



INSTITUTO SUPERIOR DE ENGENHARIA DE LISBOA

**Área Departamental de Engenharia de Electrónica e Telecomunicações e de
Computadores**

From asynchronous IO to reactive stream pipelines

Diogo Paulo de Oliveira Rodrigues

Licenciado em Engenharia Informática e de Computadores

Relatório preliminar para obtenção do Grau de Mestre
em Engenharia Informática e de Computadores

Orientador : Doutor Fernando Miguel Gamboa de Carvalho

Março, 2022

Contents

List of Figures	vii
List of Tables	ix
List of Listings	xi
1 Context and Progress status	1
2 State of the Art	3
2.1 Background	3
2.2 Asynchronous flow key concepts and design alternatives	7
2.2.1 Synchronous versus Asynchronous	7
2.2.2 Push vs Pull	8
2.2.3 Hot versus Cold	9
2.2.4 Cancelable	11
2.2.5 Error Handling	12
2.2.6 Intrinsic Keywords	13
2.2.7 Back-pressure	15
2.3 State of the Art	16
2.3.1 ReactiveX.io Project	16
2.3.1.1 RxJava Architecture	17

2.3.2	Project Reactor	18
2.3.2.1	Reactor Architecture	18
2.3.3	Asynchronous Enumerables	19
2.3.3.1	.NET Asynchronous Enumerables	19
2.3.3.2	Javascript asynchronous enumerables	20
2.3.4	Kotlin Flow	22
2.3.5	Technologies comparison	24
3	Benchmarking of Asynchronous Data Processing Algorithms	25
3.1	Strategies	26
3.1.0.1	Reactive Extensions (Rx)	27
3.1.0.2	Reactor (Flux)	27
3.1.0.3	Blocking Reader in Streams	27
3.1.0.4	Multi-threading	27
3.1.0.5	Kotlin Flow	27
3.1.0.6	Concurrency	28
3.1.0.7	Baseline	28
3.2	.NET Implementation	28
3.2.0.1	Approaches Explanation	28
3.2.1	Sync Baseline	28
3.2.2	Async Blocking Read	28
3.2.3	Async Single Task	29
3.2.4	Async Baseline	29
3.2.5	Linq Sync	29
3.2.6	Async Enumerable	29
3.2.7	Rx	29
3.2.7.1	Biggest Word Results	30
3.2.7.2	Group Word Results	31
3.3	Java/Kotlin Implementation	31

3.3.1	Results	32
3.3.1.1	Biggest Word Results	32
3.3.1.2	Group Word Results	33
3.4	JavaScript Implementation	35
3.4.1	Discussion	35
3.5	Conclusion	35

List of Figures

1.1	Planning Gant diagram	2
3.1	Processing times for different strategies for "Find the biggest word".	30
3.2	Processing times for different strategies for "Count Words".	31
3.3	Processing times for different Java/Kotlin strategies for "Biggest Word".	33
3.4	Processing times for different Java/Kotlin strategies for "Group Words".	34
3.5	Processing times for different JavaScript strategies for "Biggest Word" Graphic	35

List of Tables

2.1	Your caption	18
3.1	Processing times for different strategies for "Find the biggest word".	30
3.2	Processing times for different strategies for "Count Words".	31
3.3	Processing times for different Java/Kotlin strategies for "Biggest Word".	32
3.4	Processing times for different Java/Kotlin strategies for "Group Words".	33
3.5	Processing times for different JavaScript strategies for "Biggest Word".	35

List of Listings

2.1	Asynchronous call example	8
2.2	Synchronous call example	8
2.3	Example of Pull data handling patterns	9
2.4	Example of Pull data handling patterns	9
2.5	Cold flow example	10
2.6	Hot flow example	10
2.7	Intrinsic words .NET C#	13
2.8	Javascript Example with promises	14
2.9	Javascript syntax sugar	14
2.10	Asynchronous Enumerable, C# example	20
2.11	Mozilla's Javascript Asynchronous Enumerables example	21
2.12	Flow builder	22
2.13	Kotlin collect multicall example	23
2.14	Kotlin collect multicall example	24

Context and Progress status

This report has the objective of informing the reader, about the current status of the development of the dissertation named *From asynchronous IO to reactive stream pipelines*, that is being written by the student *Diogo Paulo de Oliveira Rodrigues*.

Firstly, to understand the current state of development, it is important to make clear what will be the objective and structure of the document that is being written.

This dissertation aims, in the first place, to make an overview on what tools were historically available to deal with non-blocking IO operations and asynchronous programming in general, and then, indentifying what are the state-of-the-art solutions available today to deal with this problem that are usually supported through asynchronous data pipeline architectures. After presenting state-of-the-art concepts and technologies, some of these technologies will be then compared in terms of performance through the development of practical benchmark test cases, written in several programming languages widely used today.

Taking in account what was explaining above, the dissertation will aim to have the following main structure:

1. Introduction and motivation
2. Concepts and State-of-the-art
3. Test case development and Benchmark results

4. Conclusions

Since this report is preliminary, this structure can be slightly changed in the course of the development.

To fulfill the objectives in a desirable timeline, were created several tasks scheduled to be done in a pre-determined time interval. This task can be viewed in the next Gant diagram, where it is possible to see the completion status of each task as they were in 15 of March:

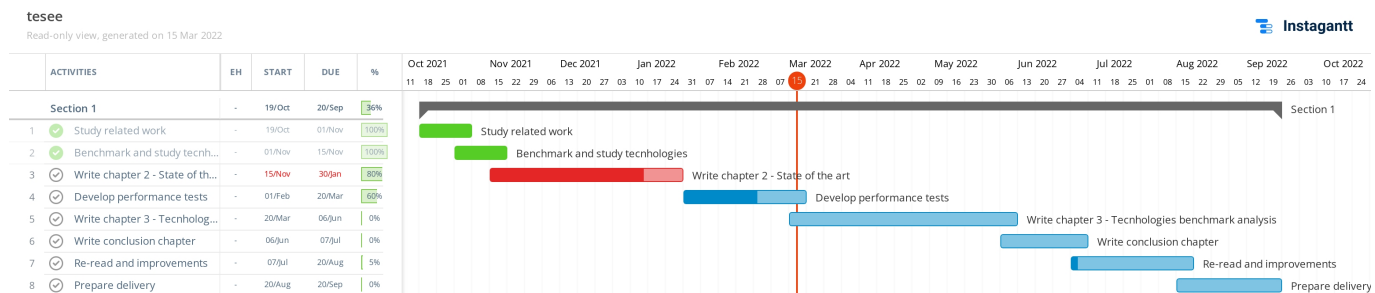


Figure 1.1: Planning Gant diagram

As we can see in the Gant diagram, 80% of the chapter 2 is completed, however, are remaining some adjustments in some examples and to be more clear in the explanation of others, these adjustments and a complete review will be completed until the final of March.

On the other side, the development of the testing solutions to benchmark the different asynchronous pipeline technologies are under course. The finish date to this task is expected to be 20 March, where is expected to begin the writing of the chapter 3 of the thesis.

On the next chapter of this report, is written the chapter 2 of the thesis "*State of the Art*" as it was on the day of 15 of March 2022.

2

State of the Art

This section, aims to show the background and the current state of the art on the subject related with asynchronous pipelining operations involving asynchronous data flows. On the section 2.1, will be made an overview on previously developed work made on this subject, then, on section 2.2, will be made a characterization of the key concepts related to it. By last, on section 2.3, are presented and explained several technologies representative of the state of the Art on how asynchronous operation pipelining are implemented in different programming realities, e.g. on Kotlin, JAVA and C#.

2.1 Background

From the end of 80's to the beginning of the 2000's, with the acceleration of Moore's Law in hardware and network bandwidth development; the creation of the web as we know today through the wide spread of use of the HTTP protocol, and the support from new operative systems to multithreading support; the necessity of high responsiveness servers started to grow. This increase in demand of new ways to handle data through parallelism involving asynchronous flow processing, caused the necessity of design new programming models.

Taking the wave initiated by the Gang of Four in [1], where 23 patterns were compiled to deal with object-oriented problems; a group of researchers wrote a paper [3], where they identify the characteristics that high available server must

have, and how the use of asynchronous IO pipe lining design patterns can help to achieve these characteristics, as described bellow:

- Concurrency - The server must process multiple client requests simultaneously.
- Efficiency - The software design must be built aiming the use the least hardware resources as possible.
- Simplicity - The code of the solution must be easy to understand, modular and avoid own built design patterns as possible.
- Adaptability - The system must be totally decoupled from client implementations, allowing it to be easily used by any client independently of the underlying technologic realities. To achieve this, may be used standards e.g. [5] or SOAP.

To achieve some of these properties, the authors propose the *Proactor Pattern* [3]. In their opinion, traditional concurrency models, until that date, failed to fully achieve the enumerated properties.

On the other side, before presenting the *Proactor Pattern*, are identified two major existing non-blocking models, namely: *multithreading* and *reactive event dispatching*.

On multithreading, the paper refers that one of the most direct solution of this approach, is the handling of multiple requests by creating a new thread every request. Each request will then be fully processed and the recently created thread is then be disposed after the work is done.

This solution has several serious issues. Firstly, creating a new thread per request is highly costly in terms of computational resources, because are involved context switches between user and kernel modes; secondly, must be taken in account synchronization to maintain data integrity. Then, the authors warn about the fact that the IO retrieved data is mainly memory-mapped, which rises the question: What happens when the data obtained through IO becomes greater than the system memory can hold? The system stalls until more memory becomes available!?

One possible solution is asynchronous pipelines that will be discussed ahead in this dissertation. On last, if the server receives a high demand of requests, the server easily blocks in the process of creating and disposing threads.

To avoid this issue, the authors, recommended the use of dynamic thread pools to process requests, where each request will be linked to a pre-existing thread, avoiding all the overhead of creating and disposing a thread per request; however, issues related with memory-mapping and overhead due to the switching of data between different threads maintains.

Another traditional concurrency model identified by the authors of the paper, is the *Reactive Synchronous Event Dispatching*. In this model, a *Dispatcher*, with a single thread in a loop, is constantly listening requests from clients and sending work requests to an entity named *Handler*. The *Handler*, will then process the IO work Synchronously and request a connection to the client in the *Dispatcher*. When the requested connection is ready to be used, the *Dispatcher* notifies the *Handler*. After the notification, the *Handler* asynchronously sends the data, that is being or has been obtained through IO, to the client.

Although the authors identifying that this approach is positive, because decouples the application logic from the dispatching mechanisms, besides with the low overhead due the use of a single thread, the authors identify several drawbacks with this approach. Firstly, since IO operation are synchronous, the code for this approach is complex because must be set in place mechanisms to avoid IO blocking through hand off mechanisms. Then, if a request processing blocks, the processing of another requests may be impacted.

To keep the positive points but mitigating the identified issues of previous approaches, is suggested the *Proactor Pattern*. This pattern is very similar to the *Reactive Synchronous Event Dispatching*, however, after the requests processed by a single threaded *Completion dispatcher*, the IO work is then dispatched asynchronously to the underlying OS IO subsystems, where multiple requests can be processed simultaneously. For the result to be retrieved, is previously registered a callback in the *Completion Dispatcher* and the OS has the responsibility to queue the finished result in a well known place.

Finally, the *Completion Dispatcher* has the responsibility to dequeue the result placed by the OS and call the correct previously registered callback. With this, this model creates a platform that provides: decoupling between application and processing mechanisms, offers concurrent work without the issues inherent with the use of threading mechanisms and since IO is managed by the OS subsystems,

is avoided code complexity in handling possible blocking and scheduling issues.

The *Proactor Pattern*, creates the ground for several models used by modern platforms that use a reduced number of threads to process client requests and parallel mechanisms to do the heavy work in the background. Some of these technologies, are, for example: *Javascript NODE.JS*, *Spring Webflux*, *vertx* and others.

From what was explained until now, is evident the tendency followed by software architects in terms of asynchronous processing from non-reactive to event driven approaches and trying to make the code simpler and easier to maintain. Initially the systems were non-reactive, where each request had to be processed in a specific thread and that thread blocked until something got ready to go further. The code was usually very complex and hard to maintain. Then, with the asynchronous systems based on events with the introduction of callback systems inspired in patterns like the *Reactor* or *Proactor* the software design started to become more event driven, allowing the servers to be more efficient in responsiveness, resources optimization and code easier to read and maintain by the average programmer.

However, are some limitations in these asynchronous models. For example, if the data to be processed is bigger than the memory available at a given moment or if the data to be calculated is from a source that produces data at a constant rate that must be processed in real time, these models work badly. The traditional models fail to comply these objectives because are mostly eager by design or not comply with the notion of a continuous source of information that requires to be processed in real time. Taken this in account, projects like project Reactor, Asynchronous Enumerable provided by Microsoft or papers like [2], try to deal with these issues, by providing API's that merge the concepts of Fluent API's, functional programming and code syntax that tries to resemble synchronous code, being the complexity inherent with asynchronous models implementations hidden from the programmer since the asynchronous events are treated like, for example, a `Stream` in JAVA would be treated.

2.2 Asynchronous flow key concepts and design alternatives

With the development of several approaches related to asynchronous IO processing, and with the growing necessity to build a name set of properties that simplify the description of asynchronous systems, a dictionary of properties, concepts and design alternatives started to grow by itself. In the following, are discussed several of the concepts related with asynchronous data flow, namely:

2.2.1 Synchronous versus Asynchronous

Before explaining more terms related with asynchronous data flow, it's important to clarify what is synchronous and asynchronous in programming.

Asynchronous in programming, is a call to a function or routine that returns immediately, not blocking the caller until the operation is finished. The operation processing, will be completely independent from the caller execution process and can even be done in another machine. This way, the caller is freed to do more work, even to start N more operations in parallel.

Meanwhile, a call to a synchronous function or routine, blocks the caller until the operation finishes. In this case, the caller has to wait for the completion of the synchronous operation before going forward, which limits the program efficiency if parallelism is applicable. To better visualize what was explained, we have the following examples:

```

1  HttpClient client = HttpClient.
    newHttpClient();
2
3  HttpRequest request = HttpRequest
4      .newBuilder(URI.create("SOME MOVIE API
5      "))....
6  Task futureResponse = client
7      .sendAsync(request, new
8      JsonBodyHandler<>(DTO.class))
9      .thenAccept(res -> {
10         Console.WriteLine( res.title);
11     });
12 Console.WriteLine("prints something")
13
14 client.Wait();
15
16 //OUTPUT:
17 //prints something
18 //movie title
19

```

Listing 2.1: Asynchronous call example

```

1  HttpClient client = HttpClient.
    newHttpClient();
2
3  HttpRequest request = HttpRequest.
4      newBuilder(
5      URI.create("SOME URL"))
6      .header("accept", "application/json")
7      .build();
8  HttpResponseMessage<Supplier<DTO>> res =
9      client.send(request, new
10         JsonBodyHandler<>(DTO.class)) //
11         synchronous
12
13 Console.WriteLine(res.title);
14
15 Console.WriteLine("prints something")
16
17 //OUTPUT:
18 //movie title
19 //prints something

```

Listing 2.2: Synchronous call example

As we can see, in the synchronous call example, the operation return only happens after the whole subsequent remote operation is finished, consequently, the caller operation is dependent from several variables to go forward e.g. : HTTP messaging latency, remote server operation speed or bandwidth issues. This causes that the processing of the next code statements to happen only after the synchronous IO call.

Meanwhile, in the asynchronous operation call, the return happens immediately after the call, however, the processing inherent with that operation will start just when the subsystem that handles the asynchronous function is ready to process that work. For example, when the OS is ready to process the received responses from a remote server that handled the operation. In this case, the statement right after the asynchronous call is processed before the asynchronous operation. Allowing the IO asynchronous operation to be processed outside the program scope and avoiding any block of the main program.

2.2.2 Push vs Pull

Another concept important to understand how asynchronous data flow is handled in programming, is the *Pull* and *Push* processing patterns. In *Pull* pattern,

usually, exists a source of data and the program iterates over that source to operate over each item.

On the other hand, in the *Push* pattern, the items of the data source are "Pushed" to a routine that will operate over that item. To help to assimilate what was just explained, we have the following example:

```

1 Flowable<Long> flow = Flowable
2   .interval(1, TimeUnit.SECONDS);
3
4 Iterator<Long> iterator =
5   flow.blockingIterable().iterator();
6
7 while (iterator.hasNext())
8   System.out.println(iterator.next());
9
10
```

Listing 2.3: Example of Pull data handling patterns

```

1 Flowable<Long> flow = Flowable
2   .interval(1, TimeUnit.SECONDS);
3
4 flow.blockingSubscribe(System.out::println
5   );
6
```

Listing 2.4: Example of Pull data handling patterns

As we can see, in the pull pattern, the items are "pulled" from a data source through an iteration mechanism.

In contrast with that, in the *Push* pattern, items are pushed to a consumer through a supplier.

2.2.3 Hot versus Cold

Another property that must be taken in account when handling with *Reactive Streams* or asynchronous data processing in general, is the nature of the data flow. There are two main adjectives to name a data flow, `Hot` or `Cold`.

A `Cold` data flow, is a flow of information that is produced just when the stream pipeline is subscribed by an observer. In this case, the producer only starts sending/producing data when someone is interested in the data from that source. For example, when program uses a IO mechanism to lazily retrieve a sequence of words from a database, the IO mechanism will only start sending information just when a consumer subscribes that data flow. Usually, the data is sent to the consumer in unicast.

On the other hand, in a `Hot` data flow, the data is produced independently of existing any observer to that information. This mechanism usually work in broadcast and the data is continuously produced and sent to possible observers. In this

case, when an observer subscribes to a publisher, exists the possibility of data items being already lost to that publisher while in the `Cold` flow, the consumer usually receives all items that were produced by the source. In the following examples, a number is produced each 100 milliseconds:

```

1  Flowable<Long> cold = Flowable.interval
    (100, TimeUnit.MILLISECONDS);
2  Thread.sleep(1000);
3  cold.blockingSubscribe(System.out::println
    );
4  //Output:
5  //1
6  //2
7  //3
8
1 ConnectableFlowable<Long> hot = Flowable
2   .intervalRange(0, 100, 0, 100,
3   TimeUnit.MILLISECONDS)
4   .publish();
5 flux.connect();
6 Thread.sleep(1000);
7 flux.blockingSubscribe(System.out::
8   println
9
10 //output:
11 //2
12 //3
13 //4
14 //

```

Listing 2.6: Hot flow example

Listing 2.5: Cold flow example

As we can see, in the `Hot` data flow example, the items are emitted from the moment the producer is created, independently of existing any subscriber or observer attached to that publisher. Notice that when a consumer is subscribed to the publisher, 1 seconds after the emission started, the numbers from 0 to 10 were not printed.

In the `Cold` example, the producer only emits data when a subscription is done, and because of that, all the produced numbers were printed, in contrast with what happen in `Hot` stream, where data loss are almost certain.

2.2.4 Cancelable

As already stated above, asynchronous operations may run outside the main program scope. This implies, that the main program loses visibility and control on what happens in the asynchronous operation contexts. Because of that, exists the need to put in place mechanisms of control that allow the main program to maintain control over an asynchronous operation to, for example, cancel the operation or put in place finishing logic that allow, for example: resource disposing, logging, decisions etc...

These mechanisms, are many times done through the concept of *cancelables*. Usually, a cancelable, is an entity that represents an operation that can canceled from an external entity, or, an entity that allows to set logic when an operation finishes by any reason.

In C#, a cancelable is an interface implemented by objects that represent asynchronous operations and provides the means to cancel asynchronous operations, on the fly. This is achieved through a mechanism named: `CancellationToken`, that is used to pass information through different execution threads.

In RxJava, a `Cancelable`, is a functional interface with the method `cancel()`. Then, the `Cancelable` can be associated to a data source representation, the `Observable`, by calling the method `Observable.setCancelable(Cancelable)`. When the `Observable` finishes or is canceled for any reason, the method `Cancelable.cancel()` will be called. This way, proper logic is put in place to handle an asynchronous operation cancelation.

As we saw, these two concepts of cancelable diverge. One, provides the means to cancel an operation on the fly and gives some control over the operation cancelation; the other, provides the means to control an operation cancelation independently of how it was cancelled.

2.2.5 Error Handling

In synchronous environments, usually, when something goes wrong, the way to handle an error in the majority of cases is by throwing an exception and propagate it until the proper code handles it, usually in a try/catch block. However, in cases when exists an asynchronous operation or when a continuous stream of data items are being received, that way of dealing with an error can imply several issues, e.g: exceptions not reaching main program, log losing or asynchronous flow blockage.

Since log losing or blocking a whole operation because of a badly handled error is unacceptable, the best way to deal with errors in asynchronous data flow it is to isolate the error. This way, the flow processing may continue in parallel while the error is properly handled.

The best way to handle this kind of errors, it is to have proper callbacks that are called when an error occurs on the stream item. This way, a function can handle the error properly, without the necessity to blocking any data stream processing, if avoidable and the proper logging and any additional measure to handle it can be put in place.

2.2.6 Intrinsic Keywords

As already stated, asynchronous code is tendentially harder to understand because, in opposition with what happens in synchronous environments, the operations inherent with the sequence of programming statements operations may not happen chronologically ordered. Because of that, many times it is difficult read, debug and sustain asynchronous software.

For that reason, many languages started to add syntax techniques that allow the programmer to build asynchronous code that resembles the synchronous syntax. Under the hoods, the virtual machines that sustain these syntax mechanisms, handle the code bounded with that 'intrinsic words' and builds asynchronous routines that the programmer will not be aware of; being this a way to abstract the programmer from the complexity of handling and sustaining complex asynchronous code.

One example of *intrinsic words* mechanisms, is the `async...await` keywords implemented in *Microsoft's .NET C#* and in *javascript*.

In the next example we can see an example of these keywords being used:

```
1  static async Task Main(string[] args)
2  {
3      IEnumerable<int> enumerable = FetchItems(1000);
4      int i = 0;
5      await foreach (var item in enumerable)
6      {
7          if (i++ == 10){ break;}
8      }
9  }
10
11 static async IEnumerable<int> FetchItems(int delay)
12 {
13     int count = 0;
14     while(true)
15     {
16         await Task.Delay(delay);
17         yield return count++;
18     }
19 }
20
```

Listing 2.7: Intrinsic words .NET C#

Where, for example, we can observe in the line 16, a call to an asynchronous operation, and, by adding the keyword `await`, the next statement although

being a call to an asynchronous operation, the code statement order looks like it is from synchronous set of instructions.

Additionally, the use of `async...await` in .NET, for example, simplifies error handling in asynchronous code. Instead of use a callback to handle an error, by using `async...await` the error can be handled by just using a simple try/catch block.

To better visualize the advantage of using intrinsic words in asynchronous code, on the next example, we can see a code comparison in ECM6 Javascript, with and without the use of intrinsic words in asynchronous code. In the next example, it is possible to observe a decrease of code complexity and increment of readability where is used the "syntax sugar" provided by the 'yield' return.

The example using promises is purposefully made with a "Pyramid of Doom" code to accentuate a difficulty of reading asynchronous if is made without any concern with readability.

<pre> 1 function ourImportantFunction(callback) 2 { 3 task1(function(val1) { 4 task2(val1, function(val2) { 5 task3(val2, callback); 6 }); 7 }); 8 }</pre>	<pre> 1 function ourImportantFunction() { 2 3 var val1 = yield task1(); 4 5 var val2 = yield task2(val1); 6 7 var val3 = yield task3(val2); 8 9 return val3; 10 };</pre>
--	---

Listing 2.9: Javascript syntax sugar

Listing 2.8: Javascript Example with promises

As we can see, with the use of `yield` keyword, the code that uses a result of several asynchronous operation, instead of being used in a "Matrioska Dool" type of code, with a code made with a chain of callback results; the simple use of a intrinsic keyword like `yield` simplifies the code a lot. Making the code previously hard to read in a easier code to understand and maintain.

2.2.7 Back-pressure

When the *pull* method is used to retrieve items from a source, the producer retrieves only the items it can process in the given time.

However, when the *push* approach is used as data retrieval method from asynchronous flows, the producers have the initiative to push items to its consumers. This can originate situations, where the producer emits items faster than the producers can handle, which can create problems like: unwanted loss of data, lack of responsiveness from consumers, etc...

To resolve these issues, were created strategies and design patterns that are commonly referred as *Backpressure*. There are four main approaches which the majority of *Backpressure* strategies are designed from and can be resumed as:

1. **DROP:** Producer drops items after a retrieving buffer gets full.
2. **Buffer everything:** A buffer, keeps all unprocessed items that are received. Usually, this strategy is used when all received items are critical for the business development and memory management has flexibility to handle the increase of storage needs.
3. **Error:** An error is thrown when the buffer threshold is reached, usually all items received after the threshold is reached are discarded.
4. **Lastest:** Only the last received item in the given moment is kept.
5. **Missing:** No back-pressure strategy it is in place, all items that can not be processed on arrival, are discarded.

2.3 State of the Art

In this section, we will elaborate on various state-of-the-art frameworks used for asynchronous IO pipeline processing across several programming languages.

Initially, in Section 2.3.1, we present the multi-language *ReactiveX.io* project and explain how the *Observer Pattern* is utilized in this approach to implement real-time asynchronous processing with and without back-pressure. In Section 2.3.1.1, we discuss *RxJava*, which is an implementation of this project in Java.

In the following section, we focus on *Project Reactor*, a pivotal component of Spring WebFlux's non-blocking web stack. We delve into how *Project Reactor* takes a different approach from *ReactiveX* but still offers similar benefits, enabling developers to work with a composable API for declarative, event-driven programming. This discussion can be found in Section 2.3.2.

Next, we introduce the *Async Enumerables* in this Section 2.3.3.1, is discussed how these approaches are implemented in .NET C# and in javascript and how it diverges from *ReactiveX.io* or Kotlin's *Kotlin Flow*. Sections 2.3.3 and 2.3.4 will present the JavaScript and Kotlin Flow strategies for asynchronous data processing, respectively.

Finally, in Section 2.3.5, we provide an overview and draw conclusions about the various technologies presented, discussing how each approach can be employed for different problems and objectives. We then make theoretical predictions on how each technology behaves in several known circumstances.

2.3.1 ReactiveX.io Project

Following the historical context explained in Chapter 2.1, various developer groups have pursued implementing solutions for handling asynchronous data flows. One significant outcome of these endeavors is the *ReactiveX* project.

This project amalgamates several concepts and recognized design patterns to implement an asynchronous data flow mechanism. Notable advantages of this project include its multi-language implementation with an open-source approach, and a design that abstracts the complexity inherent to asynchronous programming from the developer, thus yielding more intuitive and easily understandable code. In this dissertation, we will specifically examine the Java implementation of *ReactiveX*, namely *RxJava*.

2.3.1.1 RxJava Architecture

To implement an asynchronous mechanism capable of managing *Hot* and *Cold* data flows with and without back-pressure, the project has adopted the *Observer pattern* combined with the concepts encapsulated in the *Publisher/Subscriber pattern*.

The designers justify this choice by arguing that this model enables treating streams of asynchronous events with the same ease as their synchronous counterparts, such as *Collection* and *Iterator*. They further elaborate that using this approach, the code complexity associated with low-level threading, non-blocking IO, thread-safety, and so forth, is effectively abstracted away from the developer.

In the implementation, the data source/publisher of an asynchronous stream of events is represented by the classes `Observable` and `Flowable`. The only difference between `Observable` and `Flowable` lies in that `Flowable` supports back-pressure through different buffering strategies.

It is worth mentioning that in RxJava, while `Observable` and `Flowable` represent a stream of *N* events, a single event is represented by a class named `Single`.

One of the significant advantages of using this framework is that `Observable` and `Flowable` offer fluent APIs that allow operation chaining in pipelines, similar to the provision in synchronous environments like Java's *Stream* fluent API or .NET's Linq framework. These fluent APIs support stream processing operations like `filter`, `flatMap`, and `distinct`, enabling developers to treat these asynchronous event sequences as they would with synchronous streams in fluent APIs.

The observers/subscribers are represented by consumers that are attached to `Observable` via the method `Observable.subscribe()`. These consumers can either be implementations of the functional interface `Consumer<T>` or the interface *Observer*.

The *Observer* interface offers extra control over stream processing through error handling, a feature that implementations of `Consumer<T>` lack. The *Observer* can be viewed as the asynchronous representation of the Java util interface *Iterator*, with the resemblance evident in the *Observer* interface methods:

1. `onSubscription()`: This method is called immediately after a subscription is made with an `Observable`.

2. `onNext(T item)`: This method is called when an item is emitted by the asynchronous data source.
3. `onError()`: Contrary to the Iterator, the RxJava *Observer* is equipped to support error handling. This callback is invoked when an error occurs.
4. `onComplete()`: This method is called when the data source closes or the subscription finishes.

Given the possible relationship between synchronous and its asynchronous logic counterparts in many languages, we will take Java as an example to demonstrate how these objects can be directly related to their asynchronous counterparts in RxJava. The following table shows how each object type can be mapped:

	Single Item	Multiple Items
Java	<code>T getData()</code>	<code>Iterable<T> getData()</code>
RxJava	<code>Single<T> getData()</code>	<code>Observable<T> getData()</code>

Table 2.1: Your caption

2.3.2 Project Reactor

Similar to the development of *RxJava*, described in Chapter 2.3.1.1, the increasing demand for non-blocking reactive systems in the Java ecosystem has led to the creation of *Project Reactor*.

Project Reactor, as a part of the Spring portfolio, provides a fully non-blocking foundation that applies the Reactive Streams specification for Java. By doing this, Reactor offers efficient demand management (back-pressure ready) capabilities, which make it a compelling choice for scenarios that involve handling live streams of data.

2.3.2.1 Reactor Architecture

Reactor's design is based on the *Publisher/Subscriber pattern*, just like *RxJava*, enabling the use of *Flux* (a stream of 0 to N items) and *Mono* (a stream of 0 or 1 item) as representations of asynchronous data streams. The choice to include these representations is because they adhere to the Reactive Streams specification, which makes them interoperable with other implementations following the same specification.

The benefits of using Project Reactor are analogous to the ones seen with *RxJava*, including the abstraction of low-level threading, non-blocking IO, and thread-safety details from the developer. Moreover, it provides a range of operators that can transform, filter, combine, and manipulate data streams. This enables programmers to construct intuitive instruction pipelines, similar to those seen in Java's Stream API, C#'s LINQ, or Rx.NET.

In Project Reactor, developers attach subscribers to publishers using the `subscribe()` method, similar to what is done in *RxJava*. Here, the subscribers can implement error handling, which allows more precise control over stream processing.

Here's a brief comparison of how Reactor relates to the conventional synchronous counterparts in Java:

	Single Item	Multiple Items
Java	<code>T getData()</code>	<code>Iterable<T> getData()</code>
Reactor	<code>Mono<T> getData()</code>	<code>Flux<T> getData()</code>

By adhering to the Reactive Streams specification and offering a wide array of operators to handle data, Project Reactor serves as a powerful tool for developing reactive, non-blocking applications in Java.

2.3.3 Asynchronous Enumerables

Asynchronous enumerables bridge the gap between traditional enumerables, which represent collections of data, and asynchronous programming, where data often arrives over time rather than being available instantly. This concept is crucial for efficiently managing data streams from real-time sources.

In this section, we explore the implementation of asynchronous enumerables in two programming environments: .NET (specifically C#) and JavaScript.

2.3.3.1 .NET Asynchronous Enumerables

In the .NET framework, the concept of an enumerable is represented through the `IEnumerable<T>` interface, which defines a method `GetEnumerator()`. This

method returns an `IEnumerator<T>`, enabling iteration over a collection. To extend this concept to the asynchronous world, .NET introduces the `IAsyncEnumerable<T>` interface.

Similar to its synchronous counterpart, `IAsyncEnumerable<T>` returns an `IAsyncEnumerator<T>` but with an asynchronous `MoveNextAsync()` method. This minor but significant modification lets us deal with data sources where data availability is asynchronous, such as real-time feeds, network streams, etc.

Here's a simple example of how asynchronous enumerables can be used in C#:

```
1  static async Task Main(string[] args)
2  {
3      IAsyncEnumerable<int> enumerable = FetchItems(1000);
4      int i = 0;
5      await foreach (int item in enumerable)
6      {
7          if (i++ == 10){ break;}
8          Console.WriteLine(item);
9      }
10 }
11
12 static async IAsyncEnumerable<int> FetchItems(int delay)
13 {
14     int count = 0;
15     while(true)
16     {
17         await Task.Delay(delay);
18         yield return count++;
19     }
20 }
21 //
22 //1
23 //1 sec delay
24 //2
25 //1 sec delay
26 //3
27 //....
28
```

Listing 2.10: Asynchronous Enumerable, C# example

This example demonstrates how asynchronous data sources can be worked with in a similar way as synchronous collections, thanks to the use of `IAsyncEnumerable<T>` and the `await foreach` construct.

2.3.3.2 Javascript asynchronous enumerables

Node.js tries to provide a functional and weakly typed approach to asynchronous flow processing, opposed to what we saw in the C# solution. The mechanism to

provide the means to iterate over asynchronous streams is relatively similar to what we saw in C#, with the use of the intrinsic keywords `async...await` and `for await...of`. With the keyword pair `async...await`, it's provided the syntax "sugar" that enables making asynchronous calls and allowing the syntax to resemble synchronous code. The `async` mark is used in an asynchronous function and enables the use of the keyword `await`. The `await` keyword is then used in an asynchronous call to another asynchronous function. On the other side, the keywords `for await...of` provide support for iteration over asynchronous enumerables, as we can see in the next example:

```
1  async function* streamAsyncIterablestream {
2    const reader = stream.getReader;
3    try {
4      while true {
5        const { done, value } = await reader.read;
6        if done {
7          return;
8        }
9        yield value;
10       }
11     } finally {
12       reader.releaseLock;
13     }
14   }
15
16   async function getResponseSizeurl {
17     const response = await fetchurl;
18     let responseSize = 0;
19
20     const iterable = streamAsyncIterableresponse.body;
21
22     for await const chunk of iterable {
23       responseSize += chunk.length;
24     }
25     return responseSize;
26   }
27
```

Listing 2.11: Mozilla's Javascript Asynchronous Enumerables example

This example demonstrates how JavaScript's `async function*` construct can be used to define asynchronous enumerables. These can then be conveniently iterated over using the `for await...of` construct, just like you would with regular collections.

Overall, both C# and JavaScript provide convenient ways to handle asynchronous data flows that simplify coding and make the code more readable and maintainable. These tools help abstract away the complexities of asynchronous programming, such as callbacks, tasks, and completion tokens, making it easier to write

and understand asynchronous code.

2.3.4 Kotlin Flow

Kotlin asynchronous flow implementation is a [4] compliant solution, inspired in RxReactor and another projects, made with the intent to be simple as possible. Similarly to RxJava, Kotlin strongly typifies a stream of N events, which, in this case, is done through the implementation of the interface `Flow<T>`.

A `Flow<T>` data source implementation is done through a builder, like we can see in the next example:

```

1  fun simple(): Flow<Int> = flow { // flow builder
2      for (i in 1..3) {
3          delay(100) // pretend we are doing something useful here
4          emit(i) // emit next value
5      }
6  }
7
8  fun main() = runBlocking<Unit> {
9      launch {
10         for (k in 1..3) {
11             println("I'm not blocked $k")
12             delay(100)
13         }
14     }
15     simple().collect { value -> println(value) }
16 }
17
18 //output:
19 //I'm not blocked 1
20 //1
21 //I'm not blocked 2
22 //2
23 //I'm not blocked 3
24 //3
25
26
```

Listing 2.12: Flow builder

A consumer, to start listening a particular data flow has to call the method `Flow.collect()`

. Since `Flow` provides support only to cold flows, calling `collect()` has the particularity of initiating flow the sequence. On the other side, Hot flows in Kotlin are represented by the interface `SharedFlow<out T> : Flow<T>`.

The main difference between the implementation of these two interfaces, is at the result of `collect()` call. While the `Flow.collect()` starts the flow every

time its called, resulting in the limited emission of the same set of items per call; the `SharedFlow.collect()` emits an unpredicted set of items from external stream of events initiated before the call to `SharedFlow.collect()`. On the other side, while the `Flow.collect()` call context is private to the caller, the `ShareFlow.collect()` is shareable by N subscribers, which makes this solution ideal for broadcast mechanisms shared by many users. On the next example, we can see several calls to the `Flow.collect()` that results in retrieving the same set of items; as explained, the call starts a cold flow every time its called:

```

1 fun simple(): Flow<Int> = flow {
2     println("Flow started")
3     for (i in 1..3) {
4         delay(100)
5         emit(i)
6     }
7 }
8
9 fun main() = runBlocking<Unit> {
10     println("Calling simple function...")
11     val flow = simple()
12     println("Calling collect...")
13     flow.collect { value -> println(value) }
14     println("Calling collect again...")
15     flow.collect { value -> println(value) }
16 }
17
18 //Output:
19 //Calling simple function...
20 //Calling collect...
21 //Flow started
22 //1
23 //2
24 //3
25 //Calling collect again...
26 //Flow started
27 //1
28 //2
29 //3
30
31

```

Listing 2.13: Kotlin collect multicall example

As we can see, with the use of the keyword *emit*, it is achieved the same of what we saw in C# with the use of the keyword *yield return*. In this case, the same way the *yield return* returned an item that was part of an asynchronous enumeration of events through `IAsyncEnumerable`, the use of the keyword *emit* will lazily emit data, as it becomes available to be set as event of the Flow.

Likewise what happens in RxJava, `Flow<T>` provides a fluent API of intermediate operators that allow data transformation through the use of operation pipelines, where is received an upstream flow and the operators return a transformed downstream flow through the traditional push methods like: `filter`, `map()`, `zip()`, `take()` etc... On the next example, we can see an example of an asynchronous data pipeline operation from Kotlin `Flow<T>`, taking advantage of the Fluent API provided by its platform:

```

1 suspend fun performRequest(request: Int): String {
2     delay(1000) // imitate long-running asynchronous work
3     return "response $request"
4 }
5
6 fun main() = runBlocking<Unit> {
7     (1..3).asFlow() // a flow of requests
8     .map { request -> performRequest(request) }
9     .collect { response -> println(response) }
10 }
11
12 //Output:
13 //response 1
14 //response 2
15 //response 3
16
17

```

Listing 2.14: Kotlin collect multicall example

2.3.5 Technologies comparison

As we saw, each technology have a set of properties that help to characterize the solution. To help the characterization of each technology documented, we have a relation between the characteristics saw in the chapter 2.2.

	RxJava	Reactor Java (FLUX)	Kotlin Flow	Javascript	C# async enums
Pull	X	X	X		X
Push	X	X	X	X	X
Cancelable	X	X	X	X	X
Error Handling	X	X	X	X	X
Backpressure	X	X	X		
Intrinsic words			X	X	X

Benchmarking of Asynchronous Data Processing Algorithms

In this chapter, we delve into the benchmark results of two key algorithms applied using various strategies and implementations. These algorithms can be understood as abstract representations of real-world tasks that involve processing large amounts of data asynchronously. The two main algorithms under consideration are:

1. Finding the largest word in a set of files
2. Grouping of words based on their sizes, known as the "group words" operation

The "group word" operation identifies words within a specified size range in a dataset. The algorithm then returns a collection of these words, along with a count of their occurrences, and is particularly designed to output the most recurring words within the specified range. This operation presents an intriguing computational challenge as it requires efficient data retrieval, processing, and frequency analysis.

Conversely, finding the largest word in a set of files, although seemingly simple, becomes a non-trivial task when considering vast amounts of data.

The data used in these tests is a collection of text files from Project Gutenberg, a large digital library of thousands of free eBooks. Project Gutenberg offers a wide variety of books in different languages, and for this experiment, we used a random sample of hundreds of books, providing a diverse and challenging dataset for our implementations.

During the implementations, a conscious effort was made to keep the operation pipelines as similar as possible across the different technologies for each algorithm. This endeavor aimed to create a fair and representative evaluation of the behaviors of each technology. By maintaining consistency in pipeline operations, we can more accurately attribute performance differences to the underlying technology, rather than variations in the implemented code. This approach brings us closer to a true comparison of how each technology handles the challenges of asynchronous I/O data retrieval and processing.

In certain instances, particularly with Java, there was a need to incorporate external libraries for non-blocking asynchronous file retrieval. This requirement highlights the varying degrees of native support for these operations across programming environments. Some environments have native functions, while others rely heavily on third-party solutions. This aspect of the project mirrors the challenges often faced in real-world environments, where the necessity to adapt and find suitable solutions is a common part of the development process.

The implementations differ in the specific programming languages, libraries, and technologies they use. This allows us to evaluate the relative strengths and weaknesses of each approach and provides valuable insights into how these factors can affect performance.

After presenting and discussing each implementation and its results, we will be able to directly compare them. This will enable us to draw meaningful conclusions about the performance of the strategies and technologies when applied to the same task under the same conditions. Specifically, we are interested in how the performance of a given strategy or technology can vary across different programming frameworks when tasked with finding the largest word and executing the group word operation.

3.1 Strategies

The aim of this section is to define various strategies that we will be implementing across different languages and technologies. While some of these strategies can

be used universally across all platforms, others are platform-specific due to the constraints of the environment.

3.1.0.1 Reactive Extensions (Rx)

Reactive Extensions (Rx) is a library for composing asynchronous and event-based programs using observable sequences. This model promotes a focus on data streams and the propagation of change. Rx has implementations in several programming languages including .NET (Rx.NET), Java (RxJava), and JavaScript (RxJS).

3.1.0.2 Reactor (Flux)

Reactor is a fourth-generation reactive library, based on the Reactive Streams specification, for building non-blocking applications on the JVM based on Java 8 and later. The Flux class, a Reactive Streams Publisher with reactive operators, emits 0 to N elements, and then completes (successfully or with an error).

3.1.0.3 Blocking Reader in Streams

This strategy uses Java's streams interface to read the data in a blocking manner. Due to the blocking nature of the I/O operations, this approach may lead to slower processing times as it has to wait for each I/O operation to complete.

3.1.0.4 Multi-threading

This strategy leverages multi-threading to process data concurrently. By distributing the work across multiple threads, it often results in improved performance, especially for CPU-bound tasks.

3.1.0.5 Kotlin Flow

In Kotlin, the concept of 'flows' is integral to the Coroutines library. A Flow is an asynchronous data stream that sequentially emits values and completes normally or with an exception.

3.1.0.6 Concurrency

Concurrency is used in combination with several of the aforementioned strategies such as Kotlin Flow and Rx. Concurrency aims to improve the performance of data processing by executing multiple tasks in parallel.

3.1.0.7 Baseline

The Baseline strategy is a straightforward implementation without any pipelining or other overhead. It serves as a reference for comparison. This method may use traditional synchronous I/O operations and sequential processing.

These strategies will be applied to our chosen algorithms and implemented in the .NET, Java, and Kotlin environments. It's important to note that the aim is not to find the 'best' strategy but to demonstrate the relative strengths and weaknesses of each approach, how they can affect performance, and how they fit into different programming paradigms.

3.2 .NET Implementation

3.2.0.1 Approaches Explanation

3.2.1 Sync Baseline

The Sync Baseline approach is the most straightforward and is synchronous, meaning it processes one task at a time and waits for each operation to complete before moving on to the next. This strategy reads a file line by line and extracts the longest word from each line. It repeats this for each file in the directory.

3.2.2 Async Blocking Read

The Async Blocking Read approach applies asynchronous programming to read file content. This approach reads files in blocks, with each block being processed for its longest word. However, while it does use `async/await`, it does not truly take full advantage of asynchrony as it still blocks on the read operation to complete before processing the block.

3.2.3 Async Single Task

The Async Single Task approach is an asynchronous version of the baseline strategy. It uses `async` and `await` keywords to allow other tasks to execute when the current one is waiting for I/O operations to complete, improving the overall performance of the program. It reads a file line by line, extracts the longest word from each line, and repeats this for each file in the directory. The processing of each file is wrapped in a separate Task, but the processing of the files is still done sequentially.

3.2.4 Async Baseline

The Async Baseline approach is similar to the Async Single Task approach but with one major difference: the processing of files is done in parallel, meaning it doesn't wait for a file to be completely processed before starting to process the next file. This strategy allows for better utilization of system resources, especially when dealing with a large number of files.

3.2.5 Linq Sync

The Linq Sync approach uses LINQ (Language Integrated Query) to process the files. It is a synchronous approach that utilizes the powerful querying capabilities of LINQ to read and process the files. Despite being synchronous, LINQ's declarative nature can lead to more readable and concise code.

3.2.6 Async Enumerable

The Async Enumerable approach uses the `IAsyncEnumerable` interface to asynchronously stream and process data as it becomes available, instead of waiting for all data to be available as with traditional enumerables. This strategy can provide a performance boost when dealing with large datasets.

3.2.7 Rx

The Rx approach uses the Reactive Extensions (Rx) library for .NET, which provides tools for composing asynchronous and event-based programs using observable sequences and LINQ-style query operators. Like Async Enumerable, it

can also stream and process data as it becomes available.

3.2.7.1 Biggest Word Results

Strategy	Processing Time Average (s)
Sync Baseline	15.8
Multithread (Async Blocking Read)	1.6
Async Baseline (Async Single Task)	2.8
Linq Sync	22.2
Async Enumerable	30
Rx	23
Rx Async File Read	15

Table 3.1: Processing times for different strategies for "Find the biggest word".

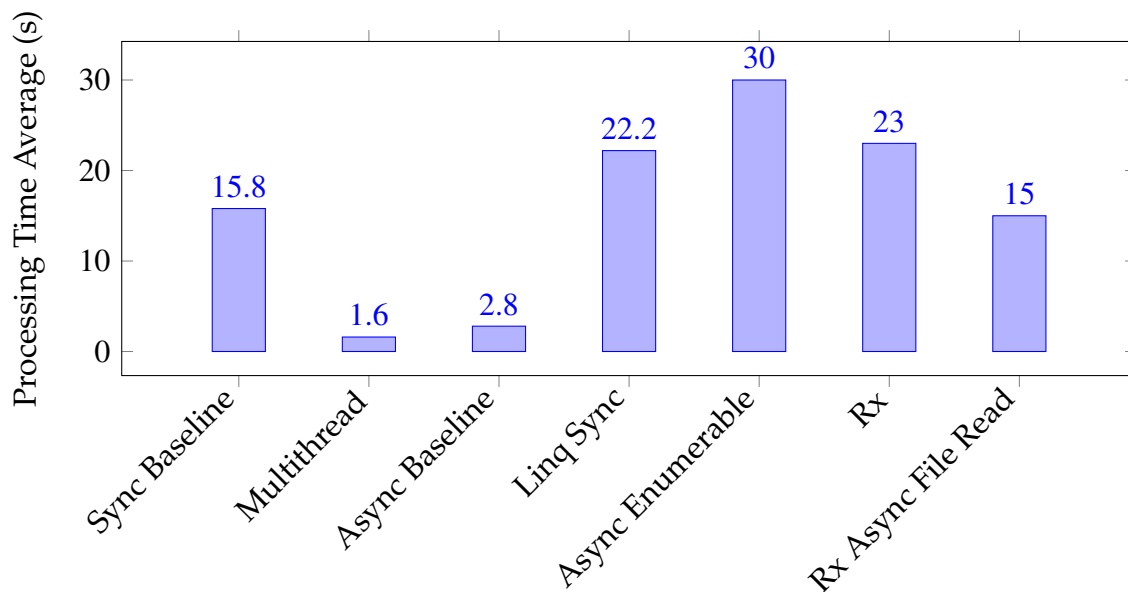


Figure 3.1: Processing times for different strategies for "Find the biggest word".

3.2.7.2 Group Word Results

Strategy	Processing Time Average (s)
RunCountWordsBaseline	26
RunSyncTest	22
RunAsyncEnumerableTest	30
RunRxTest	22

Table 3.2: Processing times for different strategies for "Count Words".

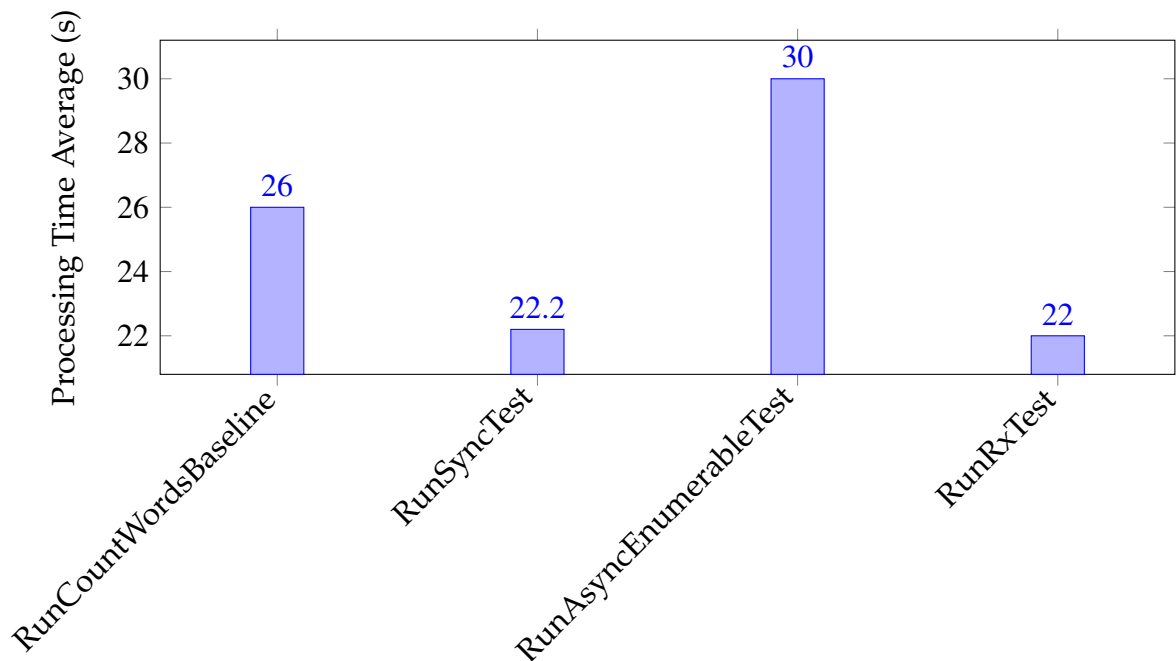


Figure 3.2: Processing times for different strategies for "Count Words".

3.3 Java/Kotlin Implementation

In this section, we will present the different strategies utilized in Java and Kotlin implementations for file processing and compare their performances. In the case of Java, asynchronous file reading isn't directly supported by the standard libraries. To overcome this limitation and maintain parity with the other implementations, we made use of a library named `AsyncFiles`. This library was

developed by Professor Jorge Martins of the Lisbon Engineering Superior Institute. It provides an efficient and straightforward way to perform asynchronous file reading operations in Java.

3.3.1 Results

3.3.1.1 Biggest Word Results

Strategy	Processing Time Average (s)
NIO Reactor Flux	1.1
NIO Reactor Flux with Concurrency (use of <code>parallel()</code>)	1.0
NIO Observable RXJava	0.9
NIO Observable with Concurrency (use of <code>parallel()</code>)	0.8
Blocking Reader in Streams	3.9
Blocking Reader MultiThread	0.7
Blocking Reader Streams with Concurrency	3.0
NIO Baseline	0.9
Concurrent Processing with Kotlin Flow	2.4
Sequential Processing with Kotlin Flow	5.1
Kotlin Flow IO Blocking Operations	5.2

Table 3.3: Processing times for different Java/Kotlin strategies for "Biggest Word".

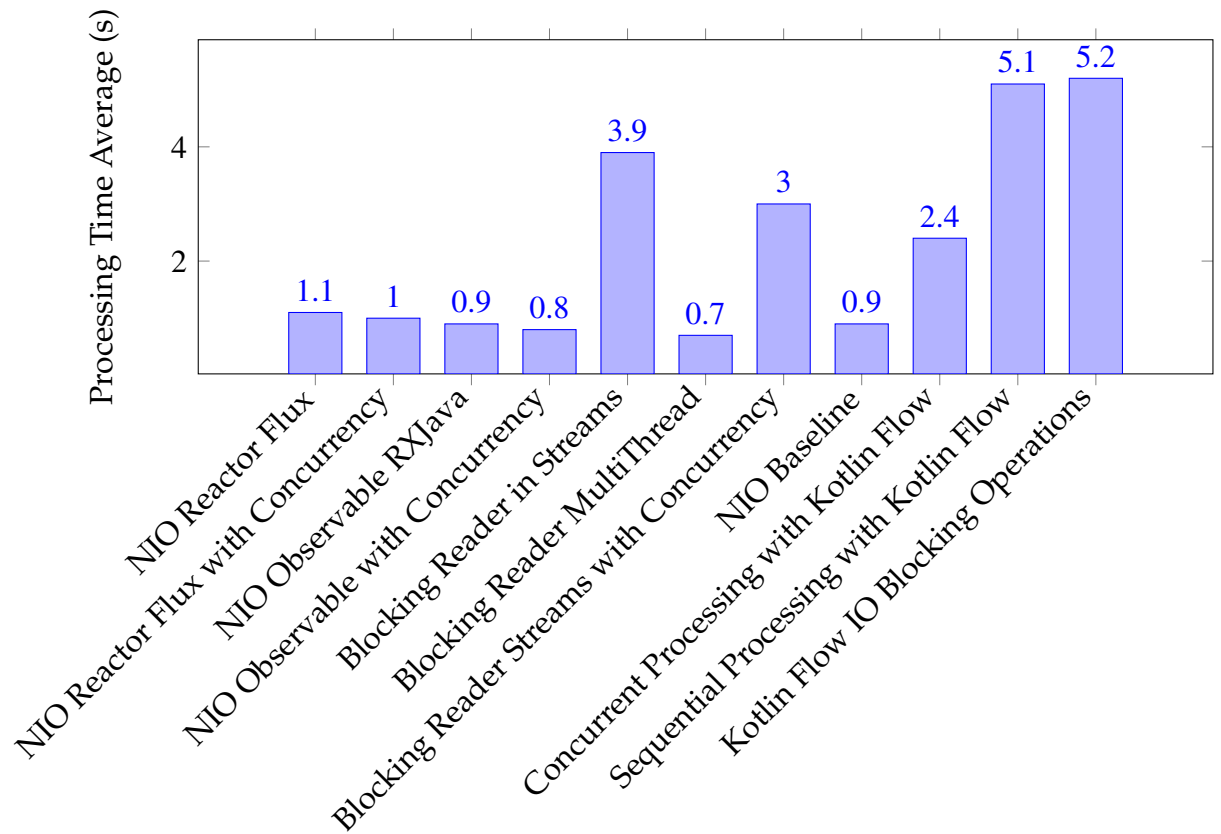


Figure 3.3: Processing times for different Java/Kotlin strategies for "Biggest Word".

3.3.1.2 Group Word Results

Strategy	Processing Time Average (s)
GroupWordsBaseline	1.3
GroupWordsRXJava	5.8
GroupWordsInRectorCoreFlux	12.0
GroupWordsBlockingReaderInMultiThread	1.2
GroupWordsBlockingReaderInStreams	1.8
GroupWordsWithFlow (Kotlin)	15.0

Table 3.4: Processing times for different Java/Kotlin strategies for "Group Words".

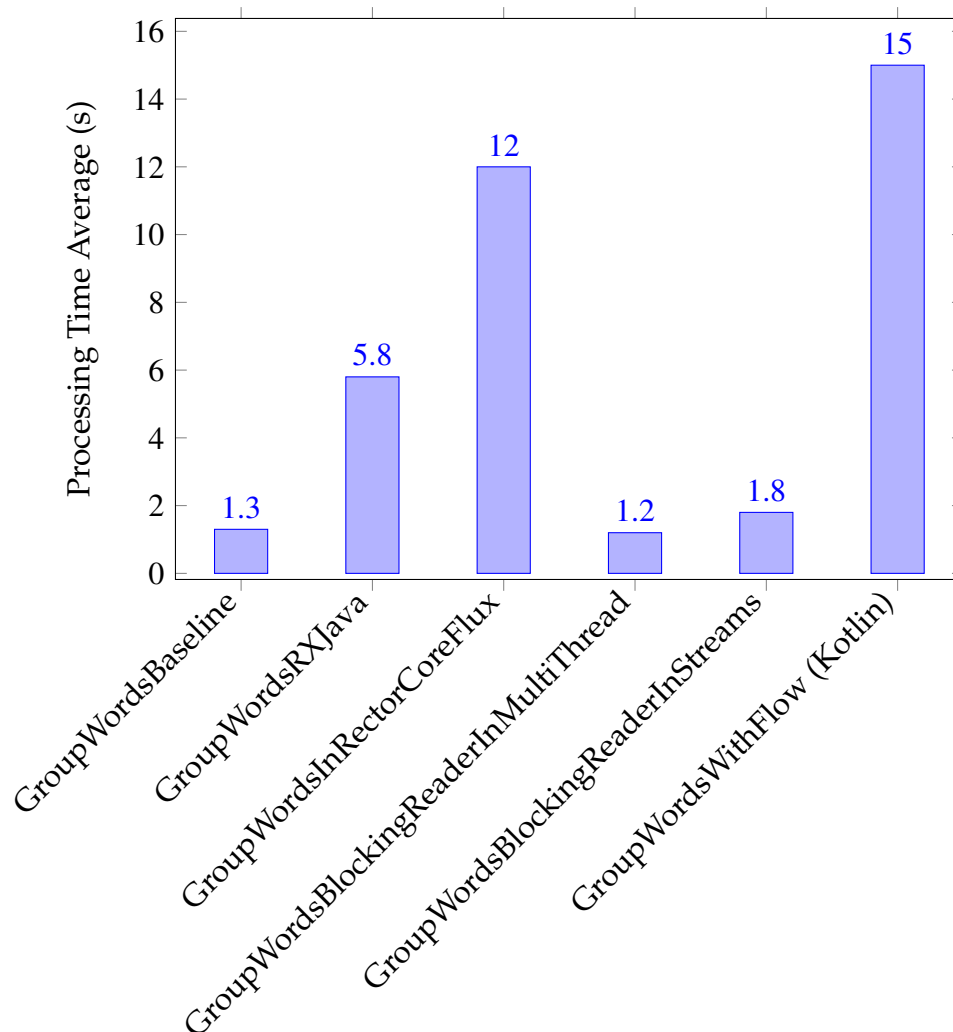


Figure 3.4: Processing times for different Java/Kotlin strategies for "Group Words".

...

3.4 JavaScript Implementation

Strategy	Processing Time Average (s)
Baseline JS	8.2
Baseline Stream JS	23.2
RxStrategy JS	13.9

Table 3.5: Processing times for different JavaScript strategies for "Biggest Word".

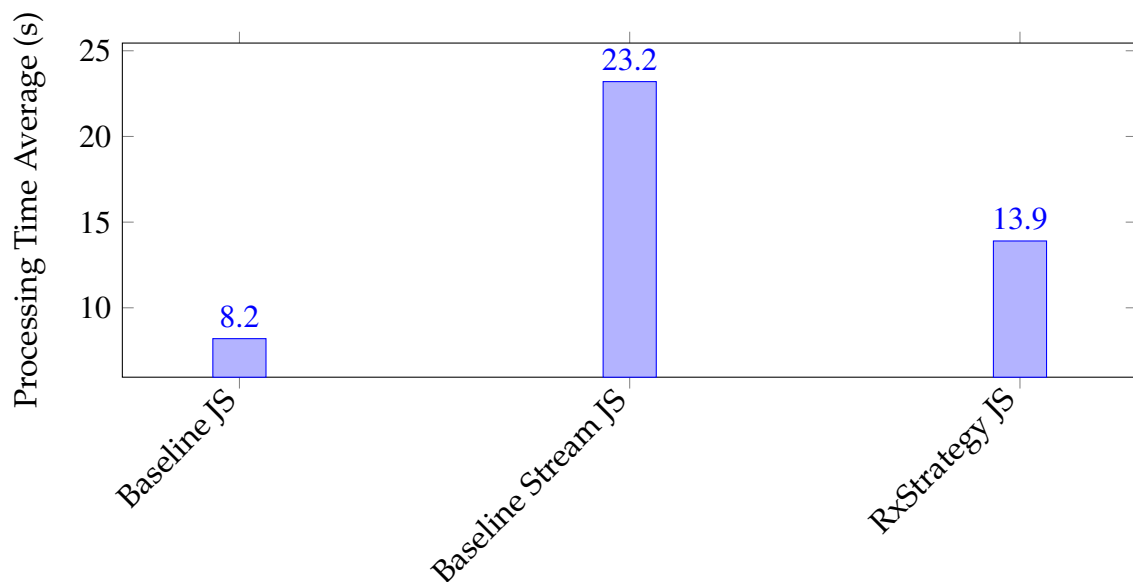


Figure 3.5: Processing times for different JavaScript strategies for "Biggest Word" Graphic

3.4.1 Discussion

3.5 Conclusion