Liquid IT : Toward a better compromise between development scalability and performance scalability not definitive

Etienne Brodu

July 10, 2015

Abstract

TODO translate from below when ready

Résumé

Internet étend l'économie à une échelle spatiale et temporelle sans précédent. Il permet à chacun de mettre un service à disposition de milliards d'utilisateurs, en seulement quelques heures. La plupart des grands services actuels ont commencés comme de simples applications créées dans un garage par une poignée de personnes. C'est cette facilité à l'entrée qui a permis jusqu'à maintenant une telle croissance sur le web. Google, Facebook ou Twitter en sont les exemples les plus flagrants. Pendant le développement d'une application, il est important de suivre cette croissance, au risque de se faire rattraper par la concurrence. Le développement est guidé par les besoins en terme de fonctionnalités, afin de vérifier rapidement si le service peut satisfaire l'audience. Des langages tel que Ruby ou Java se sont imposés comme les languages du web, justement parce qu'ils permettent d'intégrer facilement de nouvelles fonctionnalités.

Si une application répond correctement aux besoins, elle atteindra de manière virale un nombre important d'utilisateurs. Son audience peut prendre plusieurs ordres de grandeurs en quelques jours seulement, ou même en quelques heures suivant comment elle est relayée. Une application est dites scalable si elle peut absorber ces augmentations d'audience. Or il est difficile pour une application dont le développement est guidé par les fonctionnalités d'être scalable.

Au moment où l'audience commence à devenir trop importante, il est nécessaire de modifier l'approche de développement de l'application. Le plus souvent cela implique de la réécrire complètement en utilisant des infrastructures scalables qui imposent des modèles de programmation et des API spécifiques. Cela représente une charge de travail conséquente et incertaine. De plus, l'équipe de développement dois concilier cette nouvelle approche de développement scalable, avec la demande en fonctionnalités. Aucun langage n'a clairement réussi le compromis entre ces deux objectifs.

Pour ces raisons, ce changement est un risque pour la pérennité de l'application. D'autant plus que le cadre économique accorde peu de marges d'erreurs, comme c'est le cas dans la plupart des start-up, mais également dans de plus grandes structures.

Mon travail consiste à tenter d'écarter ce risque dans une certaine mesure. Ma thèse se base sur les deux observations suivantes. D'une part, Javascript

est un langage qui a énormément gagné en popularité ces dernières années. Il est omniprésent sur les clients, et commence à s'imposer également sur les serveurs avec Node.js. Il a accumulé une communauté de développeur importante, et est l'environnement d'exécution le plus largement déployé. De ce fait, il se place comme le langage principal du web, détrônant Ruby ou Java. D'autre part, l'execution de Javascript a la particularité de ressembler à un pipeline. La boucle événementielle de Javascript est un pipeline qui s'exécute sur un seul cœur pour profiter d'une mémoire globale. On observe le même flux de messages traversant cette boucle événementielle que dans un pipeline.

L'objectif de ma thèse est de permettre à des applications développées en Javascript d'être automatiquement transformées vers un pipeline d'exécutions repartis. Nous construisons un compilateur permettant d'identifier les fonctions de Javascript et de les isoler dans ce que nous appelons des Fluxions. Un conteneur qui peut exécuter une fonction à la réception d'un message, et envoyer des messages pour continuer le flux vers d'autres fluxions. Les fluxions étant indépendantes, elles peuvent être déplacées d'une machine à l'autre. En transformant automatiquement un programme Javascript en Fluxions, on le rend scalable, sans effort.

Contents

Chapter 1

Introduction

TODO 5p

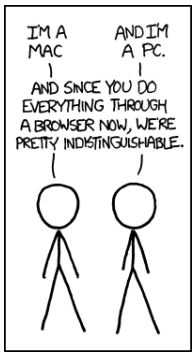
Chapter 2

Context and objectives

Contents

2.1 The Web as a platform

2.1.1 From operating systems to the web



With the invention of electronic computing machine, appeared the market for software applications. This market is not limited by marginal production cost ; software being a virtual product, the production and distribution cost for another unit is virtually null. The market is limited by the platform a software can be deployed on. The bigger the platform, the wider the market. It is interesting economically both for the provider, and for the consumer to standardize and widen the platform. For the provider because it can increase its revenue by reaching more consumers. And for the consumers, because of the competition between providers feeding innovation.

However, the first platforms for software started as products. The first commodity computers and the operating systems are economic products. It made sense for their manufacturer to increase their market share, so as to increase their revenue. The economical incentive. Microsoft successfully took over the market of operating system in the 90s, and cross the border of the monopoly more than once. However, before the internet, the distribution of software copies is still limited by physical medium. It still takes time to burn a CD, or a floppy, and to bring it to the consummer's home.

Sir Tim Berners Lee invented the world wide web in 1989. It was initially intended to share scientific documents and results. But it eventually became the distribution medium of choice for virtual products, software included. And pushed the scalability of software distribution.

Similarly to operating systems, Web browsers are products. They exposed innovative features to try to increase their market share. Among these features is the ability to run scripts directly inside the browser. Browsers partially replaced operating system as the platform to run software. But these heterogeneity in features between web browsers worsen the user experience on the web. Eventually, web technologies became specification to be respected by all browsers, so as to assure an homogeneous experience across the web. The web became the platform. It allows to deploy and run software at unprecedented scales. With web services, or Software as a Service, the distribution medium of software is so transparent that owning a product to

have an easier access is no longer relevant.

The browser and the operating system are only the foundation. The specifications assure the unification of the web as a platform. We explore now the different language to write and deploy applications on the web. *

TODO This subsection needs reviews. The message is still not clear.

2.1.2 The languages of the web

*

TODO This subsection needs reviews, it is poorly written.

In the early 90's, at the same time the web started developing, most of the more popular programming langauges were released. Python(1991), Ruby(1993), Java(1994), PHP(1995) and Javascript(1995). With Moore's law predicting exponential increase in hardware performance, the industry realized that development time is more expensive than hardware. Low-level language were replaced by higher-level language with worse performance, but easier to develop with, to decrease the development time.

Java, developed by Sun Microsystems, imposes itself early as a language of choice and never really decreased. The language is executed on a virtual machine, allowing to write an application once, and to deploy it on heterogeneous machines. Additionally, it presents many hardware abstractions, such as memory garbage collection, and references, instead of manual memory allocation and pointers. For these reasons, it is easier to develop with, but shows slower performance than lower level languages like C/C++. The software industry quickly adopted it as its main development language. It is currently the most cited language on StackOverflow, and the second used on Github. It is generally in the first place of all the language popularity indexes. However, the software industry wants stable and safe solutions. This prudence generally slows down Java evolution. *

TODO More arguments needed here. Java is overengineered, overtooled, too verbose, it is not dynamic enough : OOP is outdated, and it lacks lambdas, and so on ...

Python is the second best language for everything. It is a general purpose language, currently quite popular for data science. In 2003, the release of the Django web framworks brought the language to the web development scene.

Ruby was confined in Japan and almost unknown to the world until the release of Rails in 2005. With the release of this web framework, Ruby took-off and is still in active use. *

TODO why Ruby could have replaced Java ? dynamic language, functional paradigm

PHP stands for Personal Home Page Tools. It was designed to design web pages since its beginning. It is probably one of the easiest language to start web development. However, according to several language popularity indexes, it is on a slow decline since a few years. Because of its simplicity, it often fail to grow projects to industrial size, and is being replaced by

languages that succeed at bringing a prototype to industrial standards, like Java, Ruby or Python.

Since a few years, Javascript is slowly becoming the main language for web development. It is everywhere. It is present in every browser, and on the server as well with Node.js. Because of this unavoidable position, it became fast (V8, ASM.js) and usable (ES6, ES7). Additionally, it is a target for LLVM, allowing many languages to compile to Javascript, strengthening again its omnipresent position. I argue in this thesis, that Javascript is the language of choice to bring a prototype to industrial standards. * *

⚠ TODO un peu rapide ici

2.1.3 Explosion of Javascript popularity

2.1.3.1 In the beginning

⚠ TODO I should note the distinction between Javascript client and server side. There is no choice on the client, while there is a choice on the server. So the popularity on the client is not as significant than on the server. However, as I said, Javascript unavoidable position made its success and its performance.

Javascript was created by Brendan Eich at Netscape around May 1995, and released to the public in September. The initial name of the project was Mocha, then LiveScript, the name Javascript was finally adopted to leverage the trend around Java. Indeed, Java was considered the hot new web programming language at this time. It was quickly adopted as default language for web servers development, and everybody was betting on pushing Java to the client as well. The history proved them wrong.

When Javascript was released in 1995, the world wide web was on the rise.[1] Browsers were emerging, and started a battle to show off the best features and user experience to attract the wider public.[2] Javascript was released to be one of these features. Microsoft released their browser Internet Explorer 3 in June 1996 with a concurrent implementation of Javascript. They changed the name to JScript, to avoid trademark conflict with Oracle Corporation, who owns the name Javascript. The differences between the two implementations made difficult for a script to be compatible to both. At the time, banners and signs started to appear on web pages to inform the user about the ideal web browser to use for the best experience. This competition was fragmenting the web.

Netscape submitted Javascript to Ecma International for standardization in November 1996 to stop this fragmentation. In June 1997, ECMA International released ECMA-262, the first specification of ECMAScript, the

[1] `http://www.internetlivestats.com/internet-users/`

[2] to get an idea of the web in 1997 : `http://1x-upon.com/`

standard for Javascript. A standard to which all browser should refer for their implementations.

The initial release for this specification was designed in a rush. The version released in 1995 was finished within 10 days. Because of this precipitation, the language has often been considered poorly designed and unattractive. Moreover, Javascript was intended to be simple enough to attract unexperienced developers, by opposition to Java or C++, which targeted professional developers. For these reasons, Javascript started with a poor reputation among the developer community.

But things evolved drastically since. The success of Javascript is due to many factors ; maybe the most important of all is the View Source menu that reveals the complete source code of any web application. *The view source menu is the ultimate form of open source*[3]. It is the vector of the quick dissemination of source code to the community, which picks, emphasizes and reproduces the best techniques. This brought open source and collaborative development before github. * Moreover, all modern web browsers now include a Javascript interpreter, making Javascript the most ubiquitous runtime in history [**Flanagan2006**].

⚠ TODO neither open source nor collaborative development are the correct terms

When a language like Javascript is distributed freely with the tools to reproduce and experiment on every piece of code. When this distribution is carried during the expansion of the largest communication network in history. Then an entire generation seizes this opportunity to incrementally build and share the best tools they can. This collaboration is the reason for the popularity of Javascript on the Web.

2.1.3.2 Rising of the unpopular language

*

⚠ TODO why is Javascript unpopular ? cite some blog post like : https://wiki.theory.org/YourLan and add many blog posts titles in a mozaïc

Javascript started as a programming language to implement short interactions on web pages. The best usage example was to validate some forms on the client before sending the request to the server. This situation hugely improved since the beginning of the language. Nowadays, there is a lot of web-based application replacing desktop applications, like mail client, word processor, music player, graphics editor.

There is now more software services released to the public as web-based application compared to desktop clients.

[3] `http://blog.codinghorror.com/the-power-of-view-source/`

ECMA International released several version in the few years following the creation of Javascript. The first and second version, released in 1997 and 1998, brought minor revisions to the initial draft. However, the third version, released in the late 1999, contributed to give Javascript a more complete and solid foundation as a programming language. From this point on, the consideration for Javascript kept improving.

In 2005, James Jesse Garrett released *Ajax: A New Approach to Web Applications*, a white paper coining the term Ajax [**Garrett2005**]. This paper points the trend in using this technique, and explain the consequences on user experience. Ajax stands for Asynchronous Javascript And XML. It consists of using Javascript to dynamically request and refresh the content inside a web page. It has the advantage to avoid requesting a full page from the server. Javascript is not anymore confined to the realm of small user interactions on a terminal. It can be proactive and responsible for a bigger part in the whole system spanning from the server to the client. Indeed, this ability to react instantly to the user gave to developer the feature to develop richer applications inside the browser. At the time, the first web applications to use Ajax were Gmail, and Google maps[4].

Around this time, the Javascript community started to emerge. The third version of ECMAScript had been released, and it was homogeneously supported in the browsers. However, the DOM, and the `XMLHttpRequest` method, two components on which AJAX relies, still present heterogeneous interfaces among browsers. Javascript framework were released with the goal to straighten the differences between browsers implementations. Prototype[5] and DOJO[6] are early famous examples, and later jQuery[7] and underscore[8]. These frameworks are responsible in great part to the wide success of Javascript and of the web technologies.

In the meantime, in 2004, the Web Hypertext Application Technology Working Group[9] was formed to work on the fifth version of the HTML standard. This new version provide new capabilities to web browsers, and a better integration with the native environment. It features geolocation, file API,

[4] A more in-depth analysis of the history of Ajax, given by late Aaron Swartz `http://www.aaronsw.com/weblog/ajaxhistory`

[5] `http://prototypejs.org/`

[6] `https://dojotoolkit.org/`

[7] `https://jquery.com/`

[8] `http://underscorejs.org/`

[9] `https://whatwg.org/`

web storage, canvas drawing element, audio and video capabilities, drag and drop, browser history manipulation, and many mores. It gave Javascript the missing interfaces to become a rich environment to develop rich application in the browser. The first public draft of HTML 5 was released in 2008, and the fifth version of ECMAScript was released in 2009. These two releases, ECMAScript 5 and HTML5, represent a mile-stone in the development of web-based applications. Javascript became the programming language of this rising application platform.

Javascript, and web technologies are also used outside the web. NW.js[10] and electron[11] are two solutions to deploy application built with web technologies. They use Node.js and Chromium. The Atom text editor[12], Popcorn Time[13] and Light Table[14] are example of such applications. However, if web applications are common choice for web service client on the desktop, HTML5 is not yet widely accepted as ready to build complete application on mobile, where performance and design are crucial. Indeed web-technologies are often not as capable, and well integrated as native technologies. But even for native development, Javascript seems to be a language of choice. An example is the React Native Framework[15] from Facebook, which allow to use Javascript to develop native mobile applications. They prone the philosophy *"learn once, write anywhere"*, in opposition to the usual slogan *"write once, run everywhere"*.[16]

*

⚠

Insert in this section a summary table for HTML and Javascript

2.1.3.3 Current situation

"When JavaScript was first introduced, I dismissed it as being not worth my attention. Much later, I took another look at it and discovered that hidden in the browser was an excellent programming language."

—Douglas Crockford

The rise of Javascript is obvious on the web and particularly the open

[10] `https://github.com/nwjs/nw.js`
[11] `https://github.com/atom/electron`
[12] `https://atom.io/`
[13] `https://popcorntime.io/`
[14] `http://lighttable.com/`
[15] `https://facebook.github.io/react-native/`
[16] Used firstly by Sun for Java, but then stolen by many others

source communities. It also seems to be rising in the software industry. However, it is harder to give an accurate picture of the situation. The software industry is not as clear and open as the web. Moreover, there is no right metrics to accurately and directly measure programming language popularity. In the following paragraphs, I report some of the best metrics and indexes available freely on the web to try to represent the situation, both in the open source community and in the more opaque software industry. More detailed informations are available section A.

Available resources The TIOBE Programming Community index is a monthly indicator of the popularity of programming languages. Javascript ranks 6th on this index, as of April 2015, and it was the most rising language in 2014. It uses the number of results on many search engines as a measure of the popularity of a programming language. The results contains learning and training resources, forums logs, books and many other traces of the activity of a the community around the language. However, the measure used by the TIOBE is controversial, and might not be representative. It is a lagging indicator, and the number of pages doesn't represent the number of readers.

Alternatively, the PYPL index is based on Google trends to measure the number of requests on a programming language. Javascript ranks 7th on this index, as of May 2015. This index seems to be more accurate, as it depicts the actual interest of the community for a language. However, it is not representative as it only takes Google search into account.

From these indexes, the major programming languages are Java, C/C++ and C#. The three languages are still the most widely taught, and used to write softwares. But Javascript is rising to become an important language as well.

Developers collaboration platforms An indicator of the popularity and usage of a language is the number of developers and projects using it.

Github is the most important collaborative development platform, with around 9 millions users. Javascript is the most used language on github since mid-2011, with more than 320 000 repositories. The second language is Java with more than 220 000 repositories.

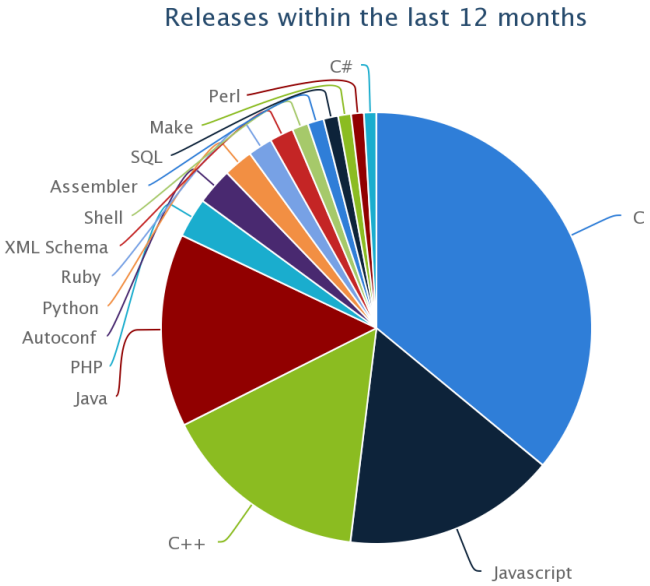
*

StackOverflow, is the most important Q&A platform for developers. It is a good representation of the activity around a language. Javascript is the

⚠ TODO : graph of Github repositories by languages

second language showing the most activity on StackOverflow, with more than 840 000 questions. The first one is Java with more than 850 000 questions.

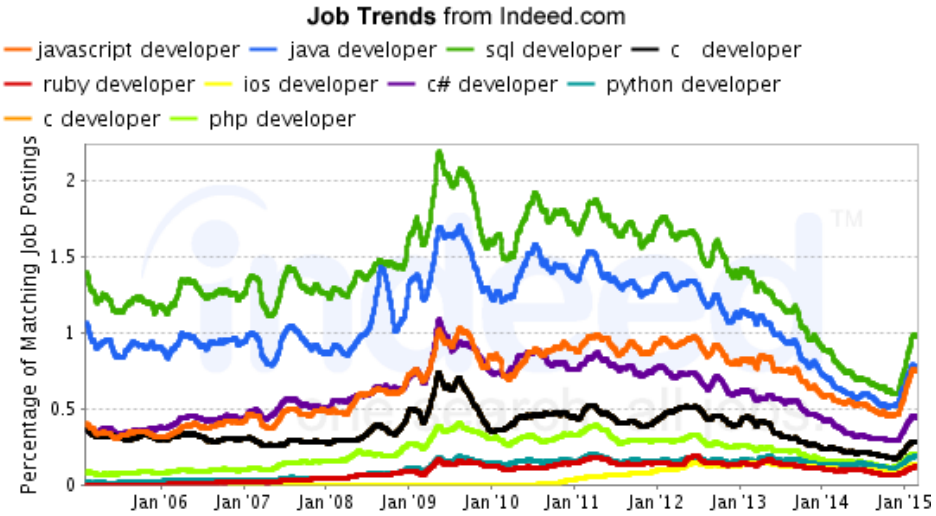
Black Duck Software helps companies streamline, safeguard, and manage their use of open source. For its activity, it analyzes 1 million repositories over various forges, and collaborative platforms to produce an index of the usage of programming language in open source communities. Javascript ranks second. C is first, and C++ third. Along with Java, the four first languages represent about 80% of all programming language usage.



*

TODO redo this graph, it is ugly.

Jobs The software industry is rather closed sourced, and its activity is rather opaque. All these previous metrics are representing the visible activity about programming language, but are not representative of the software industry. The trends on job openings gives an hint of the direction the software industry is heading towards. *Indeed* provide some trends over its database of job propositions. Javascript developers ranked at the third position, right after SQL and Java developers. Then come C# and C developers. This position means that Javascript might finally be on the edge to become a major language in the software industry, and become as important as Java and C/C++.



*

TODO redo this graph, it is ugly.

All these metrics in this section represent different faces of the current situation of Javascript adoption. With the rise of web applications, we can safely say that Javascript is one of most important language of this decade, alongside with Java and C/C++. It is widely used in open source projects, and everywhere on the web, as well as in the software industry.

2.2 Highly concurrent web servers

*

This section needs review

2.2.1 Concurrency

The Internet allows interconnection at an unprecedented scale. There is currently more than 16 billions devices connected to the internet, and it is growing exponentially[17]. This massively interconnected network gives the ability for a web applications to be reached at the largest scale. A large web application like google search receives about 40 000 requests per seconds. That is about 3.5 billions requests a day[18]. Such a web application needs to be highly concurrent to manage a large number of simultaneous request. Concurrency is the ability for an application to make progress on several

[17] http://blogs.cisco.com/news/cisco-connections-counter

[18] http://www.internetlivestats.com/google-search-statistics/

tasks at the same time. For example to respond to several simultaneous requests, a task is a part in the response to a request. *

TOOD define more clearly what is a task

In the 2000s, the limit to reach was to process 10 thousands simultaneous connections with a single commodity machine[19]. Nowadays, in the 2010s, the limit is set at 10 millions simultaneous connections at roughly the same price[20]. With the growing number of connected devices on the internet, concurrency is a very important property in the design of web applications.

2.2.1.1 Scalability

The traffic of a popular web application such as Google search is huge, and it remains roughly stable because of this popularity. There is no apparent spikes in the traffic, because of the importance of the average traffic. However, the traffic of a less popular web application is much more uncertain. If the web application fits the market need, it might become viral when it is efficiently relayed in the media. For example, when a web application appears in the evening news, it expects a huge spike in traffic. With the growth of audience, the number of simultaneous requests obviously increases, and the load of the web application on the available resources increases as well. The available resources needs to increase to meet the load. This growth might be steady and predictable enough to plan the increase of resources ahead of time, or it might be erratic and challenging. The spikes of a less popular web application are unpredictable. Therefore, the concurrency needs to be expressed in a scalable fashion. An application is scalable, if the growth of its audience is proportional to the increase of its load on the resources. For example, if a scalable application uses one resource to handle n simultaneous requests, it will use k resource to handle two times n simultaneous requests. With k being constant, for n ranging from tens to millions of simultaneous requests. Scalability assures that the resource usage is not increasing exponentially in function of the audience increase ; it increases roughly linearly.

2.2.1.2 Time-slicing and parallelism

Concurrency can be achieved on hardware with either a single or several processing units. On a single processing unit, the tasks are executed sequentially ; their executions are interleaved in time. On several processing unit,

[19] `http://www.kegel.com/c10k.html`
[20] `http://c10m.robertgraham.com/p/manifesto.html`

the tasks are executed in parallel. Parallel executions reduce computing time over sequential execution, as it uses more processing units.

If the tasks are completely independent, they can be executed in parallel as well as sequentially. This parallelism is a form of scalable concurrency, as it allows to stretch the computation on available hardware to meet the required performance, at the required cost. This parallelism is used in operating system to execute several applications concurrently to allow multi-tasking.

However, the tasks of an application are rarely independent. The tasks need to coordinate their dependencies to modify the global state of the application. This coordination limits the possible parallelism between the tasks, and might impose to execute them sequentially. The type of possible concurrency, sequential or parallel, is defined by the required coordination between the tasks. Either the tasks are independents and they can be executed in parallel, or the tasks need to coordinate a common state, and they need to be executed sequentially to avoid conflicting accesses to the state.

2.2.2 Interdependencies

It is easier to understand the possible parallelism of a cooking recipe than an application. That is because the modifications to the state are trivial in the cooking recipe, hence the interdependencies between operations. It is easy to understand that preheating the oven is independent from whipping up egg whites. While the interdependencies are not immediately obvious in an application. *

TODO is this metaphor useful here ? if yes, continue to a transition

2.2.2.1 State coordination

The interdependencies between the tasks impose the coordination of the global application state. This coordination happen either by sending events from one task to another, or by modifying a shared memory.

If the tasks are independent enough, they never need access to a state at the same time. The coordination of the state of the application can be done with message passing. They pass the states from one task to another so as to always have an exclusive access on the state. As example, applications built around a pipeline architecture define independent tasks arranged to be executed one after the other. The tasks pass the result of their computation to the next. These tasks never share a state.

If the tasks need concurrent accesses to a state, they cannot efficiently pass the state from one to the other repeatedly. They need to share and to coordinate their accesses to this state. Each access needs to be exclusive to avoid corruption. This exclusivity is assured differently depending on the scheduling strategy.

2.2.2.2 Task scheduling

There is roughly two main scheduling strategy to execute tasks sequentially on a single processing unit : preemptive scheduling and cooperative scheduling. The state coordination presented previously is highly depending on the scheduling strategy.

Preemptive scheduling is used in most execution environment in conjunction with multi-threading. The scheduler allows each task to execute for a limited time before preempting it to let another task execute. It is a fair and pessimistic scheduling, as it grant the same amount of computing time to each task. However, as the preemption might happen at any point in the execution, it is important for the developer to lock the shared state before access, so as to assure exclusivity. This protection is known to be hard to manage.

In cooperative scheduling, the scheduler allows a task to run until the task yield the execution back. Each task is an atomic execution ; it is never be preempted, and have an exclusive access on the memory. It gives back to the developer the control over the preemption. It seems to be the easiest way for developers to write concurrent programs efficiently. Indeed, I presented in the previous section the popularity of Javascript, which is often implemented on top of this scheduling strategy (DOM, Node.js).

As I explained the different paradigms for writing concurrent program in this subsection, it appears that the main problem is to assure to the developer for each task the exclusive access to the state of its application. This assurance is called invariance.

2.2.2.3 Invariance

I call invariance the assurance given that the state accessible from a task will remain unchanged during its access to avoid corruption, and more generally to allow the developer to perform atomic modifications on the state. This assurance allows the developer to regroup operations logically so as to

perform all the operations without interference from concurrent executions. The same concept is found in transactional memory.

In a multi-process application, there is no risk of corrupted state by simultaneous, conflicting accesses. The invariance is made explicit by the developer as the memory needs to be isolated inside each process. The invariance is assure at any point in time because the process remain isolated.

In a cooperative scheduling application, the developer is aware of the points in the code where the scheduler switches from one concurrent execution to the other, so it can manage its state in atomic modification. The invariance is assured, because any region in the memory can be accessed only by one task at a time.

Between these two invariances, the locking mechanisms seems to be a promising compromise. The developer defines only the shared states, and these are locked only when needed. However, it increases the complexity of the possible locked combination, leading to unpredictable situations, such as deadlock, and so on. The locking mechanisms are known to be difficult to manage, and sub-optimal. Indeed, they are eventually as efficient as a queue to share resources.

For the rest of this thesis, I focus only on the invariances provided by the multi-process paradigm and the cooperative scheduling. They are similar, because the developer defines sequence of instructions with atomic access to the memory. And in both paradigms, these sequences communicate by sending messages to each other. The difference is that in the multi-process paradigm, the developer defines the region and the isolated memory, while in the cooperative scheduling, the developer defines only the region, and the memory is isolated by the exclusivity in the execution.

This difference seems to be crucial in the adoption of the technology by the developer community. As we will see in the next subsection, the parallelism of multi-process is difficult to develop, but provide good performances, while the sequentiality of the cooperative scheduling is easier to develop, but provide poor performances compared to parallelism.

*

⚠

TODO this paragraph needs review

2.2.3 Technological shift

2.2.3.1 Scalable concurrency

Around 2004, the so-called Power Wall was reached. The clock of CPU is stuck at 3GHz because of the inability to dissipate the heat generated at higher frequencies. Additionally, the instruction-level parallelism is limited. Because of these limitations, a processor is limited in the number of instruction per second it can execute. Therefore, a coarser level of parallelism, like the task-level, multi-processes parallelism previously presented is the only option to achieve high concurrency and scalability. But as I presented previously, this parallelism requires the isolation of the memory of each independent task. This isolation is in contradiction with the best practices of software development, hence, is difficult to develop for common developers. It creates a rupture between performance and development accessibility.

2.2.3.2 The case for global memory

The best practices in software development advocate to design a software into isolated modules. This modularity allows to understand each module by itself, without an understanding of the whole application. The understanding of the whole application emerges from the interconnections between the different modules. A developer need only to understand a few modules to contribute to an application of hundreds or thousands of modules.

Modularity advocates three principles : encapsulation, a module contains the data, as well as the functions to manipulate this data ; separation of concerns, each module should have a clear scope of action, and this scope should not overlap with the scope of other modules ; and loose coupling, each module should require no, or as little knowledge as possible about the definition of other modules. The main goal followed by these principles, is to help the developer to develop and maintain a large code-base.

Modularity is intended to avoid a different problem than the isolation required by parallelism. The former intends to avoid unintelligible spaghetti code ; while the latter avoids conflicting memory accesses resulting in corrupted state. The two goals are overlapping in the design of the application. * Therefore, every language needs to provide a compromise between these two goals, and specialized in specific type of applications. I argue that the more accessible, hence popular programming languages choose to provide modularity over isolation. They provide a global memory at the sacrifice

⚠ TODO needs more explanations -> so it is hard for dev to do both ? Why exactly ?

of the performance provided by parallelism. On the other hand, the more efficient languages sacrifice the readability and maintainability, to provide a model closer to parallelism, to allow better performances. * *

TODO instead of language, use a more generic term to refer to language or infrastructure

2.2.3.3 Rupture

Between the early development, and the maturation of a web application, the development needs are radically different. In its early development, a web application needs to quickly iterate over feedback from its users. *"Release early, release often"*, and *"Fail fast"* are the punchlines of the web entrepreneurial community. The development team quickly releases a Minimum Viable Product as to get these feedbacks. The development reactivity is crucial. The first reason of startup failures is the lack of market need[21]. Therefore, the development team opt for a popular, and accessible language.

TODO justification and examples. What are modular application, or parallel applications ?

As the application matures and its audience grows, the focus shift from the development speed to the scalability of the application. The development team shift from a modular language, to a language providing parallelism.

This shift brings two problems. First, the development team needs to take a risk to be able to grow the application. This risk usually implies for the development team to rewrite the code base to adapt it to a completely different paradigm, with imposed interfaces. It is hard for the development team to find the time, hence the money, or the competences to deploy this new paradigm. Indeed, the number two and three reasons for startup failures are running out of cash, and missing the right competences. Second, after this shift the development pace is different. Parallel languages are incompatible with the commonly learned design principles. The development team cannot react as quickly to user feedbacks as with the first paradigm.

This technological rupture proves that there is economically a need for a a more sustainable solution to follow the evolution of a web application. A paradigm that it is easy to develop with, as needed in the beginning of a web application development, and yet scalable, so as to be highly concurrent when the application matures.

[21] `https://www.cbinsights.com/blog/startup-failure-post-mortem/`

2.3 Equivalence

I argue that the language should propose to the developer an abstraction to encourage the best practices of software development. Then a compiler, or the execution engine, can adapt this abstraction so as to leverage parallel architectures. So as to provide to the developer a usable, yet efficient compromise. We propose to find an equivalence between the invariance proposed by the cooperative scheduling paradigm and the invariance proposed by the multi-processes paradigm in the case of web applications.

2.3.1 Architecture of web applications

2.3.1.1 Real-time streaming web services

*

⚠

The need for invariance in the streaming applications : it can be emulated by message passing. Indeed the data flows from one processing step to the other, with few retro-progation of state (don't mention retro-propagation yet)

This equivalence intends not to be universal. It focuses on a precise class of applications : web applications processing stream of requests from users in soft real-time.

Such applications are organized in sequences of concurrent tasks to modify the input stream of requests to produce the output stream of responses. This stream of data stand out from the pure state of the application. The data flows in a communication channel between different concurrent tasks, and is never stored on any task. The state represents a communication channel between different instant in time, it remains in the memory to impact the future behaviors of the application. The state might be shared by several tasks of the application, and result in the needs for coordination presented in the previous section. In this thesis I study two programming paradigm derived directly form the cooperative scheduling and the multi-process paradigms presented in previous sections to be applied in the case of real-time web applications. The event-loop execution engine is a direct application of the cooperative scheduling, and the pipeline architecture is a direct application of the multi-process paradigm.

2.3.1.2 Event-loop

The event-loop is an execution model using asynchronous communication and cooperative scheduling to allow efficent execution of concurrent tasks on a single processing unit. It relies on a queue of event, and a loop to process each event one after the other. The communications are asynchronous to let the

application use the processor instead of waiting for a slow response. When the response of a communication is available, it queues an event. This event is composed of the result of the communication, and of a function previously defined at the communication initiation, to continue the execution with the result. In the Javascript even-loop, this function is defined following the continuation passing style, and is named a callback. After processing the result, this callback can initiate communications, resulting in the queuing of more events.

In this model, the data is the result of every communication operations - starting with the received user request - flowing through a sequence of callbacks, one after the other. The state contains all the variables remaining in memory from one request to the other, and from one callback to the other. In Javascript, it includes the closures.

*

⚠ TODO schema of an event-loop

2.3.1.3 Pipeline

The pipeline software architecture uses the multi-process paradigm and message passing to leverage the parallelism of a multi-core hardware architectures for streaming application. It consists of many processes treating and carrying the flow of data from stage to stage. This flow of data consist roughly of the requests, and associated data from the user, as well as the necessary state coordination between the stages. Each stage has its independent memory to hold its own state from one request to another.

*

*

⚠ TODO it is not universal, but multi-process paradigms are also oriented around event-loops. An Event-loop is a multi-process on one machine. A multi-process is multiple event-loop running different part of the same program.

The pipeline architecture and the event-loop model present similar execution model. Both paradigms encapsulate the execution, in callbacks or processes. Those containers are assured to have an exclusive access to the memory. However, they provide two different memory models to provide this exclusivity. It results in two distinct ways for the developer to assure the invariance, and to manage the global state of the application. The event-loop shares the memory globally through the application, allowing the best practice of software development. It is possibly the reason of the wide adoption of this programming model by the community of developers.

I argue in this thesis that it is possible to provide an equivalence between the two memory models for streaming web application. In the next subsection, I present the similarity in the execution model, and the differences in

⚠ TODO schema of a pipeline

the memory model for which an equivalence is necessary. Such equivalence would allow to transform an application following the event-loop model to be compatible with the pipeline architecture. This transformation would allow the development of an application following a programming model allowing the best practices of software development, while leverageing the parallelism of multi-core hardware architecture.

2.3.2 Equivalence

2.3.2.1 Rupture point

The execution of the pipeline architecture is well delimited in isolated stages. Each stage has its own thread of execution, and is independent of the others. On the other hand, the execution of the event-loop seems pretty linear to the developer. The continuation passing style nest callbacks linearly inside each others. The message passing linking the callbacks is transparently handled by the event-loop. However, the execution of the different callbacks are as distinct as the execution of the different stages of a pipeline. Precisely, the call stack is as distinct between two callbacks, as between two stages. Therefore, in the event-loop, an asynchronous function call represents the end of the call stack of the current callback, and the beginning of the call stack of the next. It represents what I call a rupture point. It is the equivalent to a data stream between two stages in the pipeline architecture.

Both the pipeline architecture and the event-loop present these ruptures points. To allow the transformation from the event-loop model to the pipeline architecture in the case of real-time web applications, I study in this thesis the possibility to transform the global memory of the event-loop into isolated memory to be able to execute the application on a pipeline architecture.

2.3.2.2 State coordination

The global memory used by the event-loop holds both the state and the data of the application. The invariance holds for the whole memory during the execution of each callback. As I explained in the previous section, this invariance is required to allow the concurrent execution of the different tasks. On the other hand, the invariance is explicit in the pipeline architecture, as all the stages have isolated memories. The coordination between these isolated process is made explicit by the developer through message passing.

I argue that the state coordination between the callbacks requireing a global memory could be replaced by the message passing coordination used manually in the pipeline architecture. I argue that not all applications need concurrent access on the state, and therefore, need a shared memory. Specifically, I argue that each state region remains roughly local to a stage during its modification. *

⚠

TODO review that, I don't know how to formulate these paragraphs. Identify the state and the data in the global memory.

2.3.2.3 Transformation

This equivalence should allow the transformation of an event loop into several parallel processes communicating by messages. In this thesis, I study the static transformation of a program, but the equivalence should also hold for a dynamic transformation. I present the analyzis tools I developed to identify the state and the data from the global memory.

With this compiler, it would be possible to express an application with a global memory, so as to follow the design principles of software development. And yet, the execution engine could adapt itself to any parallelism of the computing machine, from a single core, to a distributed cluster.

TODO too fast on the end of this section

TODO Transition to the chapter State of the Art

Chapter 3

State of the art

3.1 Javascript

3.1.1 Overview of the language

3.1.1.1 Functions as First-Class citizens

3.1.1.2 Lexical Scoping

3.1.1.3 Closure

3.2 Concurrency

3.2.1 Two known concurrency model

3.2.1.1 Thread

3.2.1.2 Event

3.2.1.3 Orthogonal concepts

3.2.2 Differentiating characteristics

3.2.2.1 Scheduling

3.2.2.2 Coordination strategy

3.2.3 Turn-based programming

3.2.3.1 Event-loop

3.2.3.2 Promises

3.2.3.3 Generators

3.2.4 Message-passing / pipeline parallelism -> DataFlow programming ?

3.3 Scalability

3.3.2 Scalability outside computer science (only if I have time)

If I have time, I would like to try to explain why scalability is at the core of material engagement and information theory, and is at the core of our universe : the propagation of Gravity wave is an example : it is impossible to scale

3.4 Framworks for web application distribution

3.4.1 Micro-batch processing

3.4.2 Stream Processing

3.5 Flow programming

3.5.1 Functional reactive programming

3.5.2 Flow-Based programming

3.6 Parallelizing compilers

OpenMP and so on

3.7 Synthesis

There is no compiler focusing on event-loop based applications

Chapter 4

Fluxion

4.1 Fluxionnal Compiler

Some parts of this are already written in the first paper. It needs a lot additional explanations and rewritting

4.1.1 Identification

4.1.1.1 Continuation and listeners

4.1.1.2 Dues

4.1.2 Isolation

4.1.2.1 Scope identification

Scope leaking

4.1.2.2 Execution and variable propagation

4.1.3 distribution

4.2 Fluxionnal execution model

Everything here is already written in the first paper : flx-paper. It only needs to be rewritten

4.2.1 Fluxion encapsulation

4.2.1.1 Execution

4.2.1.2 Name

4.2.1.3 Memory

4.2.2 Messaging system

Chapter 5

Evaluation

5.1 Due compiler

5.2 Fluxionnal compiler

5.3 Fluxionnal execution model

Chapter 6

Conclusion

Appendix A

Language popularity

A.1 PopularitY of Programming Languages (PYPL)

[1] The PYPL index uses Google trends[2] as a leading indicator of the popularity of a programming language. It search for the trend for each programming language by counting the number of searches of this language and the word "tutorial".

PYPL for May 2015

[1] http://pypl.github.io/PYPL.html
[2] https://www.google.com/trends/

Rank	Change	Language	Share	Trend
1		Java	24.1%	-0.9%
2		PHP	11.4%	-1.6%
3		Python	10.9%	+1.3%
4		C#	8.9%	-0.7%
5		C++	8.0%	-0.2%
6		C	7.6%	+0.2%
7		Javascript	7.1%	-0.6%
8		Objective-C	5.7%	-0.2%
9		Matlab	3.1%	+0.1%
10	2× ↑	R	2.8%	+0.7%
11	5× ↑	Swift	2.6%	+2.9%
12	1× ↓	Ruby	2.5%	+0.0%
13	3× ↓	Visual Basic	2.2%	-0.6%
14	1× ↓	VBA	1.5%	-0.1%
15	1× ↓	Perl	1.2%	-0.3%
16	1× ↓	lua	0.5%	-0.1%

A.2 TIOBE

[3]

The TIOBE index uses many search engines as an indicator of the current popularity of programming languages. It count the number of pages each search engine find when queried with the language name and the word "programming". This indicator indicates the number of resources available, and the discussions about a given programming language.

Javascript was the most rising language of 2014 in the TIOBE index.

TIOBE for April 2015

[3] http://www.tiobe.com/index.php/content/paperinfo/tpci/index.html

Apr 2015	Apr 2014	Change	Programming Language	Ratings	Change
1	2	↑	Java	16.041%	-1.31%
2	1	↓	C	15.745%	-1.89%
3	4	↑	C++	6.962%	+0.83%
4	3	↓	Objective-C	5.890%	-6.99%
5	5		C#	4.947%	+0.13%
6	9	↑	JavaScript	3.297%	+1.55%
7	7		PHP	3.009%	+0.24%
8	8		Python	2.690%	+0.70%
9	-	2× ↑	Visual Basic	2.199%	+2.20%

A.3 Programming Language Popularity Chart

[4]

The programming language popularity chart indicates the activity of a given language in the online communities. It uses two indicators to rank languages : the number of line changed in github of, and the number of questions tagged with a certain language.

Javascript is ranked number one in this index. The Javascript community is particularly active online, and in the open source.

indeed.com

A.4 Black Duck Knowledge

[5]

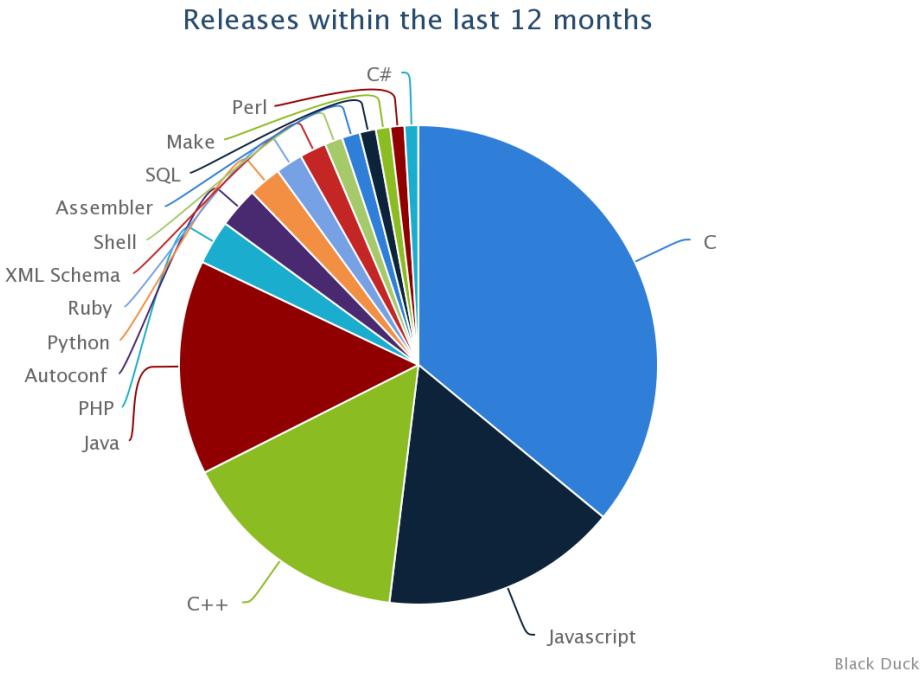
The black-duck, which analyze the usage of language on many forges, and collaborative hosts, rank Javascript number 2, after C, and with about the same usage as C++.

github.com sourceforge.net cpan.org rubyforge7.org planetsourcecode.com ddj.com

[4] http://langpop.corger.nl

[5] https://www.blackducksoftware.com/resources/data/this-years-language-use

Language	%
C	34.80
Javascript	15.45
C++	15.13
Java	14.02
PHP	2.87
Autoconf	2.65
Python	2.15
Ruby	1.77
XML Schema	1.73
Shell	1.18
Assembler	1.16
SQL	1.07
Make	0.94
Perl	0.92
C#	0.90



A.5 Github

http://githut.info/

A.6 HackerNews Poll

https://news.ycombinator.com/item?id=3746692

Language	Count
Python	3335
Ruby	1852
JavaScript	1530
C	1064
C#	907
PHP	719
Java	603
C++	587
Haskell	575
Clojure	480
CoffeeScript	381
Lisp	348
Objective C	341
Perl	341
Scala	255
Scheme	202
Other	195
Erlang	171
Lua	150
Smalltalk	130
Assembly	116
SQL	112
Actionscript	109
OCaml	88
Groovy	83
D	79
Shell	76
ColdFusion	51
Visual Basic	47
Delphi	45
Forth	41
Tcl	34
Ada	29
Pascal	28
Fortran	26
Rexx	13
Cobol	12