**CONFIGS \_LOCAL&SERVER\_DATABASE**

**SETUP PROCESS LOCAL**

|  |
| --- |
|  |

Prepared by***: NAM***

**08/12/2024**

**Mục lục**

[Local database configurations 3](#_Toc184566629)

[1.1 Introduction 3](#_Toc184566630)

[1.2 Preparation 3](#_Toc184566631)

[1.3 Local database configurations 3](#_Toc184566632)

[1.3.1 Control 3](#_Toc184566633)

[1.3.2 Staging 5](#_Toc184566634)

[1.3.3 Warehouse 6](#_Toc184566635)

[1.3.4 Datamart 7](#_Toc184566636)

[Setup Processes 9](#_Toc184566637)

[*2.1 Preparation* 9](#_Toc184566638)

[*2.2 Setup process with task scheduler in Window* 10](#_Toc184566639)

**Revision History**

|  |  |  |  |
| --- | --- | --- | --- |
| **Revision Date** | **Version No.** | **Detailed Description of Revision** | **Document Authors** |
| 08/12/2024 | 1.0 | Cấu hình local database và cài đặt để chạy tự động lấy dữ liệu | Nam |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

# Local database configurations

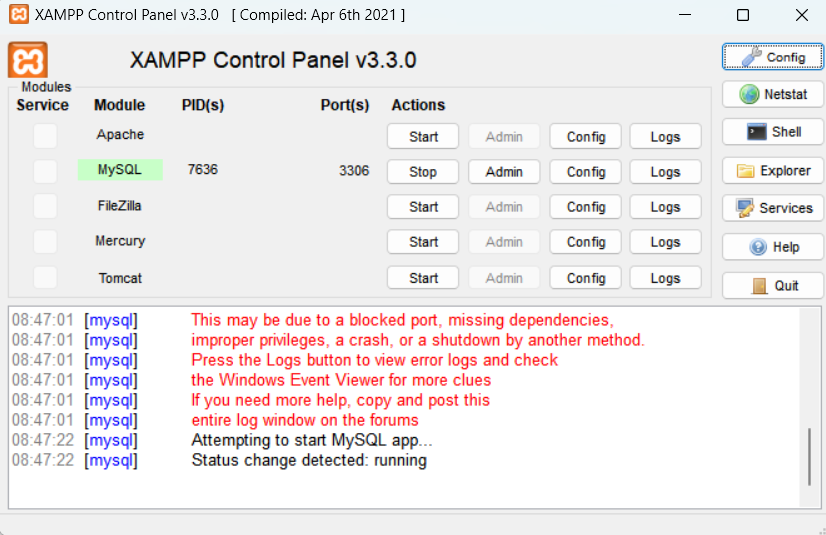
## 1.1 Introduction

## 1.2 Preparation

* MySQL version 10.4.32-MariaDB

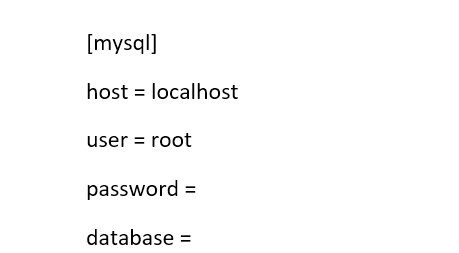


* Xampp Control Panel version 3.3.0



## 1.3 Local database configurations

Connect database vào file.ini với database name lần lượt có như phía dưới:



### 1.3.1 Control

**Khởi tạo các bản trong database:**

* **Bảng config:**

DROP TABLE IF EXISTS `config`;

CREATE TABLE `config` (

`id` int NOT NULL AUTO\_INCREMENT,

`keyword` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`saveFolder` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`scrapeTimes` int NULL DEFAULT NULL,

`website` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

PRIMARY KEY (`id`) USING BTREE

) ENGINE = InnoDB AUTO\_INCREMENT = 5 CHARACTER SET = utf8mb4 COLLATE = utf8mb4\_general\_ci ROW\_FORMAT = Dynamic;

* **Bảng datedim:**

DROP TABLE IF EXISTS `datedim`;

CREATE TABLE `datedim` (

`dateSk` int NOT NULL,

`DayOfYear` int NULL DEFAULT NULL,

`calendarMonth` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`calendarYear` int NULL DEFAULT NULL,

`calendarYearMonth` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`dayOfMonth` int NULL DEFAULT NULL,

`dayOfWeek` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`daySince2005` int NULL DEFAULT NULL,

`dayType` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`fullDate` date NULL DEFAULT NULL,

`holiday` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`quarterOfYear` int NULL DEFAULT NULL,

`quarterSince2005` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`weekMondayStart` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`weekOfYearMonday` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`weekOfYearSunday` int NULL DEFAULT NULL,

`weekSundayStart` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`yearWeekSunday` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`monthSince2005` int NULL DEFAULT NULL,

PRIMARY KEY (`dateSk`) USING BTREE,

INDEX `FKmjrv6v895m97jp33bji2c89xi`(`monthSince2005`) USING BTREE,

CONSTRAINT `FKmjrv6v895m97jp33bji2c89xi` FOREIGN KEY (`monthSince2005`) REFERENCES `monthdim` (`monthSk`) ON DELETE RESTRICT ON UPDATE RESTRICT

) ENGINE = InnoDB CHARACTER SET = utf8mb4 COLLATE = utf8mb4\_general\_ci ROW\_FORMAT = Dynamic;

* **Bảng logs:**

CREATE TABLE `logs` (

`id` int NOT NULL AUTO\_INCREMENT,

`message` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`quantity` int NULL DEFAULT NULL,

`timeEnd` datetime(6) NULL DEFAULT NULL,

`timeStart` datetime(6) NULL DEFAULT NULL,

`dateSk` int NULL DEFAULT NULL,

`statusId` int NULL DEFAULT NULL,

`websiteId` int NULL DEFAULT NULL,

PRIMARY KEY (`id`) USING BTREE,

INDEX `FK9714p8exryruau3vv4pw926fx`(`dateSk`) USING BTREE,

INDEX `FK3d5jkj33k7m98tttikpva61fh`(`statusId`) USING BTREE,

INDEX `FKnk8xfl5b2iffccafv3c10t5pa`(`websiteId`) USING BTREE,

CONSTRAINT `FK3d5jkj33k7m98tttikpva61fh` FOREIGN KEY (`statusId`) REFERENCES `status` (`id`) ON DELETE RESTRICT ON UPDATE RESTRICT,

CONSTRAINT `FK9714p8exryruau3vv4pw926fx` FOREIGN KEY (`dateSk`) REFERENCES `datedim` (`dateSk`) ON DELETE RESTRICT ON UPDATE RESTRICT,

CONSTRAINT `FKnk8xfl5b2iffccafv3c10t5pa` FOREIGN KEY (`websiteId`) REFERENCES `config` (`id`) ON DELETE RESTRICT ON UPDATE RESTRICT

) ENGINE = InnoDB AUTO\_INCREMENT = 110 CHARACTER SET = utf8mb4 COLLATE = utf8mb4\_general\_ci ROW\_FORMAT = Dynamic;

* **Bảng monthdim:**

DROP TABLE IF EXISTS `monthdim`;

CREATE TABLE `monthdim` (

`monthSk` int NOT NULL,

`calendarYearMonth` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`dateSkEnd` int NULL DEFAULT NULL,

`dateSkStart` int NULL DEFAULT NULL,

`monthSince2005` int NULL DEFAULT NULL,

PRIMARY KEY (`monthSk`) USING BTREE

) ENGINE = InnoDB CHARACTER SET = utf8mb4 COLLATE = utf8mb4\_general\_ci ROW\_FORMAT = Dynamic;

* **Bảng status:**

DROP TABLE IF EXISTS `status`;

CREATE TABLE `status` (

`id` int NOT NULL AUTO\_INCREMENT,

`name` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

PRIMARY KEY (`id`) USING BTREE,

UNIQUE INDEX `UKreccgx9nr0a8dwv201t44l6pd`(`name`) USING BTREE

) ENGINE = InnoDB AUTO\_INCREMENT = 5 CHARACTER SET = utf8mb4 COLLATE = utf8mb4\_general\_ci ROW\_FORMAT = Dynamic;

### 1.3.2 Staging

**Khởi tạo các bản trong database:**

* **Bảng bikes:**

DROP TABLE IF EXISTS `bikes`;

CREATE TABLE `bikes` (

`naturalId` int NOT NULL AUTO\_INCREMENT,

`id` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`name` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`price` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`priceSale` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`brand` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`color` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`size` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`status` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`description\_part1` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`description\_part2` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`description\_part3` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`timeStartScrape` datetime(6) NULL DEFAULT NULL,

`timeEndScrape` datetime(6) NULL DEFAULT NULL,

`timeStartInsert` datetime(6) NULL DEFAULT NULL,

PRIMARY KEY (`naturalId`) USING BTREE,

INDEX `idx\_id`(`id`) USING BTREE

) ENGINE = InnoDB AUTO\_INCREMENT = 16362 CHARACTER SET = utf8mb4 COLLATE = utf8mb4\_general\_ci ROW\_FORMAT = Dynamic;

* **Tạo các procedure:**

### 1.3.3 Warehouse

**Khởi tạo các bản trong database:**

* **Bảng dimBikesProcessed:**

DROP TABLE IF EXISTS `dim\_bikes\_processed`;

CREATE TABLE `dim\_bikes\_processed` (

`product\_sk` int NOT NULL AUTO\_INCREMENT,

`id` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`name` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`price` decimal(15, 2) NULL DEFAULT NULL,

`priceSale` decimal(15, 2) NULL DEFAULT NULL,

`brand` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`color` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`size` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`status` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`description\_part1` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`description\_part2` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`description\_part3` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`created\_at` datetime NULL DEFAULT current\_timestamp,

`isDelete` tinyint(1) NULL DEFAULT 0,

`date\_delete` date NULL DEFAULT NULL,

`date\_insert` timestamp NOT NULL DEFAULT current\_timestamp,

`expired\_date` date NULL DEFAULT '9999-12-31',

`dateSk` int NULL DEFAULT NULL,

PRIMARY KEY (`product\_sk`) USING BTREE,

INDEX `dateSk`(`dateSk`) USING BTREE,

CONSTRAINT `dim\_bikes\_processed\_ibfk\_1` FOREIGN KEY (`dateSk`) REFERENCES `datedim` (`dateSk`) ON DELETE RESTRICT ON UPDATE RESTRICT

) ENGINE = InnoDB AUTO\_INCREMENT = 1 CHARACTER SET = utf8mb4 COLLATE = utf8mb4\_general\_ci ROW\_FORMAT = Dynamic;

* **Bảng dimBikesRaw:**

DROP TABLE IF EXISTS `dim\_bikes\_raw`;

CREATE TABLE `dim\_bikes\_raw` (

`naturalId` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`id` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`name` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`price` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`priceSale` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`brand` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`color` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`size` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`status` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`description\_part1` text CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL,

`description\_part2` text CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL,

`description\_part3` text CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL,

`timeStartInsert` datetime NULL DEFAULT NULL,

`created\_at` datetime NULL DEFAULT current\_timestamp

) ENGINE = InnoDB CHARACTER SET = utf8mb4 COLLATE = utf8mb4\_general\_ci ROW\_FORMAT = Dynamic;

### 1.3.4 Datamart

**Khởi tạo các bản trong database:**

* **Bảng product:**

DROP TABLE IF EXISTS `product`;

CREATE TABLE `product` (

`id` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`name` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`price` decimal(15, 2) NULL DEFAULT NULL,

`priceSale` decimal(15, 2) NULL DEFAULT NULL,

`brand` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`color` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`size` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`status` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`description\_part1` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`description\_part2` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`description\_part3` varchar(255) CHARACTER SET utf8mb4 COLLATE utf8mb4\_general\_ci NULL DEFAULT NULL,

`created\_at` datetime NULL DEFAULT current\_timestamp,

`isDelete` tinyint(1) NULL DEFAULT 0,

`date\_insert` datetime NOT NULL,

`expired\_date` date NULL DEFAULT '9999-12-31',

`date\_sk` int NULL DEFAULT NULL,

INDEX `fk\_product\_date`(`date\_sk`) USING BTREE,

CONSTRAINT `fk\_product\_date` FOREIGN KEY (`date\_sk`) REFERENCES `date` (`date\_sk`) ON DELETE CASCADE ON UPDATE CASCADE

) ENGINE = InnoDB CHARACTER SET = utf8mb4 COLLATE = utf8mb4\_general\_ci ROW\_FORMAT = Dynamic;

* **Tạo các procedure:**

DROP PROCEDURE IF EXISTS `LoadDataFromDWToDM`;

delimiter ;;

CREATE PROCEDURE `LoadDataFromDWToDM`()

BEGIN

START TRANSACTION;

-- Xóa các bảng tạm nếu tồn tại

DROP TABLE IF EXISTS datamart.product\_temp;

DROP TABLE IF EXISTS datamart.date\_temp;

-- Tạm thời vô hiệu hóa ràng buộc khóa ngoại

SET foreign\_key\_checks = 0;

-- Tạo bảng tạm để lưu dữ liệu từ product\_dim

CREATE TABLE datamart.product\_temp AS

SELECT

p.id,

p.name,

p.price,

p.priceSale,

p.brand,

p.color,

p.size,

p.status,

p.description\_part1,

p.description\_part2,

p.description\_part3,

p.created\_at,

p.isDelete,

CURRENT\_TIMESTAMP AS date\_insert, -- Thời gian insert vào Data Mart

p.expired\_date,

p.date\_sk

FROM datawarehouse.product\_dim p;

-- Tạo bảng tạm để lưu dữ liệu từ date\_dim

CREATE TABLE datamart.date\_temp AS

SELECT \*

FROM datawarehouse.date\_dim;

-- Đổi tên bảng chính trong DM thành old

RENAME TABLE datamart.product TO datamart.product\_old;

RENAME TABLE datamart.date TO datamart.date\_old;

-- Đổi tên bảng tạm thành bảng chính trong DM

RENAME TABLE datamart.product\_temp TO datamart.product;

RENAME TABLE datamart.date\_temp TO datamart.date;

-- Xóa các bảng cũ trong DM nếu có

DROP TABLE IF EXISTS datamart.product\_old;

DROP TABLE IF EXISTS datamart.date\_old;

-- Đảm bảo các giá trị date\_sk trong datamart.product phải tồn tại trong datamart.date

-- Tạo chỉ mục cho date\_sk trong bảng date nếu chưa có

CREATE INDEX idx\_date\_sk ON datamart.date(date\_sk);

-- Thêm khóa ngoại vào bảng product

ALTER TABLE datamart.product

ADD CONSTRAINT fk\_product\_date FOREIGN KEY (date\_sk) REFERENCES datamart.date(date\_sk)

ON DELETE CASCADE ON UPDATE CASCADE;

-- Bật lại ràng buộc khóa ngoại

SET foreign\_key\_checks = 1;

COMMIT;

END

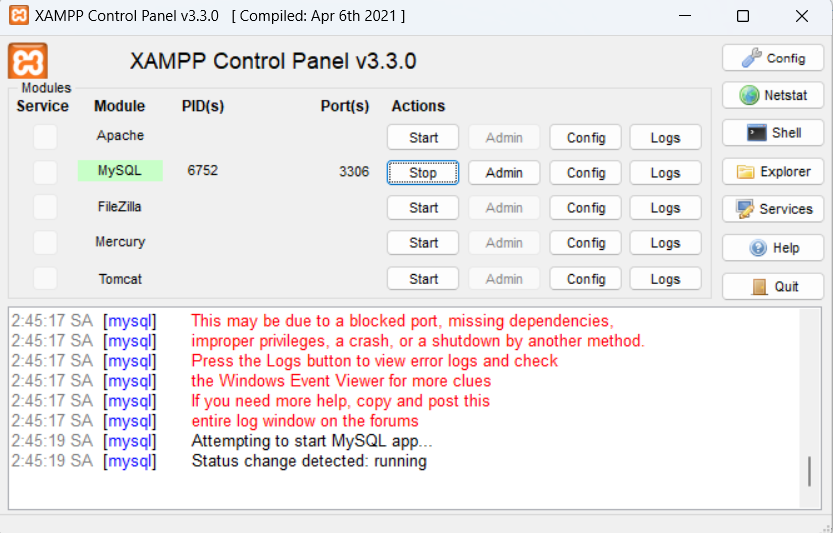
;;

delimiter ;

# Setup Processes

## *2.1 Preparation*

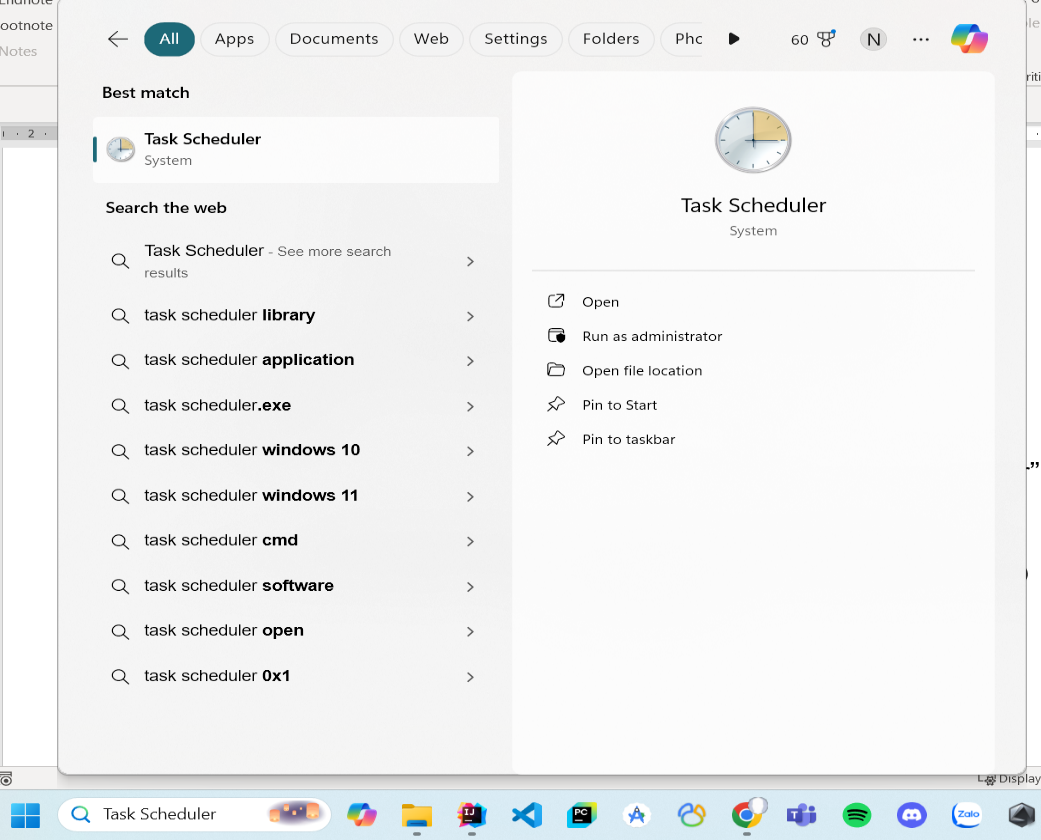
Khởi chạy xampp> start MySQL



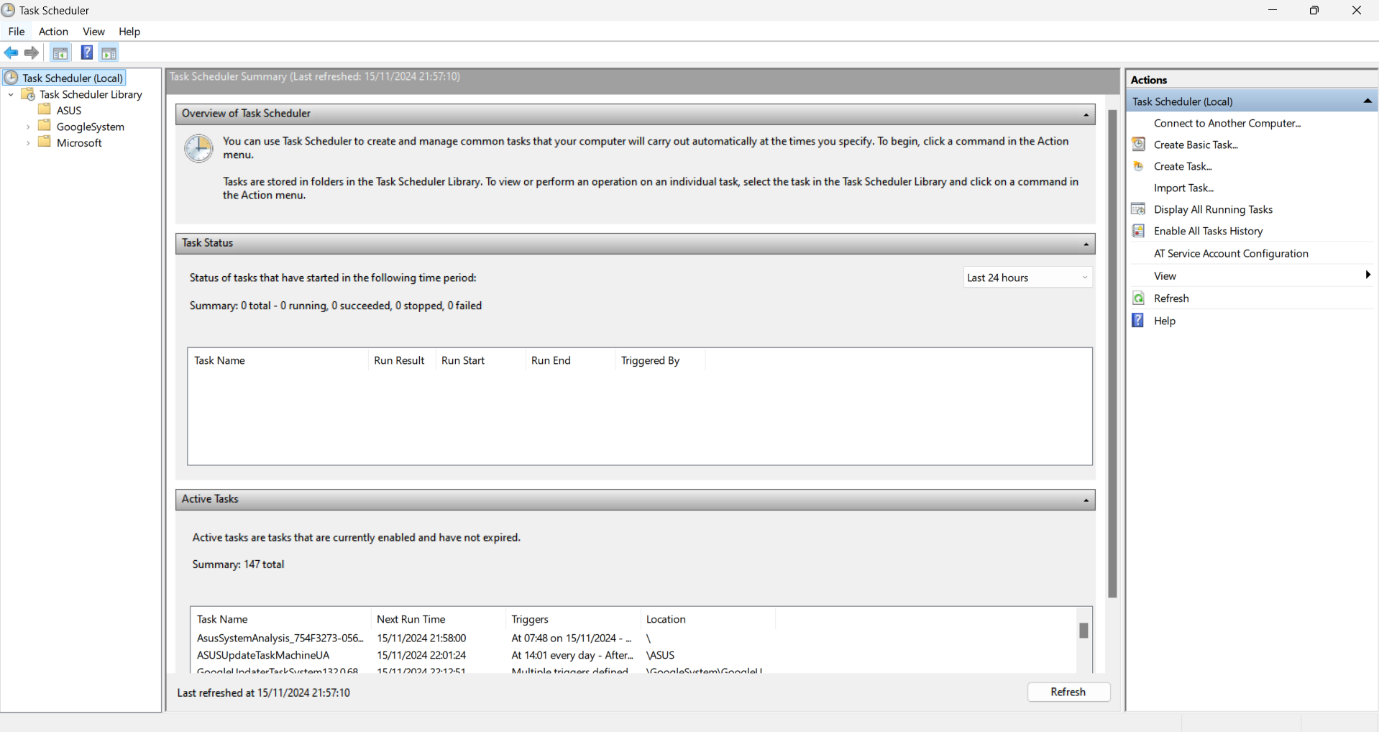
## *2.2 Setup process with task scheduler in Window*

**- Cài đặt task scheduler cho staging:**

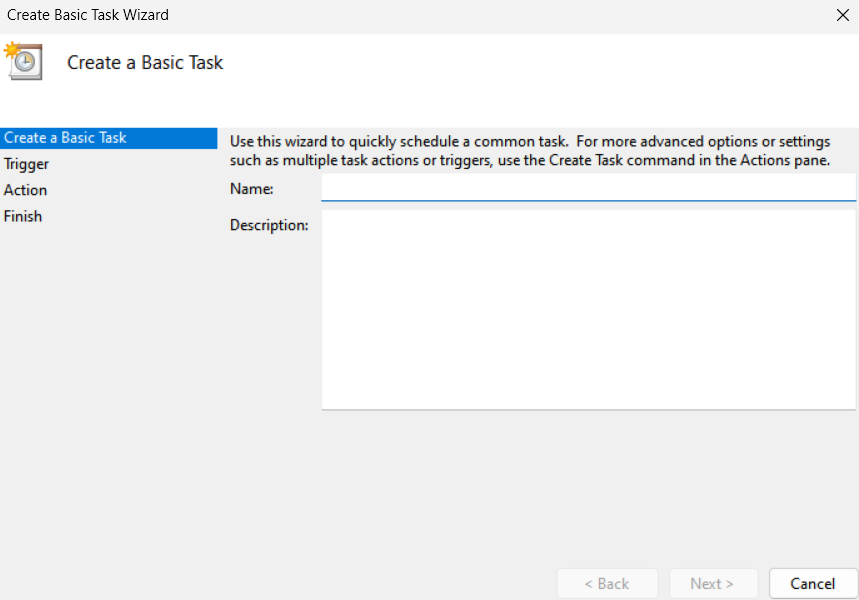
**B1:** Vào ô tìm kiếm của window nhập “Task Scheduler”



**B2:** Chọn “Create Basic Task”

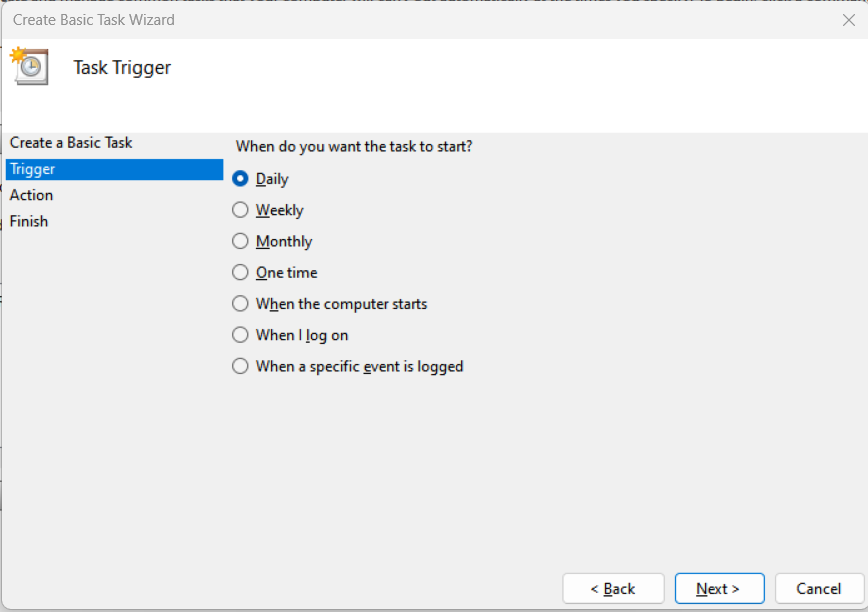


**B3:** Ở phần “Create a Basic Task”. Nhập “Name” và “Description” cho ứng dụng. Ấn “Next”

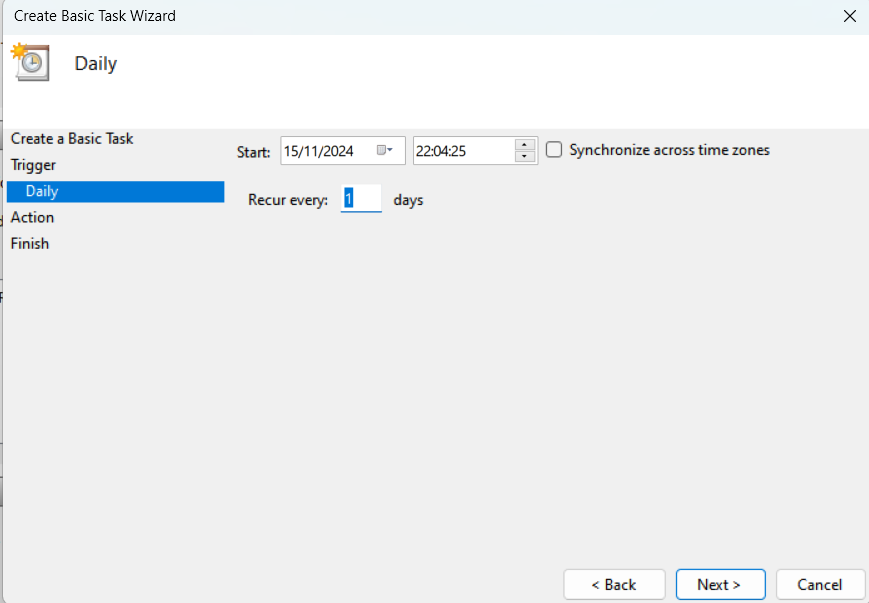


**B4:** Ở phần “Trigger”. Chọn chu kỳ lặp lại

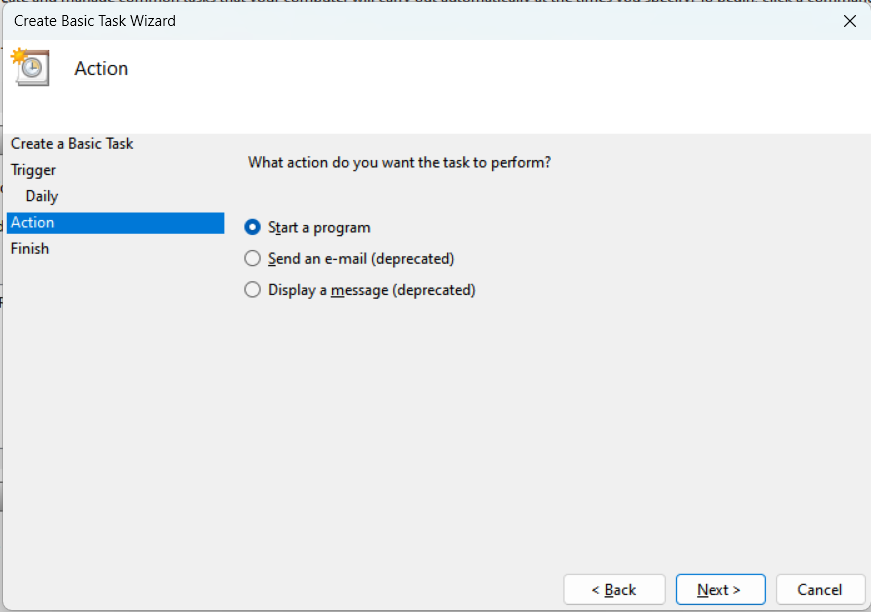
* + - “Daily”: hằng ngày
    - “Weekly”: hằng tuần
    - “Monthly”: Chạy vào các ngày cụ thể trong tháng, hoặc theo mẫu (ví dụ: ngày thứ N của tháng)
    - “One time”: Chạy một lần vào một thời điểm nhất định
* Ở đây chọn “Daily”. Ấn “Next”



**B5:** Ở phần “Trigger -> Daily”. Chọn ngày bắt đầu chạy, giờ bắt đầu và số lần lặp lại(Ví dụ chọn “Daily” thì mỗi 1 ngày 1 lần, …). Ấn “Next”

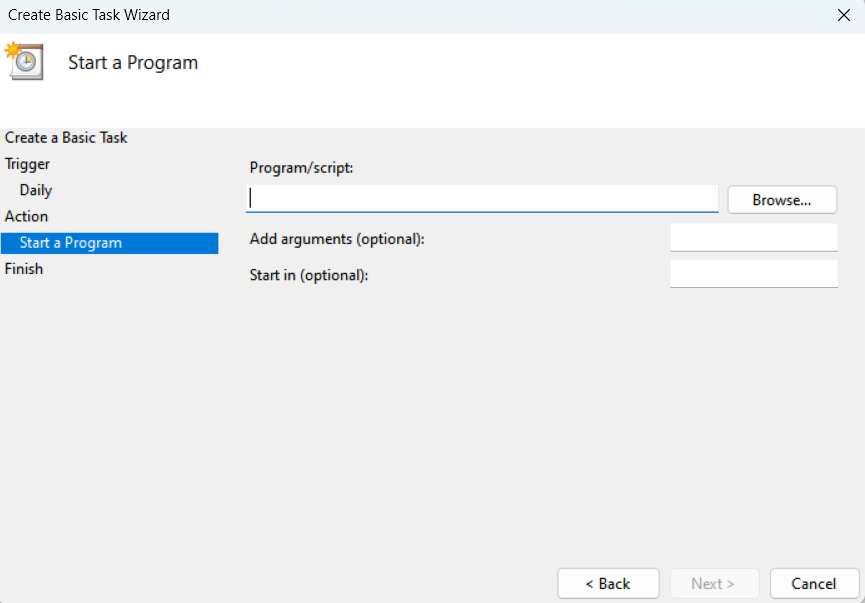
****

**B6:** Ở phần “Action”. Chọn “Start a program”. Ấn “Next”

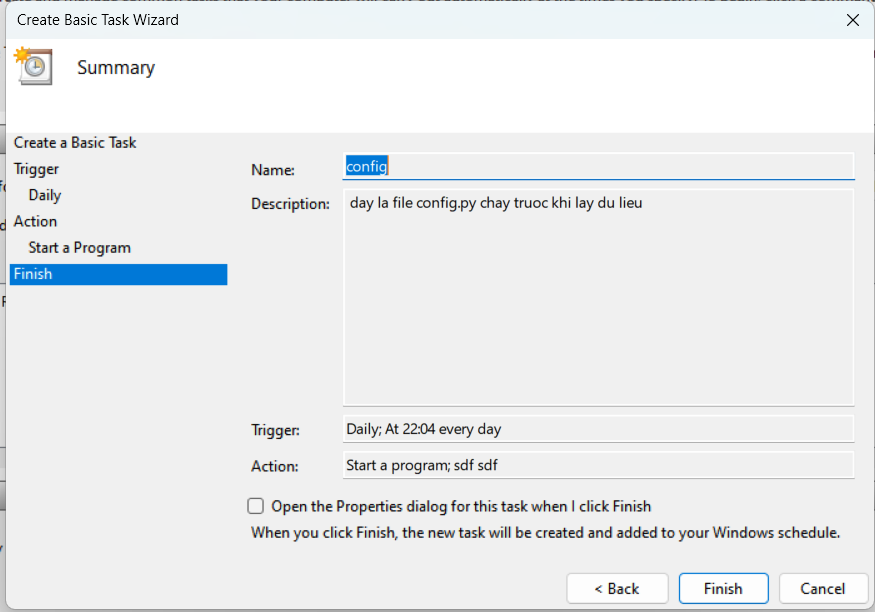
****

**B7:** Ở phần “Action -> Start a Program”

* Program/script: "D:\Python\Python312\python.exe"
* Add arguments(optional):
  + "D:\get\_data\_from\_web\config.py"
  + "D:\get\_data\_from\_web\scrape\_bike2school.py"
  + "D:\get\_data\_from\_web\scrape\_xedapgiakho.py"
* Start in (optional): D:\get\_data\_from\_web
* Ấn “Next”



**B8:** Ở phần “Finish”. Xem lại thông tin đã thiết lập. Ấn “Finish” để hoàn tất

****

**- Cài đặt task scheduler cho datawarehouse:**

Setup giống như ở trên chỉ thay đổi ở B7:

**B7:** Ở phần “Action -> Start a Program”

* Program/script: "C:\Program Files\Java\jdk-17\bin\java.exe"
* Add arguments(optional):
  + -jar "C:\Users\Nam\Desktop\loadToDatawarehouse.jar"

**- Cài đặt task scheduler cho datamart:**

Setup giống như ở trên chỉ thay đổi ở B7:

**B7:** Ở phần “Action -> Start a Program”

* Program/script: "C:\Program Files\Java\jdk-17\bin\java.exe"
* Add arguments(optional):
  + -jar "C:\Users\Nam\Desktop\loadToDatamart.jar"