Comparing a Tree-walk Interpreter with JIT compilation and embedding via Go-plugins

Evaluating the trade-offs of using the Go-plugin API for JIT compilation while comparing the approach with a Tree-walk interpreter

9525469

April 25, 2024

Contents

The goal of this paper is to evaluate whether the usage of the Go plugin API is feasible for just-in-time compilation of a query language intended for a high performance in memory data storage. This evaluation is done based upon the criteria of the ease of usability, performance and the robustness of the resulting implementation. For the sake of comparison the query language as well as its features are introduced. The code a JIT would have generated for several heavy queries is handwritten and benchmarked against the same expressions evaluated with the currently employed tree walk interpreter. The paper explores the different possibilities for accessing the Go compiler, working with the Go plugin API and highlights several benchmarks comparing the performance of the new JIT compiler and the previous language evaluation implementation.

Das Ziel dieser Arbeit besteht darin zu evaluieren ob sich das Einsetzen der Go plugin API für die JIT Kompilation einer Abfragesprache einer effizienten Datenbank, die ihre Daten ausschließlich im Arbeitsspeicher ablegt, rentiert. Diese Abwägung basiert auf den Kriterien der Verwendbarkeit, Effizienz und der Stabilität der angewendeten Lösung. Um den Vergleich unter realistischen Bedingungen zu testen wird vorerst die Abfragesprache sowie einige umfangreicher werdende Beispiele vorgestellt. Diese Abfragen werden dann per Hand zum Go code kompiliert den der JIT erzeugt hätte, dieses Ergebnis wird dann gegen die derzeitige Implementierung die das Evaluieren mit einem Tree-walk Interpreter umsetzt, verglichen. Auch werden unterschiedliche Art und Weisen des Aufrufes des Go compiler, die Verwendung des Go plugin paketes und Tests auf die Effizienz des neuen JIT und des alten Tree-walk Interpreter vorgestellt.

1 Introduction

The query language is the singular interface for accessing, reading, creating and removing data in a database. This requires the query language to provide a high degree of performance in the sense of performing processing intensive queries in a fast enough time for real-time responsiveness, especially for an in memory data store with the aspiration for high performance.

Optimisations for database query languages are common, such as in the embeddable Database *SQLite* with the *SQLite Query Optimizer*[1] and the *Next-Generation Query Planner*[1] both supporting a variety of optimisations after compiling SQL expressions to byte-code instead of walking the AST[1][2]. *PostgreSQL* is an other example of a database optimising its SQL queries, using a JIT compiler[2][4].

There are several optimisations applicable for any programming language and query languages in particular [5, 3.3 Optimisations]. Some were already applied to the query language [6]. Implementing a JIT compiler can significantly improve the performance of a long running highly processing intensive program and allows it to outperform optimisations applied to the AST before walking it or compiling to bytecode instructions and executing them in a dedicated virtual machine [3, 4. Results].

The implementation of a JIT compiler introduces complexity into the codebase for the generation of optimized assembly is inherently complex and time consuming [7, 1. Introduction] while also opening the door for potential attacks [8, 2. Challenges Securing JavaScript JIT] [8, Table 1.]. Another issue is the platform dependence, a JIT compiler has to generate native code for all platforms and operating systems the database and therefore the query language support - increasing complexity and maintenance [9, Abstract]. Furthermore a JIT compiler significantly increases the memory usage [10, Fig. 1.] of the interpreter as well as requiring a not to be disregarded startup time made up of the time to start the JIT compiler and the time to compile the language constructs to machine code. [9, 4.2.7 Breakdown of compilation times].

To evaluate whether the aforementioned positive aspects outweigh the negative points and a JIT therefore does improve the query languages performance, determines the scope of this paper. The examined performance is measured on a fork of the original language project [11], this fork was necessary to pave the way for just in time compilation [12].

---

[1]Abstract Syntax Tree: tree of syntax nodes
[2]Just in time compiler: compiling methods on demand while a program is running [3, 1. Introduction]

2 Query Language

The query language is generic by design and by usage of the newly introduced generic proposal [13]. This allows for rudimentary and complex language variations depending on the data type the desired language is created with. Such as complexity ranging from boolean algebra (see Listing 2.1 and its corresponding language definition Listing 2.2), over arithmetic operations to complex queries on lists of objects, such as filtering for values, mapping and mutating elements of the list and iterating over the entries of the list.

```
true | false
a & !b
let c=true;
if c then
    a&b
else
    a|b
```

Listing 2.1: Boolean algebra

```
var boolParser = funcGen.New[bool]()
    .AddConstant("false", false)
    .AddConstant("true", true)
    .AddSimpleOp("^", true,
            func(a, b bool) (bool, error) { return a != b, nil })
    .AddSimpleOp("=", true,
            func(a, b bool) (bool, error) { return a == b, nil })
    .AddSimpleOp("|", true,
            func(a, b bool) (bool, error) { return a || b, nil })
    .AddSimpleOp("&", true,
            func(a, b bool) (bool, error) { return a && b, nil })
    .AddUnary("!", func(a bool) (bool, error) { return !a, nil })
    .SetToBool(func(c bool) (bool, bool) { return c, true })
```

Listing 2.2: Generating boolean algebra

2.1 Feature set

To limit the scope of performance evaluation, the feature set is determined as the language definition and dialect[1], specified in the `value` package of the `parser2` [12, `value` package] project. This dialect allows for aggregation, filtering and mutating large lists of data sets consisting of objects with many key value pairs.

```
persons
    .map(p -> p.Name)
    .reduce((a,b) -> a+", "+b)
```

Listing 2.3: Reducing and formatting objects in lists - source [12]

```
persons
  .accept(p -> p.PlaceOfBirth="New York" & p.Age>21)
  .map(e -> e.Name+": "+e.Age)
  .reduce((a,b) -> a+", "+b)
```

Listing 2.4: Applying filter, mapping, join entries with comma - source [12]

Listing 2.3 and Listing 2.4 access the global `person` constant using the dialect specific facilities for creating a list of objects of type `Person` as global state (see Listing 8.1).

```
func namedFunction(argument) argument+2;
let m = (argument) -> argument+2;
let add = (a,b) -> a+b;
```

Listing 2.5: Named Functions and unnamed/anonyoums functions - source [12]

The just in time compiler targets named functions and anonyoums functions (see Listing 2.5) and will attempt to compile their contents when conditions for their compilation are hit.

2.2 Current Evaluation Approach

---

[1]Refers to the usage of [12] to define and generate a language for a specific data type, with static global functions and constant values

3 Benchmarks

- jit will use and produce handwritten non optimized go code the jit compiler would generate
- benchmarks will be done by three examples:
    - a arithmetics heavy leetcode problem
    - a string concatenating heavy problem
    - an object and array index heavy problem, which will be the most realistic

3.1 Test inputs

3.1.1 Arithmetics

3.1.2 String operations

3.1.3 Object and Array Index

3.2 Considerations

Benchmark results may be subject to various influences, including the workload on the system conducting the benchmarks, insufficiently sized data sets for comprehensive and accurate testing, as well as comparing inherently dissimilar benchmarks. Most of the aforementioned can be mitigated by using the `testing` package included in the go programming languages standard library [14].

To avoid the influences of the current workload of the system the tests are performed many times, therefore accounting for statistical outliers and external system influences on the test, which furthermore accounts for the possibility of choosing insufficiently sized data sets.

By benchmarking the execution of the same given input with the JIT-compiler enabled and with the JIT-compiler disabled an inherently comparable data set is created due to the shared purpose of evaluating the runtime performance under differing runtime configurations. Therefore, conducting benchmarks under the two previously mentioned configurations and comparing the results is a valid evaluation of the two benchmarks.

3.3 Tree-Walk interpreter

3.4 Go just-in-time compiler

4 Compiler Invocation

Instead of manually generating optimized assembly for each function to be compiled on the fly, the go compiler toolchain is invoked and receives the previously generated Go source code that was computed from the query languages AST. This omits the complexity of implementing and maintaining several platform specific machine code generator compiler backends while allowing the JIT compiler to support all platforms supported by the go toolchain.

The method of invoking the compiler tool-chain has significant effects on the startup performance, the robustness and the complexity of the JIT compiler. This chapter highlights two possible approaches for invoking the compiler tool-chain.

4.1 Including the Compiler Source Code

The first idea of invoking the compiler tool-chain, is to include the source code of the compiler as a library and simply start it while passing in the generated go code. This does not require the compiler tool-chain to exist on the target system and omits the overhead of starting the compiler process. However this approach can not be used for the source code since the go compiler is not stable nor accessible outside of the go compiler tool-chain [15, *(gcToolchain).gc*] due to the usage of `internal` packages [16].

4.2 Invoking the local Go Compiler

The remaining method is to start the locally available compiler tool-chain via the `exec.Cmd` interface [17, Overview]. This enables requesting the operating system to invoke the compiler. Approaching the problem with this method has the downside of requiring the compiler to exist on the target system, the overhead of tasking the operating system with starting the compiler, writing the generated code to a temporary file and compiling this temporary file instead of doing all of the aforementioned inside of the JIT by including the compiler as a library as introduced before.

Listing 4.1 shows a simplified implementation of a function invoking the go compiler. Error handling is omitted for the sake of simplicity.

```go
func invokeCompiler(code string) {
    f, _ := os.CreateTemp(".", "jit_*.go")
    defer os.Remove(f)
    f.WriteString(code)
    c := exec.Command("go", "build", f.Name())
    c.Run()
}
```

Listing 4.1: Tool-chain invocation

5 Plugin API

The `plugin` package enables the loading of shared objects compiled with the `-buildmode=plugin` compiler flag and the resolution of symbols contained in the plugin [18, Overview]. This allows the compilation and loading of go code while running a program.

5.1 Compiling Go Source Code to Go plugins

As introduced above the compiler tool-chain accepts different build modes via the `-buildmode` command line argument [19]. The build mode for compiling a given source file to the go plugin format is named `plugin` [19] [18, Overview].

Listing 5.1 contains a modified version of Listing 4.1, adding the compiler flags for compiling the generated source code passed via the `code` function parameter, to a go plugin. Instead of producing an executable for the target architecture and operating system the compiler now generates a shared object in the format the `plugin` package requires.

```go
func invokeCompiler(code string) {
    f, _ := os.CreateTemp(".", "jit_*.go")
    defer os.Remove(f)
    f.WriteString(code)
    pre := strings.TrimSuffix(f.Name(), ".go")
    c := exec.Command(
        "go", "build", "-buildmode=plugin", "-o", pre, f.Name())
    c.Run()
}
```

Listing 5.1: Tool-chain invocation with plugin compilation

5.2 Embedding Go plugins

The loading of plugins and the resolution of plugins uses the API exposed by the previously introduced `plugin` package.

Listing 5.2 modifies Listing 5.1 for opening the previously compiled plugin. Once opened the `*plugin.Plugin` structure can be used for resolving exported functions

and variables included in the plugin. After resolving a symbol[1] its type is `any`, therefore the function `Main`[2] has to be cast to `func()` before the go type system allows a function call. Upon type casting the function is called and the generation, compilation and calling workflow of the JIT compiler is concluded.

```go
func invokeCompiler(code string) {
    f, _ := os.CreateTemp(".", "jit_*.go")
    defer os.Remove(f)
    f.WriteString(code)
    pre := strings.TrimSuffix(f.Name(), ".go")
    c := exec.Command(
        "go", "build", "-buildmode=plugin", "-o", pre, f.Name())
    c.Run()

    plug, _ := plugin.Open(pre)
    // assumes generated code lives in func Main()
    symbol, _ := plug.Lookup("Main")
    Main, _ := symbol.(func())
    Main()
}
```

Listing 5.2: Plugin compilation, plugin opening and function resolution

5.3 Trade-offs, Issues and Considerations

The `plugin` package provides the program with the unique ability to allow for high performance on the fly code compilation and execution. It therefore fits the use case of a query language implementation well.

However the `plugin` package bears several downsides [18, Warnings], primarily the missing portability due to the package only supporting Linux, FreeBSD and MacOS. Another disadvantage is the strict requirement of both the host application and all plugins needing to be compiled with the same tool-chain version and build-tags - this is particularly difficult in the case of this JIT, due to the requirement of the existence of the local compiler that will most certainly not be of the exact same version as the compiler used for compiling the host application. Is the previously mentioned not strictly ensured runtime errors can occur. A further drawback is the increased difficulty of reasoning about program and plugin initialisation for the special `func init()` function is called upon opening a plugin [18, Overview], possibly opening the program up to race conditions and similar critical bugs due to global state initialisation [20, The init function].

---

[1]a symbol refers to a function, constant or variable
[2]for the sake of this explanation the generated code in the function parameter `code` is assumed to be contained in the `Main` function

6 Just in Time Compilation

Just in time compilation refers to the process of determining whether a segregated chunk of code is considered "hot"[1] and compiling this code segment into operating system and architecture specific machine code. This machine code is then loaded into the memory of the interpreters runtime and executed instead of interpreting the code chunk [22]. The details of just in time compilation, meta tracing, categorizing code segments as "hot", improving the performance of the just in time compiler and error handling are explored in this chapter.

Contrary to the previously introduced definition of a just in time compiler in the context of programming language interpreters, go does not support dynamically loading machine code into memory and executing these chunks. The mitigation for this is introduced and explained in chapter 5.

6.1 Meta-tracing & `JIT_CONSTANT`

```go
// Function represents a function in the interpreter runtime
type Function[V any] struct {
    // ...

    // Counter stores the amount of calls made to the function
    Counter int
}
```

Listing 6.1: `Function[V any] struct` type with meta-tracing

Meta-tracing refers to the process of tracking the actions of the programming language interpreter [23, 4.1 Meta-tracing]. The interpreter uses this functionality to determine the amount of invocations of a function and updates the `Function.Counter` field accordingly, see Listing 6.1. Once this counter reaches the threshold defined in the `JIT_CONSTANT` (see Listing 6.2) the `type Function[V any] struct` instance is passed to the just in time compiler compilation queue, in which it will be compiled with other functions waiting to be compiled. Upon the Function being compiled the interpreter executes the output of the just in time compiler for each function invocation instead of walking the abstract syntax tree and thus is no longer interpreting the function, but instead uses the compiled representation.

---

[1]hot in the context of just in time compilation refers to a code path or a segment of code that is executed massive amount of times [21], [22]

```go
// JIT_CONSTANT sets the threshold the function invocation meta tracing counter
// has to pass for the function to be considered hot and thus compilable
var JIT_CONSTANT int = 1_000
```

Listing 6.2: `JIT_CONSTANT` definition

The JIT-compiler requires some information about a function before it's being able to start the code generation step. Not only does it require the name of the function[2], but the names of its arguments and the types the JIT can use to compile. The necessary fields are stored on the previously introduced `type Function[V any] struct`, see Listing 6.3.

```go
type MetaDataParameter struct {
        Name string
        Type string
}
type MetaData struct {
        Parameters []MetaDataParameter
}
// Function represents a function in the interpreter runtime
type Function[V any] struct {
    // ...

    // ArgumentNames contains the list of parameter names of the function
    ArgumentNames []string
    // Name holds the name of the function
    Name string
    // MetaData holds the necessary data for the jit to compile valid functions
    MetaData *MetaData
}
```

Listing 6.3: `Function[V any] struct` type with meta data

6.2 Function Parameters

6.3 Type System Clashes

6.4 Reducing Compiler Invocations

6.5 Bailing out to the Interpreter

---

[2]Unnamed/anonymous functions or closures are named by prefixing a closure counter with `c`, the first encountered closure will therefore be compiled as `func c0()`

7 Conclusion

7.1 Usability and Robustness

7.2 Performance

7.3 Implementation Complexity

Bibliography

[1]   "The sqlite query optimizer overview." (), [Online]. Available: `https://www.sqlite.org/optoverview.html` (visited on 01/20/2024).

[2]   "The next-generation query planner." (2023), [Online]. Available: `https://www.sqlite.org/queryplanner-ng.html` (visited on 01/20/2024).

[3]   A. Krall, "Efficient javavm just-in-time compilation," in *Proceedings. 1998 International Conference on Parallel Architectures and Compilation Techniques (Cat. No. 98EX192)*, IEEE, 1998, pp. 205–212. [Online]. Available: `https://www.complang.tuwien.ac.at/andi/papers/pact_98.pdf`.

[4]   "Chapter 32. just-in-time compilation (jit)." (), [Online]. Available: `https://www.postgresql.org/docs/current/jit-reason.html` (visited on 01/20/2024).

[5]   T. Suganuma, T. Yasue, and T. Nakatani, "A region-based compilation technique for a java just-in-time compiler," in *Proceedings of the ACM SIGPLAN 2003 conference on Programming Language Design and Implementation*, 2003, pp. 312–323.

[6]   "Parser2.optimizer." (2023), [Online]. Available: `https://github.com/hneemann/parser2/blob/361d9bd2f6daf0ff76b47a1186ed2cfcd1a96d12/parser2.go#L23` (visited on 01/20/2024).

[7]   K. Ishizaki, M. Kawahito, T. Yasue, M. Takeuchi, *et al.*, "Design, implementation, and evaluation of optimizations in a just-in-time compiler," in *Proceedings of the ACM 1999 conference on Java Grande*, 1999, pp. 119–128.

[8]   L. Bernhard, T. Scharnowski, M. Schloegel, T. Blazytko, and T. Holz, "Jit-picking: Differential fuzzing of javascript engines," in *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*, 2022, pp. 351–364.

[9]   K. Ishizaki, M. Takeuchi, K. Kawachiya, T. Suganuma, *et al.*, "Effectiveness of cross-platform optimizations for a java just-in-time compiler," in *Proceedings of the 18th annual ACM SIGPLAN conference on Object-oriented programing, systems, languages, and applications*, 2003, pp. 187–204.

[10] J. Aycock, "A brief history of just-in-time," *ACM Computing Surveys (CSUR)*, vol. 35, no. 2, pp. 97–113, 2003.

[11] H. Neemann, *Parser2*, 2024. [Online]. Available: `https://github.com/hneemann/parser2` (visited on 04/16/2024).

[12] xnacly and dani0611, *Parser2*, 2024. [Online]. Available: `https://github.com/xnacly/parser2` (visited on 04/16/2024).

[13] I. L. Taylor. "Type parameters proposal." (2021), [Online]. Available: `https://go.googlesource.com/proposal/+/HEAD/design/43651-type-parameters.md` (visited on 04/02/2024).

[14] "Go testing package." Official documentation for the Go testing package. (), [Online]. Available: `https://pkg.go.dev/testing` (visited on 04/16/2024).

[15] "Gc.go." (2023), [Online]. Available: `https://github.com/golang/go/blob/master/src/cmd/go/internal/work/gc.go#L58` (visited on 01/20/2024).

[16] "Internal directories." (2024), [Online]. Available: `https://pkg.go.dev/cmd/go#hdr-Internal_Directories` (visited on 01/20/2024).

[17] "Os/exec package." (2024), [Online]. Available: `https://pkg.go.dev/os/exec` (visited on 01/20/2024).

[18] "Plugin package." (2024), [Online]. Available: `https://pkg.go.dev/plugin` (visited on 01/20/2024).

[19] "Build modes." (2024), [Online]. Available: `https://pkg.go.dev/cmd/go#hdr-Build_modes` (visited on 01/20/2024).

[20] "Effective go - the go programming language." (2024), [Online]. Available: `https://go.dev/doc/effective_go` (visited on 01/20/2024).

[21] "-xjit / -xnojit - ibm documentation." (Feb. 12, 2024), [Online]. Available: `https://www.ibm.com/docs/en/sdk-java-technology/8?topic=options-xjit#optlevel` (visited on 04/16/2024).

[22] A. Krall, "Efficient javavm just-in-time compilation," in *Proceedings. 1998 International Conference on Parallel Architectures and Compilation Techniques (Cat. No. 98EX192)*, IEEE, 1998, pp. 205–212.

[23] C. F. Bolz and L. Tratt, "The impact of meta-tracing on vm design and implementation," *Science of Computer Programming*, vol. 98, pp. 408–421, 2015.

8 Appendix

```go
type Person struct {
        Name         string
        Surname      string
        PlaceOfBirth string
        Age          int
}

var Persons = []Person{
        {"John", "Doe", "London", 23},
        {"Jane", "Doe", "London", 25},
        {"Bob", "Smith", "New York", 21},
        {"Frank", "Muller", "New York", 22},
        {"Mary", "Green", "Seattle", 21},
        {"Jake", "Muller", "Washington", 22},
}

var PersonToMap = value.NewToMapReflection[Person]()
var persons = value.NewListOfMaps[Person](PersonToMap, Persons)
```

Listing 8.1: **Passing Go values into the language context - source [12]**