

#### Dipartimento di Scienze Fisiche, Informatiche e Matematiche

# 3. LAB 1: Introduzione ai passi LLVM

Compilatori – Middle end [1215-014]

Corso di Laurea in INFORMATICA (D.M.270/04) [16-215] Anno accademico 2024/2025 **Prof. Andrea Marongiu** andrea.marongiu@unimore.it

### Copyright note

È vietata la copia e la riproduzione dei contenuti e immagini in qualsiasi forma.

È inoltre vietata la redistribuzione e la pubblicazione dei contenuti e immagini non autorizzata espressamente dall'autore o dall'Università di Modena e Reggio Emilia.

### Credits

- Cooper, Torczon, "Engineering a Compiler", Elsevier
- Aho, Lam, Sethi, Ullman, "Compilatori: principi, tecniche e strumenti seconda edizione", Pearson
- Gibbons, Carnegie Mellon University, "Optimizing Compilers"
- Pekhimenko, University of Toronto, "Compiler Optimization"



#### Dipartimento di Scienze Fisiche, Informatiche e Matematiche

### **INSTALLAZIONE LLVM 19**

### Opzione 1) APT – Dai repository ufficiali

https://apt.llvm.org/

#### Example (on Ubuntu Jammy Jellyfish)

```
wget -0 - https://apt.llvm.org/llvm-snapshot.gpg.key |
sudo apt-key add -
sudo apt-add-repository "deb http://apt.llvm.org/jammy/
llvm-toolchain-jammy-19 main"
sudo apt-get update
sudo apt-get install -y llvm-19 llvm-19-dev llvm-19-
tools clang-19
```

### **Opzione 2) Binaries – Dai repository ufficiali**

https://github.com/llvm/llvm-project/releases/tag/llvmorg-19.1.7

♥IIdb-19.1.7.src.tar.xz	10.2 MB	Jan 14
<b>☆</b> Ildb-19.1.7.src.tar.xz.sig	438 Bytes	Jan 14
♦ LLVM-19.1.7-Linux-X64.tar.xz	1.54 GB	Jan 15
♦ LLVM-19.1.7-Linux-X64.tar.xz.jsonl	10.2 KB	Jan 15
<b>☆LLVM-19.1.7-macOS-ARM64.tar.xz</b>	1.32 GB	Jan 15
<b>☆LLVM-19.1.7-macOS-ARM64.tar.xz.jsonl</b>	10.1 KB	Jan 15
♥LLVM-19.1.7-macOS-X64.tar.xz	1.47 GB	Jan 15
♥LLVM-19.1.7-macOS-X64.tar.xz.jsonl	10.2 KB	Jan 15
<b>☆LLVM-19.1.7-win32.exe</b>	311 MB	Jan 15
<b>☆LLVM-19.1.7-win32.exe.sig</b>	543 Bytes	Jan 15

### Opzione 3) Sorgente – Dai repository ufficiali

https://github.com/llvm/llvm-project/releases/tag/llvmorg-19.1.7

♦ polly-19.1.7.src.tar.xz.sig	In fondo alla pagina	438 Bytes	Jan 14
<b>⊘</b> runtimes-19.1.7.src.tar.xz		7.05 KB	Jan 14
<b>⊘</b> runtimes-19.1.7.src.tar.xz.sig		438 Bytes	Jan 14
<b>⇔</b> sources.jsonl		200 KB	Jan 14
♦ test-suite-19.1.7.src.tar.xz		163 MB	Jan 14
test-suite-19.1.7.src.tar.xz.sig		438 Bytes	Jan 14
third-party-19.1.7.src.tar.xz		444 KB	Jan 14
third-party-19.1.7.src.tar.xz.sig		438 Bytes	Jan 14
Source code (zip)			Jan 14
Source code (tar.gz)			Jan 14

 Creare nella vostra ROOT directory la seguente struttura

```
amarongiu@kernighan:~/workspace/LLVM_17$ ll
total 28
drwxr-xr-x 6 amarongiu compilers 4096 Mar 6 18:04 ./
drwxr-x--- 6 amarongiu compilers 4096 Mar 3 12:17 ../
drwxr-xr-x 19 amarongiu compilers 4096 Mar 7 15:59 BUILD/
drwxr-xr-x 7 amarongiu compilers 4096 Mar 4 12:49 INSTALL/
-rw-r---- 1 amarongiu compilers 88 Mar 6 18:04 setup.sh
drwxr-xr-x 3 amarongiu compilers 4096 Mar 3 12:27 SRC/
drwxr-xr-x 3 amarongiu compilers 4096 Mar 6 18:00 TEST/
```

 Il file zip/tar appena scaricato va copiato in SRC, e lì scompattato

```
amarongiu@kernighan:~/workspace/LLVM_17$ ll
total 28
drwxr-xr-x 6 amarongiu compilers 4096 Mar 6 18:04 ./
drwxr-x--- 6 amarongiu compilers 4096 Mar 3 12:17 ../
drwxr-xr-x 19 amarongiu compilers 4096 Mar 7 15:59 BUILD/
drwxr-xr-x 7 amarongiu compilers 4096 Mar 4 12:49 INSTALL/
-rw-r---- 1 amarongiu compilers 88 Mar 6 18:04 setup.sh
drwxr-xr-x 3 amarongiu compilers 4096 Mar 3 12:27 SRC/
```

 La cartella BUILD conterrà il prodotto della compilazione dei sorgenti

```
amarongiu@kernighan:~/workspace/LLVM_17$ ll
total 28
drwxr-xr-x 6 amarongiu compilers 4096 Mar 6 18:04 ./
drwxr-x--- 6 amarongiu compilers 4096 Mar 3 12:17 ../
drwxr-xr-x 19 amarongiu compilers 4096 Mar 7 15:59 BUILD/
drwxr-xr-x 7 amarongiu compilers 4096 Mar 4 12:49 INSTALL/
-rw-r---- 1 amarongiu compilers 88 Mar 6 18:04 setup.sh
drwxr-xr-x 3 amarongiu compilers 4096 Mar 3 12:27 SRC/
drwxr-xr-x 3 amarongiu compilers 4096 Mar 6 18:00 TEST/
```

- La cartella INSTALL conterrà i binari dei vari tools
  - clang, opt, llc, ...
- E le librerie, gli include files, etc.

```
amarongiu@kernighan:~/workspace/LLVM_17$ ll
total 28
drwxr-xr-x 6 amarongiu compilers 4096 Mar 6 18:04 ./
drwxr-x--- 6 amarongiu compilers 4096 Mar 3 12:17 ../
drwxr-xr-x 19 amarongiu compilers 4096 Mar 7 15:59 BUILD/
drwxr-xr-x 7 amarongiu compilers 4096 Mar 4 12:49 INSTALL/
-rw-r---- 1 amarongiu compilers 88 Mar 6 18:04 setup.sh
drwxr-xr-x 3 amarongiu compilers 4096 Mar 3 12:27 SRC/
drwxr-xr-x 3 amarongiu compilers 4096 Mar 6 18:00 TEST/
```

- La cartella TEST conterrà le nostre esercitazioni
  - Ovvero i programmi che saranno oggetto delle nostre ottimizzazioni/analisi

```
amarongiu@kernighan:~/workspace/LLVM_17$ ll
total 28
drwxr-xr-x 6 amarongiu compilers 4096 Mar 6 18:04 ./
drwxr-x--- 6 amarongiu compilers 4096 Mar 3 12:17 ../
drwxr-xr-x 19 amarongiu compilers 4096 Mar 7 15:59 BUILD/
drwxr-xr-x 7 amarongiu compilers 4096 Mar 4 12:49 INSTALL/
-rw-r--r-- 1 amarongiu compilers 88 Mar 6 18:04 setup.sh
drwxr-xr-x 3 amarongiu compilers 4096 Mar 3 12:27 SRC/
drwxr-xr-x 3 amarongiu compilers 4096 Mar 6 18:00 TEST/
```

- Potete crearvi un semplice bash script (setup.sh) per esportare la cartella dei tools INSTALL/bin
  - export PATH=\$ROOT/INSTALL/bin:\$PATH

```
amarongiu@kernighan:~/workspace/LLVM_17$ ll
total 28
drwxr-xr-x 6 amarongiu compilers 4096 Mar 6 18:04 ./
drwxr-x--- 6 amarongiu compilers 4096 Mar 3 12:17 ../
drwxr-xr-x 19 amarongiu compilers 4096 Mar 7 15:59 BUILD/
drwxr-xr-x 7 amarongiu compilers 4096 Mar 4 12:49 INSTALL/
-rw-r---- 1 amarongiu compilers 88 Mar 6 18:04 setup.sh
drwxr-xr-x 3 amarongiu compilers 4096 Mar 3 12:27 SRC/
drwxr-xr-x 3 amarongiu compilers 4096 Mar 6 18:00 TEST/
```

 Così facendo la vostra installazione può convivere con altre versioni di LLVM installate nel sistema

• Potete crearvi un semplice bash script (setup.sh)

```
per e
• e>

s which opt
ROOT/INSTALL/bin/opt
s

amarongiu(
total 28
drwxr-xr-3
```

 Così facendo la vostra installazione può convivere con altre versioni di LLVM installate nel sistema

- L'installazione della toolchain prevede tre steps:
  - Configurazione
  - Compilazione
  - Installazione

- Possiamo fare riferimento alla documentazione ufficiale
  - <a href="https://llvm.org/docs/GettingStarted.html#getting-started-with-llvm">https://llvm.org/docs/GettingStarted.html#getting-started-with-llvm</a>

# Configurazione

• cd ROOT/BUILD

cmake -G "Unix Makefiles"
 -DCMAKE\_BUILD\_TYPE=Release
 -DCMAKE\_INSTALL\_PREFIX=\$ROOT/INSTALL
 -DLLVM\_ENABLE\_PROJECTS="clang"
 -DLLVM\_TARGETS\_TO\_BUILD=host
 [other options]
 \$ROOT/SRC/llvm-project-llvmorg-17.0.6/llvm/

# Compilazione

- cd ROOT/BUILD (dovreste già trovarvi qui)
- make −j2 (l'opzione indica l'utilizzo di più cores per la compilazione)

**NOTA**: L'opzione –j comporta l'incremento dell'utilizzo della RAM. Verificare che il numero di cores indicato non comporti un sovraccarico della stessa ed eventualmente ridurlo (o disattivare completamente l'opzione)

### Installazione

- cd ROOT/BUILD (dovreste già trovarvi qui)
- make install
- Al termine del processo tutti i tools si troveranno installati in \$ROOT/INSTALL



#### Dipartimento di Scienze Fisiche, Informatiche e Matematiche

### Scrivere un passo LLVM

# Ricorda: I passi LLVM

- Abbiamo visto che il middle-end è organizzato come una sequenza di passi
  - *Passi* di analisi
    - Consumano la IR e raccolgono informazioni sul programma
  - Passi di trasformazione
    - Trasformano il programma e producono nuova IR
- Perché esiste questo isolamento?
  - Migliore leggibilità
  - Diversi passi potrebbero richiedere la stessa informazione
  - L'isolamento evita analisi ridondante
- Per analizzare il codice e trasformarlo occorre una IR espressiva che mantenga tutte le informazioni importanti da trasmettere da un passo all'altro

### La IR di LLVM

 La IR di LLVM ha una sintassi ed una semantica simile a quelle del linguaggio Assembly a cui siete abituati (es., RISC-V)

```
int main()
{
    return 0;
}
```

```
define i32 @main() ...
{
    ret i32 0
}
```

### Come scrivere un passo LLVM?

- Per rispondere a questa domanda bisogna prima comprendere i seguenti punti:
  - Moduli LLVM:
    - Come è tradotto il nostro programma in LLVM?
  - Iteratori:
    - Come attraversare il modulo?
  - Downcasting:
    - Come ricavare maggiori informazioni dagli iteratori?
  - Interfacce dei passi LLVM:
    - Che interfacce fornisce LLVM per scrivere i passi?

https://llvm.org/docs/WritingAnLLVMNewPMPass.html

https://llvm.org/docs/WritingAnLLVMPass.html

### Moduli LLVM

#### Il nostro programma

• Files • M

- Funzioni
- Basic Blocks
- Istruzioni

#### **Un Modulo LLVM**

- Module
  - lista di Function e variabili globali
- Function
  - lista di BasicBlocks e argomenti
- BasicBlock
  - lista di Instruction
- Istruction
  - Opcode e operandi

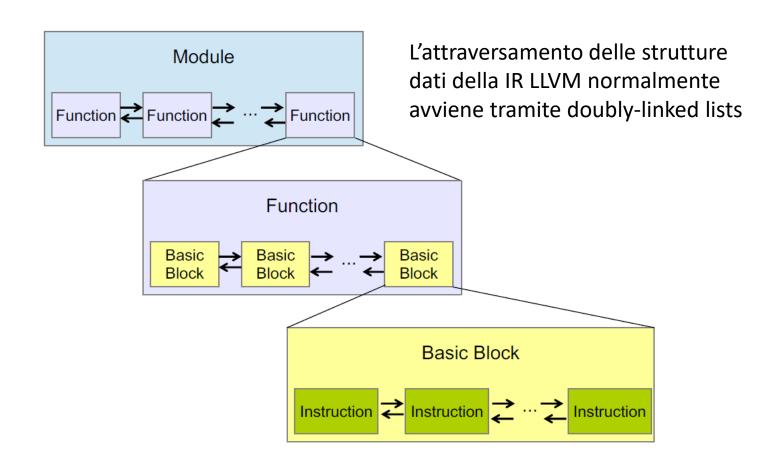
### Iteratori

```
Module &M = ...;
for (auto iter = M.begin(); iter != M.end(); ++iter) {
   Function &F = *iter;
   // do some stuff for each function
}
```

#### Nota:

• Sintassi simile a quella del container STL vector.

### Iteratori



### Downcasting

- Perché ci serve il **Downcasting**?
- Immaginiamo di avere una Instruction
  - Come facciamo a sapere se è un'istruzione unaria o binaria?
  - O capire se è una branch instruction o una call instruction?
- Il **Downcasting** ci aiuta a recuperare maggiore informazione dagli Iterators
- Es.:

# Interfacce dei passi LLVM

- LLVM fornisce diverse interfacce per i passi
  - BasicBlockPass: itera sui basic blocks
  - CallGraphSCCPass: itera sui nodi del call graph
  - FunctionPass: itera sulla lista delle funzioni nel modulo
  - LoopPass: itera sui loops, in ordine inverso di nesting
  - ModulePass: generico passo interprocedurale
  - RegionPass: itera sulle SESE regions, in ordine inverso di nesting
- Come usare queste interfacce?
  - Lo vedremo più avanti

# (New) Pass manager

- Il pass manager del middle-end ha una sequenza di default di applicazione dei passi
- Il default si può alterare invocando una sequenza arbitraria di passi tramite linea di comando

```
$ opt -passes='pass1,pass2' /tmp/a.ll -S
# -p is an alias for -passes
$ opt -p pass1,pass2 /tmp/a.ll -S
```

https://llvm.org/docs/NewPassManager.html https://llvm.org/docs/CommandGuide/opt.html



#### Dipartimento di Scienze Fisiche, Informatiche e Matematiche

### LAB 1 – TEST PASS

### Test Pass – Sorgente

- Scaricate dalla pagina Moodle i files di test per l'esercitazione 1
  - Fibonacci.c
  - Loop.c

Per produrre la IR da dare in pasto al middle-end ci serve clang (il frontend)
 produce la IR solo compilazione (genera assembly)

```
• clang -O2 -emit-llvm -S -c test/Loop.c \
-o test/Loop.ll input file
```

• Oppure si può prima produrre il *bytecode* e poi disassemblare per produrre la forma assembly

clang -02 -emit-llvm -c test/Loop.c \
 -o test/Loop.bc

• llvm-dis test/Loop.bc -o=./test/Loop.ll

 Per ognuno dei test benchmarks produrre la IR con clang e analizzarla, cercando di capire cosa significa ogni parte

Disegnare il CFG per ogni funzione

```
int g;
int g_incr(int c) {
  g += c;
  return g;
}
SORGENTE
```

```
@g = dso_local local_unnamed_addr global i32 0, align 4

define dso_local i32 @g_incr(i32 noundef %0) local_unnamed_addr #0
{
    %2 = load i32, i32* @g, align 4, !tbaa !3
    %3 = add nsw i32 %2, %0
    store i32 %3, i32* @g, align 4, !tbaa !3
    ret i32 %3
}
LLVM IR
```

```
symbol will resolve within
  int q;
                               the same local unit
  int g_incr(int c) {
   q += c;
                                    The address of the object is not
   return q;
                                    known within the module
                   SORGENTE
@g = dso local local unnamed addr global i32 0, align 4
define dso local i32 @g incr(i32 noundef %0) local unnamed addr #0
 %2 = 10ad i32, i32* @g, align 4, !tbaa !3
 %3 = add nsw i32 %2, %0
 store i32 %3, i32* @g, align 4, !tbaa !3
 ret i32 %3
                        Attributes are described later in the file
                                                                  LLVM IR
```

```
; Function Attrs: nofree norecurse nosync nounwind uwtable
define dso local i32 @loop(i32 noundef %0, i32 noundef %1, i32
noundef %2) local unnamed addr #1 {
  %4 = load i32, i32* @q, align 4, !tbaa !3
                                                          IIVM IR
  %5 = icmp sqt i32 %1, %0
 br i1 %5, label %6, label %10
6:
                             ; preds = %3
  %7 = \text{sub i} 32 %1, %0
  %8 = \text{mul i} 32 \%7, \%2
  %9 = add i32 %4, %8
  store i32 %9, i32* @q, align 4, !tbaa !3
 br label %10
10:
                             ; preds = %6, %3
  %11 = phi i32 [ %9, %6 ], [ %4, %3 ]
  ret i32 %11 h
                                           int loop(int a, int b, int c) {
                                             int i, ret = 0;
                                             for (i = a; i < b; i++)
                                               g incr(c);
                                                                 SORGENTE
```

# Esercizio 1 – IR e CFG

```
; Function Attrs: nofree norecurse nosync nounwind uwtable
define dso local i32 @loop(i32 noundef %0, i32 noundef %1, i32
noundef %2) local unnamed addr #1 {
  %4 = load i32, i32* @q, align 4, !tbaa !3
                                                          IIVM IR
  %5 = icmp sqt i32 %1, %0
                                             BB1
  br i1 %5, label %6, label %10
                             ; preds = %3
  %7 = \text{sub i} 32 %1, %0
  %8 = \text{mul i} 32 \%7, \%2
  %9 = add i32 %4, %8
  store i32 %9, i32* @g, align 4, !tbaa !3
  br label %10
10:
                             ; preds = %6, %3
  %11 = phi i32 [ %9, %6 ], [ %4, %3 ]
  ret i32 %11 h
                                           int loop(int a, int b, int c) {
                                   BB3
                                             int i, ret = 0;
                                             for (i = a; i < b; i++)
                                               g incr(c);
                                                                 SORGENTE
```

# Esercizio 1 – IR e CFG

```
; Function Attrs: nofree norecurse nosync nounwind uwtable
   define dso local i32 @loop(i32 noundef %0, i32 noundef %1, i32
   noundef %2) local unnamed addr #1 {
     %4 = load i32, i32* @q, align 4, !tbaa !3
                                                             IIVM IR
     %5 = icmp sqt i32 %1, %0
    br i1 %5, label %6, label %10
   6:
                                ; preds = %3
     %7 = \text{sub i} 32 %1, %0
     %8 = \text{mul i32 } %7, %2
     %9 = add i32 %4, %8
     store i32 %9, i32* @q, align 4, !tbaa !3
     br label %10
   10:
                                ; preds = %6, %3
     %11 = phi i32 [ %9, %6 ], [ %4, %3 ]
     ret i32 %11 h
                                              int loop(int a, int b, int c) {
                                                int i, ret = 0;
                                                for (i = a; i < b; i++)
Cos'è successo al loop?
                                                  g incr(c);
                                                                    SORGENTE
```

### Esercizio 1 – IR e CFG

- Provare a rigenerare la IR con l'opzione –00 al comando di invocazione di clang
- Cosa cambia nell'intermedio?
  - Provate a visualizzare le due IR affiancate nell'editor
- Perché?
  - Provate ad aggiungere il flag —Rpass=.\* al comando di invocazione di clang (con –O2)

https://llvm.org/docs/WritingAnLLVMNewPMPass.html

- Per questo corso scriveremo i nostri passi di analisi e ottimizzazione come plugin per il pass manager di LLVM
- Questo ci consente uno sviluppo esterno al build tree di LLVM
  - Il compilato verrà poi linkato come una dynamic library
- Createvi un workspace con una directory root
- Es.
  - mkdir Laboratori\_Compilatori
  - export ROOT\_LABS=<path/to>Laboratori\_Compilatori

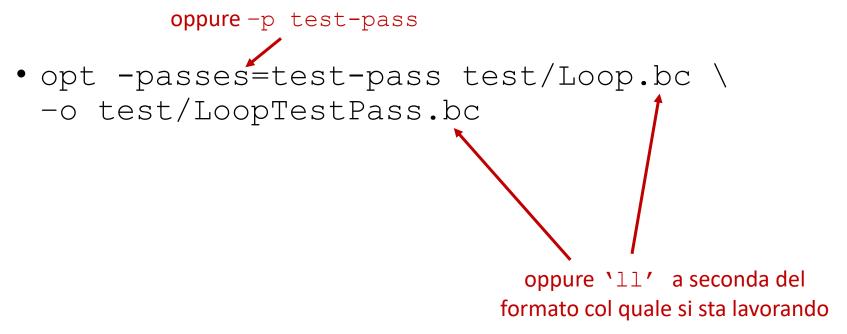
https://llvm.org/docs/WritingAnLLVMNewPMPass.html

- Scaricate da Moodle in ROOT\_LABS/Lab1 i due file
  - TestPass.cpp
  - CMakeLists.txt

- Quindi provate a settare l'environment e compilare
  - export LLVM\_DIR=<installation/dir/of/llvm/19>
  - mkdir build
  - cd build
  - cmake -DLT\_LLVM\_INSTALL\_DIR=\$LLVM\_DIR <source/dir/test/pass>/
  - make

- Scaricate da Moodle in ROOT\_LABS/Lab1/Test i due file
  - Loop.c
  - Fibonacci.c

- Possiamo ora invocare l'ottimizzatore (opt) con il flag per fare overriding del pass manager di default
  - indichiamo noi la lista di passi da applicare



- Possiamo ora invocare l'ottimizzatore (opt) con il flag per fare overriding del pass manager di default
  - indichiamo noi la lista di passi da applicare
- opt -passes=test-pass test/Loop.bc \-o test/LoopTestPass.bc

dal momento che il nostro è un passo di analisi che non produce output (non altera la IR) possiamo disabilitare l'output

Rimpiazzare il flag -o ... con -disable-output

• Se avete fatto tutto correttamente l'esecuzione del passo su Loop.ll dovrebbe produrre questo output:

```
g_incr
loop
```

• Qui comincia il nostro lavoro...

- Estendete il passo TestPass di modo che analizzi la IR e stampi alcune informazioni utili per ciascuna della funzioni che compaiono nel programma di test
  - Nome
  - 2. Numero di argomenti ('N+\*' in caso di funzione variadica)(\*)
  - 3. Numero di chiamate a funzione nello stesso modulo
  - 4. Numero di *Basic Blocks*
  - 5. Numero di *Istruzioni*
- (\*) es., per la funzione int printf (const char \*format, ...); bisogna stampare '1+\*'

https://llvm.org/docs/WritingAnLLVMNewPMPass.html

- Predisponiamo lo scheletro per un nuovo passo
- Tutti i passi LLVM ereditano da CRTP¹ mix-in PassInfoMixin<PassT>.
- Ogni passo deve implementare un metodo run()
   che ritorna un oggetto PreservedAnalysis e
   prende in ingresso una qualche unità IR e un
   analysis manager.
  - Es., per un Function pass
  - PreservedAnalyses run(Function &F, FunctionAnalysisManager &AM);

- La documentazione è nostra amica
  - Indice delle classi per la IR LLVM
  - https://llvm.org/doxygen/classes.html

• Trasformare il Function Pass in un Module Pass, preservando la sua funzionalità