

Name: Xin Peng, Yuchen Zhao

Date: 11/23/2022

NYU NetID: xp2083, yz8759

Section: 3

Total in points:

Professor's Comments:

Project 2

Affirmation of my Independent Effort: Xin Peng , Yuchen Zhao

1. Collection of unstructured data

Get two dataset of health disease prediction in the format of csv.

<https://www.kaggle.com/datasets/akshaydattatraykhare/diabetes-dataset>

<https://www.kaggle.com/datasets/rishidamarla/cancer-patients-data>

Download the data and choose several columns and several fields to be used as our unstructured data to help to predict the probability that a customer has cancer/diabetes in the future.

For the cancer data, we choose id, gender, alcohol_use, generic_risk, balance_diet, smoking, snoring fields as cancer predicting factors, choose cancer_level as predicting result.

For the diabetes data, we choose glucose, blood_pressure, skin_thickness, insulin, bmi, diabetes_pedigree_func, age as diabetes predicting factors, choose is_diabetes as predicting result.

2. Create and optimize the logical database schema for both structured data and unstructured data

a) creation of the logical database schema

During the transformation of ER diagram to Schema, follows the procedure of regular entity, weak entity, binary 1:1 relationship, binary 1:N relationship, binary M:N relationship, multivalue attribute and specialization and conduct relative operations. Especially for binary 1:1 relationship, binary 1:N relationship, choose the foreign key approach.

Also, add the unstructured data's schema to the final schema.

The final schema is shown in the schema.pdf.

2.2 optimization of the logical database schema

2.2.1 ACCOUNT_MEMBER

All attributes are atomic, so it is in 1NF. There's only one attribute in the primary key, so it is in 2NF. There's no transitively dependent relationship in all nonprime attributes, so it is in 3NF.

2.2.2 PREVIOUS_ACCOUNTS

All attributes are atomic, so it is in 1NF. Since it's derived from a multi-value attribute of ACCOUNT, the primary key is all attributes. There's no non-prime attribute, so it is in 2NF. There's no transitively dependent relationship in all nonprime attributes, so it is in 3NF.

2.2.3 ACCOUNT

All attributes are atomic, so it is in 1NF. The primary key is all attributes. There's no non-prime attribute, so it is in 2NF. There's no transitively dependent relationship in all nonprime attributes, so it is in 3NF.

2.2.4 EMPLOY

All attributes are atomic, so it is in 1NF. It has the full functional dependency {EESsn, EAccAddress, EAccCode, EAccName} -> {EmployType,} and {EESsn, EAccAddress,

`EAccCode, EAccName}` is the primary key. It's in 2NF. There's no transitively dependent relationship in all nonprime attributes, so it is in 3NF.

2.2.5 ACCOUNT_ADMIN

All attributes are atomic, so it is in 1NF. There's only one attribute in the primary key, so it is in 2NF. There's no transitively dependent relationship in all nonprime attributes, so it is in 3NF.

2.2.6 ACCT_ACCTADMIN

All attributes are atomic, so it is in 1NF. It has the full functional dependency $\{AA_AccAddr, AA_AccCode, AA_AccName, AA_AdminSsn\} \rightarrow \{AdminRole, \dots\}$ and $\{AA_AccAddr, AA_AccCode, AA_AccName, AA_AdminSsn\}$ is the primary key. It's in 2NF. There's no transitively dependent relationship in all nonprime attributes, so it is in 3NF.

2.2.7 ACCOUNT_ALIAS

All attributes are atomic, so it is in 1NF. There's only one attribute in the primary key, so it is in 2NF. There's no transitively dependent relationship in all nonprime attributes, so it is in 3NF.

2.2.8 ISMASTEROF

All attributes are atomic, so it is in 1NF. It has the full functional dependency $\{I_AccAddr1, I_AccCode1, I_AccName1, I_AccAddr2, I_AccCode2, I_AccName2\} \rightarrow \{MasterType\}$ and $\{I_AccAddr1, I_AccCode1, I_AccName1, I_AccAddr2, I_AccCode2, I_AccName2\}$ is the primary key. It's in 2NF. There's no transitively dependent relationship in all nonprime attributes, so it is in 3NF.

2.2.9 BILLING_ACCOUNT

All attributes are atomic, so it is in 1NF. There's only one attribute in the primary key, so it is in 2NF. There's no transitively dependent relationship in all nonprime attributes, so it is in 3NF.

2.2.10 ACCT_BILLCOUNT

All attributes are atomic, so it is in 1NF. It has the full functional dependency $\{AB_AccAddr, AB_AccCode, AB_AccName, AB_BAName\} \rightarrow \{Usage, CoveredPerson\}$ and $\{AB_AccAddr, AB_AccCode, AB_AccName, AB_BAName\}$ is the primary key. It's in 2NF. There's no transitively dependent relationship in all nonprime attributes, so it is in 3NF.

2.2.11 CUS_CREATE_CONTRACT

All attributes are atomic, so it is in 1NF. The primary key is all attributes. There's no non-prime attribute, so it is in 2NF. There's no transitively dependent relationship

in all nonprime attributes, so it is in 3NF.

2.2.12 CUSTOMER

All attributes are atomic, so it is in 1NF. There's only one attribute in the primary key, so it is in 2NF. There's no transitively dependent relationship in all nonprime attributes, so it is in 3NF.

2.2.13 CUS_RELATION

All attributes are atomic, so it is in 1NF. It has the full functional dependency $\{CR_CusSsn1, CR_CusSsn2\} \rightarrow \{RType\}$ and $\{CR_CusSsn1, CR_CusSsn2\}$ is the primary key. It's in 2NF. There's no transitively dependent relationship in all nonprime attributes, so it is in 3NF.

2.2.14 ACC_EMPLOY_CUS

All attributes are atomic, so it is in 1NF. The primary key is all attributes. There's no non-prime attribute, so it is in 2NF. There's no transitively dependent relationship in all nonprime attributes, so it is in 3NF.

2.2.15 ACC_CREATE_CONTRACT

All attributes are atomic, so it is in 1NF. The primary key is all attributes. There's no non-prime attribute, so it is in 2NF. There's no transitively dependent relationship in all nonprime attributes, so it is in 3NF.

2.2.16 CONTRACT

All attributes are atomic, so it is in 1NF. It has the full functional dependency $\{\text{ContractNumber}, \text{AccountAddress}, \text{CompanyCode}, \text{AccountName}\} \rightarrow \{\text{PolicyOwner}, \dots\}$ and $\{\text{ContractNumber}, \text{AccountAddress}, \text{CompanyCode}, \text{AccountName}\}$ is the primary key. It's in 2NF. There's no transitively dependent relationship in all nonprime attributes, so it is in 3NF.

2.2.17 MANAGER_CONTRACT

All attributes are atomic, so it is in 1NF. The primary key is all attributes. There's no non-prime attribute, so it is in 2NF. There's no transitively dependent relationship in all nonprime attributes, so it is in 3NF.

2.2.18 CONTRACT_POLICY_PROTECTOR

All attributes are atomic, so it is in 1NF. There's only one attribute in the primary key, so it is in 2NF. There's no transitively dependent relationship in all nonprime attributes, so it is in 3NF.

2.2.19 BENEFIT

All attributes are atomic, so it is in 1NF. The primary key is all attributes. There's no non-prime attribute, so it is in 2NF. There's no transitively dependent relationship in all nonprime attributes, so it is in 3NF.

2.2.20 BENEFIT_BENEFICIARY

All attributes are atomic, so it is in 1NF. The primary key is all attributes. There's no non-prime attribute, so it is in 2NF. There's no transitively dependent relationship in all nonprime attributes, so it is in 3NF.

2.2.21 PREMIUM

All attributes are atomic, so it is in 1NF. The primary key is all attributes. There's no non-prime attribute, so it is in 2NF. There's no transitively dependent relationship in all nonprime attributes, so it is in 3NF.

2.2.22 ASSOCIATE

All attributes are atomic, so it is in 1NF. There's only one attribute in the primary key, so it is in 2NF. There's no transitively dependent relationship in all nonprime attributes, so it is in 3NF.

2.2.23 ASSO_RELATION

All attributes are atomic, so it is in 1NF. It has the full functional dependency $\{\text{AssociateSsn1}, \text{AssociateSsn2}\} \rightarrow \{\text{RelationType}\}$ and $\{\text{AssociateSsn1}, \text{AssociateSsn2}\}$ is the primary key. It's in 2NF.

2.2.24 WRITING_NUMBER

All attributes are atomic, so it is in 1NF. There's only one attribute in the primary key, so it is in 2NF. There's no transitively dependent relationship in all nonprime attributes, so it is in 3NF.

2.2.25 SITCODE

All attributes are atomic, so it is in 1NF. There's only one attribute in the primary key, so it is in 2NF. There's no transitively dependent relationship in all nonprime attributes, so it is in 3NF.

2.2.26 GET_HERITAGE

All attributes are atomic, so it is in 1NF. The primary key is all attributes. There's no non-prime attribute, so it is in 2NF. There's no transitively dependent relationship in all nonprime attributes, so it is in 3NF.

2.2.27 ACCOUNT_ASSO_RELATION

All attributes are atomic, so it is in 1NF. It has the full functional dependency

$\{AAR_AccAddr, AAR_AccCode, AAR_AccName, AAR_AssoSsn\} \rightarrow \{AAR_AssoRule\}$
and $\{AAR_AccAddr, AAR_AccCode, AAR_AccName, AAR_AssoSsn\}$ is the primary key.
There's no transitively dependent relationship in all nonprime attributes, so it is in 3NF.

3. Put all the data in the cloud platform of Aliyun's Dataworks

a) Unstructured data upload

(1)

Create the project space of xp_caroline to save all the tables related to the project.

(2)

Create the table xp_caroline.cancer_data_online to save collected cancer data. Use the import function to import cancer_data_online.xlsx into xp_caroline.cancer_data_online table.

The imported table looks like this:

Dimension Table / cancer_data_online								
Dimension table information	id	gender	alcohol_use	generic_risk	balance_diet	smoking	snoring	cancer_level
Workspace :	1 P10	1	4	3	2	3	4	Low
xp_caroline	2 P10	1	1	4	2	2	2	Medium
Table Name :	3 P100	1	5	5	6	2	2	High
cancer_data_online	4 P1000	1	7	6	7	7	5	High
Table Description :	5 P101	1	8	7	7	8	3	High
	6 P102	1	5	5	6	2	2	High
	7 P103	2	4	3	2	3	4	Low
	8 P104	2	1	2	4	1	3	Low
	9 P105	2	5	6	5	6	1	Medium
	10 P106	1	3	4	3	2	5	Medium
	11 P107	1	7	7	7	7	3	High
	12 P108	2	8	7	7	7	4	High
	13 P109	2	5	5	6	6	6	Medium
Lifecycle :	14 P11	1	7	6	7	7	5	High
1 year	15 P110	2	1	3	3	2	2	Low
	16 P111	1	6	6	6	8	2	Medium
	17 P112	1	1	4	2	2	2	Medium
--	18 P113	1	7	6	7	7	5	High
	19 P114	1	7	7	6	7	2	High
	20 P115	1	4	5	4	5	6	Medium
	21 P116	1	8	6	7	8	6	High
	22 P117	2	5	5	6	2	2	High
	23 P118	1	8	7	7	8	3	High
	24 P119	2	7	7	7	7	1	High
	25 P12	1	7	7	6	7	2	High

(3)

Create the related xp_caroline.diabetes_data_online table. Then use the import function to import diabetes_data_online.csv into xp_caroline.diabetes_data_online table.

The imported table looks like this:

	glucose	blood_pressure	skin_thickness	insulin	bmi	diabetes_pedigree_func	age	is_diabetes
1	148	72	35	0	33.6	0.627	50	1
2	85	66	29	0	26.6	0.351	31	0
3	183	64	0	0	23.3	0.672	32	1
4	89	66	23	94	28.1	0.167	21	0
5	137	40	35	168	43.1	2.288	33	1
6	116	74	0	0	25.6	0.201	30	0
7	78	50	32	88	31	0.248	26	1
8	115	0	0	0	35.3	0.134	29	0
9	197	70	45	543	30.5	0.158	53	1
10	125	96	0	0	0	0.232	54	1
11	110	92	0	0	37.6	0.191	30	0
12	168	74	0	0	38	0.537	34	1
13	139	80	0	0	27.1	1.441	57	0
14	189	60	23	846	30.1	0.398	59	1
15	166	72	19	175	25.8	0.587	51	1
16	100	0	0	0	30	0.484	32	1
17	118	84	47	230	45.8	0.551	31	1
18	107	74	0	0	29.6	0.254	31	1
19	103	30	38	83	43.3	0.183	33	0
20	115	70	30	96	34.6	0.529	32	1
21	126	88	41	235	39.3	0.704	27	0
22	99	84	0	0	35.4	0.388	50	0
23	196	90	0	0	39.8	0.451	41	1
24	119	80	35	0	29	0.263	29	1
25	143	94	33	146	36.6	0.254	51	1
26	125	70	26	115	31.1	0.205	41	1
27	147	76	0	0	39.4	0.257	43	1

b) Structured data generalization

(1) Create the customer data.

First, using some generalization program to generalize fake ssn data. Then set all rows' if_dead field to 0, all rows' if_associate field to 0, all rows c_first_name to 'xxx', all rows c_mid_name to 'xxx', all rows c_last_name to 'xxx'. Then for glucose, blood_pressure, skin_thickness, insulin, bmi and diabetes_pedigree_function fields, we use some data from cancer_data_online.xlsx. Then for age, gender, alcohol_use, generic_risk, balance_diet, smoking and snoring fields, we use some data from diabetes_data_online.csv.

(2) Generate the contract data

Using the customer table to generate contract data.

The main idea is to generate the unique number for the primary key contract_number and use the customer table's cssn as cus_ssn because cus_ssn is the foreign key of contract table and it is actually the primary key cssn of customer table.

The sql is like this:

```

INSERT INTO TABLE contract
SELECT ROW_NUMBER() OVER (ORDER BY cssn) as
, 'xx'
, (
    ROW_NUMBER() OVER (ORDER BY cssn)
) * 10 + 1
, 'xxx'
, cssn
, 'cancer'
, 'xx'
, 'xx'
, '20221001'

```

```

        , '20321001'
        , '612566236'
FROM      customer
;

INSERT INTO TABLE contract
SELECT   ROW_NUMBER() OVER (ORDER BY cssn ) + 210
        , 'xx'
        , (
                ROW_NUMBER() OVER (ORDER BY cssn )
        ) * 10 + 2
        , 'xxx'
        , cssn
        , 'diabetes'
        , 'xx'
        , 'xx'
        , '20220901'
        , '20320901'
        , '612566236'
FROM      customer
;

```

(3) Create other tables,

First, account related tables,

1) account_member

The screenshot shows the 'account_member' table details in a database management interface. The left panel displays the table's basic information, including its name, owner, and various metadata fields like 'Views', 'Reads', and 'Favorites'. The right panel provides a detailed view of the table's structure, including its fields, their types, descriptions, and primary key status. A search bar at the top right allows users to search for keywords.

No.	Field Name	Type	Description	Business Description	Number of Reads	Primary Key	Foreign Key
1	employee_ssn	string			★★★★★	Yes	-
2	employee_lname	string			★★★★★	No	-
3	employee_mname	string			★★★★★	No	-
4	employee_fname	string			★★★★★	No	-
5	cus_ssn	string			★★★★★	No	-

2) previous_accounts

Table Basic Information

No.	Field Name	Type	Description	Business Description	Number of Reads	Primary Key	Foreign Key
1	account_address	string			★★★★★	Yes	-
2	account_code	string			★★★★★	Yes	-
3	account_name	string			★★★★★	Yes	-
4	employee_ssn	string			★★★★★	Yes	-
5	start_date	string			★★★★★	Yes	-
6	end_date	string			★★★★★	Yes	-
7	employ_type	string			★★★★★	Yes	-
8	role	string			★★★★★	Yes	-
9	employ_period	int			★★★★★	Yes	-
10	salary	float			★★★★★	Yes	-

Business Information

Partition Fields

3) account

account

account 公司账户

Apply for Permission Join/View Album Add to Favorites Data Analysis Refresh

Search by keywords, such as the table name and table description Search

Table Basic Information

No.	Field Name	Type	Description	Business Description	Number of Reads	Primary Key	Foreign Key
1	account_address	string			★★★★★	Yes	-
2	account_code	string			★★★★★	Yes	-
3	account_name	string			★★★★★	Yes	-

Partition Fields

No.	Field Name	Display Name	Type	Description	Business Description
					No Data

4) employ

employ

account 和 account member 的雇佣关系

Apply for Permission Join/View Album Add to Favorites Data Analysis Refresh

Search by keywords, such as the table name and table description Search

Table Basic Information

No.	Field Name	Type	Description	Business Description	Number of Reads	Primary Key	Foreign Key
1	e_essn	string			★★★★★	Yes	-
2	e_acc_address	string			★★★★★	Yes	-
3	e_acc_code	string			★★★★★	Yes	-
4	e_acc_name	string			★★★★★	Yes	-
5	employ_end_date	string			★★★★★	No	-
6	employ_start_date	string			★★★★★	No	-
7	employ_type	string			★★★★★	No	-
8	role	string			★★★★★	No	-

Table Basic Information

0 0 0

Views 0 Reads 0 Favorites 0

Output Node : None
MaxCompute : odps.xp_caroline
Project :
Region of Current : us-west-1
Workspace :
Region of Compute : us-west-1
Engine Instance :
Owner : carolinetfs
Created At : Nov 22, 2022, 15:01:19
TTL (Days) : 360
Partitioned Table : No
Storage : 0 B
Table Description : account 和 account member 的雇佣关系
Table Type : Table
Tag : + New Tag

Business Information

No.	Field Name	Type	Description	Business Description	Number of Reads	Primary Key	Foreign Key
1	e_essn	string			★★★★★	Yes	-
2	e_aco_address	string			★★★★★	Yes	-
3	e_aco_code	string			★★★★★	Yes	-
4	e_acc_name	string			★★★★★	Yes	-
5	employ_end_date	string			★★★★★	No	-
6	employ_start_date	string			★★★★★	No	-
7	employ_type	string			★★★★★	No	-
8	role	string			★★★★★	No	-
9	salary	double			★★★★★	No	-
10	employ_period	int			★★★★★	No	-

Partition Fields

5) account_admin

account_admin

Apply for Permission Join/View Album Add to Favorites Data Analysis Refresh

Search by keywords, such as the table name and table description Search

Table Basic Information

0 0 0

Views 0 Reads 0 Favorites 0

Output Node : None
MaxCompute : odps.xp_caroline
Project :
Region of Current : us-west-1
Workspace :
Region of Compute : us-west-1
Engine Instance :
Owner : carolinetfs
Created At : Nov 22, 2022, 14:50:59
TTL (Days) : 360
Partitioned Table : No
Storage : 0 B
Table Description :
Table Type : Table
Tag : + New Tag

Business Information

No.	Field Name	Type	Description	Business Description	Number of Reads	Primary Key	Foreign Key
1	admin_ssn	string			★★★★★	Yes	-
2	admin_suffx	string			★★★★★	No	-
3	admin_fname	string			★★★★★	No	-
4	admin_mname	string			★★★★★	No	-
5	admin_lname	string			★★★★★	No	-

Partition Fields

No.	Field Name	Display Name	Type	Description	Business Description
No Data					

6) acct_acctadmin

acct_acctadmin

account与account_admin的关系表 Apply for Permission Join/View Album Add to Favorites Data Analysis Refresh

Search by keywords, such as the table name and table description Search

Table Basic Information

0 0 0

Views 0 Reads 0 Favorites 0

Output Node : None
MaxCompute : odps.xp_caroline
Project :
Region of Current : us-west-1
Workspace :
Region of Compute : us-west-1
Engine Instance :
Owner : carolinetfs
Created At : Nov 22, 2022, 14:55:24
TTL (Days) : 360
Partitioned Table : No
Storage : 0 B
Table Description : account与account_admin的关系表
Table Type : Table

Business Information

No.	Field Name	Type	Description	Business Description	Number of Reads	Primary Key	Foreign Key
1	aa_admin_ssn	string			★★★★★	Yes	-
2	aa_aco_address	string			★★★★★	Yes	-
3	aa_aco_code	string			★★★★★	Yes	-
4	aa_acc_name	string			★★★★★	Yes	-
5	admin_role	string			★★★★★	No	-
6	s_date	string			★★★★★	No	-
7	e_date	string			★★★★★	No	-

Partition Fields

7) account_alias

account_alias

公司别名

Views 0 Reads 0 Favorites 0

Output Node : None
MaxCompute : odpsxp_caroline
Project : us-west-1
Region of Current : us-west-1
Workspace : us-west-1
Region of Compute : us-west-1
Engine Instance :
Owner : carolinetfls
Created At : Nov 22, 2022, 15:19:27
TTL (Days) : 366
Partitioned Table : No
Storage : 0 B
Table Description : 公司别名
Table Type : Table

No.	Field Name	Type	Description	Business Description	Number of Reads	Primary Key	Foreign Key
1	alias_name	string			★★★★★	Yes	-
2	aal_acc_address	string			★★★★★	No	-
3	aal_acc_code	string			★★★★★	No	-
4	aal_acc_name	string			★★★★★	No	-

Partition Fields

No.	Field Name	Display Name	Type	Description	Business Description
					No Data

8) is_master_of

is_master_of

公司和公司的关系

Views 0 Reads 0 Favorites 0

Output Node : None
MaxCompute : odpsxp_caroline
Project : us-west-1
Region of Current : us-west-1
Workspace : us-west-1
Region of Compute : us-west-1
Engine Instance :
Owner : carolinetfls
Created At : Nov 22, 2022, 15:36:10
TTL (Days) : 366
Partitioned Table : No
Storage : 0 B
Table Description : 公司和公司的关系
Table Type : Table

No.	Field Name	Type	Description	Business Description	Number of Reads	Primary Key	Foreign Key
1	i_acc_addr1	string			★★★★★	Yes	-
2	i_acc_code1	string			★★★★★	Yes	-
3	i_acc_name1	string			★★★★★	Yes	-
4	i_acc_addr2	string			★★★★★	Yes	-
5	i_acc_code2	string			★★★★★	Yes	-
6	i_acc_name2	string			★★★★★	Yes	-
7	master_type	string			★★★★★	No	-

Partition Fields

No.	Field Name	Display Name	Type	Description	Business Description
					No Data

9) billing_account

billing_account

公司财务账户

Views 0 Reads 0 Favorites 0

Output Node : None
MaxCompute : odpsxp_caroline
Project : us-west-1
Region of Current : us-west-1
Workspace : us-west-1
Region of Compute : us-west-1
Engine Instance :
Owner : carolinetfls
Created At : Nov 22, 2022, 15:23:23
TTL (Days) : 366
Partitioned Table : No
Storage : 0 B
Table Description : 公司财务账户
Table Type : Table

No.	Field Name	Type	Description	Business Description	Number of Reads	Primary Key	Foreign Key
1	b_account_name	string			★★★★★	Yes	-
2	b_account_address	string			★★★★★	No	-
3	b_account_type	string			★★★★★	No	-

Partition Fields

No.	Field Name	Display Name	Type	Description	Business Description
					No Data

10) acct_billaccount

acct_billaccount

公司和财务账户之间的关系

Views: 0 Reads: 0 Favorites: 0

Output Node: None
MaxCompute: odps.xp_caroline
Project: us-west-1
Region of Current: us-west-1
Workspace:
Region of Compute: us-west-1
Engine Instance:
Owner: carolinetfls
Created At: Nov 22, 2022, 15:27:59
TTL (Days): 366
Partitioned Table: No
Storage: 0 B
Table Description: 公司和财务账户之间的关系
Table Type: Table

No.	Field Name	Type	Description	Business Description	Number of Reads	Primary Key	Foreign Key
1	ab_acc_addr	string			★★★★★	Yes	-
2	ab_acc_code	string			★★★★★	Yes	-
3	ab_acc_name	string			★★★★★	Yes	-
4	ab_ba_name	string			★★★★★	Yes	-
5	usage	string			★★★★★	No	-
6	covered_person	string			★★★★★	No	-

Partition Fields

No.	Field Name	Display Name	Type	Description	Business Description

Secondly, customer related tables,

11) cus_create_contract

cus_create_contract

customer 创建 contract

Views: 0 Reads: 0 Favorites: 0

Output Node: None
MaxCompute: odps.xp_caroline
Project: us-west-1
Region of Current: us-west-1
Workspace:
Region of Compute: us-west-1
Engine Instance:
Owner: carolinetfls
Created At: Nov 22, 2022, 16:09:11
TTL (Days): 366
Partitioned Table: No
Storage: 0 B
Table Description: customer 创建 contract
Table Type: Table

No.	Field Name	Type	Description	Business Description	Number of Reads	Primary Key	Foreign Key
1	ccc_cus_ssn	string			★★★★★	Yes	-
2	ccc_contract_number	string			★★★★★	Yes	-

Partition Fields

No.	Field Name	Display Name	Type	Description	Business Description
					No Data

12) customer

customer

null。业务咨询: carolinetfls(null) #

Views: 0 Reads: 0 Favorites: 0

Output Node: None
MaxCompute: odps.xp_caroline
Project: us-west-1
Region of Current: us-west-1
Workspace:
Region of Compute: us-west-1
Engine Instance:
Owner: carolinetfls
Created At: Nov 22, 2022, 00:11:11
TTL (Days): 360
Partitioned Table: No
Storage: 0 B
Table Description: null。业务咨询: carolinetfls(null)
Table Type: Table

No.	Field Name	Type	Description	Business Description	Number of Reads	Primary Key	Foreign Key
1	cssn	string			★★★★★	Yes	-
2	if_dead	string			★★★★★	No	-
3	if_associate	string			★★★★★	No	-
4	c_first_name	string			★★★★★	No	-
5	c_last_name	string			★★★★★	No	-
6	c_mid_name	string			★★★★★	No	-
7	glucose	string			★★★★★	No	-
8	blood_pressure	string			★★★★★	No	-

Table Type : Table

Tag : + New Tag

Business Information

Workspace : xp_caroline

Environment : Production

Category : -

Display Name : -

Permission Information

My Permissions : - View More

Technical Information

DDL Statement Nov 22, 2022, 00:11:11

Updated At :

Data Updated At : Nov 22, 2022, 00:11:11

Data Viewed At : -

Compute Engine View

No.	Field Name	Display Name	Type	Description	Business Description
9	skin_thickness	string			★★★★★ No -
10	insulin	string			★★★★★ No -
11	bmi	string			★★★★★ No -
12	diabetes_pedigree_func	string			★★★★★ No -
13	age	string			★★★★★ No -
14	gender	string			★★★★★ No -
15	alcohol_use	string			★★★★★ No -
16	generic_risk	string			★★★★★ No -
17	balance_diet	string			★★★★★ No -
18	smoking	string			★★★★★ No -
19	snoring	string			★★★★★ No -

Partition Fields

No.	Field Name	Display Name	Type	Description	Business Description
No Data					

Some Customer table preview data

Table Basic Information

0 0 0

Views 0 Reads 0 Favorites 0

Output Node : None

MaxCompute : odps.xp_caroline

Project : -

Region of Current : us-west-1

Workspace :

Region of Compute : us-west-1

Engine Instance :

Owner : carolinetfls

Created At : Nov 22, 2022, 00:11:11

TTL (Days) : 360

Partitioned Table : No

Storage 0 B

Table Description : null, 业务咨询: carolinetfls(null)#

Table Type : Table

Tag : + New Tag

Business Information

Workspace : xp_caroline

Table Basic Information

0 0 0

Views 0 Reads 0 Favorites 0

Output Node : None

MaxCompute : odps.xp_caroline

Project : -

Region of Current : us-west-1

Workspace :

Region of Compute : us-west-1

Engine Instance :

Owner : carolinetfls

Created At : Nov 22, 2022, 00:11:11

TTL (Days) : 360

Partitioned Table : No

Storage 0 B

Table Description : null, 业务咨询: carolinetfls(null)#

Table Type : Table

Tag : + New Tag

Business Information

Details Output Lineage Usage Notes Data Quality Data Preview

C Refresh Go to DataAnalysis

Field Name	cssn 9999/0174	if_dead	if_associate	c_first_name	c_last_name	c_mid_name	glucose	blood_press ure
0	u	v	xxx	xxx	xxx	xxx	//	92
6	473438713	0	0	xxx	xxx	xxx	115	98
7	286898090	0	0	xxx	xxx	xxx	150	76
8	734290195	0	0	xxx	xxx	xxx	120	76
9	506004254	0	0	xxx	xxx	xxx	161	68
10	716566914	0	0	xxx	xxx	xxx	137	68
11	841368587	0	0	xxx	xxx	xxx	128	68
12	887619022	0	0	xxx	xxx	xxx	124	68
13	239168708	0	0	xxx	xxx	xxx	80	66
14	822058177	0	0	xxx	xxx	xxx	106	70
15	811328565	0	0	xxx	xxx	xxx	155	74
16	559888270	0	0	xxx	xxx	xxx	113	50
17	441735186	0	0	xxx	xxx	xxx	109	80
18	435845681	0	0	xxx	xxx	xxx	112	68
19	778846290	0	0	xxx	xxx	xxx	99	80
20	597311337	0	0	xxx	xxx	xxx	182	74

Details Output Lineage Usage Notes Data Quality Data Preview

C Refresh Go to DataAnalysis

Field Name	skin_thickne ss	insulin	bmi	diabetes_pe digree_func	age	gender	alcohol_use	generic_risk
3	0	0	32.3	0.839	30	2	6	3
4	33	135	31.6	0.422	25	1	1	4
5	41	42	35.8	0.156	35	2	2	2
6	0	0	52.9	0.209	28	1	8	7
7	0	0	21	0.207	37	2	8	7
8	37	105	39.7	0.215	29	2	1	2
9	23	132	25.5	0.326	47	2	2	3
10	14	148	24.8	0.143	21	2	2	2
11	19	180	30.5	1.391	25	1	8	7
12	28	205	32.9	0.875	30	1	7	7
13	30	0	26.2	0.313	41	1	8	7
14	37	148	39.4	0.605	22	1	7	7
15	17	96	26.6	0.433	27	1	3	2
16	10	85	29.5	0.626	25	1	8	6

Table Basic Information

Field Name	diabetes_pe_digree_func	age	gender	alcohol_use	generic_risk	balance_diet	smoking	snoring
2	0.84	58	2	6	7	7	3	1
3	0.839	30	2	6	3	6	3	5
4	0.422	25	1	1	4	2	1	2
5	0.156	35	2	2	2	3	3	5
6	0.209	28	1	8	7	2	1	2
7	0.207	37	2	8	7	2	1	2
8	0.215	29	2	1	2	2	1	3
9	0.326	47	2	2	3	3	2	4
10	0.143	21	2	2	2	3	3	5
11	1.391	25	1	8	7	2	1	2
12	0.875	30	1	7	7	7	7	3
13	0.313	41	1	8	7	7	7	4
14	0.605	22	1	7	7	7	4	5
15	0.433	27	1	3	2	2	2	4
16	0.626	25	1	8	6	7	8	6

Business Information

13) cus_relation

cus_relation

customer之间的关系

server adopts GMT-8. The time gap is 3 hours.

Search by keywords, such as the table name and table description |

Table Basic Information

Field Name	Type	Description	Business Description	Number of Reads	Primary Key	Foreign Key
cr_cus_ssn1	string			★★★★★	Yes	-
cr_cus_ssn2	string			★★★★★	Yes	-
r_type	string			★★★★★	No	-

Partition Fields

No.	Field Name	Display Name	Type	Description	Business Description
No Data					

14) acc_employ_cus

acc_employ_cus

公司和客户customer之间的关系

Search by keywords, such as the table name and table description |

Table Basic Information

Field Name	Type	Description	Business Description	Number of Reads	Primary Key	Foreign Key
aec_acc_address	string			★★★★★	Yes	-
aec_acc_code	string			★★★★★	Yes	-
aec_acc_name	string			★★★★★	Yes	-
aec_cus_ssn	string			★★★★★	Yes	-

Partition Fields

No.	Field Name	Display Name	Type	Description	Business Description
No Data					

Thirdly, contract related tables,

15) acc_create_contract

acc_create_contract

account 创建 contract

Views: 0 Reads: 0 Favorites: 0

Output Node: None

MaxCompute: odps.xp_caroline

Project: us-west-1

Region of Current: us-west-1

Workspace:

Region of Compute: us-west-1

Engine Instance:

Owner: carolinetfs

Created At: Nov 22, 2022, 16:07:04

TTL (Days): 366

Partitioned Table: No

Storage: 0 B

Table Description: account 创建 contract

Table Type: Table

No.	Field Name	Type	Description	Business Description	Number of Reads	Primary Key	Foreign Key
1	acc_acc_addr	string			★★★★★	Yes	-
2	acc_acc_code	string			★★★★★	Yes	-
3	acc_acc_name	string			★★★★★	Yes	-
4	acc_contract_number	string			★★★★★	Yes	-

Partition Fields

No.	Field Name	Display Name	Type	Description	Business Description
No Data					

16) contract

contract

保险合同

Views: 0 Reads: 0 Favorites: 0

Output Node: None

MaxCompute: odps.xp_caroline

Project: us-west-1

Region of Current: us-west-1

Workspace:

Region of Compute: us-west-1

Engine Instance:

Owner: carolinetfs

Created At: Nov 22, 2022, 20:45:44

TTL (Days): 366

Partitioned Table: No

Storage: 0 B

Table Description: 保险合同

Table Type: Table

Tag: + New Tag

No.	Field Name	Type	Description	Business Description	Number of Reads	Primary Key	Foreign Key
1	contract_number	string			★★★★★	No	-
2	account_address	string			★★★★★	No	-
3	account_code	string			★★★★★	No	-
4	account_name	string			★★★★★	No	-
5	cus_ssn	string			★★★★★	No	-
6	policy	string			★★★★★	No	-
7	policy_owner	string			★★★★★	No	-
8	policy_funder	string			★★★★★	No	-
9	c_end_date	string			★★★★★	No	-
10	c_start_date	string			★★★★★	No	-
11	selling_asso	string			★★★★★	No	-

Business Information

No.	Field Name	Display Name	Type	Description	Business Description
No Data					

Permission Information

My Permissions: - View More

Technical Information

DDL Statement: Nov 22, 2022, 20:45:44

Updated At: Nov 22, 2022, 20:45:44

Data Updated At: Nov 22, 2022, 20:45:44

Data Viewed At: -

Some contract table preview data:

contract

保险合同

Apply for Permission Join/View Album Add to Favorites Data Analysis Refresh

Search by keywords, such as the table name and table description

Table Basic Information

0 0 0

Views Reads Favorites

Output Node : None
MaxCompute : odps.xp_caroline
Project :
Region of Current : us-west-1
Workspace :
Region of Compute : us-west-1
Engine Instance :
Owner : carolinetfls
Created At : Nov 22, 2022, 20:45:44
TTL (Days) : 366
Partitioned Table : No
Storage : 0 B
Table Description : 保险合同
Table Type : Table
Tag : + New Tag

Data Preview

Field Name	contract_number	account_address	account_code	account_name	cus_ssn	policy	policy_owner	policy_funder
1	211	xx	12	xxx	200281499	diabetes	xx	xx
2	212	xx	22	xxx	206176379	diabetes	xx	xx
3	213	xx	32	xxx	221107278	diabetes	xx	xx
4	214	xx	42	xxx	224471346	diabetes	xx	xx
5	215	xx	52	xxx	227556266	diabetes	xx	xx
6	216	xx	62	xxx	229396476	diabetes	xx	xx
7	217	xx	72	xxx	230170987	diabetes	xx	xx
8	218	xx	82	xxx	230428183	diabetes	xx	xx

contract

保险合同

Apply for Permission Join/View Album Add to Favorites Data Analysis Refresh

Search by keywords, such as the table name and table description

Table Basic Information

0 0 0

Views Reads Favorites

Output Node : None
MaxCompute : odps.xp_caroline
Project :
Region of Current : us-west-1
Workspace :
Region of Compute : us-west-1
Engine Instance :
Owner : carolinetfls
Created At : Nov 22, 2022, 20:45:44
TTL (Days) : 366
Partitioned Table : No
Storage : 0 B
Table Description : 保险合同
Table Type : Table

Data Preview

Field Name	id	account_name	cus_ssn	policy	policy_owner	policy_funder	c_end_date	c_start_date	selling_association
1	xxx	200281499	diabetes	xx	xx	xx	20220901	20320901	612566236
2	xxx	206176379	diabetes	xx	xx	xx	20220901	20320901	612566236
3	xxx	221107278	diabetes	xx	xx	xx	20220901	20320901	612566236
4	xxx	224471346	diabetes	xx	xx	xx	20220901	20320901	612566236
5	xxx	227556266	diabetes	xx	xx	xx	20220901	20320901	612566236
6	xxx	229396476	diabetes	xx	xx	xx	20220901	20320901	612566236
7	xxx	230170987	diabetes	xx	xx	xx	20220901	20320901	612566236

17) manager_contract

manager_contract

manager保险合同

Apply for Permission Join/View Album Add to Favorites Data Analysis Refresh

Search by keywords, such as the table name and table description

Table Basic Information

0 0 0

Views Reads Favorites

Output Node : None
MaxCompute : odps.xp_caroline
Project :
Region of Current : us-west-1
Workspace :
Region of Compute : us-west-1
Engine Instance :
Owner : carolinetfls
Created At : Nov 22, 2022, 15:59:48
TTL (Days) : 366
Partitioned Table : No
Storage : 0 B
Table Description : manager保险合同
Table Type : Table

Data Preview

No.	Field Name	Type	Description	Business Description	Number of Reads	Primary Key	Foreign Key
1	mc_contract_number	string			★★★★★	Yes	-
2	mc_account_address	string			★★★★★	Yes	-
3	mc_account_code	string			★★★★★	Yes	-
4	mc_account_name	string			★★★★★	Yes	-
5	mc_cus_ssn	string			★★★★★	Yes	-
6	sitcode	string			★★★★★	Yes	-

Partition Fields

No.	Field Name	Display Name	Type	Description	Business Description
-----	------------	--------------	------	-------------	----------------------

18) contract_policy_protector

contract_policy_protector

contract_policy_protector

Views 0 Reads 0 Favorites 0

Output Node : None

MaxCompute odps.xp_caroline

Project : Region of Current us-west-1

Workspace : Region of Compute us-west-1

Engine Instance : Owner : carolinetfls

Created At : Nov 22, 2022, 16:16:29

TTL (Days) : 366

Partitioned Table : No

Storage 0 B

Table Description : contract_policy_protect or

Table Type : Table

Details							
Field Information		Partition Information		Change Records			
No.	Field Name	Type	Description	Business Description	Number of Reads	Primary Key	Foreign Key
1	cpp_contract_number	string			★★★★★	Yes	-
2	cpp_acc_addr	string			★★★★★	Yes	-
3	cpp_acc_code	string			★★★★★	Yes	-
4	cpp_acc_name	string			★★★★★	Yes	-
5	cpp_cus_ssn	string			★★★★★	Yes	-
6	policy_protector	string			★★★★★	No	-

Partition Fields

No.	Field Name	Display Name	Type	Description	Business Description

19) benefit

benefit

contract中的benefit

Views 0 Reads 0 Favorites 0

Output Node : None

MaxCompute odps.xp_caroline

Project : Region of Current us-west-1

Workspace : Region of Compute us-west-1

Engine Instance : Owner : carolinetfls

Created At : Nov 22, 2022, 16:13:38

TTL (Days) : 366

Partitioned Table : No

Storage 0 B

Table Description : contract中的benefit

Table Type : Table

Details							
Field Information		Partition Information		Change Records			
No.	Field Name	Type	Description	Business Description	Number of Reads	Primary Key	Foreign Key
1	be_contract_number	string			★★★★★	Yes	-
2	be_acc_addr	string			★★★★★	Yes	-
3	be_acc_code	string			★★★★★	Yes	-
4	be_acc_name	string			★★★★★	Yes	-
5	be_cus_ssn	string			★★★★★	Yes	-
6	benefit_number	string			★★★★★	Yes	-

Partition Fields

No.	Field Name	Display Name	Type	Description	Business Description

20) benefit_beneficiary

benefit_beneficiary

contract中的benefit的beneficiary

Views 0 Reads 0 Favorites 0

Output Node : None

MaxCompute odps.xp_caroline

Project : Region of Current us-west-1

Workspace : Region of Compute us-west-1

Engine Instance : Owner : carolinetfls

Created At : Nov 22, 2022, 16:26:16

TTL (Days) : 366

Partitioned Table : No

Storage 0 B

Table Description : contract中的benefit的 beneficiary

Table Type : Table

Details							
Field Information		Partition Information		Change Records			
No.	Field Name	Type	Description	Business Description	Number of Reads	Primary Key	Foreign Key
1	bef_contract_number	string			★★★★★	Yes	-
2	bef_acc_addr	string			★★★★★	Yes	-
3	bef_acc_code	string			★★★★★	Yes	-
4	bef_acc_name	string			★★★★★	Yes	-
5	bef_cus_ssn	string			★★★★★	Yes	-
6	bef_benefit_number	string			★★★★★	Yes	-
7	beneficiary	string			★★★★★	Yes	-

Partition Fields

No.	Field Name	Display Name	Type	Description	Business Description

21) premium

premium

contract中的benefit的premium

Views 0 Reads 0 Favorites 0

Output Node : None
MaxCompute : odps.xp_caroline
Project :
Region of Current : us-west-1
Workspace :
Region of Compute : us-west-1
Engine Instance :
Owner : carolinetfs
Created At : Nov 22, 2022, 16:31:55
TTL (Days) : 366
Partitioned Table : No
Storage : 0 B
Table Description : contract中的benefit的 premium
Table Type : Table

No.	Field Name	Type	Description	Business Description	Number of Reads	Primary Key	Foreign Key
1	p_acc_addr	string			★★★★★	Yes	-
2	p_acc_code	string			★★★★★	Yes	-
3	p_acc_name	string			★★★★★	Yes	-
4	p_cus_ssn	string			★★★★★	Yes	-
5	p_contract_number	string			★★★★★	Yes	-
6	p_benefit_number	string			★★★★★	Yes	-
7	preminimum_number	string			★★★★★	Yes	-
8	amount	double			★★★★★	No	-

The fourth, associate related tables.

22) Associate

associate

associate 保险销售员

Views 0 Reads 0 Favorites 0

Output Node : None
MaxCompute : odps.xp_caroline
Project :
Region of Current : us-west-1
Workspace :
Region of Compute : us-west-1
Engine Instance :
Owner : carolinetfs
Created At : Nov 22, 2022, 16:38:17
TTL (Days) : 366
Partitioned Table : No
Storage : 0 B
Table Description : associate 保险销售员
Table Type : Table

No.	Field Name	Type	Description	Business Description	Number of Reads	Primary Key	Foreign Key
1	associate_ssn	string			★★★★★	Yes	-
2	asso_fname	string			★★★★★	No	-
3	asso_lname	string			★★★★★	No	-
4	asso_mname	string			★★★★★	No	-
5	reporting_chain	string			★★★★★	No	-
6	commission_rate	float			★★★★★	No	-
7	production_credit	float			★★★★★	No	-
8	sum_commissions	double			★★★★★	No	-
9	selling_area	string			★★★★★	No	-

23) Asso_relation

asso_relation

asso之间关系

Views 0 Reads 0 Favorites 0

Output Node : None
MaxCompute : odps.xp_caroline
Project :
Region of Current : us-west-1
Workspace :
Region of Compute : us-west-1
Engine Instance :
Owner : carolinetfs
Created At : Nov 22, 2022, 16:41:19
TTL (Days) : 366
Partitioned Table : No
Storage : 0 B
Table Description : asso之间关系
Table Type : Table

No.	Field Name	Type	Description	Business Description	Number of Reads	Primary Key	Foreign Key
1	associate_ssn1	string			★★★★★	Yes	-
2	associate_ssn2	string			★★★★★	Yes	-
3	relation_type	string			★★★★★	No	-

Partition Fields

No.	Field Name	Display Name	Type	Description	Business Description
					No Data

24) writing_number

writing_number

writing number

Views 0 Reads 0 Favorites 0

Output Node : None
MaxCompute : odps.xp_caroline
Project :
Region of Current : us-west-1
Workspace :
Region of Compute : us-west-1
Engine Instance :
Owner : carolinetfls
Created At : Nov 22, 2022, 16:54:07
TTL (Days) : 366
Partitioned Table : No
Storage : 0 B
Table Description : writing number
Table Type : Table

Field Information							
No.	Field Name	Type	Description	Business Description	Number of Reads	Primary Key	Foreign Key
1	writing_number	string			★★★★★	Yes	-
2	auth_asso_ssn	string			★★★★★	No	-

Partition Fields

No.	Field Name	Display Name	Type	Description	Business Description
No Data					

25) sitcode

sitcode

server adopts GMT+8, the time gap is 3 hours.

Views 0 Reads 0 Favorites 0

Output Node : None
MaxCompute : odps.xp_caroline
Project :
Region of Current : us-west-1
Workspace :
Region of Compute : us-west-1
Engine Instance :
Owner : carolinetfls
Created At : Nov 22, 2022, 16:56:44
TTL (Days) : 366
Partitioned Table : No
Storage : 0 B
Table Description : sitcode
Table Type : Table

Field Information							
No.	Field Name	Type	Description	Business Description	Number of Reads	Primary Key	Foreign Key
1	sitcode	string			★★★★★	Yes	-
2	writing_number	string			★★★★★	No	-

Partition Fields

No.	Field Name	Display Name	Type	Description	Business Description
No Data					

26) account_asso_relation

account_asso_relation

account asso relation

Views 0 Reads 0 Favorites 0

Output Node : None
MaxCompute : odps.xp_caroline
Project :
Region of Current : us-west-1
Workspace :
Region of Compute : us-west-1
Engine Instance :
Owner : carolinetfls
Created At : Nov 22, 2022, 19:41:37
TTL (Days) : 366
Partitioned Table : No
Storage : 0 B
Table Description : account asso relation
Table Type : Table

Field Information							
No.	Field Name	Type	Description	Business Description	Number of Reads	Primary Key	Foreign Key
1	aar_acc_addr	string			★★★★★	Yes	-
2	aar_acc_code	string			★★★★★	Yes	-
3	aar_acc_name	string			★★★★★	Yes	-
4	aar_asso_ssn	string			★★★★★	Yes	-
5	asso_rule	string			★★★★★	No	-

Partition Fields

No.	Field Name	Display Name	Type	Description	Business Description
No Data					

27) get_heritage

get_heritage

asso relatives get heritage and become customer

Apply for Permission Join/View Album Add to Favorites Data Analysis Refresh

Search by keywords, such as the table name and table description Search

Table Basic Information		Details						Output		Lineage		Usage Notes		Data Quality		Data Preview	
Views	Reads	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Views	Reads	Favorites	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Output Node :		None						MaxCompute		odpsxp_caroline							
Project :								Region of Current		us-west-1							
Workspace :								Region of Compute		us-west-1							
Engine Instance :								Owner :		carolnetfls							
Created At :								Created At :		Nov 22, 2022, 19:37:52							
TTL (Days) :								TTL (Days) :		366							
Partitioned Table :								Storage :		0 B							
Table Description :		asso relatives get heritage and become customer						Table Description :		asso relatives get heritage and become customer							

Field Information Partition Information Change Records

Edit Upload Download Generate SELECT Statement Generate DDL Statement

No.	Field Name	Type	Description	Business Description	Number of Reads	Primary Key	Foreign Key
1	gh_asso_ssn	string			★★★★★	Yes	-
2	gh_cus_ssn	string			★★★★★	Yes	-

Partition Fields

No.	Field Name	Display Name	Type	Description	Business Description
No Data					

28) cancer_data_online

just as shown before

DataWorks DataAnalysis Homepage SQL Query Spreadsheet Dimension Table More English

Dimension Table / cancer_data_online

Dimension table information	id	gender	lcohol_us	generic_risalance_di	smoking	snoring	ancer_level	
Workspace : xp.caroline	1 P1	1	4	3	2	3	4	Low
Table Name : cancer_data_online	2 P10	1	1	4	2	2	2	Medium
Table Description :	3 P100	1	5	5	6	2	2	High
Lifecycle : 1 year	4 P1000	1	7	6	7	7	5	High
	5 P101	1	8	7	7	8	3	High
	6 P102	1	5	5	6	2	2	High
	7 P103	2	4	3	2	3	4	Low
	8 P104	2	1	2	4	1	3	Low
	9 P105	2	5	6	5	6	1	Medium
	10 P106	1	3	4	3	2	5	Medium
	11 P107	1	7	7	7	7	3	High
	12 P108	2	8	7	7	7	4	High
	13 P109	2	5	5	6	6	6	Medium
	14 P11	1	7	6	7	7	5	High
	15 P110	2	1	3	3	2	2	Low
	16 P111	1	6	6	6	8	2	Medium
	17 P112	1	1	4	2	2	2	Medium
	18 P113	1	7	6	7	7	5	High
	19 P114	1	7	7	6	7	2	High
	20 P115	1	4	5	4	5	6	Medium
	21 P116	1	8	6	7	8	6	High
	22 P117	2	5	5	6	2	2	High
	23 P118	1	8	7	7	8	3	High
	24 P119	2	7	7	7	7	1	High
	25 P12	1	7	7	6	7	2	High
	26 P120	1	8	7	2	1	2	Medium
	27 P121	2	1	2	2	1	3	Medium

29) diabetes_data_online

DataWorks DataAnalysis Homepage SQL Query Spreadsheet Dimension Table More English

Dimension Table / diabetes_data_online

Dimension table information	glucose	blood_pressure	skin_thickness	insulin	bmi	diabetes_pedigree_func	age	is_diabetes
Workspace : xp.caroline	1 148	72	35	0	33.6	0.627	50	1
Table Name : diabetes_data_online	2 85	66	29	0	26.6	0.351	31	0
Table Description :	3 183	64	0	0	23.3	0.672	32	1
Lifecycle : 1 year	4 89	66	23	94	28.1	0.167	21	0
	5 137	40	35	168	43.1	2.288	33	1
	6 116	74	0	0	25.6	0.201	30	0
	7 78	50	32	88	31	0.248	26	1
	8 115	0	0	0	35.3	0.134	29	0
	9 197	70	45	543	30.5	0.158	53	1
	10 125	96	0	0	0	0.232	54	1
	11 110	92	0	0	37.6	0.191	30	0
	12 168	74	0	0	38	0.537	34	1
	13 139	80	0	0	27.1	1.441	57	0
	14 189	60	23	846	30.1	0.398	59	1
	15 166	72	19	175	25.8	0.587	51	1
	16 100	0	0	0	30	0.484	32	1
	17 118	84	47	230	45.8	0.551	31	1
	18 107	74	0	0	29.6	0.254	31	1
	19 103	30	38	83	43.3	0.183	33	0
	20 115	70	30	96	34.6	0.529	32	1
	21 126	88	41	235	39.3	0.704	27	0
	22 99	84	0	0	35.4	0.388	50	0
	23 196	90	0	0	39.8	0.451	41	1
	24 119	80	35	0	29	0.263	29	1
	25 143	94	33	146	36.6	0.254	51	1
	26 125	70	26	115	31.1	0.205	41	1
	27 147	76	0	0	39.4	0.257	43	1

4. data insights

We can calculate median or average values with diseases for single factors. We can combine the data we get from customers and from online data so that we can have more data to use.

(1) for cancer

the sql is like this:

```
SELECT  is_cancer
        ,AVG(alcohol_use)  avgAlcoholUse
        ,MEDIAN(alcohol_use)  medAlcoholUse
        ,AVG(generic_risk)  avgGenericRisk
        ,MEDIAN(generic_risk)  medGenericRisk
        ,AVG(balance_diet)  avgBalanceDiet
        ,MEDIAN(balance_diet)  medBalanceDiet
        ,AVG(smoking)  avgSmoking
        ,MEDIAN(smoking)  medSmoking
        ,AVG(snoring)  avgSnoring
        ,MEDIAN(snoring)  medSnoring
FROM    (
            SELECT  alcohol_use
                    ,generic_risk
                    ,balance_diet
                    ,smoking
                    ,snoring
                    ,CASE WHEN cancer_level = 'Low' THEN 0
                          ELSE 1
                      END AS is_cancer
            FROM    xp_caroline.cancer_data_online
            UNION ALL
            SELECT  alcohol_use
                    ,generic_risk
                    ,balance_diet
                    ,smoking
                    ,snoring
                    ,b.is_pay AS is_cancer
            FROM    xp_caroline.customer a
            JOIN    xp_caroline.contract b
            ON      a.cssn = b.cus_ssn
            WHERE   b.policy = 'cancer'
            AND     b.is_pay = 1
        )  tmp
```

```

GROUP BY is_cancer
;

```

the result is like this:

is_cancer	avgAlcohol	medAlcohol	avgGenericRisk	medGenericRisk	avgBalanceDiet	medBalanceDiet	avgSmoking	medSmoking	avgSnoring	medSnoring
0	2.25	2	2.71875	2	3	3	2.96875	3	2.21875	2
1	4.804034582 13257	5	4.760806916 42651	5	4.6138328 5302594	5	4.0547550 4322767	3	3.051873198 84726	3

(2) for diabetes

the sql is like this:

```

SELECT  is_diabetes
        ,AVG(glucose)  avgGlucose
        ,MEDIAN(glucose)  medGlucose
        ,AVG(blood_pressure)  avgBloodPressure
        ,MEDIAN(blood_pressure)  medBloodPressure
        ,AVG(insulin)  avgInsulin
        ,MEDIAN(insulin)  medInsulin
        ,AVG(bmi)  avgBmi
        ,AVG(diabetes_pedigree_func)  avgPedigree
        ,AVG(age)  avgAge
        ,MEDIAN(age)  medAge
FROM    (
        SELECT  glucose
                ,blood_pressure
                ,skin_thickness
                ,insulin
                ,bmi
                ,diabetes_pedigree_func
                ,age
                ,cast(is_diabetes as int)  is_diabetes
        FROM    xp_caroline.diabetes_data_online
        UNION ALL
        SELECT  glucose
                ,blood_pressure
                ,skin_thickness
                ,insulin
                ,bmi

```

```

,diabetes_pedigree_func
,age
,is_pay AS is_diabetes
FROM      xp_caroline.customer a
JOIN      xp_caroline.contract b
ON        a.cssn = b.cus_ssn
WHERE     b.policy = 'diabetes'
AND      b.is_pay = 1
) tmp
GROUP BY is_diabetes
;

```

the result is like this :

is_diabetes	avgGlucose	medGlucose	avgBloodPressure	medBloodPressure	avgInsulin	medInsulin	avgBmi	avgPedigree	avgAge	medAge
0	106.56	105	65.98	68	52.57	0	28.79	0.42	31.22	27
1	123.28	119	68.98	72	83.54	50	32.19	0.5	34.14	30

5. data management

5.1 metadata management

Aliyun provides a data map product which can manage table's metadata quite well.
It allows users to see the field, primary key and field descriptions.

Views	0	Reads	0	Favorites	0	
Output Node	None	MaxCompute	odps.xp_caroline	Project		
Region of Current	us-west-1	Workspace		Region of Compute	us-west-1	
Engine Instance		Owner	carolinettls	Created At	Nov 22, 2022, 00:11:11	
TTL (Days)	360	Partitioned Table	No	Storage	0 B	
Table Description	null, 业务咨询: carolinettls(null) #					

No.	Field Name	Type	Description	Business Description	Number of Reads	Primary Key	Foreign Key
1	cssn	string			★★★★★	Yes	-
2	if_dead	string			★★★★★	No	-
3	if_associate	string			★★★★★	No	-
4	c_first_name	string			★★★★★	No	-
5	c_last_name	string			★★★★★	No	-
6	c_mid_name	string			★★★★★	No	-
7	glucose	string			★★★★★	No	-
8	blood_pressure	string			★★★★★	No	-

It also allows users to see the field lineage(the function is paid)

It allows users to see usage notes, if there is any.

Server adopts GMT+8. The time gap is 3 hours.

customer

null。业务咨询: carolinetts(null)#
 [Apply for Permission](#) [Join/View Album](#) [Add to Favorites](#) [Data Analysis](#) [Refresh](#)

Search by keywords, such as the table name and table description [Search](#)

Table Basic Information

Views	Reads	Favorites
1	0	0

Output Node: None
MaxCompute: odps.xp_caroline
Project:
Region of Current: us-west-1
Workspace:
Region of Compute: us-west-1
Engine Instance:
Owner: carolinetts
Created At: Nov 22, 2022, 16:11:11
TTL (Days): 360
Partitioned Table: No
Storage: 6.54 KB

[Edit](#) [View Versions](#) [View Markdown Syntax](#)

For security purposes, the system automatically verifies and converts the content that you entered.

Details **Output** **Lineage** **Usage Notes** **Data Quality** **Data Preview**

It can show data quality situations.

customer

null。业务咨询: carolinetts(null)#
 [Apply for Permission](#) [Join/View Album](#) [Add to Favorites](#) [Data Analysis](#) [Refresh](#)

Search by keywords, such as the table name and table description [Search](#)

Table Basic Information

Views	Reads	Favorites
0	0	0

Output Node: None
MaxCompute: odps.xp_caroline
Project:
Region of Current: us-west-1
Workspace:
Region of Compute: us-west-1
Engine Instance:
Owner: carolinetts
Created At: Nov 22, 2022, 00:11:11
TTL (Days): 360
Partitioned Table: No
Storage: 0 B
Table Description: null。业务咨询: carolinetts(null)#
 [Edit](#)

Data Quality

[Configure Rules](#)

Rule Details

Rule Type	Rules
No Data	

Data Quality Rules

[View More](#)

Rule ID	Attribute	Strong or Weak	Rule Type	Check Type	Expected Value	Actual Value
No Data						

And it gives a preview of the data.

customer

null。业务咨询: carolinetts(null)#
 [Apply for Permission](#) [Join/View Album](#) [Add to Favorites](#) [Data Analysis](#) [Refresh](#)

Search by keywords, such as the table name and table description [Search](#)

Table Basic Information

Views	Reads	Favorites
0	0	0

Output Node: None
MaxCompute: odps.xp_caroline
Project:
Region of Current: us-west-1
Workspace:
Region of Compute: us-west-1
Engine Instance:
Owner: carolinetts
Created At: Nov 22, 2022, 00:11:11
TTL (Days): 360
Partitioned Table: No
Storage: 0 B
Table Description: null。业务咨询: carolinetts(null)#
 [Edit](#)

[C Refresh](#) [Go to DataAnalysis](#)

Data Preview

Field Name	cssn	if_dead	if_associate	c_first_name	c_last_name	c_mid_name	glucose	blood_press ure	ski ss
1	251963128	0	0	xxx	xxx	xxx	100	78	25
2	402521267	0	0	xxx	xxx	xxx	112	72	0
3	672038652	0	0	xxx	xxx	xxx	167	0	0
4	765424251	0	0	xxx	xxx	xxx	144	58	33
5	999878174	0	0	xxx	xxx	xxx	77	82	41
6	473438713	0	0	xxx	xxx	xxx	115	98	0
7	286898090	0	0	xxx	xxx	xxx	150	76	0

More conveniently, especially when the table is really big, it generates select sql or creates sql for users so that one doesn't need to find and type each field.

The screenshot shows the Aliyun Operation Center interface with two modal windows open over a table's basic information page.

Table Basic Information (Left Panel):

- Views: 0
- Reads: 0
- Favorites: 0
- Output Node: None
- MaxCompute: odps.xp_caroline
- Project: (empty)
- Region of Current: us-west-1
- Region of Compute: us-west-1
- Engine Instance: (empty)
- Owner: carolinetfls
- Created At: Nov 22, 2022, 00:11:11
- TTL (Days): 360
- Partitioned Table: No
- Storage: 0 B
- Table Description: null, 业务咨询: carolinetfls(null)#
 - Table Type: Table
 - Tag: + New Tag

Generate SELECT Statement (Top Modal):

```

1 SELECT
2   csn,
3   if_dead,
4   if_associate,
5   c_first_name,
6   c_last_name,
7   c_mid_name,
8   glucose,
9   blood_pressure,
10  skin_thickness,
11  insulin,
12  bmi,
13  diabetes_pedigree_func,
14  age,
15  gender,
16  alcohol_use,
17  generic_risk,
18  balance_diet,
19  smoking,
20  snoring
21 FROM xp_caroline.customer
22
  
```

Generate DDL Statement (Bottom Modal):

```

1 CREATE TABLE IF NOT EXISTS customer
2 (
3   csn          STRING,
4   if_dead      STRING,
5   if_associate STRING,
6   c_first_name STRING,
7   c_last_name  STRING,
8   c_mid_name   STRING,
9   glucose      STRING,
10  blood_pressure STRING,
11  skin_thickness STRING,
12  insulin     STRING,
13  bmi          STRING,
14  diabetes_pedigree_func STRING,
15  age          STRING,
16  gender       STRING,
17  alcohol_use  STRING,
18  generic_risk  STRING,
19  balance_diet  STRING,
20  smoking      STRING,
21  snoring      STRING,
22 )
23 COMMENT 'null, 业务咨询: carolinetfls(null)#
24 LIFECYCLE 360';
  
```

5.2 data quality and data governance

Aliyun's operation center provides functions for the check of cycle tasks. Where each table which updates as a cycle will be a virtual node. And all the tasks will form the graph of nodes by the dependencies between them. As you can see clearly how the graph looks like. Also, you can set rules to check whether the invariants of the table have been violated. As soon as the violation happens, it will send a text/email to the related person immediately.

