



# Refactoring at Google Scale

The Why and How

Hyrum Wright, PhD <[hwright@google.com](mailto:hwright@google.com)>



Refactoring:

A controlled technique for improving the design of an existing code base. Its essence is applying a series of small behavior-preserving transformations, each of which "too small to be worth doing."

-Martin Fowler

```
void foo(bool x) {  
    // Analyse the colour of the fibre  
    Frob(x);  
}
```

```
void foo(bool x) {  
    // Analyze the color of the fiber  
    Frob(x);  
}
```

```
class File {  
    static string JoinPath(const string& path1, const string& path2);  
}
```

```
string path1 = File::JoinPath("a", "b");  
assert(path1 == "a/b");
```

```
string path2 = File::JoinPath("a", "/b");  
assert(path2 == "/b");    // WAT
```

```
namespace file {  
namespace internal {  
    string JoinPathImpl(std::initializer_list<StringPiece> paths);  
    string JoinPathAbsoluteImpl(  
        std::initializer_list<StringPiece> paths);  
}
```

```
template<typename... T>  
string JoinPath(const T&... args) {  
    return internal::JoinPathImpl({args...});  
}
```

```
template<typename... T>  
string JoinPathAbsolute(const T&... args) {  
    return internal::JoinPathAbsoluteImpl({args...});  
}  
}
```

```
void foo(bool x) {  
    string path = File::JoinPath(path1, path2);  
}
```

```
void foo(bool x) {  
    string path = file::JoinPath(path1, path2);  
}
```



```
void foo(bool x) {  
    string path = File::JoinPath(path1, path2);  
}
```

```
void foo(bool x) {  
    string path = file::JoinPathAbsolute(path1, path2);  
}
```

```
void foo(bool x) {  
    string path = File::JoinPath(path1, File::JoinPath(path2, path3));  
}
```

```
void foo(bool x) {  
    string path = file::JoinPath(path1, path2, path3);  
}
```

```
void foo(bool x) {  
    vector<string> v;  
    SplitStringUsing("a|b,c|d", "|,", &v);  
}
```

```
void foo(bool x) {  
    vector<string> v =  
        strings::Split("a|b,c|d", strings::AnyOf("|,"));  
}
```

```
void foo(bool x) {  
    vector<string> v;  
    SplitStringUsing("a|b,c|d", ",", &v);  
}
```

```
void foo(bool x) {  
    vector<string> v =  
        strings::Split("a|b,c|d", ",");  
}
```

```
void foo(bool x) {  
    vector<string> v;  
    vector<string> v2;  
    if (x)  
        SplitStringUsing("a,b,c,d", ",", &v2);  
    SplitStringUsing("a,b,c,d", ",", &v);  
}
```

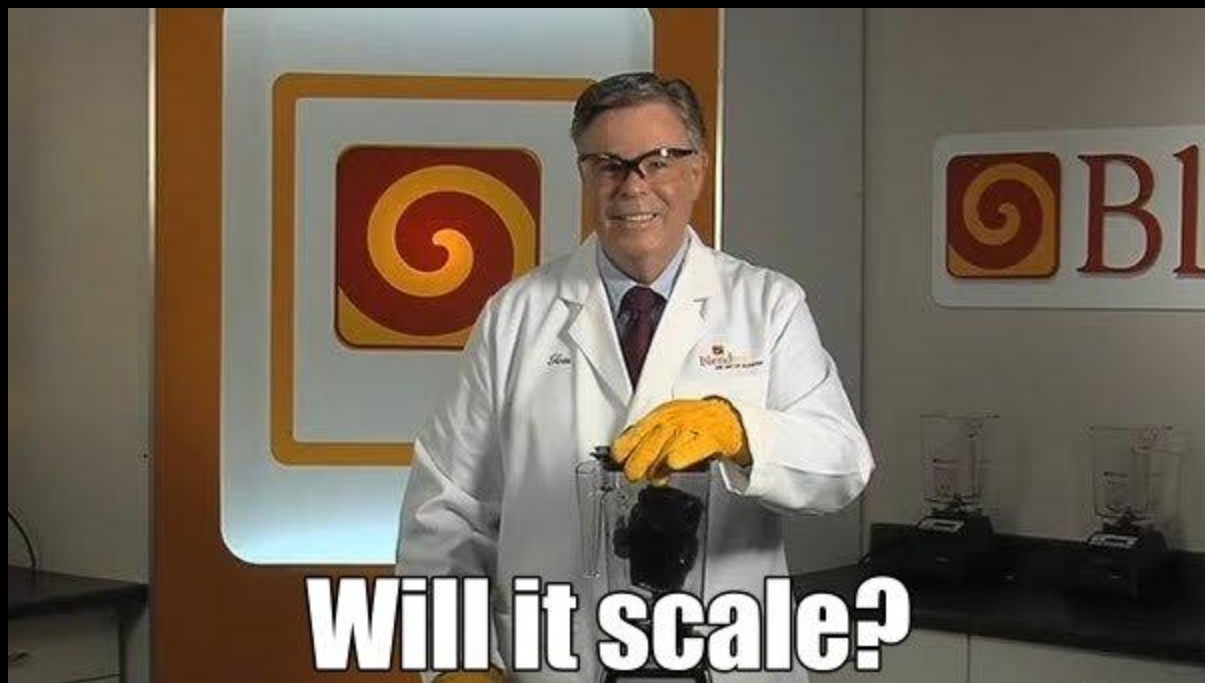
```
void foo(bool x) {  
    vector<string> v2;  
    if (x)  
        SplitStringUsing("a,b,c,d", ",", &v2);  
    vector<string> v = strings::Split("a,b,c,d", ",");  
}
```

```
void foo(bool x) {  
    std::unique_ptr<Foo> f(new Foo());  
    f.get()->DoStuff();  
}
```

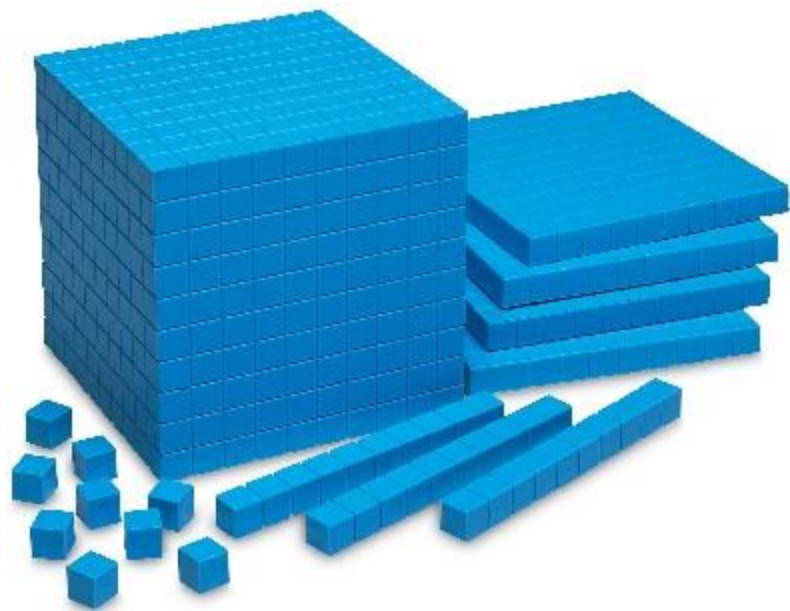
```
void foo(bool x) {  
    std::unique_ptr<Foo> f(new Foo());  
    f->DoStuff();  
}
```

```
void foo(bool x) {  
    map<int, string> m = ...;  
    const string& foo = FindWithDefault(m, "key", "default");  
}
```

```
void foo(bool x) {  
    map<int, string> m = ...;  
    string default = "default";  
    const string& foo = FindWithDefault(m, "key", default);  
}
```









XXXXXXXXXX

[illegible]

[illegible]

[illegible]

WHAT IF WE  
TRIED *MORE*?



10x





# 1 Monolithic Codebase

4,000+ C++ engineers  
working concurrently

100M lines of code  
developed over 15+ years

All changes subject to review-  
before-commit

Massive testing and  
continuous integration  
infrastructure

Prefer the standard thing over  
homegrown solutions







~~By Hand~~

~~By Regular Expression~~

~~By IDE~~





1. Generate the change

```
void foo(bool x) {  
    string path = File::JoinPath(path1, path2);  
}
```

```
void foo(bool x) {  
    string path = file::JoinPath(path1, path2);  
}
```



```
void foo(bool x) {  
    string path = File::JoinPath(path1, path2);  
}
```

```
void foo(bool x) {  
    string path = file::JoinPath(path1, path2);  
}
```

OR

```
void foo(bool x) {  
    string path = file::JoinPathAbsolute(path1, path2);  
}
```

ClangMR

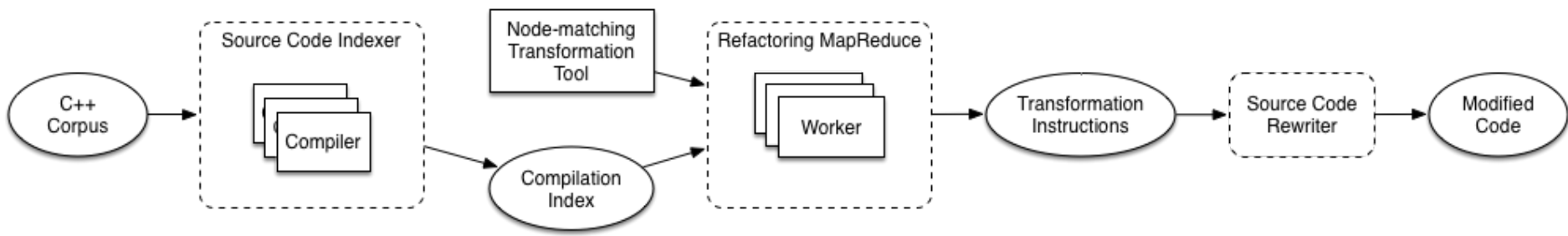
# ClangMR = Clang + MapReduce

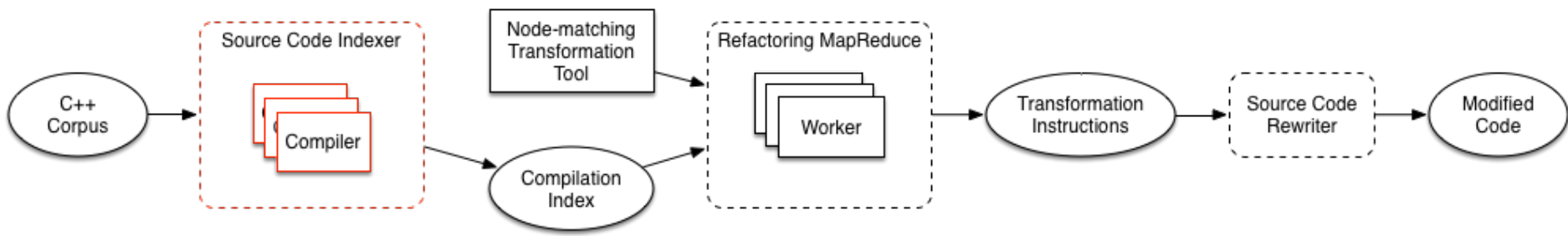
## Clang:

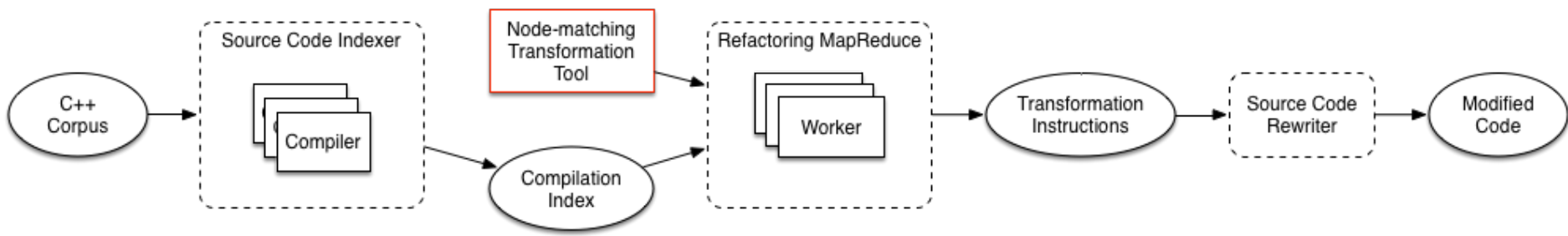
- Provides library infrastructure for AST node matching and traversal

## MapReduce:

- A framework of processing data on a massively parallel scale [Dean & Ghemawat, OSDI '04]







Matcher< <a href="#">Decl</a> >	constructorDecl	Matcher< <a href="#">CXXConstructorDecl</a> >...
Matcher< <a href="#">Decl</a> >	decl	Matcher< <a href="#">Decl</a> >...
Matcher< <a href="#">Decl</a> >	declaratorDecl	Matcher< <a href="#">DeclaratorDecl</a> >...
Matcher< <a href="#">Decl</a> >	destructorDecl	Matcher< <a href="#">CXXDestructorDecl</a> >...

Matches explicit C++ destructor declarations.

```
Example matches Foo::~~Foo()
class Foo {
public:
    virtual ~Foo();
};
```

---

Matcher< <a href="#">Decl</a> >	enumConstantDecl	Matcher< <a href="#">EnumConstantDecl</a> >...
Matcher< <a href="#">Decl</a> >	enumDecl	Matcher< <a href="#">EnumDecl</a> >...
Matcher< <a href="#">Decl</a> >	fieldDecl	Matcher< <a href="#">FieldDecl</a> >...
Matcher< <a href="#">Decl</a> >	friendDecl	Matcher< <a href="#">FriendDecl</a> >...
Matcher< <a href="#">Decl</a> >	functionDecl	Matcher< <a href="#">FunctionDecl</a> >...
Matcher< <a href="#">Decl</a> >	functionTemplateDecl	Matcher< <a href="#">FunctionTemplateDecl</a> >...
Matcher< <a href="#">Decl</a> >	methodDecl	Matcher< <a href="#">CXXMethodDecl</a> >...
Matcher< <a href="#">Decl</a> >	namedDecl	Matcher< <a href="#">NamedDecl</a> >...
Matcher< <a href="#">Decl</a> >	namespaceDecl	Matcher< <a href="#">NamespaceDecl</a> >...

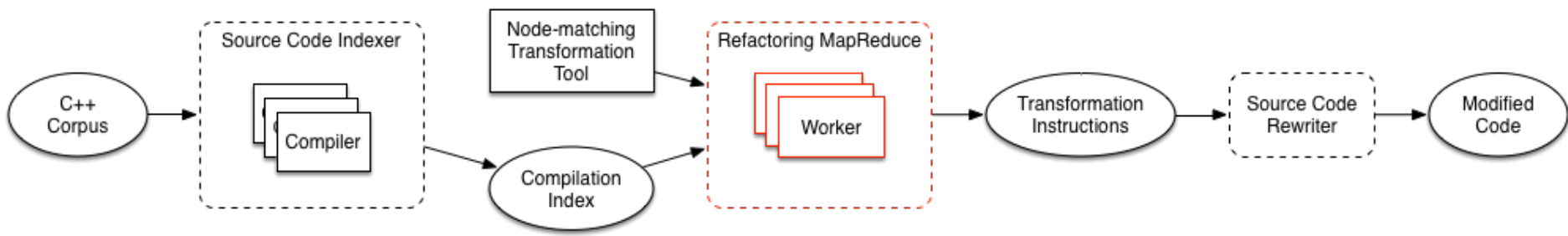
```
StatementMatcher joinpath_call =  
    callExpr(callee(functionDecl(hasName("::File::JoinPath"))));  
  
// For nested JoinPath-calls, only refactor the innermost one.  
StatementMatcher match =  
    callExpr(  
        callee(functionDecl(hasName("::File::JoinPath"))),  
        unless(hasAnyArgument(hasDescendant(joinpath_call)))  
        .bind("call");  
  
match_finder->addMatcher(match, callback_.get());
```

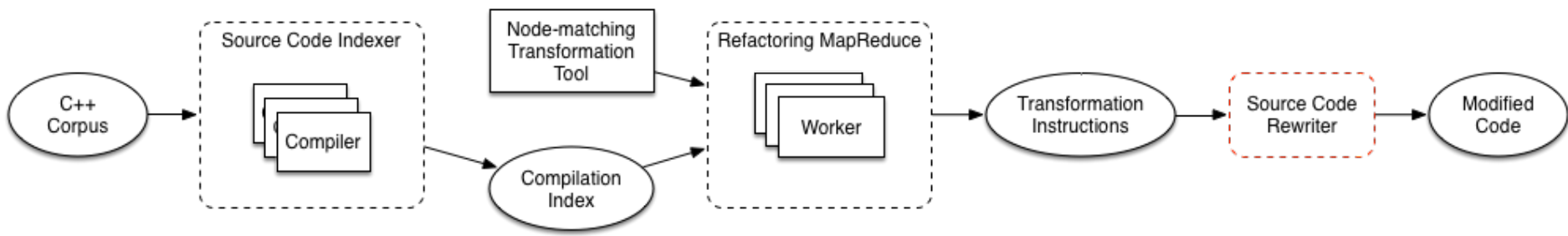


```
const clang::CallExpr* call =  
    result.Nodes.getStmtAs<clang::CallExpr>("call");  
std::unique_ptr<cymbal::EditState> edit(CreateEditState(call));
```

```
const clang::Expr* arg2 = call->getArg(1);  
switch (GetPathArgumentKind(result, arg2)) {  
    // Argument specific behavior  
}
```

```
Report(edit.get());
```





## 1.5 Format the resulting change

## 2. Shard and test the change

3. Review each shard

4. Submit to the version control system

# Flaky Tests



Fast moving codebase

Correct transformation isn't always known

# War story: `shared_ptr` → `std::shared_ptr` migration

## The Goal:

- Migrate all references from a custom `shared_ptr` class to use `std::shared_ptr` from `<memory>`

## War story: `shared_ptr` → `std::shared_ptr` migration

What worked:

- Old class defined in a common header file
- Old class implementation a strict subset of standard version
- typedef'ed the new type to the existing one
- Add `std::` prefix everywhere (and reformat line overflows)

War story: `shared_ptr` → `std::shared_ptr` migration

What didn't:

- Nothing (almost)

We got *really* lucky.

# War story: `scoped_ptr` → `std::unique_ptr` migration

## The Goal:

- Migrate all references from a custom `scoped_ptr` class to use `std::unique_ptr` from `<memory>`

```
void foo(bool x) {  
    scoped_ptr<Foo> f(new Foo());  
}
```

```
#include <memory>
```

```
void foo(bool x) {  
    std::unique_ptr<Foo> f(new Foo());  
}
```

# War story: `scoped_ptr` → `std::unique_ptr` migration

## The Problems:

- Subtle API differences between `scoped_ptr` and `std::unique_ptr`
- Use spanning API boundaries means changes are not independent
- Some code still built with non-C++11 toolchains
- Many different `scoped_ptr`s defined in our codebase, but only references to one should be updated



Given sufficient use,  
there is no such thing as a private implementation.

Refactoring at Google Scale:  
A typical problem that gets a whole lot more interesting.

# Questions?

[hwright@google.com](mailto:hwright@google.com)

