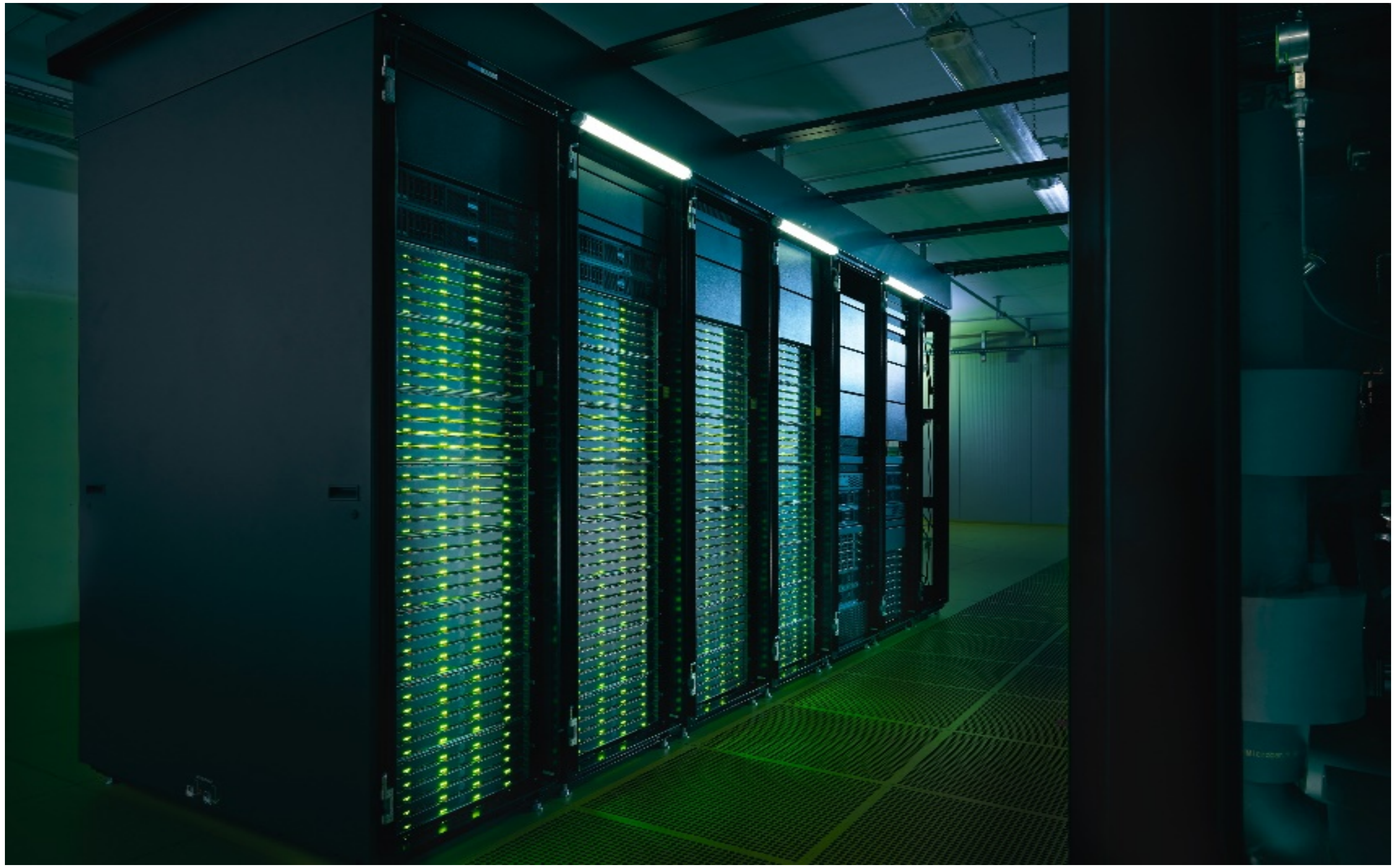
SC20

Everywhere we are | more than hpc.

Extending an Open-Source Federated Identity Management System for Enhanced HPC Security

Nov. 17th 2020



Dr. Jennifer Buchmüller

Head of the Department Scientific Computing and Simulation

HPC Core Facility Leader

Contact:

jennifer.buchmueller@kit.edu

https://twitter.com/SCC_KIT

Character of the attack

Character of attack

- Many European HPC centers have been compromised between November 2019 and March 2020 (Tier-1, Tier-2, Tier-3)
- On the majority of systems, a backdoor was installed that allowed unprivileged uses to gain a root shell.
- Motive of the attackers still unknown.
- Partners from around the world observed similar attacks and break-in attempts as well.

Possible attack vectors

- HPC: Successful SSH login using stolen credentials.
- User workstations, laptops and third-party servers: Unknown.

SSH attack vector in detail

- Attackers extracted credentials in very sophisticated ways (e.g. from key stores, manipulated SSH binaries etc.)
- SSH private keys not protected with a passphrase were used right away. Some indication that passphrases might have been cracked offline.
- Attackers analyzed sshd log files, ssh configs, bash history and other data to find more targets.

SSH attack vector in detail

- Two actually used exploit binaries have been found on two different HPC systems.

- Both exploits follow the same structure. The actual exploits seemed to be derivatives of publicly available proof-of-concept (PoC) codes for CVE-2017-889 and CVE-2018-9568.

- Attackers seem to have access to a continuously updated collection of exploits.

Short-term security mitigations

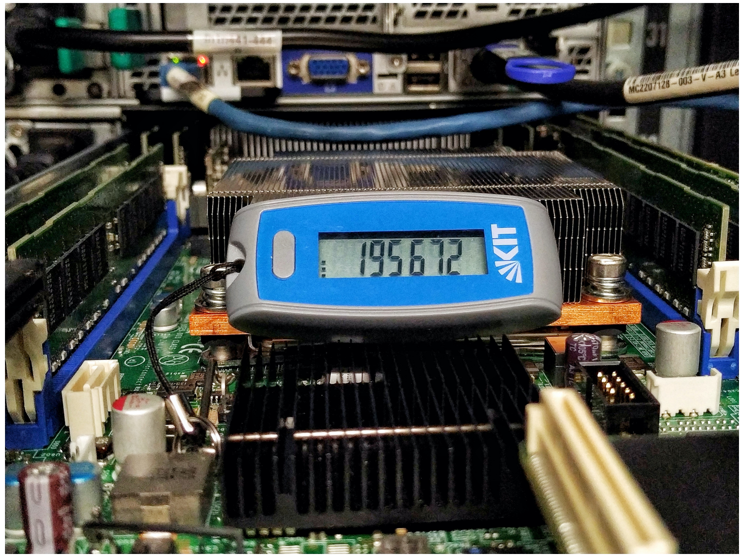
- Full Re-installation of the respective system

- Reset all passwords (Users + Admins)
- Enforced log-in only via service password

- Deactivated and collected SSH keys
- Deactivated SSH agent forwarding

- Secured mount points (nodev, nosuid etc.)

- Continuous security checks and monitoring (SUID binary scanner, auditd, Yara etc.)

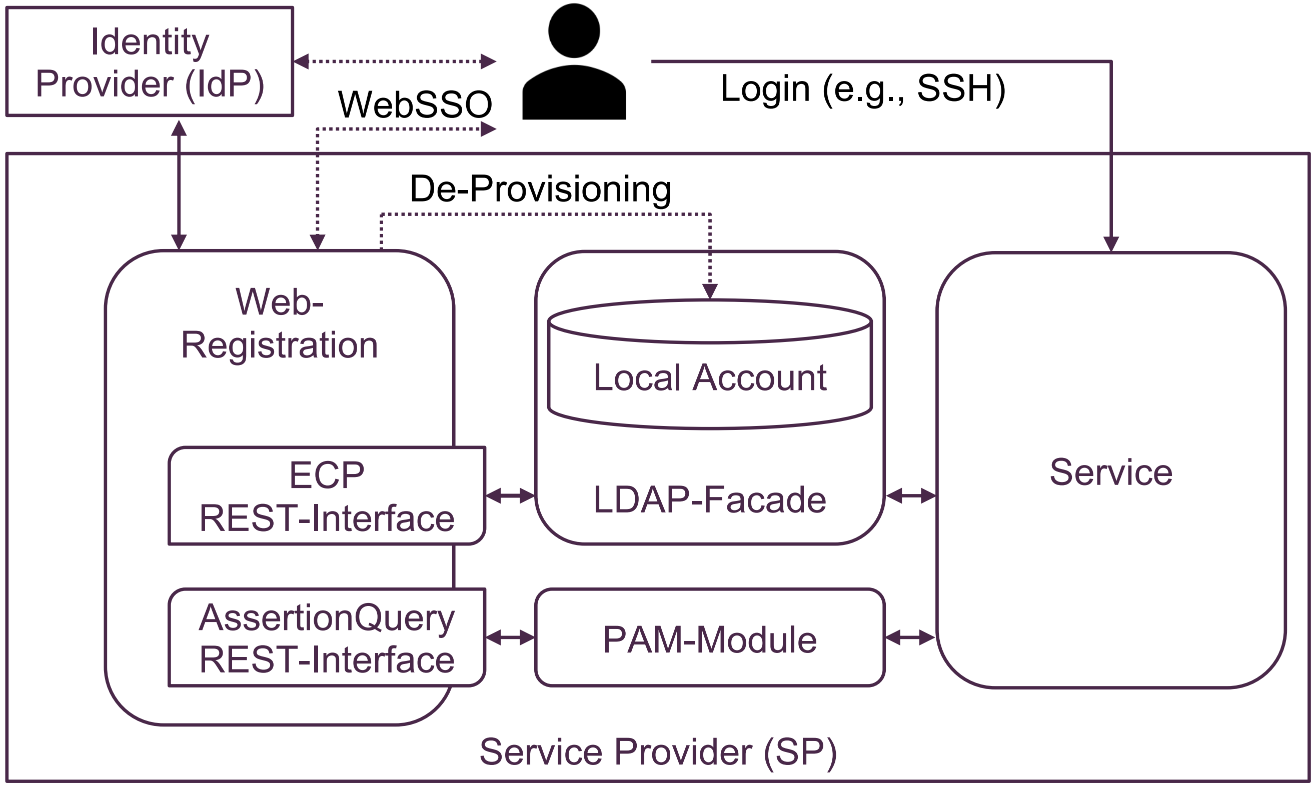


Introduction of bwIDM and reg-app

bwIDM

Federation of IT services within the German federal state of Baden-Wuerttemberg (“bw”)

- … acts as a single point of authentication for all services
- … acts as a gateway between the university’s existing Identity Provider (IdP) servers and the individual services.



J. Köhler, S. Labitzke, M. Simon, T. Dussa, M. Nussbaumer, and H. Hartenstein. bwIDM - federated access to it-based services at the universities of the state of Baden-wüttemberg. *Praxis der Informationsverarbeitung und Kommunikation*, 37(1):15–21, 2014.

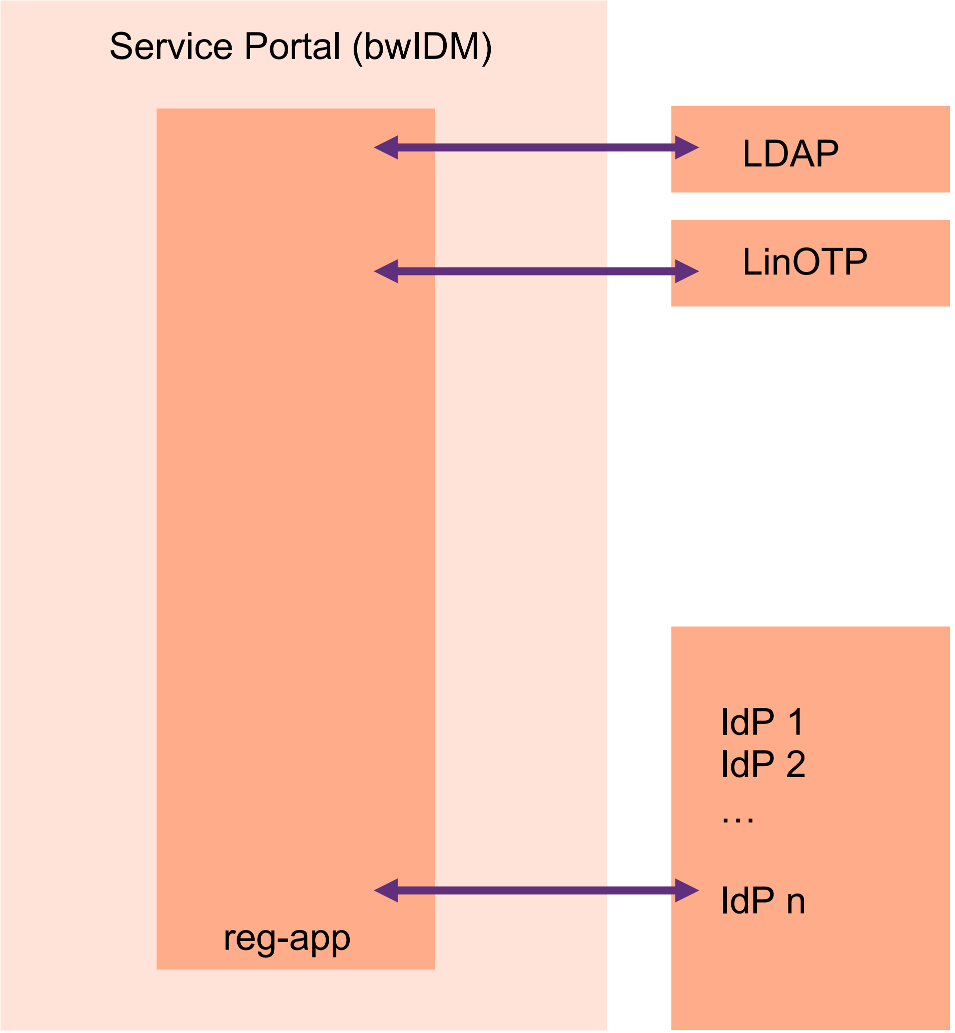
reg-app

… is an open-source Java application

… provides an LDAP facade to the services

… provides SAML/Shibboleth and OpenID Connect enpoints

… manages individual service passwords for every service or forwards password requests to other IdPs

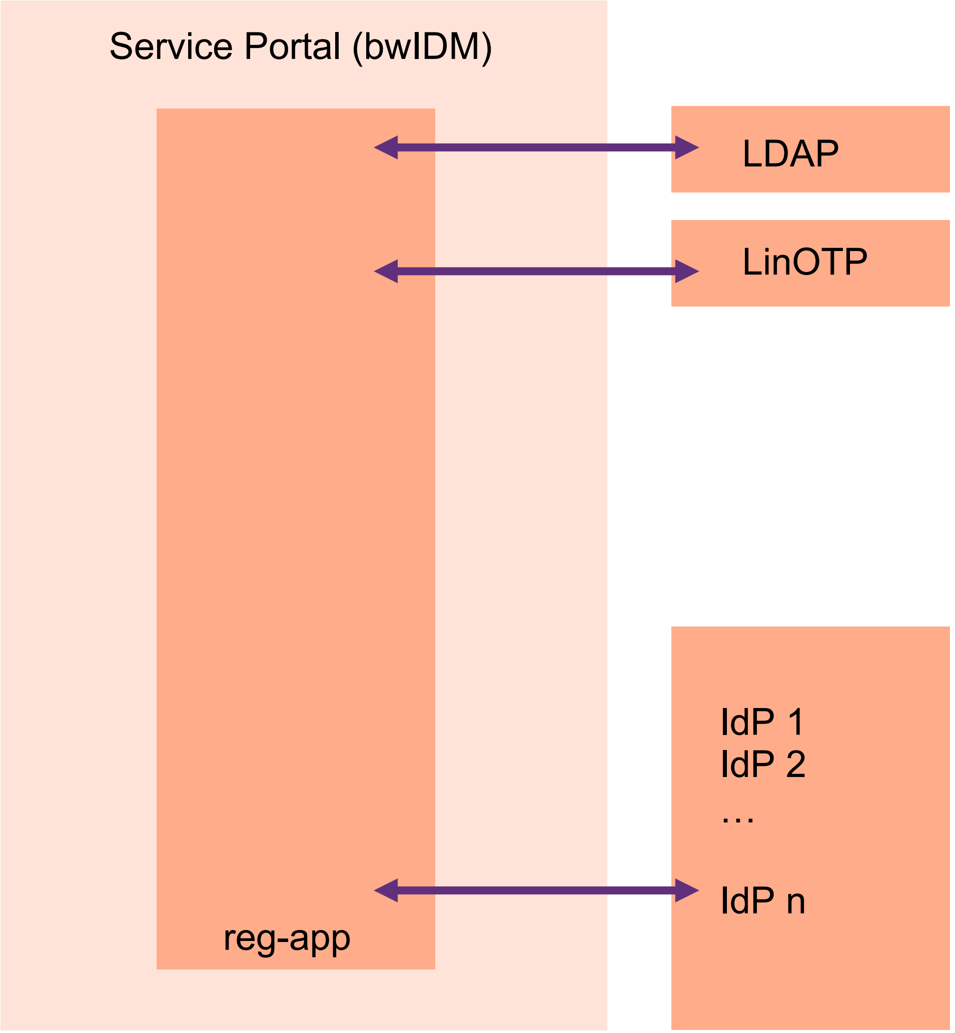


reg-app

… includes a user- and administrator-facing web frontend for user registration and user/group/service management

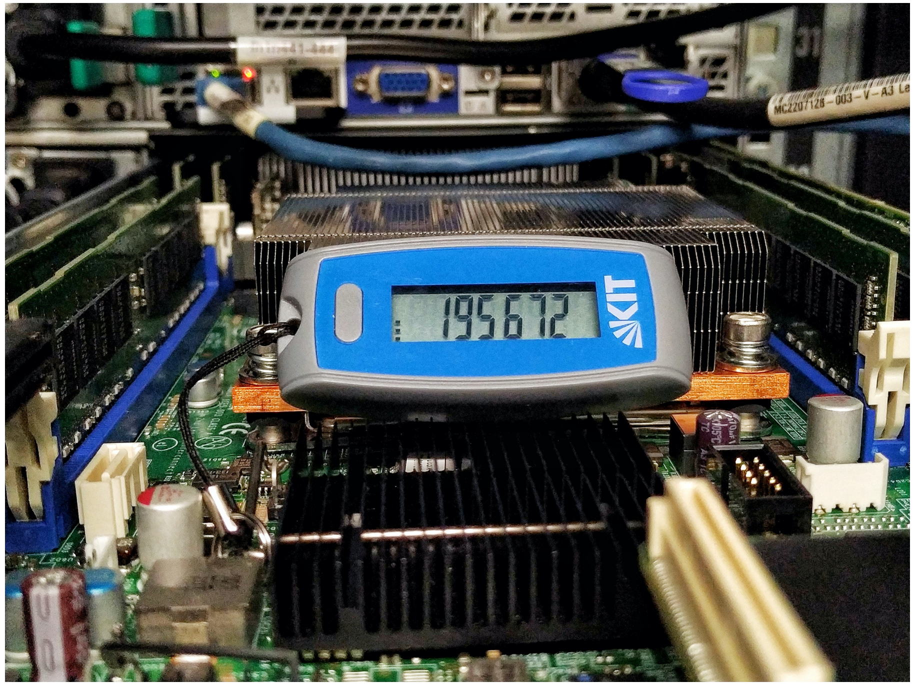
… provides a custom HTTP REST interface for everything not possible/feasible with the other endpoints

… is covered by 13 load-balanced servers



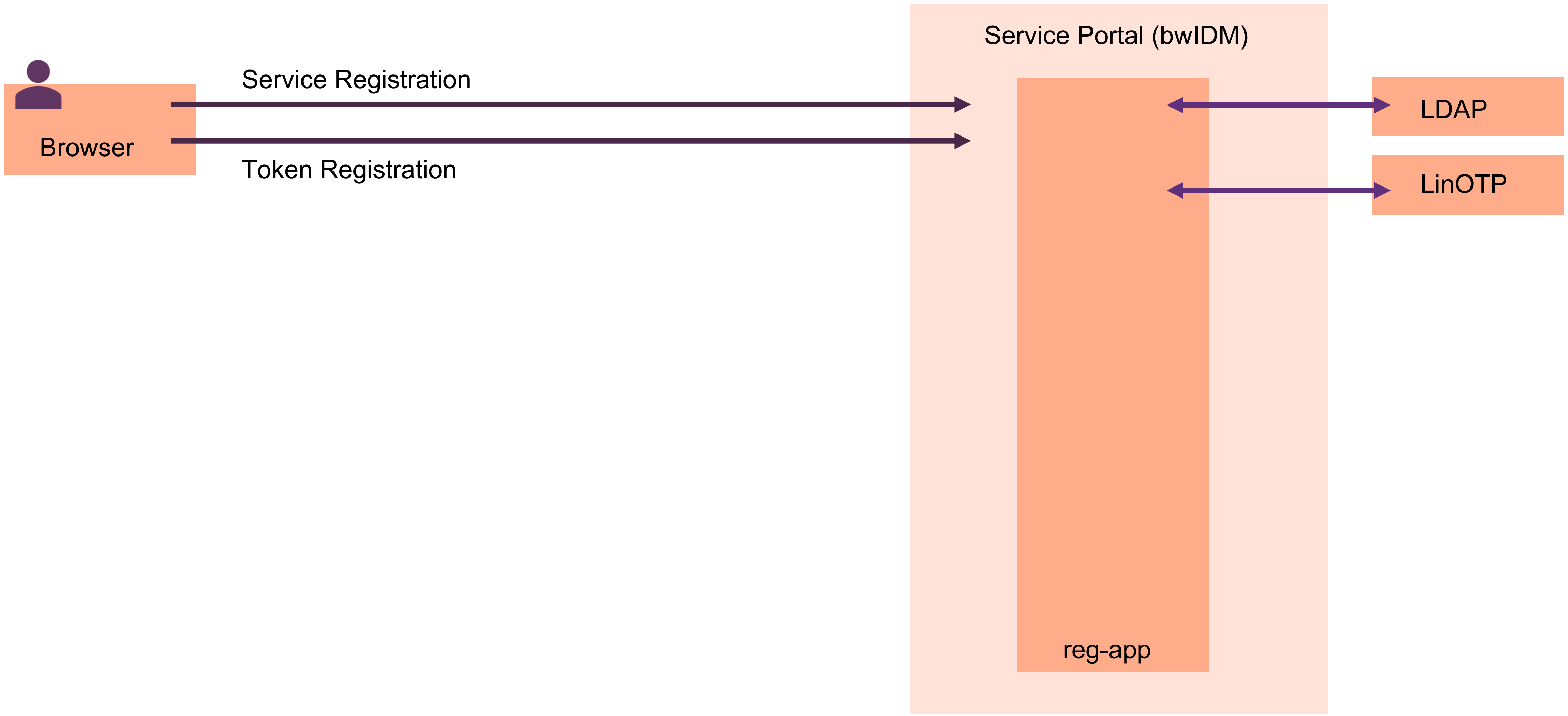
Introducing the Second Factor

- Hardware- or Software-Tokens
- Time-based one-time passwords (OATH/TOTP, Yubico OTP), Backup-TAN lists.
- Centralized and easy self-managment system(bwIDM).
- SSH-keys are no longer taken from HPC file-systems.

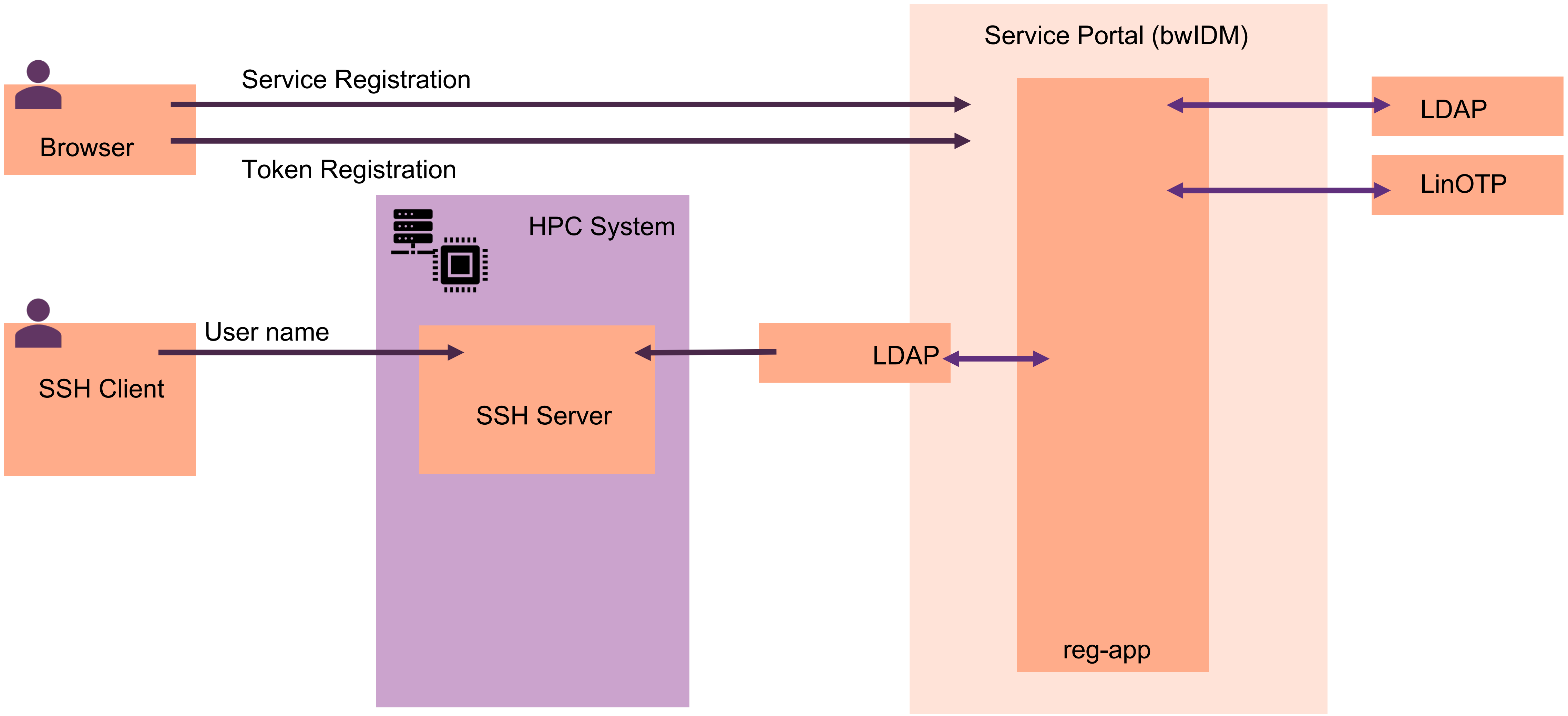


Where reg-app and HPC come together

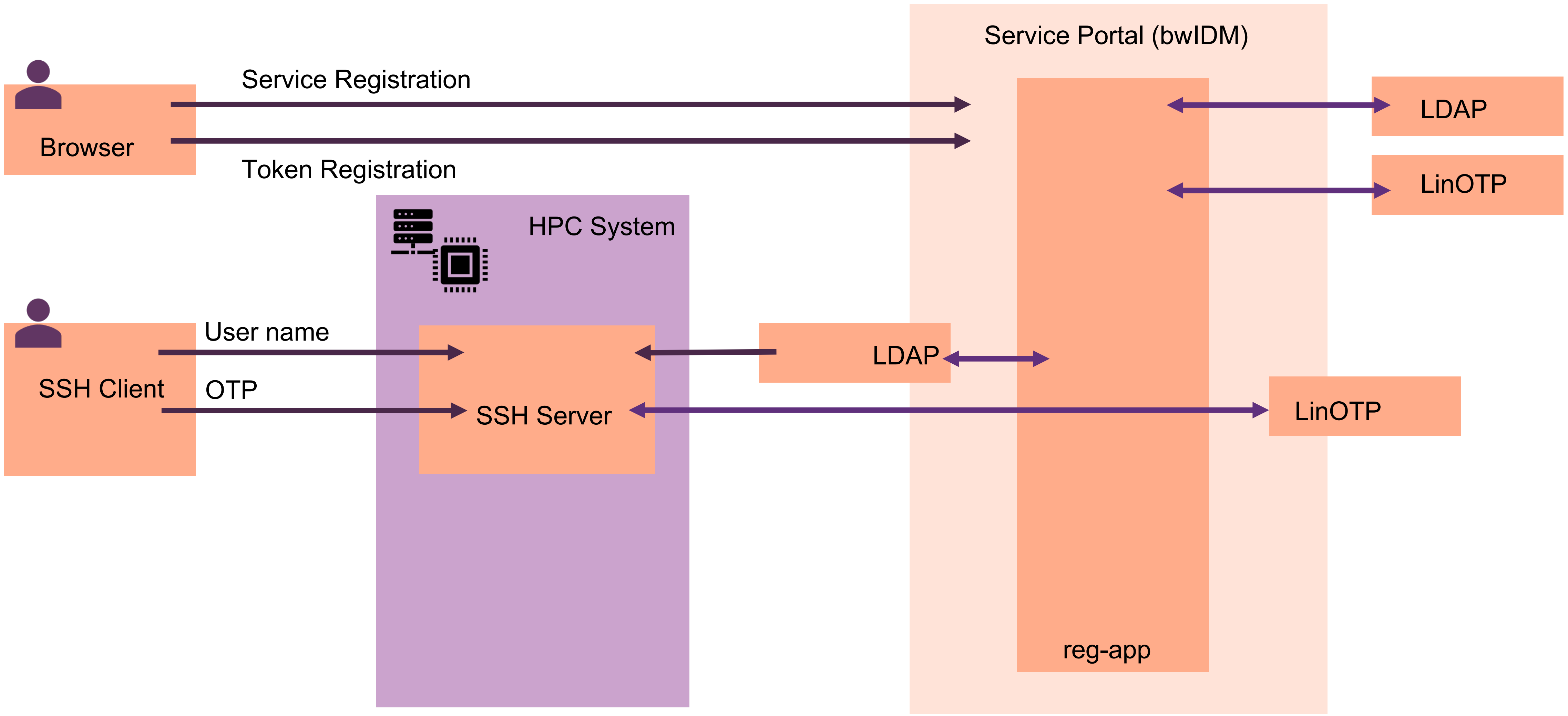
Where reg-app and HPC come together



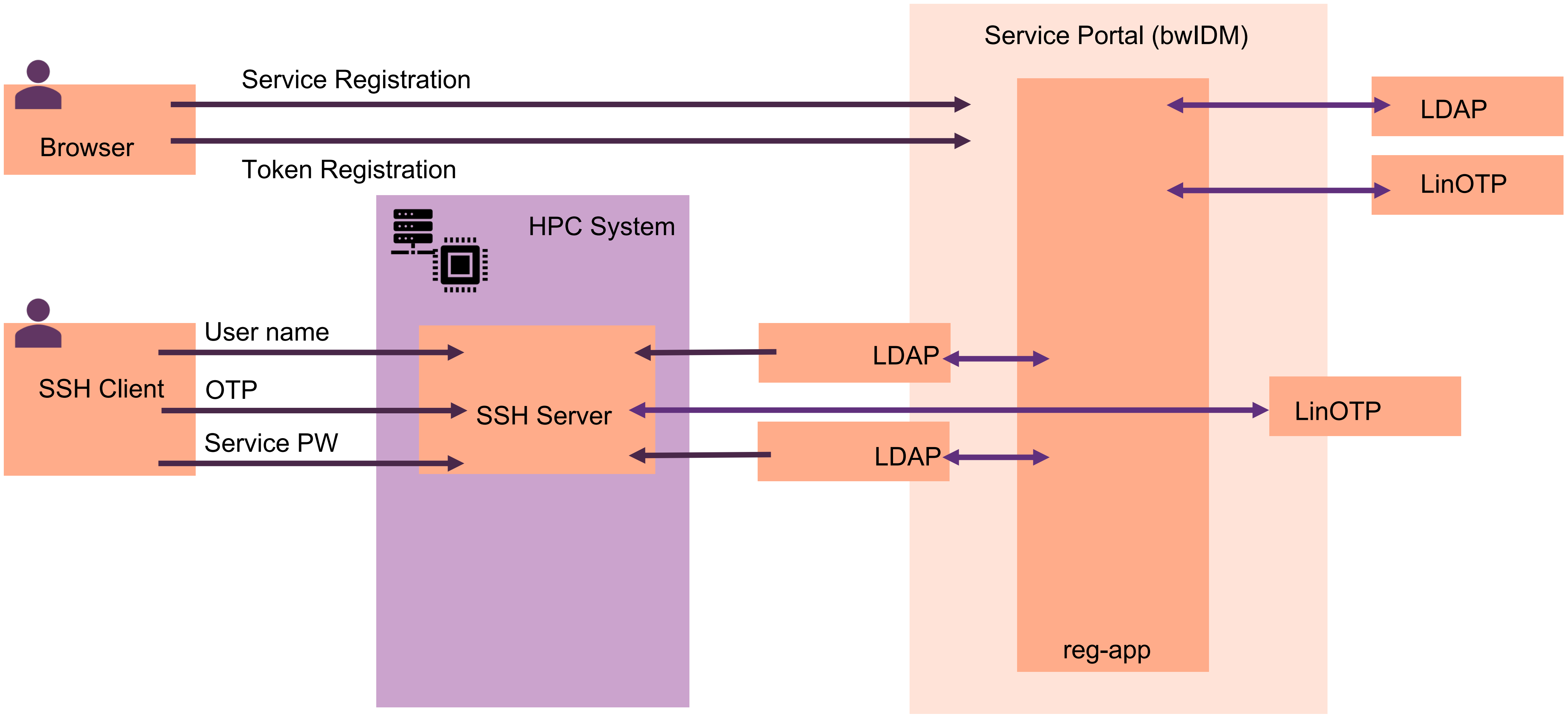
Where reg-app and HPC come together



Where reg-app and HPC come together

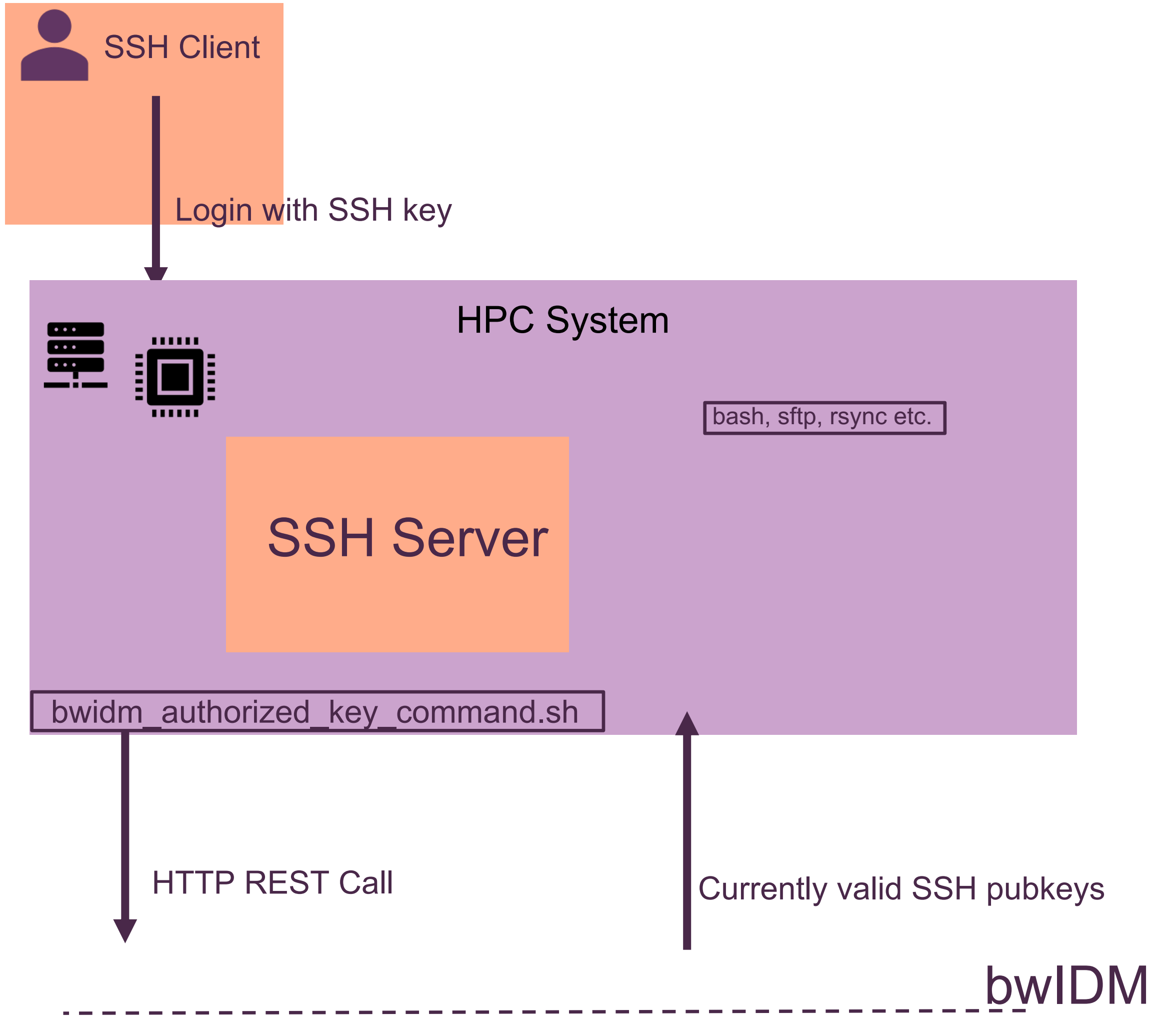


Where reg-app and HPC come together



Where reg-app and HPC come together

- 2FA: Time-Based One-Time Passwords (OATH/TOTP) managed by a dedicated LinOTP server.
- reg-app: token and SSH key management
- Reg-app: LinOTP-compatible HTTP interface and a custom HTTP REST API
- Linux servers use pam_linotp for PAM and a custom AuthorizedKeyCommand for OpenSSH



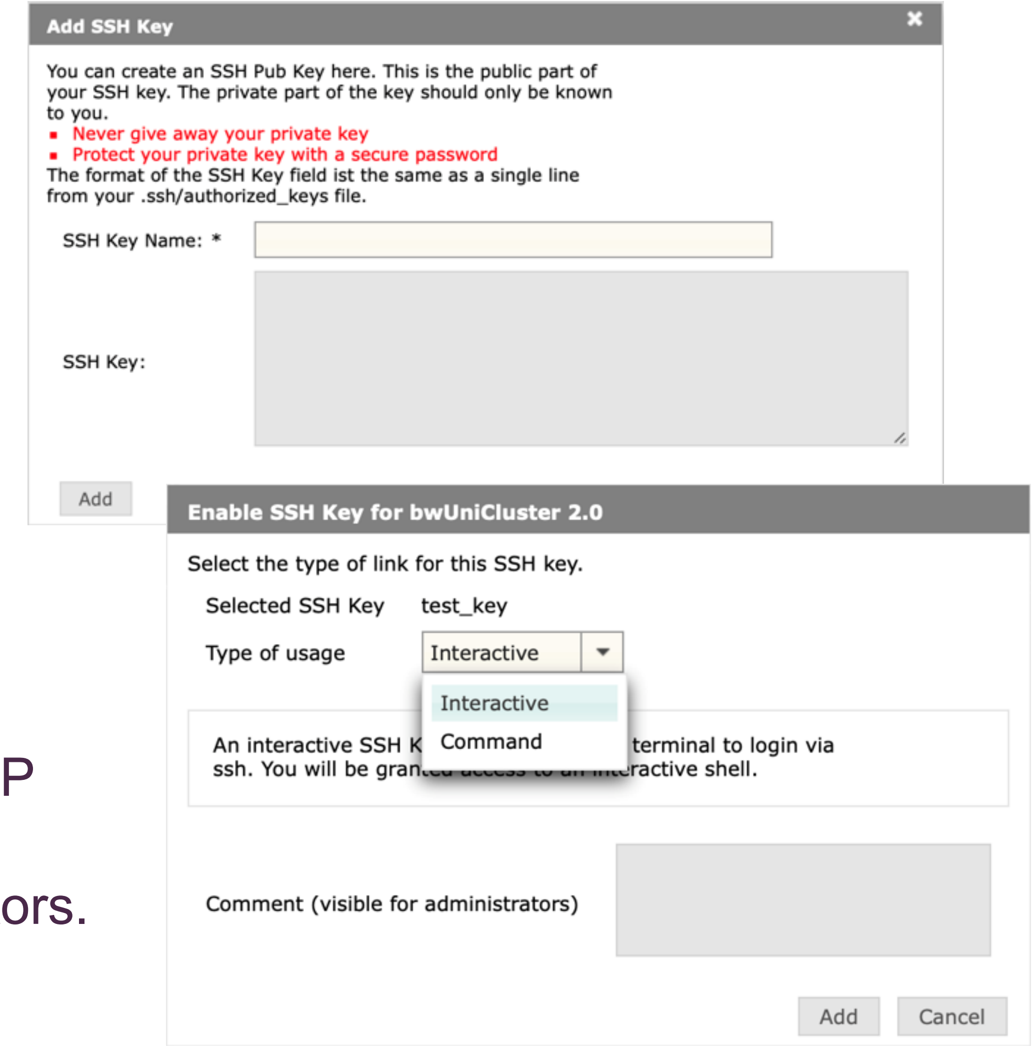
Where 2FA and SSH keys come together

Interactive keys

- For normal interactive logins
- Unrestricted as to which commands can be executed
- Usability is limited to a period of one hour after the last successful two-factor login. Has to be “unlocked”
- Valid for six months.

Command keys

- Intended for scientific workflow systems, continuous integration, interactive data exploration e.g.
- Always valid, do not have to be “unlocked” with 2FA.
- Restricted to a single command and to either a single IP address or a small IP subnet.
- Have to be checked and approved by HPC administrators.
- Valid for one month.



Add SSH Key

You can create an SSH Pub Key here. This is the public part of your SSH key. The private part of the key should only be known to you.

- Never give away your private key
- Protect your private key with a secure password

The format of the SSH Key field ist the same as a single line from your .ssh/authorized_keys file.

SSH Key Name: *

SSH Key:

Enable SSH Key for bwUniCluster 2.0

Select the type of link for this SSH key.

Selected SSH Key test_key

Type of usage Interactive

An interactive SSH K terminal to login via ssh. You will be gran ractive shell.

Comment (visible for administrators)

And this is what happened in real life…

And this is what happened in real life…

Example #1

”This isn’t working at all – I cant’ register any token!”

Problem: Edge case due to misconfiguration of production server.

Lesson: Allow end users to perform preparatory steps before 2FA goes live. Luckily we did it right the first time and fixed this two days before.

And this is what happened in real life…

Example #1

”This isn’t working at all – I cant’ register any token!”

Problem: Edge case due to misconfiguration of production server.

Lesson: Allow end users to perform preparatory steps before 2FA goes live. Luckily we did it right the first time and fixed this two days before.

Example #2

”I’ve lost the only token I’ve ever registered…”

Problem: Users lose their only active token because they lose the device, uninstall the TOTP app, reset or reinstall the OS, etc.
Lesson: Emphasized the importance of registering a second token or use a Backup TAN list in our user documentation and training materials.

And this is what happened in real life…

Example #1

”This isn’t working at all – I cant’ register any token!”

Problem: Edge case due to misconfiguration of production server.

Lesson: Allow end users to perform preparatory steps before 2FA goes live. Luckily we did it right the first time and fixed this two days before.

Example #2

”I’ve lost the only token I’ve ever registered…”

Problem: Users lose their only active token because they lose the device, uninstall the TOTP app, reset or reinstall the OS, etc.
Lesson: Emphasized the importance of registering a second token or use a Backup TAN list in our user documentation and training materials.

Example #3

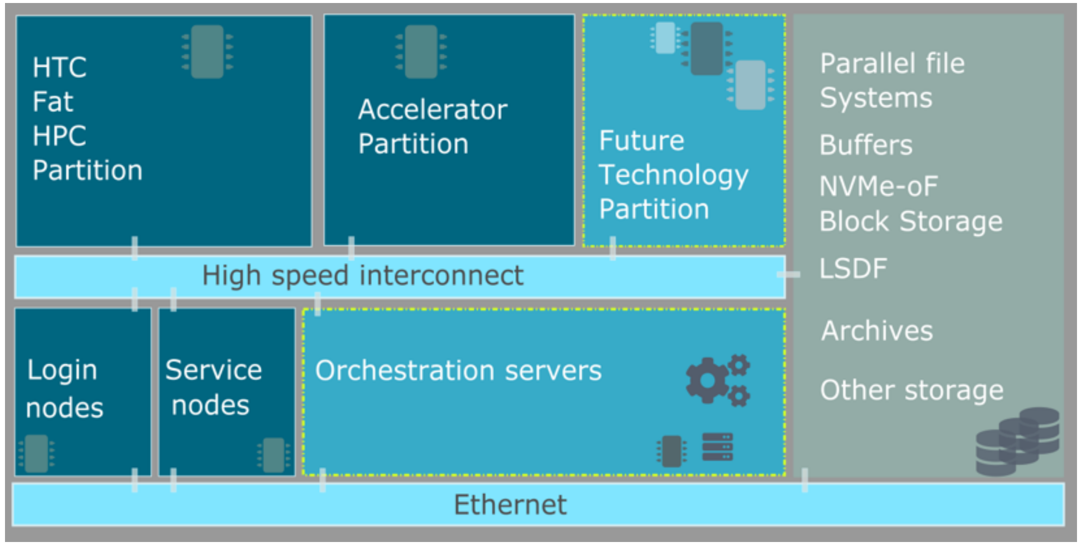
”I do have multiple tokens, but sometimes they get rejected?”

Problem: LinOTP server checks the provided OTP against every active token registered for a user. We set the limits too low.

Lesson: Expect „power users“ to take 2FA seriously. Two Yubikeys, two mobile devices and a Backup TAN list per user is actually not that much.

Introduction of Orchestration Servers - HoreKa

- HoreKa: New Tier-2 system, to be up and running in Q1/2021.
- New core layout of the HPC system (storage, compute, login)
- The orchestration servers include e.g. JupyterHub servers and Continuous Integration environment.
- Everything is covered by 2FA, enabled through OpenID Connect.



HTC Fat HPC Partition

Accelerator Partition

Future Technology Partition

Parallel file Systems

Buffers

NVMe-oF Block Storage

LSDF

Archives

Other storage

High speed interconnect

Login nodes

Service nodes

Orchestration servers

Ethernet

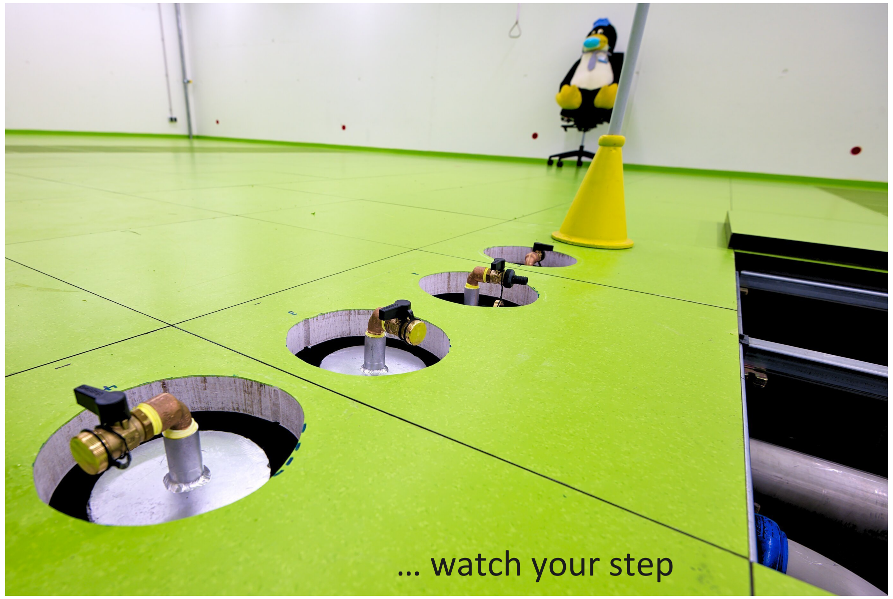
Conclusion and Future Developments

Addressing questions like:

- How to transfer data?
- Monitoring and performance indicators?
- Abstract ways of job submission?

Lesson learned:
Communication is key!

Contact: jennifer.buchmueller@kit.edu



... watch your step