Notes on Software Patterns: Pipes and Filters Architectural Pattern

H. Conrad Cunningham

16 April 2022

Contents

2 Pipes and Filters Architectural Pattern **2**
- 2.1 Chapter Introduction . . . 2
- 2.2 Definition . . . 2
- 2.3 Context . . . 2
- 2.4 Problem . . . 2
- 2.5 Solution . . . 3
- 2.6 Structure . . . 3
- 2.7 Implementation . . . 4
- 2.8 Example . . . 6
- 2.9 Variants . . . 8
- 2.10 Consequences . . . 9
  - 2.10.1 Benefits . . . 9
  - 2.10.2 Liabilities . . . 9
- 2.11 What Next? . . . 10
- 2.12 Acknowledgements . . . 10
- 2.13 Terms and Concepts . . . 11
- 2.14 References . . . 11

Professor of Computer and Information Science
University of Mississippi
214 Weir
P.O. Box 1848
University, MS 38677
(662) 915-7396 (dept. office)

Browser Advisory: The HTML version of this textbook requires a browser that supports the display of MathML. A good choice as of April 2022 is a recent version of Firefox from Mozilla.

2 Pipes and Filters Architectural Pattern

2.1 Chapter Introduction

The goal of this chapter is to present the PIPES AND FILTER architectural design pattern. This presentation follows the presentation in Buschmann et al. [1].

The accompanying set of Powerpoint slides (not fully updated in 2022) is the following:

- Pipe and Filters Architectural Pattern (Powerpoint)

2.2 Definition

Buschmann et al. [1] defines the PIPES AND FILTERS architectural pattern as follows:

> “The *Pipes and Filters* architectural pattern provides a structure for systems that process a stream of data. Each processing step is encapsulated in a filter component. Data *are* passed through pipes between adjacent filters. Recombining filters allows you to build families of related filters.”

2.3 Context

The context consists of programs that must process streams of data.

2.4 Problem

Suppose we need to build a system to solve a problem:

- that must be built by several developers
- that decomposes naturally into several independent processing steps
- for which the requirements are likely to change

The design of the components and their interconnections must consider the following forces [1]:

- It should be possible to enhance the system by substituting new filters for existing ones or by recombining the steps into a different communication structure.
- Components implementing small processing steps are easier to reuse than components implementing large steps.
- If two steps are not adjacent, then they share no information.
- Different sources of input data exist.
- It should be possible to display or store the final results of the computation in various ways.

- If the user stores intermediate results in files, then the likelihood of errors increases and the file system may become cluttered with junk.
- Parallel execution of the steps should be possible.

2.5 Solution

The Solution involves the following steps:

- Divide the task into a sequence of processing steps.
- Let each step be implemented by a filter program that consumes from its input and produces data on its output incrementally.
- Connect the output of one step as the input to the succeeding step by means of a pipe.
- Enable the filters to execute concurrently.
- Connect the input to the sequence to some data source, such as a file.
- Connect the output of the sequence to some data sink, such as a file or display device.

2.6 Structure

The *filters* are the processing units of the pipeline. A filter may enrich, refine, or transform its input data [1].

- It may enrich the data by computing new information from the input data and adding it to the output data stream.
- It may refine the data by concentrating or extracting information from the input data stream and passing only that information to the output stream.
- It may transform the input data to a new form before passing it to the output stream.
- It may, of course, do some combination of enrichment, refinement, and transformation.

A filter may be active (the more common case) or passive.

- An *active* filter runs as a separate process or thread; it actively *pulls* data from the input data stream and *pushes* the transformed data onto the output data stream.
- A *passive* filter is activated by either being called:
  - as a function, a *pull* of the output from the filter
  - as a procedure, a *push* of output data into the filter

The *pipes* are the connectors—between a data source and the first filter, between filters, and between the last filter and a data sink. As needed, a pipe synchronizes the active elements that it connects together.

A *data source* is an entity (e.g., a file or input device) that provides the input data to the system. It may either actively push data down the pipeline or passively supply data when requested, depending upon the situation.

A *data sink* is an entity that gathers data at the end of a pipeline. It may either actively pull data from the last filter element or it may passively respond when requested by the last filter element.

(See the Class-Responsibility-Collaborator (CRC) cards for these elements on page 56 of Buschmann et al. [1])

TODO: Consider whether the above paragraph is needed or can be replaces without depending upon the book completely.

2.7 Implementation

Implementation of the pipes-and-filters architecture is usually not difficult. It often includes the following steps [1]:

1. *Divide the functionality of the problem into a sequence of processing steps.*

   Each step should only depend upon the outputs of the previous step in the sequence. The steps will become the filters in the system.

   In dividing up the functionality, be sure to consider variations or later changes that might be needed—a reordering of the steps or substitution of one processing step for another.

2. *Define the type and format of the data to be passed along each pipe.*

   For example, Unix pipes carry an unstructured sequence of bytes. However, many Unix filters read and write streams of ASCII characters that are structured into lines (with the newline character as the line terminator).

   Another important formatting issue is how the end of the input is marked. A filter might rely upon a system end-of-input condition or it may need to implement their own "sentinel" data value to mark the end.

3. *Determine how to implement each pipe connection.*

   For example, a pipe connecting active filters might be implemented with operating system or programming language runtime facility such as a message queue, a Unix-style pipe, or a synchronized-access bounded buffer.

   A pipe connecting to a passive filter might be implemented as a direct call of the adjacent filter: a push connection as a call of the downstream filter as a procedure or a pull connection as a call of the upstream filter as a function.

4. *Design and implement the filters.*

   The design of a filter is based on the nature of the task to be performed and the natures of the pipes to which it can be connected.

   - An active filter needs to run with its own *thread of control.* It might run as as a "heavyweight" operating system process (i.e., having its own address space) or as a "lightweight" thread (i.e., sharing an address space with other threads).

   - A passive filter does not require a separate thread of control (although it could be implemented with a separate thread).

   The selection of the *size of the buffer* inside a pipe is an important performance tradeoff. Large buffers may use up much available memory but likely will involve less synchronization and context-switching overhead. Small buffers conserve memory at the cost of increased overhead.

   To make filters flexible and, hence, increase their potential reusability, they often will need different *processing options* that can be set when they are initiated. For example, Unix filters often take command line parameters, access environment variables, or read initialization files.

5. *Design for robust handling of errors.*

   Error handling is difficult in a pipes-and-filters system since there is no global state and often multiple asynchronous threads of execution. At the least, a pipes-and-filters system needs mechanisms for detecting and reporting errors. An error should not result in incorrect output or other damage to the data.

   For example, a Unix program can use the `stderr` channel to report errors to its environment.

   More sophisticated pipes-and-filters systems should seek to recover from errors. For example, the system might discard bad input and resynchronize at some well-defined point later in the input data. Alternatively, the system might back up the input to some well-defined point and restart the processing, perhaps using a different processing method for the bad data.

6. *Configure the pipes-and-filters system and initiate the processing.*

   One approach is to use a standardized main program to create, connect, and initiate the needed pipe and filter elements of the pipeline.

   Another approach is to use an end-user tool, such as a command shell or a visual pipeline editor, to create, connect, and initiate the needed pipe and filter elements of the pipeline.

2.8 Example

An example pipes-and-filter system might be a retargetable compiler for a programming language. The system might consist of a pipeline of processing elements similar to the following:

1. A *source* element reads the program text (i.e., source code) from a file (or perhaps a sequence of files) as a stream of characters.

2. A *lexical analyzer* converts the stream of characters into a stream of lexical tokens for the language—keywords, identifier symbols, operator symbols, etc.

3. A *parser* recognizes a sequence of tokens that conforms to the language grammar and translates the sequence to an abstract syntax tree.

4. A *"semantic" analyzer* reads the abstract syntax tree and writes an appropriately augmented abstract syntax tree.

   Note: This element handles context-sensitive syntactic issues such as type checking and type conversion in expressions.

5. A *global optimizer* (usually optionally invoked) reads an augmented syntax tree and outputs one that is equivalent but corresponds to program that is more efficient in space and time resource usage.

   Note: A global optimizer may transform the program by operations such as factoring out common subexpressions and moving statements outside of loops.

6. An *intermediate code generator* translates the augmented syntax tree to a sequence of instructions for a virtual machine.

7. A *local optimizer* converts the sequence of intermediate code (i.e., virtual machine) instructions into a more efficient sequence.

   Note: A local optimizer may transform the program by removing unneeded loads and stores of data.

8. A *backend code generator* translates the sequence of virtual machine instructions into a sequence of instructions for some real machine platform (i.e., for some particular hardware processor augmented by operating system calls and a runtime library).

9. If the previous step generated symbolic assembly code, then an *assembler* is needed to translate the sequence of symbolic instructions into a relocatable binary module.

10. If the previous steps of the pipeline generated a sequence of separate binary modules, then a *linker* might be needed to bind the separate modules with library modules to form a single executable (i.e., object code) module.

11. A *sink* element outputs the resulting binary module into a file.

The pipeline can be reconfigured to support a number of different variations:

- If source code preprocessing is to be supported (e.g., as in C), then a *preprocessor* filter (or filters) can be inserted in front of the lexical analyzer.
- If the language is to be interpreted rather than translated into object code, then the backend code generator (and all components after it in the pipeline) can be replaced by an *interpreter* that implements the virtual machine.
- If the compiler is to be retargeted to a different platform, then a backend code generator (and assembler and linker) for the new platform can be substituted for the old one.
- If the compiler is to be modified to support a different language with the same lexical structure, then only the parser, semantic analyzer, global optimizer, and intermediate code generator need to be replaced.

  Note: If the parser is driven by tables that describe the grammar, then it may be possible to use the same parser with a different table.
- If a load-and-go compiler is desired, the file-output sink can be replaced by a *loader* that loads the executable module into an address space in the computer's main memory and starts the module executing.

Of course, a pure active-filters system as described above for a compiler may not be very efficient or convenient.

- Sometimes a system of filters can be made more efficient by directly sharing a global state. Otherwise the global information must be encoded by one filter, passed along a pipe to an adjacent filter, decoded by that filter, and so forth on downstream.

  In the compiler pipeline, the symbol table is a key component of the global state that is constructed by the lexical analyzer and needed by the phases downstream through (at least) the intermediate code generator.
- Sometimes performance can be improved by combining adjacent active filters into one program and replacing the pipe by an upstream function call (a passive pull connection) or a downstream procedure call (a passive push connection).

  In the compiler pipeline, it may be useful to combine the phases from lexical analysis through intermediate code generation into one program because they share the symbol table. Performance can be further improved by having the parser directly call the lexical analyzer when the next token is needed.
- Although a piece of information may not be required at some step, the availability of that information may be useful. For example, the symbol table information is not usually required during backend code generation,

interpretation, or execution. However, some of the symbol table information, such as variable and procedure names, may be useful in generation of error messages and execution traces or for use by a runtime debugging tools.

2.9 Variants

So far we have focused on single-input single-output filters. A generalization of the pipes-and-filters pattern allows filters with multiple input and/or multiple output pipes to be connected in any directed graph structure.

In general, such dataflow systems are difficult to design so that they compute the desired result and terminate cleanly. However, if we restrict ourselves to directed acyclic graph structures, the problem is considerably simplified.

In the UNIX operating system shell, the `tee` filter provides a mechanism to split a stream into two streams, named pipes provide mechanisms for constructing network connections, and filters with multiple input files/streams provide mechanisms for joining two streams.

TODO: Bring the following up to date?

Consider the following UNIX shell commands. On a Solaris “Unix” machine (late 1990’s), this sequence sets up a pipe to build a sorted list of all words that occur more than once in a file:

```
# create two named pipes
mknod pipeA p
mknod pipeB p
# set up side chain computation (running in the background)
cat pipeA >pipeB &
# set up main pipeline computation
cat filename | tr -cs "[:alpha:]" "[\n*256]" \
             | tr "[:upper:]" "[:lower:]" | sort | tee pipeA | uniq \
             | comm -13 - pipeB | uniq
```

- The `mknod` commands set up two named pipes, `pipeA` and `pipeB`, for connecting to a “side chain” computation.

- The “side chain” command starts a `cat` program running in a *background* fork (note the `&`). The program takes its input from the pipe named `pipeA` and writes its output to the pipe named `pipeB`.

- The main pipeline uses a `cat` filter as a source for the stream. The next two stages use filter `tr` to translate each sequence of non-alphabetic characters to a single newline character and to map all uppercase characters to lowercase, respectively. The words are now in a standard form—in lowercase, one per line.

- The fourth stage of the main pipeline sorts the words into ascending order using the `sort` filter.
- After the sort, the main pipeline uses a `tee` filter to replicate the stream, sending one copy down the main pipeline and another copy onto the side chain via `pipeA`.
- The side chain simply copies the words from `pipeA` onto `pipeB`. Meanwhile the main pipeline uses the `uniq` filter to remove adjacent duplicate words.
- The main pipeline stream and the side chain stream are then joined by the `comm` filter. The `comm` filter takes two inputs, one from main pipeline's stream (note the `-` parameter) and another from `pipeB`.
- Invoking the `comm` filter with the `-13` option cause it to output the lines that appear in the second stream (i.e., `pipeB`) but not the first stream (i.e., the main pipeline). Thus, the output is an alphabetical list of words that appear more than once in the input file.
- The final stage, another `uniq` filter, removes duplicates from the final output.

2.10 Consequences

2.10.1 Benefits

The pipes-and-filters architectural pattern has the following benefits [1]:

- *Intermediate files unnecessary, but possible.* File system clutter is avoided and concurrent execution is made possible.
- *Flexibility by filter exchange.* It is easy to exchange one filter element for another with the same interfaces and functionality.
- *Flexibility by recombination.* It is not difficult to reconfigure a pipeline to include new filters or perhaps to use the same filters in a different sequence.
- *Reuse of filter elements.* The ease of filter recombination encourages filter reuse. Small, active filter elements are normally easy to reuse if the environment makes them easy to connect.
- *Rapid prototyping of pipelines.* Flexibility of exchange and recombination and ease of reuse enables the rapid creation of prototype systems.
- *Efficiency by parallel processing.* Since active filters run in separate processes or threads, pipes-and-filters systems can take advantage of a multiprocessor.

2.10.2 Liabilities

The pipes-and-filters architectural pattern has the following liabilities [1]:

- *Sharing state information is expensive or inflexible.* The information must be encoded, transmitted, and then decoded.

- *Efficiency gain by parallel processing is often an illusion.* The costs of data transfer, synchronization, and context switching may be high. Non-incremental filters, such as the Unix `sort`, can become the bottleneck of a system.

- *Data transformation overhead.* The use of a single data channel between filters often means that much transformation of data must occur, for example, translation of numbers between binary and character formats.

- *Error handling.* It is often difficult to detect errors in pipes-and-filters systems. Recovering from errors is even more difficult.

2.11 What Next?

TODO

2.12 Acknowledgements

I wrote the first version of these notes for my Spring 1998 graduate Special Topics in Software Architecture class. I based the notes, in part, on:

- The Pipes and Filters pattern description in Section 2.2 of the Buschmann et al. (i.e., “Siemens”) book [1]

- The Pipes and Filters pattern description in Section 2.2 of Shaw and Garlan [2]

I revised the notes somewhat for related courses in 1999, 2000, 2001, 2002, and 2004. Also, in 2004 I revised the notes, created slides, and included them as a part of the materials supported by a grant from the Acxiom Corporation titled “The Acxiom Laboratory for Software Architecture and Component Engineering (ALSACE)”. My ALSACE research team included PhD students Yi Liu and Pallavi Tadepalli and MS students Mingxian Fu and “Melody” Hui Xiong.

In Spring 2017, I adapted the earlier notes to use Pandoc-flavored Markdown. In Spring 2018 I revised the notes slightly to fit in with the other documents for the CSci 658 course.

I retired from the full-time faculty in May 2019. As one of my post-retirement projects, I am continuing work on possible textbooks based on the course materials I had developed during my three decades as a faculty member. In January 2022, I began refining the existing content, integrating separately developed materials together, reformatting the documents, constructing a unified bibliography (e.g., using citeproc), and improving my build workflow and use of Pandoc.

TODO: Modify acknowledgements appropriately for 2022+ updates.

I maintain this chapter as text in Pandoc's dialect of Markdown using embedded LaTeX markup for the mathematical formulas and then translate the document to HTML, PDF, and other forms as needed

2.13 Terms and Concepts

TODO

2.14 References

[1] Frank Buschmann, Regine Meunier, Hans Rohnert, Peter Sommerlad, and Michael Stal. 1996. *Pattern-oriented software architecture: A system of patterns.* Wiley, Hoboken, New Jersey, USA.

[2] Mary Shaw and David Garlan. 1996. *Software architecture: Perspectives on an emerging discipline.* Prentice-Hall, Englewood Cliffs, New Jersey, USA.