



Contribution ID: 16

Type: poster presentation

A Case Study in Preserving a High Energy Physics Application

The reproducibility of scientific results increasingly depends upon the preservation of computational artifacts. Although preserving a computation to be used later sounds trivial, it is surprisingly difficult due to the complexity of existing software and systems. Implicit dependencies, networked resources, and shifting compatibility all conspire to break applications that appear to work well. Tools are needed which can automatically identify both local and remote dependencies, so that they can be captured and preserved.

To investigate these issues, we present a case study of preserving a CMS application using Parrot. We analyze the application and attempt several methods at extracting its dependencies for the purposes of preservation. We demonstrate a fine-grained dependency management toolkit which can observe both the local filesystem and remote network dependencies, using the system call tracing capabilities of Parrot. We observe that even a simple application depends upon 22,068 files and directories totalling 21 GB of data and software drawn from 8 different sources including cvmfs, hdfs, afs, git, http, cvs, PanFS and local root filesystem.

Once the dependencies are observed, a portable execution package can be generated. This package is not tied to any particular technology and can be re-run using Parrot, Docker, a chroot Jail, or as a virtual machine image, depending on the technology available at the execution site. We will report on the performance and completeness of re-execution using both public and private clouds and offer some guidance for future work in application preservation.

Primary authors: Prof. THAIN, Douglas (University of Notre Dame); MENG, Haiyan (U)

Co-authors: Ms WOODARD, Anna Elizabeth (University of Notre Dame (US)); WOLF, Matthias (University of Notre Dame (US)); Prof. HILDRETH, Michael (University of Notre Dame); Mr IVIE, Peter (University of Notre Dame)

Presenters: Prof. THAIN, Douglas (University of Notre Dame); MENG, Haiyan (U); Prof. HILDRETH, Michael (University of Notre Dame)

Track Classification: Track5: Computing activities and Computing models