

User Manual for Inferenz Unify

(Snowflake Native App for Enterprise-Grade Data Deduplication)

1. Introduction

Inferenz Unify leverages GenAI technology to intelligently generate rules for identifying duplicate data within healthcare systems. By automating the detection process, it ensures that duplicate records are flagged efficiently, reducing the manual effort typically required for data cleansing. This approach not only streamlines data management but also enhances the overall integrity and reliability of healthcare information.

2. Prerequisites

Before getting started, ensure that application have necessary permissions:

Example: Granting Permissions to a Snowflake Native App in a Consumer Account

Below are example SQL statements to grant a Snowflake Native App access to tables, schemas, and databases within a consumer account. Replace `your_app_name`, `your_db`, `your_schema`, and `your_table` with your actual object and app names.

1. Granting Privileges on a Database

```
-- Grant USAGE privilege on a database to the application
GRANT USAGE ON DATABASE your_db TO APPLICATION your_app_name;
```

2. Granting Privileges on a Schema

```
-- Grant USAGE privilege on a schema to the application
GRANT USAGE ON SCHEMA your_db.your_schema TO APPLICATION
your_app_name;
```

3. Granting Privileges on a Table

```
-- Grant SELECT privilege on a table to the application
GRANT SELECT ON TABLE your_db.your_schema.your_table TO
APPLICATION your_app_name;
```

4. Granting Global Privileges (e.g., EXECUTE TASK)

```
-- Grant EXECUTE TASK privilege at the account level to the
application
GRANT EXECUTE TASK ON ACCOUNT TO APPLICATION your_app_name;
```

5. Granting IMPORTED PRIVILEGES on the SNOWFLAKE Database

```
-- Grant IMPORTED PRIVILEGES on the SNOWFLAKE database to the
application
GRANT IMPORTED PRIVILEGES ON DATABASE SNOWFLAKE TO APPLICATION
your_app_name;
```

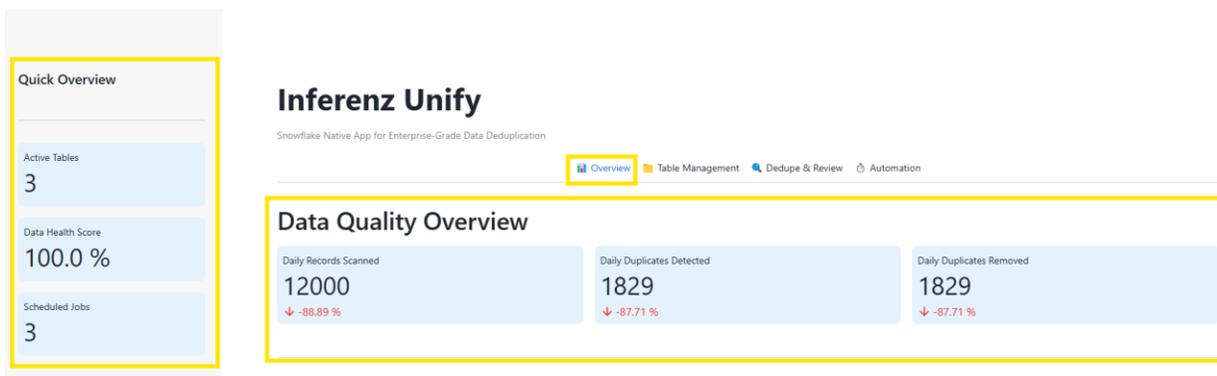
Note:

You can view the privileges requested by the app using:
SHOW PRIVILEGES IN APPLICATION your_app_name;

3. Using the Inferenz Unify

Overview and stats

In the overview tab you will get the complete stats of deduplication of your database.



I. Quick Overview

- No. of table configured for the deduplication and are active under Active Table.
- Health score indicating the Health Score of overall Deduplication.
- No. of job scheduled for automatic run are specified under the Scheduled Jobs (daily/weekly/monthly/cron).

II. Data Quality Overview

- In the Data Quality Overview, you will get information of Daily changes captured.
 - Total record scanned Daily.
 - Total duplicate records detected.
 - Total duplicate records deleted.

III. Table Level Statistics

- There are two types of stats available in Stats.
 - When you select **ALL** then you will get all the stats of overall deduplication.
 - When you select any **Table** from the dropdown list you will get table level stats.
- Funnel View of data shows Daily stats.
- The line chart shows monthly stats.



Step-by-Step guide to add new data source for deduplication

- Navigate to the **Table Management** tab to choose the dataset you want to deduplicate.
- Select the **Database** where your table is present.
- Select the **Schema** and then **Table** from the drop-down list.

Quick Overview

- Active Tables: 3
- Data Health Score: 100.0 %
- Scheduled Jobs: 3

Inferenz Unify
Snowflake Native App for Enterprise-Grade Data Deduplication

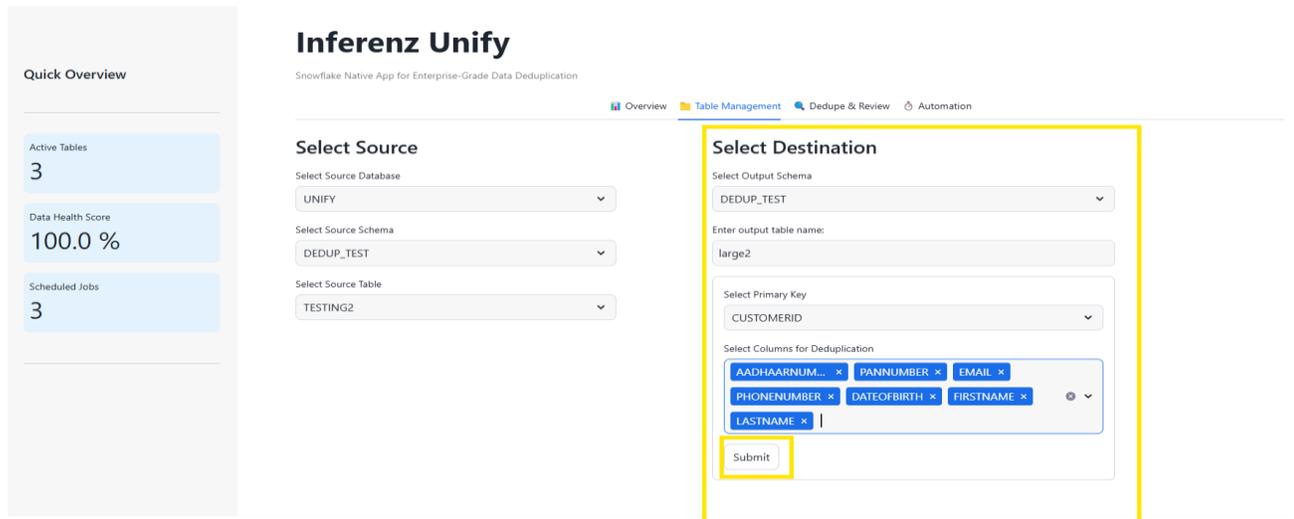
Overview **Table Management** Dedupe & Review Automation

Select Source

Select Source Database
Select a database

© 2025 Inferenz. All rights reserved. | inferenz.ai

- After you select the source details, you need to provide the destination details.



- e. System detected Primary key and Columns for deduplication will appear Automatically,
- f. You can review and change according to the requirement.
- g. You can add or remove columns for deduplication.
Check- (Primary key should not be in the columns selected for deduplication)
 Press the Submit button to generate the deduplication rules and add the info.
- h. This process will take a few seconds.
- i. You can review the rules and make changes if required. (According to your data)

Rule No	Rules
1	(PANNUMBER = 1 AND AADHAARNUMBER = 1 AND (FIRSTNAME >= 0.90 OR LASTNAME >= 0.90))
2	(PANNUMBER >= 0.90 AND AADHAARNUMBER = 1 AND FIRSTNAME = 1 AND LASTNAME >= 0.90 AND (PHONENUMBER = 1 OR EMAIL = 1))
3	(PANNUMBER = 1 AND AADHAARNUMBER >= 0.90 AND (FIRSTNAME = 1 OR LASTNAME = 1))
4	(PANNUMBER >= 0.90 AND AADHAARNUMBER >= 0.90 AND FIRSTNAME = 1 AND LASTNAME = 1 AND (PHONENUMBER = 1 OR EMAIL = 1))
5	(PANNUMBER = 1 OR AADHAARNUMBER = 1 AND FIRSTNAME = 1 AND (PHONENUMBER = 1 OR EMAIL = 1))
6	(AADHAARNUMBER = 1.00 OR PHONENUMBER = 1.00 AND (FIRSTNAME >= 0.90 OR LASTNAME >= 0.90))
7	(AADHAARNUMBER >= 0.90 OR PHONENUMBER = 1.00 AND (FIRSTNAME >= 0.90 OR LASTNAME = 1.00))
8	(AADHAARNUMBER >= 0.90 OR PHONENUMBER = 1.00 AND (FIRSTNAME = 1.00 OR LASTNAME = 1.00))
9	(EMAIL = 1.00 OR PHONENUMBER = 1.00 AND (FIRSTNAME >= 0.90 OR LASTNAME >= 0.90))
10	(EMAIL = 1.00 OR PHONENUMBER = 1.00 AND (FIRSTNAME = 1.00 OR LASTNAME = 1.00))

Submit Edited Rules

- j. Once you submit the edited rules, your table is now configured to run Deduplication

Note- After submitting all the information in the **Table Management** tab, now to run the dedupe you need to switch to **Dedupe and Review** tab.

Steps to run Deduplication.

- To run the dedupe select the source you have already configured.
- Once you select your source click on the **Run Dedupe Now** button.
- After you click on Run Dedupe it will take a few seconds to complete the deduplication.

Inferenz Unify

Snowflake Native App for Enterprise-Grade Data Deduplication

Overview Table Management **Dedupe & Review** Automation

Select Table

DEDUPE.TEST.CUSTOMER_DATA_WITH_IND_REC_TEN_THOUSAND_WITH_DUPLICATES_LARGE_2

Run Dedupe Now

- You will get the option to keep or delete the Latest/Oldest records.

Choose one record per cluster to keep:

- Keep latest
 Keep oldest

Duplicates to Review (Edit / if needed)

	URESPECIMENPATH	AADHAARNUMBER	PANNUMBER	REGION	CREATEDAT	UPDATEDAT	CLUSTER_ID	RULE_ID	DEDUPE_INSERTED_DATE	KEEP
5.827	MENTS/1/SIGNATURE_9344557.PNG	7023 3789 4335	DLDPG6931N	EAST	07/11/2018 14:54:59	27/11/2018 04:03:44	1	1	2025-05-02 08:17:51+05:00	<input checked="" type="checkbox"/>
11.232	MENTS/1/SIGNATURE_9344557.PNG	7023 3789 4335	DLDPG6931N	EAST	07/11/2018 14:54:59	27/11/2018 04:03:44	1	1	2025-05-02 08:17:51+05:00	<input checked="" type="checkbox"/>
862	MENTS/10/SIGNATURE_4636818.PNG	718537571234	MQHPA6611P	NORTH	30/11/2018 13:11:58	16/12/2018 13:12:07	2	1	2025-05-02 08:17:51+05:00	<input checked="" type="checkbox"/>
11.525	MENTS/10/SIGNATURE_4636818.PNG	718537571234	MQHPA6611P	NORTH	30/11/2018 13:11:58	16/12/2018 13:12:07	2	1	2025-05-02 08:17:51+05:00	<input checked="" type="checkbox"/>
1.740	MENTS/22/SIGNATURE_5008762.PNG	8620 8776 8849	CLAPR5101E	SOUTH	02/02/2021 04:17:14	27/02/2021 16:20:08	3	10	2025-05-02 08:17:51+05:00	<input checked="" type="checkbox"/>
11.349	MENTS/22/SIGNATURE_5008762.PNG	8620 3776 8849	CQAPR5101E	SOUTH	02/02/2021 04:17:14	27/02/2021 16:20:08	3	10	2025-05-02 08:17:51+05:00	<input checked="" type="checkbox"/>
10.056	MENTS/27/SIGNATURE_1259879.PNG	279729485240	FVUPS6834S	SOUTH	11/02/2020 22:53:59	12/03/2020 02:27:56	4	1	2025-05-02 08:17:51+05:00	<input checked="" type="checkbox"/>

- You can delete the identified duplicates by clicking on the **Delete Selected Duplicates Button**.

Duplicates to Review (Edit / if needed)

	URESPECIMENPATH	AADHAARNUMBER	PANNUMBER	REGION	CREATEDAT	UPDATEDAT	CLUSTER_ID	RULE_ID	DEDUPE_INSERTED_DATE	KEEP
5.864	MENTS/68/SIGNATURE_9242432.PNG	3259 4745 6225	KIZPF2037J	NORTH	25/09/2019 00:31:31	10/10/2019 00:16:35	9	10	2025-05-02 08:44:09+05:30	<input checked="" type="checkbox"/>
11.127	MENTS/70/SIGNATURE_7459749.PNG	165256405035	AHYPZ0401H	NORTH	20/09/2023 12:06:43	29/09/2023 01:43:16	10	1	2025-05-02 08:44:09+05:30	<input checked="" type="checkbox"/>
10.917	MENTS/70/SIGNATURE_7459749.PNG	165256405035	AHYPZ0401H	NORTH	20/09/2023 12:06:43	29/09/2023 01:43:16	10	1	2025-05-02 08:44:09+05:30	<input checked="" type="checkbox"/>
2.758	MENTS/1008/SIGNATURE_9732482.PNG	798938550675	FLEPR2137T	EAST	03/06/2024 23:14:21	01/07/2024 05:57:40	11	1	2025-05-02 08:44:09+05:30	<input checked="" type="checkbox"/>
4.189	MENTS/1008/SIGNATURE_9732482.PNG	798938550675	FLEKR2197T	EAST	03/06/2024 23:14:21	01/07/2024 05:57:40	11	1	2025-05-02 08:44:09+05:30	<input checked="" type="checkbox"/>
621	MENTS/90/SIGNATURE_8314634.PNG	6683 2026 2156	VIUPS0530P	EAST	24/08/2018 03:40:52	07/09/2018 01:57:48	12	10	2025-05-02 08:44:09+05:30	<input checked="" type="checkbox"/>
534	MENTS/90/SIGNATURE_8314634.PNG	6683 2026 2116	VIUPS0530P	EAST	24/08/2018 03:40:52	07/09/2018 01:57:48	12	10	2025-05-02 08:44:09+05:30	<input checked="" type="checkbox"/>
7.650	MENTS/91/SIGNATURE_7021052.PNG	5110 8483 0721	XKVPW2005R	SOUTH	07/02/2021 07:10:34	11/02/2021 22:40:58	13	1	2025-05-02 08:44:09+05:30	<input checked="" type="checkbox"/>
565	MENTS/91/SIGNATURE_7021052.PNG	5110 8483 0721	XKVPW2005R	SOUTH	07/02/2021 07:10:34	11/02/2021 22:40:58	13	1	2025-05-02 08:44:09+05:30	<input checked="" type="checkbox"/>
7.833	MENTS/98/SIGNATURE_3084703.PNG	680650072561	PDAPV2396K	NORTH	06/06/2024 23:01:39	29/06/2024 02:45:13	14	1	2025-05-02 08:44:09+05:30	<input checked="" type="checkbox"/>

Delete Selected Duplicates

- Records marked as 'X' will be deleted.
- To get complete stats of the dedupe go to the overview tab.

Note - Your deduplication result will be stored at destination details provided during configuration.

Automation Scheduler

Schedule/Update Job:

- a. To schedule the task and automate the dedupe process select the source table(s) you have already configured.

Inferenz Unify

Snowflake Native App for Enterprise-Grade Data Deduplication

Overview Table Management Dedupe & Review **Automation**

Automation Scheduler

Select Table(s)
Choose an option

Frequency
Daily

Time to Run
18:06

Enter Cron Expression (e.g., '0 10 * * * UTC')
e.g., 0 10 * * * UTC

Schedule / Update Dedupe Jobs Suspend Scheduled Jobs

- b. Once you select your source table(s), next select the frequency to run the job.
 - If the frequency selected Daily, Weekly or Monthly
 - Select time to run from the dropdown slot list provided and similarly Day of the Week/Day of the month respective to the frequency chosen.
- c. If the frequency selected Cron, then enter the Cron Expression (e.g., '0 10 * * * UTC')
- d. Click on the **Schedule/Update Dedupe Job** Button to automate the job.

Suspend Job:

- To suspend the existing Scheduled Dedupe Tasks/Jobs, select the source table(s) you have already configured and click on **Suspend Scheduled Jobs**.

Inferenz Unify

Snowflake Native App for Enterprise-Grade Data Deduplication

Overview Table Management Dedupe & Review Automation

Automation Scheduler

Select Table(s)

DEDUPE.TEST.C... x UNIFY.DEDUP_T... x

Frequency

Daily

Time to Run

11:38

Enter Cron Expression (e.g., '0 10 * * * UTC')

e.g., 0 10 * * * UTC

A task 'CUSTOMER_DATA_WITH_IND_REC_TEN_THOUSAND_WITH_DUPLICATES_LARGE_task' is already scheduled to run daily at 10:22 and the current status of the task is INACTIVE.

Schedule / Update Dedupe Jobs Suspend Scheduled Jobs

Last Executed Jobs:

- The Last Executed Jobs table will depict the details such as Table name, and latest task execution time for upto 10 records.

Last Executed Jobs

	Table	Last Executed At
0	CUSTOMER_DATA_WITH_IND_REC_TEN_THOUSAND_WITH_DUPLICATES_LARGE_2	2025-05-15 13:27:51
1	CUSTOMER_DATA_WITH_IND_REC_TEN_THOUSAND_WITH_DUPLICATES_LARGE_2	2025-05-15 11:45:28
2	CUSTOMER_DATA_WITH_IND_REC_TEN_THOUSAND_WITH_DUPLICATES_LARGE_2	2025-05-15 11:29:07
3	CUSTOMER_DATA_WITH_IND_REC_TEN_THOUSAND_WITH_DUPLICATES_LARGE_2	2025-05-15 11:18:47
4	CUSTOMERS_WITH_DUPLICATES1	2025-05-15 06:13:56
5	CUSTOMER_DATA_WITH_IND_REC_TEN_THOUSAND_WITH_DUPLICATES_LARGE	2025-05-15 06:12:03
6	CUSTOMERS_WITH_DUPLICATES1	2025-05-15 06:08:44
7	CUSTOMER_DATA_WITH_IND_REC_TEN_THOUSAND_WITH_DUPLICATES_LARGE_2	2025-05-15 05:44:28
8	CUSTOMER_DATA_WITH_IND_REC_TEN_THOUSAND_WITH_DUPLICATES_LARGE_2	2025-05-15 05:19:10
9	CUSTOMER_DATA_WITH_IND_REC_TEN_THOUSAND_WITH_DUPLICATES_LARGE_2	2025-05-14 13:17:08

6. Support

For further assistance, contact our support team at hello@inferenz.ai

