Software Architecture

Lecture 5. Case Study on Dataflow System

Professor:
Yushan (Michael) Sun
Fall 2020

Legacy File Update
旧文件更新系统

旧文件更新系统 (Legacy File Update)

问题描述：利用管道-过滤器体系结构设计一个旧文件更新系统

In a text file,

1) **Fix the Year 2000 problem: Find and change all year expressions "xy" into "19xy", e.g. "89" into "1989", and**
2) **Fix political issues: find and change all expressions "Republic of China" into "Taiwan"**

旧文件更新系统 (Legacy File Update)

Solution: Pipe-and-filter architecture
设计：使用Pipe-and-filter软件体系结构

Design four filters:

Input, Y2K, Politics, output

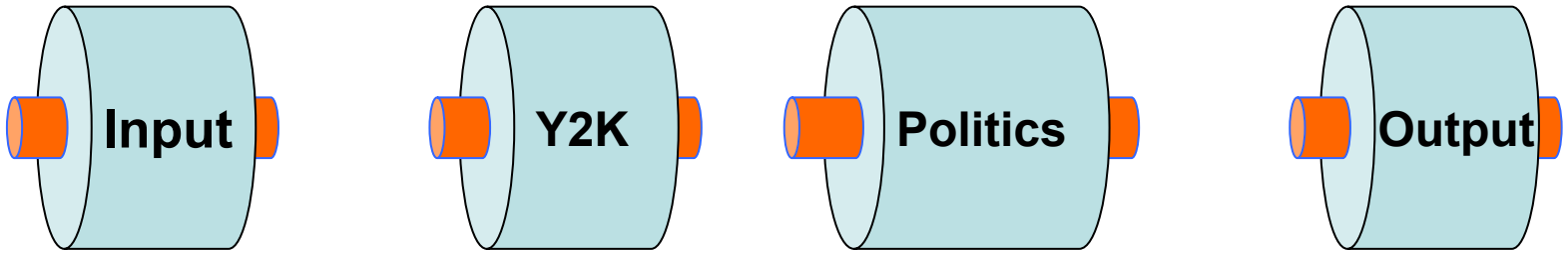
Each filter processes the data and sends it to the next filter. The four filters are connected by three pipes.

Control: The control is distributed.

- Each filter can run whenever it has data on which to compute. Data sharing between filters is strictly limited to that transmitted on pipes

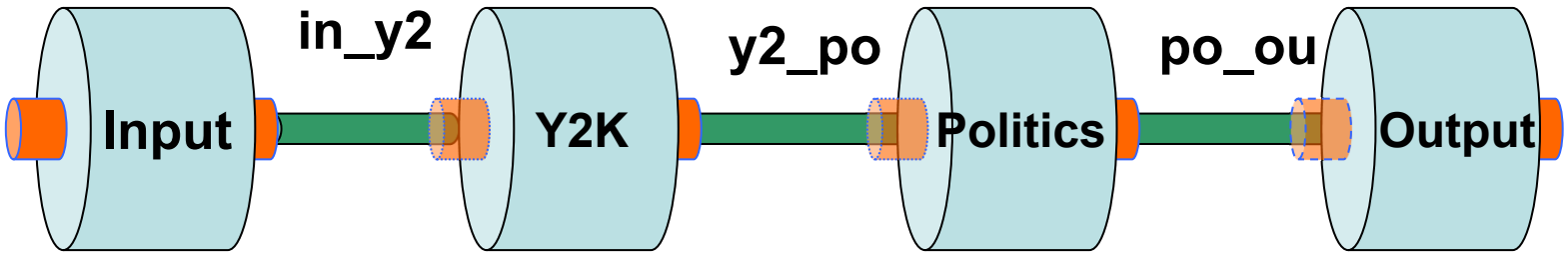
旧文件更新系统 (Legacy File Update)

Four filters with in-ports and out-ports



旧文件更新系统 (Legacy File Update)

使用3个管道将4个过滤器连接起来



- 设计要求1: 像生产线上的机器一样，所有的过滤器应该同时工作
  - Q1: How to implement this functionality?
- 设计要求2:像水流一样，数据应该源源不断的从左端流向右端
  - Q2: How to implement this functionality?

旧文件更新系统 (Legacy File Update)

过滤器的功能：

- Input filter: is connected to a source file
  a) read the content of an input file, in character stream
  b) parse it, and
  c) write the parsed contents to its output stream, character by character (确保空格与回行符也写入下游数据流).

- Y2K filter: connects to pipe in_y2, output stream of Input filter.
  a) read character stream from pipe in_y2
  b) Organize them into a line
  c) Find and fix Y2k problem in this line
  d) Once a line is fixed, then write it to the y2_po, writing one character at a time (确保空格与回行符也写入下游数据流).

旧文件更新系统 (Legacy File Update)

- Politics filter: connected to pipe y2_po, the output stream of Y2K filter.
  a) read character stream from the pipe y2_po
  b) organize them into a line
  c) Find and fix the political problem in this line
  d) Once a line is fixed and then write it to the pipe po_ou, character by character (确保空格与回行符也写入下游数据流).
- Output filter: connected to pipe po_ou, the output stream of Politics filter.
  – read character stream from po_ou
  – writes them to the standard output.

旧文件更新系统 (Legacy File Update)

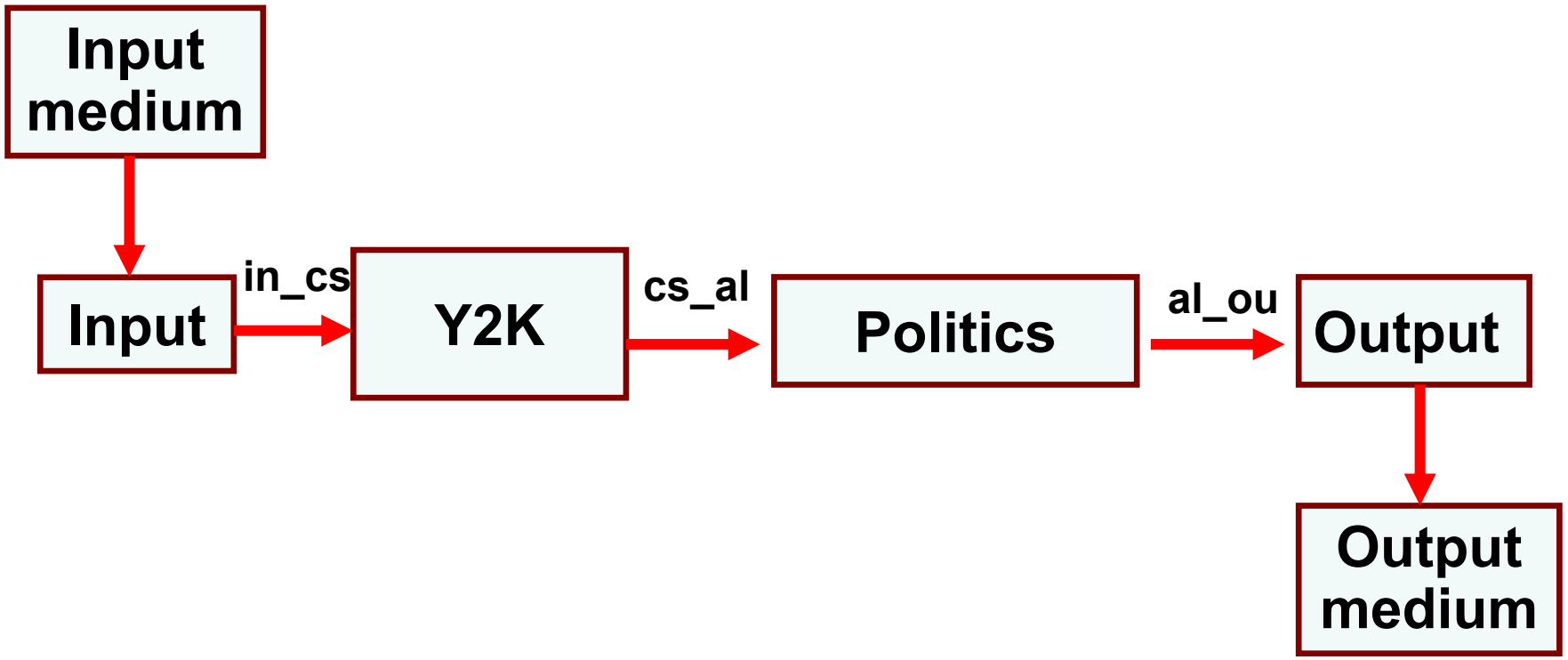
- 主要控制类（主程序）的责任：
  1）创建Filter对象
  2）创建Pipe对象
  3）将它们组装（链接）起来
  4）启动运行过滤器对象
- Master Control manages filter and pipe mechanism, by creating a pipeline of the above described filters.

旧文件更新系统 (Legacy File Update)

- **Functionalities of the pipes:**
- **in_y2 pipe:** shared between Input filter and Y2K filter,
  - **this pipe is the output pipe for Input filter and the input pipe for Y2K filter.**
- **y2_po pipe:** shared between Y2K filter and Politics filter,
  - **this pipe is the output pipe for Y2K filter and the input pipe for Politics filter.**
- **po_ou pipe:** shared between Politics and Output filter,
  - **this pipe is the output pipe for** Politics **filter and the input pipe for Output filter.**

旧文件更新系统 (Legacy File Update)

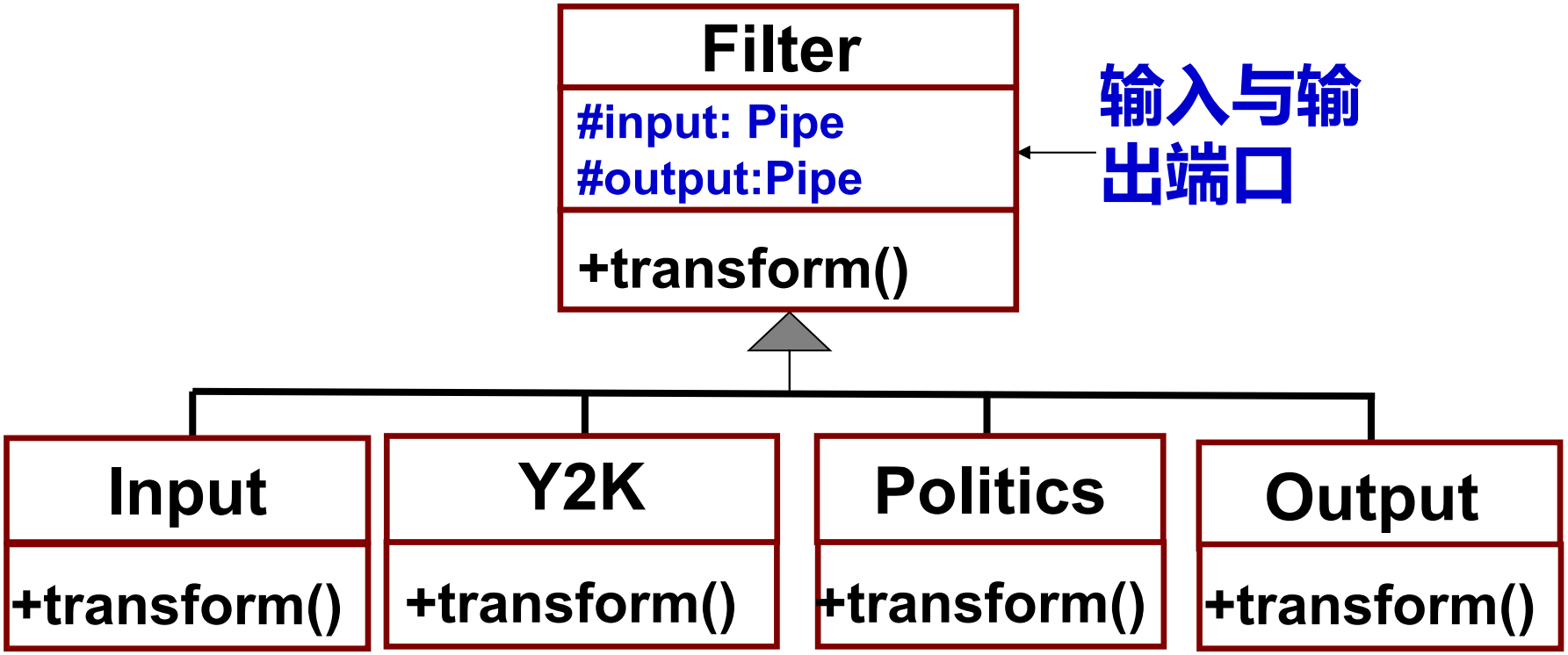
从而，我们得到程序的逻辑结构如下



Architecture of the program-the program is formed by the filters and pipes

旧文件更新系统 (Legacy File Update)

过滤器的设计: Need a common interface Filter with 2 pipes, one for input and the other for output.



Class hierarchy of Filters

旧文件更新系统 (Legacy File Update)

过滤器的结构 （Structure of the filters）

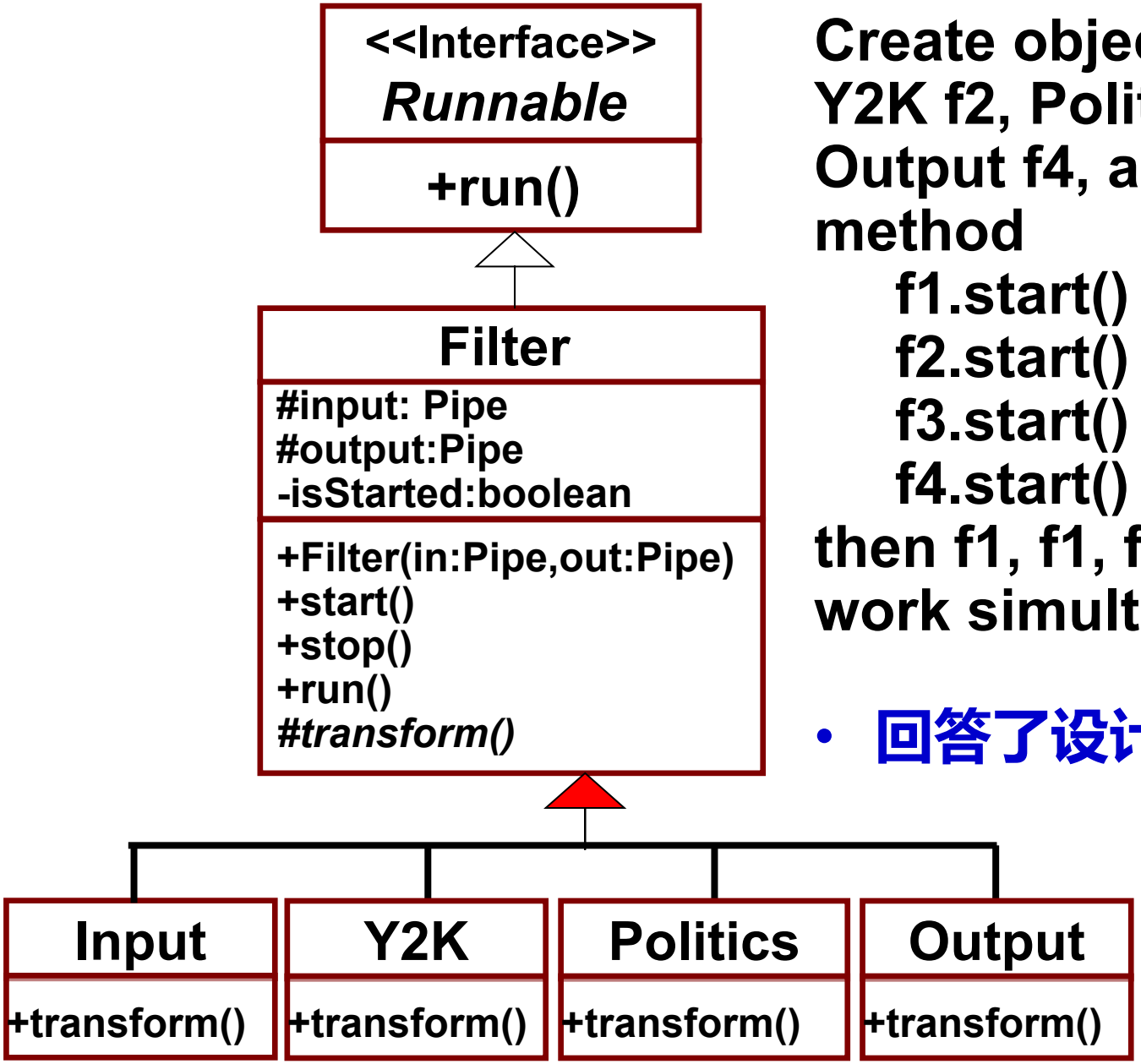
- Filters in the system are represented by an abstract class named Filter. An instance of Filter class is composed of:
  - Input pipe object, from which the filter object reads the data
  - Output pipe object, to which the filter object writes the processed data

旧文件更新系统 (Legacy File Update)

- 设计要求1: 怎样使4个过滤器对象同时工作?

- **解决方案(Solution): 利用多线程机制**
- **All the 4 filters should have multi-thread properties. So, let Filter class implements interface Runnable.**

旧文件更新系统 (Legacy File Update)



Create objects of Input f1, Y2K f2, Politics f3 and Output f4, and call start method

f1.start()
f2.start()
f3.start()
f4.start()

then f1, f1, f3 and f4 will work simultaneously

- 回答了设计要求1

旧文件更新系统 (Legacy File Update)

- **Methods**:
  - **start**(): start a thread and run the thread
  - **run**(): called by the thread when needed
  - **stop**(): stop a thread
  - **Transform**(): the abstract method that
    - **Inputs data (in Input filter),**
    - **Finds Y2K and processes Y2K (in Y2K filter),**
    - **Finds and fixes political problem (in Politics filter)**
    - **outputs data**
- **Process of the run:**

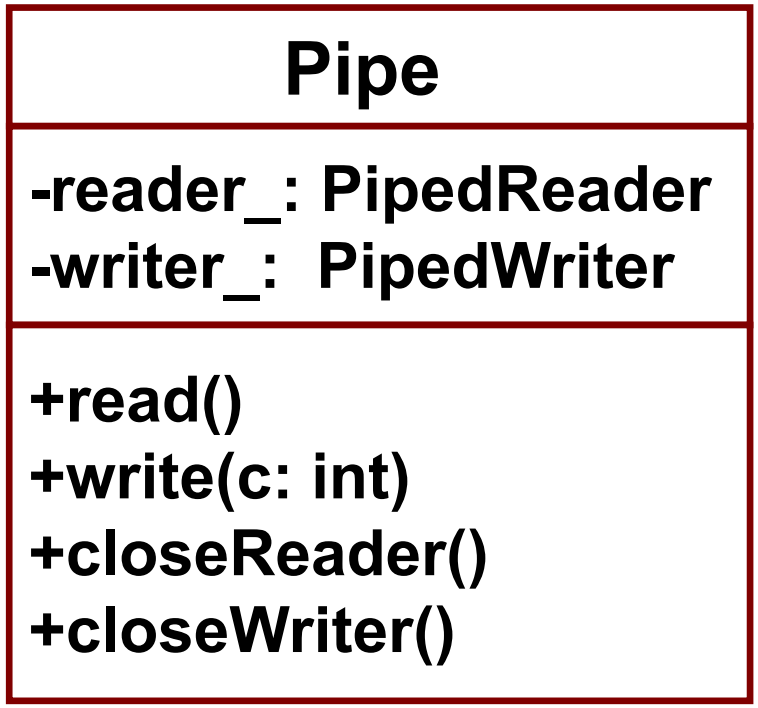
start➔ run➔ transform()

旧文件更新系统 (Legacy File Update)

Design of the pipe class (考虑回答设计要求2）:

- Pipes in the Legacy file update system are represented as instances of the Pipe class.
- An instance of the Pipe class is a composition of two streams: an input and an output stream.
- Pipe class connects these two streams in the following way.
  - **The data that is written to the input stream is transmitted to the output stream. In this way these data become available for reading from the output stream.**

旧文件更新系统 (Legacy File Update)



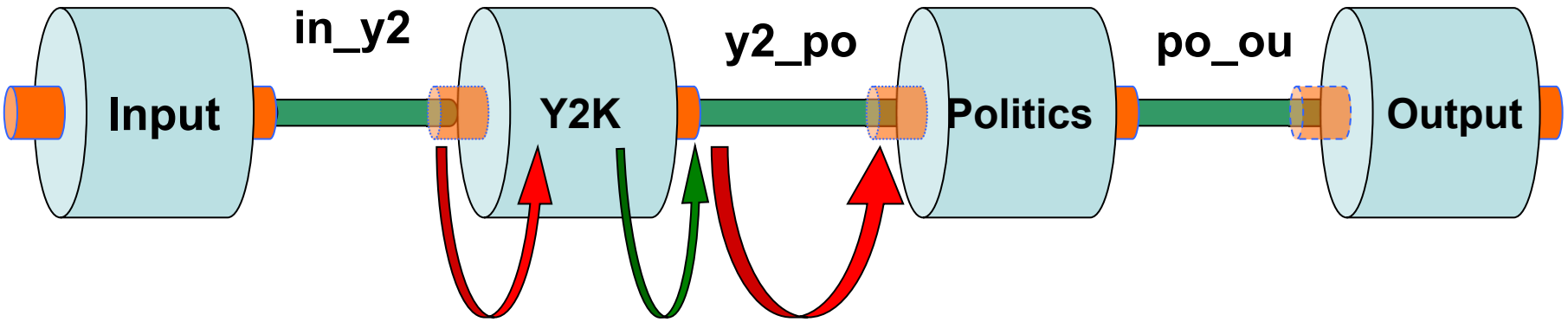
Java API classes PipedReader and PipedWriter are used to implement data stream to keep data flowing continuously

The Pipe class encapsulates the input and output streams as its private variables and provide just a simple public Interface for writing data to and reading data from a pipe object.

旧文件更新系统 (Legacy File Update)

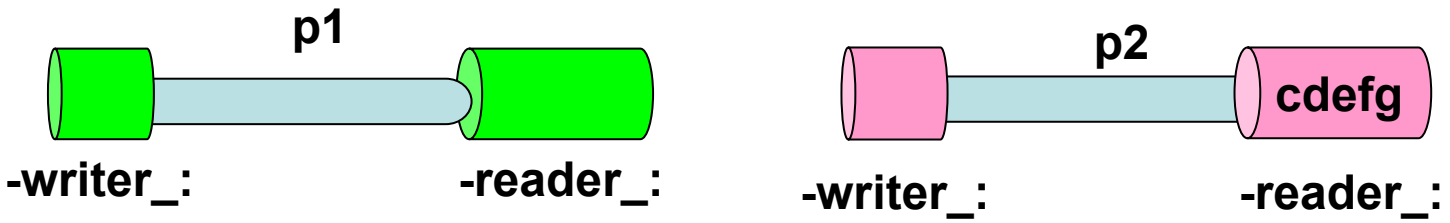
回答设计要求2：怎样使得数据像水流一样从左向右移动?

一串管道是怎样搬运数据的呢?

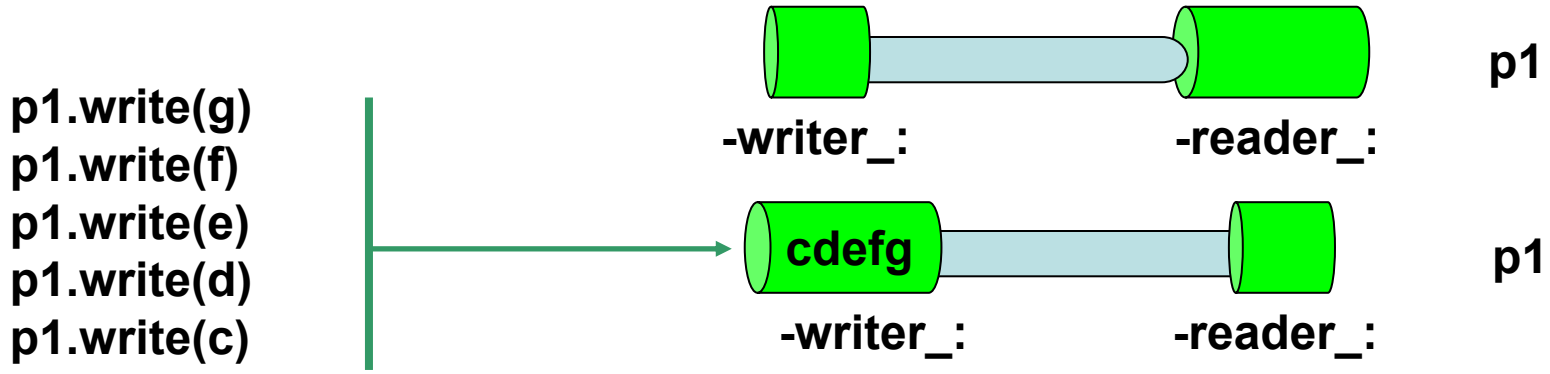


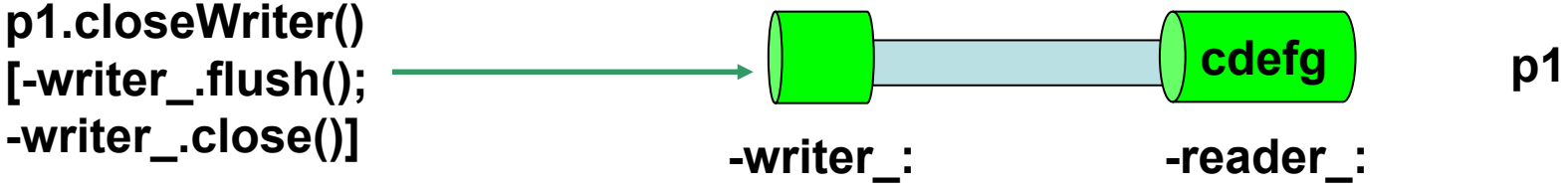
旧文件更新系统 (Legacy File Update)

先看看相邻两个管道是怎样搬运数据的?

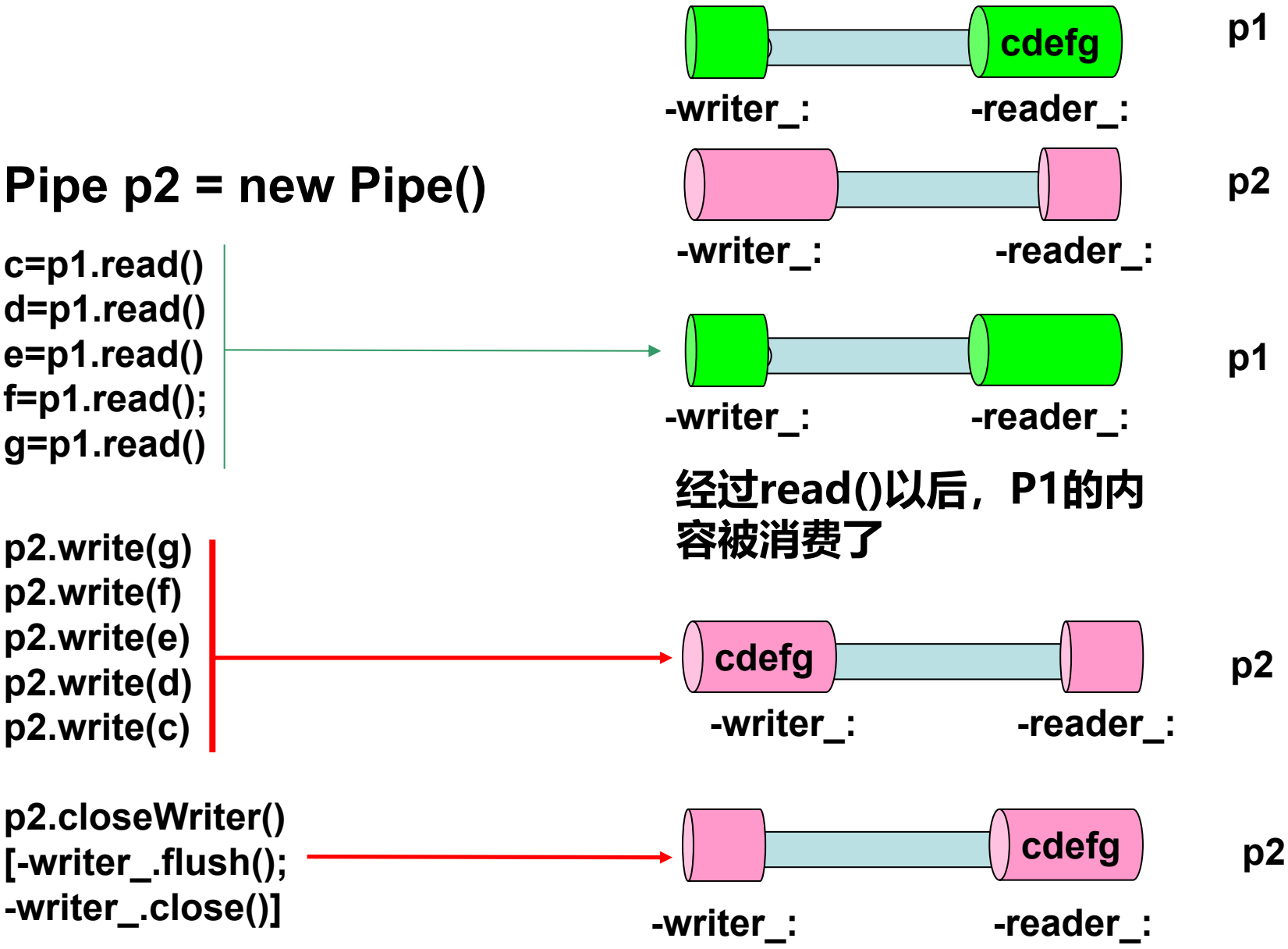


Pipe p1 = new Pipe()





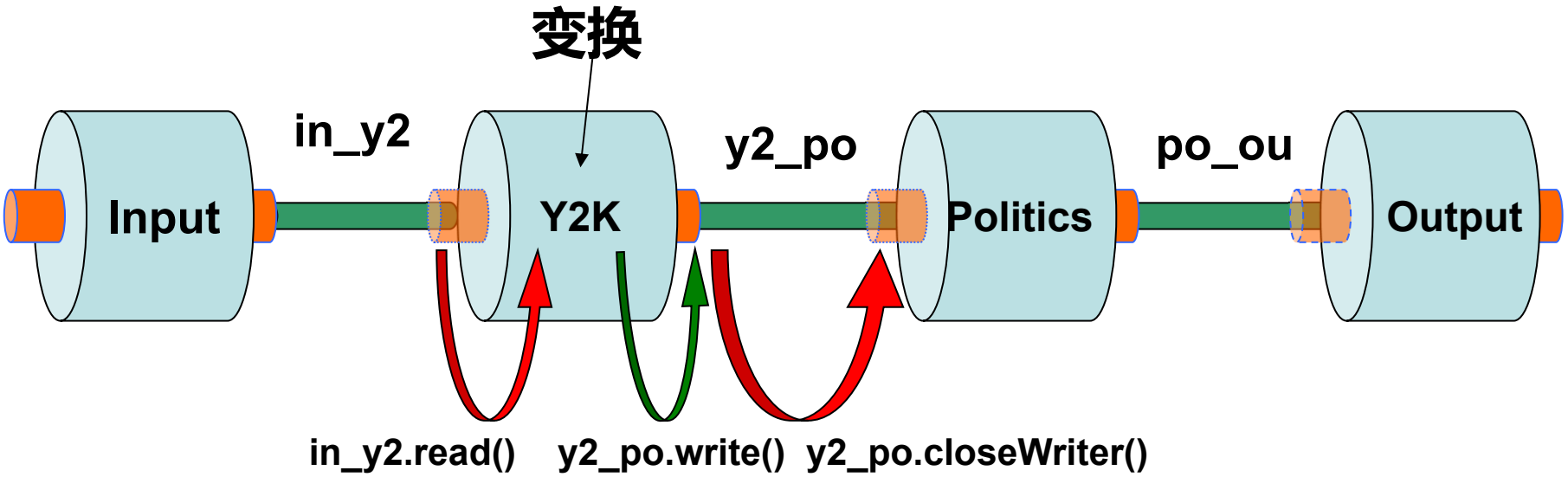
旧文件更新系统 (Legacy File Update)



旧文件更新系统 (Legacy File Update)

How Pipe object works?

- So a filter simply write some data to a pipe object by calling its write method. These data become then available for the next filter, which can call the read method of the pipe object to retrieve the data.



设计要求2得到满足：以这种方式，数据就像水流一样，从左向右不断地流动。

旧文件更新系统 (Legacy File Update)

关于管道(Some issues about a pipe)

a) **数据量. Amount of data in a pipe**: Pipe objects limit the amount of data they can hold.

b) **数据消费. Data consumption.** However, whenever clients read the data from a pipe, this data is considered "consumed" and the space that was occupied by the retrieved data is made free again.

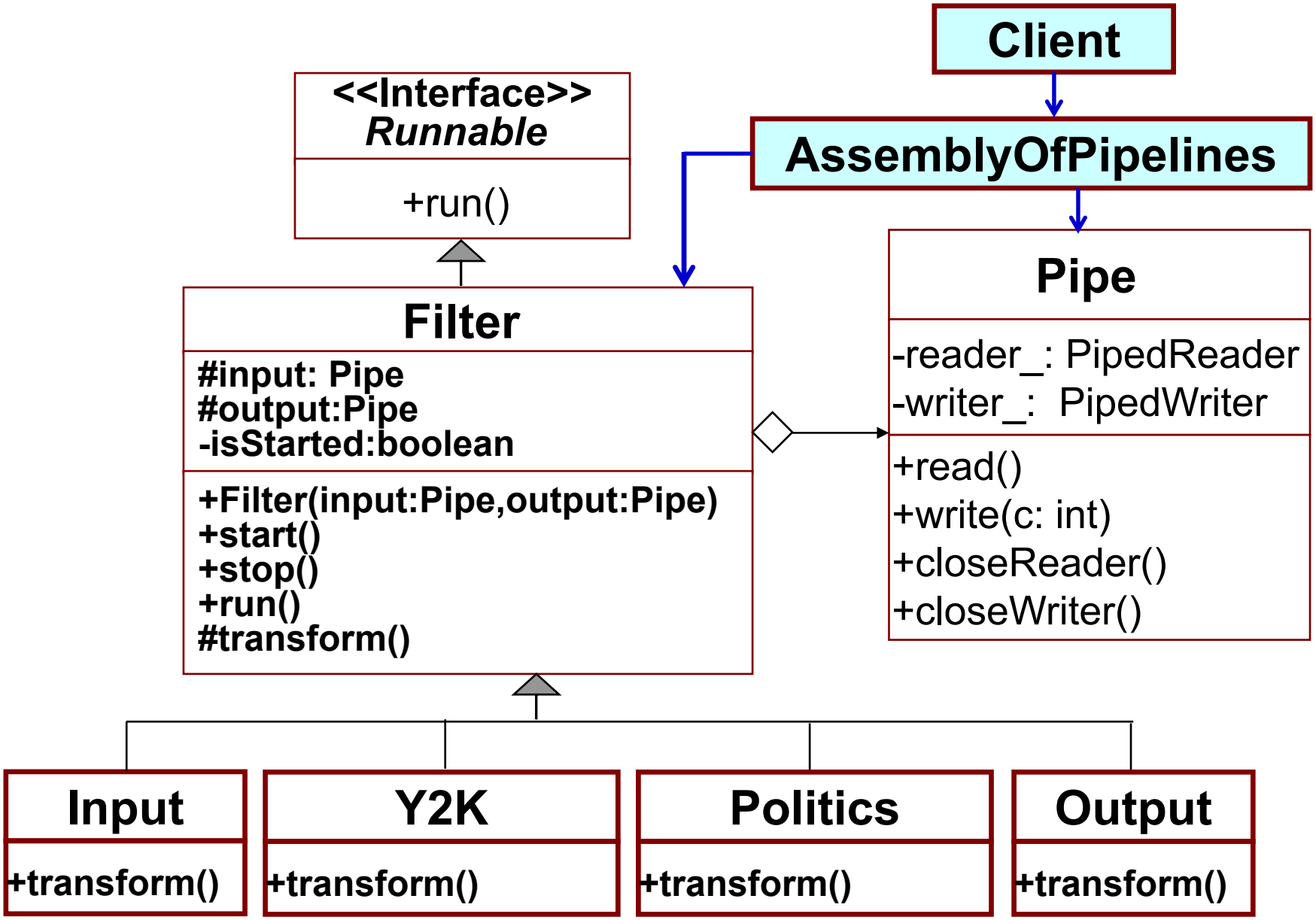
旧文件更新系统 (Legacy File Update)

Some issues about a pipe (cont)

c) **同步性.** Pipe objects are synchronized. There can be only one thread working with a pipe object at a specific moment, i.e, there can be only one thread currently writing or reading the data from a pipe object.

d) **空管道、满管道**. Threads reading from an empty pipe or writing to a full pipe are blocked as long as there are not some data to read from the pipe, or there is no free space in the pipe to write the data.

旧文件更新系统 (Legacy File Update)



旧文件更新系统 (Legacy File Update)

设计优点(Advantages):

- **直观性强(intuitive)**: It maintains the intuitive flow of processing.
- **复用(reuse)性好**: It supports reuse, since each filter can function in isolation
- **容易扩展(Easy evolution)**: New functions are easily added to the system by inserting filters at the appropriate points in the processing sequence.
- **容易修改(Easy modification)**: it’s easy to modify the system since filters are logically independent of other filters.

旧文件更新系统 (Legacy File Update)

Disadvantages:

- **人-机交互差 (Poor Interaction)**: It is virtually impossible to modify the design to support an **interactive** system. For example, deleting a line would require some persistent (**持久的**) shared storage, violating basic principle of this approach.
- **浪费内存 (Waste of space)**: The solution uses space inefficiently, since each filter must copy all of the data to its output ports.

Back