Long term data preservation demonstrator using CernVM

Jakob Blomer for the CernVM Team

ALICE Offline Week
March 2014

Agenda

1. Components and Functionality
2. ALEPH software on CernVM / SL4
3. CMS Open Data Pilot on CernVM / SL5

Base Technology: Virtual Machines

Isolation; spawn (historic) software environment on any physical host.

Base Technology: Virtual Machines

Isolation; spawn (historic) software environment on any physical host.

Add-On 1: CernVM File System

CernVM-FS is a *versioning* and *snapshotting* file system used to make the virtual machine's content accountable.
It is used to distribute

- Operating System (Scientific Linux 4–6)
- Experiment software (`/cvmfs/alice.cern.ch/...`)
- Possibly: conditions data (`/cvmfs/alice-ocdb.cern.ch/...`)

Base Technology: Virtual Machines

Isolation; spawn (historic) software environment on any physical host.

Add-On 1: CernVM File System

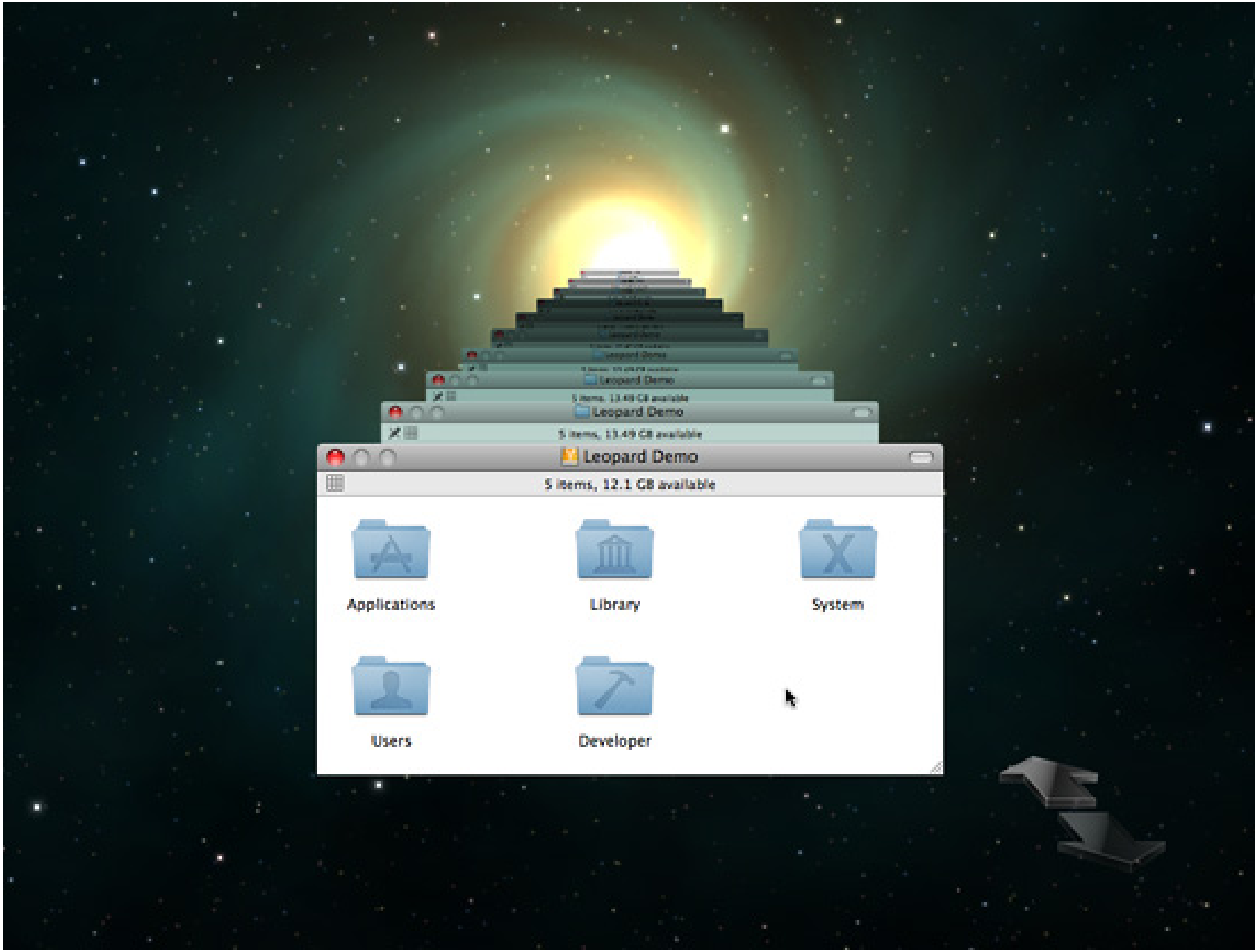
CernVM-FS is a *versioning* and *snapshotting* file system used to make the virtual machine's content accountable.
It is used to distribute

- Operating System (Scientific Linux 4–6)
- Experiment software (`/cvmfs/alice.cern.ch/...`)
- Possibly: conditions data (`/cvmfs/alice-ocdb.cern.ch/...`)

Add-On 2: CernVM Contextualization Agent

Interprets a textual specification for customizing CernVMs.
A generic operating system installation can convert into many roles.

A Time Machine for the Analysis Environment



Example

- For LHCb software stored in CernVM-FS, we can go back to essentially every day until October 2010
- This capability becomes more powerful since we can associate meaningful tags with snapshots

Tag list for the CernVM 3 operating system repository:

jakob — cernvm@cernvm002:~ — ssh — 109×36

```
cernvm@cernvm002:~$ sudo cvmfs_server lstags cernvm-prod.cern.ch
NAME | HASH | SIZE | REVISION | TIMESTAMP | CHANNEL | DESCRIPTION
cernvm-system-3.1.0.0 | fb17e39ca21729a9509fe836fc7f30d26cae1c82 | 14kB | 11 | 28 Jan 2014 14:31:17 | 0 |
cernvm-system-3.1.1.0 | d855c3c05e4fcdb9d5c6f1d0b08c74094f4f5008 | 14kB | 13 | 30 Jan 2014 00:11:10 | 0 |
cernvm-system-3.1.1.1 | 3a06202aadc3b3163b9c5bd36f48b25744f3f204 | 14kB | 16 | 5 Feb 2014 21:03:00 | 0 |
cernvm-system-3.1.1.2 | fc2faf3bc87a2f74da7db22525189b5c582975de | 14kB | 18 | 16 Feb 2014 13:01:32 | 0 |
cernvm-system-3.1.1.3 | fc0d2515c9e79f9fd3cf8b01eac0a16746f4f6cb | 14kB | 20 | 4 Mar 2014 09:26:27 | 0 |
cernvm-system-3.1.1.4 | 314d93015ce473d9a6c99a7365dd4ce38b4e7b13 | 14kB | 22 | 17 Mar 2014 11:07:02 | 0 |
HEAD | 314d93015ce473d9a6c99a7365dd4ce38b4e7b13 | 14kB | 22 | 17 Mar 2014 11:07:10 | 0 |
```

Use Cases for a Preserved Software Environment

1. Processing of legacy data
   - Software implicitly encodes knowledge about the correct interpretation of the data
   - After substantial upgrades and modifications of the detector, the new software might lose this legacy knowledge
   - After experiment decommission, porting and validation of software is likely to end
   - Porting and validation will at some point become prohibitively expensive or just impossible

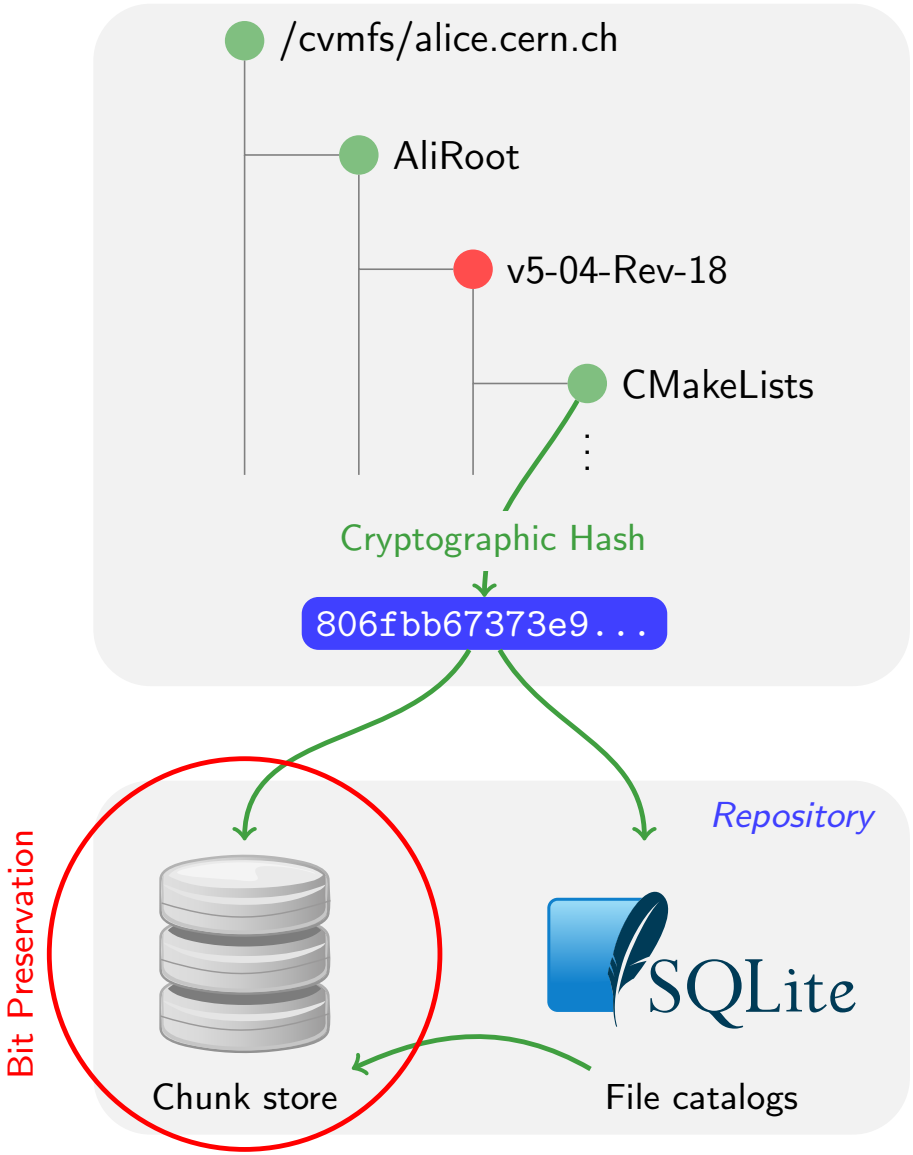
Use Cases for a Preserved Software Environment

1. Processing of legacy data
   - Software implicitly encodes knowledge about the correct interpretation of the data
   - After substantial upgrades and modifications of the detector, the new software might lose this legacy knowledge
   - After experiment decommission, porting and validation of software is likely to end
   - Porting and validation will at some point become prohibitively expensive or just impossible
2. Validation of new software versions (see talk by S. Roiser)
   - Comparison with historic version provides input for validation

Use Cases for a Preserved Software Environment

1. Processing of legacy data
   - Software implicitly encodes knowledge about the correct interpretation of the data
   - After substantial upgrades and modifications of the detector, the new software might lose this legacy knowledge
   - After experiment decommission, porting and validation of software is likely to end
   - Porting and validation will at some point become prohibitively expensive or just impossible
2. Validation of new software versions (see talk by S. Roiser)
   - Comparison with historic version provides input for validation
3. Stable environment for education (cf. CMS Open Data Pilot)
   - Stable operating system and experiment software version accompanies "open data" set and well-defined analysis tasks
   - Driver for data preservation:
     Opportunity to streamline data format and documentation
     Disentangle from grid environment

Versioning and Snapshots in CernVM-FS



Data Store

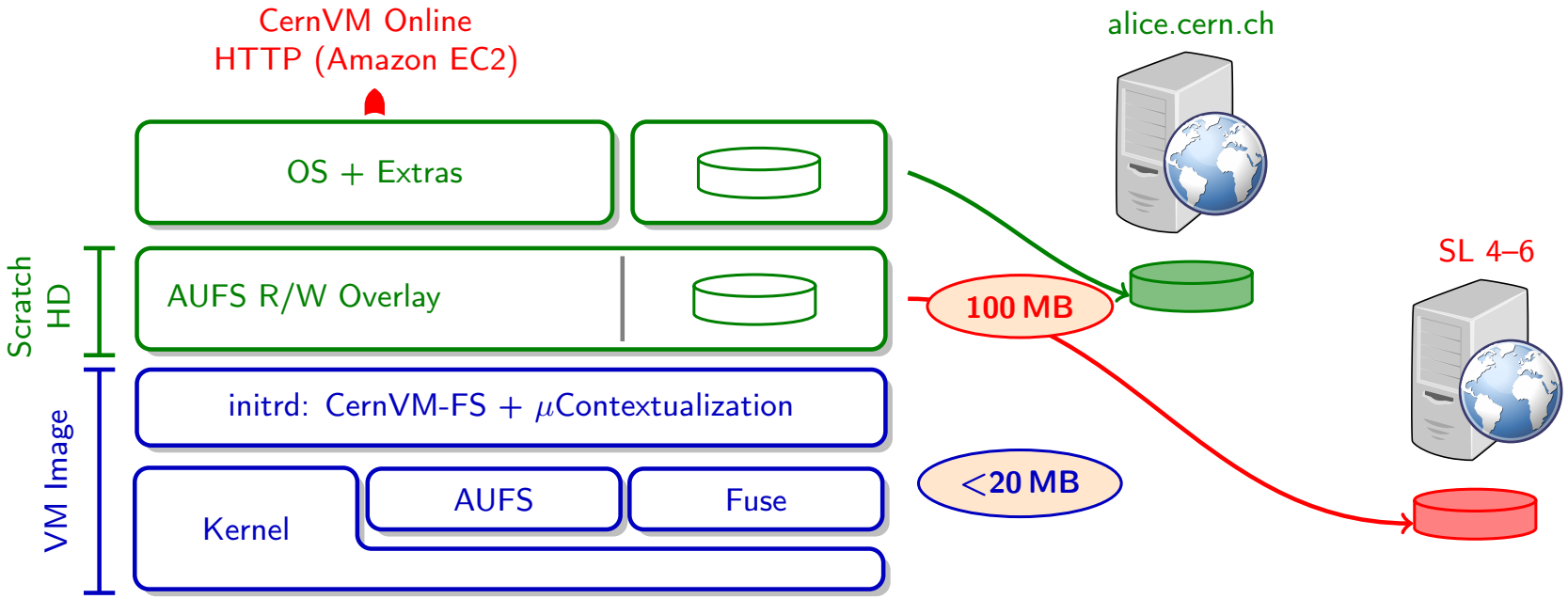
- Eliminates duplicates
- Archiving

File Catalog

- Directory structure, symlinks
- Content hashes of regular files
- Digitally signed
- Plain files

Reduces preservation of software environment to bit preservation

CernVM Components



Twofold system: μCernVM boot loader + OS delivered by CernVM-FS

- μCernVM has a modern Linux kernel,
  support for all relevant hypervisors and clouds
- The very same image can be *contextualized* to run
  Scientific Linux 4 32bit as well as the latest Scientific Linux 6 64bit
- $\approx$ 10 years with a single image

user-data.txt

```
[cernvm]
organisations=ALICE
repositories=alice,alice-ocdb,sft
shell=/bin/bash
config_url=http://cernvm.cern.ch/config
users=alice:alice:ion
edition=Desktop
keyboard=us
startXDM=on
auto_login=on

[ucernvm-begin]
cvmfs_tag=cernvm-system-3.1.1.4
[ucernvm-end]
```

Boot on CERN OpenStack

```
nova boot AliceVM -image "cvm3" -flavor m1.small \
  -key-name ssh-key -user-data user-data.txt
```

ALEPH software on CernVM / SL4

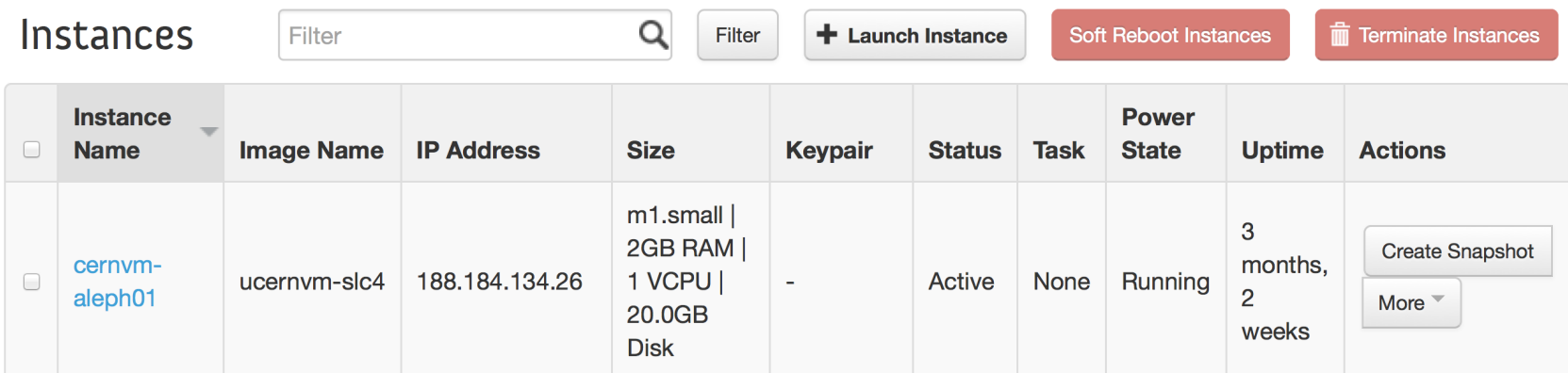
Work in progress

Exercise: resurrecting the ALEPH environment

- Can we use CernVM on current CERN OpenStack infrastructure to do ALEPH physics?
- Backport of CernVM-FS to Scientific Linux 4
- Template installation of Scientific Linux 4 for use with μCernVM

Exercise: resurrecting the ALEPH environment

- Can we use CernVM on current CERN OpenStack infrastructure to do ALEPH physics?
- Backport of CernVM-FS to Scientific Linux 4
- Template installation of Scientific Linux 4 for use with μCernVM



ALEPH software on CernVM / SL4

Work in progress

jakob — aleph@cernvm-aleph01:~/test/ALPHA — ssh — 66×18

```
pb-d-128-141-134-74:~ jakob$ ssh -X aleph@cernvm-aleph01
aleph@cernvm-aleph01's password:
[aleph@cernvm-aleph01 ~]$ source setaleph.sh
[aleph@cernvm-aleph01 ~]$ cd test/ALPHA/
[aleph@cernvm-aleph01 ALPHA]$ sh alpha.sh
     ************************************************
     **********         ALPHA RUN       **** 11.6 ***
     ************************************************
     ************************************************

Wed Mar 19 16:10:27 CET 2014

   ************************************************************
   ***    Compilation  and creation of the makefile 6lep.mk
   ************************************************************
gmake -f /home/aleph/test/ALPHA/6lep.mk
gmake: `6lep' is up to date.
```

CMS Open Data Pilot on CernVM / SL5

Work in progress

Purpose: Provide an easy-to-use virtual machine with CMS computing environment for CMS Open Data

Data:

- Frozen data set
- Remote data access
  Initially through XrootD, eventually DPHEP portal

Software:

- Frozen CMS software framework (CMSSW.4.2.8.patch7)
- *Complete* analysis environment required (compile + run)
- Requires Scientific Linux 5 compatible virtual machine

Virtual machine, user interface:

- Graphical environment
- Easy-to-install and easy-to-use

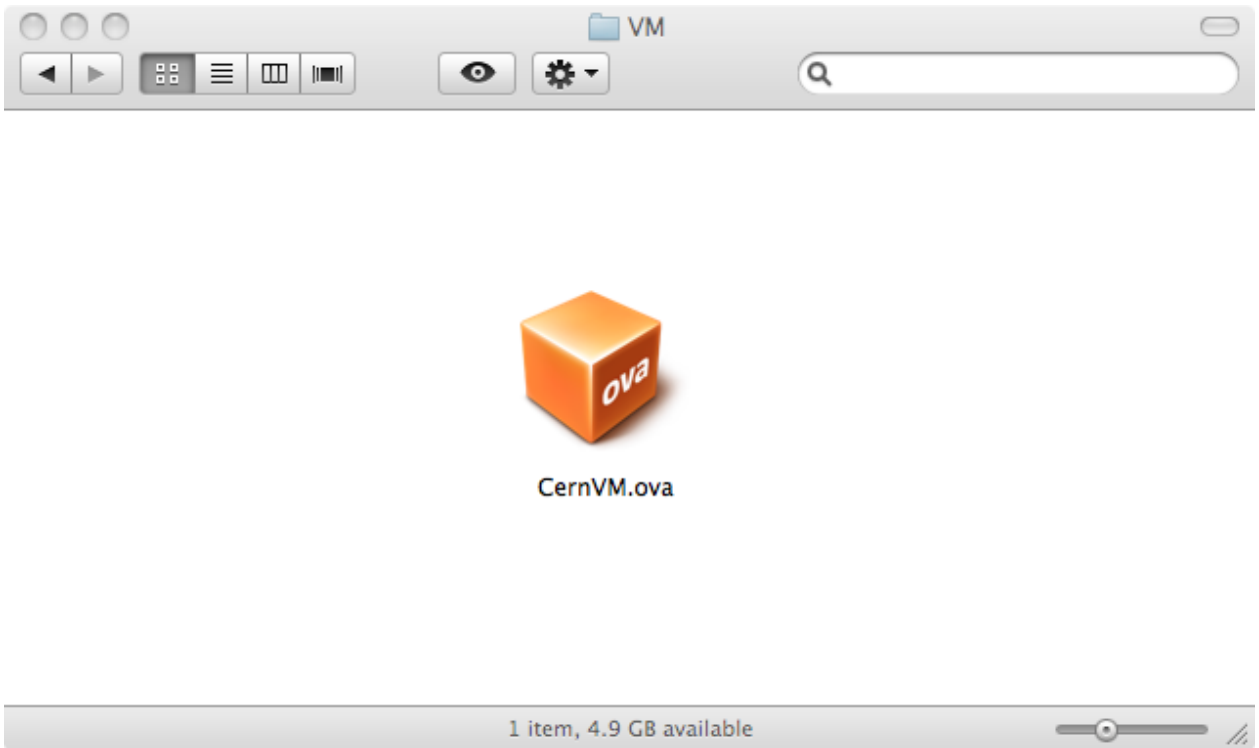
Deployment: as OVF/OVA bundle[1]

- Open specification for bundling VMs, stable since 2009
- OVA: tarball containing hard disk image and an XML specification

[1] Open Virtualization Format / Open Virtual Appliance, http://www.dmtf.org/standards/ovf

Deployment: as OVF/OVA bundle[1]

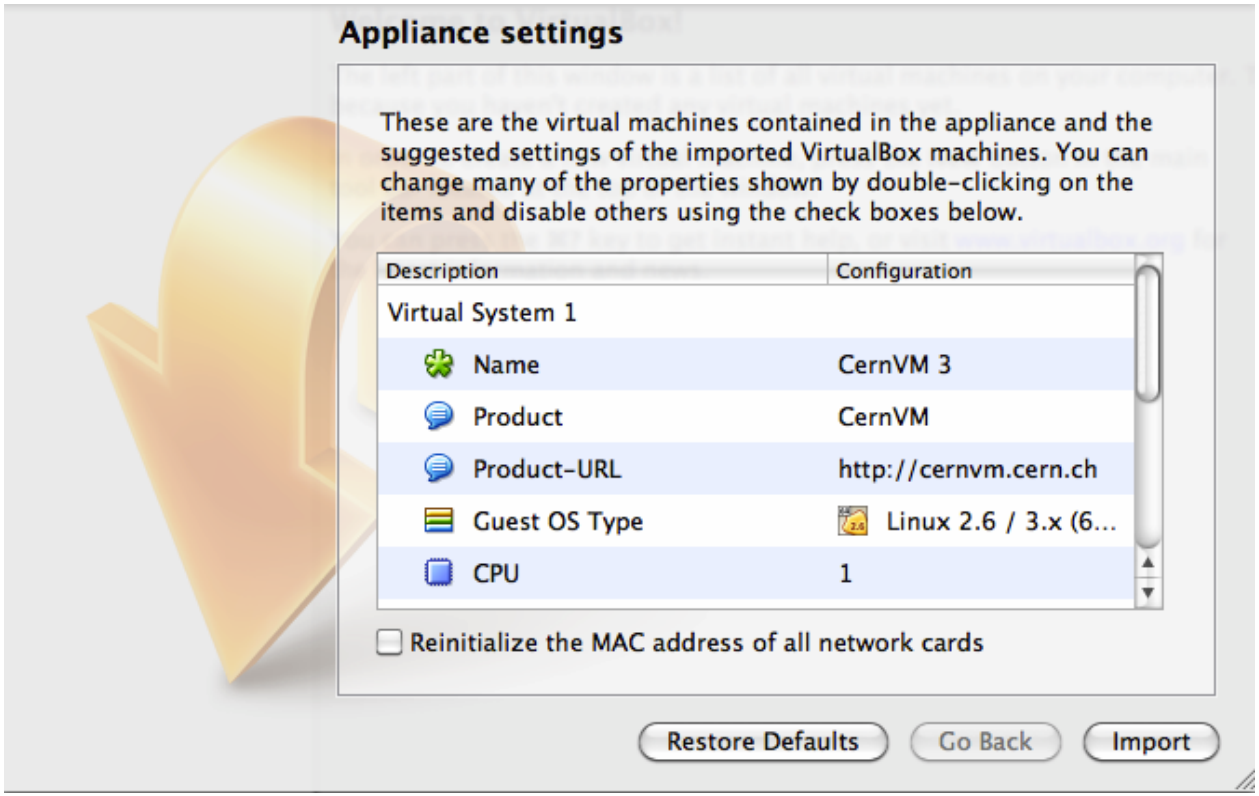
- Open specification for bundling VMs, stable since 2009
- OVA: tarball containing hard disk image and an XML specification



[1] Open Virtualization Format / Open Virtual Appliance, http://www.dmtf.org/standards/ovf

Deployment: as OVF/OVA bundle[1]

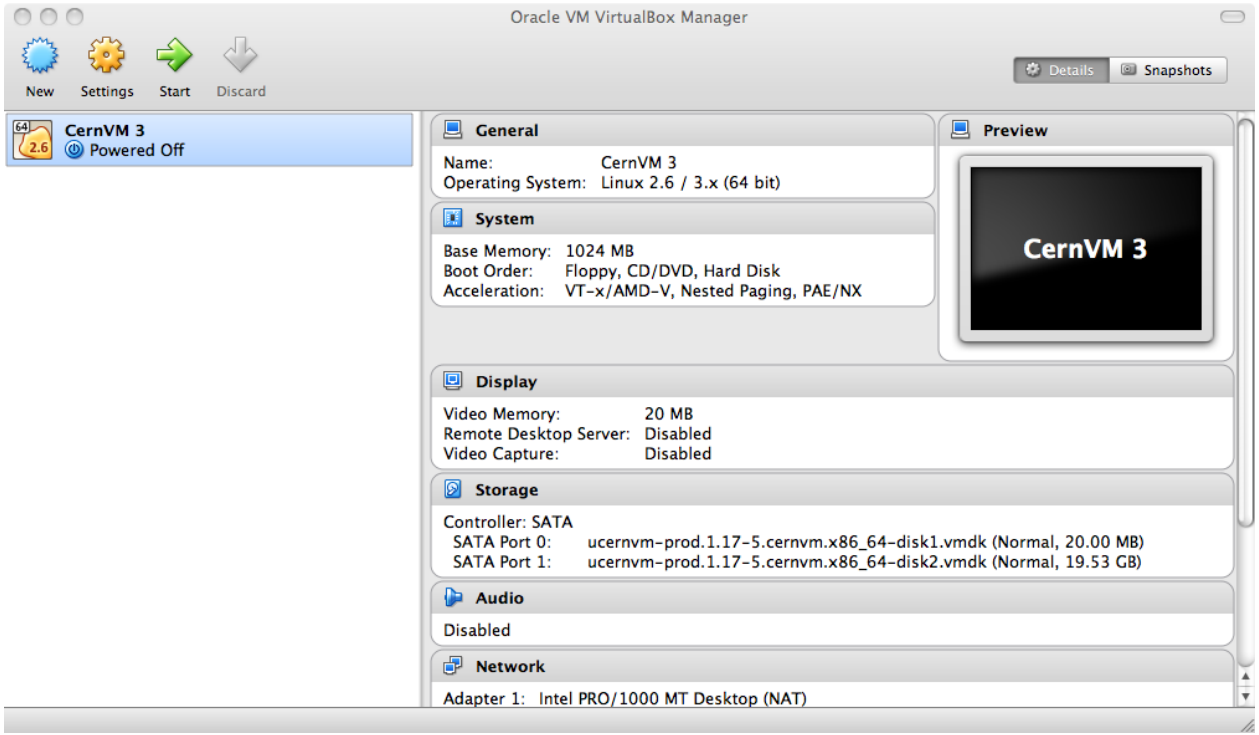
- Open specification for bundling VMs, stable since 2009
- OVA: tarball containing hard disk image and an XML specification



[1] Open Virtualization Format / Open Virtual Appliance, http://www.dmtf.org/standards/ovf

Deployment: as OVF/OVA bundle[1]

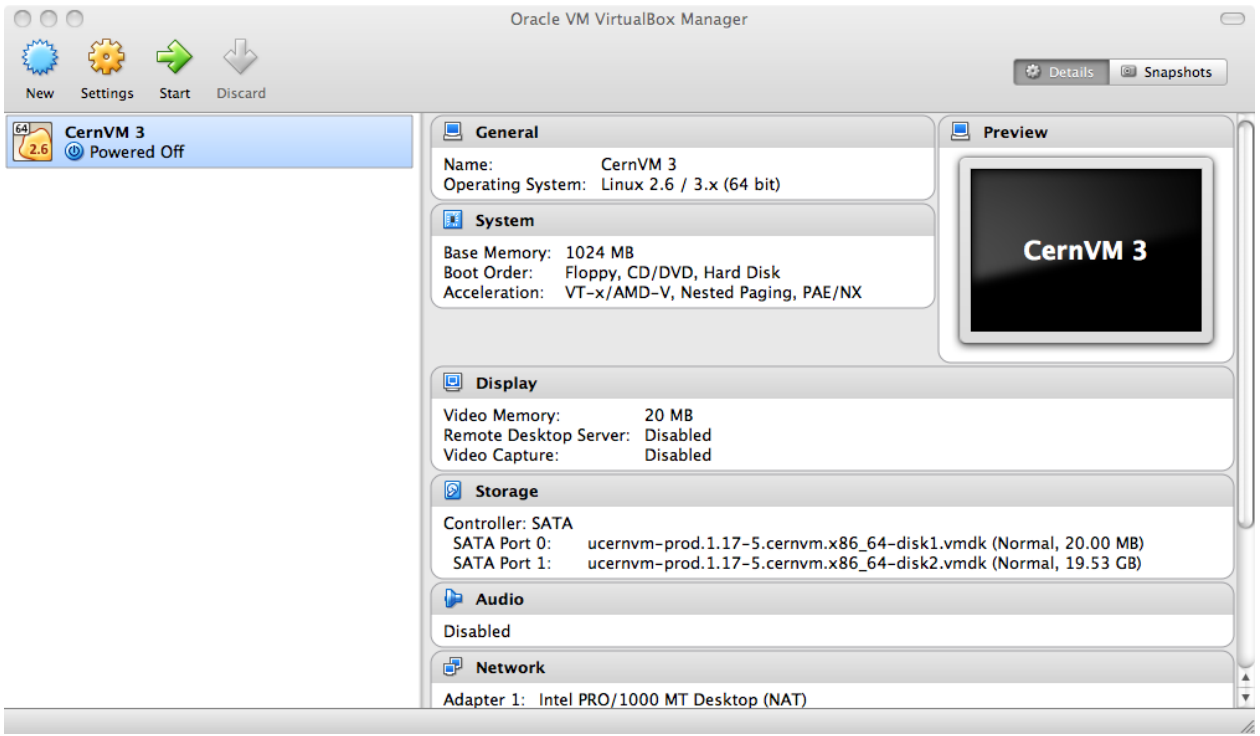
- Open specification for bundling VMs, stable since 2009
- OVA: tarball containing hard disk image and an XML specification



[1] Open Virtualization Format / Open Virtual Appliance, http://www.dmtf.org/standards/ovf

Deployment: as OVF/OVA bundle[1]

- Open specification for bundling VMs, stable since 2009
- OVA: tarball containing hard disk image and an XML specification

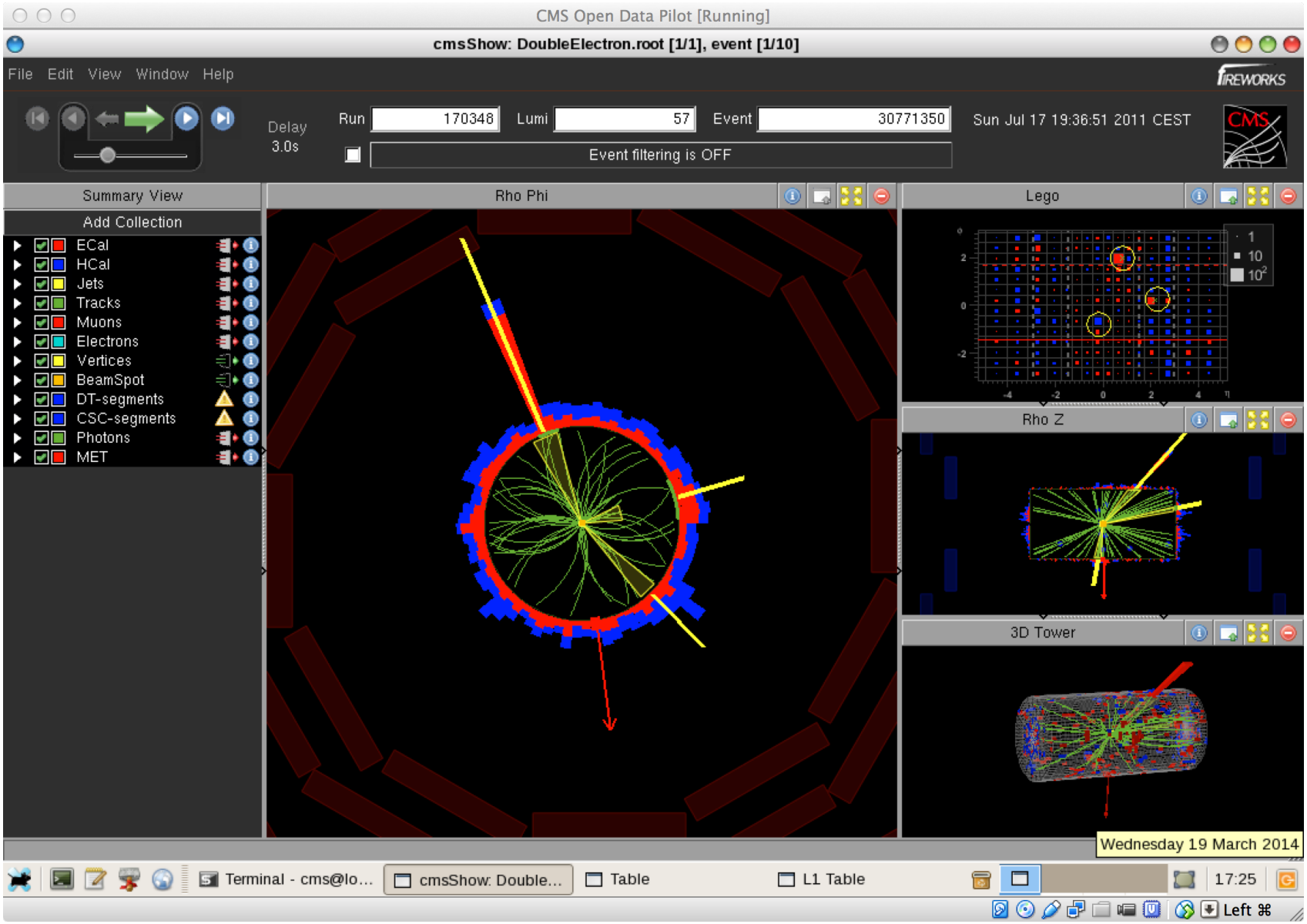


- OVA packaging fed back into CernVM baseline

[1] Open Virtualization Format / Open Virtual Appliance, http://www.dmtf.org/standards/ovf

CMS Open Data Pilot on CernVM / SL5

Work in progress

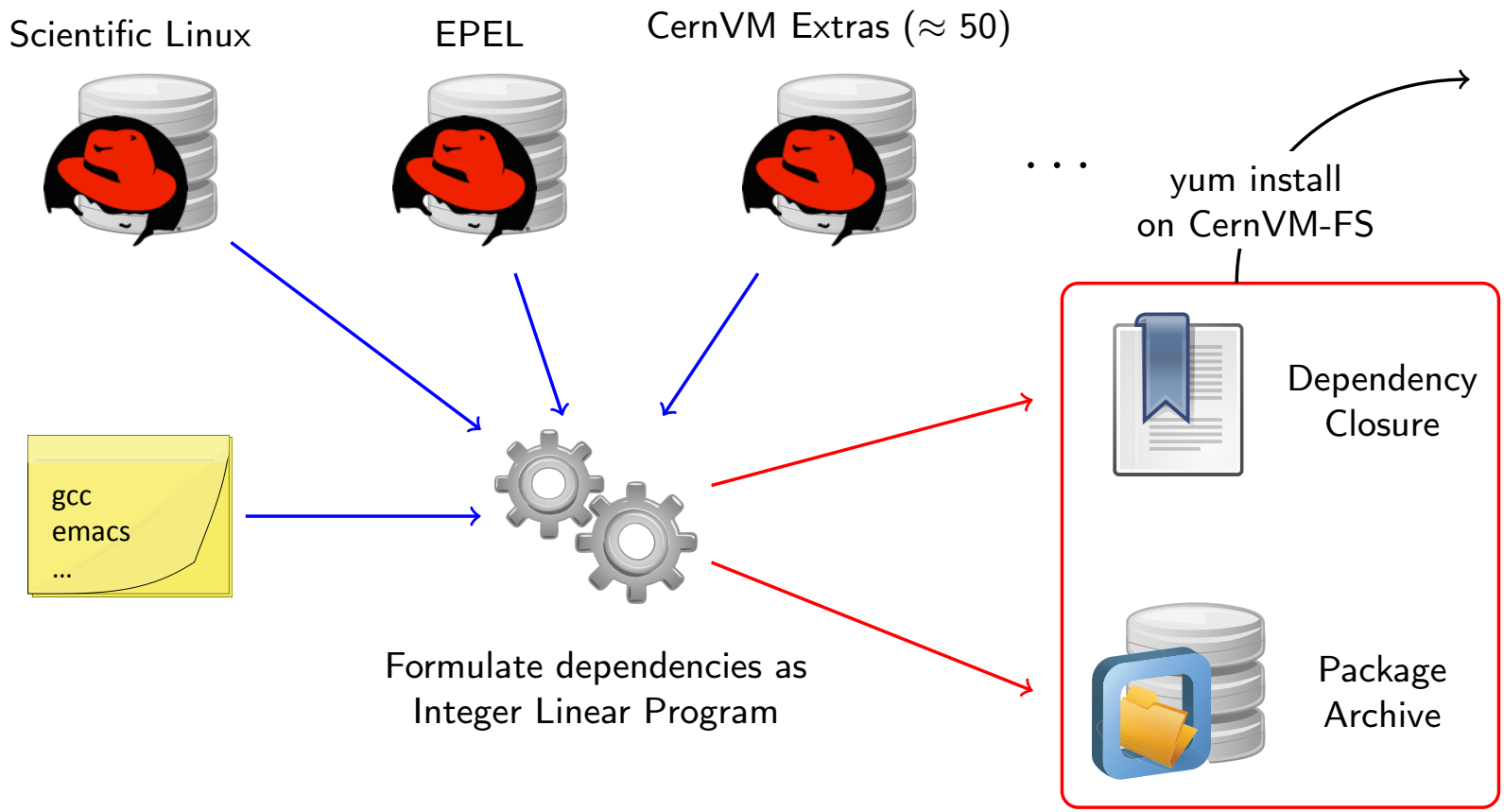


Backup Slides

Build Process: Scientific Linux on CernVM-FS

Maintenance of the repository **should not** become a Linux distributor's job
But: should be reproducible and well-documented

Idea: automatically generate a **fully versioned**, **closed** package list
from a "shopping list" of unversioned packages



Build Process: Package Dependency ILP

Normalized (Integer) Linear Program:

$$\text{Minimize } (c_1 \cdots c_n) \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad \text{subject to } \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \leq \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$$

Here: every available (package, version) is mapped to a $x_i \in \{0, 1\}$.
Cost vector: newer versions are cheaper than older versions.
(Obviously: less packages cheaper than more packages.)
Dependencies:
Package x_a requires x_b or x_c: $x_b + x_c - x_a \geq 0$.
Packages x_a and x_b conflict: $x_a + x_b \leq 1$.
(...)

Figures

≈17 000 available packages ($n = 17000$), 500 packages on "shopping list"
≈160 000 inequalities ($m = 160000$), solving time <10 s (`glpk`)
Meta RPM: ≈1 000 fully versioned packages, dependency closure

Idea: Mancinelli, Boender, di Cosmo, Vouillon, Durak (2006)

Hypervisor Support Status

Hypervisor / Cloud Controller	Status
VirtualBox	✓
VMware	✓
KVM	✓
Xen	✓
Microsoft HyperV	✓
Parallels	↯[3]
OpenStack	✓
OpenNebula	✓[2]
Amazon EC2	✓[1]
Google Compute Engine	✓

[1] Only tested with ephemeral storage, not with EBS backed instances
[2] Only amiconfig contextualization
[3] Unclear license of the guest additions