

**CZ 4031:**

**Database System Principles**

**Project 1:**

**Querying Databases Efficiently**

**Group Members:**

Zhang Qing Mei

Li Chu Qiao

Qian Lei

Apple Tan Xuan Ning

**Submission date:** 11th October 2016

**Semester 1 2016/2017**

**SCHOOL OF COMPUTER SCIENCE AND ENGINEERING**

**NANYANG TECHNOLOGICAL UNIVERSITY**

Table of Contents

[1 Introduction 3](#_Toc400874110)

[2 Schema Design 3](#_Toc400874111)

[2.1 E/R Diagram 3](#_Toc400874112)

[2.2 Commands for Creating Tables 4](#_Toc400874113)

[2 Data Acquisition 6](#_Toc400874114)

[3 Queries 6](#_Toc400874115)

[3.1 SQL Queries 6](#_Toc400874116)

[3.2 Screen Captures of SQL Queries Results 10](#_Toc400874117)

[3.3 Analysis of the Effect of Database Size on Query Time 16](#_Toc400874118)

[4. Index Building 23](#_Toc400874119)

[4.1 CREATE INDEX Statements 23](#_Toc400874120)

[4.2 Comparison of Query Execution Time 23](#_Toc400874121)

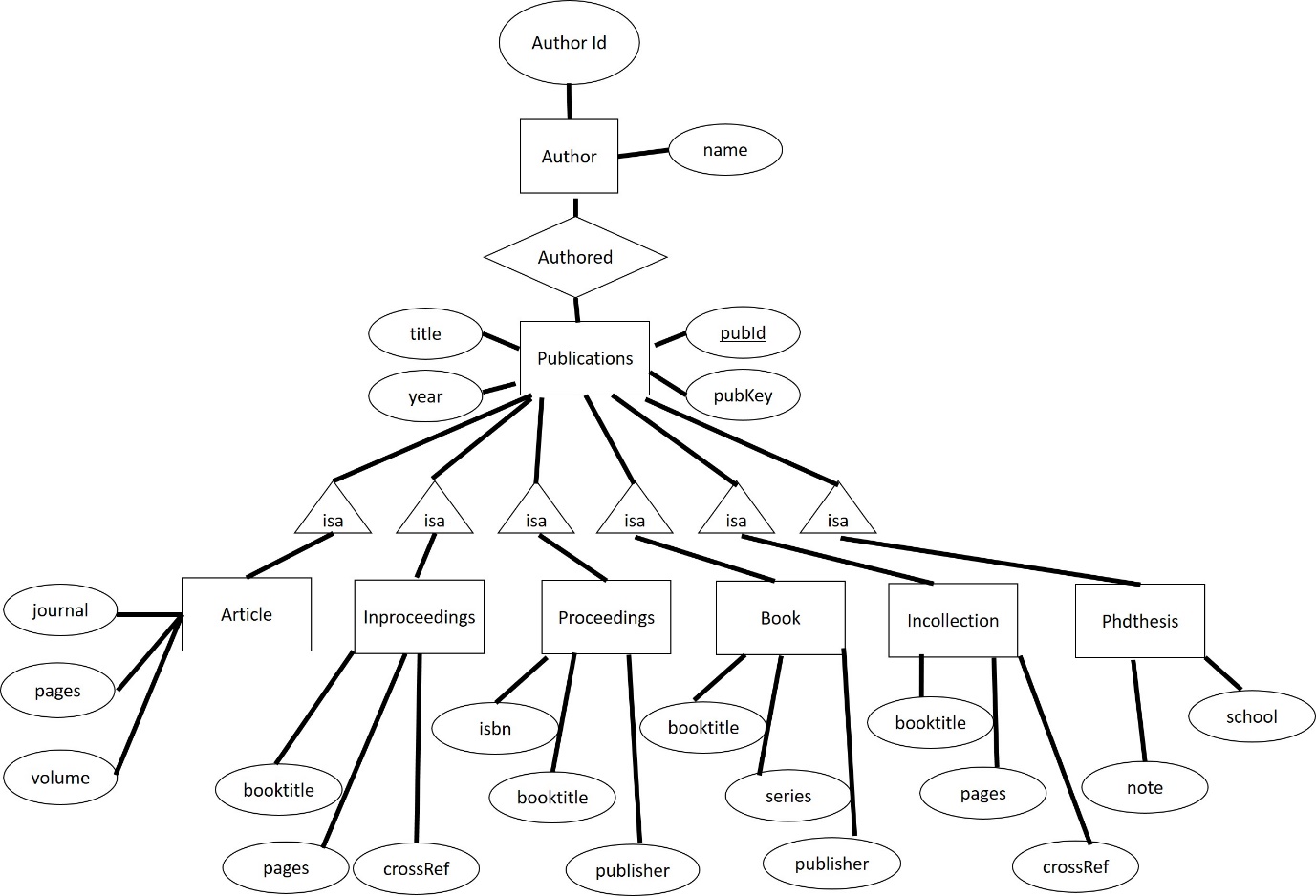
# 1 Introduction

This project aims in allowing us to get familiarized with the main components of the database design: schema design, data acquisition, data transformation, querying, and indexing. We are required to self-learn how to write a program to extract data from XML file into CSV file, to deal with a new database, and which it involves huge lump of data. For this project, we have chosen Postgresql as our database and java language to extract data from XML file to CSV file.

# 2 Schema Design

This section consists of an E/R diagram and the SQL commands for creating the tables.

## 2.1 E/R Diagram

****

## 2.2 Commands for Creating Tables

Article Table

CREATE TABLE "article" (

"pubId" int NOT NULL,

"journal" VARCHAR(255) DEFAULT NULL,

"pages" VARCHAR(255) DEFAULT NULL,

"volume" VARCHAR(255) DEFAULT NULL,

PRIMARY KEY ("pubId")

);

ALTER TABLE article

ADD CONSTRAINT FK\_Article\_PubID FOREIGN KEY (pubid)

    REFERENCES publication (pubid)

    ON DELETE CASCADE

    ON UPDATE CASCADE

;

Author Table

CREATE TABLE "author" (

"authorId" int NOT NULL,

"name" varchar(255) DEFAULT NULL,

PRIMARY KEY ("authorId")

);

Authored Table

CREATE TABLE "authored" (

 "pubId" int NOT NULL,

 "authorId" int NOT NULL,

 PRIMARY KEY ("pubId","authorId")

);

Book Table

CREATE TABLE "book" (

"pubId" int NOT NULL,

“booktitle” varchar(255) DEFAULT NULL,

"series" varchar(255) DEFAULT NULL,

"publisher" varchar(255) DEFAULT NULL,

PRIMARY KEY ("pubId")

);

Incollection Table

CREATE TABLE "incollection" (

CREATE TABLE "incollection" (

"pubId" int NOT NULL,

"pages" VARCHAR(255) DEFAULT NULL,

"booktitle" VARCHAR(255) DEFAULT NULL,

"crossref" VARCHAR(255) DEFAULT NULL,

PRIMARY KEY ("pubId")

);

Inproceedings Table

CREATE TABLE "inproceedings" (

"pubId" int NOT NULL,

"booktitle" VARCHAR(255) DEFAULT NULL,

"pages" VARCHAR(255) DEFAULT NULL,

"crossref" VARCHAR(255) DEFAULT NULL,

PRIMARY KEY ("pubId")

);

Publication Table

CREATE TABLE "publication" (

"pubId" int NOT NULL,

"title" varchar(255) DEFAULT NULL,

"year" int DEFAULT NULL,

"pubKey" varchar(255) DEFAULT NULL,

PRIMARY KEY ("pubId")

);

Proceedings Table

CREATE TABLE "proceedings" (

 "pubId" int NOT NULL,

 "isbn" VARCHAR(255) DEFAULT NULL,

 "booktitle" VARCHAR(255) DEFAULT NULL,

 "publisher" VARCHAR(255) DEFAULT NULL,

 PRIMARY KEY ("pubId")

);

Phdthesis Table

CREATE TABLE "phdthesis" (

"pubId" int NOT NULL,

"school" VARCHAR(255) DEFAULT NULL,

"note" VARCHAR(255) DEFAULT NULL,

 PRIMARY KEY ("pubId")

);

# 2 Data Acquisition

This section is submitted as zip file.

# 3 Queries

This section includes our SQL queries and the screenshots of the results after running these queries. Also, there will be an analysis on how the database size affects the query time for each query.

## 3.1 SQL Queries

Question 1

Question 2

Question 3a

Question 3b

Question 4

Question 5a

Question 5b

Question 6

Question 7

Question 8

Question 9

Question 10

Question 11

## 3.2 Screen Captures of SQL Queries Results

Question 1

Question 2

Question 3a

Question 3b

Question 4

Question 5a

Question 5b

Question 6

Question 7

Question 8

Question 9

Question 10

Question 11

## 3.3 Analysis of the Effect of Database Size on Query Time

Question 1

Question 2

Question 3a

Question 3b

Question 4

Question 5a

Question 5b

Question 6

Question 7

Question 8

Question 9

Question 10

Question 11

Analysis

From all the result shown above, as data size reduces by half, execution time reduced by half or even more, and was even further reduced when data size was cut to a quarter size. This result has shown that data size is a one of the determining factor where execution time is concerned.

# 4. Index Building

## 4.1 CREATE INDEX Statements

The following is the statement we used for creating the index.

## 4.2 Comparison of Query Execution Time

Question 3a

Question 7

Question 9

Analysis