Consolidation and Performance measurements of ROOT Multiproc Core

Cern Openlab Lightning Talk

Student: Anda-Catalina Chelba
Supervisor: Gerardo GANIS (EP-SFT)

August 18, 2016

Anda-Catalina Chelba

CERNopenlab

"*A software framework for data processing, storage and analysis widely adopted by HEP and other scientific communities.*"

"A software framework for data processing, storage and analysis widely adopted by HEP and other scientific communities."

Data Processing in Root

Serial || Parallel

Anda-Catalina Chelba

Parallel Data Processing in Root

Anda-Catalina Chelba

Parallel Data Processing in Root

Proof   Proof-Lite   MultiProc

Parallel Data Processing in Root

Proof	Proof-Lite	MultiProc

Multi-node parallelism

Drawbacks

- environment settings
- limited scope

Parallel Data Processing in Root

Proof	Proof-Lite	MultiProc
Multi-node parallelism	A re-adaptation of PROOF	
Drawbacks	1 machine, multiple processes	
• environment settings	Successful	
• limited scope	Drawback	
	• inherits the setup technology from PROOF	

Anda-Catalina Chelba

Parallel Data Processing in Root

Proof	Proof-Lite	MultiProc
Multi-node parallelism	A re-adaptation of PROOF	Introduced last year
Drawbacks	1 machine, multiple processes	1 machine, multiple processes
• environment settings	Successful	Goal : fix the problems listed above (generic lists of tasks)
• limited scope	Drawback	More powerful & user-friendly
	• inherits the setup technology from PROOF	Tree processing is a vital functionality, but was never benched => existent bugs, not efficient

MultiProc module : My mission

Consolidate the existent version
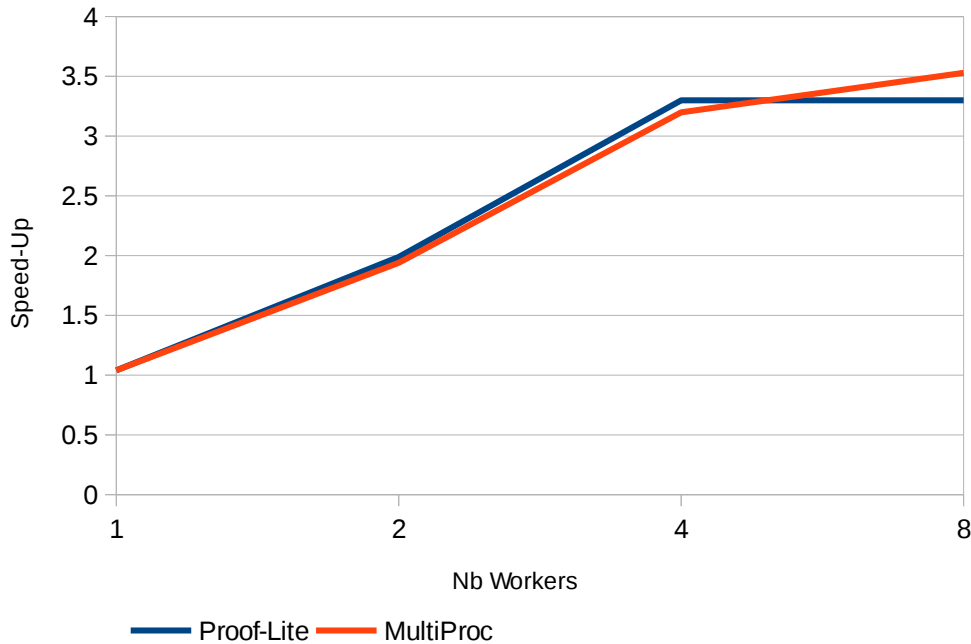
- Complete the tree processing interface & bug fix
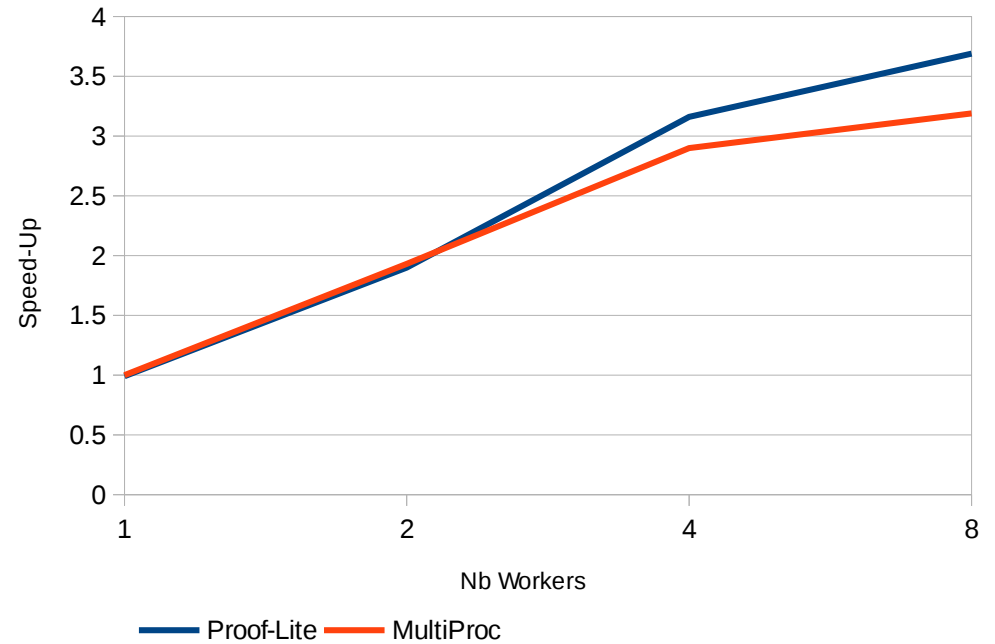
Develop a Bench Tool

Make it more efficient

Anda-Catalina Chelba

The first results

4 files processed

10 files processed

Aug 18, 2016

Anda-Catalina Chelba

MultiProc module

- How it works ?
  - 2 workers, 3 files (same size)
    - Each worker gets one file
    - The third file will be given to the first free worker
  - Problem
    - Work is not split equally

- Current packetizing technology is not efficient

Next steps

New packetizing technology

Finish the Bench Tool development

Anda-Catalina Chelba

Thank you.

Anda-Catalina Chelba