# InTune Device Compliance Reporter

Contents

[InTune Device Compliance Reporter 1](#_Toc195264909)

[Introduction 1](#_Toc195264910)

[1. Initial Setup 2](#_Toc195264911)

[1.1. Local Repository 2](#_Toc195264912)

[1.2. Version Control 4](#_Toc195264913)

[2. Requirement Gathering 4](#_Toc195264914)

[2.1. Business Problem 4](#_Toc195264915)

[2.2. Objective 5](#_Toc195264916)

[2.3. Stakeholder Engagement 5](#_Toc195264917)

[2.4. Functional Requirements 6](#_Toc195264918)

[2.5. Technical Requirements 6](#_Toc195264919)

[2.6. Constraints and Assumptions 6](#_Toc195264920)

[3. Analysis and Initial Design 6](#_Toc195264921)

[3.1. Introduction 7](#_Toc195264922)

[3.2 MoSCoW 7](#_Toc195264923)

[3.3. Risks and Constraints 7](#_Toc195264924)

[3.4. Identify System Inputs, Processes, and Outputs 8](#_Toc195264925)

[3.5. High-Level Architecture 8](#_Toc195264926)

[3.6. Database Design 9](#_Toc195264927)

[3.7 Chosen Technologies 9](#_Toc195264928)

[4. Development 10](#_Toc195264929)

[4.1. Project Setup and Environment 10](#_Toc195264930)

[4.2. API Connection to Microsoft Graph 11](#_Toc195264931)

[4.3 Storing Credentials Securely with .env and .gitignore 12](#_Toc195264932)

[4.4 Data Retrieval and Authentication 14](#_Toc195264933)

[4.5. Store and Query Device Data 18](#_Toc195264934)

[4.6. Querying and Visualising the Data 22](#_Toc195264935)

[4.7. Optional Enhancements (e.g. Notifications, Scheduling, AI Summary) 23](#_Toc195264936)

## Introduction

This project, *Intune Device Compliance Reporter*, was initiated as part of the **A1.4 Apply** stage of my Level 4 Software Engineer apprenticeship with QA. It aims to automate the process of retrieving, processing, and reporting on device compliance data for Windows and iOS devices managed via Microsoft Intune- a cloud-based endpoint management solution used by Assura to secure and monitor company devices.

Currently, reporting on device compliance is a **manual, time-consuming task** performed periodically by the IT team. This project will reduce administrative overhead, improve accuracy, and provide timely insights to enable faster IT triage.

My goal is to develop a maintainable, testable, and structured Python-based tool that retrieves compliance data via the Microsoft Graph API, stores it in an SQL database, and outputs clear reports for IT analysis.

The project directly benefits my organisation while also allowing me to evidence key apprenticeship learning outcomes:

* **S1:** Writing logical and maintainable code
* **S4:** Unit testing and debugging
* **S7:** Structured problem solving
* **B4:** Collaborating with internal teams
* **B10:** Continuing professional development through real-world tooling and automation

## 1. Initial Setup

### 1.1. Local Repository

To begin, I designed a clean and modular folder structure to support separation of concerns during development. Each folder has a clearly defined role.

A screenshot of a computer

AI-generated content may be incorrect.

#### Folder-by-Folder Review

|  |  |
| --- | --- |
| **Folder** | **Purpose** |
| Database/ | Storage for my .db file (SQLite) and SQL schema files. |
| Docs/ | Great for requirement gathering notes, design diagrams, and SDLC write-up drafts. |
| Reports/ | Ideal for storing auto-generated output: CSVs, charts (from matplotlib), or summary logs. |
| Screenshots/ | Useful to evidencing progress along the way (e.g., API calls, testing, terminal outputs, debugging, stakeholder communication). |
| Scripts/ | My code will live here with separate modules for fetching, transforming, storing, etc. |
| Tests/ | Where I will write unit tests testing and debug my code. Could use pytest at a later point. |

#### README.md File Creation

A markdown-based README file was added to the root of the project as well as to each sub-folder. It includes a brief overview of the project’s purpose (or folder’s purpose), the technologies used, and how the solution aligns with my apprenticeship learning outcomes (S1, S4, S7, B4, B10). This also serves as documentation for anyone reviewing the code or assessing the project via GitHub.

A screenshot of a computer program

AI-generated content may be incorrect.

### 1.2. Version Control

To manage this project and demonstrate good version control practices, Git and GitHub were used. Git was used locally to initialise and track changes to the project, while GitHub served as the remote repository, allowing for secure cloud-based backup and versioning.

As part of this project:

* The default branch was renamed to **main**, following modern Git conventions.
* Commits were made regularly using **clear, meaningful messages** to describe changes.
* Each core folder (e.g. Docs, Scripts, Tests) includes a placeholder README.md file to ensure they are tracked from the start.

#### GitHub Repository

I created a new **private GitHub repository** titled **intune-device-compliance-reporter**. I made the repository private as it will include sensitive information such as client secrets from Azure.

#### Connecting Local Repository to GitHub

The newly created repository was linked to my local folder using Git, and the initial commit (including the full folder structure and README) was pushed successfully. This ensures all future code changes are tracked and version controlled, supporting maintainability and collaboration.

Later on in this project I also leveraged Visual Studio Code’s built-in Source Control features as it will allow for quicker visual management of changes, commits, branches, and synchronisation with GitHub, all within a single interface.

Finally, I tested it had worked by navigating back to my GitHub repo page and confirming that I was able to see the folders and my README file.

## 2. Requirement Gathering

### 2.1. Business Problem

Currently, the process of extracting device compliance data from Microsoft Intune is manual, repetitive, and inefficient, relying on IT staff to log in, navigate the portal, and export reports manually. This slows down triage efforts, limits visibility of non-compliant devices, and introduces risk through inconsistent reporting intervals.

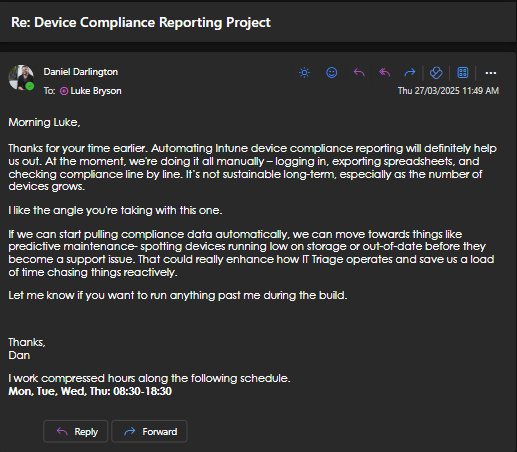
This process is time-consuming and does not support historical trend analysis or automated alerts. Due to its time-consuming nature, it is also only run on a monthly basis- I would like to change this to run weekly.

### 2.2. Objective

To design and build a Python-based tool that automates the retrieval, storage, and reporting of compliance data for Windows and iOS devices managed by Microsoft Intune, providing consistent and timely insights.

### 2.3. Stakeholder Engagement

To shape the requirements and ensure the project delivers value, I consulted with our Head of Digital & Transformation and IT Support Manager regarding reporting needs as well as discussing report structure and compliance categories.



*Figure A.1 – Email from IT Support Manager confirming manual process and value of predictive maintenance.*

A screenshot of a computer

AI-generated content may be incorrect.

*Figure A.2 – Email from Head of Digital Transformation with initial reporting requirements and strategic alignment.*

The project scope and plan were reviewed and approved by my Digital Learning Consultant (DLC) at QA.

### 2.4. Functional Requirements

|  |  |
| --- | --- |
| **Requirement** | **Description** |
| Retrieve Compliance Data | Use Microsoft Graph API to fetch device compliance data (platform, status, last check-in) |
| Store Data | Insert structured records into a SQL database for long-term reference |
| Generate Reports | Provide summaries such as number of compliant vs non-compliant devices, platform breakdown |
| Automate | Schedule periodic execution to keep data up to date |
| Alerting | Email notifications if non-compliance thresholds are exceeded |

### 2.5. Technical Requirements

|  |  |
| --- | --- |
| **Area** | **Technology / Decision** |
| Language | Python 3 |
| API Access | Microsoft Graph API with OAuth 2.0 |
| Data Storage | SQLite (lightweight, file-based SQL database) |
| Data Processing | pandas for transformation and aggregation |
| Visual Reporting | matplotlib or seaborn for charts (if time allows) |
| Security | Use .env file to store sensitive API credentials securely |

### 2.6. Constraints and Assumptions

* API access is assumed to be pre-authorised for device compliance endpoints.
* The script will run in a local or controlled internal environment with access to the Graph API.
* Delivery time is capped at ~30 hours across 4 weeks, with 7.5 hours per week.

With a clear understanding of the current challenges, requirements, and limitations, I was ready to begin the Analysis and Design planning phase.

## 3. Analysis and Initial Design

### 3.1. Introduction

Before starting development, I analysed the functional and non-functional requirements gathered from stakeholders, along with the technical constraints and project scope. This helped me structure the solution around high-priority features, identify potential blockers early, and align the development effort to Assura’s needs.

### 3.2 MoSCoW

To prioritise the project requirements, I applied the MoSCoW method, a well-known Agile technique used to categorise features based on their importance to project success:

* Must Have – Essential features required for the solution to function
* Should Have – Important features that add value but aren't critical
* Could Have – Nice-to-have additions if time and resources permit
* Won’t Have – Explicitly excluded from this phase or project scope

This helped me focus on delivering a Minimum Viable Product (MVP) while remaining realistic about time constraints and project scope.

|  |  |
| --- | --- |
| Priority | Requirement |
| Must Have | Fetch device compliance data from Microsoft Intune using the Graph API |
| Must Have | Store the data in a SQL database for later reference and reporting |
| Must Have | Generate summary reports showing compliant vs non-compliant devices |
| Should Have | Filter reports by platform (Windows/iOS) |
| Should Have | Support trend analysis by storing historical data |
| Could Have | Automated email notifications for non-compliant devices |
| Won’t Have | User interface or web dashboard (out of scope) |

### 3.3. Risks and Constraints

During analysis, I also identified potential risks and technical constraints that could affect development or project delivery. These are documented below with mitigation plans.

|  |  |  |
| --- | --- | --- |
| Risk/Constraint | Impact | Mitigation |
| API Authentication | OAuth2 token access may require approval or permissions | Request delegated access and test in a secure environment |
| Rate Limiting | Frequent API calls could be throttled | Batch requests sensibly and cache where possible |
| Data Sensitivity | Compliance data may contain identifiers | Avoid storing PII; use secure storage |
| Time Constraints | Project must be completed in 4 weeks (~30 hours) | Time-boxed features and focused Minimum Viable Product delivery |
| No UI Planned | Project is CLI and reporting-based only | Stakeholders aware this is a backend/reporting tool |

### 3.4. Identify System Inputs, Processes, and Outputs

To design the application effectively, I analysed the expected data inputs, the internal processes the system must perform, and the desired outputs for end users. This Input-Process-Output (IPO) model helps ensure the solution is well structured, maintainable, and aligned with user and business needs.

#### Inputs

* Microsoft Intune device compliance data retrieved via the Microsoft Graph API
* Authentication credentials (OAuth2 access token)
* Configuration parameters (e.g., filters, date range, environment variables)
* Historical compliance records (for trend analysis, if available)

#### Processes

* Authenticate with the Microsoft Graph API using secure credentials/ client secret
* Fetch device compliance data and parse the response
* Transform and clean the data using pandas (e.g., convert timestamps, standardise values)
* Insert the structured data into a SQL database (SQLite)
* Run summary queries to calculate compliance rates, trends, and other metrics
* Optionally trigger automated notifications for non-compliance thresholds

#### Outputs

* Tabular compliance summary reports (e.g., device ID, status, last check-in date)
* Aggregate statistics (e.g., number of non-compliant devices per OS)
* Visual charts (e.g., bar charts showing compliance over time)
* Email alerts or log messages for internal IT reporting  
  Persistent record storage for historical analysis and audit

### 3.5. High-Level Architecture

#### System Architecture Diagram

The following diagram shows a high-level overview of the components that make up the solution and how they interact. It includes the Python scripts for data retrieval and processing, Microsoft Graph API as the external data source, the SQLite database for local storage, and the reporting layer used to generate outputs for the IT team. This architecture supports a modular, testable, and scalable design that can evolve over time as our needs grow.

A blue cylinder with white text

AI-generated content may be incorrect.

#### Data Flow Diagram (DFD)

A group of blue rectangular objects with text

AI-generated content may be incorrect.This DFD shows how data moves through the system from external source (Microsoft Graph API) through internal processing, storage, and reporting workflows initiated by the relevant member of IT. The diagram reflects a clean, modular structure to support maintainability and testability.

### 3.6. Database Design

This section outlines the design of the database that will store compliance data retrieved from Microsoft Intune. A clear and logical schema is essential for ensuring that data is stored efficiently, is easy to query, and supports meaningful reporting and analysis. This database design also contributes to the maintainability and scalability of the overall system.

#### Table: device\_compliance

A single table named device\_compliance will be designed to hold the retrieved data from the Graph API. This table includes the most critical fields required for reporting on device compliance and monitoring trends over time.

#### Entity Relationship Diagram (ERD)

I decided that an ERD was not necessary as my design only involves one table and the table structure I have written (with clear fields, types, and descriptions) already meets the expectations for demonstrating structured database design.

### 3.7 Chosen Technologies

This section explains the technologies selected for this project and provides a rationale for each, alongside an evaluation of possible alternatives. The chosen tools enable me to develop a solution that is maintainable, testable, and aligned with both the project's requirements and my current technical skillset at an early stage of my apprenticeship.

#### Chosen Tools and Rationale

|  |  |  |  |
| --- | --- | --- | --- |
| **Technology** | **Reasoning** | **Alternatives** | **Trade-offs** |
| **Python** | Easy to learn, excellent for scripting and APIs, strong library ecosystem (e.g. requests, pandas) | Java, C# | Slower than compiled languages; whitespace-sensitive |
| **SQLite** | Lightweight, no server required, ideal for local data storage; integrates easily with Python | PostgreSQL, MySQL | Not ideal for large-scale or multi-user systems |
| **Microsoft Graph API** | Official method to retrieve Intune data; secure and well-documented | Manual Intune exports | Requires authentication setup and API learning curve |
| **Pandas / Matplotlib** | Efficient tools for data manipulation and visualisation in Python | Excel, Power BI, seaborn | May require additional learning for complex plotting |
| **Task Scheduler / Cron** | Enables scheduled automation of scripts | Manual script execution | Platform dependent; needs basic config |

## 4. Development

### 4.1. Project Setup and Environment

To begin development, I ensured my local environment was fully configured for Python development, version control, and structured project management.

* **IDE Used:** I used Visual Studio Code for its Python support, Git integration, and terminal access.
* **Folder Structure:** As defined in Section 1, I created a modular folder structure (Scripts, Docs, Reports, etc.) to separate concerns and support scalability.
* **Version Control:** The project was tracked using Git from the beginning. All changes were committed with meaningful messages, and the repository was hosted privately on GitHub.
* **Python Environment:** I planned to use the system-wide Python 3.9.13 installation and create a virtual environment with built-in modules as well as additional packages.

This setup phase allowed me to begin development with confidence, knowing that the core tooling was in place and aligned with the design decisions laid out in earlier sections.

### 4.2. API Connection to Microsoft Graph

The first development task was establishing a secure connection to the **Microsoft Graph API**, which is the recommended method for retrieving compliance and device management data from Microsoft Intune.

To configure API access logged into Azure, navigated to **App registrations** and selected **New registration** before entering a name and selecting **Register**. I then granted it the required Graph API permissions (**DeviceManagementManagedDevices.Read.All**)

#### Authentication Approach

Microsoft Graph uses OAuth 2.0 for secure access. I selected **Application permissions** as this will allow it to run as a background service without a signed-in user.

A screenshot of a computer

AI-generated content may be incorrect.

#### Client Secret

A client secret was generated during the Microsoft Azure app registration process to authenticate against the Microsoft Graph API. A **client secret** is essentially a password used by your application to authenticate against Microsoft’s identity platform. It's used in combination with your **Client ID** and **Tenant ID** when requesting access tokens.

The secret was set to expire in 12 months, following best practices for development environments. This expiry window balances security with convenience and ensures the application’s credentials are rotated regularly. The expiry date has been documented to prevent unexpected failures- When a client secret **expires**, your application will **fail to authenticate**, meaning:

* You will **not be able to retrieve tokens** from Microsoft Graph
* Any scripts, integrations, or automations depending on it will **stop working**
* You will need to **create a new secret** and update your code

### 4.3 Storing Credentials Securely with .env and .gitignore

#### To follow secure coding practices and protect sensitive API credentials, I implemented a .env file for secure local storage, excluded it from source control using .gitignore, and managed Python dependencies within a virtual environment. This ensured that my authentication secrets were never exposed publicly and that my project could be easily reproduced on any machine.

#### Step 1: Creating and Populating .gitignore

To avoid accidentally committing sensitive information, I ensured relevant details were excluded from Git version control:

# Python cache and virtual environment

\_\_pycache\_\_/

\*.pyc

\*.pyo

\*.pyd

.venv/

env/

# Ignore environment variables

.env

# Ignore virtual environment folder

.venv/

# OS-specific

.DS\_Store

Thumbs.db

# VS Code settings

.vscode/

# Ignore device output

cleaned\_devices.json

# Ignore screenshots

Screenshots/

# Ignore all SQLite database files

\*.db

\*.db-journal

\*.sqlite3

This was added to my .gitignore file at the root of the repository.

#### Step 2: Setting up the Virtual Environment

I followed Python best practices by isolating my project dependencies using a virtual environment. This avoids conflicts with global packages and increases reproducibility.

python -m venv .venv

To create and activate the virtual environment:

To activate it in PowerShell:

powershell -ExecutionPolicy Bypass -File .\.venv\Scripts\Activate.ps1

#### Step 3: Install required packages inside the virtual environment

The next step was to install my required packages for the project using pip.

pip install python-dotenv requests pandas matplotlib sqlite-utils

These packages were chosen because:

* python-dotenv loads environment variables securely from .env.
* requests handles HTTP requests to the Microsoft Graph API.
* pandas and matplotlib will later be used to manipulate and visualise compliance data.
* sqlite-utils provides extra utilities for working with SQLite databases via Python.

#### Step 4: Freeze Dependencies for Reproducibility

I generated a requirements.txt file so that I can easily recreate the same environment:

pip freeze > requirements.txt

python-dotenv

requests

pandas

matplotlib

sqlite-utils

Contents of the file:

#### 

#### Step 5: Create a .env File to Store API Credentials

CLIENT\_ID=my-client-id-here

TENANT\_ID=my-tenant-id-here

CLIENT\_SECRET\_ID=my-client-secret-id-here

CLIENT\_SECRET\_VALUE=my-client-secret-value-here

This file was used to store key-value pairs for CLIENT\_ID, TENANT\_ID, CLIENT\_SECRET\_ID and CLIENT\_SECRET\_VALUE, which are loaded securely using the python-dotenv package.

### 4.4 Data Retrieval and Authentication

To securely access the Microsoft Graph API and retrieve device compliance data from Intune, I implemented a modular Python-based authentication and data retrieval workflow. This section outlines how I developed, structured, and validated the components responsible for secure API access and data extraction.

#### Step 1: Authentication via auth.py

To access Microsoft Graph securely using the client credentials flow, I created a Python script (auth.py) responsible for authenticating against the Microsoft Identity Platform.   
  
This script performs the following:

* Loads credentials from the .env file using python-dotenv

# auth.py

import os

from dotenv import load\_dotenv

import requests

# Load environment variables from .env file

load\_dotenv()

* Reads TENANT\_ID, CLIENT\_ID, and CLIENT\_SECRET\_VALUE from environment variables

# Get credentials from environment variables

TENANT\_ID = os.getenv("TENANT\_ID")

CLIENT\_ID = os.getenv("CLIENT\_ID")

CLIENT\_SECRET\_VALUE = os.getenv("CLIENT\_SECRET\_VALUE")

* Next, I created a get\_access\_token() function in auth.py which sends a POST request to the Microsoft Identity token endpoint to retrieve a bearer token

# Microsoft Identity token endpoint

TOKEN\_URL = f"https://login.microsoftonline.com/{TENANT\_ID}/oauth2/v2.0/token"

def get\_access\_token():

    """

    Retrieves an OAuth2 access token from Microsoft Identity Platform

    using the client credentials flow.

    """

    headers = {

        "Content-Type": "application/x-www-form-urlencoded"

    }

    body = {

        "client\_id": CLIENT\_ID,

        "scope": "https://graph.microsoft.com/.default",

        "client\_secret": CLIENT\_SECRET\_VALUE,

        "grant\_type": "client\_credentials"

    }

    response = requests.post(TOKEN\_URL, headers=headers, data=body)

* Then parsing and returning the access\_token for use in subsequent API calls

    if response.status\_code == 200:

        token = response.json().get("access\_token")

        print("Access token retrieved.")

        return token

    else:

        print(" Failed to retrieve access token.")

        print(response.text)

        return None

if \_\_name\_\_ == "\_\_main\_\_":

    token = get\_access\_token()

    print(token)

Step 2: Device Data Retrieval (fetch\_devices.py)Once authenticated, I developed fetch\_devices.py to retrieve data on all registered Intune devices. Here is a breakdown of the script:

* Firstly, I imported and called the get\_access\_token() function from auth.py

# fetch\_devices.py

import requests

import json

from auth import get\_access\_token  # Reuse token function from auth.py

* The raw response from Microsoft Graph contains many fields, so I created transform\_device() to extract only the required attributes. This includes device name, operating system (OS), compliance status, OS version, total and free storage (in bytes), and last check-in date. It also calculated values such as free storage percentage and compliance status.

# Transform and clean the raw device data

def transform\_device(device):

    """Cleans and extracts only the fields needed for reporting."""

    try:

        total = int(device.get("totalStorageSpaceInBytes") or 0)

        free = int(device.get("freeStorageSpaceInBytes") or 0)

        free\_pct = (free / total \* 100) if total else None

    except (ValueError, TypeError, ZeroDivisionError):

        total, free, free\_pct = 0, 0, None

    return {

        "deviceId": device.get("id"),

        "deviceName": device.get("deviceName") or "unknown",

        "emailAddress": device.get("emailAddress") or "unknown",

        "model": device.get("model"),

        "operatingSystem": device.get("operatingSystem"),

        "osVersion": device.get("osVersion"),

        "complianceState": device.get("complianceState"),

        "isCompliant": device.get("complianceState") == "compliant",

        "totalStorageBytes": total,

        "freeStorageBytes": free,

        "freeStoragePct": round(free\_pct, 2) if free\_pct is not None else None,

        "lowStorage": free\_pct is not None and free\_pct < 10,

        "lastSyncDateTime": device.get("lastSyncDateTime")

    }

* After that, I sent a POST request to the Microsoft Graph API's /deviceManagement/managedDevices endpoint and stored devices in a list.

# Step 1: Get access token

access\_token = get\_access\_token()

if access\_token:

    # Step 2: Define the Microsoft Graph API endpoint

    endpoint = "https://graph.microsoft.com/v1.0/deviceManagement/managedDevices"

    headers = {

        "Authorization": f"Bearer {access\_token}",

        "Content-Type": "application/json"

    }

    all\_devices = []  # Store all device records here

* The @odata.nextLink property was also included in the script to paginate responses- this is also discussed in the next step of this section.

    # Step 3: Handle pagination to fetch all devices

    while endpoint:

        response = requests.get(endpoint, headers=headers)

        if response.status\_code == 200:

            data = response.json()

            devices = data.get("value", [])

            all\_devices.extend(devices)

            # Continue if there's another page of results

            endpoint = data.get("@odata.nextLink", None)

        else:

            print("❌ Failed to retrieve device data:", response.status\_code)

            print(response.text)

            break

* Finally, the program outputs the total device count, transforms and cleans the data and saves the cleaned data to cleaned\_devices.JSON.

    # Step 4: Output total device count

    print(f"✅ Retrieved {len(all\_devices)} devices")

    # Step 5: Transform and clean the data

    cleaned\_devices = [transform\_device(device) for device in all\_devices]

    # Step 6: Save cleaned data to JSON

    with open("cleaned\_devices.json", "w") as f:

        json.dump(cleaned\_devices, f, indent=2)

    print("✅ Cleaned device data saved to cleaned\_devices.json")

else:

    print("❌ Authentication failed. Cannot retrieve device data.")

* The program also prints key status messages to confirm successful data acquisition and storage or to confirm it was unable to authenticate.

#### Step 3: Output Verification and Pagination Handling

Upon initial testing, only a subset of devices was retrieved. By inspecting the raw JSON output, I identified the need for pagination handling- a common scenario when querying large datasets from Graph. I updated the script to recursively follow the @odata.nextLink property and retrieve all available pages, ultimately capturing a full dataset of 180 devices.

All device data was successfully exported into cleaned\_devices.json, stored in the root of the project directory, for use in later processing and visualisation stages.

#### Step 4: Reflections and Testing Considerations

Although detailed testing and debugging are discussed in Section 5, I used frequent print() statements and controlled output preview such as:

print(json.dumps(devices[:3]))

During development this helped me to isolate issues such as partial responses, empty tokens, or indentation errors which allowed for rapid identification and resolution of logical flaws.

### 4.5. Store and Query Device Data

With the device compliance data cleaned and transformed in Section 4.4, the next step was to store it in a structured format to enable filtering, reporting, and future data queries. I selected SQLite as the storage solution due to its simplicity, portability, and suitability for a lightweight, read-intensive reporting tool.

#### Step 1: Selecting a Database Solution

SQLite was chosen for the following reasons:

* No server configuration required- ideal for local development or small-scale automation.
* Lightweight and embeddable- integrates directly with Python without needing a running service.
* Supports SQL syntax- allowing the use of powerful queries, indexes, and joins if required in future enhancements.

This aligns with the project's scope and reduces complexity. It also makes it easier to port the logic to a cloud-native database later if scalability is required.

#### Step 2: Designing the Table Schema

I designed a schema to match the fields prepared in cleaned\_devices.json, which were the result of the transformation step. These fields provide a foundation for compliance reporting across both Windows and iOS devices.

The table was named device\_compliance and includes the following fields:

|  |  |  |
| --- | --- | --- |
| **Column** | **Data Type** | **Description** |
| deviceId | TEXT | Unique device identifier (used as primary key) |
| emailAddress | TEXT | User’s email address |
| Model | TEXT | Device model |
| deviceName | TEXT | User-friendly device name |
| operatingSystem | TEXT | OS type (e.g. Windows, iOS) |
| osVersion | TEXT | Operating system version |
| complianceState | TEXT | Raw compliance state string from Intune |
| isCompliant | BOOLEAN | Boolean flag derived from complianceState |
| totalStorageBytes | INTEGER | Total storage reported by the device |
| freeStorageBytes | INTEGER | Free space reported by the device |
| freeStoragePct | REAL | Calculated % of storage space remaining |
| lowStorage | BOOLEAN | True if free space < 10% |
| lastSyncDateTime | TEXT | Last reported sync date from InTune |

The primary key was set to deviceId as each record in Intune is uniquely identified by this field.

#### Step 3: Writing Cleaned Data to SQLite

To produce the transformed device data for reporting, I developed a Python script called insert\_into\_sqlite.py. Th*e* script loads the structured device records from cleaned\_devices.json and writes them into a local SQLite database (intune\_devices.db).

The script performs the following steps:

* Connects to SQLite and creates the database file intune\_devices.db if it doesn't exist.

# insert\_into\_sqlite.py

import json

import sqlite3

# Load the cleaned device data

with open("cleaned\_devices.json", "r") as file:

    devices = json.load(file)

# Connect to the SQLite database (or create it)

conn = sqlite3.connect("intune\_devices.db")

cursor = conn.cursor()

* Defines the table schema (device\_compliance) and creates it if not already present.

# Create the table (if it doesn't already exist)

cursor.execute("""

CREATE TABLE IF NOT EXISTS device\_compliance (

    deviceId TEXT PRIMARY KEY,

    deviceName TEXT,

    emailAddress TEXT,

    model TEXT,

    operatingSystem TEXT,

    osVersion TEXT,

    complianceState TEXT,

    isCompliant BOOLEAN,

    totalStorageBytes INTEGER,

    freeStorageBytes INTEGER,

    freeStoragePct REAL,

    lowStorage BOOLEAN,

    lastSyncDateTime TEXT

);

""")

* Reads the cleaned JSON file and iteratively inserts each device record into the table using INSERT OR REPLACE, ensuring no duplicates on deviceName.

# Insert each device one at a time

for device in devices:

    cursor.execute(f"""

    INSERT OR REPLACE INTO device\_compliance (

        deviceId, deviceName, emailAddress, model, operatingSystem, osVersion,

        complianceState, isCompliant, totalStorageBytes, freeStorageBytes,

        freeStoragePct, lowStorage, lastSyncDateTime

    ) VALUES (

        ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?

    );

    """, (

        device["deviceId"],

        device["deviceName"],

        device["emailAddress"],

        device["model"],

        device["operatingSystem"],

        device["osVersion"],

        device["complianceState"],

        device["isCompliant"],

        device["totalStorageBytes"],

        device["freeStorageBytes"],

        device["freeStoragePct"],

        device["lowStorage"],

        device["lastSyncDateTime"]

    ))

* Commits the changes and closes the database connection once all records are processed, whilst printing a message to confirm the number of records inserted into device\_compliance.

# Save changes and close the connection

conn.commit()

conn.close()

print(f"✅ {len(devices)} records inserted into device\_compliance table.")

I then verified that intune\_devices.db had been created by viewing it with Visual Studio Code’s SQLite Viewer Extension.

A screenshot of a computer

AI-generated content may be incorrect.

Choosing SQLite kept the solution self-contained and aligned with the constraints of a single-user, internal reporting tool. Storing the data prepares the project for integration with visualisation tools or automation layers.

### 4.6. Querying and Visualising the Data

The goal of this stage was to extract actionable compliance insights from the device dataset and present them in an accessible and visual manner using SQL queries and data visualisations.

I used the SQLite extension within VS Code to write and run queries against intune\_devices.db.

-- Query to get all data

SELECT

    deviceId AS "Device ID",

    deviceName AS "Device Name",

    emailAddress AS "User",

    operatingSystem AS "Operating System",

    osVersion AS "OS Version",

    complianceState AS "Compliance State",

    isCompliant AS "Is Compliant",

    totalStorageBytes AS "Total Storage (Bytes)",

    freeStorageBytes AS "Free Storage (Bytes)",

    freeStoragePct AS "Free Storage (%)",

    lowStorage AS "Low Storage",

    lastSyncDateTime AS "Last Sync Date"

FROM device\_compliance;

-- Query to get Windows devices with low storage that have checked in within the last 3 months

SELECT

    emailAddress AS "User",

    deviceName AS "Device Name",

    freeStoragePct AS "Free Storage (%)"

FROM device\_compliance

WHERE lowStorage = 1 AND operatingSystem = 'Windows' AND date(lastSyncDateTime) >= date('now', '-3 months');

-- Query to get Windows devices that are not compliant and have checked in within the last 3 months

SELECT

    emailAddress AS "User",

    deviceName AS "Device Name",

    Model AS "Model"

FROM device\_compliance

WHERE isCompliant = 0 AND operatingSystem = 'Windows' AND date(lastSyncDateTime) >= date('now', '-3 months');

-- Query to get iOS devices with low storage that have checked in within the last 3 months

SELECT

    emailAddress AS "User",

    deviceName AS "Device Name",

    freeStoragePct AS "Free Storage (%)"

FROM device\_compliance

WHERE model != "iPhone 13" AND lowStorage = 0 AND operatingSystem = 'iOS' AND date(lastSyncDateTime) >= date('now', '-3 months');

-- Query to get iOS devices that are not compliant and have checked in within the last 3 months

SELECT

    emailAddress AS "User",

    deviceName AS "Device Name",

    Model AS "Model"

FROM device\_compliance

WHERE isCompliant = 0 AND operatingSystem = 'iOS' AND date(lastSyncDateTime) >= date('now', '-3 months');

### 4.7. Optional Enhancements (e.g. Notifications, Scheduling, AI Summary)

Testing/ Debugging:  
Indentation Errors

Packages not installed

Incorrect amount of devices listed- deviceName instead of deviceId selected as PRIMARY KEY

Improvements:

**Automatically Pull Latest Versions (Advanced/Future Work)**

This is **out of scope for now**, but good to mention in your **Reflection or Future Enhancements** section:

* **Microsoft releases latest Windows OS builds** via public update history:  
  <https://learn.microsoft.com/en-us/windows/release-health/>
* **Apple lists iOS versions here**:  
  <https://support.apple.com/en-us/HT201222>

You could theoretically scrape or API-poll those in a future version of your tool.

**1. Anomaly Detection for Low-Storage Devices**

**Overview**

* **Goal**: Detect “outliers” among devices based on freeStoragePct or OS version.
* **Method**: A simple **k-means clustering** or **Z-score** approach to see which devices differ significantly from the group (e.g., extremely low storage).
* **Libraries**: pandas, numpy

**2. Predicting Future Non-Compliance (Logistic Regression)**

**Overview**

* **Goal**: Predict if a device might become **noncompliant** soon, based on patterns in OS version, storage usage, or last check-in time.
* **Method**: A basic **logistic regression** with historical data for each device. If you keep snapshots over time in your database, you can label examples as “stayed compliant” or “became noncompliant.”
* **Libraries**: pandas, scikit-learn

**Why It Might Be More Advanced**

* Requires **labelled historical data** (did the device become noncompliant later?).
* If you only have a single snapshot so far, you’d need some assumptions or partial data.

**(If You Want to Summarise Compliance Text)**

* **LiteLLM** or other minimal Large Language Model wrappers can do:
  + Summaries of device compliance state.
  + Natural language reports (e.g., “We have 18 Windows devices out of compliance. iOS devices show 30% free storage on average.”)

**Example**: Use a library to prompt an LLM to generate a summary from your numeric stats. This is less about ML in the classical sense (e.g. scikit-learn) and more about **language-based summarisation**.

**But** be aware that production LLM usage can raise data privacy concerns, especially with device-level info.