Introduction

User Requirements

Data

Data Oues

Conclusion

## Python Package Metadata Management

Nguyễn Gia Phong—BI9-184 Nguyễn Quốc Thông—BI9-214 Nguyễn Văn Tùng—BI9-229 Trần Minh Vương—BI9-239

University of Science and Technology of Hà Nội

July 8, 2020

Introduction

Requirement

2 User Requirements

1 Introduction

Data Quer

3 Data Definition
Entity Relationship Diagram
Database Schema

4 Data Query

Introduction

- Python package managers download whole packages just for metadata
- Mirroring PyPI is expensive (6 GB)
- Middle approach: Mirroring metadata

Introduction

User Requirements

Data Ouer

\_ . .

- 1 Introduction
- 2 User Requirements
- 3 Data Definition
  Entity Relationship Diagram
  Database Schema
- 4 Data Query
- **6** Conclusion

miroduction

User Requirements

Data Definition

Data Quer

- list\_projects()
   List of registered project names.
- project\_releases(project)
   List of releases for given project, ordered by version.
- project\_release\_latest()
  Latest release of given project.
- belong\_to(name)
   List of projects whose author is name.

miroduction

User Requirements

Data

Data Quer

- browse(classifier): List of (project, version) of all releases classified by classifier.
- release\_data(project, version): Metadata of given release: project, version, homepage, author, author's email, summary, keywords, classifiers and dependencies
- search\_name(pattern): List of (project, version, summary) where name matches pattern.
- search\_summary(pattern): List of (project, version, summary) where summary matches pattern.

Data

Definition

- 3 Data Definition Entity Relationship Diagram Database Schema

Introduction

User

Requiremen

Definition

Entity Relationship

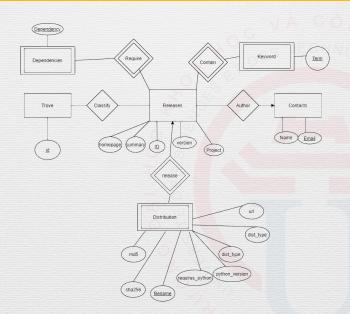
Diagram

Database Schem

Data Quer

Conclusion

## Entity Relationship Diagram



Introduction

Requirements

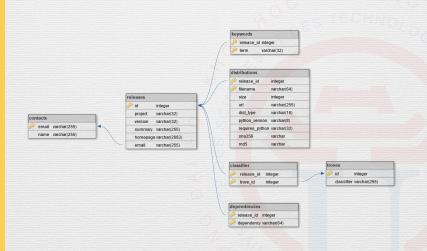
Requirement

Delinition

Diagram

Database Schema

Data Quer



Python Package Metadata Management

Group 8

Introduction

User Requirements

Data Query

1 Introduction

2 User Requirements

3 Data Definition Entity Relationship Diagram Database Schema

4 Data Query

Data Query

## SELECT DISTINCT project FROM releases;

Introduction

Requirement

Requirement

Data Query

Data Que

SELECT version
FROM releases
WHERE project = 'spam'
ORDER BY version;

Data Query

SELECT version FROM releases WHERE project = 'spam' ORDER BY version DESC LIMIT 1;

Data Query

CREATE VIEW authorships AS SELECT name as author, project FROM contacts NATURAL JOIN releases GROUP BY author, project;

Introduction

Requirement

Requiremen

Data Query

\_\_\_\_

SELECT project
FROM authorships
WHERE author='Monty Python';

.....

Requirement

Definition BEGI

Data Query

```
DELIMITER //
CREATE PROCEDURE browse(class varchar(255))
BEGIN
  SELECT project, version
  FROM releases, classifiers
  WHERE id = release id AND trove id = (
    SELECT id
    FROM troves
    WHERE classifier = class);
END//
DELIMITER ;
```

User

Requiremen

Data Querv

Data que

```
Conclusion
```

```
DELIMITER //
CREATE PROCEDURE release data(
  project varchar(32), version varchar(32))
BEGIN
  DECLARE i smallint;
  SET i = (
    SELECT id
    FROM releases
    WHERE releases.project = project
      AND releases.version = version);
  SELECT project, version, homepage,
    name as author, email, summary
  FROM releases NATURAL JOIN contacts
  WHERE id = i;
```

...

Requirement

Data Query

```
SELECT term as keyword
FROM keywords
WHERE release_id = i;
```

```
SELECT classifier
FROM classifiers, troves
WHERE release id = i AND trove id = troves.id;
```

```
SELECT dependency
FROM dependencies
WHERE release_id = i;
END//
DELIMITER :
```

Data Query

SELECT project, version, summary FROM releases WHERE project LIKE 'py%';

Data Query

SELECT project, version, summary FROM releases WHERE summary LIKE '%num%';

Introduction

User Requirements

Data Ouen

- 1 Introduction
- 2 User Requirements
- 3 Data Definition
  Entity Relationship Diagram
  Database Schema
- 4 Data Query
- **5** Conclusion

Introduction

Requirement

Dete

Data Quer

- Relational databases
- SQL—MySQL in particular
- Python package metadata format

Conclusion



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.