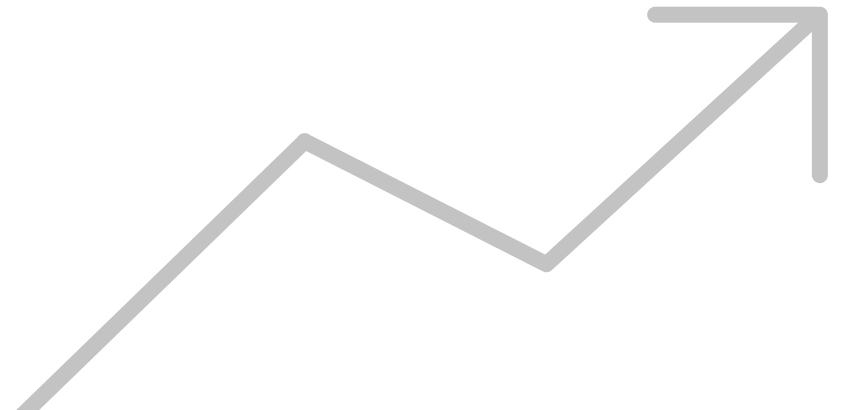


Managing a Central Server for R-Programming: Practical Insights and Challenges

Bernhard Fischer (Destatis)

uRos2024 Athens



Agenda

- » History of R use at Destatis
- » Our infrastructure
 - » Overview and sizing
 - » Challenges
 - » Automated deployment process
- » User management and with custom shiny app

History of R use at Destatis

- » R use as grassroots movement, self installation of IDE by interested users
- » First meetings of community-organised useR-Group → push for easier availability
- » Local RStudio-IDE as official software package
- » Leadership sees growing usage and pushes for centralisation of resources
- » First central R-Server established end of 2018 using Rstudio Server Professional (now Posit Workbench)
- » Number of users climbs from 40 to 400 over the the following 5 years

R infrastructure at Destatis

Sizing old server

- » Single Server (bare metal) 72 CPU Cores + 768GB RAM
- » Installed software: Posit Workbench Advanced
- » 8 TB NFS storage
- » Local Git installation
- » OS: SLES 12

R infrastructure at Destatis

Challenges

- » Resources
 - » 400 users means ~40 users active at all (day)times
 - » One user starting a grid search can use up all compute power
 - » Informal rules for cpu-usage necessary
- » Maintenance and bugs
 - » Single users are able to make the system unusable for everyone
 - » System maintenance requires downtime

R infrastructure at Destatis

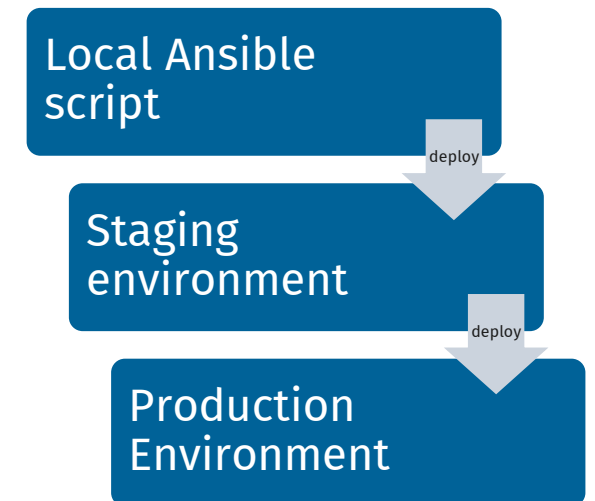
Sizing new server

- » Load Balancing Cluster (Bare Metal) with 10 x (2 x NVIDIA RTX-A6000 P 48GB VRAM + 44 CPU Cores + 2TB RAM)
- » Installed Software: Posit Workbench Advanced
- » 20 TB NFS Storage
- » Gitlab Server (Community Edition) for CI/CD
- » Posit Package Manager for Package Management (Python (PyPI) & R (CRAN, Bioconductor))
- » OS: Red Hat Enterprise Linux 8.10

R infrastructure at Destatis

Automated infrastructure deployment

- » Large cluster creates necessity for automated deployment of configuration and updates
- » We use Ansible as deployment tool
- » Git as central configuration repository for Ansible
- » New configuration committed to main branch is automatically deployed to staging environment via Jenkins
- » Deployment to Production environment is manually started by service provider



R infrastructure at Destatis

Automated infrastructure deployment

- » Automatic deployment allows easy testing of new configurations
- » Staging-environment allows for thorough testing before release of new Package, R and Workbench-Versions
- » Load-Balancing Cluster can be patched partially so no downtime for users

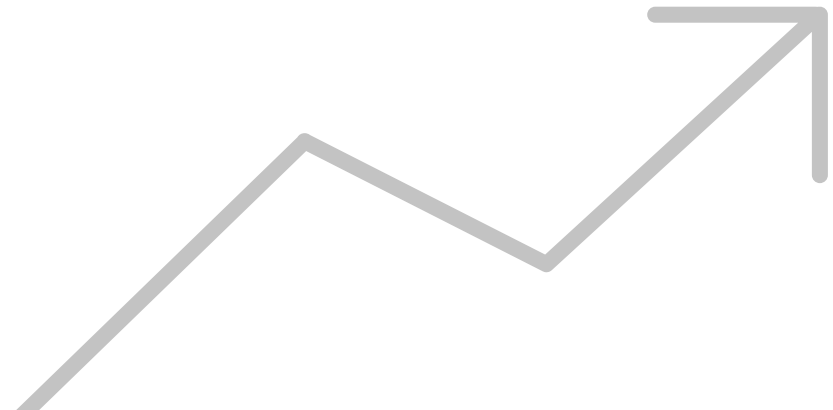
User management and monitoring

Shiny app for user management

- » Shiny app reads in linux users and uses additional sqlite database to keep track
- » Shiny app writes changes to csv-file
- » Shellscript is started regularly via cron-job
- » csv-file is sanitized and changes executed

The screenshot displays the 'R-Server-Nutzerverwaltung' (R-Server User Management) web interface. At the top, there are four tabs: 'R-Server-Nutzerverwaltung', 'Nutzer verwalten', 'Nutzerliste anzeigen', and 'Platzhalter'. The 'Nutzer verwalten' tab is currently selected. On the left side, there is a vertical menu with the following options: 'Nutzer hinzufügen' (highlighted in blue), 'E-Mail ändern', 'Passwort zurücksetzen', 'Nutzer deaktivieren', 'Nutzer aktivieren', 'Nutzer löschen', 'Nutzer einer Gruppe hinzufügen', 'Nutzer aus Gruppe löschen', and 'Datenbanken abgleichen'. The main content area on the right is titled 'Nutzer hinzufügen'. It contains several input fields: 'Familienname', 'Vorname:', 'Benutzername:', and 'E-Mail Adresse:'. Below these fields, there is a section for 'Gruppen:' with two checkboxes: 'Poweruser' and 'Admin'. At the bottom of this section is a button labeled 'Nutzer anlegen'.

Questions?



Contact

Statistisches Bundesamt
65180 Wiesbaden
Germany

www.destatis.de

www.destatis.de/kontakt

Bernhard Fischer
bernhard.fischer@destatis.de
Phone +49 611 75-3738

