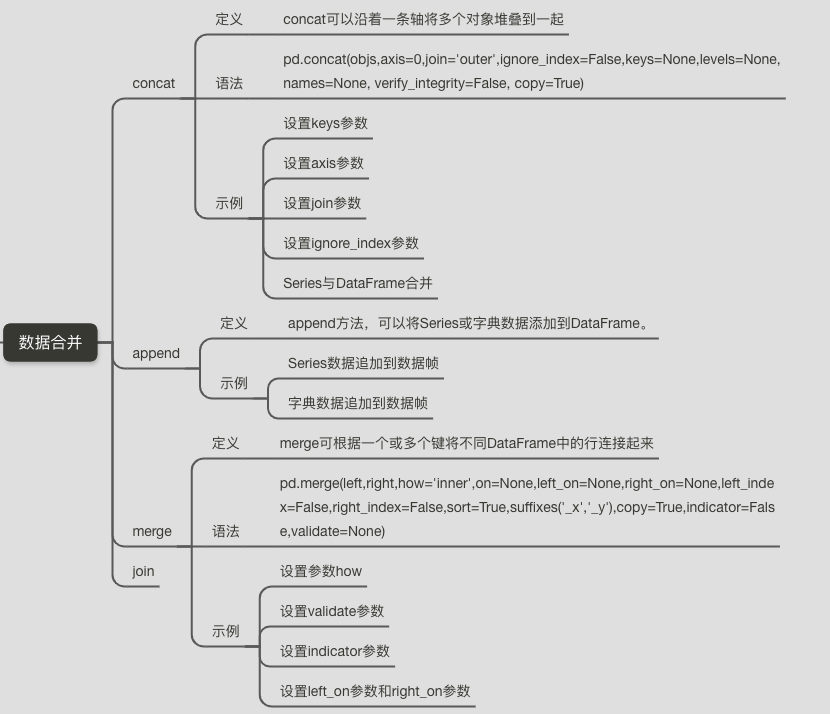
基础|Pandas常用知识点汇总(四)

Pandas 是一套用于 Python 的快速、高效的数据分析工具。它可以用于数据挖掘和数据分析，同时也提供数据清洗功能。本文将详细讲解数据合并与连接，目录如下：



①concat

**一.定义**

concat函数可以在两个维度上对数据进行拼接，默认纵向拼接（axis=0），即按行拼接。拼接方式默认为外连接（outer)，即取所有的表头字段或索引字段。

**二.语法**

pd.concat(objs,axis=0, join='outer', ignore\_index=False,

keys=None,levels=None, names=None, verify\_integrity=False,

copy=True)

**参数释义：**

***objs****：需要用于连接合并的对象列表*

***axis****：连接的方向，默认为0（按行），按列为1*

***join****：连接的方式，默认为outer，可选inner只取交集*

***ignore\_index****：合并后的数据索引重置，默认为False，可选True*

***keys****：列表或数组，也可以是元组的数组，用来构造层次结构索引*

***levels****：指定用于层次化索引各级别上的索引，在有keys值时*

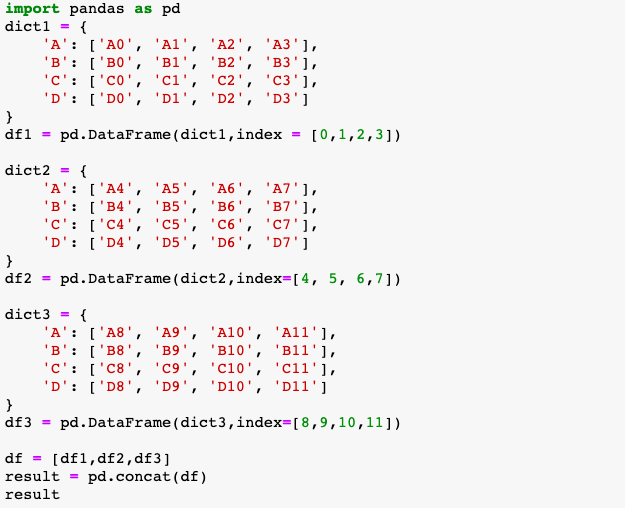
***names****：用于创建分层级别名称，在有keys和levels时*

***verify\_integrity****：检查连接对象中新轴是否重复，若是则异常，默认为False允许重复*

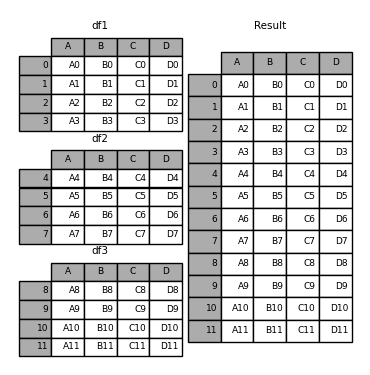
***copy****：默认为True，如果是False，则不会复制不必要的可以提高效率*

**三.代码示例**

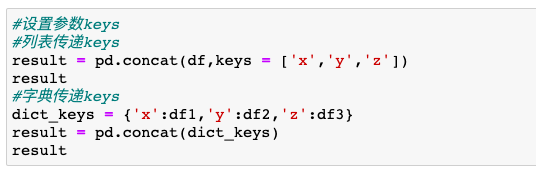
首先，我们看一个简单的例子：

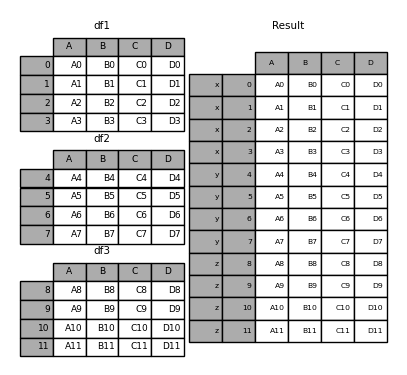


以上代码构建了3个字典并转为DataFrame，然后通过concat实现默认方式合并。合并演示如下：

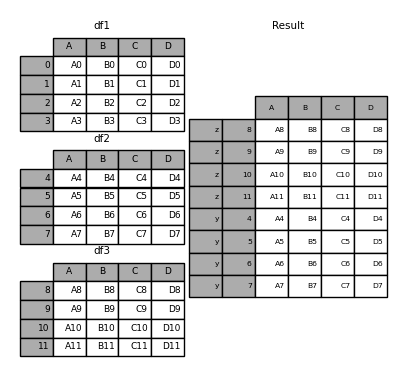


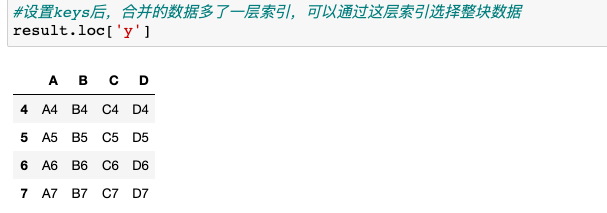
**1.设置参数keys**





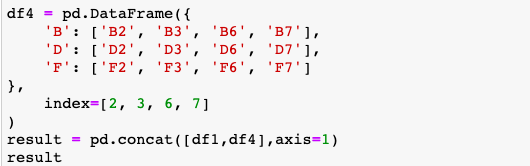
图片

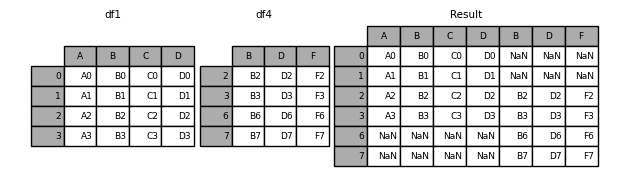


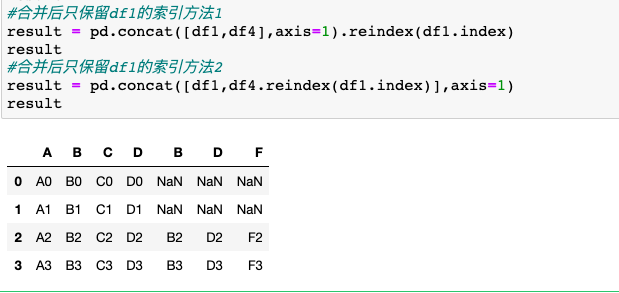


**2.设置axis参数**

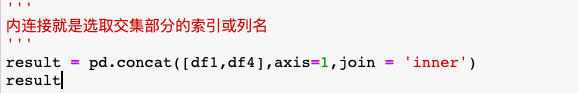
axis=1按列合并，默认情况下，join='outer'，合并时索引全部保留，对于不存在值的部分会默认赋NaN。

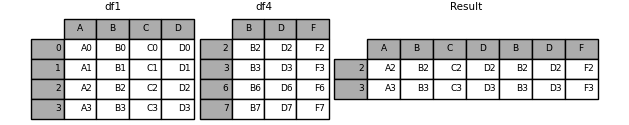




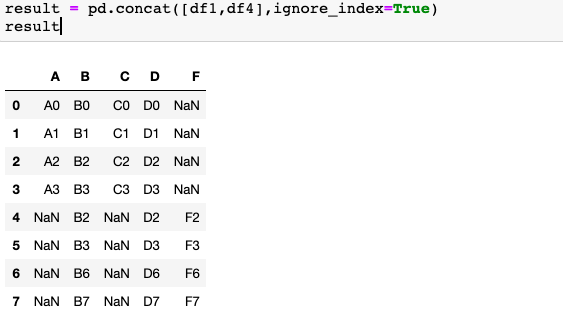


**3.设置join参数**

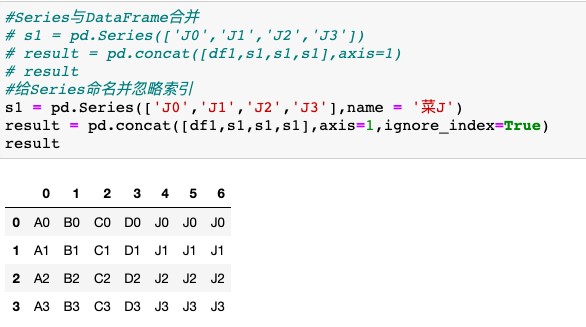




**4.设置ignore\_index参数**



**5.Series与DataFrame合并**

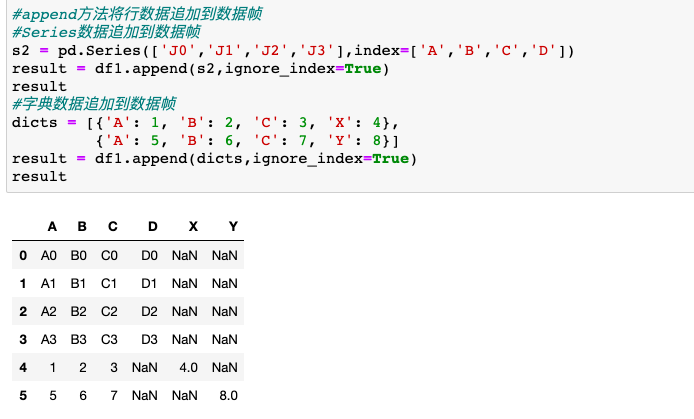


②  
append

**一.定义**

运用append方法，可以将Series或字典数据添加到DataFrame。

**二.代码示例**



③  
  
merge

**一.定义**

merge函数可根据一个或多个键（列）相同进行DataFrame拼接。类似于关系型数据库的join操作。

**二.语法**

pd.merge(left, right, how='inner', on=None, left\_on=None,

right\_on=None,left\_index=False,right\_index=False, sort=True,

suffixes=('\_x', '\_y'), copy=True, indicator=False,validate=None)

**参数释义：**

***left****：参与合并的左侧数据*

***right****：参与合并的右侧数据*

***how****：合并类型：inner(默认内连接)、outer（外连接）、left（左连接）、right（右连接）*

***on****：用于连接的列名，默认为左右侧数据共有的列名，指定时需要为左右侧数据都存在的列名*

***left\_on****：左侧数据用于连接的列*

***right\_on****：右侧数据用于连接的列*

***left\_index****：将左侧索引作为连接的列*

***right\_index****：将右侧索引作为连接的列*

***sort****：排序，默认为True，设置为False可提高性能*

***suffixes****：默认为('\_x', '\_y')，可以自定义如('date\_x','date\_y')*

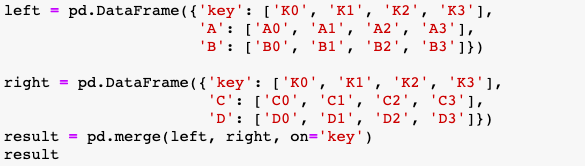
***copy****：默认为True，如果是False，则不会复制不必要的可以提高效率*

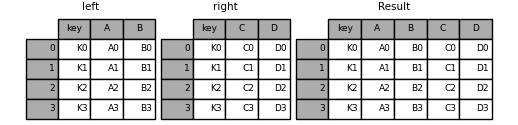
***indicator****：指示器，默认False，设置为True时会新增一列标识*

***validate****：字符串，如果指定则会检测合并的数据是否满足指定类型（1对1，1对多，多对1，多对多）*

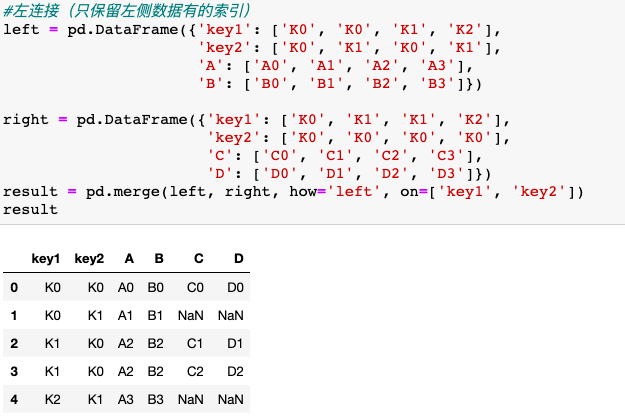
***三.示例***

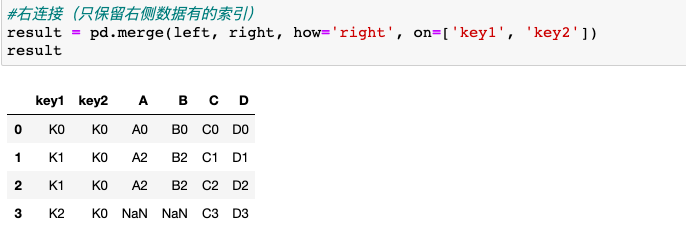
首先，我们看一个简单的例子：

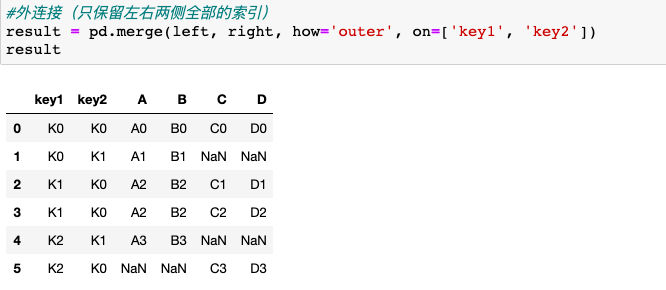


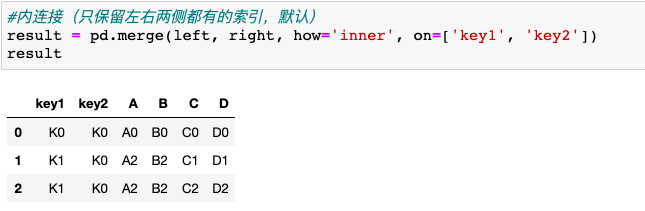


**1.设置参数how**

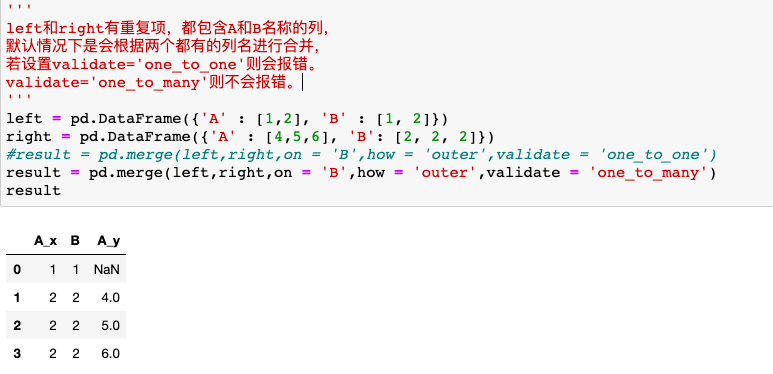




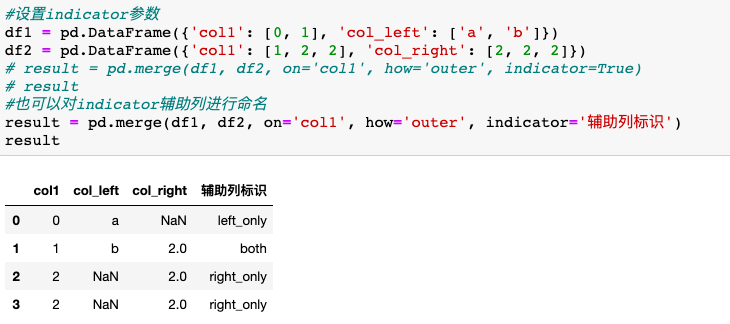




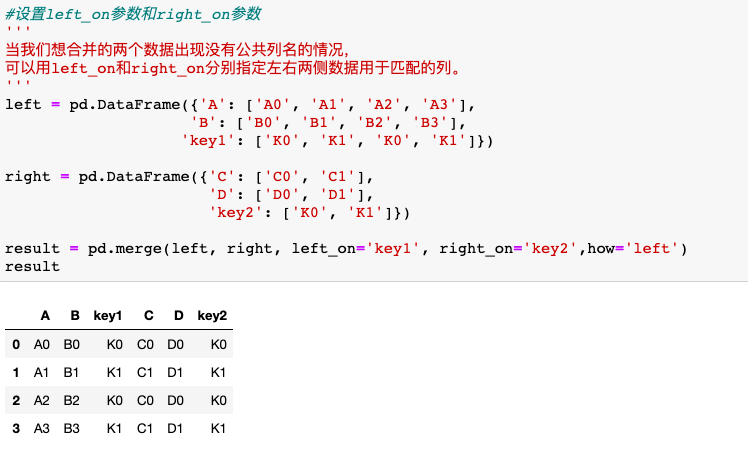
**2.设置validate参数**



**3.设置indicator参数**



**4.设置left\_on参数和right\_on参数**

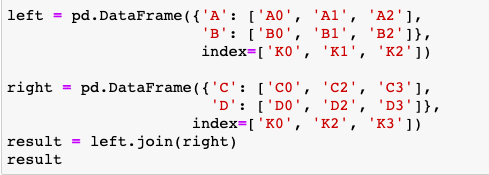


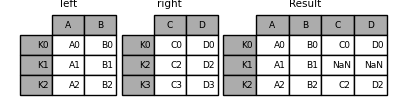
④  
  
  
join

**一.定义**

join可以将两个没用共同列名的数据进行快速合并，默认是保留被合并的数据索引。join接受的参数有how、on和suffix等。

**二.示例**





**结语**

OK，第四篇完！敬请期待下篇，下篇内容预告：数据分组与透视。欢迎关注菜J学Python，我们坚持认真写Python基础，幽默写Python实战。你可在公众号后台免费领取相关学习资料或学习交流。

1. [*睡地摊or租房？爬取某大型房产网站24685个房源信息并分析，助你选择*](http://mp.weixin.qq.com/s?__biz=MzAwMjYxNjcwMw==&mid=2456405069&idx=1&sn=f74df8b3f4fc9383d3da2da7211dc612&chksm=8d5ad85dba2d514b7b128b5b361b8d496f1bac5498950ffc073cf714151b3917e948854bbe04&scene=21#wechat_redirect)
2. [*摆地摊or打工，爬取某大型招聘网站3万条招聘信息并分析，作何选择?*](http://mp.weixin.qq.com/s?__biz=MzAwMjYxNjcwMw==&mid=2456404450&idx=1&sn=4cee4c349aaadfaba69a9ec3919b42e3&chksm=8d5addf2ba2d54e43b6306f68524c9c602a9a9e77041071b792d0bd4ebe96a0061d3c8d7616e&scene=21#wechat_redirect)
3. [*实战|手把手教你用Python爬虫(附详细源码)*](http://mp.weixin.qq.com/s?__biz=MzAwMjYxNjcwMw==&mid=2456403634&idx=1&sn=41e3e71b16be2a0a25c1eaff063e16af&chksm=8d5aa2a2ba2d2bb4dc8e13bac62d8fef84910535d3c88c1d346de4f30f0995687cef810a3f74&scene=21#wechat_redirect)
4. [*实战|利用Python获取财务金融数据的几种实用方法*](http://mp.weixin.qq.com/s?__biz=MzAwMjYxNjcwMw==&mid=2456404246&idx=1&sn=1e3b647baf6b2c208033cc2b5376f9fa&chksm=8d5add06ba2d5410810b39467900035efcd9acf240998ceebbe1456b44631069d72b8bc12b1e&scene=21#wechat_redirect)
5. [*实战|Python爬虫并用Flask框架搭建可视化网站*](http://mp.weixin.qq.com/s?__biz=MzAwMjYxNjcwMw==&mid=2456404005&idx=1&sn=381e740a8169ff5ac31a5df587383cb0&chksm=8d5adc35ba2d5523121f17d7ddb7bc06ae262548f0b8ca9e6020928fdc78e5bfdee9a40f5edc&scene=21#wechat_redirect)