A First Look at Reproducibility and Non-Determinism in CMS Software and ROOT Data

Peter Ivie, Charles Zheng, and Douglas Thain
University of Notre Dame
CHEP 2016
What we usually work on:

- HTCondor
- Lobster Data Analysis
- Global Filesystems

Today, a different question:

What happens if we attempt to run the same thing twice?
No Cure

When Bayer tried to replicate results of 67 studies published in academic journals, nearly two-thirds were only partially replicated. Only 11.9% were not applicable.

Source: Nature
The massive data sets accumulated by High Energy Physics (HEP) experiments represent the most direct result of the often decades-long process of construction, commissioning and data acquisition that characterize this science. Many of these data are unique and represent an irreplaceable resource for potential future studies. Forward-thinking efforts for preservation are necessary now in order to achieve the relevant parameters, analysis paths and software to preserve the usefulness of these rich and varied data sets.

Data and Software Preservation for Open Science. DASPOS, represents an initial exploration of the key technical problems that must be solved to provide appropriate data, software and algorithmic preservation for HEP, including the contexts necessary to understand, trust and reuse the data. While the archiving of HEP data may require some HEP-specific technical solutions, DASPOS will create a template for preservation that will be useful across many different disciplines, leading to a broad, coordinated effort.

“Ten or 20 years ago we might have been able to repeat an experiment. They were simpler, cheaper and on a smaller scale. Today that is not the case. So if we need to re-evaluate the data we collect to test a new theory, or adjust it to a new development, we are going to have to be able to reuse it. That means we are going to need to save it as open data…”

Rolf-Dieter Heuer 2008
Director General, CERN

**First Workshop Scheduled**
The first DASPOS Workshop has been scheduled for Thursday - Friday, March 21-22, 2013, at CERN. More information

**Workshop 1**
2012-12-17 16:11:04
WORKSHOP 1 Establishment of Use Cases for Archived Data and Software in HEP Date: Thursday-Friday...

**Workshop 2**
2012-12-17 16:11:04
WORKSHOP 2 Survey of Commonality with other Disciplines Attendees: Broad participation from many...

---

**Discovery and Coordination**
Series of highly-structured public workshops to define, discuss and document the details of data and software preservation

**Prototyping and Experimentation**
Key areas of research: data and query models and software sustainability models

**The DASPOS Team**
Computer science experts, experienced digital librarians, and experts in data-intensive fields, such as physics, astrophysics and bioinformatics
Different Layers of Reproducibility

(1) Can someone else follow simple instructions and get the same software environment?
   – Requires careful dependency management.

(2) Can I run the exact same code on the same computer twice and get the same result?
   – Requires deterministic computing techniques.

... 

(99) Can someone else construct equivalent expts and observe the same physical laws?
   – Difficult to do this without steps 1-98 working.
Some Reasons for Non-Determinism

• Provenance data intermixed with output.
• Unique identifiers embedded in objects.
• Pseudo-random numbers seeded by environmental data (pid,uid,time)
• Entropic randomness (/dev/urandom)
• Parallelism results in re-ordering of data.
• I/O interactions with OS not consistent.
• OS/Compiler details affect ordering/precision of floating point operations.
A very preliminary look...
An initial experiment.

CVMFS + SCRAM for consistent environment

Same machine, same day, same user, same shell, . . .

Not one of the output files was the same!
What constitutes the same output?

The same bytes? (take a checksum)

The same data? (ignore metadata)

The same physics? (fit to a model)
Structure of a Typical ROOT File

- **TStorageFile**
  - UniqueID
  - CycleID
  - Timestamp

- **TTree**

- **TBranch**

- **TBucket**
  - UniqueID
  - Data
  - Data
  - ...
1. Different structure.
2. Same structure, different content.
3. Same structure, same content, different metadata.
4. Same bytes everywhere.
Now, what’s different?

<table>
<thead>
<tr>
<th>Stage</th>
<th># Objects</th>
<th>Ignored</th>
<th>Not Equal</th>
<th>Structural Equal</th>
<th>Content Equal</th>
<th>Bitwise Equal</th>
</tr>
</thead>
<tbody>
<tr>
<td>LHE</td>
<td>605</td>
<td>9</td>
<td>4</td>
<td>592</td>
<td>580</td>
<td>0</td>
</tr>
<tr>
<td>GENSIM</td>
<td>717</td>
<td>9</td>
<td>1</td>
<td>707</td>
<td>694</td>
<td>0</td>
</tr>
<tr>
<td>DIGI – 1</td>
<td>1171</td>
<td>9</td>
<td>1</td>
<td>1161</td>
<td>1158</td>
<td>0</td>
</tr>
<tr>
<td>DIGI – 2</td>
<td>6563</td>
<td>9</td>
<td>21</td>
<td>6533</td>
<td>6526</td>
<td>0</td>
</tr>
<tr>
<td>DIGI – 3</td>
<td>204</td>
<td>21</td>
<td>7</td>
<td>176</td>
<td>175</td>
<td>0</td>
</tr>
<tr>
<td>MINIAOD</td>
<td>2723</td>
<td>9</td>
<td>4</td>
<td>2710</td>
<td>2671</td>
<td>0</td>
</tr>
</tbody>
</table>

Physics data is not the same.
What is the cost of root-diff?

Running time for comparing various size of root files

For files < RAM, root-diff is faster than standard checksums.

For files > RAM, root-diff is slower, due to non-sequential I/O.
Can we eliminate or prevent some kinds of non-determinism?
Parrot Modifies Syscalls

Use Parrot to mount remote filesystems without requiring root access.

```bash
% parrot_run bash
% cd /cvmfs/cms.cern.ch
% source cmsset_default.sh
% scram
...
```

http://ccl.cse.nd.edu/software/parrot
Use Parrot to Eliminate Some Sources of Non-Determinism

parrot_run --time-warp
  --uid 0
  --hostname thematrix
  --mount /dev/urandom /some/fixed/data

% whoami
root
% uname
thematrix
% date --utc
Mon Jan  1  00:00:05 UTC 2001

http://ccl.cse.nd.edu/software/parrot
Does it work? Almost!

“Time Warp” makes LHE produce bitwise-equal results.

(But only for <121 events.)
**Results of a First Look**

- ROOT-DIFF is a general algorithm that can determine equivalence of ROOT files, ignoring the incidental metadata inherent to ROOT.
- CMS codes (at first look) *do not* produce identical physics data on consecutive runs, when compared with ROOT-DIFF.
- “Time Warp” mode causes LHE to give *bitwise equal* results for a small number of events, therefore the code is sensitive to OS effects, even with a constant seed.
- We have not fully isolated the factors that result in non-deterministic behavior!
- Going forward:
  - Push towards one good deterministic example to be shared.
  - Clearly identify all conventional sources of non-determinism.
  - Evaluate effect of increasing concurrency on determinism.
Example code here:

Questions and Discussion Welcome

Peter Ivie
pivie@nd.edu

Charles Zheng
czheng2@nd.edu

Douglas Thain
dthain@nd.edu

@ProfThain