

Alibaba Cloud

数据安全中心 FAQ

Document Version: 20220217

Legal disclaimer

Alibaba Cloud reminds you to carefully read and fully understand the terms and conditions of this legal disclaimer before you read or use this document. If you have read or used this document, it shall be deemed as your total acceptance of this legal disclaimer.

1. You shall download and obtain this document from the Alibaba Cloud website or other Alibaba Cloud-authorized channels, and use this document for your own legal business activities only. The content of this document is considered confidential information of Alibaba Cloud. You shall strictly abide by the confidentiality obligations. No part of this document shall be disclosed or provided to any third party for use without the prior written consent of Alibaba Cloud.
2. No part of this document shall be excerpted, translated, reproduced, transmitted, or disseminated by any organization, company or individual in any form or by any means without the prior written consent of Alibaba Cloud.
3. The content of this document may be changed because of product version upgrade, adjustment, or other reasons. Alibaba Cloud reserves the right to modify the content of this document without notice and an updated version of this document will be released through Alibaba Cloud-authorized channels from time to time. You should pay attention to the version changes of this document as they occur and download and obtain the most up-to-date version of this document from Alibaba Cloud-authorized channels.
4. This document serves only as a reference guide for your use of Alibaba Cloud products and services. Alibaba Cloud provides this document based on the "status quo", "being defective", and "existing functions" of its products and services. Alibaba Cloud makes every effort to provide relevant operational guidance based on existing technologies. However, Alibaba Cloud hereby makes a clear statement that it in no way guarantees the accuracy, integrity, applicability, and reliability of the content of this document, either explicitly or implicitly. Alibaba Cloud shall not take legal responsibility for any errors or lost profits incurred by any organization, company, or individual arising from download, use, or trust in this document. Alibaba Cloud shall not, under any circumstances, take responsibility for any indirect, consequential, punitive, contingent, special, or punitive damages, including lost profits arising from the use or trust in this document (even if Alibaba Cloud has been notified of the possibility of such a loss).
5. By law, all the contents in Alibaba Cloud documents, including but not limited to pictures, architecture design, page layout, and text description, are intellectual property of Alibaba Cloud and/or its affiliates. This intellectual property includes, but is not limited to, trademark rights, patent rights, copyrights, and trade secrets. No part of this document shall be used, modified, reproduced, publicly transmitted, changed, disseminated, distributed, or published without the prior written consent of Alibaba Cloud and/or its affiliates. The names owned by Alibaba Cloud shall not be used, published, or reproduced for marketing, advertising, promotion, or other purposes without the prior written consent of Alibaba Cloud. The names owned by Alibaba Cloud include, but are not limited to, "Alibaba Cloud", "Aliyun", "HiChina", and other brands of Alibaba Cloud and/or its affiliates, which appear separately or in combination, as well as the auxiliary signs and patterns of the preceding brands, or anything similar to the company names, trade names, trademarks, product or service names, domain names, patterns, logos, marks, signs, or special descriptions that third parties identify as Alibaba Cloud and/or its affiliates.
6. Please directly contact Alibaba Cloud for any errors of this document.

Document conventions









Style	Description	Example
 Danger	A danger notice indicates a situation that will cause major system changes, faults, physical injuries, and other adverse results.	 Danger: Resetting will result in the loss of user configuration data.
 Warning	A warning notice indicates a situation that may cause major system changes, faults, physical injuries, and other adverse results.	 Warning: Restarting will cause business interruption. About 10 minutes are required to restart an instance.
 Notice	A caution notice indicates warning information, supplementary instructions, and other content that the user must understand.	 Notice: If the weight is set to 0, the server no longer receives new requests.
 Note	A note indicates supplemental instructions, best practices, tips, and other content.	 Note: You can use Ctrl + A to select all files.
>	Closing angle brackets are used to indicate a multi-level menu cascade.	Click Settings > Network > Set network type .
Bold	Bold formatting is used for buttons, menus, page names, and other UI elements.	Click OK .
Courier font	Courier font is used for commands	Run the <code>cd /d C:/window</code> command to enter the Windows system folder.
<i>Italic</i>	Italic formatting is used for parameters and variables.	<code>bae log list --instanceid</code> <i>Instance_ID</i>
[] or [a b]	This format is used for an optional value, where only one item can be selected.	<code>ipconfig [-all -t]</code>
{ } or {a b}	This format is used for a required value, where only one item can be selected.	<code>switch {active stand}</code>

Table of Contents

1.Overview	05
2.Data security	06
3.Data authorization	07
4.Sensitive data scan and detection	08
5.Data de-identification	10
6.Supported OSS files	11
7.Supported data de-identification algorithms	15
8.Built-in anomalous activity detection rules	22
9.Limits on using watermarks	25

1. Overview

The following part lists the topics that provide answers to frequently asked questions about Data Security Center (DSC).

Features

Data security

Data authorization

Sensitive data scan and detection

Data de-identification

Limits on using watermarks

Supported items

Supported sensitive data

Supported OSS files

Built-in anomalous activity detection rules

Supported data de-identification algorithms

2.Data security

This topic provides answers to commonly asked questions about data security in Sensitive Data Discovery and Protection (SDDP).

Does DSC store my data?

Data Security Center (DSC) does not store your data. After you authorize SDDP to access data assets, DSC scans data in the data assets. Then, SDDP displays the scan results in the DSC console for your use.

How can I audit the operations that are related to DSC?

DSC records all operations in ActionTrail by using the ActionTrail API.

DSC is integrated with ActionTrail. After you activate ActionTrail, you can view all the operations that are related to DSC and review their security in ActionTrail. For more information about how to activate ActionTrail, see [Billing](#).

3.Data authorization

When you authorize Data Security Center (DSC) to access data in MaxCompute, ApsaraDB RDS, or Object Storage Service (OSS), the authorization may fail. This topic describes the possible causes of an authorization failure to help you troubleshoot the failure.

What are the possible causes for the failure to authorize DSC to access ApsaraDB RDS?

- The username or password for accessing the ApsaraDB RDS database is invalid.
- The service IP addresses of DSC are deleted from the whitelist of the ApsaraDB RDS database.
- The ApsaraDB RDS database resides on the classic network, but the public endpoint of the ApsaraDB RDS database is inaccessible due to access control.

What are the possible causes for the failure to authorize DSC to access MaxCompute?

- The name of the MaxCompute project is invalid.
- The DSC account fails to be added to the MaxCompute project.

References

[Grant access to data assets](#)

4. Sensitive data scan and detection

This topic provides answers to frequently asked questions about sensitive data scan and detection.

- [What types of data assets can scan?](#)
- [How long does it take to scan data in my data asset after I authorize DSC to access the data asset?](#)
- [How does scan data in an unstructured data asset, such as an OSS bucket?](#)
- [Can DSC rescan an OSS object after the object is scanned?](#)
- [How does scan data in a structured data asset, such as a MaxCompute project or an ApsaraDB RDS database?](#)
- [Does log on to a database to obtain data?](#)
- [When will a rescan be triggered?](#)

What types of data assets can DSC scan?

DSC can scan data assets that store structured data or unstructured data. DSC can scan the following types of data assets:

- ApsaraDB RDS databases and self-managed databases, which store structured data
- MaxCompute projects, which store structured data
- Object Storage Service (OSS) buckets, which store unstructured data

How long does it take to scan data in my data asset after I authorize DSC to access the data asset?

DSC starts to scan your data asset within 2 hours after it is authorized to access the data asset. The time taken to scan your data depends on the data volume. If a data asset contains a large number of tables, for example, more than 10,000 tables, it takes a long period of time to scan the data asset. If the total size of objects stored in an OSS bucket is large, for example, more than 1 PB, it also takes a long period of time to scan the OSS bucket. When DSC scans your data, the scan results are progressively updated on the **Overview** page in the [DSC console](#). For more information, see [View summary information](#).

How does DSC scan data in an unstructured data asset, such as an OSS bucket?

DSC scans data that is stored in an unstructured data asset and determines whether the objects are sensitive.

- **First scan:** After you authorize DSC to access an OSS bucket, DSC scans all objects that are stored in the OSS bucket.
- **Scan of incremental data:** If you add objects to or modify objects stored in the OSS bucket, DSC scans the new or modified objects.

Can DSC rescan an OSS object after the object is scanned?

If the object remains unchanged, DSC does not rescan it. If you modify the object, DSC rescans the object within 4 to 8 hours after the modification.

How does DSC scan data in a structured data asset, such as a MaxCompute project or an ApsaraDB RDS database?

DSC scans the names and values of fields in databases or projects, and determines whether the fields are sensitive. For example, DSC scans the name and values of the age field. If DSC cannot determine whether a field is sensitive based only on the values of the field, DSC also checks the name of the field to determine whether the field is sensitive.


- **First scan:** After you authorize DSC to access a database or project, DSC scans all tables in the database or project.
- **Scan of incremental data:** If you add tables to the database or project, DSC scans the new tables. If you modify the schema of an existing table by changing fields, DSC rescans the table.

Does DSC log on to a database to obtain data?

If authorized, DSC logs on to a database and samples data to detect sensitive data. DSC does not save data from databases or MaxCompute projects.

When will a rescan be triggered?

DSC automatically rescans data in an authorized data asset in the scenarios described in the following table.

Scenario	Scan logic	Billing
You authorize DSC to access your data asset for the first time.	DSC scans all data in the data asset.	DSC charges you for a full scan on data in the data asset.
You modify data in a data asset after DSC has scanned the data asset with authorization.	If you add fields to or delete fields from a MaxCompute or database table, DSC automatically rescans the table. If you add rows to or delete rows from a table, DSC does not automatically rescan the table.	DSC charges you for a full scan on data in the data asset.
	If you add objects to or modify objects stored in an OSS bucket, DSC automatically scans the new or modified objects. <div> Note If you only delete objects from an OSS bucket, DSC does not automatically rescan the bucket.</div>	DSC charges you for scanning the new or modified objects.
You change sensitive data detection rules. For example, you create, delete, enable, or disable rules.	DSC automatically scans all data in all authorized data assets.	DSC charges you for a full scan on data in all authorized data assets.

5.Data de-identification

Does static de-identification affect original data?

Static de-identification does not affect original data. The static de-identification feature only reads data, de-identifies the data, and saves the de-identified data to the location that you specify. The feature does not modify the original data.

6.Supported OSS files

This topic describes the types of Object Storage Service (OSS) files from which Data Security Center (DSC) can detect sensitive data.

No.	File	No.	File
1	C or C++ source file	85	Tokyo Cabinet database file
2	Lua file	86	X3D model XML file
3	JavaScript source file	87	XML file
4	VRML source file	88	XML sitemap file
5	BCPL source file	89	DBF file
6	Windows initialization file	90	PGP file
7	Java source file	91	FTP session file
8	BAT file	92	Binary file
9	Objective-C source file	93	EML file
10	Pascal source file	94	Visio file
11	Perl source file	95	iWork file
12	Python source file	96	WPD file
13	Ruby source file	97	WPS file
14	TCL source file	98	Microsoft Works file
15	Java JCE KeyStore file	99	Microsoft Office file
16	Java KeyStore file	100	XPS file
17	Shell script	101	Email file
18	HTML file	102	VCF file
19	Go source file	103	Microsoft Reader file
20	DataX configuration file	104	Excel file
21	Internet Information Services (IIS) configuration file	105	Outlook file
22	Tomcat configuration file	106	Word file
23	MaxCompute configuration file	107	PDF file

No.	File	No.	File
24	OpenVPN configuration file	108	FDF file
25	OSS configuration file	109	Lotus Multi-Byte Character Set (LMBCS) file
26	Tomcat application configuration file	110	Lotus Word Pro file
27	Tomcat users configuration file	111	PowerPoint file
28	WebLogic configuration file	112	SVG image XML file
29	SSH configuration file	113	CAD file
30	DICOM data file	114	MP4 file
31	HDF file	115	MPEG file
32	GIF file	116	Audio and video files
33	JP2 file	117	MP4A file
34	JPEG file	118	3GPP file
35	JPM file	119	GnuCash XML file
36	JPX file	120	3GPP2 file
37	JXR file	121	H.264 file
38	PCX file	122	MJ2 file
39	PNG file	123	MP2 file
40	TIFF file	124	MP2T video file
41	Photoshop file	125	SSH public key
42	DjVu file	126	MP4V file
43	Icon file	127	Text file
44	BIOS logo file	128	MPEG-4 file
45	BMP file	129	MPV file
46	Canon CR2 file	130	QuickTime video file
47	Canon CR file	131	DVB file
48	Cartesian Perceptual Compression file	132	FLC file

No.	File	No.	File
49	DPX file	133	FLI file
50	EPS file	134	JNG file
51	OpenEXR file	135	M4V file
52	Gem file	136	MNG file
53	macOS icon file	137	ASF file
54	Full-text index in the Windows help system	138	SG file
55	Windows installation information	139	ACC file
56	NIFF file	140	SSH private key
57	Olympus ORF file	141	M4A file
58	Paint.NET file	142	DOS executable file
59	Windows Help documentation	143	Empty file
60	Polar Monitor Bitmap file	144	COM file
61	Windows precompiled file	145	ELF file
62	Greymap file	146	Object file
63	Pixmap file	147	tcpdump capture file
64	QuickTime image file	148	7-ZIP file
65	TGA file	149	BZIP2 file
66	Windows INF file	150	Cabinet file
67	X3F original image file	151	Common compressed file
68	XPM file	152	LHA file
69	XWD file	153	LRZIP file
70	BPG file	154	LZ4 file
71	MongoDB database file	155	LZIP file
72	MySQL database file	156	UNIX compressed file
73	Oracle database file	157	LZMA file
74	PostgreSQL database file	158	GZIP file

No.	File	No.	File
75	SQLite3 database file	159	XZ file
76	Redis database file	160	Zstandard compressed file
77	SQL Server database file	161	Zstandard dictionary file
78	Berkeley database file	162	ZIP file
79	dBase database file	163	ZLIB file
80	GNU dbm or ndbm database file	164	Qpress compressed file
81	Access database file	165	Snappy file
82	Scalable storage engine database file	166	RAR file
83	MSVC database file	167	TAR file
84	Windows application compatibility database	N/A	N/A

7. Supported data de-identification algorithms

This topic describes the data de-identification algorithms that are supported by Data Security Center (DSC).

Category	Description	Algorithm	Input	Applicable sensitive data and scenario
Hashing	<p>Raw data cannot be retrieved after it is de-identified by using this type of algorithm.</p> <p>This type of algorithm is applicable to password protection or scenarios in which you must check whether data is sensitive by comparison.</p> <p>You can use common hash algorithms and specify a salt value.</p>	MD5	Salt value	<ul style="list-style-type: none"> • Sensitive data: keys • Scenario: data storage
		Secure Hash Algorithm 1 (SHA-1)	Salt value	
		SHA-256	Salt value	
		Hash-based Message Authentication Code (HMAC)	Salt value	
		Keeps the first N characters and the last M characters	Values of N and M	
		Keeps characters from the Xth position to the Yth position	Values of X and Y	
		Redacts the first N characters and the last M characters	Values of N and M	

Category	Description	Algorithm	Input	Applicable sensitive data and scenario
Redaction by using asterisks (*) or number signs (#)	<p>Raw data cannot be retrieved after it is de-identified by using this type of algorithm.</p> <p>This type of algorithm is applicable to scenarios in which sensitive data is to be shown on a user interface or shared with others.</p> <p>This type of algorithm redacts specified text in sensitive data with asterisks (*) or number signs (#).</p>	Redacts characters from the Xth position to the Yth position	Values of X and Y	<ul style="list-style-type: none"> • Sensitive data: sensitive personal information • Scenarios: <ul style="list-style-type: none"> ◦ Data usage ◦ Data sharing
		Redacts characters that precede a special character when the special character appears for the first time	At sign (@), ampersand (&), or period (.)	
		Redacts characters that follow a special character when the special character appears for the first time	At sign (@), ampersand (&), or period (.)	

Category	Description	Algorithm	Input	Applicable sensitive data and scenario
		Substitutes specific content in ID card numbers with mapped values	Mapping table for substituting the IDs of administrative regions	
		Randomly substitutes specific content in ID card numbers	Code table for randomly substituting the IDs of administrative regions	
		Randomly substitutes specific content in the IDs of military officer cards	Code table for randomly substituting type codes	
		Randomly substitutes specific content in passport numbers	Code table for randomly substituting purpose fields	
		Randomly substitutes specific content in permit numbers of Exit-Entry Permits for Travelling to and from Hong Kong and Macao	Code table for randomly substituting purpose fields	

Category	Description	Algorithm	Input	Applicable sensitive data and scenario
Substitution (customization supported)	<p>Raw data can be retrieved after it is de-identified by using some of the algorithms.</p> <p>This type of algorithm can be used to de-identify fields in fixed formats, such as ID card numbers.</p> <p>This type of algorithm substitutes the entire value or part of the value of a field with a mapped value by using a mapping table. In this case, raw data can be retrieved after it is de-identified. This type of algorithm also substitutes the entire value or part of the value of a field randomly based on a random interval. In this case, raw data cannot be retrieved after it is de-identified. DSC provides multiple built-in mapping tables and allows you to customize substitution algorithms.</p>	Randomly substitutes specific content in bank card numbers	Code table for randomly substituting Bank Identification Numbers (BINs)	<ul style="list-style-type: none"> • Sensitive data: <ul style="list-style-type: none"> ◦ Sensitive personal information ◦ Sensitive information of enterprises ◦ Sensitive information of devices • Scenarios: <ul style="list-style-type: none"> ◦ Data storage ◦ Data sharing
		Randomly substitutes specific content in landline telephone numbers	Code table for randomly substituting the IDs of administrative regions	
		Randomly substitutes specific content in mobile numbers	Code table for randomly substituting mobile network codes	
		Randomly substitutes specific content in unified social credit codes	Code table for randomly substituting the IDs of registration authorities, code table for randomly substituting type codes, and code table for randomly substituting the IDs of administrative regions	

Category	Description	Algorithm	Input	Applicable sensitive data and scenario
		Substitutes specific content in general tables with mapped values	Mapping table for substituting uppercase letters, mapping table for substituting lowercase letters, mapping table for substituting digits, and mapping table for substituting special characters	
		Randomly substitutes specific content in general tables	Code table for randomly substituting uppercase letters, code table for randomly substituting lowercase letters, code table for randomly substituting digits, and code table for randomly substituting special characters	

Category	Description	Algorithm	Input	Applicable sensitive data and scenario
Rounding	<p>Raw data can be retrieved after it is de-identified by using some of the algorithms.</p> <p>This type of algorithm can be used to analyze and collect statistics on sensitive datasets.</p> <p>DSC provides two types of rounding algorithms. One algorithm rounds numbers and dates, and raw data cannot be retrieved after it is de-identified. The other algorithm bit-shifts text, and raw data can be retrieved after it is de-identified.</p>	Rounds numbers	Numbers are rounded to the Nth digit before the decimal point. Valid values of N: 1 to 19.	<ul style="list-style-type: none"> • Sensitive data: general sensitive information • Scenarios: <ul style="list-style-type: none"> ◦ Data storage ◦ Data usage
		Rounds dates	Dates are rounded to the year, month, day, hour, or minute level.	
		Shifts characters	Number of places by which specific bits are moved and shift direction (left or right)	
Encryption	<p>Raw data can be retrieved after it is de-identified by using this type of algorithm.</p> <p>This type of algorithm can be used to encrypt sensitive fields that need to be retrieved after encryption.</p> <p>Common symmetrical encryption algorithms are supported.</p>	Data Encryption Standard (DES) algorithm	Encryption key	<ul style="list-style-type: none"> • Sensitive data: <ul style="list-style-type: none"> ◦ Sensitive personal information ◦ Sensitive information of enterprises • Scenario: data storage
		Triple Data Encryption Standard (3DES) algorithm	Encryption key	
		Advanced Encryption Standard (AES) algorithm	Encryption key	

Category	Description	Algorithm	Input	Applicable sensitive data and scenario
Shuffling	<p>Raw data cannot be retrieved after it is de-identified by using this type of algorithm.</p> <p>This type of algorithm can be used to de-identify structured data columns.</p> <p>This type of algorithm extracts values of a field in a specified range from the source table and rearranges the values in a specific column. Alternatively, this type of algorithm randomly selects values from a specific column within the value range and rearranges the selected values. This way, the values are mixed up and de-identified.</p>	Randomly shuffles data	Shuffle method: rearrangement or random selection	<ul style="list-style-type: none">• Sensitive data:<ul style="list-style-type: none">◦ Sensitive information of devices◦ Sensitive location information• Scenario: data storage

8. Built-in anomalous activity detection rules

This topic describes the built-in anomalous activity detection rules that are supported by Sensitive Data Discovery and Protection (SDDP).

Model type	Model name	Anomaly description	Supported service
Anomalous data flow	Sensitive data download in an unusual location	An external attacker obtains the logon credentials of an account and uses the account to download sensitive data.	Object Storage Service (OSS), ApsaraDB RDS, and MaxCompute
	Sensitive data download on an unusual terminal	An external attacker obtains the logon credentials of an account and uses the account to download sensitive data, or an employee downloads sensitive data to a personal terminal.	OSS
	Sensitive data download during an unusual period	An external attacker obtains the logon credentials of an account and uses the account to download sensitive data, or an employee downloads sensitive data after working hours.	OSS, ApsaraDB RDS, and MaxCompute
	Sensitive data download for the first time	An account is mistakenly granted the permission to download sensitive data.	OSS, ApsaraDB RDS, and MaxCompute
	Anomalous volume of downloaded sensitive data	An external attacker obtains the logon credentials of an account and uses the account to download sensitive data, or an employee maliciously backs up sensitive data.	OSS, ApsaraDB RDS, and MaxCompute
	Download of unnecessary sensitive tables	An account is mistakenly granted the permission to download sensitive data.	ApsaraDB RDS and MaxCompute
	Unusual low log output	The log feature encounters a failure. As a result, anomalous data operations cannot be effectively detected.	OSS, ApsaraDB RDS, and MaxCompute
	Anomalous volume of downloaded objects	An external attacker obtains the logon credentials of an account and uses the account to download sensitive data, or an employee maliciously backs up sensitive data.	OSS

Model type	Model name	Anomaly description	Supported service
	Anomalous volume of downloaded data	An external attacker obtains the logon credentials of an account and uses the account to download sensitive data, or an employee maliciously backs up sensitive data.	ApsaraDB RDS and MaxCompute
Anomalous permission access	Unusual logon time	An external attacker obtains the logon credentials of an account and uses the account to log on to the service, or an employee logs on to the service after working hours.	OSS, ApsaraDB RDS, and MaxCompute
	Unusual logon terminal	An external attacker obtains the logon credentials of an account and uses the account to log on to the service, or an employee logs on to the service on a personal terminal.	OSS, ApsaraDB RDS, and MaxCompute
	Unusual logon location	An external attacker obtains the logon credentials of an account and uses the account to log on to the service.	OSS, ApsaraDB RDS, and MaxCompute
	Download of sensitive objects from an unusual OSS bucket	An account is mistakenly granted the permission to download sensitive data.	OSS
	No protection for a sensitive MaxCompute project	Protection is disabled for a sensitive MaxCompute project. As a result, the MaxCompute project is not protected when data flows out of it. For more information, see Project data protection .	MaxCompute
	LabelSecurity disabled for a sensitive MaxCompute project	LabelSecurity is disabled for a sensitive MaxCompute project. As a result, the workspace administrator cannot control the access of users to sensitive data in the MaxCompute project. For more information, see Column-level access control .	MaxCompute
	Sensitive OSS bucket at the security level of public	The security level of a sensitive OSS bucket is set to public. As a result, external users can access sensitive data in the OSS bucket by calling an API operation.	OSS
	Beyond the maximum idle period for a permission	An unnecessary permission is granted, which violates the principle of minimum authorization. It is difficult to detect external attackers who have obtained such permissions.	OSS, ApsaraDB RDS, and MaxCompute

Model type	Model name	Anomaly description	Supported service
	Access to an object that does not exist for multiple times	An external attacker repeatedly makes access attempts.	OSS
	Access to an unauthorized object for multiple times	An external attacker repeatedly makes access attempts.	OSS
	Multiple failed access attempts	An external attacker repeatedly makes access attempts.	OSS, ApsaraDB RDS, and MaxCompute
Anomalous data operation	Anomalously low risk level marked for a MaxCompute project	The risk level marked for a MaxCompute project is maliciously lowered. As a result, permission control loses effectiveness and data security protection cannot cover all sensitive data.	MaxCompute
	Sensitive field modification in the SDDP console	An employee maliciously modifies sensitive fields in the SDDP console. Data modification through applications is more risky than data modification in the SDDP console.	MaxCompute

9.Limits on using watermarks

This topic describes the limits when you embed watermarks into data sources.

What types of data sources can be embedded with watermarks?

You can embed watermarks into all types of data sources for which Data Security Center (DSC) supports static de-identification. For more information about the types of data sources for which DSC supports static de-identification, see [Supported data assets](#).

What requirements must the data sources meet before I can embed watermarks into the data sources?

The principle of embedding watermarks is to embed the watermark information into the columns that have different characteristics. The more characteristics of the source data, the more complete the watermark information that can be embedded, and the higher the success rate of watermark extraction. You can extract watermarks even if specific data is missing. When you embed watermarks into a data source, take note of the following items:

- Make sure that the number of rows in the data source is greater than or equal to 1,000.

If the data source contains fewer than 1,000 rows, you may fail to extract the watermarks due to insufficient characteristics.

- Embed watermarks into the columns that have a wide range of values. If you embed watermarks into a column that has only a few enumerated values, you may fail to extract the watermarks due to insufficient characteristics.

Typically, you can embed watermarks into columns whose name is address, name, UUID, amount, or total. Do not embed watermarks into columns whose name is gender or status.

- After you embed watermarks into a column, the values of the column may change. Therefore, before you embed watermarks into a column, make sure that the value changes of the column are acceptable.

For more information, see the "[Does the source data change after watermarks are embedded?](#)" section of this topic.

Which watermark embedding algorithm do I need to select?

If you embed watermarks into a column of the string type, select the space algorithm. If you embed watermarks into a column of the numeric type, select the least significant bit algorithm.

Does the source data change after watermarks are embedded?

After you configure a task in the DSC console to embed watermarks into the columns in Table A and write the data in Table A to Table B, the data in Table A remains unchanged. Only the data that is written to Table B changes.

Specific data embedded with watermarks is different from the source data. You can use the space algorithm and least significant bit algorithm to embed watermarks.