

Preserving Run Time Environments on Parallel Computers: An Essential aspect for research data management

Gerhard Schneider
Rechenzentrum Universität Freiburg
direktor@rz.uni-freiburg.de

Albert-Ludwigs-Universität Freiburg



UNI
FREIBURG

Science needs data

- Scientific data: generated, collected, measured
 - To serve as a basis to derive concepts
 - Data can be valuable
 - Human time to regain the data
 - Unavailability of sources
 - People may be dead
 - weather
 - Comets come back after 75 years
- Keep the data
 - Once well understood → publication on time&technology independent devices (like books)

Science needs data



- Keep the Digital Data:
 - Storage is not made for eternity
 - Reinvestment necessary
 - Note: archives/libraries do not run at zero cost
 - How to read/display digital data?
 - Word 2.0 – my favourite problem☺
 - Migrate to new formats
 - Very costly – migrate a million to access one file later??
 - Keep the data generating software
 - Emulation as a service

06.10.2016

Preserving Run Time Environments on Parallel Computers

3

Science needs data



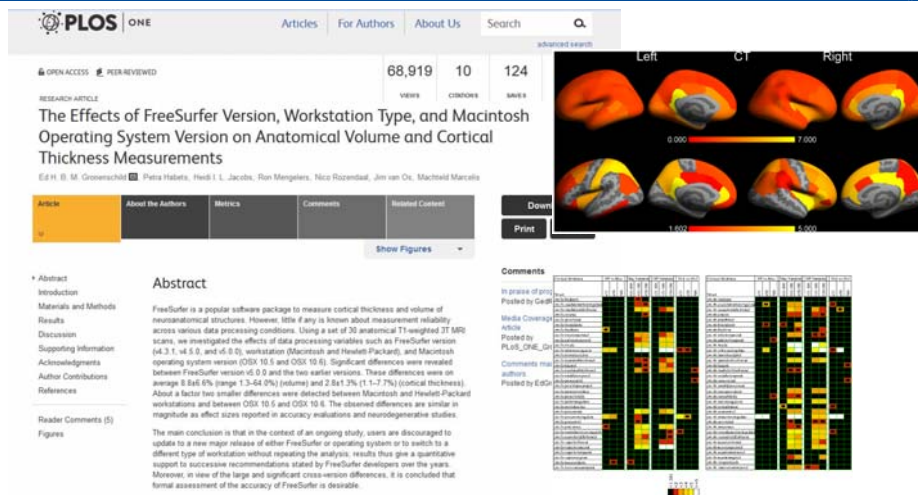
- Keep the (data generating) Software
 - Perhaps you can recompile it later
 - If you can't recompile, keep the binaries
 - build virtual machines to run the binaries with the data
 - Preferably „ready to click“
 - See my talk in 2014 on „emulation as a service“
 - Long term functional archiving
- Keep the complete IT processes
 - Maybe something was wrong in the setup?
- Changing computer centres
 - Just to confuse everyone, we change this setting too...

06.10.2016

Preserving Run Time Environments on Parallel Computers

4

From Data to Processes



<http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0038234>

Source: Andreas Rauber, TU Wien (Vienna, Austria)

Say this again!?

- The display of an x-ray image may depend on the software?
 - A diagnosis may depend on mathematics??? YES!
- real numbers are stored as 64bit or 128bit entities
 - There are infinitely many real numbers but only 2^{128} entities to store them
 - So you have to „cut them short“ – rounding errors
 - $1.01/0.99=1.020202\dots$
 - Round to one decimal: $1.0/1.0 = 1.0$
- When developing mathematical software you must decide on how to treat rounding errors
 - Several theoretically sound approaches! Quite well understood!
 - Things may change after a software update ..

consequences

- It is not enough to „preserve a virtual machine running word“
 - Preserve all the other programs, their interaction, etc
- On the road to virtual research environments
 - Better „virtualized research environments“
 - The programs, the labbook the microscope connectors,....
 - Are these fine with virtualisation?
 - What about licenses?
 - What about users?

14.12.2016

Preserving Run Time Environments on Parallel Computers

7

bwLehrpool

- Unpronounceable word – sorry
- Rapid deployment of run time image
 - Only PXE boot system on the PC
 - Boots a VM image from a server
 - PC runs a virtual environment
- Excellent for class rooms
 - Change rapidly from Maths to Mandarin
- Excellent for e-exams
 - Provide a stripped-down image (security) with only the software needed
 - Large scale possible – several hundred PCs in one exam

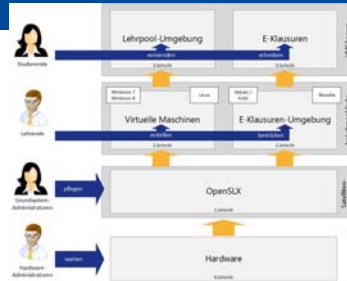
14.12.2016

Preserving Run Time Environments on Parallel Computers

8

bwLehrpool: flexible PC infrastrukt

- separate the tasks:
 - Operation: Computer Center
 - Contents: lecturer



zentrale Infrastruktur	• Kostenreduzierung • Verringerung Administrationsaufwand • Konsolidierung
Raum- und Hardware-unabhängigkeit	• Flexibilität • Räumliche und zeitliche Unabhängigkeit
Abgesicherte Umgebungen	• e-Prüfungen • Komplexe Systemumgebungen • Eigenes Netzwerk
Kooperative Entwicklung von Lehrumgebungen	• Einbindung und Anpassung in Echtzeit • Administrative Unabhängigkeit der Professoren und Lehrbeauftragten

25.05.2016

Preserving Run Time Environments on Parallel Computers

9

Strategic recalibration of IT services

- Aiming for a reasonable redistribution of responsibilities
- No replication of IT-services in the faculty
 - Not an easy job
- But faculty IT may then concentrate on science rather than infrastructure
- An IT-centres can concentrate on fundamental research infrastructures rather than services
- What is the point of this?
 - For this conference?
- Well...

25.05.2016

Preserving Run Time Environments on Parallel Computers

10

Tales from a different world

- Changing computer centers
- A report on recent developments in the German State of Baden-Württemberg
- Parallel computers
 - How to use them --- and what are the consequences for electronic data management ?

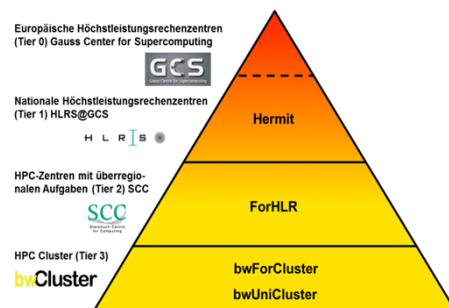
06.10.2016

HPC an der Uni Freiburg

11

HPC in Ba-Wü

- Leistungspyramide
 - Wie Formel-1: den Anfänger lässt man besser nicht auf die Piste
 - Macht zuviel kaputt!!!
 - i.e. vergeudet Rechenzeit
 - i.e. verschwendet Geld
 - (ein Rechner kostet Geld und ist nach fünf Jahren auf Null abgeschrieben)
 - Jede CPU-Sekunde hat einen finanziellen Wert
- Erst mal üben!!
 - Auf dem Desktop, der Workstation
 - Dann auf dem UniversalCluster
 - Dann auf dem ForCluster
 - Dann auf dem ForHLR
 - Und erst dann ...



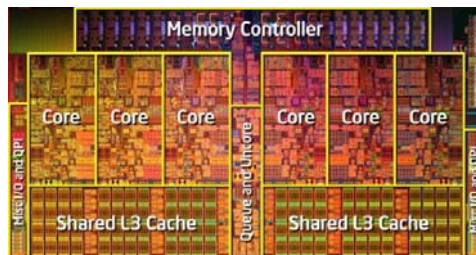
06.10.2016

HPC an der Uni Freiburg

12

Und heute?

- Intel-Prozessoren sind so schnell
 - proprietäre Entwicklungen haben keine Chance mehr!
- Heute so viel Platz auf dem „Die“, dass man mehrere CPUs darauf packen kann
 - Nur „basic CPU“, nicht Cache, Controller, etc
 - → Multi-Core CPU
- Beschleunigung durch „Masse“, nicht durch Spitze!
 - Mehrspännige Kutsche
 - Auch nicht schneller als Einspänner
- Wenn man es schafft, dass alle Kerne unbehindert voneinander an Problemen arbeiten können...
 - „Getrappel“ beim memory controller?



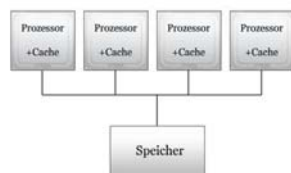
06.10.2016

HPC an der Uni Freiburg

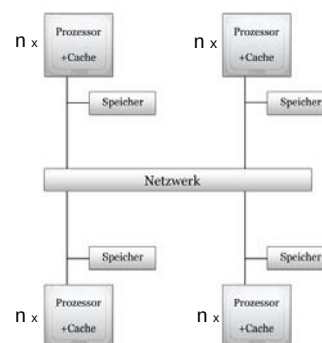
13

Und wenn das nicht reicht?

- Multicore / Multi-CPU
 - Mehrere Prozessoren (mit Cores) mit gemeinsamem Speicher
 - Idealerweise können auch diese unabhängig voneinander arbeiten
- Noch mehr?
 - Dann viele über ein schnelles Netz zusammenschalten



- In der Praxis gibt es wechselseitige „Störungen“ – die Performance wird schlechter ☹



06.10.2016

HPC an der Uni Freiburg

14

HPC: Freiburg



UNI
FREIBURG

- Konfiguration
 - Hochleistungs-Storage mit ½ Petabyte Kapazität und 14 Gigabyte/s Durchsatz
- Aggregierte Ressourcen:
 - 15040 CPU-Kerne
 - 96256 Gigabyte RAM
 - Fast 100 TB RAM ☺
 - Oder: $9,6 \cdot 10^9 \cdot 10\text{KB}$
 - 180480 Gigabyte lokale SSD
 - 180 TB SSD-Speicher
 - Infrastruktur-Betriebsfragen!
 - Für viele Informatiker neues Thema...
- Man könnte für jeden Menschen auf der Welt 10kB Daten im Speicher halten, für schnelle Suche
 - Dies ist eine kleine Maschine



13/04/2016

15

Aber....



UNI
FREIBURG

- „Das Problem“ muss dazu aufgeteilt werden
 - Auf die Knoten...
 - Auf die Cores...
 - So, dass sich möglichst wenig behindert!
 - Nachdenken!
 - Ganz wie in der Fertigung
 - Dort werden die Komponenten auch unabhängig voneinander hergestellt
 - Endmontage
 - Just in time...
 - Hier auch: die Teilergebnisse müssen „eingesammelt“ werden
 - Erst dann weiterrechnen!
- Ein Intel-CPU ist heute mit 2,2 GHz und mehr getaktet
 - $2,2 \cdot 10^9 \text{ Hz}$
- Und wie schnell ist ein Signal?
 - 300000 km/s – Lichtgeschwindigkeit ☺
 - In Kupfer 220.000km/s
 - $220 \cdot 10^6 \text{ m/s}$
 - $2,2 \cdot 10^9 \text{ dm/s}$
- Das heißt: in einem Takt kommt das Signal zur nächsten CPU gerade 10cm weit
 - Die Kabel sind **viel** länger!

06.10.2016

HPC an der Uni Freiburg

16

Neuorientierung – warum?

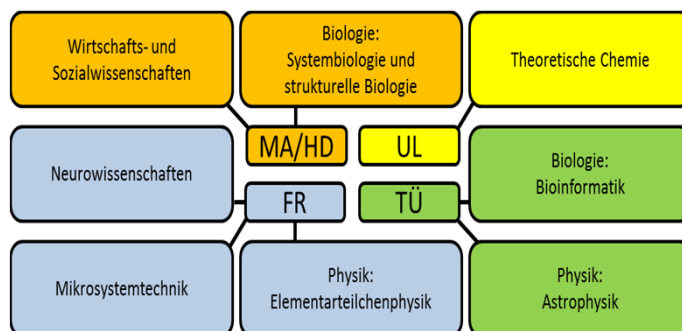
- Klassisch: Das Uni-Rechenzentrum hat den Hochleistungsrechner für die Uni
 - Ein Gerät für alle Disziplinen (?!?)
 - Die wenigen Mitarbeiter (Folge der Kürzungen...)
 - Sollen den Rechner betreiben
 - Alle Software-Pakete betreuen
 - Die unterschiedlichsten Anwender beraten
 - → Überlast / Mehr Personal gibt es nicht
- Spezialisierung der RZs als letzte Antwort auf die gestiegenen Anforderungen

06.10.2016

HPC an der Uni Freiburg

17

HPC in Ba-Wü



- **Allgemeine Ressource**
in KA: bw-
UniCluster

07.12.2016

18

How to archive?

- 15000 cores – 15000 jobs....
- Run virtualisation
- Jobs depend on the number of cores, but good jobs adapt to the resources
- A 15000 cores job can be rerun on 200 cores
 - But will take much longer

06.10.2016

HPC an der Uni Freiburg

19

Challenges

- Loss of performance – lose synchronisation?
- Out of sync anyway, so losses may be minute
- Challenge: high performance network not virtualised, so adapter only in emulation mode
- So far: CERN with machines that use a small number of cores

06.10.2016

HPC an der Uni Freiburg

20

Consequences



- Archive even more data...
- Storage becomes an issue
 - It does not fade away...
 - If you don't get CPU time – fine
 - If you don't get your storage allocated – what to do with the data?
 - Real storage stays expensive – and large datasets make it worse

