Principles of Software Construction: Objects, Design, and Concurrency

IDEs, Build system, Continuous Integration, Libraries

Vincent Hellendoorn Claire Le Goues



Administrivia

institute for software RESEARC

What did we talk about Thursday?



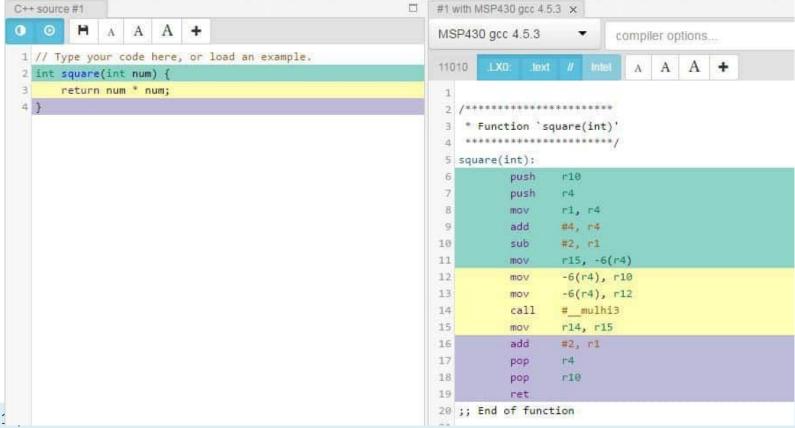
Productivity Requires Automation Requires Abstraction





17-214/514

Productivity Requires Automation Requires Abstraction



Automation Requires Abstraction

We all treat familiar levels of abstraction as normal/natural

- That's fine if you only drive your car
 - Not so much if you are a mechanic
 - How to debug a broken transmission?
- Also slow to evolve
 - o Conf. people adamantly refusing to use an automatic
- Abstractions simplify engineering work, and allow engineers to focus on the hard parts.
 - But, effective engineers also know what is beneath the abstractions

Automation Requires Abstraction

Today's "normal":

- Integrated-development environments (IDEs) galore
 - Web-based too! Press "." on a GitHub (file) page ?
- Frequent build, test, release
 - o In some companies, every commit is a "release"
- Never write code for which there is a useful library
 - Define "useful"?
- All of the above, entangled



Abstraction, Reuse, and Programming Tools

- For each in {Build systems, IDE, libraries, CI}:
 - O What is it?
 - What happens under the hood?
 - What are some practical ways to use it?
- What is next?

institute for SOFTWARI

Abstraction, Reuse, and Programming Tools

- For each in {Build systems, IDE, libraries, CI}:
 - O What is it?
 - What happens under the hood?
 - What are some practical ways to use it?
- What is next?

institute for SOFTWARI

Quick overview of today's toolchain: Build Systems

How does this happen?



Starting a program: Java

All Java code is in classes, so how to create an object and call a method?

Special syntax for *main* method in class (java X calls *main* in X)

```
start with: java Printer
                                    in Java,
class Printer {
                                  everything is
    void print() {
                                     a class
         System.out.println("hi");
    public static void main(String[] args) {
         Printer obj = new Printer();
         obj.print();
                                    main must be
                                     public and
                                       static
```

Main method to be executed, here used to create object and invoke method

Static methods belong to class not the object, generally avoid them

Starting a Program: Javascript

Objects do not do anything on their own, they wait for method calls

Every program needs a starting point, or waits for events

```
// start with: node file.js
function createPrinter() {
    return {
                                                                 Defining interfaces,
         print: function() { console.log("hi"); }
                                                                 functions, classes
const printer = createPrinter();
                                                                 Starting:
                                                                 Creating objects and
printer.print()
                                                                 calling methods
 // hi
```

SimpleHello.java/SimpleHello.ts

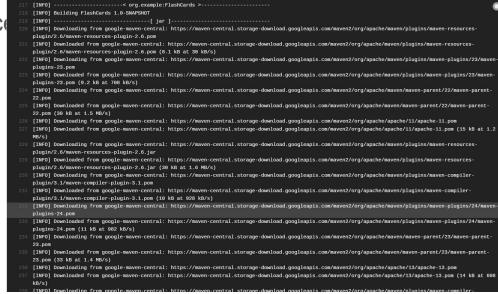
- Java:
 - Is compiled to bytecode, which is then run in a runtime environment (JVM + GC, etc).
 - Command line execution vs. in VSCode
 - What's the -cp doing?
- Typescript:
 - Is transpiled to javascript, which is then executed inside a runtime environment.
 - Outside the browser: we use node.js
- Why do we encourage you to run everything in an IDE/use Maven/npm?
 - Hints: what are the roles of -cp in Java? Or package.json in typescript?

Quick overview of today's toolchain: Build Systems

Compiling is "easy" when all your source code is here

Nowadays, your code is not "here"

- Even libraries that you use in the IDE!
- Interfaces make that possible



IST institute for SOFTWARI

Quick overview of today's toolchain: Build Systems

- Building a few basic tasks:
 - Compiling & linking, to produce an executable
 - Creating secondary *artifacts*, e.g. documentation-pages, linter reports, test suite reports
 - Different levels of "depth" may be appropriate, for large code bases (e.g. Google)
- Popular options:
 - For Java: Maven and Gradle -- historically Ant.
 - You could do any homework in either, in principle; we just standardize to make helping you and grading easier.
 - For JS/TS: Node(JS)
 - Generally coupled with the Node Package Manager (NPM)
- Often built into IDEs, as plugins

Under the Hood: Build Systems

- These days: intricately tied with IDEs, package managers
- Projects often come with a build config file or two
 - o 'pom.xml' for Maven
 - 'tsconfig.json' + 'package.json' for TypeScript+NPM -- the second deals with packages
 - These can be nested, one per (sub-)directory, to compose larger systems
 - On GitHub, you can create links across repositories

Under the Hood: Build Systems

Projects typically require build config files

- Checked into source control
 - o 'pom.xml' for Maven
 - 'tsconfig.json' + 'package.json' for TypeScript+NPM -- the second deals with packages
- Relevant components:
 - Compilation source and target version
 - High-level configuration options
 - Targets for various development phases
 - Relevant plugins
 - Dependencies with versions





Maven Phases

Although hardly a comprehensive list, these are the most common default lifecycle phases executed.

- validate: validate the project is correct and all necessary information is available
- compile: compile the source code of the project
- test: test the compiled source code using a suitable unit testing framework. These tests should not require the code
- package: take the compiled code and package it in its distributable format, such as a JAR.
- Integration-test: process and deploy the package if necessary into an environment where integration tests can be
- verify: run any checks to verify the package is valid and meets quality criteria
- install: i stall the package into the local repository, for use as a dependency in other projects locally
- deploy: done in an integration or release environment, copies the final package to the remote repository for sharing

There are two other Maven lifecycles of note beyond the default list above. They are

- clean: cleans up artifacts created by prior builds
- . site: generates site documentation for this project

https://maven.apache.org/guides/getting-started/maven-in-five-minutes.html

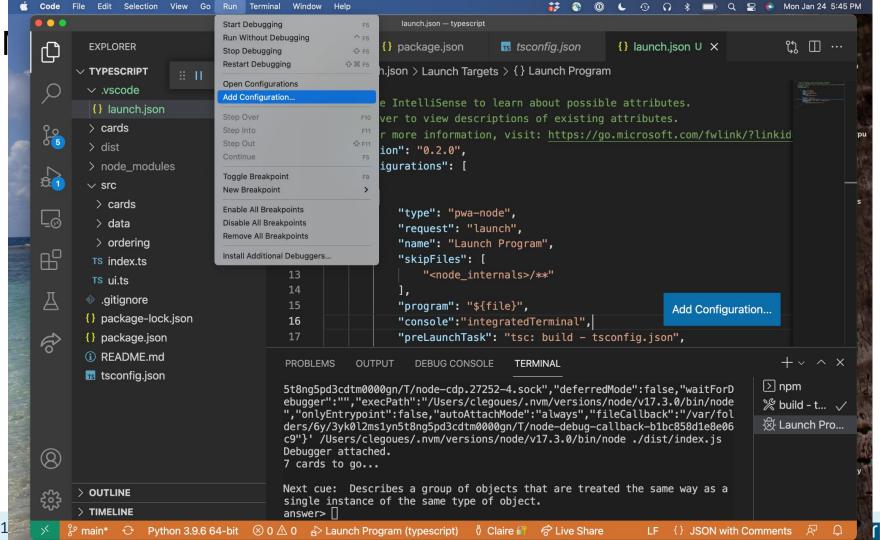
17-214/514 **18** is

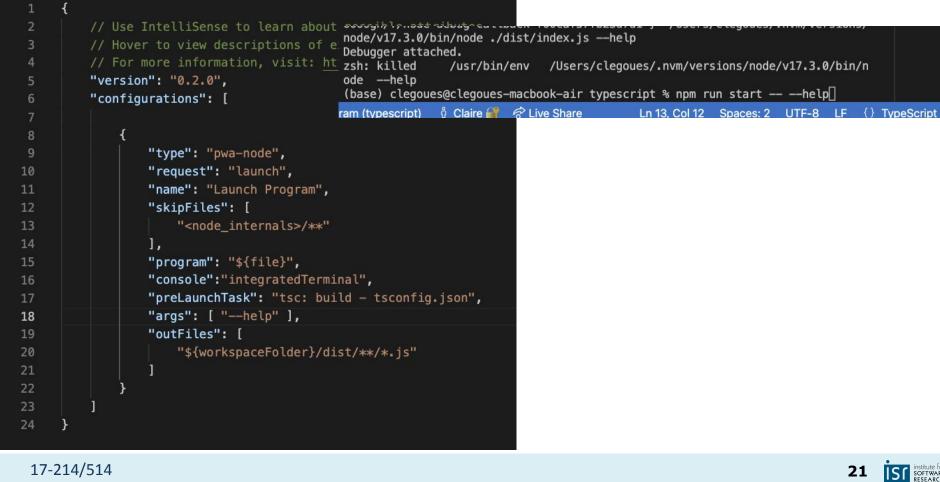


 Node.js is a JS runtime. npm is its package manager.

```
} package.json 1, M ×
{} package.json > {} dependencies
        "name": "hw1-flashcards",
        "version": "1.0.0",
        "description": "",
        "main": "index.js",
        ▶ Debug
        "scripts": {
          "compile": "tsc",
 8
          "lint": "ts-standard",
          "start": "node dist/index.js"
10
        "author": "",
11
12
        "license": "ISC",
13
        "devDependencies": {
14
          "@types/node": "^17.0.8",
15
          "@types/readline-sync": "^1.4.4",
          "ts-standard": "^10.0.0",
          "typescript": "^4.4.2"
17
18
        "dependencies": {
19
20
          "readline-sync": "^1.4.10",
21
22
23
```

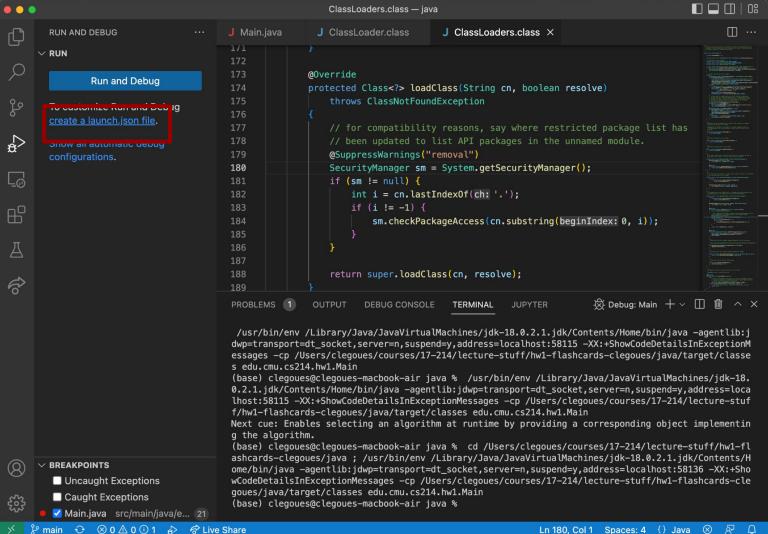
package.json — claire-hw1-js





{} launch.json U X

.vscode > {} launch.json > Launch Targets > {} Launch Program















What's going on in HW1?

Also: quick Maven setup demo.



Abstraction, Reuse, and Programming Tools

- For each in {Build systems, IDE, libraries, CI}:
 - O What is it?
 - What happens under the hood?
 - What are some practical ways to use it?
- What is next?

IST institute for

Abstraction, Reuse, and Programming Tools

- For each in {Build systems, IDE, libraries, CI}:
 - O What is it?
 - What happens under the hood?
 - What are some practical ways to use it?
- What is next?

Quick overview of today's toolchain: IDEs

Integrated Development Environments bundle development workflows in a single UI

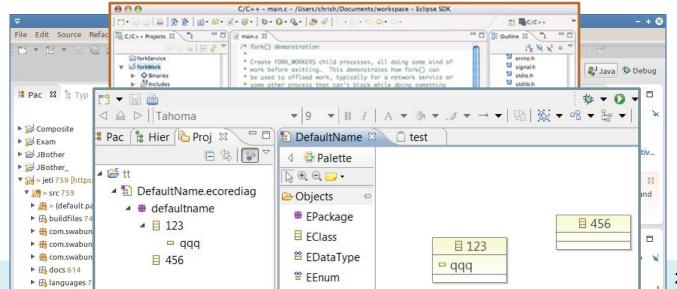
- Editing, refactoring, running & debugging, adding dependencies, compiling, deploying, plugins, you name it
- They often try to be everything, with mixed results
- Leverage them to the fullest extent, to automate and check your work

IDEs are key to managing complexity.

IST institute for

Quick overview of today's toolchain: IDEs

Eclipse was the dominant player in Java for 20-odd years, owing to its powerful backbone and plugin architecture



Quick overview of today's toolchain: IDEs

Recently, IntelliJ has been more dominant

 Packs a lot of "recipes" to create certain types of projects (e.g., web-app with Spring & Maven)

VSCode is surging in popularity

Local & web, lightweight but with a massive plugin ecosystem

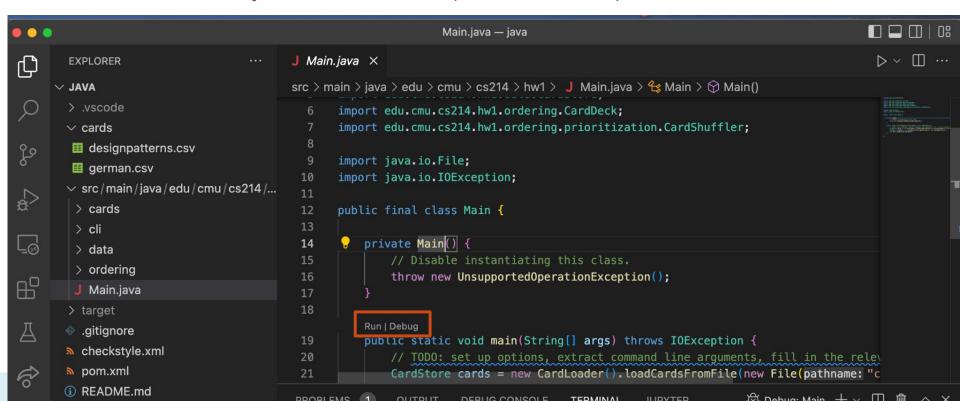
In general: choose based on need!

- You can relearn key-bindings; "killer features" are rare and temporary
- E.g., Android: might want Android Studio (itself built on IntelliJ) since Google supports it

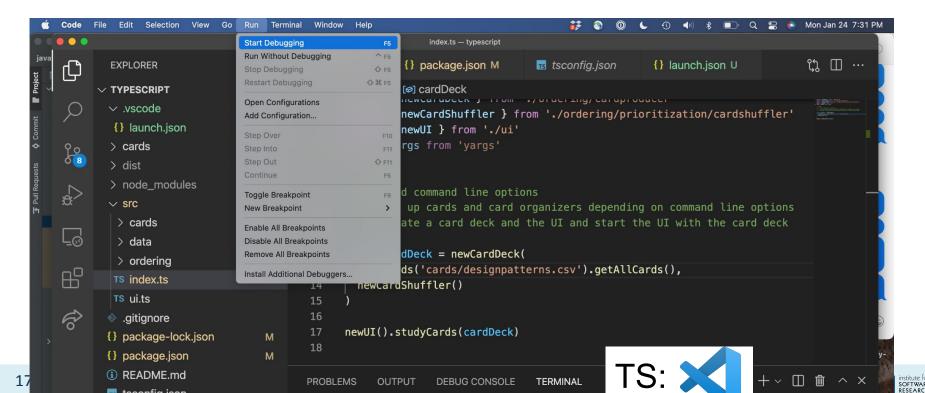
We recommended VSCode because you're programming in two languages, it's very current, and then it's easier for us to help you.

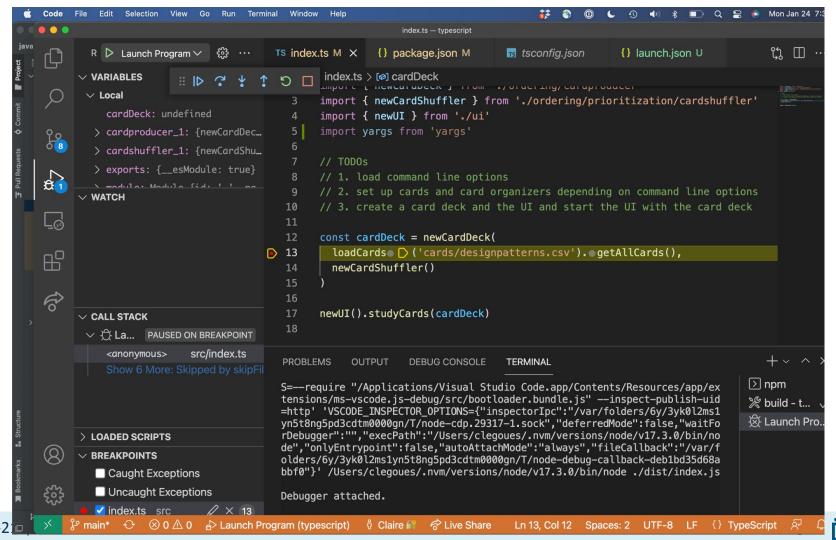
- The engine: continuous parsing, building
 - Key feature: most partial programs don't parse, but IDEs make sense of them
 - That allows quickly relaying compile warnings/errors and useful suggestions
 - Same with API resolution
- Powered by rapid <u>incremental compilation</u>
 - Only build what has been updated
 - Virtually every edit you make triggers a compilation, re-linking
 - Of just the changed code and its dependencies
 - Works because very little of the code changes most of the time
 - But no free lunch: tends to drop optimizations (mostly fine), may struggle with big projects
 - Just try it: call an API with the wrong parameters & see how fast it triggers an alert; contrast with running a full Maven build (e.g., with `mvn install`)

Automate common programming actions, like debugging, which is often the default mode when you run in the IDE (like in VSCode)

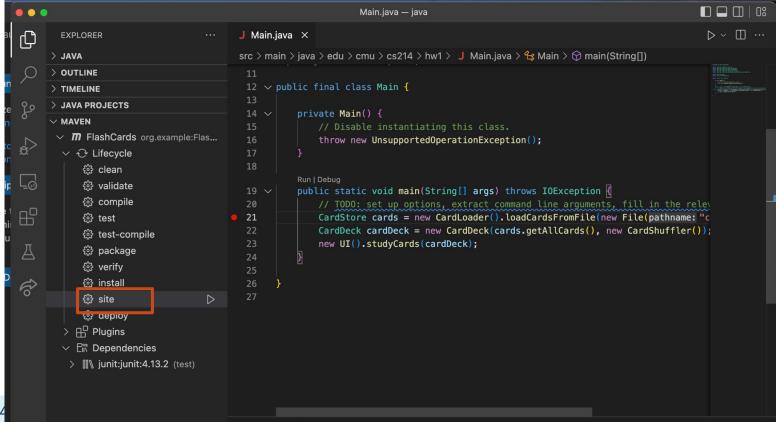


Debugging allows setting breakpoints in the GUI, access to rich execution info.





Combine build systems + IDEs + plugins (checkstyle example/demo!)



Automate common programming actions:

- Handy refactorings, suggestions
 - E.g., just press `alt+enter` in IntelliJ while highlighting nearly any code
 - Keyboard shortcuts are super useful: explore your IDE!
 - These can make you a better programmer: encode a lot of best-practices
 - Though, don't read into them too much

```
public final class Main {

private Main() {
    // Disable instantiating this class.
    throw new UnsupportedOperationException();
}

Add runtime exception(s) to method signature >

Press Ctrl+Shift+| to open preview

public static void main(String[] args) throws IUException {
    // TODO: set up options, extract command line arguments, fill in the relevant object
    CardStore cards = new CardLoader().loadCardsFromFile(new File( pathname: "cards/designation").
```

- IDE designers spend a lot of time automating common development tasks
 - Sometimes they get a little too helpful (modifying pom's)
 - Many plugins provide customized experience
 - Mostly evolve with new tools, prioritizing emerging routines
- Useful to know how these actions work
 - Often not much more than invoking commands for you
 - VSCode is very explicit about this in the terminal -- great for customization

"C:\Program Files\Java\jdk-16.0.1\bin\java.exe" -ea -Didea.test.cyclic.buffer.size=1048576 "-javaagent:C:\Program Files\JetBrains\IntelliJ ID

Process finished with exit code 0

Abstraction, Reuse, and Programming Tools

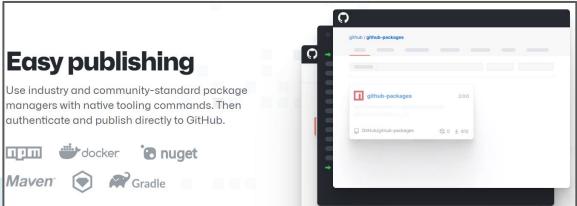
- For each in {Build systems, IDE, libraries, CI}:
 - O What is it?
 - What happens under the hood?
 - What are some practical ways to use it?
- What is next?

Quick overview of today's toolchain: Libraries

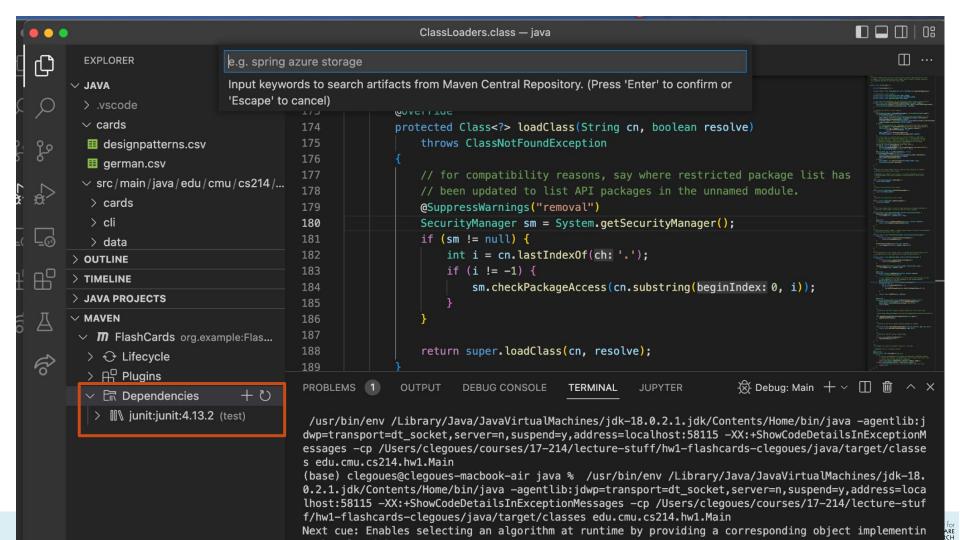
Reusable packages of code.

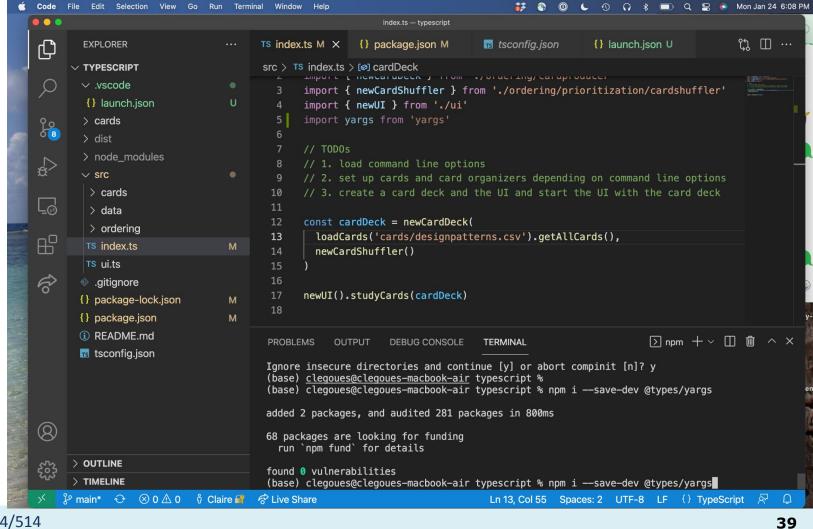
Publicly hosted on various package managers

- Often tied, but not inextricably linked, to build tools, and languages
- Maven/Gradle for Java, NPM for JS/TS, Nuget for C#, ...
- Registries of managers, e.g., GitHub Packages









17-214/514

Quick overview of today's toolchain: Libraries

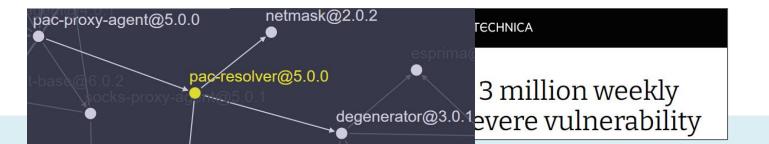
- Myriad. Publicly hosted on various managers
 - Often tied, but not inextricably linked, to build tools, and languages
 - o Maven, Gradle, NPM, Nuget, Docker, ...
 - Registries of managers, e.g., GitHub Packages
- Releases are generally fast-paced or frigid
 - Almost all volunteer-based, so support waivers, as does documentation quality
 - Often open-source, so you can check out the status & details on GitHub
 - o Beware of vulnerabilities and bugs, esp. with minor-releases and nightly's, old packages



NPM package with 3 million weekly downloads had a severe vulnerability

Quick overview of today's toolchain: Libraries

- A Case-Study:
 - o 'pac-resolver' (3M weekly downloads) has a major security vulnerability
 - Uses 'degenerator' (same author), which misuses a Node module
 - "The vm module is not a security mechanism. Do not use it to run untrusted code."
 - (a mistake that's been made before: people rarely read disclaimers)
 - o 'pac-proxy-agent' (2M weekly downloads, same author) uses the above
 - Is widely popular, the main reason people use 'degenerator'
 - Most people using this package have never heard of the latter -- many never will



Log4j software bug: What you need to know

Casual computer users have probably logging software, but it's used across





https://threatpost.com > Vulnerabilities :

Third Log4J Bug Can Trigger DoS; Apache Issues Patch

Dec 20, 2021 — The latest **bug** isn't a variant of the Log4Shell remote-code execution (RCE) **bug** that's plagued IT teams since Dec. 10, coming under active ...

https://www.scmagazine.com > Application security

Log4j, again, needs patching as new bug is found and ...

Dec 28, 2021 — Researchers at Checkmarx discovered a way to use **Log4j** to launch malicious code, forcing yet another round of patching for affected users.

Packages can be either:

- Libraries:
 - A set of classes and methods that provide reusable functionality
 - Typically: programmer calls, library returns data, that's it.

Packages can be either:

- Libraries:
 - A set of classes and methods that provide reusable functionality
 - Typically: programmer calls, library returns data, that's it.
- Frameworks:
 - Reusable skeleton code that can be customized into an application
 - Framework calls back into client code
 - The Hollywood principle: "Don't call us. We'll call you."
 - E.g., Android development: you declare your UI elements, activities to be composed
 - Principle: <u>inversion of control</u>



Packages can be either:

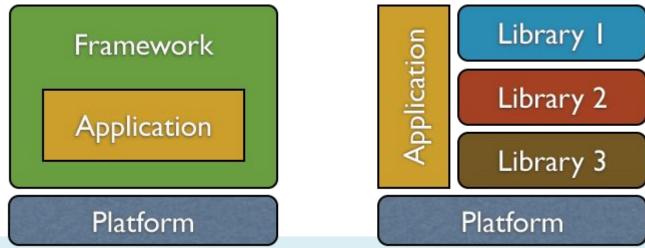
- Libraries:
 - A set of classes and methods that provide reusable functionality
 - Typically: programmer calls, library returns data, that's it.
- Frameworks:
 - Reusable skeleton code that can be customized into an application
 - Framework calls back into client code
 - The Hollywood principle: "Don't call us. We'll call you."
 - o E.g., Android development: you declare your UI elements, activities to be composed
 - Principle: <u>inversion of control</u>
- You typically use zero/one framework and many libraries
 - Frameworks might be especially constraining, but for good reason.
 - Some tools are a bit of both, and not all frameworks quite invert control

institute for SOFTWARE RESEARCH

Which kind is a command-line parsing package?

Which kind is Android?

How about a tool that runs tests based on annotations you add in your code?



Look into:

Stated Goal:

- A simple interface ("get started in one line!") also means lots of abstraction
- That's neither good nor bad; know what you need
- Docs with "advanced use cases" are always neat

Maintenance:

- Active release cycle, recent updates to documentation
- GitHub build status, issue tracker (filled with unmerged 'dependabot' PRs?)
- Lots of companies deliberately lag by one minor (or even major) version

Recursive dependencies

Myriad, beyond inspection. Using OSS in corporate environments is a headache

IST institute for SOFTWARE RESEARCH

Frameworks

Whitebox:

- Extension via subclassing and overriding methods
- Common design pattern(s):
 - Template method
- Subclass has main method but gives control to framework

Blackbox:

- Extension via implementing a plugin interface
- Common design pattern(s):
 - Command
 - Observer
- Plugin-loading mechanism loads plugins and gives control to the framework

Abstraction, Reuse, and Programming Tools

- For each in {Build systems, IDE, libraries, CI}:
 - O What is it?
 - What happens under the hood?
 - What are some practical ways to use it?
- What is next?

IST institute for SOFTWARE RESEARCH

Quick overview of today's toolchain: Continuous Integration

CI: Automates standard build, test, deploy pipelines

(Technically, the latter is "CD")

Typically builds from scratch in a clean *container*

Often tied to code-review; triggers on new commits, pull requests

Ideally, official releases pass the build

Produces (long) logs with debugging outputs

Under the Hood: Continuous Integration

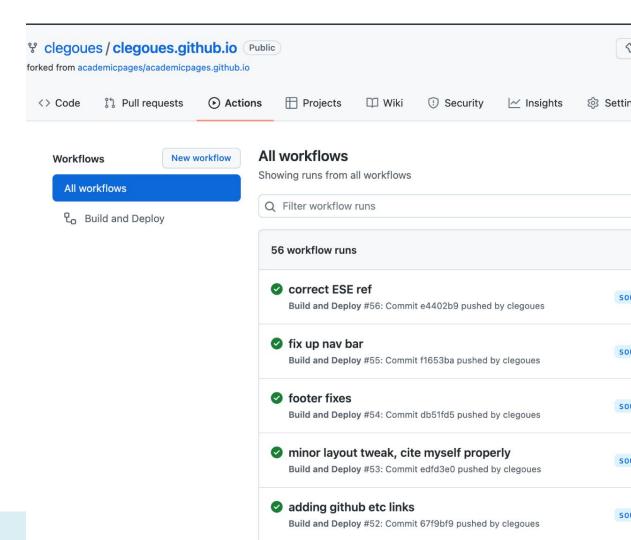
Defines a series of <u>actions</u> to be run in a <u>clean build</u>:

- Actions start from the very top:
 - Clone repository, checkout branch
 - Download & install Java/Node
 - Invoke commands with timeouts
- Travis allocates a new (Docker) container for each build
 - Think of this like a fresh, temporary computer
 - Usually with a few default libraries present (i.e., based on an image)
- That means: fully replicable builds



Continuous integration – GitHub Actions

You can see the results of builds over time



Installing SSH key from: default repository key ssh_key Using /home/travis/.netrc to clone repository. \$ git clone --depth=50 --branch=TypeScript https://github.com/CMU-17-214/template-21f-hw1.git CMU-17-214/template-21f-hw1 1.75s Cloning into 'CMU-17-214/template-21f-hw1'... remote: Enumerating objects: 117, done. remote: Counting objects: 100% (117/117), done. remote: Compressing objects: 100% (73/73), done. remote: Total 117 (delta 50), reused 104 (delta 37), pack-reused 0 Receiving objects: 100% (117/117), 69.89 KiB | 2.25 MiB/s, done. Resolving deltas: 100% (50/50), done. \$ cd CMU-17-214/template-21f-hw1 \$ git checkout -qf 0d657225c8cbdd52751c2f88527f93f4099b041e 0.01s \$ nvm install 16 nvm.install 3.65s Downloading and installing node v16.8.0... Downloading https://nodejs.org/dist/v16.8.0/node-v16.8.0-linux-x64.tar.xz... Computing checksum with sha256sum Checksums matched! Now using node v16.8.0 (npm v7.21.0) Setting up build cache cache.1 cache.npm \$ node --version 196 v16.8.0 \$ npm --version 198 7.21.0 199 \$ nvm --version 0.38.0 \$ npm ci 3.04s install.npm \$ timeout 5m npm run compile 4.93s 213 > hw1-flashcards@1.0.0 compile 214 > tsc

Under the Hood: Continuous Integration

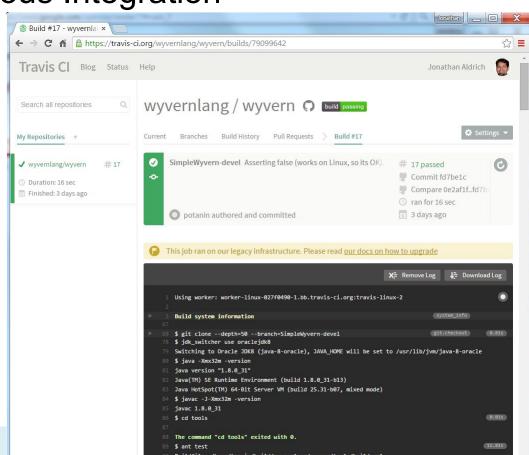
Automatically builds, tests, and displays the result

We – and everyone else – used to use Travis CI.

 Until they randomly stopped supporting OSS.

GitHub has native CI support, and it's pretty good: GitHub Actions.

 Sidebar on how our GH Actions are configured for HW1



Quick overview of today's toolchain: not mentioned

Docker: containerize applications for coarse-grained reuse

Cloud: deploy and scale rapidly, release seamlessly

Bug/Issue trackers, often integrated with reviews



Behind the Abstraction: Some Nuance

- Automation vs. Reuse
 - We tend to automate common chains of actions
 - Gear-up := {Press clutch, switch gear, release clutch while accelerating}
 - To facilitate reusing such "subroutines", we introduce abstractions
 - Accelerate in 'D' => Gear-up when needed
- Reuse vs. Interfaces
 - Interfaces facilitate reuse through abstraction
 - Allow upgrading implementation without breaking things
 - Provide explicit & transparent contract

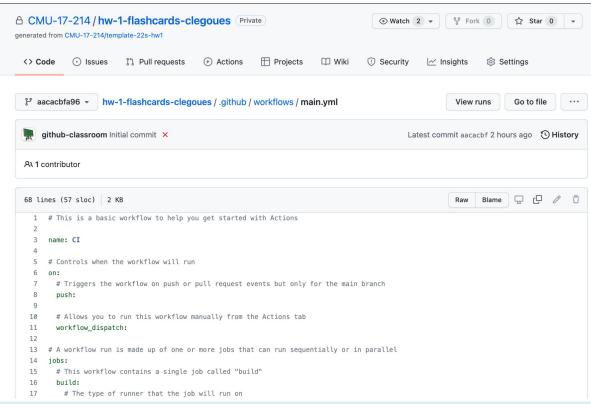


17-214/514 56 IST institute

Behind the Abstraction, Some Nuance

Most tools are abstractions of common commands

- Typically operated via GUI and/or a DSL
- Obvious for GitHub Actions: just read the Yaml
 - Script-like languages are common
 - Involving a vocabulary of "targets"
 - E.g., `mvn site`





Behind the Abstraction, Some Nuance

Most tools are abstractions of common commands

- Typically operated via GUI and/or a DSL
- Obvious for GitHub Actions: just read the Yaml
 - Script-like languages are common
 - Involving a vocabulary of "targets"
 - E.g., `mvn site`

Abstraction can also "trap" us

- When/how do we leave the abstraction?
- Command-line comes built into IDEs for a reason
- Non-trivial in general! May require switching/"patching" libraries
 - E.g., Maven → Gradle for more unusual build routines

institute for SOFTWARE RESEARCH

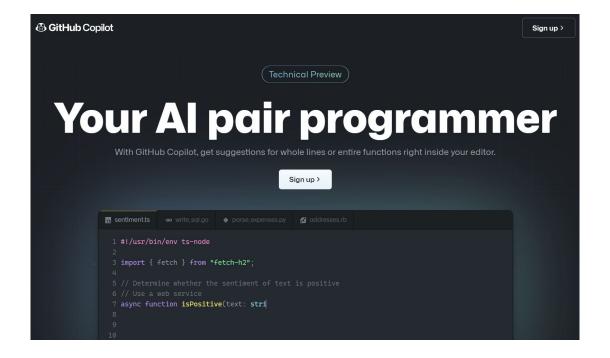
Abstraction, Reuse, and Programming Tools

- For each in {Build systems, IDE, libraries, CI}:
 - O What is it?
 - What happens under the hood?
 - What are some practical ways to use it?
- What is next?

institute for softwar RESEARCH

What's Next: Al Powered Programming

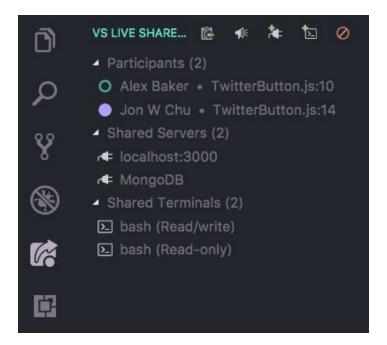
- Easier in Web IDEs
 - Which are themselves "next"



institute for SOFTWARI

What's Next: Collaborative online coding

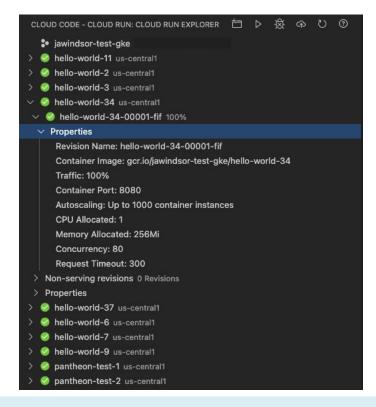
- Think: Google Docs for code
- E.g. VS Live Share
- How will this change "commits"?



institute for SOFTWARE RESEARCH

What's Next: Tighter IDE-to-cloud integration

- Google Cloud is pushing on this with VSCode
- We will (lightly) touch on Containers & Clouds in this course



IST institute for SOFTWARE RESEARCH

Summary

- Programming Tools are abundant, and rapidly evolving
 - Learn multiple; you will have to inevitably
- They rely on abstractions through interfaces to facilitate reuse
 - Which come in many shapes: GUI, API, DSL
 - And can be a limitation -- choose wisely
- Your HW1 toolchain sets you up for all homeworks
 - With modest variations (frameworks, new build targets)
 - Self-discovery is a big asset
 - Recitation should be helpful!

institute for SOFTWARE RESEARCH