Data Scramble Process Model in SaaS Application

Rajagopal Arputham Chetty

Senior Workday Systems Analyst – Payroll – LifeTime
rajgopal.78@gmail.com

Abstract:

With the emergence SaaS applications wherein organizations are transforming to SaaS application from conventional application, the biggest challenge these organizations are currently facing the how data can be protected between the environments. Organizations do not want the real data of production into lower environments when the environment is copied as the risk of exposing the PII data from lower environment is high as the data flows to other systems for testing. The solution is Data Scrambling. This paper explores the functions of how data can be scrambled in SaaS applications especially in Workday ERP. The paper is organized in a way which explains what data scrambling is. steps of data scrambling and highlights a case study of how we can perform data scrambling. This paper concludes with the prerequisite and considerations to be followed before we implement the scrambling process.

Keywords: Workday, Data, PII, Production, Scrambler.

1. INTRODUCTION

Organizations using SaaS Applications like Workday, SAP, Oracle Fusion will have different environments for development, testing and Production. Development and testing environments will be known as lower environments. The lower environments will be refreshed with the actual production environment data on a particular interval based on the solution which will result in actual employee PII and other confidential data. Having real data in lower environment will lead to considerable risk of data exposing as these environments may not have the same strict security reasons and these data will go to other application systems for any testing. The best way to avoid this data risk is by scrambling the data without changing the data model structure. As Organizations using SaaS application cannot do database level scramble as they do not have the access to the database and the data structures.

2. WHAT IS DATA SCRAMBLING?

Data Scrambling refers to the technique of identifying sensitive information and protecting the sensitive information by replacing the original values with random generative values based on data types like text, number date etc. This technique ensures that the sensitive information is secure but usable in non-production environment for development, testing and training.

3. DATA MASKING VS DATA SCRAMBLING?

Data Masking and Data Scrambling are often used and discussed interchangeably. There are key differences between Data Masking and Data Scrambling.

Data Masking is the procedure where the real data is not getting updated in the database but the data when we viewed in UI does change and it is masked.

Data Scrambling is the procedure for whether the real data itself is changed in the database and the real data no longer exists in the database. As the data is already updated with the scrambled value the UI will display the scrambled data value.

4.   NEED FOR SCRAMBLE DATA

There are many reasons why we need scrambled data. Some of the need for scrambling data is as follows: -

- **Configurations, Integrations and Reports:** Customers require non-production tenants to develop new functions like configurations and build new integrations. Also, customers build many custom reports based on their business needs. In these scenarios, we do not want to expose real data from production and hence Data scrambling is needed.
- **Testing and Troubleshooting:** Testing of functionality and troubleshooting of the issues is done in non-production tenants and the individual performing these functions should not have the real production data.
- **Training and Demos:** Often clients use non-production tenants to provide training and demos for the employees, and it does not require to have actual production data.

5.   WORKDAY ERP

Workday is a powerful cloud-based ERP platform (SaaS solution) that helps organizations to have a centralized platform for managing all the company's business functions. Workday has various products like Human Capital Management, Financial Management, Workday Payroll, Adaptive Planning, Workforce Management, Analytics and Reporting, Spend Management, Student, Talent Management, Peakon Employee Voice etc., Workday ERP provides better scalability, lower costs, better data security and frequent updates on regulatory rules.

6.   WORKDAY Data Scrambler

Prior to Workday Data Scrambler, Customers need to identify their own data scrambling algorithms and download the data from the tenant and apply their data scrambling algorithms and manually load the data into tenants. Also, they need to create complex security configurations to restrict the user from having access to sensitive data.

Workday Data Scrambling is built in such a way that the customers can identify the exact data elements that they want to scramble and the method they want to use for each data element value and initiate the process in mass. The data scramble process is stored in the tenant and the process can be reused. The process is built for efficiency. Another benefit of Workday data scrambler is the functionality which resides inside the customer tenant. This method is ensuring your data is not leaving the tenant and the data is not getting compromised.

Workday Data Scrambler is completely free for Workday Customer which is included in the base subscription, and it is very user friendly, and the customer does not require any service assistance to enable and use this feature.

Figure 1: Workday Data Scrambler

7. PREREQUISITE SETUP

The prerequisite setup is enabling the configure of Scrambler Administration Domain. The recommendation is that either the security administrator or the user who is responsible for handling Workday user accounts should be provided with this domain policy access to who can manage the scrambling activities in the tenant. The reason for this recommendation is that you may decide to scramble the user accounts and only the security administrator or the equivalent domain access user can activate and reset the user account.
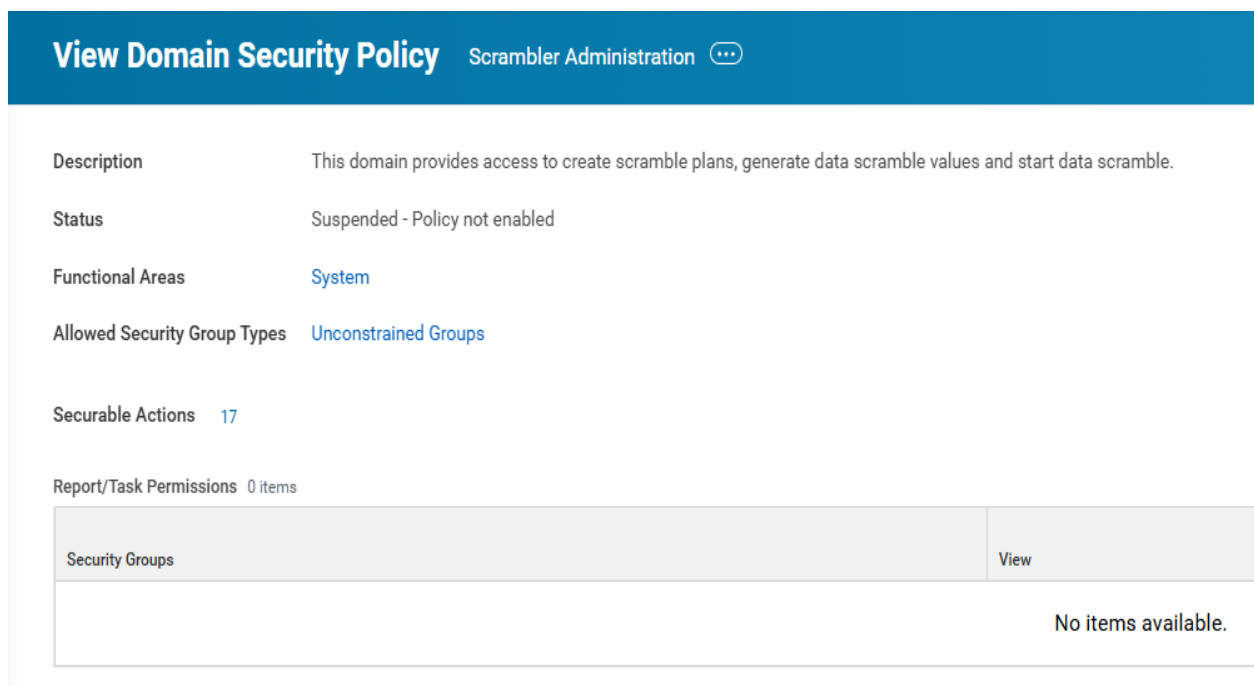
Figure 2: Data Scrambler – Domain Security Policy

8. DATA SCRAMBLING PROCESS
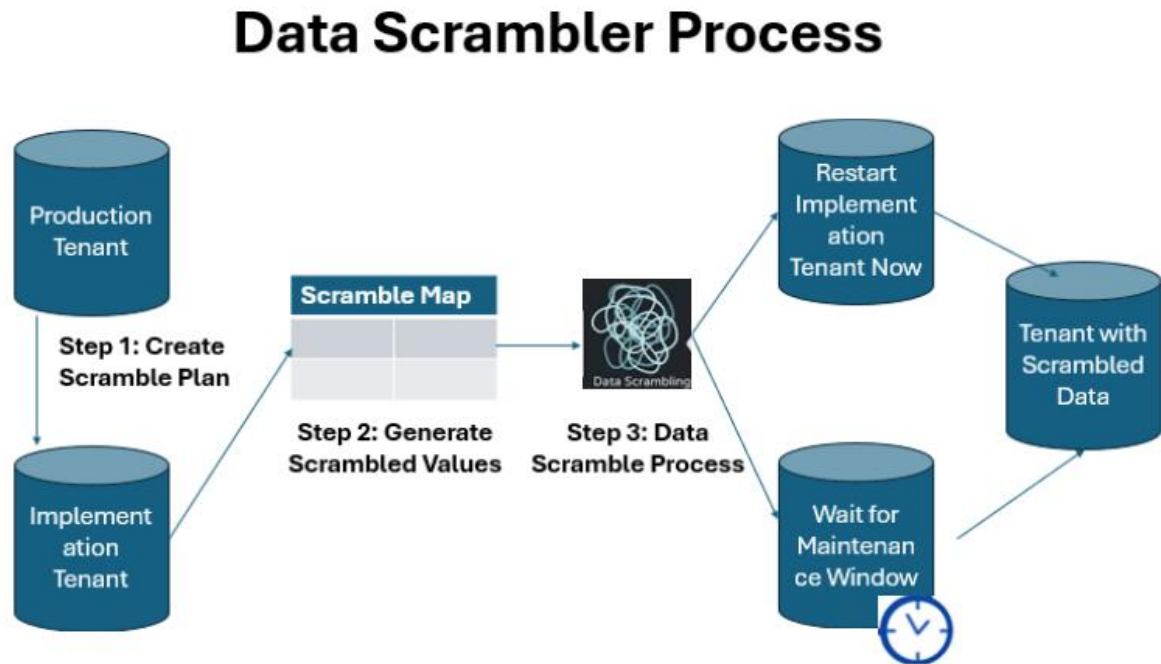
Figure 3: Data Scrambling Process

Step 1 → Create Scramble Plan: You can create the scramble plan in production which identifies the data elements which are selected for data scrambling and the methods to be used for each of the data elements.
Copy Scramble Plan to Implementation Tenant: Copy of the scramble plan to the implementation tenant. You can also create the scramble plan directly in the implementation tenant, but it may get overridden in case the implementation tenant got refreshed by production tenant. Creating the scramble plan in production and copying to implementation tenant ensured the scramble plan is reusable and not forget to scramble certain fields.

Step 2 →Scramble Map: This step is used to generate the scrambled data value. This step uses the scramble plan created in Step 1 to identify all the data elements selected for data scrambling. The data is staged, and the scrambled value is created for each original data element value based on the scrambled method selected.

Step 3 → Data Scramble: The third step is the actual data scramble process. This is the step where the original data is converted into scrambled data based on our methods and this step is irreversible.

Final Step: Once the data scrambling is completed, the tenant needs to restart, to have the scrambled data to be persistent. You can either restart the implementation tenant immediately or wait for the maintenance window period. If you opt to wait for the maintenance window for the tenant to restart, then the real value will still exist till the tenant is restarted after the maintenance window. If you opt to restart immediately then the scrambled data is visible once the tenant is restarted.

9. SCRAMBLING METHODS

Scrambling Methods are the algorithm which is used by Workday to convert the real data to scrambled data. Workday has already provided pre-defined scrambling methods which you can attach to different data elements. Each data element is provided with their own scrambling methods, and some data elements can have more than one scrambling method to choose from. These methods are specifically designed for the data

field. Customers cannot create their own scrambling methods. Please find some examples of the scrambling methods.

Example 1: First Name

The First Name data element has only one scrambling method which is "Consonant-vowel-consonant-vowel-consonant-vowel" method. This method will scramble the original value to a random Consonant-Vowel combination whereas the length is preserved.

Example 1:First Name

Field	Method(s)
First Name	Consonant-vowel-consonant-vowel-consonant-vowel

Original Value	Scrambled Value
Logan	Wepur

Figure 4: First Name

Example 2: Address Line 1

The Address Line 1 of the worker has only method for data Scrambling which is "Random 5 digit number + Maple Lane". This method will scramble the original address 1 with a random 5 digit number followed by Maple Lane. In this method, the length of the original address is not preserved.

Example 2: Address Line 1

Field	Method(s)
Address Line 1	Random 5 digit number + Maple Lane

Original Value	Scrambled Value
1099 14th St NW	87419 Maple Lane

Figure 5: Address Line 1

Example 3: Date of Birth

The Data of birth has only method which is "Within six months of original date." This method will scramble the original date of birth with a random date which will be either six months before or after the original date.

Example 3: Date of Birth

Field	Method(s)
Date of Birth	Within six month of original date

Original Value	Scrambled Value
10/17/1991	04/17/1991 – 4/7/1992

Figure 6: Date of Birth

Example 4: Gender

The Gender has only method which is "Random Gender (from the defined list of values)." This method will scramble the original gender with the random predefined list of gender defined in the tenant. This could be different values of gender, or the original gender value is retained.

Example 4: Gender

Field	Method(s)
Gender	Random Gender (from the defined list of values)

Original Value	Scrambled Value
Male	Possible Values: Male Female Not Declared

Figure 7: Gender

Example 5: Marital Status

The Marital Status has only method which is "Random Marital Status for Country." This method will scramble the original marital status with the random marital status defined in the tenant for the specific country. This could be different values of marital status, or the original value is retained.

Example 5: Marital Status

Field	Method(s)
Marital Status	Random Marital Status for Country

Original Value	Scrambled Value
Country: United States of America Marital Status: Married Marital Status Date: 10/17/1990	Country: United States of America Marital Status: Divorced Married Partnered Separated Single Widowed Marital Status Date: 10/17/2021

Figure 8: Marital Status

Example 6: Government IDs

The Government IDs has only method which is "Random String for a Government ID Matching Existing Character Length (Up to 20 Characters)". This method will scramble the original government ids with the random government id for the specific country definition and specific to government id type matching the exact 20 characters.

Example 6: Government IDs

Field	Method(s)
Government IDs	Random String for a Government ID Matching Existing Character Length (Up to 20 Characters)
Original Value	**Scrambled Value**
Country: United States of America Government ID Type: Permanent Resident Alien card (Form I-151) Identification #: ABC9810456679	Country: United States of America Government ID Type: Permanent Resident Alien card (Form I-151) Identification #: CPQ7810293485

Figure 9: Government IDs

Example 7: National Identifier

The National Identifier has only method which is "Same NID Type Respecting Regex." This method will scramble the original national identifier with the random national identifier for the specific country definition and specific to National id type.

Example 7: National Identifier

Field	Method(s)
National Identifier	Same NID Type Respecting Regex

Original Value	Scrambled Value
Country: United States of America National ID Type: Social Security Number (SSN) Identification #: 341-11-0990	Country: United States of America National ID Type: Social Security Number (SSN) Identification #: 230-01-9192

Figure 10: National Identifier

Example 8: Documents, Attachments, and Images

The documents, attachments and images have 3 methods wherein we can choose based on customer needs. The methods are "File Name with no File Access," "Include or Exclude Specified File Types [Custom Parameter]" and "Masked File Name with no File Access."

Example 8: Documents, Attachments & Images

Field	Method(s)
Documents, Attachments, and Images	1. File Name with no File Access 2. Include or Exclude Specified File Types [Custom Parameter] 3. Masked File Name with no File Acces

Original Value	Scrambled Value
Test_Certificate.pdf	1.Test_Certificate.pdf 2.***************.pdf 3.***************.pdf

Figure 11: Documents, Attachments & Images
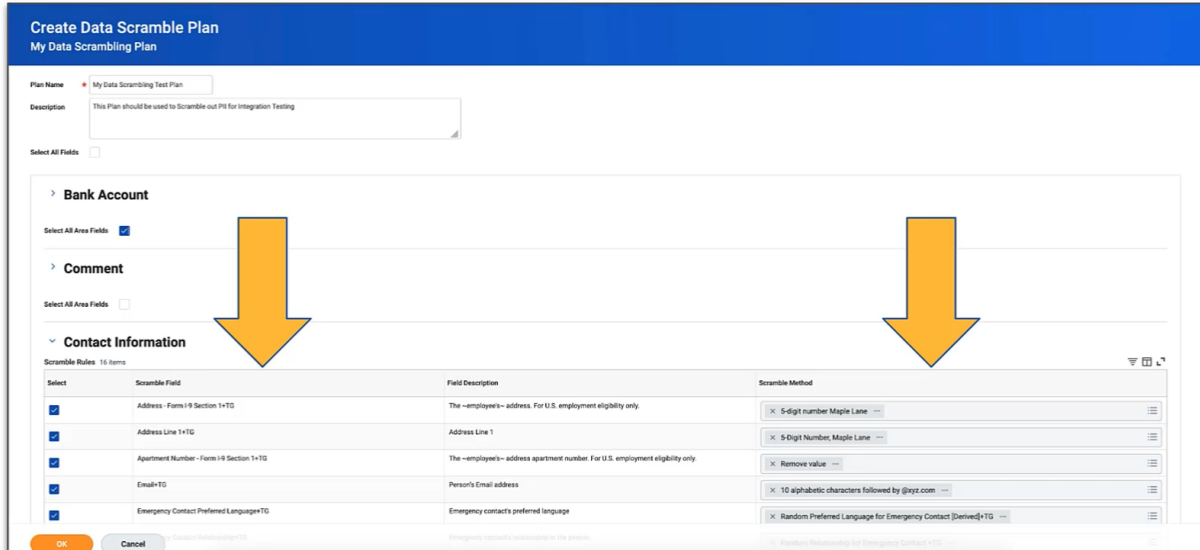
Case Study: Scrambling Process of Contact Information

A MNC company which is in Retail and fitness company wants to scramble the contact information of the employees in the implementation tenant to configure and implement a new functionality of Change Job feature in Workday. So, the production copy was copied into the sandbox tenant and executed the Workday Data Scramble process successfully to scramble the data.

Created a scrambling plan by choosing the required contact information data fields and the scrambling methods for each of the data fields. Processed the scrambling process to stage the scrambled values and completed the process.

Step 1: Create Scramble Plan

This step is where the user selects which values to scramble and the method to use to scramble each value.

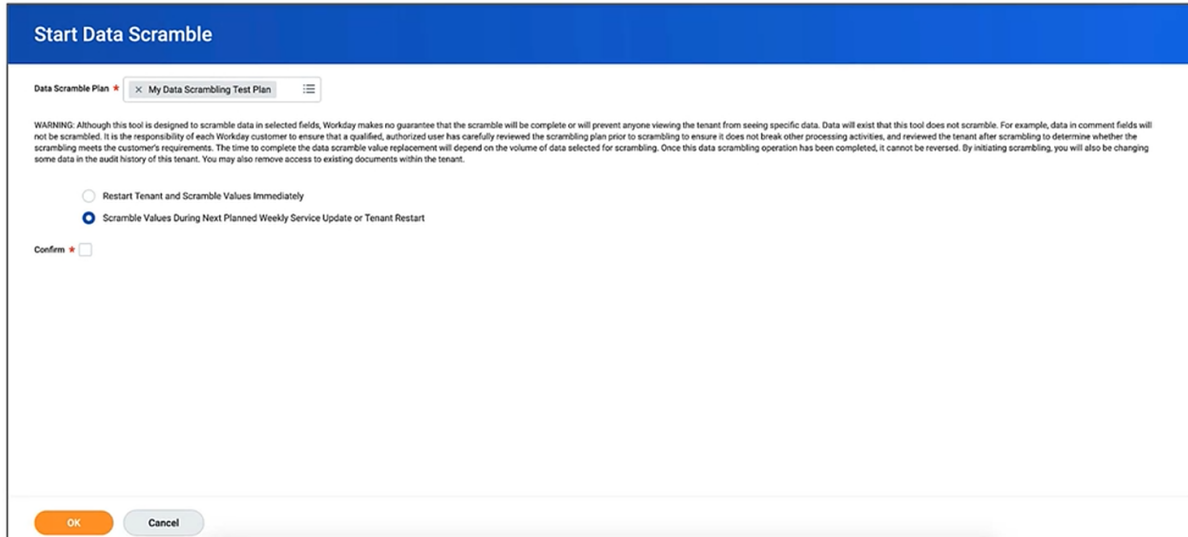**Figure 12: Contant Information Scramble Process**

Step 2: Generate Scrambled Values

This step is the staging step where the system takes each value and produces a scrambled version of each value for each data element being scrambled.

Figure 13: Contant Information – Generate Scrambled Values

Step 3: Start Data Scramble

This step replaces the unscrambled values with the scrambled values

Figure 14: Contant Information – Data Scramble Process

Scramble Completed

Figure 15: Contant Information –Scramble Completion

Consideration and Best Practices

Based on the research /process of data scrambling for the above case study, there are lot of learnings which is listed below as consideration and best practices.

1. Limit the users having access to create scramble plans and run the scramble process.
2. The user who is performing the scramble process should have view access for user accounts as sometimes the user names are selected for scrambling.

3. As mentioned above in the document, create the scramble plan in production so that it is reusable and data elements are not forgettable.
4. Based on the business function, create separate scramble plans for compensation, payroll etc. so that it is processed in a phased manner
5. Creating multiple scrambling plans in case of large volume of data is considered for scrambling.
6. Based on the tenant restart option, generate the scrambling plans with plenty of time before the maintenance window
7. Carefully review the scrambling plan prior to execution.
8. The scrambling process is irreversible, hence make careful decisions before initiating the process.
9. Make sure no integration or any process is running while executing data scrambling.
10. Data Scrambling can be performed only in Implementation and Implementation preview tenants.
11. Sandbox and Sandbox preview is also excluded from Data Scrambling process.

CONCLUSION

This paper has researched about the needs of data scrambling especially in SaaS Application with a specific focus on Workday. This paper explains the concept of data scrambling and distinguishes the difference between data masking and data scrambling. This paper explored the option in Workday for data scrambling and the complete process flow. This paper also done complete research on the scrambling methods available in Workday for each of the data element types and provided specific examples for each of the data element types. This research document also provides the prerequisites for this process and provided the live case study done for an MNC customer. Based on the case study research, this paper also provides insights into the best practices and considerations before initiating the data scrambling process in Workday.

REFERENCES:

[1] Workday Community Portal [Online]. https://resourcecenter.workday.com/ (accessed July 10, 2025)
[2] Data Scrambling: Data Protection Techniques [Online] (accessed June 20, 2025)
[3] Managing Data Quality in MDM SaaS [Online] -
    https://www.informatica.com/content/dam/informatica-cxp/ (accessed June 21, 2025)
[4] Data Scrambling as a Service in Cloud [Online]
    https://www.youtube.com/watch?v=HVSx9NYc_0Y&pp=ygUPZGF0YSBzY3JhbWJsaW5n
    (accessed July 14, 2025)
[5] Merit Letter Load and Data Scramble in Workday Advanced Compensation [Online] Snowflake
    1920x1080 15sec (accessed July 14, 2025)