**Using MS-Excel with ODBC Database Tables**

**to do Updates and Reporting**

|  |  |
| --- | --- |
| **SCHOOL:** | Algonquin College |
| **COURSE NUMBER:** | CST8276\_030 Advanced Database Topics |
| **PROFESSOR:** | Amal Ahmed Anda |
| **PREPARED BY:** | Group 3 |
| **TEAM MEMBER:** | Cao, Dandan [cao00102@algonquinlive.com](mailto:cao00102@algonquinlive.com)  Song, Xiao [song0137@algonquinlive.com](mailto:song0137@algonquinlive.com)  Chang, lijun [chan0529@algonquinlive.com](mailto:chan0529@algonquinlive.com)  Li, Jinwei [li000464@algonquinlive.com](mailto:li000464@algonquinlive.com)  Yang, Haiyan [yang0348@algonquinlive.com](mailto:yang0348@algonquinlive.com)  Huang, Ziyang [huan0269@algonquinlive.com](mailto:huan0269@algonquinlive.com) |
| **DATE OF SUBMISSION:** | June 15, 2024 |

Table of Contents

[Introduction 2](#_Toc169387090)

[Topic Description Paragraphs 2](#_Toc169387091)

[Why we chose this topic 3](#_Toc169387092)

[Problem Description 5](#_Toc169387093)

[Who 5](#_Toc169387094)

[What 5](#_Toc169387095)

[Where 6](#_Toc169387096)

[When 7](#_Toc169387097)

[Why 7](#_Toc169387098)

[How 8](#_Toc169387099)

[Solution Demonstration Description 9](#_Toc169387100)

[Introduction of solution demonstration 9](#_Toc169387101)

[Data Source 9](#_Toc169387102)

[Solution Components 9](#_Toc169387103)

[Scripts 10](#_Toc169387104)

[Key Steps 10](#_Toc169387105)

[Risks 12](#_Toc169387106)

[Work Plan 14](#_Toc169387107)

[Table with components and deliverables 14](#_Toc169387108)

[Reference 17](#_Toc169387109)

# Introduction

## Topic Description Paragraphs

Microsoft Excel is an essential tool for daily office work due to its user-friendly interface and powerful features. It is one of the most widely used spreadsheet application for data-related tasks. Excel's capabilities and scalability make it useful for tasks such as data collection, cleaning, cataloging, and visualization. It plays a crucial role in various data-related tasks and can be integrated with databases through Open Database Connectivity (ODBC). This allows users to establish a direct connection between Excel and different database systems, enabling real-time data loading to ensure the timeliness and completeness of data sets.

By utilizing ODBC, Excel can serve as a robust front-end interface for interacting with back-end databases, facilitating seamless data exchange and ensuring that the information in Excel is always current and consistent with the database. Users can also execute SQL queries, retrieve data sets, update records, and generate detailed reports directly from Excel, streamlining workflows and enhancing productivity [1][2].

Additionally, Excel's powerful reporting tools, such as pivot tables, charts, and data models, can be linked to databases through ODBC. This allows for the creation of interactive and up-to-date reports, making it easier to monitor and analyze data [3]. By integrating Excel with databases via ODBC, users can efficiently perform data updates and reporting tasks [4].

## Why we chose this topic

We have chosen the topic of Using MS-Excel with ODBC Database Tables to do Updates and Reporting due to its practical relevance and wide applicability. Excel is a commonly used tool for data collection and management, with a user-friendly interface and powerful features that make it essential for various data tasks. As a Database Administrator, its functionality, integrating Excel with databases through ODBC, can be greatly enhanced, allowing for real-time data loading and ensuring the accuracy and timeliness of the data set. This integration also simplifies workflows significantly.

In both academic projects and co-op work, we have encountered situations where we needed to load raw data from Excel into databases, making this a highly applicable technology. Therefore, this topic not only deepens our understanding of database management and data analysis, but also prepares us to handle practical challenges in a variety of professional settings.

# Problem Description

## Who

* **Office workers:** Employees in administrative roles responsible for maintaining data integrity, accuracy, and accessibility within an organization.
* **Data analysts:** Professionals who rely on accurate data to perform analyses that support business decisions.
* **Business managers:** Decision-makers who depend on up-to-date reports to make informed strategic and operational decisions.
* **IT professionals:** Individuals tasked with ensuring the integrity, security, and proper functioning of data systems and infrastructure [5].

## What

The problem is the inefficiency and potential for errors in data management and reporting due to the lack of real-time integration between Microsoft Excel and databases. Specifically:

* Users often work with outdated data.
* Manual data entry leads to errors.
* Data synchronization is time-consuming.
* Generating accurate and timely reports is challenging.

The data aspect of the issue involves ensuring that data collected and managed in Excel is accurately transferred, synchronized, and updated within the database systems. Data administrators must ensure that the data pipelines work smoothly, which includes handling data consistency, managing data loads, and implementing data quality control measures. This responsibility is crucial to maintain the reliability and usability of the data across various systems [6].

## Where

This problem is predominantly client-side, where DBAs need to establish and maintain efficient connections between Excel and backend databases using ODBC. It also involves server-side considerations, such as ensuring the database server can handle real-time data requests and replication processes to prevent discrepancies between the database and Excel.

This problem is encountered in various environments, including:

* **Corporate offices:** Where data management is crucial for daily operations.
* **Financial institutions:** Where real-time data is essential for transactions and reporting.
* **Research facilities:** Where data accuracy and timely updates are critical.
* **Healthcare organizations:** Where patient data management requires up-to-date information.
* **Educational institutions:** Where student records and administrative data need to be current.

## When

Timing aspects include scheduling regular data updates and synchronizations between Excel and the database to ensure data timeliness. DBAs need to manage periodic data loads, real-time data retrieval during operational hours, and scheduled report generations to ensure data is always current and accurate.

The issue arises:

* During daily data entry and updates.
* When preparing for regular reporting cycles, such as daily, weekly, monthly, or quarterly reports.
* During strategic planning sessions requiring real-time data insights.
* Whenever quick decision-making is based on data analysis.

## Why

DBAs must adhere to various laws and regulations, and other industry-specific data protection standards to ensure data privacy, security, and integrity. This adherence is to avoid legal ramifications, or compromised decision-making.

The lack of real-time data integration between Excel and databases causes several problems:

* Outdated data: Manually entered or periodically imported data can become outdated quickly.
* Errors: Manual data entry and transfers increase the risk of mistakes.
* Inefficiency: Manual processes are time-consuming and labor-intensive.
* Inconsistent data: Different versions of data may exist, leading to discrepancies in reports and analyses.
* Decision-making impact: Outdated or inaccurate data hampers effective decision-making

## How

Issues arise when there is a failure in the data exchange process between Excel and the database, leading to data inconsistencies and inefficiencies. DBAs can resolve these issues by implementing automated data synchronization, regular testing, monitoring processes, and logging and reviewing.

This integration enables:

* Real-time data access: Ensuring that data in Excel is always current by directly connecting to the database.
* Data updates: Users can update database records directly from Excel.
* SQL execution: Users can run SQL queries within Excel to retrieve and manipulate data.
* Reduced errors: Minimizing manual data entry reduces the likelihood of errors.
* Improved efficiency: Streamlining data management processes saves time and resources [7].

# Solution Demonstration Description

## Introduction of solution demonstration

In the solution demo, we intend to create an application displaying COVID-19 statistics with MS-Excel as the front end for data manipulation and visualization and Oracle as the backend database, connected through ODBC (Open Database Connection).

The presentation will cover the following key components and steps:

### Data Source

We will use downloadable CSV files for the data on COVID-19 in Canada provided by Canadian government [8].

### Solution Components

* MS-Excel
* ODBC Driver
* Oracle Database 19c

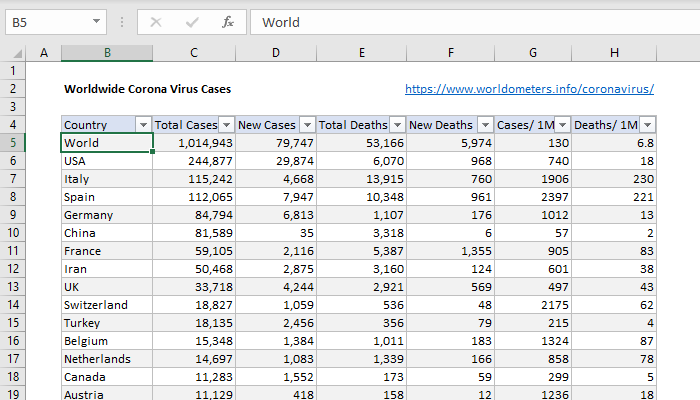
### Scripts

* **Database initialization script:** SQL script to create the necessary database, define the structure of tables based on the format of source file.
* **Data entry script:** Use SQL script to import COVID-19 data downloaded from related health organizations to the database.
* **ODBC Configuration Script:** Instructions or scripts that configure the ODBC driver and data source name (DSN) to ensure that MS-Excel can connect to the database.

### Key Steps

* **Create Database environment:** Install and initialize Oracle database. Install Oracle ODBC driver.
* **Deciding Table Structure:** Decide the basic structure of tables used to store COVID-19 statistics, such as daily cases, recoveries, deaths, and vaccination status. The table structure will be decided by our source data. Example shown below in Figure 1 [9].

**Worldwide Corona Virus Cases**



*Figure1: Covid statistic table structure example*

* **Data Entry:** Import data into database for demonstration purposes.
* **Set up ODBC connectivity:** Set up ODBC connection with Oracle Database in Excel using Data Connection Wizard.
* **Use Excel as frontend for display and manipulation:** Use Excel features to create necessary interfaces for data analysis, visualization and manipulation.

**Daily New Confirmed COVID-19 Cases per Million People**

A screenshot of a graph

Description automatically generated

*Figure 2: Covid cases display example [10]*

## Risks

**Known risks ranges from technical issues to project management mishaps during the development process:**

* **Data Source:** Choosing wrong data sources and affecting the validity of data; Potential difficulty in importing data to database
* **Connectivity between MS-Excel, ODBC Driver, and Oracle database:** Possible need for data format conversion between MS-Excel and Database, ODBC connection could fail or not synchronized
* **Software Limitation**: if the dataset is too large for Excel, performance could deteriorate.
* **Security:** Without proper security setup, system integrity could be compromised due to incidents.
* **Project Management:** Group miscommunication, inadequate planning and accidents like injury might lead to scope creep, delay or undelivered parts.
* **Failure in Collaboration and Knowledge Sharing:** Lack of documentation and communication could cause difficulties in collaborative working on the project.
* **Quality Assurance:** Insufficient testing and unexpected technical issues may cause delays. Counterintuitive user interface could be hard to use.

# Work Plan

## Table with components and deliverables

|  |  |  |  |
| --- | --- | --- | --- |
| **Component/**  **Deliverable Name** | **Time Period** | **Hours Excepted** | **Responsible Individual** |
| Initial Proposal | Week3-4 | 10 | Cao, Dandan  Li, Jinwei  Yang, Haiyan  Huang, Ziyang  Song, Xiao  Chang, Lijun |
| Final Proposal | Week5 | 10 | Cao, Dandan  Li, Jinwei  Yang, Haiyan  Huang, Ziyang  Song, Xiao  Chang, Lijun |
| Database Environment Setup | Week6 | 20 | Cao, Dandan  Li, Jinwei  Yang, Haiyan  Huang, Ziyang  Song, Xiao  Chang, Lijun |
| Search data related to COVID-19 | Week7 | 20 | Cao, Dandan  Li, Jinwei  Yang, Haiyan  Huang, Ziyang  Song, Xiao  Chang, Lijun |
| Generate statistics & graphic reports | Week8 | 30 | Cao, Dandan  Li, Jinwei  Yang, Haiyan  Huang, Ziyang  Song, Xiao  Chang, Lijun |
| Testing and Validation | Week9-10 | 25 | Cao, Dandan  Li, Jinwei  Yang, Haiyan  Huang, Ziyang  Song, Xiao  Chang, Lijun |
| Report Writing | Week11 | 10 | Cao, Dandan  Li, Jinwei  Yang, Haiyan  Huang, Ziyang  Song, Xiao  Chang, Lijun |
| Presentation | Week12-13 | 10 | Cao, Dandan  Li, Jinwei  Yang, Haiyan  Huang, Ziyang  Song, Xiao  Chang, Lijun |

# Reference

[1]. Phil Factor, "Getting Data between Excel and SQL Server using ODBC - Simple Talk," Redgate Hub, Aug. 16, 2013. [Online]. Available: https://www.red-gate.com/simple-talk/databases/sql-server/database-administration-sql-server/getting-data-between-excel-and-sql-server-using-odbc/ [Accessed: May 23, 2024]

[2]. "Use Microsoft Query to retrieve external data," Microsoft Support. Available: <https://support.microsoft.com/en-us/office/use-microsoft-query-to-retrieve-external-data-42a2ea18-44d9-40b3-9c38-4c62f252da2e> [Accessed: May 23, 2024]

[3]. "How to Use Excel to Script Database Table Updates," Segue Technologies, Dec. 3, 2013. [Online]. Available: <https://www.seguetech.com/excel-script-database-table-updates/> [Accessed: May 23, 2024]

[4]. "Connect to an ODBC Data Source (SQL Server Import and Export Wizard)," Microsoft Learn, Sep. 1, 2023. [Online]. Available: <https://learn.microsoft.com/en-us/sql/integration-services/import-export-data/connect-to-an-odbc-data-source-sql-server-import-and-export-wizard?view=sql-server-ver16> [Accessed: May 23, 2024]

[5]. A. Lancaster, “Top 10 uses of Microsoft Excel in business,” Mallory, Available: <https://mallory.com.au/cm/ms-excel-uses-business-workplace/> [ Accessed Jun. 14, 2024].

[6]. “What is ODBC? - open database connectivity - definition,” Codeless Platforms, Available: <https://www.codelessplatforms.com/docs/knowledge-base/glossary-of-terms/what-is-odbc/> [Accessed Jun. 14, 2024].

[7] “Microsoft,” Microsoft Support, Available: [https://support.microsoft.com/en-us/office/using-access-or-excel-to-manage-your-data-09576147-47d1-4c6f-9312-e825227fcaea](https://support.microsoft.com/en-us/office/using-access-or-excel-to-manage-your-data-09576147-47d1-4c6f-9312-e825227fcaea%20) [Accessed Jun. 14, 2024].

[8] T. B. of C. Secretariat and T. B. of C. Secretariat, “Public Health Infobase - Data on COVID-19 in Canada - Open Government Portal,” open.canada.ca. Available: https://open.canada.ca/data/en/dataset/261c32ab-4cfd-4f81-9dea-7b64065690dc ‌[Accessed Jun. 14, 2024].

[9] H. Ritchie, E. Mathieu, L. Rodés-Guirao, C. Appel, C. Giattino, J. Hasell, B. Macdonald, S. Dattani, D. Beltekian, and M. Roser, “Coronavirus Pandemic (COVID-19),” Our World in Data. Available: <https://ourworldindata.org/covid-cases> [Accessed May 24, 2024].

[10] Exceljt, “Coronavirus data for Excel,” exceljet.net. Available: <https://exceljet.net/articles/download-coronavirus-data-to-excel> [Accessed Jun. 14, 2024].