Pandas架构

1、数据结构

Series: 均匀数据，尺寸大小不变，数据的值可变

Dataframe： 异构数据，大小可变，数据可变

Panel：异构数据，大小可变，数据可变

2、对象创建

import pandas as pd

import numpy as np

Series

s = pd.Series([1,3,5,np.nan,6,8])

Dataframe

dates = pd.date\_range('20170101', periods=7)

df = pd.DataFrame(np.random.randn(7,4), index=dates, columns=list('ABCD'))

df2 = pd.DataFrame({ 'A' : 1.,

'B' : pd.Timestamp('20170102'),

'C' : pd.Series(1,index=list(range(4)),dtype='float32'),

'D' : np.array([3] \* 4,dtype='int32'),

'E' : pd.Categorical(["test","train","test","train"]),

'F' : 'foo' })

3、查看数据

df.head()

df.tail(3)

df.index

df.values

df.describe()

df.T

df.sort\_index(axis=1, ascending=False)

df.sort\_values(by='B')

生产代码，但建议使用优化的Pandas数据访问方法.at，.iat，.loc，.iloc和.ix

4、选择区块

df['A']

df['20170102':'20170103']

按标签选择

df.loc[dates[0]]

df.loc[:,['A','B']]

df.loc['20170102':'20170104',['A','B']]

df.loc['20170102',['A','B']]

df.loc[dates[0],'A']

df.at[dates[0],'A'

通过位置选择

df.iloc[3]

df.iloc[3:5,0:2]

df.iloc[[1,2,4],[0,2]]

df.iloc[1:3,:]

df.iloc[:,1:3]

df.iloc[1,1]

df.iat[1,1]

布尔索引

df[df.A > 0]

df[df > 0]

df2[df2['E'].isin(['two','four'])]