

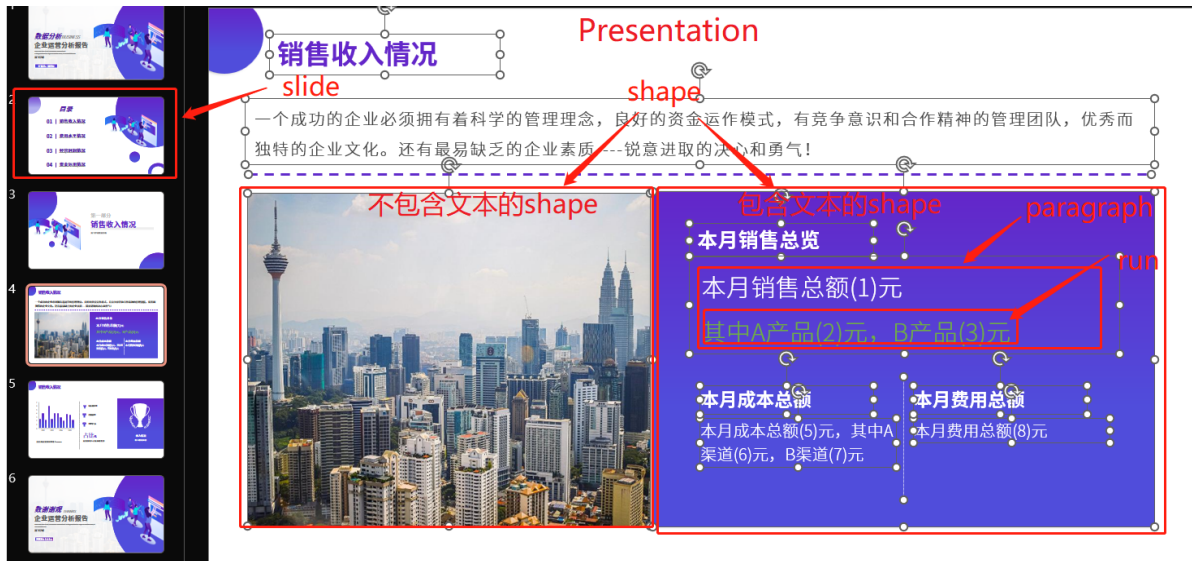
# 一、Python操作ppt

pptx 是一个非标准库，需要在命令行中安装

```
pip install python-pptx
```

## 1. 基本结构

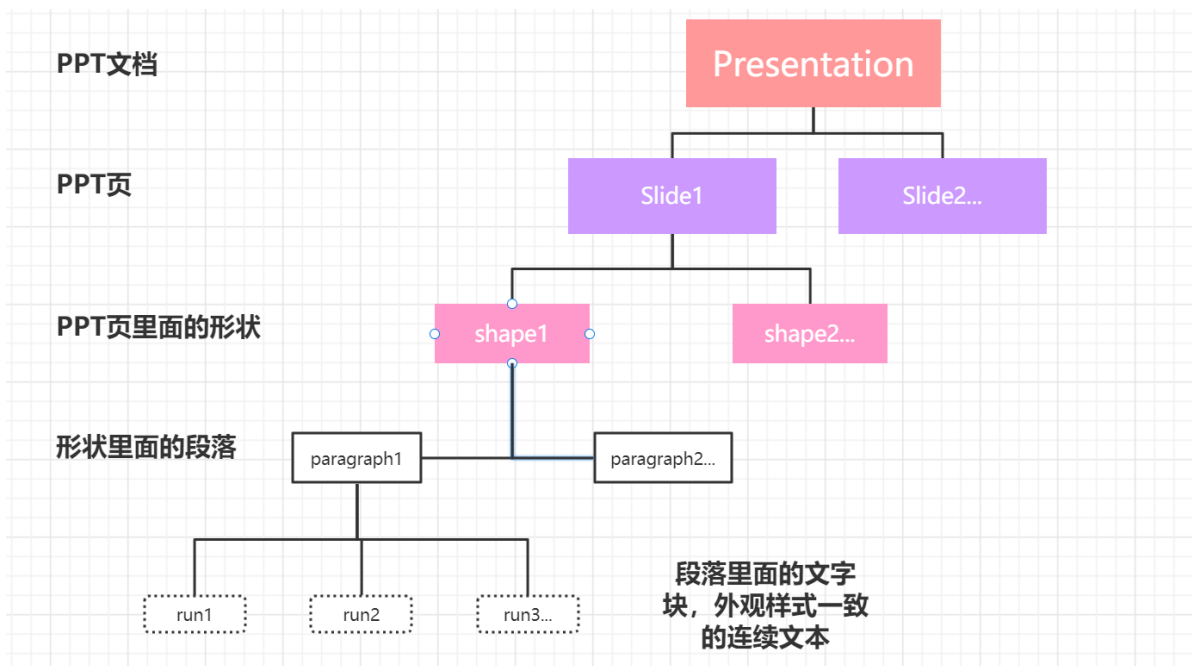
ppt结构组成如下：



简单来说，一个PPT文件为**presentation**，基本的结构为展示

**文件presentation-幻灯片页slide-形状shape**组成，形状就需要区分开，是包含文本的形状还是不包含文本的形状(纯图片等)。

如果是包含文本的形状，则可以通过获取内部的文本框，一个文本框又可以看作是一个小的word文档，包含**段落paragraph - 文字块run**



## 二、Python访问PPT

### 1. 打开PPT文件

```
from pptx import Presentation
# 这里给出需要打开的文件路径
file_path = r'...'
pptx = Presentation(file_path)
```

### 2. 获取幻灯片页

用 `pptx.slides` 可以获得一个列表，包括所有的幻灯片页 `slide` 对象

```
for slide in pptx.slides:
    print(slide)
```

### 3. 获取形状

只要熟悉了类似 Excel 和 Word 的 `多级结构`，PPT 的结构就很好理解了。每一个幻灯片页都有一个或者多个形状 `shape`

```
for slide in pptx.slides:
    for shape in slide.shapes:
        print(shape)
```

### 4. 获取文本框内容

要获取文字内容，很容易就联系到文字在形状 `shape` 的下级结构了 从 Word 中的学习我们也可以推知，文字的承载单位是 `段落 paragraph` 和 `文字块 run`

很自然可以想到用下列的代码获取文字

```
for slide in pptx.slides:
    for shape in slide.shapes:
        for paragraph in shape.text_frame.paragraphs:
            print(paragraph.text)
```

或者

```
for slide in pptx.slides:
    for shape in slide.shapes:
        for paragraph in shape.text_frame.paragraphs:
            for run in paragraph.runs:
                print(run.text)
```

## 5. 判断shape是否有文字

从上图可以看到，左侧的图片的shape是没有任何文字的

一个形状中有没有文字，关键就在于它有没有包含文本框

下面是与文本框有关的操作：

- `shape.has_text_frame` 判断形状中是否有文字框
- `shape.text_frame` 获取文字内容

在PPT中，文字框才是文字的载体，因此可以先判断shape是否包含文字`has_text_frame`，获取文字的代码如下：

```
for slide in pptx.slides:
    for shape in slide.shapes:
        if shape.has_text_frame:
            text_frame = shape.text_frame
            print(text_frame.text)
```

## 5. 访问段落和文字块

每一个文本框都可以看成是一个小的 Word 文件，里面有段落和文字块两级结构：

```
for slide in pptx.slides:
    for shape in slide.shapes:
        if shape.has_text_frame:
            text_frame = shape.text_frame
            for paragraph in text_frame.paragraphs:
                for run in paragraph.runs:
                    run.text = '新的文字内容'
```

注意，要保留原始的ppt格式，需要使用run来设置文字，否则格式会被paragraph重置。

## 三、写入 PPT

创建全新 PPT 的代码可以类比创建 Word 文件的代码，实例化的过程中不给予具体路径则为创建空白文件

```
ppt = pptx.Presentation(pptx='数据报告模板.pptx')
ppt.save('存储路径.pptx')
```