2023

Data Management and Reporting

MADHUKAR PRASAD
PRIYANKA BEJGAM
OLUBUNMI MERCY
SRAVANI VEMULA
SHUBHASHREE SAHOO

Executive Summary

Bottom Line Upfront:
As a complimentary service provided by Enterprise Systems, it is essential to demonstrate the value of the services being offered to students and faculty around the globe. By effectively managing and reporting on the data generated by requests, Enterprise Systems can provide insights into how its services are being used, what benefits they are providing, and where improvements can be made. This can help to build trust and establish a positive reputation for Enterprise Systems as a provider of valuable academic enterprise curriculum and systems.
The project aimed to address the challenge of the complexity in the collection and usage of data generated from requests raised by global university faculty to access Enterprise Systems' services and track student accounts. To address these concerns, our goal was to re-design the storage and management process for data generated from the Enterprise Systems requests form, reducing visualization creation time from 3 hours per week to under a minute.

Summarized Discussion Points and Why:
A complete redesign of the process was proposed to establish a more efficient system and overcome challenges.

- The new system enables tracking of student users using the service along with the request IDs.

The previous system lacked tracking of student accounts, which resulted in an inability to determine effective utilization and identify areas for improvement/expansion.

- We created a single table for extracting data used in visualizations.

When data is spread out across multiple tables, it can become challenging to retrieve the required information, as one would need to search and analyze data from different sources.

- The redesign of the data storage and management process allowed for improved reporting and analysis capabilities.

Reducing the time to create visualization from 3 hours to 1 minute, identify areas for improvement/expansion, better detailed tracking system and usage analytics.

Conclusions:
In conclusion, the project aimed to address the challenges faced by the Enterprise Systems team at the University of Arkansas in managing and analyzing the data generated from faculty requests and student accounts. By re-designing the storage and management process for this data, we were able to significantly reduce the time needed for creating visualizations from 3 hours weekly to less than a minute. Our solution involved storing the data into a single table, which enables fast and effective analysis of collected data using visualizations. Overall, our project provides a simple and standardized method for storing and analyzing data, which will help the Enterprise Systems team make more informed decisions and improve the services provided to faculty members and students.

CONTENTS

1. Introduction of the Client

Enterprise Systems is a group within the IS department of UA whose mission is to be the premier worldwide source of academic enterprise curriculum and systems. They provide access to platforms like Microsoft, SAP, IBM etc. and datasets to students and faculty worldwide to facilitate teaching and research. They support education and research efforts at UA and around the world. The programs produce globally competitive students for teaching and research. The current process at the university for fulfilment of request takes on average 15 days from the time of request. We aim to improve the service operations of the university so that they can better serve the students, faculty, and staff that they provide services to. Streamlining processes and reducing wait times to create a more positive and productive learning environment for everyone involved.

2. Background of the client organization and IS Department

The Enterprise Systems group at the University of Arkansas manages several enterprise systems, including SAP S/4 HANA, Teradata, SAS Viya, and Microsoft SQL Server.
The Walton IT department, which is a part of the University, is responsible for managing these enterprise systems and receives requests from faculty members of universities worldwide for access to them. After receiving these requests, the Walton IT department forwards them to the Central IT department for processing. Their primary goal is to enhance their services and expand their reach to universities globally. Additionally, they focus on fostering and maintaining strong relationships with other universities.

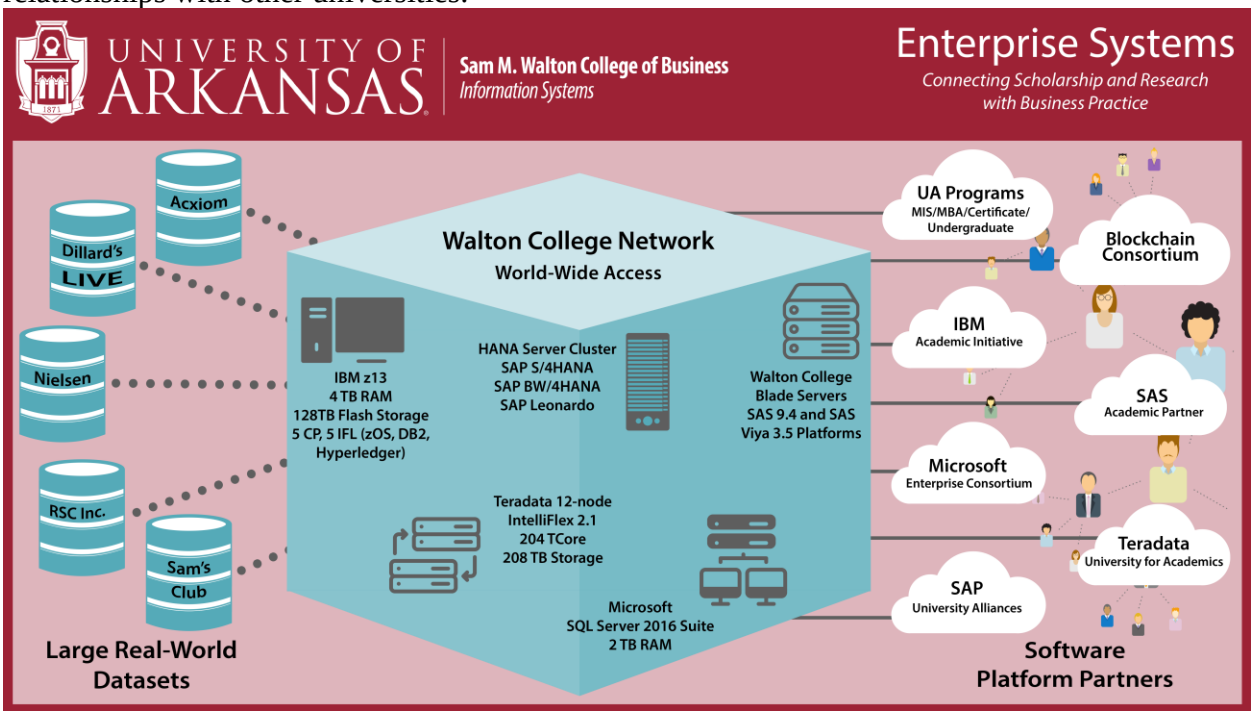

Figure 1: Enterprise Systems Network Diagram

3. Description of the Current Structure and Process

In the current process, a user creates a request through the link "https://request.information-systems.uark.edu/login", and upon receiving it, the Walton department comprehends the user's requirements and sends the request to Central IT for processing. After receiving all the necessary account information from Central IT, Walton IT sends the relevant information to the user via mail, which is the account username and password.
Whenever Walton IT sends the user information through mail, the entire email including the student account information is captured in a table as a free text field. As a result, it becomes challenging for our client to track each individual account that has been created to access their systems.

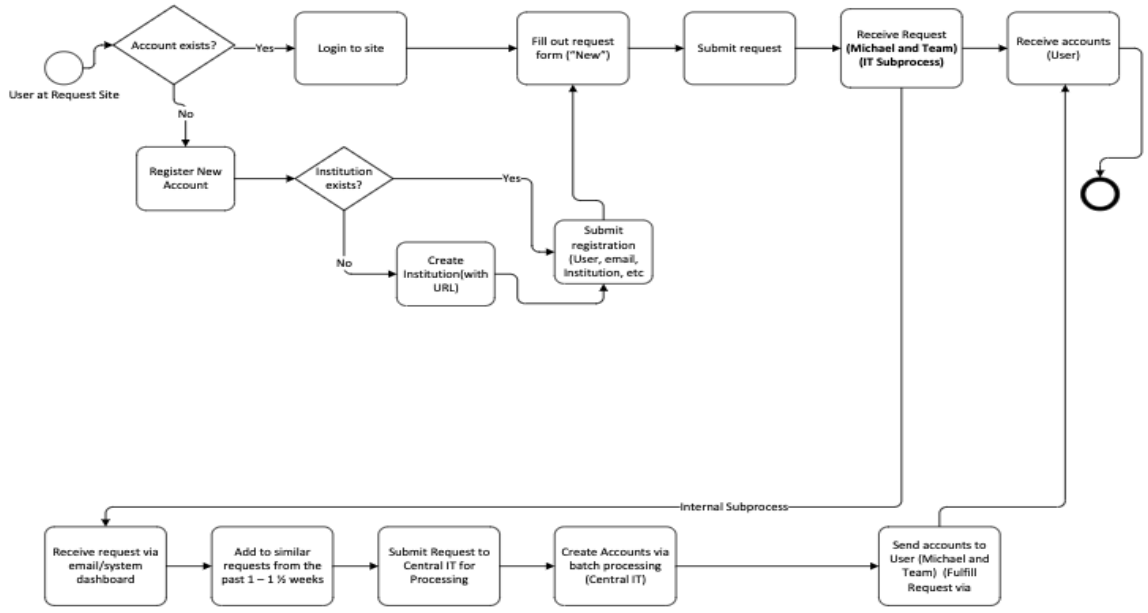

Figure 2: Process flow of Request System

Furthermore, each time a request is raised, the created requests are stored on one table while the related information is stored on different tables. Therefore, it becomes challenging to segregate data and obtain findings when the client wants to gather data to visualize and analyze the number of faculty and student accounts created worldwide.
So, the second challenge is to collect the data generated from requests made by faculty members of universities worldwide to access the services provided by Enterprise Systems.
To overcome these two major challenges, we have come up with an automated process that facilitates the extraction of student credentials to fulfill the client's requirement of tracking individual student information. This process allows for monitoring of service utilization and identification of inactive accounts that need to be terminated. By employing Python to connect with the SQL database, we are now able to extract student login information from the free text field that was previously not being tracked due to the lack of standardization and was stored in a column as free text data.
Additionally, we have created one denormalized table containing all the required data needed for visualizations along with the student account information.

4. Problem Statement and Detailed Description of the Information System

The report identifies a few key issues with the previous system, which lacks the tracking of student accounts, absence of usage analytics, difficulty in retrieving data, and security and compliance. The inefficiency of the previous system resulted in an inability to determine whether services were being utilized effectively, and the lack of tracking made it impossible to identify areas for improvement or expansion.

The problem our client was facing was a major inefficiency in their process and as a result, they were unable to effectively track how many student accounts were accessing their resources. However, we have addressed this issue by making the process more efficient and implementing a tracking system would allow the client to monitor the number of student accounts accessing our services. The lack of efficiency and inability to track student accounts was problematic for our client as they were unable to determine if their services were being utilized effectively, nor could they identify areas for improvement or expansion. Consequently, reporting and analysis also became difficult without proper tracking. The previous system required faculty members to submit requests that were then processed centrally, and email notifications were sent out to faculty members containing student account and password information.

The following are the below pain points for the environment:

- **Lack of detailed student information management:** Without a comprehensive and detailed tracking system, the business is unable to monitor each student's account individually. As a result, the degree of service utilization cannot be accurately determined and any inactive accounts that might require termination cannot be identified. (Business has requested that the number of students utilizing the service, along with their student IDs, be provided.)
- **Absence of usage analytics:** Due to the inability of the business to monitor each individual user of the system, they are unable to collect information regarding service usage by each student account which is crucial to identify universities with excessive system utilization. Information gleaned from such data can inform the IT group's reputation-building initiatives and aid in anticipating future requirements, given the provision of a free service.
- **Security and compliance:** The inability to track student account information poses a problem for the IT team who need to ensure that the university's IT resources are secure and compliant. This includes verifying that students are only accessing data and systems for which they have been authorized and ensuring that they are only performing transactions within authorized systems.
- **Reputation Building:** The provision of complimentary services is a valuable opportunity for the IT group to establish a favorable reputation. The expressions of gratitude from recipient universities could facilitate further support and acknowledgement from the Dean's Office, amplifying the IT group's standing. However, due to the absence of a detailed tracking system and beneficial usage analytics, the business is unable to achieve this. To address these issues, the report outlines the implementation of a more efficient process and a tracking system that allows for monitoring of the number of student accounts accessing services. The new system will enable better reporting and analysis, ultimately leading to improved decision-making.

To address these issues, the report outlines the implementation of a more efficient process and a tracking system that allows for monitoring of the number of student accounts accessing services.

With the implementation of the new process, we aim to address the above issues and enhance reporting and analysis, which will result in better decision-making capabilities.

5. Proposed IT Solution

Our proposal involved a complete redesign of the entire process to establish a more efficient system that would allow the client to capture the required information, and which overcomes the above challenges. Our solution involved creating a single table that would contain all the necessary information. We used tools including MSSQL, SQL Server Integration Services and Python scripts to extract and transform the data. We started with the existing system that had several tables, extracted the necessary data, and combined it into a single normalized table. The result was a solution that would enable the client to efficiently track the number of student users and improve their reporting and analysis capabilities.

To understand more on our approach, we have integrated SQL into the extraction process, allowing the query to store results in a table within SQL Server. Additionally, we aimed to extract student account details from a text block of information. To achieve this, we used SSIS as our main tool to build the package, along with integrating Python scripts into a single solution for efficiency. We gathered data from ten different tables and consolidated it into one flat table. Then, we created a concept to break out the block of information into the starting and ending range for the student account, which runs a Python script to extract the individual account details and feed them into a new table. Then integrated all the information into a single final normalized table. Instead of running this manually, we have gone a step further to automate the entire process for our client. We have set it up on a SQL server and scheduled it to run once a week, so the package runs seamlessly every time it is scheduled, with no involvement required from the client.

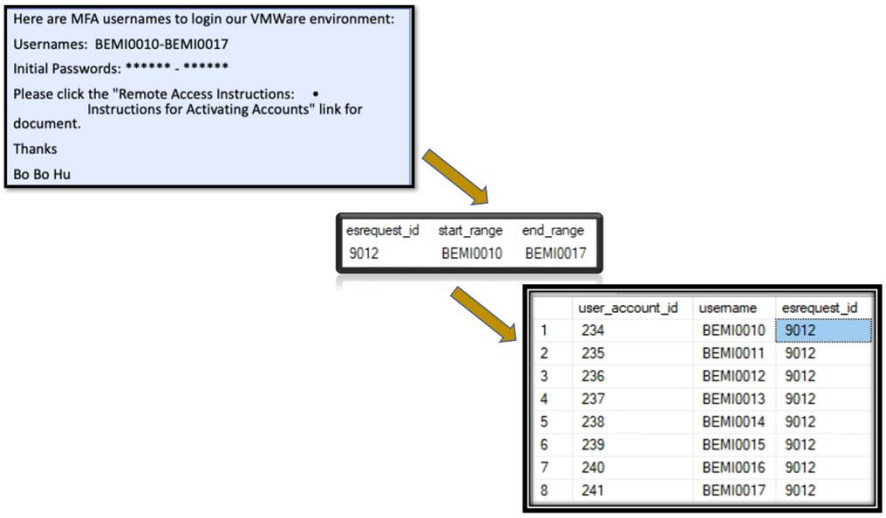

Figure 3: Data Extraction of individual Student Accounts from the email message in the 'note' free text column.

We have created 3 new tables for achieving the objectives:

1. new accounts – We have taken ESRequest_Id and Note column from ESRequests table and then split the note column into Start Range and End Range. It has the columns ESRequest_Id, Start Range and End Range.
2. user_accounts - We further tried to keep all the ranges of user accounts in each separate row. It has the column UserName and ESRequest_Id.

3. new_es_requests – We have gathered the information from the selected 6 tables esrequests, users, institutions, esrequests_platform, platforms, and faculty_accounts to create a new denormalized table with all the required columns.

We have successfully created a SQL Query that identifies the necessary fields for visualization, facilitating valuable business decisions. Currently these fields are being captured from multiple tables on SQL server database gathering the latest information captured by the query and integrated python scripts to get a new denormalized table that can be directly linked to SAS Viya for visualization purposes.

Below is the flow chart explaining the process of extracting the data from multiple tables into one denormalized table.

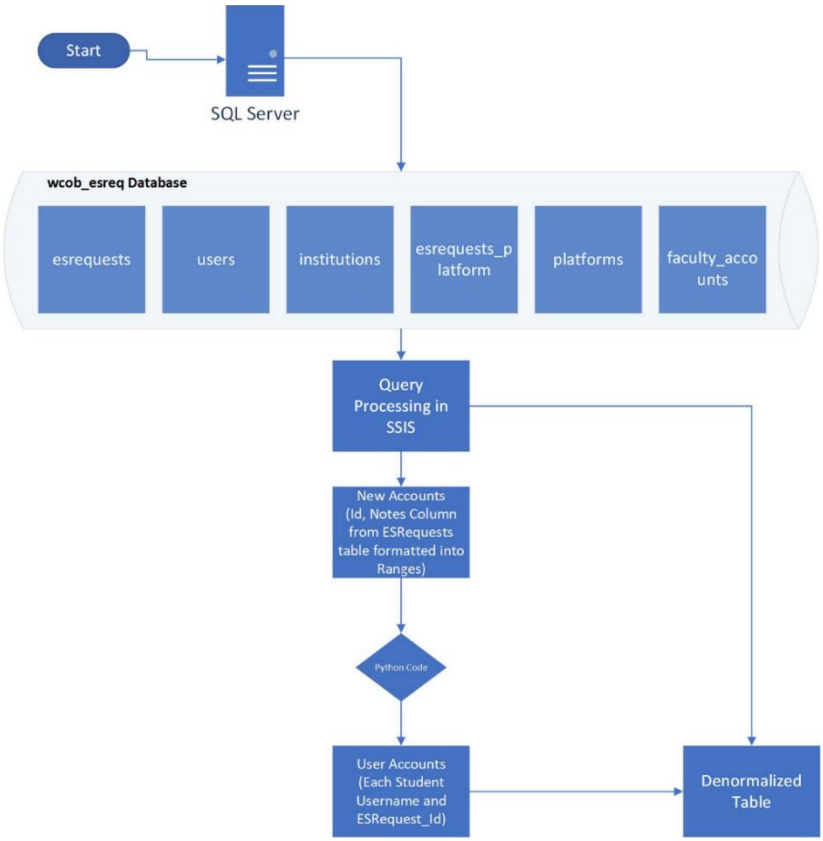

Figure 43: Process Flow of our Solution

Technologies Used:

We have utilized several tools and programming languages to design the solution for our client. The major software's used includes Microsoft SQL server as this is the primary platform being used by the client to store data generated from their request's platform. We also utilized SQL Server Integration Services (SSIS) which is a data integration and transformation tool provided by Microsoft SQL Server. It is often employed to facilitate the extraction, transformation, and loading (ETL) of data from numerous sources to a destination, such as a data warehouse, data mart, or a database. With the use of SSIS, we have created a workflow package that would collect data from the various tables, transform the data to extract the individual student accounts, and then load the transformed data along with other relevant data into one single table in the database. We have taken it a step further ensuring we automate the process for our client by scheduling the workflow,

using SQL Agent, to be executed at certain times in this case once a week as requested by our client.
We have also used SQL in the data cleaning and extraction by writing queries to aggregate data from the various tables and to extract the starting and ending range of the student accounts provided in the free text field 'note' column.
Lastly, we utilized python as part of our data transformation and generation process by writing a python script that would use the start range and end range to create a list of individual student accounts.

Cost & Time – Benefit Analysis:

Based on the above analysis, the project has delivered significant benefits to the client while saving them a considerable amount of time and money. The implementation of a comprehensive tracking system has improved data quality, provided the ability to collect usage analytics, and enhanced security and compliance. Additionally, the processing time for the entire process has been reduced from 3 hours to less than 1 minute. These benefits greatly outweigh the costs of the project, which were limited to the time invested in the project and the potential cost to the client. Overall, the project has been a success and has delivered substantial benefits to the client.

Cost:

- We have invested a total time of around 50 hours in completing the project.
- Based on the number of hours spent and the value of our services, considering that we are not professionals in the field, we have estimated the potential cost to the client for the project to be around $3000 - $3500.
- The client saved a significant amount of money by completing the project in just 50 hours (about 2 days) instead of hiring an external contractor for an estimated 240-280 hours (assuming an hourly rate of $60-$70) which will have costed around approximately $14,400 - $16,800.

This shows that the team's expertise and efficient work processes delivered significant cost and time savings for clients.

Benefits:

- Comprehensive Tracking System: A detailed tracking system was now implemented to allow the client to better manage student information. This would help the client to keep track of student data and usage patterns, which could be valuable for improving the quality of educational services offered by the client. The tracking system also helps in identifying any errors or issues with the student data, allowing for timely correction.
- Improved Data Quality: The team was able to retrieve useful information from complex text and convert it into a list of usernames. This has improved the data quality by making it more organized and easier to understand. This can be valuable for the client in understanding student usage patterns and improving the quality of educational services offered.
- Usage Analytics: The system allows for the collection of usage analytics to determine excessive or under-usage of systems. This can help the client to identify areas where improvements can be made to optimize the use of educational systems and resources. This can help in providing better educational services to students while reducing unnecessary costs.
- Enhanced Security and Compliance: The team ensured authorized access to data and systems, thereby enhancing security and compliance. This is important as educational institutions often deal with sensitive student data that needs to be protected from

unauthorized access. The team's work has helped the client to ensure that the data is accessed only by authorized personnel, which is essential for maintaining the privacy and security of student data.

What Went Well:

- Team's resilience in setbacks: During the project, the team demonstrated remarkable resilience in overcoming obstacles, setbacks, and challenges that arose during the project. The team's ability to stay motivated and focused on achieving the project objectives, even in the face of adversity, was noteworthy. Despite encountering issues such as software glitches, technical difficulties, and other unexpected roadblocks, the team's resilience allowed them to successfully complete the project.
- Prioritization: Another strength of our team was our ability to prioritize effectively. We concentrated on our most critical tasks and were able to allocate our time and resources accordingly. This helped us to stay on track and meet our project milestones despite the challenges we faced.
- Client's availability and responsiveness: We were also fortunate to have a highly engaged and available client throughout the project. The client's willingness to provide feedback and input was instrumental.
- Technical Knowhow: Our team's technical know-how was a key factor in the success of the project. We were able to leverage our combined knowledge to design and implement a solution that was both effective and efficient.

What Did not Go Well:

- Change of Scope: There were some changes in the scope of the project which had an impact on the team's plans and timelines. We had to adjust the approach to accommodate these changes, which required additional resources and time. This affected the team's ability to meet the initial project objectives within the original timeframe. The team had to work closely with the client to re-evaluate the project scope and timeline to ensure that the new objectives were achievable.
- Additional Requirements: The second issue faced during the project was related to additional requirements. As the project progressed, the client requested additional features and functionalities to be added to the system. The team had to adjust the timeline and work plan to accommodate these changes. However, the team worked closely with the client to prioritize these additional requirements and ensure that they were implemented in a timely and effective manner.
- Cooperation: Although we encountered a minor hurdle in the beginning, when we needed to standardize the email format, we worked together with the client and other team members to resolve it. Eventually, we found a collaborative solution that satisfied the client's requirements.
- Access Issues: We encountered access issues that hindered our ability to obtain certain data and access certain systems. This caused a temporary delay in our progress, as we had to wait for a few days to gain access. However, we worked closely with the client to address these issues and ultimately obtain the necessary access to successfully complete the project.
- Knowledge Gap: we encountered a knowledge gap in some areas of the project. While we had technical expertise in many areas, there were certain aspects of the project that required additional research and learning. We addressed this gap by leveraging our team's collective knowledge and expertise, as well as seeking external resources as needed.

Lessons Learned:

1) Understanding of the overall environment in Walton UARK like Test, Prod, Staging.
2) Understanding of the overall elevation permission, Azure AD single sign on, Roles for triggering SSIS Tool with new automation pipeline.
3) The scope of the project and additional requirements did change with the phases of the Project.
4) Agile Methodology was used in the execution of the Project.
5) A clear understanding of business problems is crucial for effective solutions.
6) Defining project scope is essential to manage expectations and deliver successful outcomes.
7) Consistent communication with clients is paramount for better project delivery

6. Appendix

1 – Python Codes

1.1: Python code to retrieve individual student usernames

```
# building a connection to sql server
import pandas as pd
import pyodbc
# replace the server_name and database_name with your own details
server_name = 'sql-4.uark.edu'
database_name = 'wcob_esreq'
# establish a connection to the SQL server using Windows authentication
conn = pyodbc.connect('DRIVER={ODBC Driver 17 for SQL Server};' +
                      f'SERVER={server_name};' +
                      f'DATABASE={database_name};' +
                      'Trusted_Connection=yes;')
# string variable to store all columns from new_accounts table
student_query = "SELECT * FROM new_accounts;"
# read the SQL query into a pandas dataframe
student_df = pd.read_sql(student_query, conn)
#delete only rows having NULL in both the ranges
student_df.dropna(subset=['start_range', 'end_range'], how='all', inplace=True)

#producing list of usernames
import re
usernames_df = pd.DataFrame(columns=['username', 'esrequest_id'])
for index, row in student_df.iterrows():
    # extract the start and end range from the row
    begin = row['start_range']
    end = row['end_range']
    esrequest_id = row['esrequest_id']
    # extract the numeric portion of the start and end range using regex
    start_match = re.search(r'\d+', begin)
```

```
            start_num = int(start_match.group()) if start_match else None
            end_match = re.search(r'\d+', str(end)) if pd.notnull(end) else start_match
            end_num = int(end_match.group()) if end_match else start_num
            # generate a list of all usernames in the range
            try:
               usernames = [f"{begin[:4]}{num:04}" for num in range(start_num, end_num+1)]
            except TypeError:
               pass
            # add the usernames to the dataframe
            usernames_df = usernames_df.append(pd.DataFrame({'username': usernames, 'esrequest_id':
[esrequest_id]*len(usernames)}), ignore_index=True)
         print(usernames_df)
```

2 – SQL Codes

2.1: SQL Code to extract username information from "Note" column

```
SELECT b.id, SUBSTRING(Start_Range, 2, 9) AS Start_Range, SUBSTRING(End_Range, 1, 9) AS
End_Range
FROM
(SELECT a.id,
REVERSE(PARSENAME(REPLACE(REVERSE(extracted_text), '-', '.'), 1)) AS [Start_Range],
REVERSE(PARSENAME(REPLACE(REVERSE(extracted_text), '-', '.'), 2)) AS [End_Range]
FROM
(SELECT id, SUBSTRING(note, 71, 21) AS extracted_text
FROM esrequests
WHERE note IS NOT NULL AND created_at >= '2022-11-08 22:45:06.000')a)b
```

2.2: SQL Code to extract relevant information from the 6 tables

```
SELECT esrequest_id, E.user_id, faculty_accounts, student_accounts, student_accounts_created,
sas_user_account,
why_question, where_question, platform_id, P.name AS platform_Name, platform_comment,
platform_services_list,
course_name, institution_id, I.name AS institution_name, url AS institution_url, first_name, last_name,
FA.username AS faculty_username, FA.password AS faculty_Password,latitude, longitude, E.updated_at,
E.fulfilled_at
FROM esrequests E FULL JOIN users U
ON E.id = U.id
FULL JOIN institutions I
ON U.id = I.id
FULL JOIN esrequest_platform EP
ON EP.esrequest_id = E.id
FULL JOIN platforms P
ON EP.platform_id = P.id
```

```
FULL JOIN faculty_accounts FA
ON U.id = FA.id
WHERE esrequest_id IS NOT NULL
ORDER BY esrequest_id
```

2.3: SQL Code to produce Final Denormalized table

```
SELECT a.esrequest_id, user_id, faculty_accounts, student_accounts, username, start_range, end_range,
student_accounts_created,
sas_user_account, why_question, where_question, platform_id, platform_Name, platform_comment,
platform_services_list,
course_name, institution_id, institution_name, institution_url, first_name, last_name, faculty_username,
faculty_Password, latitude, longitude, updated_at, fulfilled_at
FROM new_es_requests a LEFT JOIN new_accounts b
ON a.esrequest_id=b.esrequest_id
LEFT JOIN user_accounts c
ON b.esrequest_id=c.esrequest_id
```