



a distro for industrial R&D and engineering

DebConf 2017
August 8th, 2017

Mehdi DOGGUY
mehdi@debian.org



a distro for scientific computing

DebConf 2017
August 8th, 2017

Mehdi DOGGUY
mehdi@debian.org



a Debian derivative

DebConf 2017
August 8th, 2017

Mehdi DOGGUY
mehdi@debian.org

Intro

Since 2003, EDF has been using a custom distro on scientific workstations, HPC clusters and servers for industrial R&D and engineering.

EDF decided to make its distro publicly available and turn it into an Open Source community-driven project.

Outline of this presentation:

- ▶ business context
- ▶ scientific and engineering IT needs
- ▶ our solution

About EDF

- ▶ World nuclear energy leader
- ▶ Europe leader in hydro power
- ▶ Industrial operator in Asia and United States

- ▶ Some key figures
 - ▶ 71.2 billion euros annual revenue
 - ▶ 37.6 M clients worldwide
 - ▶ 584.7 TWh produced annually
 - ▶ 136 GW production capacity
 - ▶ 73 nuclear reactors, 78% of production
 - ▶ 154 845 employees worldwide

- ▶ Large R&D and engineering divisions



<https://www.edf.fr/en/the-edf-group/who-we-are/edf-at-a-glance>

1 Scientific computing needs

Scientific computing needs at EDF

- ▶ R&D
 - ▶ Conception
 - ▶ Information technology
 - ▶ Renewable energies
 - ▶ Electrical networks
 - ▶ ...

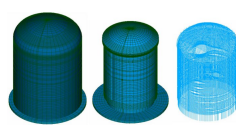
- ▶ Engineering

- ▶ Energy management
 - ▶ Reduce downtime on existing reactors
 - ▶ Planning consumption and production weeks in advance

Scientific computing at a glance

- ▶ Modeling

- ▶ Approximate reality with a model
- ▶ Often need for a modeler



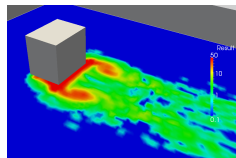
- ▶ Simulation

- ▶ Execution of a numerical code computing the behavior of the model system
- ▶ A whole area of software development
- ▶ Need for the fastest hardware to work on large arrays of floats



- ▶ Visualization

- ▶ Results exploration and analysis
- ▶ Need for the best graphics hardware and displays



Typical IT user needs

- ▶ Scientific workstation / laptop
 - ▶ Modeling and visualizing
 - ▶ Developing scientific simulation codes
 - ▶ Accessing the HPC clusters
- ▶ High performance clusters
 - ▶ Thousands of compute nodes connected using fast network
 - ▶ More or less specialized depending on applications
- ▶ Computing chains
 - ▶ Servers or small clusters
 - ▶ Regular execution of the same code
 - ▶ Coupling with other components

2 Scibian

Our solution: Scibian



Why Debian?

- ▶ **One OS for workstations, servers and clusters**
 - ▶ Full binary compatibility
- ▶ Appropriate release cycle
 - ▶ One major version approx. every 3 years
 - ▶ Frequent updates (for critical bugs and security fixes)
- ▶ Largest scientific software offering
 - ▶ Only Ubuntu matches, by following Debian repositories
- ▶ Designed for customization
 - ▶ Custom repositories, easy deployment
 - ▶ Modular and hookable installer
- ▶ Community openness
 - ▶ Easy to get interesting changes into the distribution
- ▶ Easy to integrate applications
 - ▶ Cool packaging helpers
 - ▶ Abundant documentation

Why Debian?

- ▶ One OS for workstations, servers and clusters
 - ▶ Full binary compatibility
- ▶ **Appropriate release cycle**
 - ▶ One major version approx. every 3 years
 - ▶ Frequent updates (for critical bugs and security fixes)
- ▶ Largest scientific software offering
 - ▶ Only Ubuntu matches, by following Debian repositories
- ▶ Designed for customization
 - ▶ Custom repositories, easy deployment
 - ▶ Modular and hookable installer
- ▶ Community openness
 - ▶ Easy to get interesting changes into the distribution
- ▶ Easy to integrate applications
 - ▶ Cool packaging helpers
 - ▶ Abundant documentation

Why Debian?

- ▶ One OS for workstations, servers and clusters
 - ▶ Full binary compatibility
- ▶ Appropriate release cycle
 - ▶ One major version approx. every 3 years
 - ▶ Frequent updates (for critical bugs and security fixes)
- ▶ **Largest scientific software offering**
 - ▶ Only Ubuntu matches, by following Debian repositories
- ▶ Designed for customization
 - ▶ Custom repositories, easy deployment
 - ▶ Modular and hookable installer
- ▶ Community openness
 - ▶ Easy to get interesting changes into the distribution
- ▶ Easy to integrate applications
 - ▶ Cool packaging helpers
 - ▶ Abundant documentation

Why Debian?

- ▶ One OS for workstations, servers and clusters
 - ▶ Full binary compatibility
- ▶ Appropriate release cycle
 - ▶ One major version approx. every 3 years
 - ▶ Frequent updates (for critical bugs and security fixes)
- ▶ Largest scientific software offering
 - ▶ Only Ubuntu matches, by following Debian repositories
- ▶ **Designed for customization**
 - ▶ Custom repositories, easy deployment
 - ▶ Modular and hookable installer
- ▶ Community openness
 - ▶ Easy to get interesting changes into the distribution
- ▶ Easy to integrate applications
 - ▶ Cool packaging helpers
 - ▶ Abundant documentation

Why Debian?

- ▶ One OS for workstations, servers and clusters
 - ▶ Full binary compatibility
- ▶ Appropriate release cycle
 - ▶ One major version approx. every 3 years
 - ▶ Frequent updates (for critical bugs and security fixes)
- ▶ Largest scientific software offering
 - ▶ Only Ubuntu matches, by following Debian repositories
- ▶ Designed for customization
 - ▶ Custom repositories, easy deployment
 - ▶ Modular and hookable installer
- ▶ **Community openness**
 - ▶ Easy to get interesting changes into the distribution
- ▶ Easy to integrate applications
 - ▶ Cool packaging helpers
 - ▶ Abundant documentation

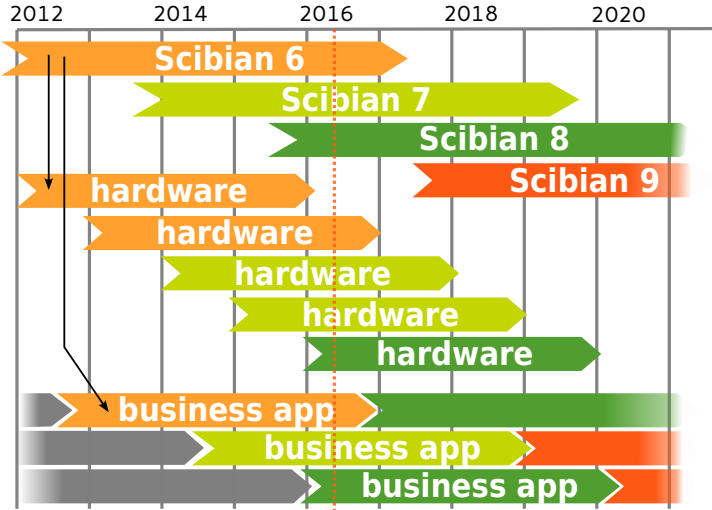
Why Debian?

- ▶ One OS for workstations, servers and clusters
 - ▶ Full binary compatibility
- ▶ Appropriate release cycle
 - ▶ One major version approx. every 3 years
 - ▶ Frequent updates (for critical bugs and security fixes)
- ▶ Largest scientific software offering
 - ▶ Only Ubuntu matches, by following Debian repositories
- ▶ Designed for customization
 - ▶ Custom repositories, easy deployment
 - ▶ Modular and hookable installer
- ▶ Community openness
 - ▶ Easy to get interesting changes into the distribution
- ▶ **Easy to integrate applications**
 - ▶ Cool packaging helpers
 - ▶ Abundant documentation

Why a derivative?

- ▶ 6+ year support
- ▶ Custom security support
- ▶ Workstation and HPC hardware support backports (mainly for Infiniband, OmniPath, nVidia GPUs and newer Intel micro-architectures)
- ▶ Support of some libs removed from Debian
- ▶ Upgrade to newer major versions of some software

Life cycle



Business Applications

Packaging of business applications follows a few rules:

- ▶ No maintainer scripts and no services
- ▶ No files outside `/opt/$name-$version`, except:
 - ▶ a script in `/usr/bin`
 - ▶ manpages
 - ▶ copyright file
 - ▶ icons and a desktop file
- ▶ Pre-Depends must be empty
- ▶ No alternative dependencies and no Provides
- ▶ All files are owned by root and writable only by root
- ▶ No setuid/setgid binaries (or other means to escalate privileges)
- ▶ ...

Each application is made available on each published version of Scibian.

3 Scibian for HPC Clusters

Deploying an HPC cluster

HPC clusters can be tricky to deploy because:

- ▶ Every manufacturer has its own solution
- ▶ No binary compatibility between clusters
- ▶ Conflicting software stacks between clusters (e.g. Infiniband vs OmniPath)
- ▶ Large number of compute nodes (a few thousands) to deploy
- ▶ Not easy to get Debian support
- ▶ Full (physical and software) installation cannot last more than 3 months
- ▶ High performance is the ultimate goal

Scibian HPC Installation Guide

Standardize Scibian cluster deployments by:

- ▶ Defining a general architecture for an HPC Cluster
- ▶ Describing how to install it using tools packaged in Scibian
- ▶ Showing how to configure it using our Puppet modules

Document available at:

<https://edf-hpc.github.io/scibian-hpc-install-guide/>

Puppet HPC

The main goal of Puppet-HPC is to provide a common generic configuration management system that can be used effortlessly across multiple HPC clusters and organizations.

The Puppet-HPC software stack notably provides:

- ▶ Many generic Puppet modules (>80) for all technical components required on a HPC cluster
- ▶ Defined data model for representing the description of an HPC cluster based on Hieradata
- ▶ Tools to easily deploy and manage the configuration with high-scalability requirements

It is heavily tested on Debian and used in production on thousands of Debian machines.

Goals behind Puppet HPC

- ▶ The code base can be re-used and the development effort is shared.
- ▶ The same code is run on many different environments, it is therefore more tested and more reliable.
- ▶ The code can be easily tested on a small testing environment even if the data is different from the production environment.

More details available at:

- ▶ https://edf-hpc.github.io/puppet-hpc/puppet_hpc_reference-0.1.html
- ▶ <https://github.com/edf-hpc/puppet-hpc>

4 Why should I use Scibian?



Why should I use Scibian?

- ▶ Still use Debian (We don't even change the kernel!)
- ▶ More relaxed update rules
- ▶ A set of integrated tools for deployment and scientific computing infrastructures
- ▶ Meet and work with other industrial users
- ▶ Well tested solution in a high demanding production environment
- ▶ Integrate specific proprietary software (that is not possible to integrate in Debian)
- ▶ ...

5 **What's next?**

What's next?

- ▶ Publish our packages (tools, open source business applications, puppet hpc. . .)
- ▶ Update our website and add up-to-date installation guides
- ▶ Convince Infiniband and OmniPath manufacturers to publish APT-gettable repositories, or let us do it
- ▶ Switch to collaborative development of Scibian
- ▶ Share, build and maintain specific tools
- ▶ Create a community around the use of Linux for business R&D
- ▶ ...



Thanks!

Useful links:

- ▶ <https://twitter.com/ScibianLinux>
- ▶ <https://github.com/scibian/>
- ▶ <https://github.com/edf-hpc>
- ▶ <irc://irc.debian.org/#debian-hpc>
- ▶ <https://lists.debian.org/debian-hpc/>

Contact us: contact@scibian.org



We're hiring!

Useful links:

- ▶ <https://twitter.com/ScibianLinux>
- ▶ <https://github.com/scibian/>
- ▶ <https://github.com/edf-hpc>
- ▶ <irc://irc.debian.org/#debian-hpc>
- ▶ <https://lists.debian.org/debian-hpc/>

Contact us: contact@scibian.org