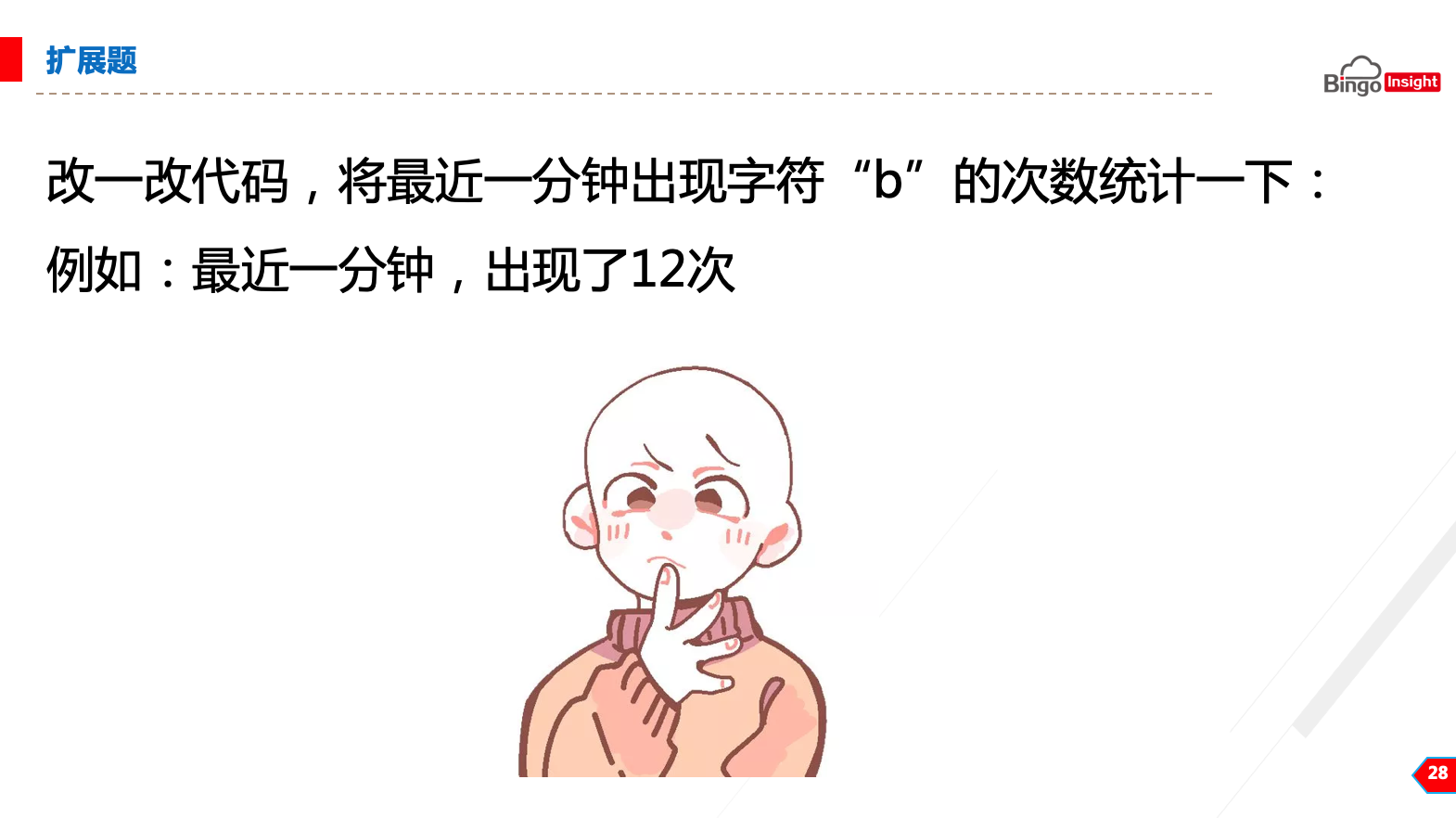
3、作业提交方式：

在github现有仓库中，创建一个“大数据实时计算扩展题作业”文件夹，并在文件夹中按照每个实操扩展题创建一个文件夹，将作业代码提交上来即可。

**课后扩展题目1:**



1. 题目描述

在实操1中，我们对socket实时输入的内容进行了对关键字“b”的判断，将含有关键字“b”的字符输出来。现在我们需要改一改代码，统计最近一分钟出现字符“b”的次数。

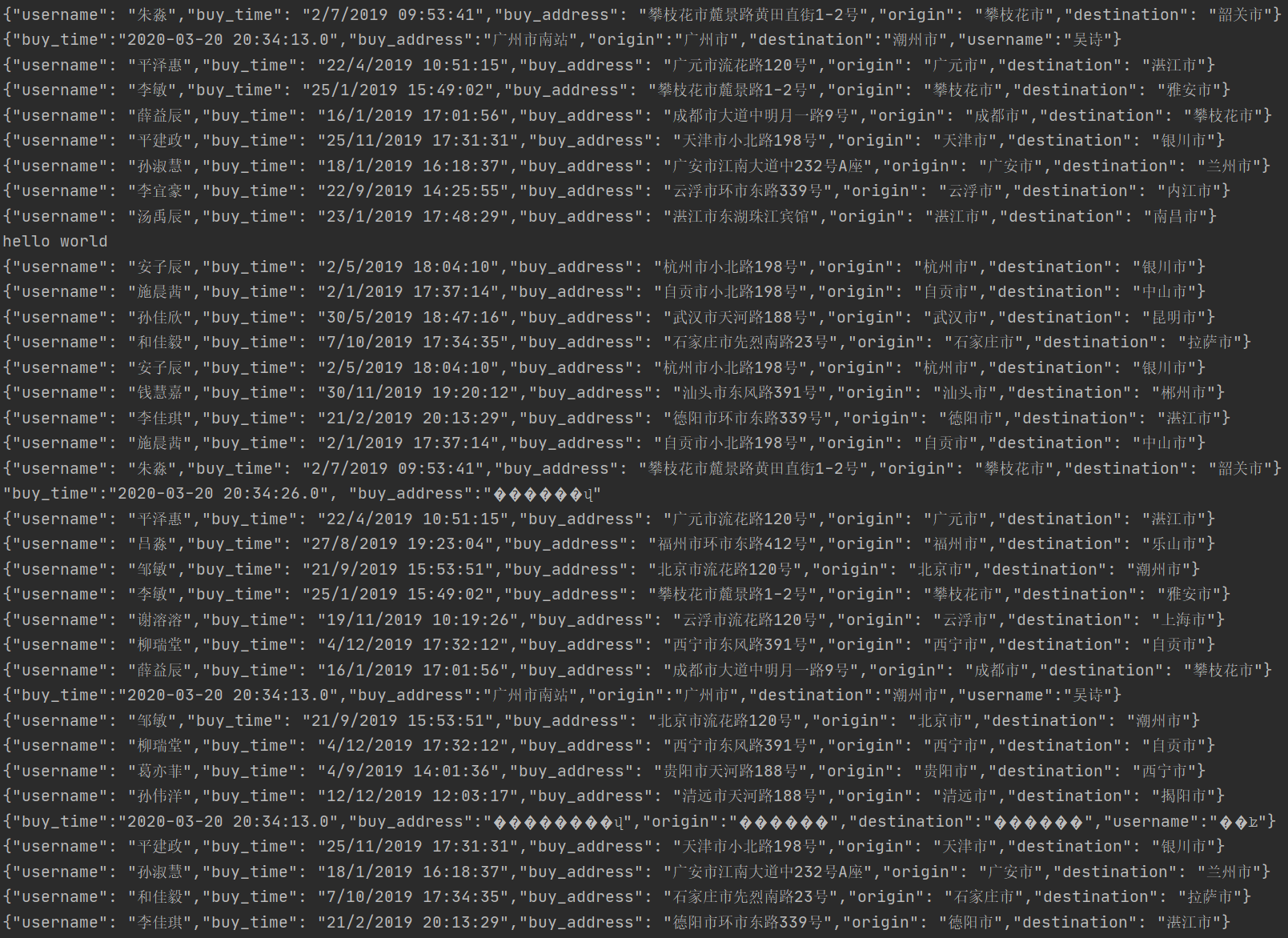
1. 题目提示：

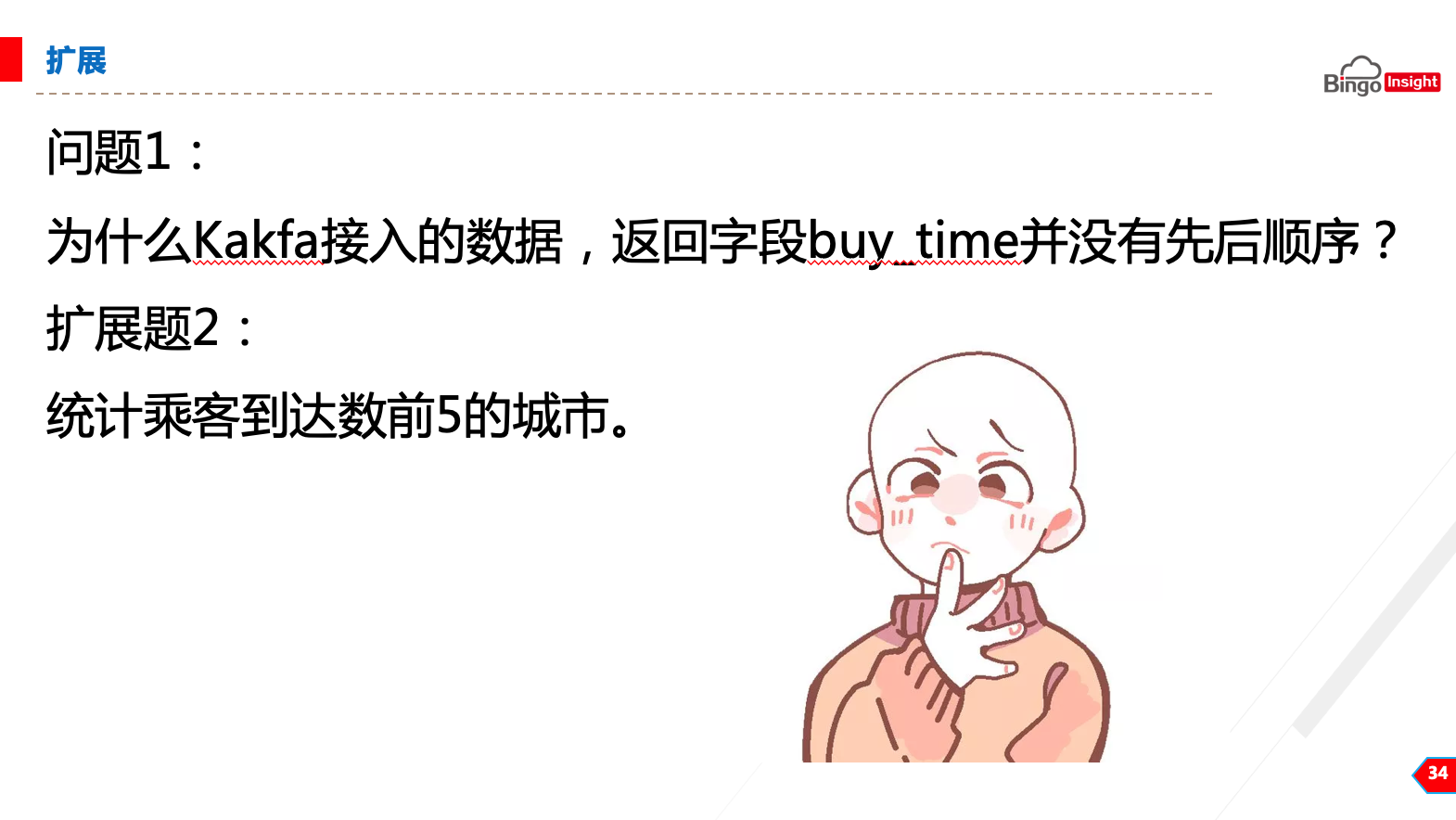
最近一分钟，这个描述需要按时间的分段统计，需要使用到时间窗口函数。

统计出现的次数，这个我们需要对MapFunction定义实现，将输入的字符“b”运行统计

**课后扩展题目2:**

**QQ浏览器截图20200605223534**



****

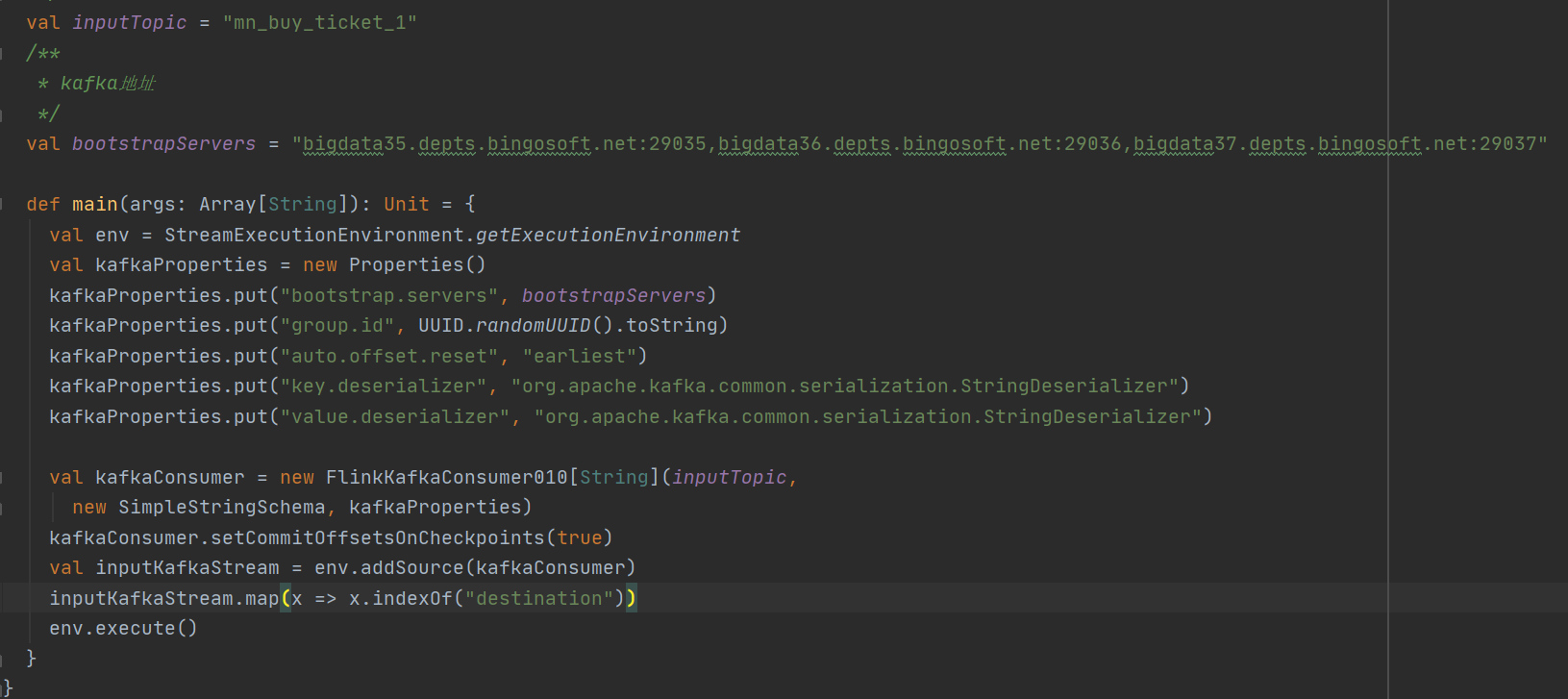
1. 题目描述

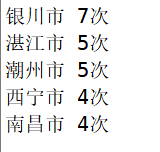
在实操2中，我们接入了来自kafka的消息，在消息中，我们读取了来自主题的数据，那么我们如何统计到达数为前五的城市，通过结果输出示例：（广州市，12）。

1. 题目提示：

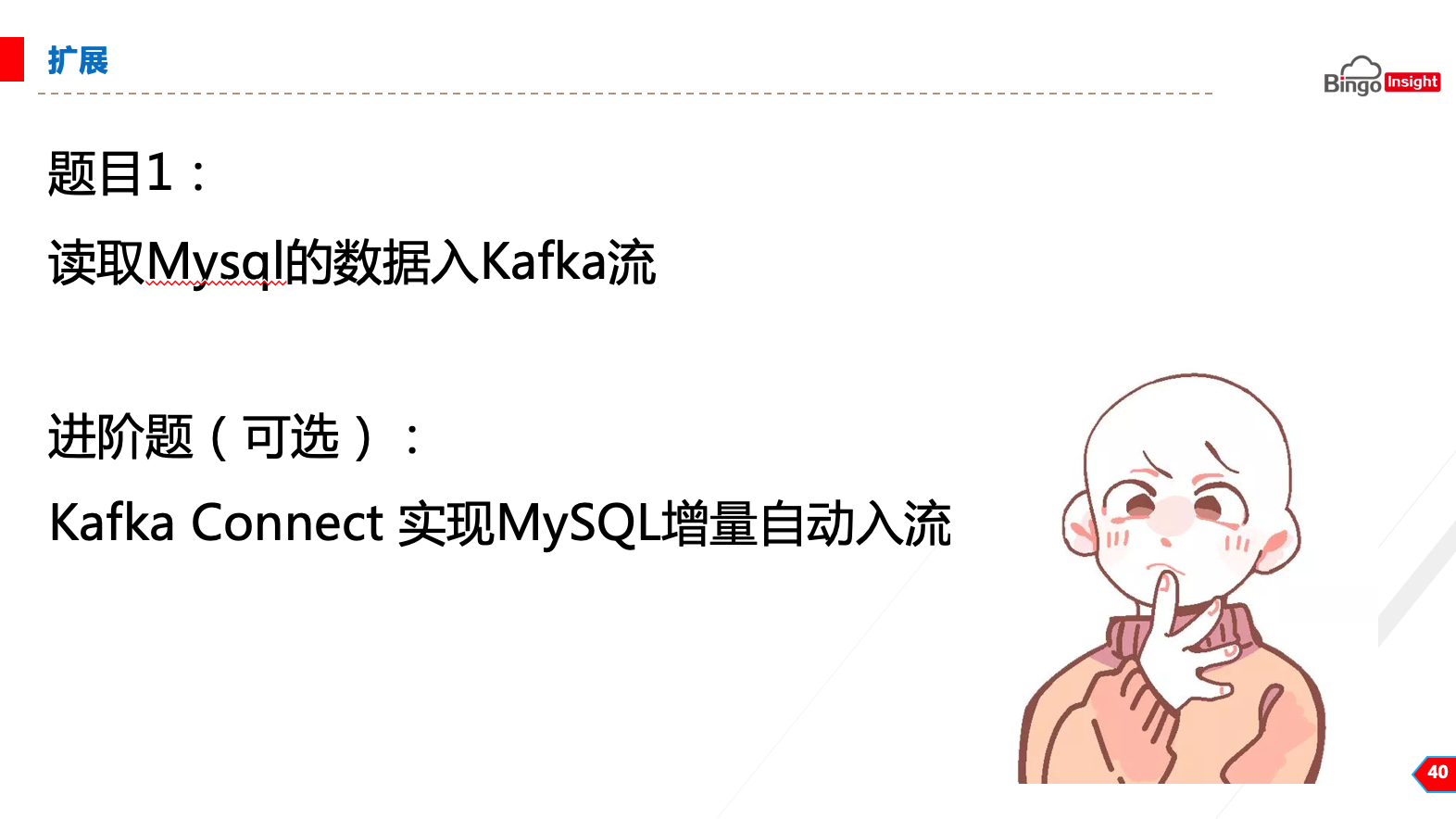
在实操2中，我们体会了Kafka的消数据，那么现在需要大家统计到达目的地城市前五的城市。

1. 我觉得可能是因为我们并没有对Kafka接入的数据做任何的处理（比如筛选、排序之类的），所以该字段并没有先后顺序





**课后扩展题目3:**

****

1. 题目描述

在实操3中，我们使用kafka的Produce 开发了从数据湖读取到数据，再将该数据提交到kafka中。现在我们需要将Mysql的表读取成一条条数据，并表的数据提交到kafka消息中。

1. 题目提示：

读取到Mysql的数据并且入Kafka中，该题目和和实操3中的区别在于，实操是使用了S3client读取的

4、其他要求：

为保障大家的性能，Mysql的表数据量条数不能大于1000条。

