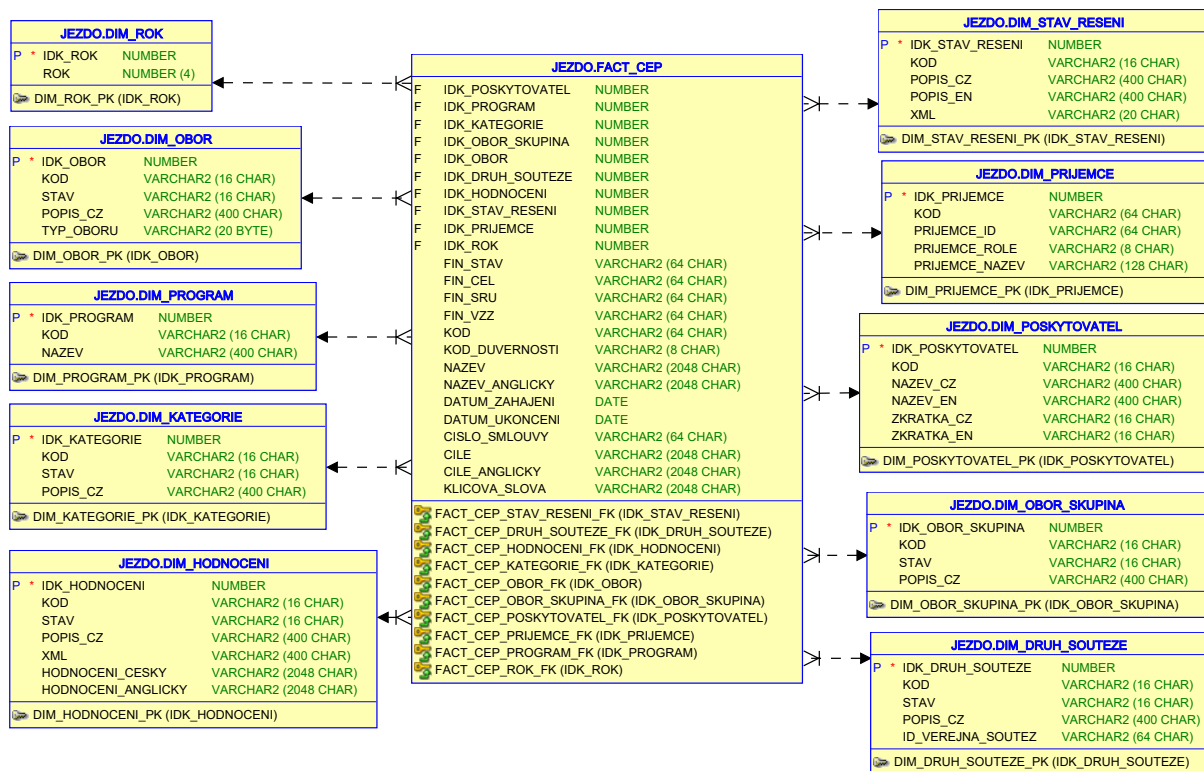
3 Transformace dat v databázi

Během nahrávání dat z formátu JSON do databázové struktury bylo nutné prozkoumat strukturu číselníků a detailů. Tento proces lze rozdělit do několika menších úkolů – tj. vytvoření tabulek o stejné struktuře jako formát souboru ve formátu JSON, transformace během nahrání dat do nově vytvořené tabulky, čištění dat a vytvoření závislostí mezi ostatními tabulkami (Kimball a Ross (2013)). Do relační databáze Oracle byla data vložena programovým způsobem, kde data byla nahrána na server a odsud se volaly skripty z vývojového prostředí.

Pro automatizované vytvoření datové kostky a její průběžné aktualizování byl vytvořen programový balík s názvem PG_CEP v jazyce PL/SQL. Tělo balíku obsahuje uložené procedury

¹ student bakalářského studijního programu Informatika a výpočetní technika, obor Informatika, e-mail: jezdo0@students.zcu.cz

pro naplnění nebo aktualizaci dat tabulek dimenzí a znova naplnění faktové tabulky datové kostky (obrázek 1).



Obrázek 1: Vytvořené schéma datové kostky CEP

4 Vizualizace dat a závěr

Analytické dotazy slouží k analýze aktuálních a historických dat – lze odhalit různé korelace, trendy/poznatky. Pro vizualizaci analytických dat byl použit nástroj Power BI, který slouží pro vizualizaci datové kostky. Nástroje BI (Business Intelligence) umožňují práci s velkými daty a jejich následnou prezentaci (Laberge (2011)). Byly vytvořené dashboards, na kterých se nachází různorodé grafy (sloupcové, koláčové, řádkové), treemapy, tabulky a další.

Datová kostka byla úspěšně vytvořena, vizualizována a otestována. Z vizualizace lze např. zjistit že:

- nejoblíbenější obor v průběhu let je Genetika a molekulární biologie,
- v kategorii experimentálního vývoje se nejčastěji vyskytuje program Operační program Podnikání a inovace pro konkurenceschopnost, který se využívá čím dál častěji.

Literatura

IS VaVaI (2022) *O IS VAVAI*. Dostupné na odkazu: <https://www.isvavai.cz/>.

Kimball, R., a Ross M. (2013) *Warehouse Toolkit*. John Wiley and Sons, Inc., Indianapolis, Indiana.

Laberge, R. (2011) *Datové sklady, Agilní metody a business intelligence*. Computer Press v Brně ve společnosti Albatros Media a. s..