

Python Package Metadata Management

Basic Databases

by Nguyễn Gia Phong, Nguyễn Quốc Thông,
Nguyễn Văn Tùng and Trần Minh Vương

July 5, 2020

Contents

1	Introduction	2
1.1	Brief Description	2
1.2	Authors and Credits	2
2	User Requirements	3
3	Data Definition	3
3.1	Entity Relationship Diagram	3
3.2	Database Schema	3
4	Data Query	3
5	Conclusion	3
6	References	3

1 Introduction

1.1 Brief Description

In traditional Unix-like operating systems like GNU/Linux distributions and BSD-based OSes, package managers tries to synchronize the packages meta-data (such as available versions and dependencies) with that of central repositories. While this proves to be reliable and efficient, language-specific package managers do not usually have such synchronized databases, since they focus on development libraries which have more flexible constraints.

Within the Python packaging ecosystem, this is the case, where package managers like `pip` needs to fetch metadata of each package to be installed to find out dependencies and other information. This turns out to have heavy performance penalty on the dependency resolution process alone, which is already a NP-hard problem. This project explores ways to store these meta-data in an efficient in a database, to be used in practice either locally or in a local/regional network, to avoid Python package managers from having to fetch (and potentially build) entire packages just to find out if it conflicts with others.

1.2 Authors and Credits

The work has been undertaken by group number 8, whose members are listed in the following table.

Full name	Student ID
Nguyễn Gia Phong	BI9-184
Nguyễn Quốc Thông	BI9-214
Nguyễn Văn Tùng	BI9-229
Trần Minh Vương	BI9-239

This report is licensed under a CC BY-SA 4.0 license, while the source code is available on GitHub* under AGPLv3+.

We would like to express our special thanks to Dr. Nguyễn Hoàng Hà, whose lectures gave us basic understanding on the key principles of relational databases. In addition, we also recieved a lot of help from the Python packaging community over `#pypa` on Freenode on understanding the structure of the metadata as well as finding a way to fetch these data from package indices.

*<https://github.com/McSinyx/cheese-shop>

2	User Requirements
3	Data Definition
3.1	Entity Relationship Diagram
3.2	Database Schema
4	Data Query
5	Conclusion
6	References