

维度表数据文件规范

请您按照下述规范准备数据文件，文件不规范将导致维度表创建失败或数据入库失败。

(一) 文件格式：目前仅支持 UTF-8 或 GBK 编码 的 CSV 格式文件。

(二) 文件大小：最大 10 M。

(三) 数据内容格式：

总体上须遵循通用的 CSV 格式规范。

此外，在易观方舟中使用 CSV 格式文件上传维度表数据时，需特别注意：

1. 如下图所示，文件中须包含至少三行内容，第一行为字段名，第二行为字段类型，第三行开始为字段值，每行应该拥有相同的列数和字段序列。

```
1 primary_key_1,primary_key_2,dimension_1,dimension_2,dimension_3,dimension_4,dimension_5
2 number,string,string,number,boolean,array,date
3 1,key1,易观方舟,12345,true,"["易观","方舟","千帆","Argo"]",20200101
4 2,key2,Argo,12345.12,false,"["易观","方舟","千帆","Argo"]",20200101 12:12:12
5 3,key3,易观千帆,12345,true,"["易观","方舟","千帆","Argo"]",20200101 12:12:12.999
6 4,key4,易观,12345,true,"["易观","方舟","千帆","Argo"]",
```

文本编辑器中的显示格式

	A	B	C	D	E	F	G
1	primary_key_1	primary_key_2	dimension_1	dimension_2	dimension_3	dimension_4	dimension_5
2	number	string	string	number	boolean	array	date
3	1	key1	易观方舟	12345	TRUE	["易观","方舟","千帆","Argo"]	20200101
4	2	key2	Argo	12345.12	FALSE	["易观","方舟","千帆","Argo"]	20200101 12:12:12
5	3	key3	易观千帆	12345	TRUE	["易观","方舟","千帆","Argo"]	20200101 12:12:12.999
6	4	key4	易观	12345	TRUE	["易观","方舟","千帆","Argo"]	

电子表格中的显示格式

2. 全部内容必须从第一行第一列开始顶格存放。

3. 第一行和第二行内容组合起来可以理解为文件的“表头”，表头中必须确保字段名和字段类型一一对应、成对出现，不能出现字段名或字段类型空缺的情况。

4. 字段名命名规范：

- 1) 支持大小写字母 a~z A~Z、数字 0~9、下划线_、符号\$ 组合命名，但请注意字段名仅允许以 \$ 或 字母 开头；

2) 最大长度 100 个字符。

5. 维度表目前支持五种数据类型作为字段类型，包括：`string`、`number`、`boolean`、`date`、`array`。

6. 字段值的数据格式须严格遵循相应数据类型的规范要求。详细说明见下表：

数据类型	字段类型	格式规范说明	示例
字符串	string	UTF-8编码后不超过2048字节	易观方舟
数值	number	支持整数和小数，小数点后最多保留3位 取值范围：-9E14到9E14	123 或者 123.123 (不支持使用科学记数法表达)
布尔	boolean	TRUE / FALSE，不区分大小写	TRUE
日期	date	格式：yyyyMMdd HH:mm:ss.SSS 或 yyyyMMdd HH:mm:ss 或 yyyyMMdd	20180618 18:18:18.188
集合	array	字符串集合，去重后最大元素个数100，其中每个元素使用UTF-8编码后最大长度255字节；元素输入可有重复，导入过程中会去重，最终形成去重的、无序的字符串集合	文本编辑器中的显示格式应为： ["元素1","元素2","元素3"]
		使用文本编辑器和使用电子表格（如Microsoft Excel）编辑时的显示格式不同，请注意区分	电子表格（如Excel）中的显示格式： ["元素1","元素2","元素3"]

【特别说明】

一般情况下，建议您优先使用纯文本编辑器软件，如 VS Code、Notepad++、EditPlus 来创建和编辑维度表数据文件。

虽然使用电子表格软件（如 Microsoft Excel、WPS 表格）也可以编辑 CSV 文件，但由于电子表格软件会自动对某些类型的数据进行格式转换，从而导致数据出错。

因此，如有需要使用电子表格软件来创建维度表数据文件时，请特别注意：

1) 在开始写入数据前，须先将全部单元格格式设置为“**文本**”。



2) 对于 集合类型 (array) 的数据:

在电子表格和文本编辑器中输入和显示的格式不同, 请注意区分

◆ 电子表格中的格式: ["元素 1","元素 2","元素 3"] (注: 如果在文本编辑器中按此格式写入数据, 用电子表格打开文件会导致数据出错)

◆ 文本编辑器中的格式应为: "["元素 1","元素 2","元素 3"]"

```
1 primary_key_1,primary_key_2,dimension_1,dimension_2,dimension_3,dimension_4,dimension_5
2 number,string,string,number,boolean,array,date
3 1,key1,易观方舟,12345,true,"["易观","方舟","千帆","Argo"]",20200101
4 2,key2,Argo,12345.12,false,"["易观","方舟","千帆","Argo"]",20200101 12:12:12
5 3,key3,易观千帆,12345,true,"["易观","方舟","千帆","Argo"]",20200101 12:12:12.999
6 4,key4,易观,12345,true,"["易观","方舟","千帆","Argo"]"
```

使用文本编辑器打开时的样式

	A	B	C	D	E	F	G
1	primary_key_1	primary_key_2	dimension_1	dimension_2	dimension_3	dimension_4	dimension_5
2	number	string	string	number	boolean	array	date
3	1	key1	易观方舟	12345	TRUE	["易观","方舟","千帆","Argo"]	20200101
4	2	key2	Argo	12345.12	FALSE	["易观","方舟","千帆","Argo"]	20200101 12:12:12
5	3	key3	易观千帆	12345	TRUE	["易观","方舟","千帆","Argo"]	20200101 12:12:12.999
6	4	key4	易观	12345	TRUE	["易观","方舟","千帆","Argo"]	

使用电子表格打开时的样式

3) 对于 数值类型 (number) 的数据:

◆ 请首先确保在使用“文本”类型“单元格格式”的前提下处理数值数据

◆ 为了避免文件在再次打开时被 Excel 强转格式, 须在数字前用半角英文单引号开头 (在文本编辑器中写入数据时不必带引号)

A2	fx	'1000
	A	B
1	单元格格式为文本类型且数字前加单引号	单元格格式为默认类型
2	1000	10000
3	1000000000000000000	1E+16
4	1000000000000000000	1E+17
5	确保数值型数据格式正确的方式	格式错误数据无法入库
6		

4) 对于 日期类型 (date) 的数据:

仅支持 yyMMdd HH:mm:ss.SSS 或 yyMMdd HH:mm:ss 或 yyMMdd 3 种格式

7. 一张维度表最多可存放 300 列 字段 (含关联字段)。

8. 文件中必须包含关联字段列, 关联字段的字段名和字段类型须与事件表中的字段名和字段类型完全一致。

9. 关联字段列必须全部有值, 不得出现关联字段值 (key 值) 为空的情况。

10. 也不得出现关联字段值（key 值）重复的情况，也就是说文件中的每行数据应具有唯一性。

【特别说明】

方舟通过关联字段的取值（key 值）来确认维度表中每条数据的唯一性，一条数据应具有唯一的 key 值，或者说每条数据都应具有唯一的 ID。

对于多关联字段的场景而言，多个字段取值组合起来构成 key 值，作为该条数据的唯一 ID。如下图所示，key_1、key_2、key_3 为关联字段，则第 3 行、第 4 行、第 5 行数据是 3 条完全不同的数据，其 key 值（唯一 ID）分别：001,argo,true、001,argo,false、001,argouba,true

1	key_1,key_2,key_3,dim_1,dim_2,dim_3
2	number,string,boolean,string,string,string
3	001,argo,true,a,b,c
4	001,argo,false,a,b,c
5	001,argouba,true,a,b,c