Languages + Build Systems

Martin Kellogg

Languages and Build Systems

Today's agenda:

- **Finish slides on Languages**
  - paradigms, type systems, multilanguage projects, performance, team and process factors, when to rewrite
- What is a build system? How does one work?
- How to choose a build system + best practices

Functional advantages

Functional advantages

- Tractable program semantics
  - Procedures are functions (simplifies reasoning)
  - Formulate and prove assertions about code more easily
  - More readable (if you like math)

Functional advantages

- Tractable program semantics
  - Procedures are functions (simplifies reasoning)
  - Formulate and prove assertions about code more easily
  - More readable (if you like math)
- ***Referential transparency***
  - Replace any expression by its value without changing the result

Functional advantages

- Tractable program semantics
  - Procedures are functions (simplifies reasoning)
  - Formulate and prove assertions about code more easily
  - More readable (if you like math)
- ***Referential transparency***
  - Replace any expression by its value without changing the result
- “No” side-effects
  - Fewer errors

Functional disadvantages

Functional disadvantages

- Efficiency
  - Copying takes time

Functional disadvantages

- Efficiency
  - Copying takes time

Language	Speed	Space
C (gcc)	1.0	1.1
C++ (g++)	1.0	1.6
OCaml	1.5	2.9
Java (JDK -server)	1.7	9.1
Lisp	1.7	11
C# (mono)	2.4	5.6
Python	6.5	3.9
Ruby	16	5.0

17 small benchmarks

Functional disadvantages

- Efficiency
  - Copying takes time
- Compiler implementation
  - Frequent memory allocation

Language	Speed	Space
C (gcc)	1.0	1.1
C++ (g++)	1.0	1.6
OCaml	1.5	2.9
Java (JDK -server)	1.7	9.1
Lisp	1.7	11
C# (mono)	2.4	5.6
Python	6.5	3.9
Ruby	16	5.0

17 small benchmarks

Functional disadvantages

- Efficiency
  - Copying takes time
- Compiler implementation
  - Frequent memory allocation
- Unfamiliar (to you, and maybe those you're hiring!)
  - New programming style

Language	Speed	Space
C (gcc)	1.0	1.1
C++ (g++)	1.0	1.6
OCaml	1.5	2.9
Java (JDK -server)	1.7	9.1
Lisp	1.7	11
C# (mono)	2.4	5.6
Python	6.5	3.9
Ruby	16	5.0

17 small benchmarks

Functional disadvantages

- Efficiency
  - Copying takes time
- Compiler implementation
  - Frequent memory allocation
- Unfamiliar (to you, and maybe those you're hiring!)
  - New programming style
- Not appropriate for every program
  - Some programs are inherently stateful

Language	Speed	Space
C (gcc)	1.0	1.1
C++ (g++)	1.0	1.6
OCaml	1.5	2.9
Java (JDK -server)	1.7	9.1
Lisp	1.7	11
C# (mono)	2.4	5.6
Python	6.5	3.9
Ruby	16	5.0

17 small benchmarks

Object-oriented programming

Definition: in the ***object-oriented*** paradigm, programs are composed of interacting objects, each of which is responsible for some well-defined part of the program's state

Object-oriented programming

Definition: in the ***object-oriented*** paradigm, programs are composed of interacting objects, each of which is responsible for some well-defined part of the program's state

- underlying mathematical formalism:

Object-oriented programming

Definition: in the ***object-oriented*** paradigm, programs are composed of interacting objects, each of which is responsible for some well-defined part of the program's state

- underlying mathematical formalism: type systems? dictionaries?
    - still something of an open research problem

Object-oriented programming

Definition: in the ***object-oriented*** paradigm, programs are composed of interacting objects, each of which is responsible for some well-defined part of the program's state

- underlying mathematical formalism: type systems? dictionaries?
    - still something of an open research problem
- **extraordinarily common**

Object-oriented programming

Definition: in the ***object-oriented*** paradigm, programs are composed of interacting objects, each of which is responsible for some well-defined part of the program's state

- underlying mathematical formalism: type systems? dictionaries?
    - still something of an open research problem
- **extraordinarily common**
- models **the real world** well
    - objects are good abstractions for real-world entities and concepts

Object-oriented programming gotchas

- classes vs prototypes

Object-oriented programming gotchas

- classes vs prototypes
  - a ***class*** is a template for building objects (but is not itself an object!)
  - a ***prototype*** is an object that is used as a template for building other objects

Object-oriented programming gotchas

- classes vs prototypes
  - a ***class*** is a template for building objects (but is not itself an object!)
  - a ***prototype*** is an object that is used as a template for building other objects
- similar, but lead to **subtle differences**
  - prototypes can be modified at run time!

Object-oriented programming gotchas

- classes vs prototypes
  - a ***class*** is a template for building objects (but is not itself an object!)
  - a ***prototype*** is an object that is used as a template for building other objects
- similar, but lead to **subtle differenc**
  - prototypes can be modified at

Which of the two does Java use? What about JavaScript?

How can programming languages differ?

- programming paradigm
- **whether they have a type system**
  - and, if they do, what kind of type system they have
- library support
  - the standard library is especially important
- performance
- team/process factors
  - how well do you know the language
  - how easy it'll be to hire other developers who do

What is a type system, anyway?

What is a type system, anyway?

Definition: a ***type system*** is a set of rules that give every program element a ***type***, which is an upper bound on the set of possible values that that element can take on at run time

What is a type system, anyway?

Definition: a ***type system*** is a set of rules that give every program element a ***type***, which is an upper bound on the set of possible values that that element can take on at run time

- goal of a type system: **prevent errors** at run time due to unexpected values

What is a type system, anyway?

Definition: a ***type system*** is a set of rules that give every program element a ***type***, which is an upper bound on the set of possible values that that element can take on at run time

- goal of a type system: **prevent errors** at run time due to unexpected values
- **type theory** is the discipline of math (yes!) that studies the formal properties of type systems

What is a type system, anyway?

Definition: a ***type system*** is a set of rules that give every program element a ***type***, which is an upper bound on the set of possible values that that element can take on at run time

- goal of a type system: **prevent errors** at run time due to unexpected values
- **type theory** is the discipline of math (yes!) that studies the formal properties of type systems
- most programming languages include some kind of type system
  - exceptions: assembly, Lisp, a few others

Kinds of type systems

- Static vs dynamic checking

Kinds of type systems

- Static vs dynamic checking
    - ***statically typed*** languages have their types checked before the program runs, typically **at compile time**

Kinds of type systems

- Static vs dynamic checking
  - ***statically typed*** languages have their types checked before the program runs, typically **at compile time**
    - shares advantages/disadvantages with other static analyses

Kinds of type systems

- Static vs dynamic checking
  - ***statically typed*** languages have their types checked before the program runs, typically **at compile time**
    - shares advantages/disadvantages with other static analyses
  - ***dynamically typed*** languages have their types checked **at run time**, typically by a special interpreter or language runtime

Kinds of type systems

- Static vs dynamic checking
  - ***statically typed*** languages have their types checked before the program runs, typically **at compile time**
    - shares advantages/disadvantages with other static analyses
  - ***dynamically typed*** languages have their types checked **at run time**, typically by a special interpreter or language runtime
    - shares advantages/disadvantages with other dynamic analyses

Kinds of type systems

- Static vs dynamic checking
  - ***statically typed*** languages have their types checked before the program runs, typically **at compile time**
    - shares advantages/disadvantages with other static analyses
  - ***dynamically typed*** languages have their types checked **at run time**, typically by a special interpreter or language runtime
    - shares advantages/disadvantages with other dynamic analyses
- **Insight**: typechecking is just another program analysis

Static vs dynamic types

- Both are **common in practice**

Static vs dynamic types

- Both are **common in practice**
    - examples of each?

Static vs dynamic types

- Both are **common in practice**
  - examples of each?
    - Static: Java, C, Rust, OCaml, TypeScript, etc.
    - Dynamic: Python, Ruby, JavaScript, etc.

Static vs dynamic types

- Both are **common in practice**
  - examples of each?
    - Static: Java, C, Rust, OCaml, TypeScript, etc.
    - Dynamic: Python, Ruby, JavaScript, etc.
- **Ongoing debate** about the benefits

Static vs dynamic types

- Both are **common in practice**
  - examples of each?
    - Static: Java, C, Rust, OCaml, TypeScript, etc.
    - Dynamic: Python, Ruby, JavaScript, etc.
- **Ongoing debate** about the benefits
  - Benefits of static typing:
    - early detection of errors, types are documentation
  - Benefits of dynamic typing:
    - faster prototyping, no false positives

Other ways type systems differ

Other ways type systems differ

- **Implicit** vs **explicit**

Other ways type systems differ

- **Implicit** vs **explicit**
  - “do you write the types yourself”
  - almost all mainstream, static languages are explicit

Other ways type systems differ

- **Implicit** vs **explicit**
  - "do you write the types yourself"
  - almost all mainstream, static languages are explicit
- **Strength** of the type system
  - not all type systems can prove the same properties

Other ways type systems differ

- **Implicit** vs **explicit**
    - "do you write the types yourself"
    - almost all mainstream, static languages are explicit
- **Strength** of the type system
    - not all type systems can prove the same properties
    - e.g., Kotlin **guarantees no null-pointer dereferences**, but Java doesn't (both compile to Java bytecode)

Other ways type systems differ

- **Implicit** vs **explicit**
  - “do you write the types yourself”
  - almost all mainstream, static languages are explicit
- **Strength** of the type system
  - not all type systems can prove the same properties
  - e.g., Kotlin **guarantees no null-pointer dereferences**, but Java doesn’t (both compile to Java bytecode)
  - stronger types can be added to a language (**ask me more**)
    - “pluggable types”

How can programming languages differ?

- programming paradigm
- whether they have a type system
  - and, if they do, what kind of type system they have
- **library support**
  - the standard library is especially important
- performance
- team/process factors
  - how well do you know the language
  - how easy it’ll be to hire other developers who do

Library support

- **Key question**: do the right tools for the job you need to do exist in the language?

Library support

- **Key question**: do the right tools for the job you need to do exist in the language?

Remember: **Don't Repeat Yourself**
If someone else has already built what you need, don't build it again

Library support

- **Key question**: do the right tools for the job you need to do exist in the language?
- Tied to **language popularity**: languages that are more popular have better libraries, so people are more likely to use them
  - positive feedback loop!

Library support

- **Key question**: do the right tools for the job you need to do exist in the language?
- Tied to **language popularity**: languages that are more popular have better libraries, so people are more likely to use them
  - positive feedback loop!
- Common situation: you need library A and library B, but A is written in language L and B is written in language M
  - What to do?

Multi-language projects

- In a given project, not all code needs to be written in the same language!

Multi-language projects

- In a given project, not all code needs to be written in the same language!

Multi-language projects are common!
Developer quote: ““My last 4 jobs have been apps that called: Java from C#, and C# from F#; Java from Ruby; Python from Tcl, C++ from Python, and C from Tcl; Java from Python, and Java from Scheme (And that's not even counting SQL, JS, OQL, etc.)””

Multi-language projects

- In a given project, not all code needs to be written in the same language!
- Multi-language projects allow you to **choose the right language** for each part of your application

Multi-language projects

- In a given project, not all code needs to be written in the same language!
- Multi-language projects allow you to **choose the right language** for each part of your application

For example, concurrency might be better handled in F#/OCaml (immutable functional) or Ruby (designed to hide such details), while low-level OS or hardware access is much easier in C or C++, while rapid prototyping is much easier in Python or Lua, etc.

Multi-language projects

- In a given project, not all code needs to be written in the same language!
- Multi-language projects allow you to **choose the right language** for each part of your application
  - but **complicate** many parts of software engineering

Multi-language projects

- In a given project, not all code needs to be written in the same language!
- Multi-language projects allow you to **choose the right language** for each part of your application
  - but **complicate** many parts of software engineering
- Traditional architecture:

Multi-language projects

- In a given project, not all code needs to be written in the same language!
- Multi-language projects allow you to **choose the right language** for each part of your application
  - but **complicate** many parts of software engineering
- Traditional architecture:
  - Application **kernel** is written in a statically typed, optimized, compiled language

Multi-language projects

- In a given project, not all code needs to be written in the same language!
- Multi-language projects allow you to **choose the right language** for each part of your application
    - but **complicate** many parts of software engineering
- Traditional architecture:
    - Application **kernel** is written in a statically typed, optimized, compiled language
    - **Scripts** are written in a dynamically typed, interpreted language

Multi-language projects

- In a given project, not all code needs to be written in the same l
- N **anguage** for e

> **Examples**: Emacs (C / Lisp), Adobe Lightroom (C++ / Lua), NRAO Telescope (C / Python), Google Android (C / Java), most games (C++ / Lua), etc.

  - but **complicate** many parts of software engineering
- Traditional architecture:
  - Application **kernel** is written in a statically typed, optimized, compiled language
  - **Scripts** are written in a dynamically typed, interpreted language

Multi-language projects

C/C++ is a lingua franca

- In a given project, not all code needs to be written in the same l…
- N… anguage for e…
  - but complicate many parts of software engineering
- Traditional architecture:
  - Application kernel is written in a statically typed, optimized, compiled language
  - Scripts are written in a dynamically typed, interpreted language

Examples: Emacs (C / Lisp), Adobe Lightroom (C++ / Lua), NRAO Telescope (C / Python), Google Android (C / Java), most games (C++ / Lua), etc.

Multi-language projects

- Another common approach: ***common language infrastructure***
    - enables easy integration and interoperability

Multi-language projects

- Another common approach: ***common language infrastructure***
  - enables easy integration and interoperability
- Examples:
  - .NET framework (Microsoft)
    - C++, C#, J#, F#, Visual Basic, etc.
  - Java bytecode + Java virtual machine
    - Java, Scala, Kotlin, Closure, etc.
  - LLVM bytecode
  - etc.

Multi-language projects: complications

Multi-language projects: complications

- **Integrating data and control flow** across languages can be difficult

Multi-language projects: complications

- **Integrating data and control flow** across languages can be difficult
- **Debugging** can be harder
    - Especially as values flow and control flow from language A to language B

Multi-language projects: complications

- **Integrating data and control flow** across languages can be difficult
- **Debugging** can be harder
  - Especially as values flow and control flow from language A to language B
- **Build process** becomes more complicated

Multi-language projects: complications

- **Integrating data and control flow** across languages can be difficult
- **Debugging** can be harder
  - Especially as values flow and control flow from language A to language B
- **Build process** becomes more complicated
- **Developer expertise** is required in multiple languages
  - Must understand types (etc.) in **all** languages

Multi-language projects: complications

- **Integrating data and control flow** across languages can be difficult
- **Debugging** can be harder
  - Especially as values flow and control flow from language A to language B
- **Build process** becomes more complicated
- **Developer expertise** is required in multiple languages
  - Must understand types (etc.) in **all** languages
- Most **tools are language specific**: testing frameworks (+ generation, coverage, etc.), static analysis, build systems, debuggers, etc.

How can programming languages differ?

- programming paradigm
- whether they have a type system
  - and, if they do, what kind of type system they have
- library support
  - the standard library is especially important
- **performance**
- team/process factors
  - how well do you know the language
  - how easy it'll be to hire other developers who do

Language performance

- Three **main axes** to trade-off between languages:

Language performance

- Three **main axes** to trade-off between languages:
  - **Performance** (“how fast do programs run”)

Language performance

- Three **main axes** to trade-off between languages:
  - **Performance** (“how fast do programs run”)
  - **Safety** (“how easy is it to make mistakes”)

Language performance

- Three **main axes** to trade-off between languages:
  - **Performance** (“how fast do programs run”)
  - **Safety** (“how easy is it to make mistakes”)
  - Developer **Effort** (“how hard do I have to think to write a program in this language”)

Language performance

- Three **main axes** to trade-off between languages:
  - **Performance** (“how fast do programs run”)
  - **Safety** (“how easy is it to make mistakes”)
  - Developer **Effort** (“how hard do I have to think to write a program in this language”)
- Different languages choose different trade-offs. Examples:

Language performance

- Three **main axes** to trade-off between languages:
  - **Performance** (“how fast do programs run”)
  - **Safety** (“how easy is it to make mistakes”)
  - Developer **Effort** (“how hard do I have to think to write a program in this language”)
- Different languages choose different trade-offs. Examples:
  - Rust: good performance and safety, hard to write

Language performance

- Three **main axes** to trade-off between languages:
  - **Performance** (“how fast do programs run”)
  - **Safety** (“how easy is it to make mistakes”)
  - Developer **Effort** (“how hard do I have to think to write a program in this language”)
- Different languages choose different trade-offs. Examples:
  - Rust: good performance and safety, hard to write
  - Python: easy to write, okay safety, slow

Language performance

- Three **main axes** to trade-off between languages:
  - **Performance** (“how fast do programs run”)
  - **Safety** (“how easy is it to make mistakes”)
  - Developer **Effort** (“how hard do I have to think to write a program in this language”)
- Different languages choose different trade-offs. Examples:
  - Rust: good performance and safety, hard to write
  - Python: easy to write, okay safety, slow
  - C: good performance, easy-ish to write, very unsafe

What impacts performance

What impacts performance

- #1: **safety features enforced at run time**

What impacts performance

- #1: safety features enforced at run time
  - dynamic type checking: type safety
  - garbage collection: memory safety
  - exceptions: segfault safety

What impacts performance

- #1: safety features enforced at run time
  - dynamic type checking: type safety
  - garbage collection: memory safety
  - exceptions: segfault safety
- Also relevant: optimizations

What impacts performance

- #1: **safety features enforced at run time**
    - dynamic type checking: type safety
    - **garbage collection**: memory safety
    - exceptions: segfault safety
- Also relevant: **optimizations**
    - **interpreted** languages almost always slower: no optimizing compiler

What impacts performance

- #1: **safety features enforced at run time**
    - dynamic type checking: type safety
    - **garbage collection**: memory safety
    - exceptions: segfault safety
- Also relevant: **optimizations**
    - **interpreted** languages almost always slower: no optimizing compiler
    - JITs (***just-in-time compilers***) can produce surprisingly fast code
        - e.g., Java Virtual Machine

Trade-off: safety features

- #1 performance problem: **safety features enforced at run time**

Trade-off: safety features

- #1 performance problem: **safety features enforced at run time**
- So, why not **enforce safety at compile time** instead?

Trade-off: safety features

- #1 performance problem: **safety features enforced at run time**
- So, why not **enforce safety at compile time** instead?
    - requires **static analysis** (= there will be false positives)

Trade-off: safety features

- #1 performance problem: **safety features enforced at run time**
- So, why not **enforce safety at compile time** instead?
    - requires **static analysis** (= there will be false positives)
    - harder for programmers (trades off against **effort**)

Trade-off: safety features

- #1 performance problem: **safety features enforced at run time**
- So, why not **enforce safety at compile time** instead?
  - requires **static analysis** (= there will be false positives)
  - harder for programmers (trades off against **effort**)
    - the garbage collector in Java/Go/etc. is automatic
    - but writing Rust code requires follows its (complex) type discipline

Trade-off: safety features

- #1 performance problem: **safety features enforced at run time**
- So, why not **enforce safety at compile time** instead?
  - requires **static analysis** (= there will be false positives)
  - harder for programmers (trades off against **effort**)
    - the garbage collector in Java/Go/etc. is automatic
    - but writing Rust code requires follows its (complex) type discipline
  - bottom line: statically safe languages **can be faster**, but are **generally harder to program in**

How can programming languages differ?

- programming paradigm
- whether they have a type system
  - and, if they do, what kind of type system they have
- library support
  - the standard library is especially important
- performance
- **team/process factors**
  - how well do you know the language
  - how easy it'll be to hire other developers who do

Team/process factors

- **Learning** a new programming language takes time

Team/process factors

- **Learning** a new programming language takes time
  - Becoming productive shouldn't take that long
    - but, this scales with how hard the language is to program in (+ access to mentors, etc.)

Team/process factors

- **Learning** a new programming language takes time
  - Becoming productive shouldn't take that long
    - but, this scales with how hard the language is to program in (+ access to mentors, etc.)
  - Becoming an expert takes a long time!

Team/process factors

- **Learning** a new programming language takes time
  - Becoming productive shouldn't take that long
    - but, this scales with how hard the language is to program in (+ access to mentors, etc.)
  - Becoming an expert takes a long time!
- If you need performance, you usually need **at least one expert**
  - cf. AWS employs some JVM experts to tune the garbage collector for AWS services that use Java

Team/process factors

- Learning a new programming language takes time

> **Implication**: if you're going to need an expert, make sure you have one! This often seriously limits your choice of languages in practice :(

program

  - Becoming an expert takes a long time!
- If you need performance, you usually need **at least one expert**
  - cf. AWS employs some JVM experts to tune the garbage collector for AWS services that use Java

Team/process factors

- Because learning a new language takes time, the **popularity** of a language is also a plus:

Team/process factors

- Because learning a new language takes time, the **popularity** of a language is also a plus:
  - it's **easier to hire** new engineers who already know the language, and therefore can ramp up faster

Team/process factors

- Because learning a new language takes time, the **popularity** of a language is also a plus:
  - it's **easier to hire** new engineers who already know the language, and therefore can ramp up faster
  - but this impact is relatively small over a typical engineer's tenure at a company

Team/process factors

- Because learning a new language takes time, the **popularity** of a language is also a plus:
  - it's **easier to hire** new engineers who already know the language, and therefore can ramp up faster
  - but this impact is relatively small over a typical engineer's tenure at a company
- Implication: if all else is equal, **choose the more popular** language

When to rewrite

- the reading talked about moving a service from one language to another
  - why?

When to rewrite

- the reading talked about moving a service from one language to another
  - why? **Performance problems**.

When to rewrite

- the reading talked about moving a service from one language to another
  - why? **Performance problems**.
- This is usually a **risky thing** to do:
  - you’re not building new features
  - integration problems
  - will the benefits be worth it?

When to rewrite

- the reading talked about moving a service from one language to another
  - why? **Performance problems**.
- This is usually a **risky thing** to do:
  - you're not building new features
  - integration problems
  - will the b

Implication: rewriting is a good idea if you're confident that the benefits of the new language are worthwhile, but be cautious: it can expensive!

Takeaways

- there is a wider world of languages than just imperative and object-oriented (but those are the most popular)
  - learning to write functional code can make you a better programmer
- different programming languages have different trade-offs
  - performance vs safety vs ease of use vs …
- when starting a new project, think carefully about the requirements before choosing a language
- rewrite a project in a new language only after careful consideration

Reading Quiz: Build Systems

Q1: The “F5 key” in the title of the reading represent substituting ____________ for a proper build process?

A. shell scripts
B. the IDE
C. developer knowledge
D. testing

Q2: **TRUE** or **FALSE**: the author thinks that the length of time that it takes for you to get a new team member working productively on your project is a good measure of the health of a software project

Reading Quiz: Build Systems

Q1: The “F5 key” in the title of the reading represent substituting ____________ for a proper build process?

A. shell scripts
B. the IDE
C. developer knowledge
D. testing

Q2: **TRUE** or **FALSE**: the author thinks that the length of time that it takes for you to get a new team member working productively on your project is a good measure of the health of a software project

Reading Quiz: Build Systems

Q1: The "F5 key" in the title of the reading represent substituting ____________ for a proper build process?

A. shell scripts
B. the IDE
C. developer knowledge
D. testing

Q2: TRUE or **FALSE**: the author thinks that the length of time that it takes for you to get a new team member working productively on your project is a good measure of the health of a software project

Build Systems

Today's agenda:

- Finish slides on Languages
- **What is a build system? How does one work?**
- How to choose a build system + best practices

What does a developer do?

What does a developer do?

- Get the source code
- Install dependencies
- Compile the code
- Run static analysis
- Generate documentation
- Run tests
- Create artifacts for customers
- Ship!

What does a developer do?

- Get the source code
- Install dependencies
- Compile the code
- Run static analysis
- Generate documentation
- Run tests
- Create artifacts for customers
- Ship!

Which should be handled manually?

What does a developer do?

- Get the source code
- Install dependencies
- Compile the code
- Run static analysis
- Generate documentation
- Run tests
- Create artifacts for customers
- Ship!

Which should be handled manually?

NONE!

From the reading

“Here's how most clients I work with build a project:

1. Open the IDE
2. Load the solution
3. Get latest
4. Press F5 (or CTRL+SHIFT+B)”

From the reading

“Here's how most clients I wo

1. Open the IDE
2. Load the solution
3. Get latest
4. Press F5 (or CTRL+SHIF

“**The F5 key is not a build process**. It's a quick and dirty substitute. If that's how you build your software, I regret that I have to be the one to tell you this, but *your project is not based on solid software engineering practices.*”

From the reading

“Here's how most clients I wo

1. Open the IDE
2. Load the solution
3. Get latest
4. Press F5 (or CTRL+SHIF

“**The F5 key is not a build process**. It's a quick and dirty substitute. If that's how you build your software, I regret that I have to be the one to tell you this, but *your project is not based on solid software engineering practices.*”

Key objective of a build system: avoid this problem!

What is a build system?

What is a build system?

Definition: A ***build system*** is a tool for orchestrating software engineering tasks

What is a build system?

Definition: A ***build system*** is a tool for orchestrating software engineering tasks

- Getting the source code
- Installing dependencies
- Compiling the code
- Running static analysis
- Generating documentation
- Running tests
- Creating artifacts for customers
- Shipping!

What is a build system?

Definition: A ***build system*** is a tool for orchestrating software engineering tasks

- Getting the source code
- Installing dependencies
- Compiling the code
- Running static analysis
- Generating documentation
- Running tests
- Creating artifacts for customers
- Shipping!

A good build system handles all these

Tasks

Definition: a *task* is anything that the build system can do

Tasks

Definition: a *task* is anything that the build system can do

- Getting the source code
- Installing dependencies
- Compiling the code
- Running static analysis
- Generating documentation
- Running tests
- Creating artifacts for customers
- Shipping!

Tasks

Definition: a ***task*** is anything that the build system can do

- Getting the source code
- Installing dependencies
- Compiling the code
- Running static analysis
- Generating documentation
- Running tests
- Creating artifacts for customers
- Shipping!

All tasks!

Tasks

- #1 thing to know about tasks: **tasks are code**, too!

Tasks

- #1 thing to know about tasks: **tasks are code**, too!
    - Should be checked into version control
    - Should be code-reviewed
    - Should be tested

Tasks

- #1 thing to know about tasks: **tasks are code**, too!
  - Should be checked into version control
  - Should be code-reviewed
  - Should be tested
- Tasks also commonly have **dependencies**

Tasks

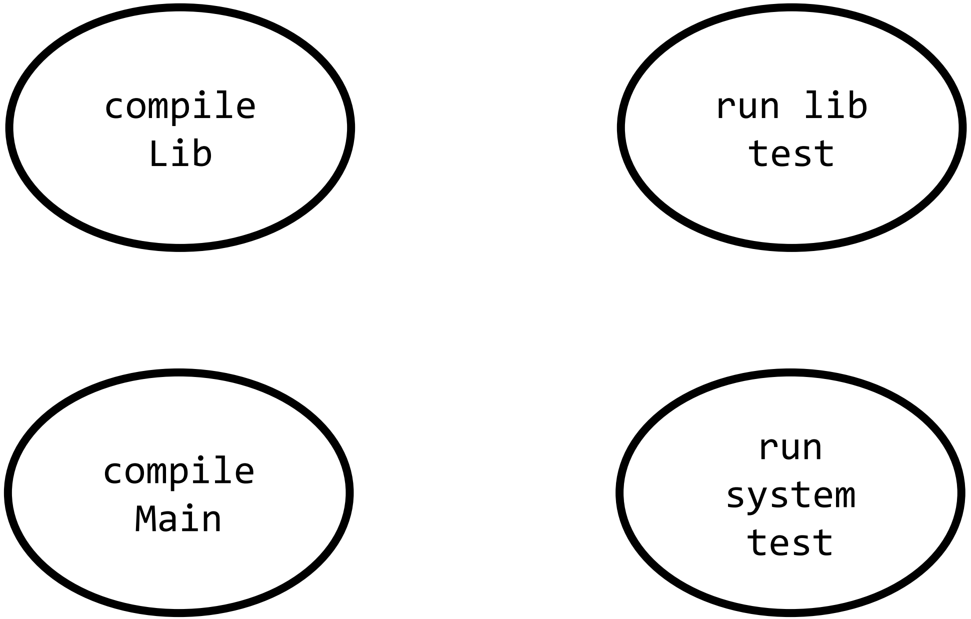
- #1 thing to know about tasks: **tasks are code**, too!
  - Should be checked into version control
  - Should be code-reviewed
  - Should be tested
- Tasks also commonly have **dependencies**
  - Dependency management is a key build system responsibility!

Dependencies between tasks

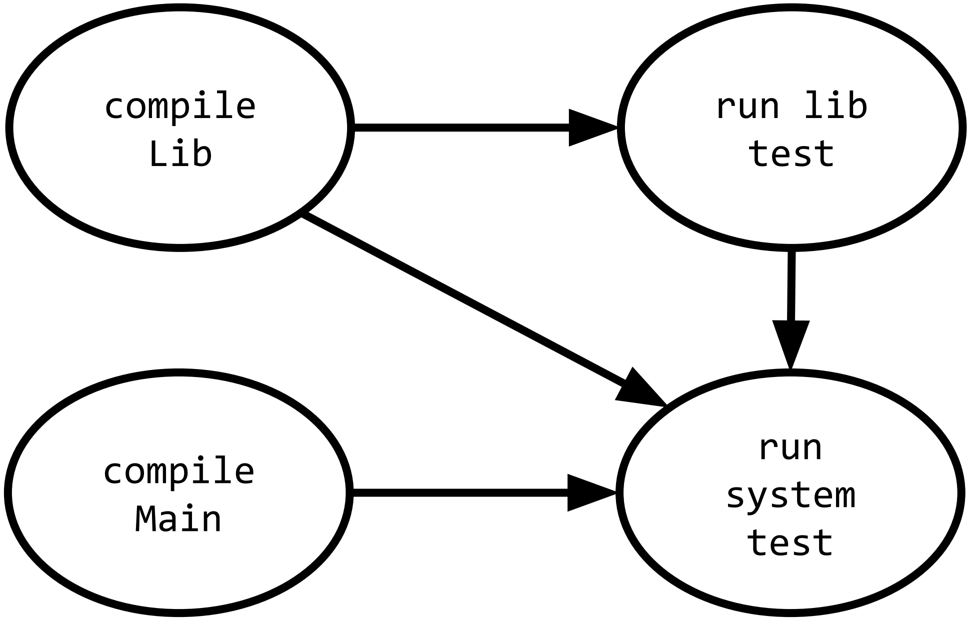
```
> ls src/

Lib.java   LibTest.java   Main.java    SystemTest.java
```

Dependencies between tasks



Dependencies between tasks



Dependencies between tasks

- A large project may have thousands of tasks

Dependencies between tasks

- A large project may have thousands of tasks
  - What order to run in?
  - How to speed up?

Dependencies between tasks

- A large project may have thousands of tasks
    - **What order to run in?**
    - How to speed up?

Determining task ordering

- Dependencies between tasks form a directed acyclic graph

Determining task ordering

- Dependencies between tasks form a directed acyclic graph

Topological sort!

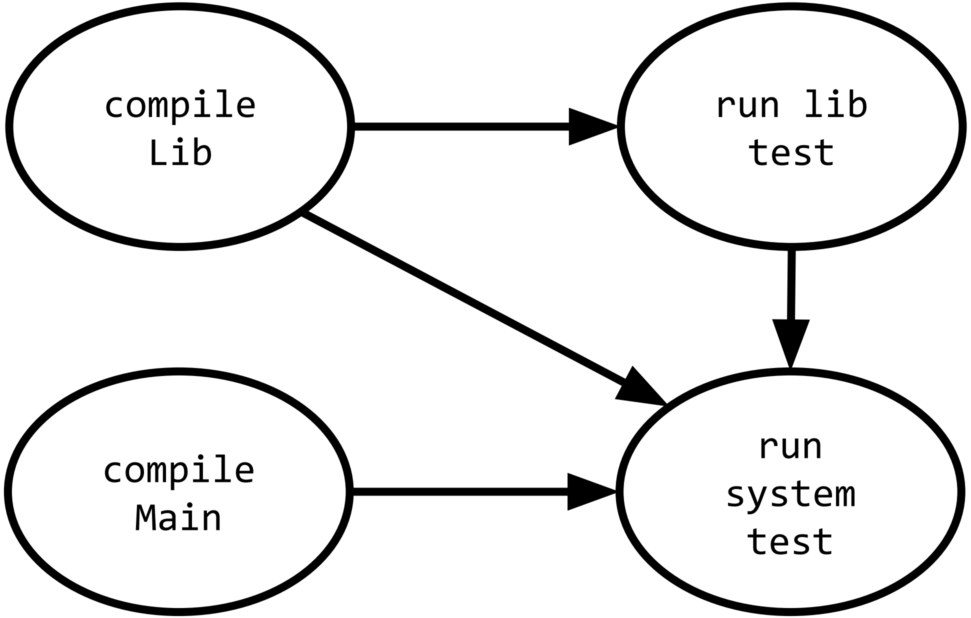
Topological sort

- Any ordering on the nodes such that all dependencies are satisfied

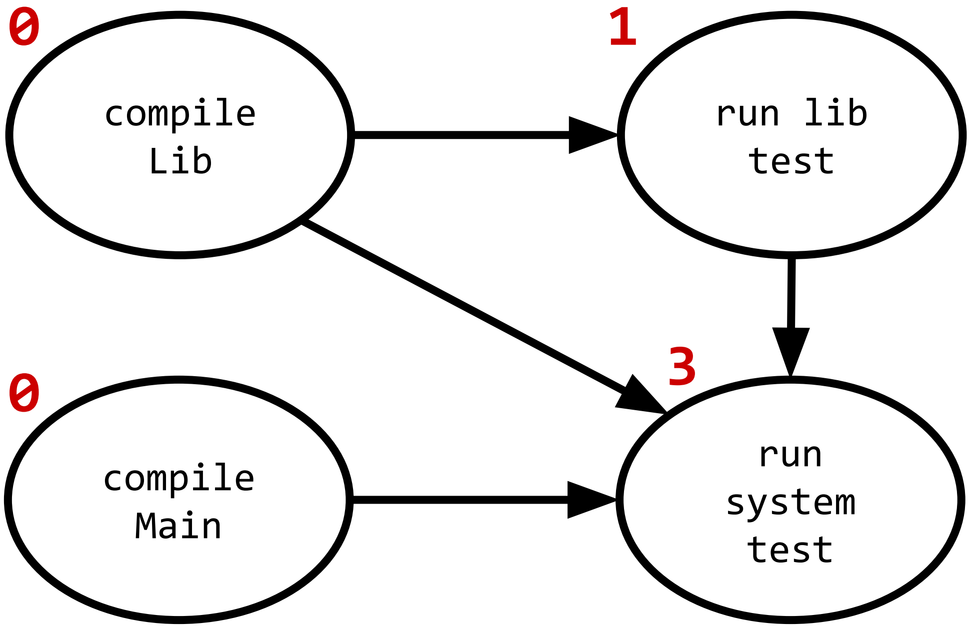
Topological sort

- Any ordering on the nodes such that all dependencies are satisfied
- Implement by computing ***indegree*** (number of incoming edges) for each node

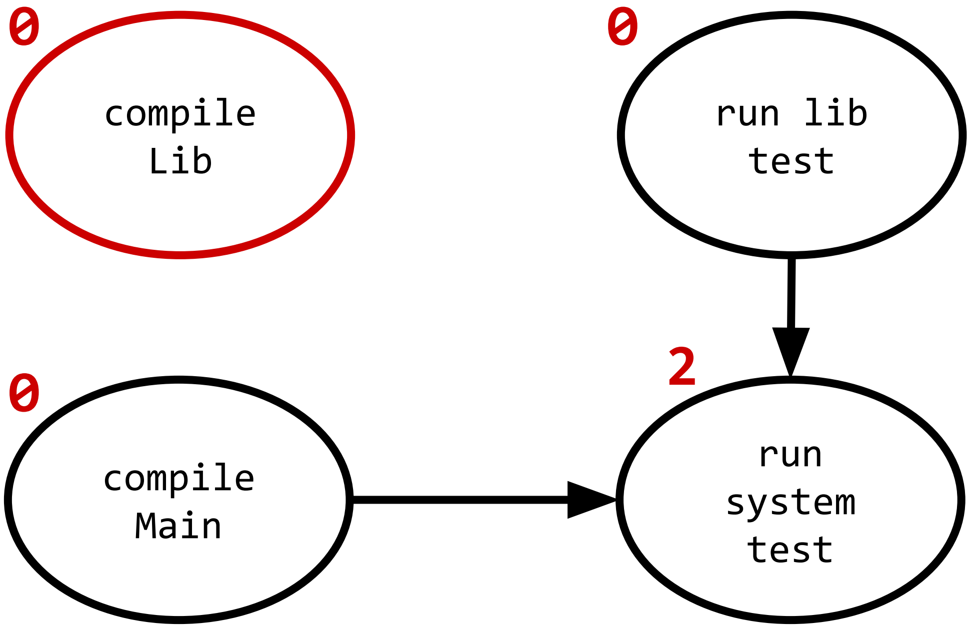
Topological sort



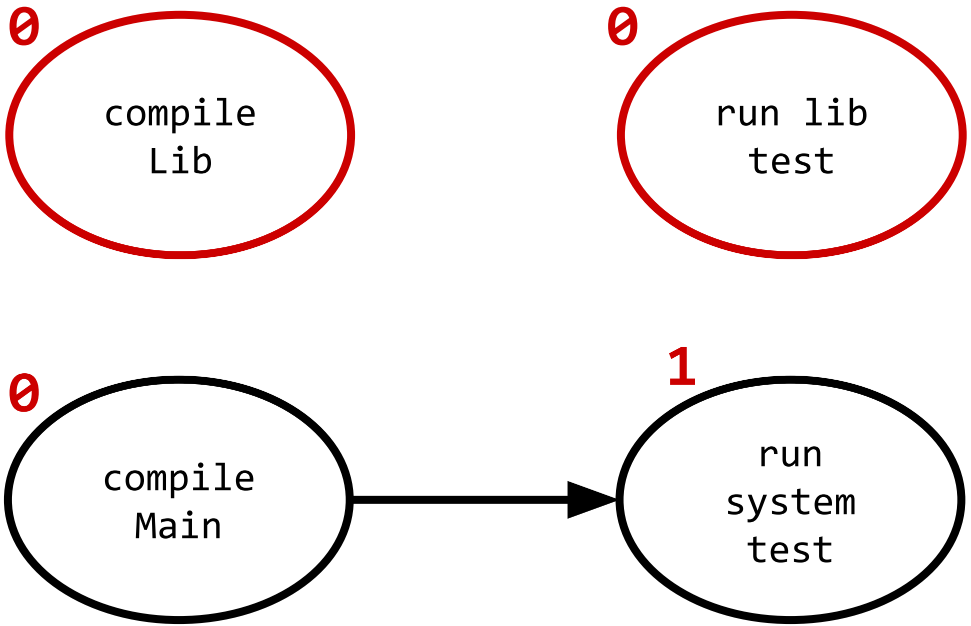
Topological sort



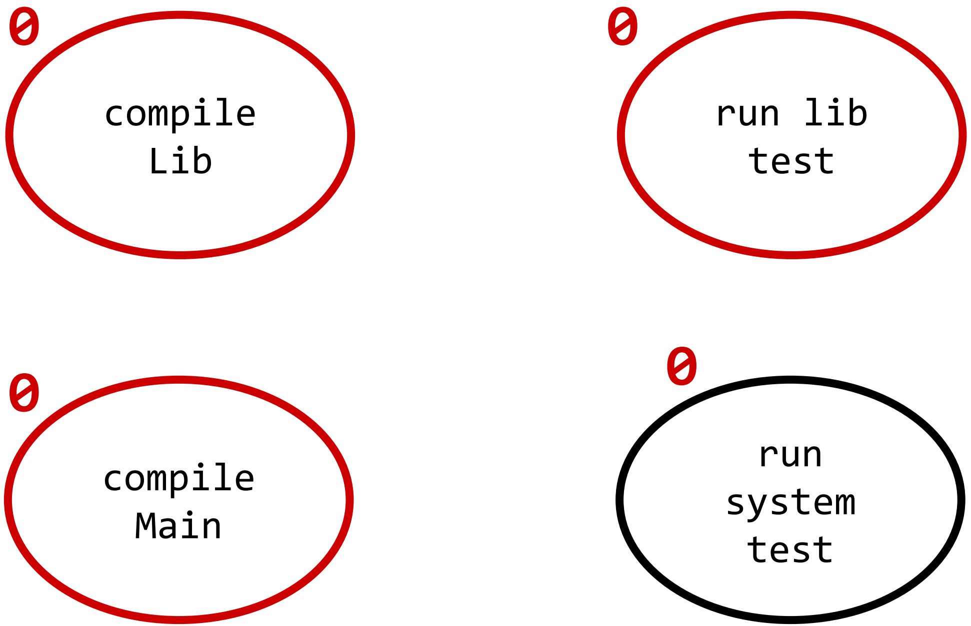
Topological sort



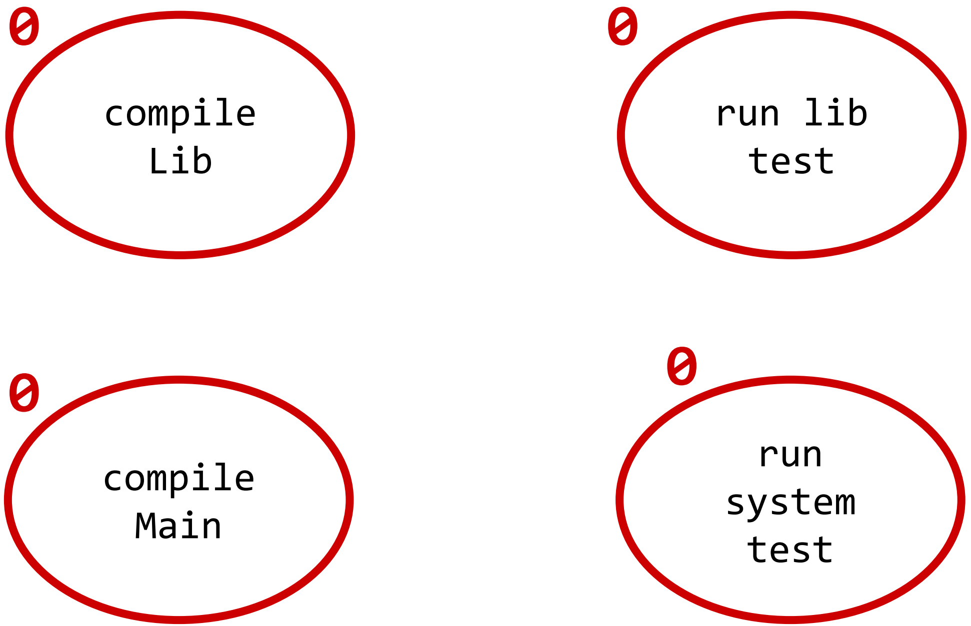
Topological sort



Topological sort



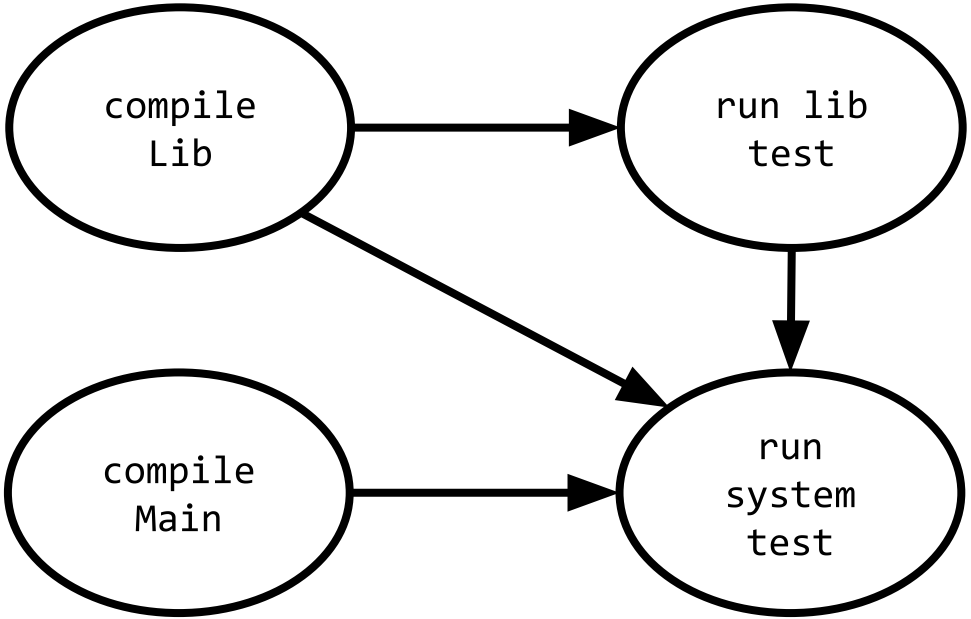
Topological sort



Topological sort

Valid sorts:

1. compile Lib, run lib test, compile Main, run system test

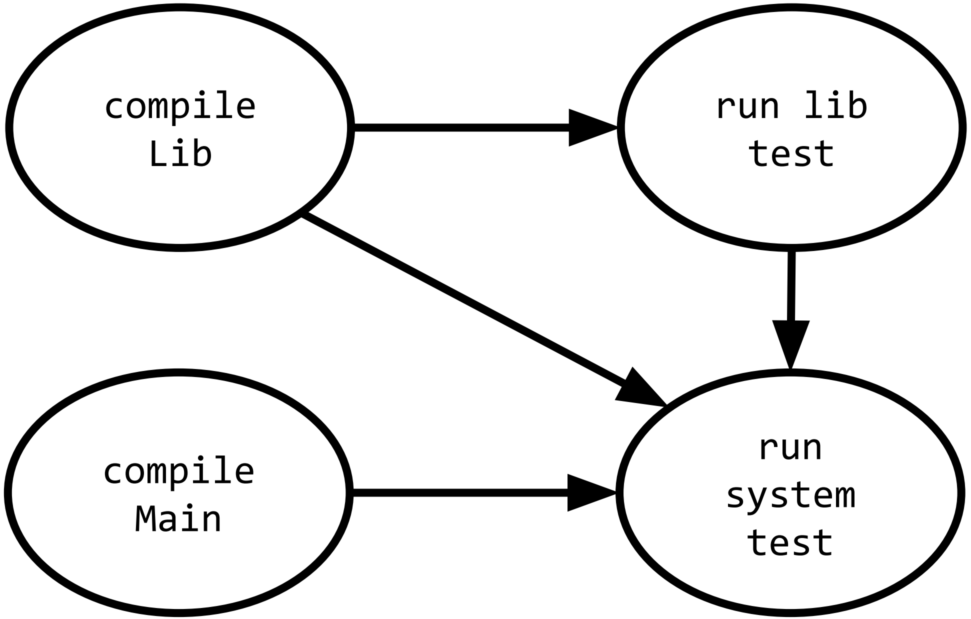


Topological sort

Valid sorts:

1. compile Lib, run lib test, compile Main, run system test

2. compile Main, compile Lib, run lib test, run system test



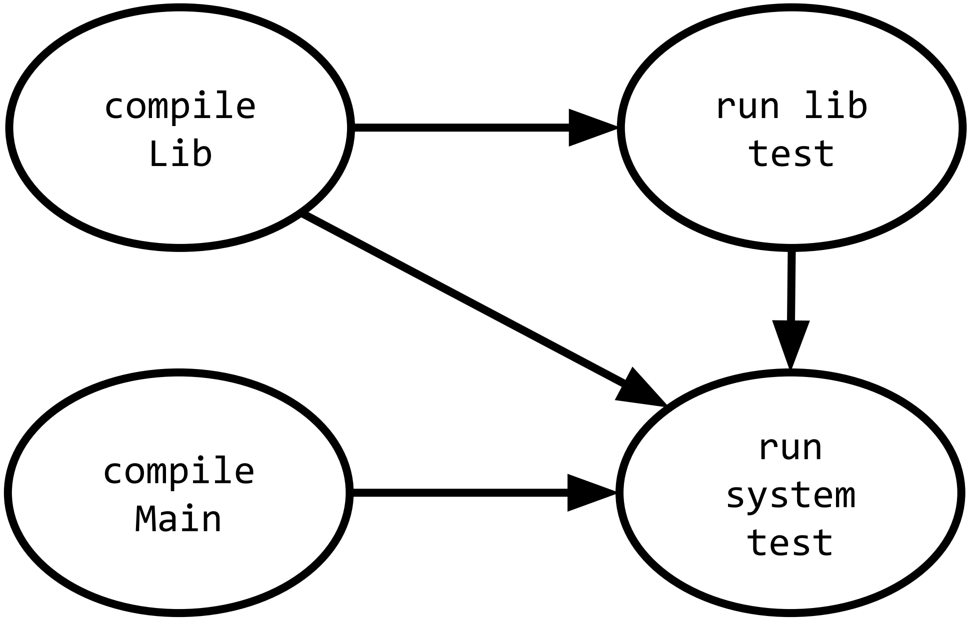
Topological sort

Valid sorts:

1. compile Lib, run lib test, compile Main, run system test

2. compile Main, compile Lib, run lib test, run system test

3. compile Lib, compile Main, run lib test, run system test



Examples of modern build systems

`gradle`

https://gradle.org/

Apache's open-source successor to ant, maven

`bazel`

https://www.bazel.build/

Google's internal build tool, now open-source

Example task: gradle

```
task reformat(type: Exec, dependsOn: getCodeFormatScripts, group: 'Format') {
   description 'Format the Java source code'
   // jdk8 and checker-qual have no source, so skip
   onlyIf { !project.name.is('jdk8') && !project.name.is('checker-qual') }
   executable 'python'
   doFirst {
       args += "${formatScriptsHome}/run-google-java-format.py"
       args += "--aosp" // 4 space indentation
       args += getJavaFilesToFormat(project.name)
   }
}
```

Example task: gradle

```
task reformat(type: Exec, dependsOn: getCodeFormatScripts, group: 'Format') {
   description 'Format the Java source code'
   // jdk8 and checker-qual have no source, so skip
   onlyIf { !project.name.is('jdk8') && !project.name.is('checker-qual') }
   executable 'python'
   doFirst {
       args += "${formatScriptsHome}/run-google-java-format.py"
       args += "--aosp" // 4 space indentation
       args += getJavaFilesToFormat(project.name)
   }
}
```

kind of rule

Example task: gradle

```
task reformat(type: Exec, dependsOn: getCodeFormatScripts, group: 'Format') {
    description 'Format the Java source code'
    // jdk8 and checker-qual have no source, so skip
    onlyIf { !project.name.is('jdk8') && !project.name.is('checker-qual') }
    executable 'python'
    doFirst {
        args += "${formatScriptsHome}/run-google-java-format.py"
        args += "--aosp" // 4 space indentation
        args += getJavaFilesToFormat(project.name)
    }
}
```

explicitly specified dependencies

Example task: gradle

```
task reformat(type: Exec, dependsOn: getCodeFormatScripts, group: 'Format') {
    description 'Format the Java source code'
    // jdk8 and checker-qual have no source, so skip
    onlyIf { !project.name.is('jdk8') && !project.name.is('checker-qual') }
    executable 'python'
    doFirst {
        args += "${formatScriptsHome}/run-google-java-format.py"
        args += "--aosp" // 4 space indentation
        args += getJavaFilesToFormat(project.name)
    }
}
```

code!

Example task: `bazel`

```
java_binary(
    name = "dux",
    main_class = "org.dux.cli.DuxCLI",
    deps = ["@google_options//:compile",
            "@checker_qual//:compile",
            "@google_cloud_storage//:compile",
            "@slf4j//:compile",
            "@logback_classic//:compile"],
    srcs = glob(["src/org/dux/cli/*.java",
                 "src/org/dux/backingstore/*.java"]),
)
```

Example task: `bazel`

```
java_binary(
    name = "dux",
    main_class = "org.dux.cli.DuxCLI",
    deps = ["@google_options//:compile",
            "@checker_qual//:compile",
            "@google_cloud_storage//:compile",
            "@slf4j//:compile",
            "@logback_classic//:compile"],
    srcs = glob(["src/org/dux/cli/*.java",
                 "src/org/dux/backingstore/*.java"]),
)
```

kind of rule

Example task: bazel

```
java_binary(
    name = "dux",
    main_class = "org.dux.cli.DuxCLI",
    deps = ["@google_options//:compile",
            "@checker_qual//:compile",
            "@google_cloud_storage//:compile",
            "@slf4j//:compile",
            "@logback_classic//:compile"],
    srcs = glob(["src/org/dux/cli/*.java",
                 "src/org/dux/backingstore/*.java"]),
)
```

explicitly specified dependencies

Example task: `bazel`

```
java_binary(
    name = "dux",
    main_class = "org.dux.cli.DuxCLI",
    deps = ["@google_options//:compile",
            "@checker_qual//:compile",
            "@google_cloud_storage//:compile",
            "@slf4j//:compile",
            "@logback_classic//:compile"],
    srcs = glob(["src/org/dux/cli/*.java",
                 "src/org/dux/backingstore/*.java"]),
)
```

explicitly specified dependencies (also bazel tasks)

External and internal dependencies

- A list of tasks (internal) or libraries (external)

External and internal dependencies

- A list of tasks (internal) or libraries (external)

```
deps = ["@google_options//:compile",
        "@checker_qual//:compile",
        "@google_cloud_storage//:compile",
        "@slf4j//:compile",
        "@logback_classic//:compile"],
```

```
dependencies {
    compile group:
        'org.hibernate',
        name: 'hibernate-core',
        version : '3.6.7.Final'
    testCompile group:
        'junit',
        name: 'junit',
        version : '4.+'
}
```

https://docs.gradle.org/current/userguide/artifact_dependencies_tutorial.html

Why list dependencies?

- Reproducibility!

Why list dependencies?

- Reproducibility!
- ***Hermetic builds***: "they are insensitive to the libraries and other software installed on the build machine"[1]

[1] https://landing.google.com/sre/sre-book/chapters/release-engineering/

Why list dependencies?

- Reproducibility!
- ***Hermetic builds***: "they are insensitive to the libraries and other software installed on the build machine"[1]
    - critical if you want to get new developers working quickly (remember the reading!)
    - useful for debugging problems users encounter with old versions (can always get back to exactly the code they're using)
    - prevents "it works on my machine" syndrome

[1] https://landing.google.com/sre/sre-book/chapters/release-engineering/

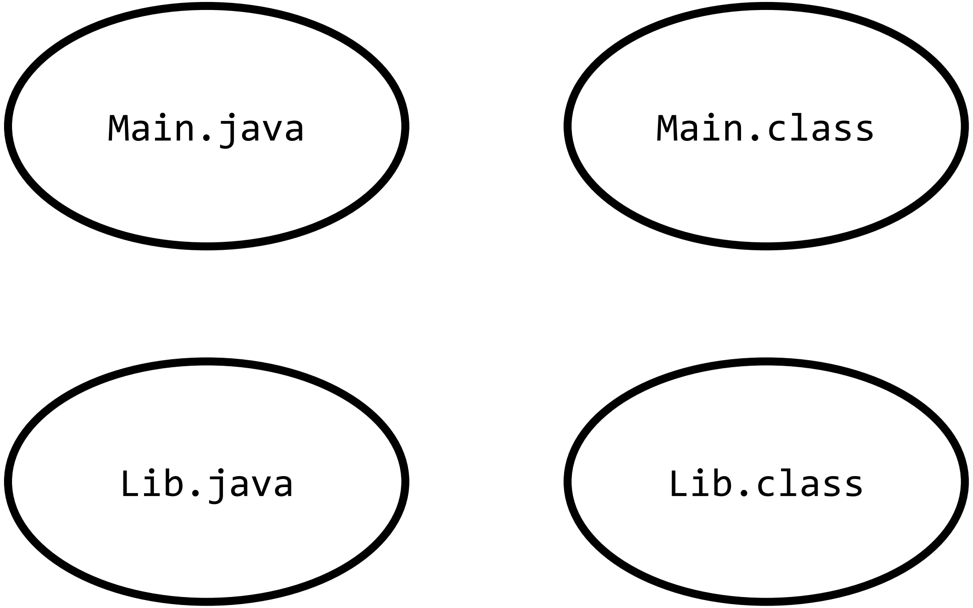
Dependencies between tasks

- A large project may have thousands of tasks
  - What order to run in?
  - **How to speed up?**

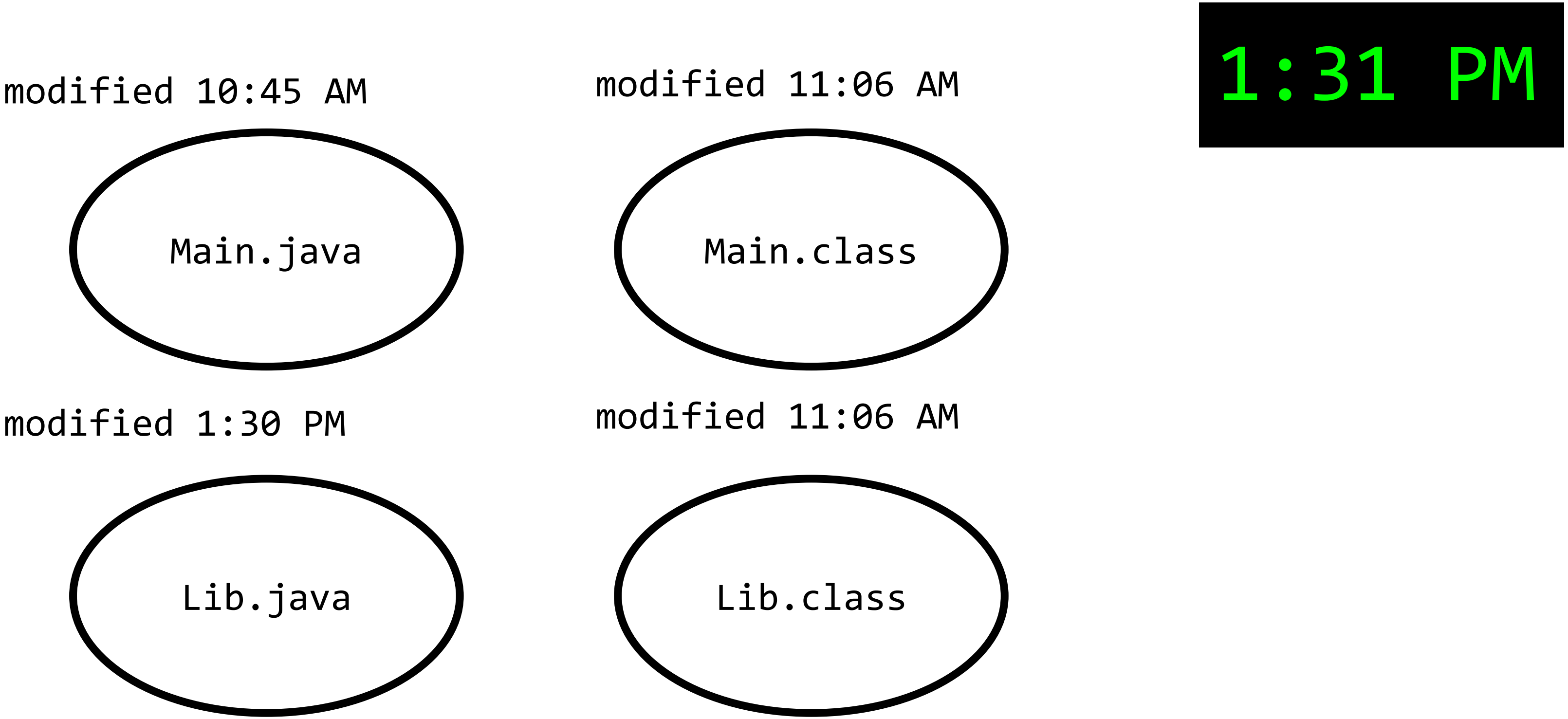
How to speed up builds?

- **Incrementalize** - only rebuild what you have to

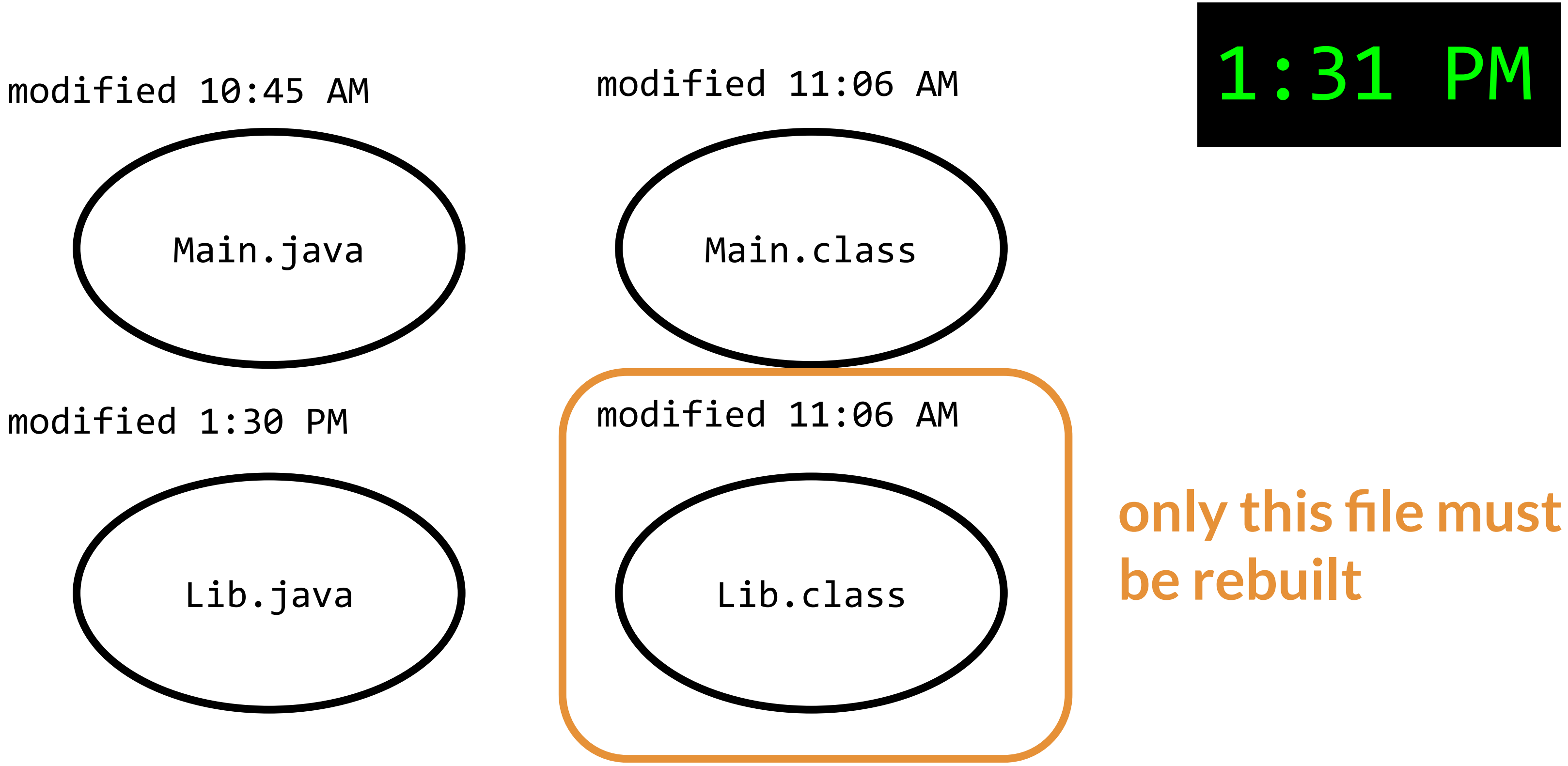
Incrementalization



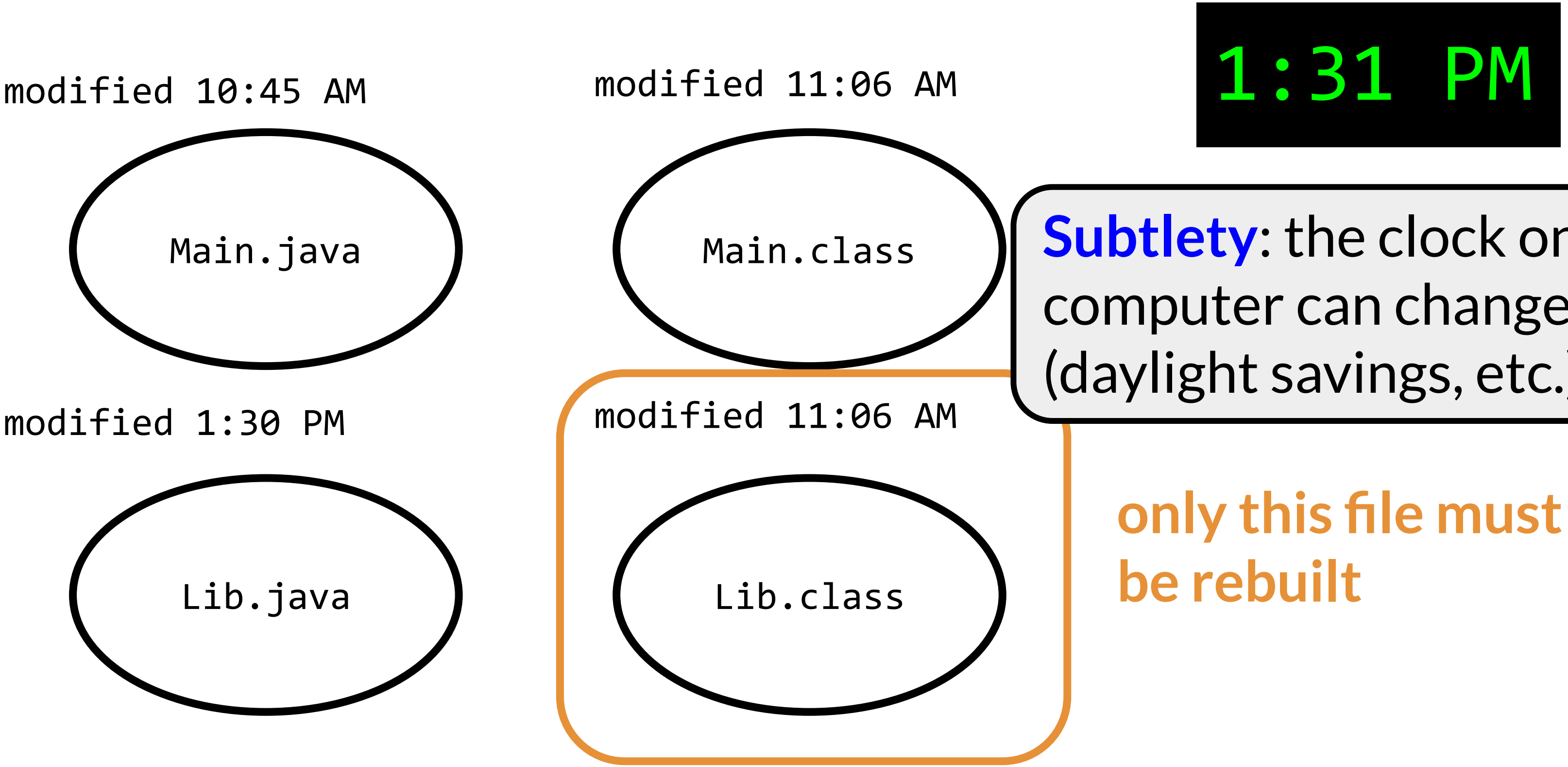
Incrementalization: time stamps



Incrementalization: time stamps



Incrementalization: time stamps



Subtlety: the clock on a computer can change… (daylight savings, etc.)

only this file must be rebuilt

Incrementalization: hashing

Incrementalization: hashing

- Compute hash codes for inputs to each task
- When about to execute a task, check input hashes - if they match the last time the task was executed, skip it!

How to speed up builds?

- **Incrementalize** - only rebuild what you have to
- Execute many tasks in **parallel**

How to speed up builds?

- **Incrementalize** - only rebuild what you have to
- Execute many tasks in **parallel**
  - some build system tasks are ***embarrassingly parallel***: they can be reordered without explicit synchronization
    - is this true of all tasks?

How to speed up builds?

- **Incrementalize** - only rebuild what you have to
- Execute many tasks in **parallel**
  - some build system tasks are ***embarrassingly parallel***: they can be reordered without explicit synchronization
    - is this true of **all** tasks? **No**: some tasks depend on each other. The problem of scheduling tasks with no unbuilt dependencies is embarrassingly parallel, though.

How to speed up builds?

- **Incrementalize** - only rebuild what you have to
- Execute many tasks in **parallel**
  - some build system tasks are ***embarrassingly parallel***: they can be reordered without explicit synchronization
    - is this true of all tasks? **No**: some tasks depend on each other. The problem of scheduling tasks with no unbuilt dependencies is embarrassingly parallel, though.
- **Cache** artifacts in the cloud

How do build systems differ

How do build systems differ

- Scheduling algorithm

How do build systems differ

- Scheduling algorithm
  - We've already seen topological scheduling (used by e.g. make), which is a **static** scheduling algorithm

How do build systems differ

- Scheduling algorithm
    - We've already seen topological scheduling (used by e.g. make), which is a **static** scheduling algorithm
    - **Dynamic** scheduling algorithms are also possible

How do build systems differ

- Scheduling algorithm
  - We've already seen topological scheduling (used by e.g. make), which is a **static** scheduling algorithm
  - **Dynamic** scheduling algorithms are also possible
    - **Key idea**: compute what dependencies are necessary as you go

How do build systems differ

- Scheduling algorithm
  - We've already seen topological scheduling (used by e.g. make), which is a static scheduling algorithm
  - Dynamic scheduling algorithms are also possible
    - **Key idea**: compute what dependencies are necessary as you go
    - this is how e.g., Bazel actually schedules tasks

How do build systems differ

- Rebuilding strategy

How do build systems differ

- Rebuilding strategy
  - We’ve seen two:

How do build systems differ

- Rebuilding strategy
  - We’ve seen two:
    - a ***dirty bit*** strategy (make’s timestamps)

How do build systems differ

- Rebuilding strategy
  - We’ve seen two:
    - a ***dirty bit*** strategy (make’s timestamps)
    - a ***verifying trace*** strategy (storing hashes of each object)

How do build systems differ

- Rebuilding strategy
  - We've seen two:
    - a ***dirty bit*** strategy (make's timestamps)
    - a ***verifying trace*** strategy (storing hashes of each object)
  - Other options:
    - ***constructive traces***: store all intermediate objects (usually in the cloud) along with the hashes of the inputs used to produce them. If we ever see the same input hashes again, just return the intermediate object

How do build systems differ

- How are tasks expressed?

How do build systems differ

- How are tasks expressed?
  - traditionally **declarative** (e.g., make, Ant, Maven)
    - "declarative" = you tell the build system what you want, it figures out how to build that thing
    - call back to languages: programming languages can also be from the ***declarative paradigm*** (e.g., Prolog)

How do build systems differ

- How are tasks expressed?
  - traditionally **declarative** (e.g., make, Ant, Maven)
    - "declarative" = you tell the build system what you want, it figures out how to build that thing
    - call back to languages: programming languages can also be from the ***declarative paradigm*** (e.g., Prolog)
  - most modern build systems have **scripting languages**
    - e.g., Groovy in Gradle, Starlark in Bazel, etc.
    - enables us to write tasks as if they are other code

How to choose a build system

How to choose a build system

High level idea: same rules apply to choosing a language

How to choose a build system

High level idea: same rules apply to choosing a language

- **don’t change what’s already there** unless there is a good reason

How to choose a build system

High level idea: same rules apply to choosing a language

- **don't change what's already there** unless there is a good reason
- **follow convention** and prefer the tooling that's "idiomatic" to your language
    - e.g., use Gradle or Maven when working in Java

When to switch build systems

- developers rarely choose to change build systems **except** when **build performance** is a problem

When to switch build systems

- developers rarely choose to change build systems **except** when **build performance** is a problem
  - common causes include:

When to switch build systems

- developers rarely choose to change build systems **except** when **build performance** is a problem
  - common causes include:
    - poor incrementalization (e.g., Maven's per-module incremental compilations)

When to switch build systems

- developers rarely choose to change build systems **except** when **build performance** is a problem
  - common causes include:
    - poor incrementalization (e.g., Maven's per-module incremental compilations)
    - lack of support for artifact caching (= **cloud builds**)

When to switch build systems

- developers rarely choose to change build systems **except** when **build performance** is a problem
  - common causes include:
    - poor incrementalization (e.g., Maven’s per-module incremental compilations)
    - lack of support for artifact caching (= **cloud builds**)
    - build has become too complex for a declarative task language

When to switch build systems

- developers rarely choose to change build systems **except** when **build performance** is a problem
    - common causes include:
        - poor incrementalization (e.g., Maven's per-module incremental compilations)
        - lack of support for artifact caching (= **cloud builds**)
        - build has become too complex for a declarative task language
    - most projects keep the same build system **forever**

Best practices

- Automate everything

Best practices

- Automate everything
- Always use a build tool

Best practices

- Automate everything
- Always use a build tool
- Have a build server that builds and tests your code on every commit (continuous integration)

Best practices

Your CI server is a good place to test that your build is hermetic. **Standard practice**: spin up a new CI server for **each build**.

- Automate everything
- Always use a build tool
- Have a build server that builds and tests your code on every commit (continuous integration)

Best practices

- Automate everything
- Always use a build tool
- Have a build server that builds and tests your code on every commit (continuous integration)
- Don’t depend on anything that’s not in the build file (hermetic)

Best practices

- Automate everything
- Always use a build tool
- Have a build server that builds and tests your code on every commit (continuous integration)
- Don’t depend on anything that’s not in the build file (hermetic)
- Don’t break the build

Best practices

- Automate everything
- Always use a build tool
- Have a build server that builds and tests your code on every commit (continuous integration)
- Don't depend on anything that's not in the build file (hermetic)
- Don't break the build

A **common mistake to avoid**: allowing the CI server to fail for a long time because "we know what the problem is." Don't do this: leads to complacency, missing real bugs.