Kafka幂等性

幂等性实现

PID 和 Sequence Number

为了实现Producer的幂等性，Kafka引入了Producer ID（即PID）和Sequence Number。

1. PID。每个新的Producer在初始化的时候会被分配一个唯一的PID，这个PID对用户是不可见的。
2. Sequence Numbler。（对于每个PID，该Producer发送数据的每个<Topic, Partition>都对应一个从0开始单调递增的Sequence Number。

Broker端在缓存中保存了这seq number，对于接收的每条消息，如果其序号比Broker缓存中序号大于1则接受它，否则将其丢弃。这样就可以实现了消息重复提交了。但是，只能保证单个Producer对于同一个<Topic, Partition>的Exactly Once语义。不能保证同一个Producer一个topic不同的partion幂等。