时间处理基础

Pandas提供了四种类型的生成日期时间的对象：日期时间、时间增量、时间跨度、日期偏移量

（1） 日期时间(Date Times)：具有时区支持的特定日期和时间。与Python标准库中的datetime.datetime类似。如2020年12月6日13点37分50秒；

（2） 时间增量(Time Deltas)：绝对持续时间，用于在指定时间点基础上增加指定的增量，如在某年月日的基础上增加2天、增加2个月、减少4小时等，最后产生一个新的时间点；

（3） 时间跨度(Time Span)：由时间点及其相关周期定义的时间跨度，如连续产生一年四个季度的时间序列；

（4） 日期偏移(Date Offsets)：以日历计算的相对持续时间，表示时间间隔，两个时间点之间的长度，如日、周、月、季度、年。

概念	标量类	数组类	Pandas数据类型	主要建立方法
日期时间(Date Times)	Timesstamp时间戳	DatetimeIndex时间索引	datetime64[ns]、datetime64[ns,tz]	to_datetime()、date_range()
时间增量(Time Deltas)	Timedelta时间增量	TimedeltaIndex时间增量索引	timedelta[ns]	to_timedelta()、timedelta_range()
时间跨度(Time Span)	Period时间周期	PeriodIndex周期索引	period[freq]	Period()、period_range()
日期偏移(Date Offsets)	DateOffset	None	None	DateOffset()

一般情况下，时间序列主要是 `Series` 或 `DataFrame` 的时间型索引，可以用时间元素进行操控。

```
import pandas as pd

import numpy as np

# python内置的模块
import time

import datetime
```

一、时间序列类型

1.时间戳

python datetime的替代品

时间戳相当于python的Datetime，在大多数情况下可以与之互换。

该类型用于组成DatetimeIndex的条目，以及pandas中其他面向时间序列的数据结构。

```
pandas.Timestamp(ts_input=<object object>, freq=None, tz=None, unit=None, year=None, month=None, day=None, hour=None, minute=None, second=None, microsecond=None, nanosecond=None, tzinfo=None, *, fold=None)
```

- ts_input:要转换为时间戳的值
- freq:时间戳将具有的偏移量
- unit:用于转换的单位

例子:

- 转换类似日期时间的字符串

```
pd.Timestamp('2022-01-01')
```

```
Timestamp('2022-01-01 00:00:00')
```

```
pd.Timestamp('2021-12-15 12')
```

```
Timestamp('2021-12-15 12:00:00')
```

```
pd.Timestamp('01-01-2022 12')
```

```
Timestamp('2022-01-01 12:00:00')
```

```
pd.Timestamp('2022-01')
```

```
Timestamp('2022-01-01 00:00:00')
```

```
pd.Timestamp('2022')
```

```
Timestamp('2022-01-01 00:00:00')
```

```
# 不能将错误日期转换为Timestamp ,
# ValueError: could not convert string to Timestamp
# pd.Timestamp('2022-01-50')
# pd.Timestamp('2022-02-31')
```

- unit参数为s:

  这将转换以秒为单位表示Unix历元的浮点值

  1970年1月1日这个时间正是Unix系统的起始时间

```
pd.Timestamp(time.time(), unit="s")
```

```
Timestamp('2022-04-21 02:29:27.827368736')
```

```
pd.Timestamp(int(time.time()), unit="s")
```

```
Timestamp('2022-04-21 02:29:43')
```

```
pd.Timestamp(time.time())
```

```
Timestamp('1970-01-01 00:00:01.650508212')
```

- year, month, day,hour, minute, second, microsecond单独设置时间

```
pd.Timestamp(2022, 1, 1, 12)
```

```
Timestamp('2022-01-01 00:00:00')
```

```
# 提供年月日
pd.Timestamp(year=2022, month=1,day=1)
```

```
pd.Timestamp(year=2022, month=1,day=1)
```

```
Timestamp('2022-01-01 00:00:00')
```

```
pd.Timestamp(year=2022, month=1, day=1, hour=12)
```

```
使用单独设置的形式,最小到日 ,否则报错:

            TypeError: an integer is required (got type NoneType)
```

总结:

Timestamp

- 将字符串或者unix时间转化为 pandas对应的时间戳

2.时间间隔--实现datetime加减

表示持续时间，即两个日期或时间之间的差异。

相当于python的datetime.timedelta，在大多数情况下可以与之互换

```
pd.Timedelta(
    value=<object object at 0x000001BE55DCFE80>,
    unit=None,
    **kwargs,
)
```

- value:数值 或者Timedelta

- unit：如果输入是整数，则表示输入的单位‘M’,'W', ‘D’, ‘T’, ‘S’,

```
ts = pd.Timestamp('2022-01-01 12')
ts
```

```
Timestamp('2022-01-01 12:00:00')
```

```
ts + pd.Timedelta(-1, "D")
```

```
Timestamp('2021-12-31 12:00:00')
```

```
#时间间隔
td = pd.Timedelta(days=5, minutes=50, seconds=20)
td
```

```
Timedelta('5 days 00:50:20')
```

```
ts + td
```

```
Timestamp('2022-01-06 12:50:20')
```

```
# 总秒数
td.total_seconds()
```

```
datetime.datetime.now()
```

```
datetime.datetime(2022, 4, 21, 10, 40, 32, 610955)
```

```
time.time()+ td.total_seconds()
pd.Timestamp(int(time.time()+ td.total_seconds()),unit='s')
```

```
Timestamp('2022-04-26 03:30:12')
```

```
import datetime
# 取得当前时间
now = datetime.datetime.now()
print(now)

#计算当前时间往后100天的日期
dt = now + pd.Timedelta(days=100)
#dt = now + datetime.timedelta(days=100)
print(dt,type(dt))
#只显示年月日
dt.strftime('%Y-%m-%d')
```

```
2022-04-21 10:41:44.686268
2022-07-30 10:41:44.686268 <class 'datetime.datetime'>
```

```
'2022-07-30'
```

```
dt = pd.Timestamp('2022-01-11')
#dt + pd.Timedelta(days=100)
dt + datetime.timedelta(days=100)
```

```
Timestamp('2022-04-21 00:00:00')
```

3.时间转化

`to_datetime` 转换时间戳

你可能会想到，我们经常要和文本数据（字符串）打交道，能否快速将文本数据转为时间戳呢？

```
    答案是可以的，通过 to_datetime 能快速将字符串转换为时间戳。当传递一个Series时，它会返回一个Series（具有相同的索引），而类似列表的则转换为DatetimeIndex。
```

```
to_datetime(arg, errors='raise', dayfirst=False, yearfirst=False, utc=None, format=None, unit=None, infer_datetime_format=False, origin='unix')
```

函数用户将数组、序列或dict的对象转换为datetime对象

- `arg` 要转换为日期时间的对象
- `errors` :错误处理

  If 'raise',将引发异常.

  If 'coerce', 无效的转换,使用NaT.

  If 'ignore', 无效的转换,将使用输入的数据.
- `dayfirst` :转换时指定日期分析顺序 `yearfirst`
- `utc` :控制与时区相关的解析、本地化和转换(忽略)
- `format` : 用于分析时间的strftime，例如“%d/%m/%Y”,自定义格式
- `unit` : D,s,ms 将时间戳转datetime
- `origin` : 定义参考日期。数值将被解析为自该参考日期起的单位数

1). 处理各种输入格式

```
从一个数据帧的多个列中组装日期时间。

这些键可以是常见的缩写，

如['year'、'month'、'day'、'minute'、'second'、'ms'、'us'、'ns']) ，

也可以是相同的复数形式
```

```
df = pd.DataFrame({'year': [2015, 2016],'month': [2, 3],'day': [4, 5]})
df
```

```
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

	year	month	day
0	2015	2	4
1	2016	3	5

```
pd.to_datetime(df)
```

```
0   2015-02-04
1   2016-03-05
dtype: datetime64[ns]
```

2). 将字符串转datetime

```
pd.to_datetime(['11-12-2021'])
```

```
DatetimeIndex(['2021-11-12'], dtype='datetime64[ns]', freq=None)
```

```
pd.to_datetime(["2005/11/23", "2010.12.31"])
```

```
DatetimeIndex(['2005-11-23', '2010-12-31'], dtype='datetime64[ns]',
freq=None)
```

3). 除了可以将文本数据转为时间戳外，还可以将 unix 时间转为时间戳。

```
pd.to_datetime([1349720105, 1349806505, 1349892905], unit="s")
```

```
DatetimeIndex(['2012-10-08 18:15:05', '2012-10-09 18:15:05',
               '2012-10-10 18:15:05'],
              dtype='datetime64[ns]', freq=None)
```

```
pd.to_datetime([1349720105100, 1349720105200, 1349720105300],
unit="ms")
```

```
DatetimeIndex(['2012-10-08 18:15:05.100000', '2012-10-08
18:15:05.200000',
               '2012-10-08 18:15:05.300000'],
              dtype='datetime64[ns]', freq=None)
```

4). 自动识别异常

- 210605

```
pd.to_datetime('210605')
```

```
Timestamp('2005-06-21 00:00:00')
```

```
pd.to_datetime('210605',yearfirst=True)
```

```
Timestamp('2021-06-05 00:00:00')
```

5). 配合unit参数,使用非unix时间

```
# origin参考起始时间
pd.to_datetime([1, 2, 3], unit='D', origin=pd.Timestamp('2020-01-11'))
```

```
DatetimeIndex(['2020-01-12', '2020-01-13', '2020-01-14'],
dtype='datetime64[ns]', freq=None)
```

```
pd.to_datetime([1, 2, 3], unit='d')
```

```
DatetimeIndex(['1970-01-02', '1970-01-03', '1970-01-04'],
dtype='datetime64[ns]', freq=None)
```

```
# origin参考起始时间
pd.to_datetime([1, 2, 3], unit='h', origin=pd.Timestamp('2020-01'))
```

```
DatetimeIndex(['2020-01-01 01:00:00', '2020-01-01 02:00:00',
               '2020-01-01 03:00:00'],
              dtype='datetime64[ns]', freq=None)
```

```
# origin参考起始时间 m--分钟
pd.to_datetime([1, 2, 3], unit='m', origin=pd.Timestamp('2020-01'))
```

```
DatetimeIndex(['2020-01-01 00:01:00', '2020-01-01 00:02:00',
               '2020-01-01 00:03:00'],
              dtype='datetime64[ns]', freq=None)
```

```
# origin参考起始时间 m--分钟
pd.to_datetime([1, 2, 3], unit='s', origin=pd.Timestamp('2020-01'))
```

```
DatetimeIndex(['2020-01-01 00:00:01', '2020-01-01 00:00:02',
               '2020-01-01 00:00:03'],
              dtype='datetime64[ns]', freq=None)
```

6). 不可转换日期/时间

```
    如果日期不符合时间戳限制，则传递errors='ignore'将返回原始输入，而不是引发任何异常。

    除了将非日期（或不可解析的日期）强制传递给NaT之外，传递errors='coerce'还会将越界日期强制传递给NaT
```

- errors`:错误处理

  If 'raise',将引发异常.

  If 'coerce', 无效的转换,使用NaT.

  If 'ignore', 无效的转换,将使用输入的数据.

```
#ParserError: year 120211204 is out of range: 120211204
pd.to_datetime(['120211204','20210101'])
```

```
# 无效的转换,将使用输入的数据
pd.to_datetime(['120211204','2021.02.01'], errors="ignore")
```

```
Index(['120211204', '2021.02.01'], dtype='object')
```

```
# 无效的转换,使用NaT
pd.to_datetime(['120211204','2021.02.01'], errors="coerce")
```

```
DatetimeIndex(['NaT', '2021-02-01'], dtype='datetime64[ns]',
freq=None)
```

```
# 自动识别
pd.to_datetime(pd.Series(["Jul 31, 2018", "2018.05.10", None]))
```

```
0   2018-07-31
1   2018-05-10
2          NaT
dtype: datetime64[ns]
```

```
1
```

> 对于float arg，可能会发生精确舍入。要防止意外行为，请使用固定宽度的精确类型。

4.生成时间戳范围

有时候，我们可能想要生成某个范围内的时间戳。例如，我想要生成 "2018-6-26" 这一天之后的8天时间戳

我们可以使用 `date_range` 和 `bdate_range` 来完成时间戳范围的生成。

date_range() 返回固定频率的DatetimeIndex

`date_range(start=None, end=None, periods=None, freq=None, tz=None, normalize=False, name=None, closed=None, **kwargs)`

返回等距时间点的范围（其中任意两个相邻点之间的差值由给定频率指定），以便它们都满足开始<[=]x<[=]end，其中第一个和最后一个分别为。，该范围内的第一个和最后一个时间点位于freq的边界（如果以频率字符串的形式给出）或对freq有效

- start :生成日期的左边界
- end : 生成日期的左边界
- periods: 要生成的周期数
- freq : 频率, default 'D' ,频率字符串可以有倍数, '5H'
- tz: 时区用于返回本地化日期时间索引的时区名称，例如“Asia/Hong_Kong”。默认情况下，生成的DatetimeIndex是时区初始索引。
- normalize: 默认 False, 在生成日期范围之前，将开始/结束日期标准化
- name: 默认 None 设置返回DatetimeIndex name

1). 指定值

- 默认是包含开始和结束时间, 默认频率使用的D(天)

```
1 pd.date_range(start='1/1/2021', end='1/08/2021')
```

```
DatetimeIndex(['2021-01-01', '2021-01-02', '2021-01-03', '2021-01-04',
               '2021-01-05', '2021-01-06', '2021-01-07', '2021-01-
08'],
              dtype='datetime64[ns]', freq='D')
```

```
pd.date_range(start='2010', end='2011')
```

```
DatetimeIndex(['2010-01-01', '2010-01-02', '2010-01-03', '2010-01-04',
               '2010-01-05', '2010-01-06', '2010-01-07', '2010-01-08',
               '2010-01-09', '2010-01-10',
               ...
               '2010-12-23', '2010-12-24', '2010-12-25', '2010-12-26',
               '2010-12-27', '2010-12-28', '2010-12-29', '2010-12-30',
               '2010-12-31', '2011-01-01'],
              dtype='datetime64[ns]', length=366, freq='D')
```

2). 指定开始日期,设置期间数

```
pd.date_range(start='1/1/2018', periods=8)
```

```
DatetimeIndex(['2018-01-01', '2018-01-02', '2018-01-03', '2018-01-04',
               '2018-01-05', '2018-01-06', '2018-01-07', '2018-01-
08'],
              dtype='datetime64[ns]', freq='D')
```

- 指定开始、结束和期间；频率自动生成（线性间隔）

```
pd.date_range(start='2018-04-24', end='2018-04-27', periods=3)
```

```
DatetimeIndex(['2018-04-24 00:00:00', '2018-04-25 12:00:00',
               '2018-04-27 00:00:00'],
              dtype='datetime64[ns]', freq=None)
```

```
pd.date_range(start='2018-04-24', end='2018-04-27', periods=4)

```

```
DatetimeIndex(['2018-04-24', '2018-04-25', '2018-04-26', '2018-04-
27'], dtype='datetime64[ns]', freq=None)
```

```
pd.date_range(start='2018-04-24', periods=4)
```

```
DatetimeIndex(['2018-04-24', '2018-04-25', '2018-04-26', '2018-04-
27'], dtype='datetime64[ns]', freq='D')
```

3). 频率 freq

freq	描述
Y	年
M	月
D	日(默认)
T(MIN)	分钟
S	秒
L	毫秒
U	微秒
A-DEC	每年指定月份的最后一个日历日
W-MON	指定每月的哪个星期开始
WOM_2MON	指定每个月的第几个星期(这里是第二个星期)
Q-DEC(Q-月)	指定月为季度末，每个季度末最后一月的最后一个日历日
B,(M,Q,A),S	分别代表了工作日，（以月为频率，以季度为频率，以年为频率），最接近月初的那一天
B	工作日

- 月缩写：JAN/FEB/MAR/APR/MAY/JUN/JUL/AUG/SEP/OCT/NOV/DEC
- 星期几缩写：MON/TUE/WED/THU/FRI/SAT/SUN
- 所以Q-月只有三种情况：1-4-7-10,2-5-8-11,3-6-9-12

```
pd.date_range('2022/1/1','2022/2/1', freq = 'W-MON')
# W-MON：从指定星期几开始算起，每周
# 星期几缩写：MON/TUE/WED/THU/FRI/SAT/SUN
```

```
DatetimeIndex(['2022-01-03', '2022-01-10', '2022-01-17', '2022-01-24',
               '2022-01-31'],
              dtype='datetime64[ns]', freq='W-MON')
```

```
pd.date_range('2022/1/1','2022/5/1', freq = 'WOM-2MON')
# WOM-2MON：每月的第几个星期几开始算，这里是每月第二个星期一
```

```
DatetimeIndex(['2022-01-10', '2022-02-14', '2022-03-14', '2022-04-
11'], dtype='datetime64[ns]', freq='WOM-2MON')
```

```
# 默认freq = 'D'：每日历日
pd.date_range('2022/1/1','2022/1/4')
```

```
DatetimeIndex(['2022-01-01', '2022-01-02', '2022-01-03', '2022-01-
04'], dtype='datetime64[ns]', freq='D')
```

```
pd.date_range('2022/1/1','2022/1/5', freq = 'B')  # B：每工作日
```

```
DatetimeIndex(['2022-01-03', '2022-01-04', '2022-01-05'],
dtype='datetime64[ns]', freq='B')
```

```
pd.date_range('2022/1/1','2022/1/2', freq = 'H')  # H：每小时
```

```
DatetimeIndex(['2022-01-01 00:00:00', '2022-01-01 01:00:00',
               '2022-01-01 02:00:00', '2022-01-01 03:00:00',
               '2022-01-01 04:00:00', '2022-01-01 05:00:00',
               '2022-01-01 06:00:00', '2022-01-01 07:00:00',
               '2022-01-01 08:00:00', '2022-01-01 09:00:00',
               '2022-01-01 10:00:00', '2022-01-01 11:00:00',
               '2022-01-01 12:00:00', '2022-01-01 13:00:00',
               '2022-01-01 14:00:00', '2022-01-01 15:00:00',
               '2022-01-01 16:00:00', '2022-01-01 17:00:00',
               '2022-01-01 18:00:00', '2022-01-01 19:00:00',
               '2022-01-01 20:00:00', '2022-01-01 21:00:00',
               '2022-01-01 22:00:00', '2022-01-01 23:00:00',
               '2022-01-02 00:00:00'],
              dtype='datetime64[ns]', freq='H')
```

```
pd.date_range('2022/1/1 12:00','2022/1/1 12:10', freq = 'T')  # T/MIN：每分
```

```
DatetimeIndex(['2022-01-01 12:00:00', '2022-01-01 12:01:00',
               '2022-01-01 12:02:00', '2022-01-01 12:03:00',
               '2022-01-01 12:04:00', '2022-01-01 12:05:00',
               '2022-01-01 12:06:00', '2022-01-01 12:07:00',
               '2022-01-01 12:08:00', '2022-01-01 12:09:00',
               '2022-01-01 12:10:00'],
              dtype='datetime64[ns]', freq='T')
```

```
pd.date_range('2022/1/1 12:00:00','2022/1/1 12:00:10', freq = 'S')  # S：每秒
```

```
DatetimeIndex(['2022-01-01 12:00:00', '2022-01-01 12:00:01',
               '2022-01-01 12:00:02', '2022-01-01 12:00:03',
               '2022-01-01 12:00:04', '2022-01-01 12:00:05',
               '2022-01-01 12:00:06', '2022-01-01 12:00:07',
               '2022-01-01 12:00:08', '2022-01-01 12:00:09',
               '2022-01-01 12:00:10'],
              dtype='datetime64[ns]', freq='S')
```

```
pd.date_range('2022/1/1 12:00:00','2022/1/1 12:00:10', freq = 'L')  # L：每毫秒（千分之一秒）
```

```
pd.date_range('2022/1/1 12:00:00','2022/1/1 12:00:10', freq = 'U')  # U：每微秒（百万分之一秒）
```

```
DatetimeIndex([       '2022-01-01 12:00:00', '2022-01-01
12:00:00.000001',
               '2022-01-01 12:00:00.000002', '2022-01-01
12:00:00.000003',
               '2022-01-01 12:00:00.000004', '2022-01-01
12:00:00.000005',
               '2022-01-01 12:00:00.000006', '2022-01-01
12:00:00.000007',
               '2022-01-01 12:00:00.000008', '2022-01-01
12:00:00.000009',
               ...
               '2022-01-01 12:00:09.999991', '2022-01-01
12:00:09.999992',
               '2022-01-01 12:00:09.999993', '2022-01-01
12:00:09.999994',
               '2022-01-01 12:00:09.999995', '2022-01-01
12:00:09.999996',
               '2022-01-01 12:00:09.999997', '2022-01-01
12:00:09.999998',
               '2022-01-01 12:00:09.999999',        '2022-01-01
12:00:10'],
              dtype='datetime64[ns]', length=10000001, freq='U')
```

```
# M：每月最后一个日历日
pd.date_range('2017','2018', freq = 'M')
```

```
DatetimeIndex(['2017-01-31', '2017-02-28', '2017-03-31', '2017-04-30',
               '2017-05-31', '2017-06-30', '2017-07-31', '2017-08-31',
               '2017-09-30', '2017-10-31', '2017-11-30', '2017-12-
31'],
              dtype='datetime64[ns]', freq='M')
```

```
# Q-月：指定月为季度末，每个季度末最后一月的最后一个日历日
print(pd.date_range('2017','2020', freq = 'Q-DEC'))
```

```
DatetimeIndex(['2017-03-31', '2017-06-30', '2017-09-30', '2017-12-31',
               '2018-03-31', '2018-06-30', '2018-09-30', '2018-12-31',
               '2019-03-31', '2019-06-30', '2019-09-30', '2019-12-
31'],
              dtype='datetime64[ns]', freq='Q-DEC')
```

```
# A-月：每年指定月份的最后一个日历日
print(pd.date_range('2017','2020', freq = 'A-DEC'))
```

```
DatetimeIndex(['2017-12-31', '2018-12-31', '2019-12-31'],
dtype='datetime64[ns]', freq='A-DEC')
```

```
# B,(M,Q,A),S   分别代表了工作日，（以月为频率，以季度为频率，以年为频率），最接近月初的那一天
pd.date_range('2017','2018', freq = 'BM')
# BM：每月最后一个工作日
# BQ-月：指定月为季度末，每个季度末最后一月的最后一个工作日
# BA-月：每年指定月份的最后一个工作日
```

```
pd.date_range('2017','2020', freq = 'BQ-DEC')
```

```
pd.date_range('2017','2020', freq = 'BA-DEC')
```

```
# pd.date_range()-日期范围：复合频率

print(pd.date_range('2017/1/1','2017/2/1', freq = '7D'))  # 7天
print(pd.date_range('2017/1/1','2017/1/2', freq = '2h30min'))  # 2小时30分钟
print(pd.date_range('2017','2018', freq = '2M'))  # 2月，每月最后一个日历日
```

```
DatetimeIndex(['2017-01-01', '2017-01-08', '2017-01-15', '2017-01-22',
               '2017-01-29'],
              dtype='datetime64[ns]', freq='7D')
DatetimeIndex(['2017-01-01 00:00:00', '2017-01-01 02:30:00',
               '2017-01-01 05:00:00', '2017-01-01 07:30:00',
               '2017-01-01 10:00:00', '2017-01-01 12:30:00',
               '2017-01-01 15:00:00', '2017-01-01 17:30:00',
               '2017-01-01 20:00:00', '2017-01-01 22:30:00'],
              dtype='datetime64[ns]', freq='150T')
DatetimeIndex(['2017-01-31', '2017-03-31', '2017-05-31', '2017-07-31',
               '2017-09-30', '2017-11-30'],
              dtype='datetime64[ns]', freq='2M')
```

```
pd.date_range(start='1/1/2018', periods=5, freq='3M')
```

4) normalize: 在生成日期范围之前，将开始/结束日期标准化

```
pd.date_range(start = '1/1/2021 15:30', periods =10)
```

```
DatetimeIndex(['2021-01-01 15:30:00', '2021-01-02 15:30:00',
               '2021-01-03 15:30:00', '2021-01-04 15:30:00',
               '2021-01-05 15:30:00', '2021-01-06 15:30:00',
               '2021-01-07 15:30:00', '2021-01-08 15:30:00',
               '2021-01-09 15:30:00', '2021-01-10 15:30:00'],
              dtype='datetime64[ns]', freq='D')
```

```
pd.date_range(start = '1/1/2021 15:30', periods =10, name = 'mypd')
```

```
DatetimeIndex(['2021-01-01 15:30:00', '2021-01-02 15:30:00',
               '2021-01-03 15:30:00', '2021-01-04 15:30:00',
               '2021-01-05 15:30:00', '2021-01-06 15:30:00',
               '2021-01-07 15:30:00', '2021-01-08 15:30:00',
               '2021-01-09 15:30:00', '2021-01-10 15:30:00'],
              dtype='datetime64[ns]', name='mypd', freq='D')
```

```
pd.date_range(start = '1/1/2021 15:30', periods =10, name =
'mypd',normalize=True)
```

```
DatetimeIndex(['2021-01-01', '2021-01-02', '2021-01-03', '2021-01-04',
               '2021-01-05', '2021-01-06', '2021-01-07', '2021-01-08',
               '2021-01-09', '2021-01-10'],
              dtype='datetime64[ns]', name='mypd', freq='D')
```

```
pd.date_range(start = '1/1/2021 15:30', periods=10, name='mypd',
freq="H")
```

```
DatetimeIndex(['2021-01-01 15:30:00', '2021-01-01 16:30:00',
               '2021-01-01 17:30:00', '2021-01-01 18:30:00',
               '2021-01-01 19:30:00', '2021-01-01 20:30:00',
               '2021-01-01 21:30:00', '2021-01-01 22:30:00',
               '2021-01-01 23:30:00', '2021-01-02 00:30:00'],
              dtype='datetime64[ns]', name='mypd', freq='H')
```

```
pd.date_range(start = '1/1/2021 15:30', periods=10,
name='mypd',normalize=True, freq="H")
```

```
DatetimeIndex(['2021-01-01 00:00:00', '2021-01-01 01:00:00',
               '2021-01-01 02:00:00', '2021-01-01 03:00:00',
               '2021-01-01 04:00:00', '2021-01-01 05:00:00',
               '2021-01-01 06:00:00', '2021-01-01 07:00:00',
               '2021-01-01 08:00:00', '2021-01-01 09:00:00'],
              dtype='datetime64[ns]', name='mypd', freq='H')
```

```
#closed  包含left  还是right  ,默认start和end都包含
pd.date_range(start='1/1/2021', end='1/08/2021')
```

```
DatetimeIndex(['2021-01-01', '2021-01-02', '2021-01-03', '2021-01-04',
               '2021-01-05', '2021-01-06', '2021-01-07', '2021-01-
08'],
              dtype='datetime64[ns]', freq='D')
```

```
pd.date_range(start='1/1/2021', end='1/08/2021',closed='left')
```

```
DatetimeIndex(['2021-01-02', '2021-01-03', '2021-01-04', '2021-01-05',
               '2021-01-06', '2021-01-07', '2021-01-08'],
              dtype='datetime64[ns]', freq='D')
```

bdate_range(start=None, end=None, periods=None, freq='B')

返回固定频率的DatetimeIndex，默认频率为工作日

```
pd.bdate_range(start='2022-04-01', end='2022-05-01')
```

```
DatetimeIndex(['2022-04-01', '2022-04-04', '2022-04-05', '2022-04-06',
               '2022-04-07', '2022-04-08', '2022-04-11', '2022-04-12',
               '2022-04-13', '2022-04-14', '2022-04-15', '2022-04-18',
               '2022-04-19', '2022-04-20', '2022-04-21', '2022-04-22',
               '2022-04-25', '2022-04-26', '2022-04-27', '2022-04-28',
               '2022-04-29'],
              dtype='datetime64[ns]', freq='B')
```

```

```

```
pd.date_range(start = '1/1/2021', end='1/2/2021', name='mypd',
freq="H",closed='left')
```

```
DatetimeIndex(['2021-01-01 00:00:00', '2021-01-01 01:00:00',
               '2021-01-01 02:00:00', '2021-01-01 03:00:00',
               '2021-01-01 04:00:00', '2021-01-01 05:00:00',
               '2021-01-01 06:00:00', '2021-01-01 07:00:00',
               '2021-01-01 08:00:00', '2021-01-01 09:00:00',
               '2021-01-01 10:00:00', '2021-01-01 11:00:00',
               '2021-01-01 12:00:00', '2021-01-01 13:00:00',
               '2021-01-01 14:00:00', '2021-01-01 15:00:00',
               '2021-01-01 16:00:00', '2021-01-01 17:00:00',
               '2021-01-01 18:00:00', '2021-01-01 19:00:00',
               '2021-01-01 20:00:00', '2021-01-01 21:00:00',
               '2021-01-01 22:00:00', '2021-01-01 23:00:00'],
              dtype='datetime64[ns]', name='mypd', freq='H')
```

5.时期频率转换

```
asfreq(freq, method=None, how=None, normalize=False,
fill_value=None)
```

将时间序列转换为指定的频率

- 如果此数据帧的索引是PeriodIndex，则新索引是使用PeriodIndex转换原始索引的结果
- 否则，新指数将相当于pd。date_range（start，end，freq=freq），其中start和end分别是原始索引中的第一个和最后一个条目
- 与新索引中未出现在原始索引中的任何时间步对应的值将为空（NaN），除非提供了填充此类未知值的方法

```
# asfreq：时期频率转换

ts = pd.Series(np.random.rand(4),
               index = pd.date_range('20170101','20170104'))
print(ts)

# 改变频率，这里是D改为4H
# method：插值模式，None不插值，ffill用之前值填充，bfill用之后值填充
```

```
2017-01-01    0.356101
2017-01-02    0.740144
2017-01-03    0.900764
2017-01-04    0.880999
Freq: D, dtype: float64
```

```
ts.asfreq('4H')
```

```
2017-01-01 00:00:00    0.356101
2017-01-01 04:00:00         NaN
2017-01-01 08:00:00         NaN
2017-01-01 12:00:00         NaN
2017-01-01 16:00:00         NaN
2017-01-01 20:00:00         NaN
2017-01-02 00:00:00    0.740144
2017-01-02 04:00:00         NaN
2017-01-02 08:00:00         NaN
2017-01-02 12:00:00         NaN
2017-01-02 16:00:00         NaN
2017-01-02 20:00:00         NaN
2017-01-03 00:00:00    0.900764
2017-01-03 04:00:00         NaN
2017-01-03 08:00:00         NaN
2017-01-03 12:00:00         NaN
2017-01-03 16:00:00         NaN
```

```
2017-01-03 20:00:00         NaN
2017-01-04 00:00:00    0.880999
Freq: 4H, dtype: float64
```

```
ts.asfreq('4H', method = 'ffill')
```

```
2017-01-01 00:00:00    0.356101
2017-01-01 04:00:00    0.356101
2017-01-01 08:00:00    0.356101
2017-01-01 12:00:00    0.356101
2017-01-01 16:00:00    0.356101
2017-01-01 20:00:00    0.356101
2017-01-02 00:00:00    0.740144
2017-01-02 04:00:00    0.740144
2017-01-02 08:00:00    0.740144
2017-01-02 12:00:00    0.740144
2017-01-02 16:00:00    0.740144
2017-01-02 20:00:00    0.740144
2017-01-03 00:00:00    0.900764
2017-01-03 04:00:00    0.900764
2017-01-03 08:00:00    0.900764
2017-01-03 12:00:00    0.900764
2017-01-03 16:00:00    0.900764
2017-01-03 20:00:00    0.900764
2017-01-04 00:00:00    0.880999
Freq: 4H, dtype: float64
```

```
ts.asfreq('4H', method = 'bfill')
```

```
2017-01-01 00:00:00    0.356101
2017-01-01 04:00:00    0.740144
2017-01-01 08:00:00    0.740144
2017-01-01 12:00:00    0.740144
2017-01-01 16:00:00    0.740144
2017-01-01 20:00:00    0.740144
2017-01-02 00:00:00    0.740144
2017-01-02 04:00:00    0.900764
2017-01-02 08:00:00    0.900764
2017-01-02 12:00:00    0.900764
```

```
2017-01-02 16:00:00    0.900764
2017-01-02 20:00:00    0.900764
2017-01-03 00:00:00    0.900764
2017-01-03 04:00:00    0.880999
2017-01-03 08:00:00    0.880999
2017-01-03 12:00:00    0.880999
2017-01-03 16:00:00    0.880999
2017-01-03 20:00:00    0.880999
2017-01-04 00:00:00    0.880999
Freq: 4H, dtype: float64
```

6. shift()-时间频率进行移位

`shift(periods=1, freq=None, axis=0, fill_value=None)`
按所需的时段数和可选的时间频率进行移位索引。

如果未传递freq，则在不重新调整数据的情况下移动索引。如果传递了freq（在这种情况下，索引必须是date或datetime，否则将引发NotImplementedError），只要在索引中设置了freq或推断的_freq属性，就可以推断freq

- periods: 要转换的时段数。可以是正面的，也可以是负面的
- freq:如果指定了freq，则索引值会移位，但数据不会重新对齐。也就是说，如果要在移动时扩展索引并保留原始数据
- axis:{0 or 'index', 1 or 'columns', None} 转换方向.
- fill_value: 填充值

```
df = pd.DataFrame(np.random.rand(16).reshape((4,4)),
                  index = pd.date_range('20210101','20210104'),
                  columns=list('ABCD'))
df
```

```
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

	A	B	C	D
2021-01-01	0.973223	0.995945	0.151857	0.765527
2021-01-02	0.273757	0.964698	0.179043	0.181121
2021-01-03	0.058434	0.077550	0.366556	0.744166
2021-01-04	0.534782	0.004702	0.300140	0.495053

```
# 正数：数值后移（滞后）,模式为行
df.shift(periods=2)
```

```
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

	A	B	C	D
2021-01-01	NaN	NaN	NaN	NaN
2021-01-02	NaN	NaN	NaN	NaN
2021-01-03	0.973223	0.995945	0.151857	0.765527
2021-01-04	0.273757	0.964698	0.179043	0.181121

```
# 正数：数值后移（滞后） 设置为列
df.shift(periods=1, axis="columns")
```

```
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

	A	B	C	D
2021-01-01	NaN	0.973223	0.995945	0.151857
2021-01-02	NaN	0.273757	0.964698	0.179043
2021-01-03	NaN	0.058434	0.077550	0.366556
2021-01-04	NaN	0.534782	0.004702	0.300140

```
# 正数: 数值后移, NaN填充为0
df.shift(periods=3, fill_value=0)
```

```
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

	A	B	C	D
2021-01-01	0.000000	0.000000	0.000000	0.000000
2021-01-02	0.000000	0.000000	0.000000	0.000000
2021-01-03	0.000000	0.000000	0.000000	0.000000
2021-01-04	0.973223	0.995945	0.151857	0.765527

```
# 当设置freq时 ,对时间索引移动
df.shift(periods=3, freq="D")
```

```
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

	A	B	C	D
2021-01-04	0.973223	0.995945	0.151857	0.765527
2021-01-05	0.273757	0.964698	0.179043	0.181121
2021-01-06	0.058434	0.077550	0.366556	0.744166
2021-01-07	0.534782	0.004702	0.300140	0.495053

```
df.shift(1)
```

```
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

	A	B	C	D
2021-01-01	NaN	NaN	NaN	NaN
2021-01-02	0.973223	0.995945	0.151857	0.765527
2021-01-03	0.273757	0.964698	0.179043	0.181121
2021-01-04	0.058434	0.077550	0.366556	0.744166

```
# 计算变化百分比，这里计算：该时间戳与上一个时间戳相比，变化百分比
per = df/df.shift(1) - 1
print(per)
```

```
                   A         B         C         D
2021-01-01       NaN       NaN       NaN       NaN
2021-01-02 -0.718711 -0.031374  0.179025 -0.763403
2021-01-03 -0.786548 -0.919612  1.047304  3.108655
2021-01-04  8.151918 -0.939363 -0.181190 -0.334754
```

```

```

7. Pandas时期： Period()

```
p = pd.Period('2017')
p
```

```
Period('2017', 'A-DEC')
```

```
p = pd.Period('2017-1', freq = 'M')
print(p, type(p))
```

```
2017-01 <class 'pandas._libs.tslibs.period.Period'>
```

- 通过加减整数可以实现对Period的移动

```
print(p + 1)
print(p - 2)
```

```
2017-02
2016-11
```

- 如果两个Period对象拥有相同频率，则它们的差就是它们之间的单位数量

```
p = pd.Period('2017-1', freq = 'M')
print(p, type(p))
```

```
pd.Period('2018', freq='M') - p
```

```
<12 * MonthEnds>
```

period_range函数可用于创建规则的时期范围

```
rng = pd.period_range('2021-1-1', '2021-6-1')
rng
```

```
PeriodIndex(['2021-01-01', '2021-01-02', '2021-01-03', '2021-01-04',
             '2021-01-05', '2021-01-06', '2021-01-07', '2021-01-08',
             '2021-01-09', '2021-01-10',
             ...
             '2021-05-23', '2021-05-24', '2021-05-25', '2021-05-26',
             '2021-05-27', '2021-05-28', '2021-05-29', '2021-05-30',
             '2021-05-31', '2021-06-01'],
            dtype='period[D]', length=152, freq='D')
```

```
pd.Series(np.random.rand(6), index=rng)
```

PeriodIndex类的构造函数允许直接使用一组字符串表示一段时期

```
values = ['200103', '200104', '200105']
# 必须指定 freq
index = pd.PeriodIndex(values, freq='M')
index
```

```
PeriodIndex(['2001-03', '2001-04', '2001-05'], dtype='period[M]',
freq='M')
```

时期的频率转换 asfreq

```
# A-月：每年指定月份的最后一个日历日
p = pd.Period('2021', freq='A-DEC')
p
```

```
Period('2021', 'A-DEC')
```

```
p.asfreq('M')
```

```
Period('2021-12', 'M')
```

```
p.asfreq('M', how="start") #  # 也可写 how = 's'
```

```
Period('2021-01', 'M')
```

```
p.asfreq('M', how="end") # 也可写 how = 'e'
```

```
Period('2021-12', 'M')
```

```
p.asfreq('H')
```

对于PeriodIndex或TimeSeries的频率转换方式相同

```
rng = pd.period_range('2006', '2009', freq='A-DEC')
rng
```

```
PeriodIndex(['2006', '2007', '2008', '2009'], dtype='period[A-DEC]',
freq='A-DEC')
```

```
ts = pd.Series(np.random.rand(len(rng)), rng)
ts
```

```
2006    0.094354
2007    0.675481
2008    0.001803
2009    0.434551
Freq: A-DEC, dtype: float64
```

```
ts.asfreq('M', how='s')
```

```
2006-01    0.094354
2007-01    0.675481
2008-01    0.001803
2009-01    0.434551
Freq: M, dtype: float64
```

```
ts.asfreq('M')
```

```
2006-12    0.094354
2007-12    0.675481
2008-12    0.001803
2009-12    0.434551
Freq: M, dtype: float64
```

```
ts2 = pd.Series(np.random.rand(len(rng)), index = rng.asfreq('D', how = 'start'))
ts2
```

```
2006-01-01    0.568333
2007-01-01    0.509033
2008-01-01    0.618067
2009-01-01    0.772937
Freq: D, dtype: float64
```

时间戳与时期之间的转换：pd.to_period()、pd.to_timestamp()

```
rng = pd.date_range('2017/1/1', periods = 10, freq = 'M')
prng = pd.period_range('2017','2018', freq = 'M')
print(rng)
print(prng)
```

```
DatetimeIndex(['2017-01-31', '2017-02-28', '2017-03-31', '2017-04-30',
               '2017-05-31', '2017-06-30', '2017-07-31', '2017-08-31',
               '2017-09-30', '2017-10-31'],
              dtype='datetime64[ns]', freq='M')
PeriodIndex(['2017-01', '2017-02', '2017-03', '2017-04', '2017-05',
'2017-06',
             '2017-07', '2017-08', '2017-09', '2017-10', '2017-11',
'2017-12',
             '2018-01'],
            dtype='period[M]', freq='M')
```

```
ts1 = pd.Series(np.random.rand(len(rng)), index = rng)
ts1
```

```
2017-01-31    0.596303
2017-02-28    0.817321
2017-03-31    0.347044
2017-04-30    0.906550
2017-05-31    0.630162
2017-06-30    0.754577
2017-07-31    0.742694
2017-08-31    0.790809
2017-09-30    0.812810
2017-10-31    0.749713
Freq: M, dtype: float64
```

```
# 每月最后一日，转化为每月 t
# o_period()参数为空， 推断每日频率
ts1.to_period().head()
```

```
2017-01    0.596303
2017-02    0.817321
2017-03    0.347044
2017-04    0.906550
2017-05    0.630162
Freq: M, dtype: float64
```

```
ts1.to_period("M")
```

```
2017-01    0.596303
2017-02    0.817321
2017-03    0.347044
2017-04    0.906550
2017-05    0.630162
2017-06    0.754577
2017-07    0.742694
2017-08    0.790809
2017-09    0.812810
2017-10    0.749713
Freq: M, dtype: float64
```

```
ts2 = pd.Series(np.random.rand(len(prng)), index = prng)
print(ts2.head())

```

```
2017-01    0.137227
2017-02    0.730822
2017-03    0.828779
2017-04    0.330024
2017-05    0.191158
Freq: M, dtype: float64
```

```
print(ts2.to_timestamp().head())
```

```
2017-01-01    0.137227
2017-02-01    0.730822
2017-03-01    0.828779
2017-04-01    0.330024
2017-05-01    0.191158
Freq: MS, dtype: float64
```

时间序列 - 重采样resample

Pandas中的resample，重新采样，是对原样本重新处理的一个方法，是一个对常规时间序列数据重新采样和频率转换的便捷的方法。

重新取样时间序列数据。

方便的时间序列的频率转换和重采样方法。

对象必须具有类似datetime的索引(DatetimeIndex、PeriodIndex或TimedeltaIndex)，或将类似datetime的值传递给on或level关键字。

`DataFrame.resample(rule, closed=None, label=None, level=None)`

- rule: 表示目标转换的偏移量字符串或对象
- closed:在降采样时，各时间段的哪一段是闭合的，‘right’或‘left’，默认‘right’
- label:在降采样时，如何设置聚合值的标签，例如，9：30-9：35会被标记成9：30还是9：35,默认9：35

```
rng = pd.date_range('20170101', periods = 12)
ts = pd.Series(np.arange(12), index = rng)
print(ts)
```

```
2017-01-01     0
2017-01-02     1
2017-01-03     2
2017-01-04     3
2017-01-05     4
2017-01-06     5
2017-01-07     6
2017-01-08     7
2017-01-09     8
2017-01-10     9
2017-01-11    10
2017-01-12    11
Freq: D, dtype: int32
```

```
# 将序列下采样到5天的数据箱中，并将放入数据箱的时间戳的值相加
ts.resample('5D').sum()
```

```
#  得到一个新的聚合后的Series，聚合方式为求和
print(ts.resample('5D').mean(),'→ 求平均值\n')
print(ts.resample('5D').max(),'→ 求最大值\n')
print(ts.resample('5D').min(),'→ 求最小值\n')
print(ts.resample('5D').median(),'→ 求中值\n')
print(ts.resample('5D').first(),'→ 返回第一个值\n')
print(ts.resample('5D').last(),'→ 返回最后一个值\n')
print(ts.resample('5D').ohlc(),'→ OHLC重采样\n')
# OHLC:金融领域的时间序列聚合方式 → open开盘、high最大值、low最小值、close收盘
```

```
# 降采样
rng = pd.date_range('20170101', periods = 12)
ts = pd.Series(np.arange(1,13), index = rng)
print(ts)
```

```
ts.resample('5D').sum()  #'→ 默认\n'
```

closed：各时间段哪一端是闭合（即包含）的，默认 左闭右闭

- 详解：这里values为0-11，按照5D重采样 → [1,2,3,4,5],[6,7,8,9,10],[11,12]
- left指定间隔左边为结束 → [1,2,3,4,5],[6,7,8,9,10],[11,12]
- right指定间隔右边为结束 → [1],[2,3,4,5,6],[7,8,9,10,11],[12]

```
# 将系列降采样到5天的箱中，但关闭箱间隔的左侧
print(ts.resample('5D', closed = 'left').sum(), '→ left\n')
```

```
# 将系列降采样到5天的箱中，但关闭箱间隔的右侧
print(ts.resample('5D', closed = 'right').sum(), '→ right\n')
```

```
print(ts.resample('5D', label = 'left').sum(),'→ leftlabel\n')
# label：聚合值的index，默认为取左
# 值采样认为默认（这里closed默认）
```

```
print(ts.resample('5D', label = 'right').sum(),'→ rightlabel\n')
```

```
# 升采样及插值

rng = pd.date_range('2017/1/1 0:0:0', periods = 5, freq = 'H')
ts = pd.DataFrame(np.arange(15).reshape(5,3),
                  index = rng,
                  columns = ['a','b','c'])
print(ts)

print(ts.resample('15T').asfreq())
print(ts.resample('15T').ffill())
print(ts.resample('15T').bfill())
```

```
# 低频转高频，主要是如何插值
# .asfreq()：不做填充，返回Nan
# .ffill()：向上填充
# .bfill()：向下填充
```

```

```

作业

作业1：请输出以下时间序列

```
---时间序列1:-----
2022-01-01    0.123626
2022-01-02    0.932822
2022-01-03    0.251188
2022-01-04    0.079741
2022-01-05    0.328489
Freq: D, dtype: float64
---时间序列2:-----
2022-01-31 00:00:00    0.242894
2022-01-31 00:00:01    0.377202
2022-01-31 00:00:02    0.046502
2022-01-31 00:00:03    0.873964
Freq: S, dtype: float64
---时间序列3:-----
                       value1    value2    value3
2021-12-01 00:00:00  0.748819  0.368575  0.286854
2021-12-01 00:10:00  0.231864  0.828566  0.117412
2021-12-01 00:20:00  0.342888  0.609301  0.296714
2021-12-01 00:30:00  0.936361  0.448391  0.349759
```

```

```