VISPA: Web Platform for Scientific Data Analysis Needs

Martin Erdmann, Peter Fackeldey, Benjamin Fischer, Marcel Rieger
for the VISPA Team

Physics Institute 3A, RWTH Aachen University

ESCAPE Science Platform - technical evaluation
21.10.2019

RWTH AACHEN UNIVERSITY

VISPA (VISual Physics Analysis)

- platform for data analysis
- comprehensive work environment
- configurable and integratable
- allows for specialized tools

Outline

- Audience & requirements
- Concept
- Components
- What we are doing
- Implementation & Demo
  - VISPA Web
  - New VISPA
- Comparison to Alternatives
- Discussion

Broad spectrum of users

Undergraduate student

- quick & easy access
  - ➜ no installation
- basic analysis environment
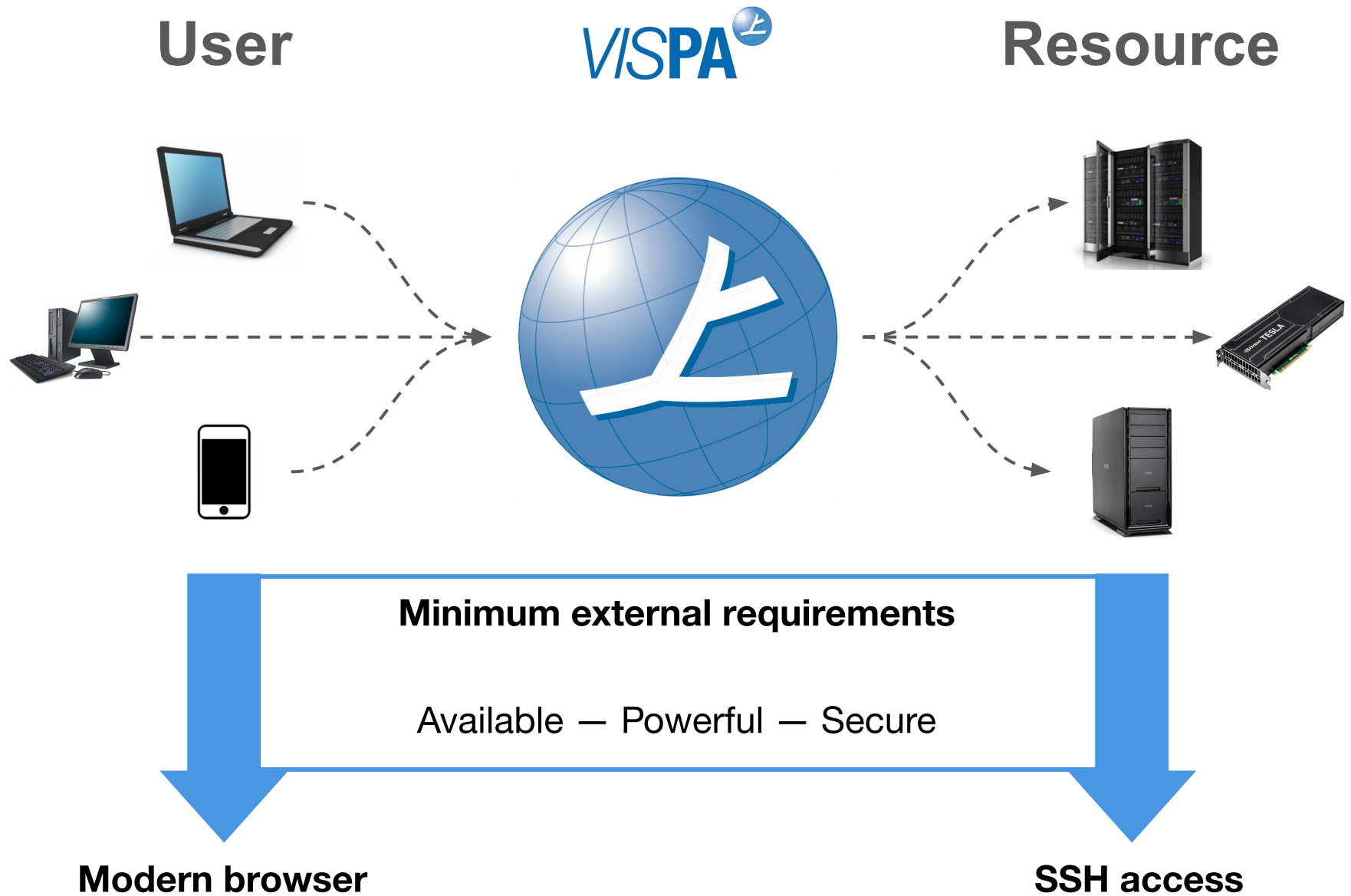 - ➜ numpy, matplotlib, …

Senior researcher

- full (software) environment control
  - ➜ Virtualenv, …
- custom resources
  - ➜ Accelerators of CampusIT
- specialized tools
  - ➜ Tensorboard, HTCondor, …

Summarized requirements

high accessibility & low entrance barrier

flexible & extendable

User

VISPA

Resource

Minimum external requirements

Available — Powerful — Secure

Modern browser

SSH access

User

VISPA

Resource

- does not want to install/setup anything
- wants to work from anywhere
- needs persistence to continue where interrupted
- uses basic tools: file browser, code editor, terminal
- profits from custom/specialized tools
- may use their credentials for custom Resources

- anything SSH reachable
- may need intermediate connections (SSH ProxyJump)
- needs automatic bootstrapping
- consists of both:
  ○ storage
  ○ computing power

User

Resource

- User Interface: HTML + CSS + JS
  - desktop like interface
  - dynamic: no (re)loading
  - reactive & multi-tasking
- VISPA: HTTP(S)S server
  - authentication
  - session & preferences storage
  - multi-resource & -user handling
  - APIs for translation to/through SSH
  - delivers UI/bootstrapping code
- Resource side RPC
  - implements complexer RPC directly on Resource

- **Software**: VISPA Web
  - https://git.rwth-aachen.de/3pia/vispa/vispa-web
  - 7+ years of development
- **Service**: VISPA Web
  - https://vispa.physik.rwth-aachen.de/server/
  - 5+ years of wide active usage
  - instance of Software
  - default Resource:
    - 200 CPU cores
    - 160 TFlops in GPUS
    - 30 TB storage

- Track record:
  - Bachelor & Master level physics lectures
  - Workshops & Tutorials (e.g. Big Data Science in Astroparticle Research)
  - Day-to-day work from Bachelor Student to Professor
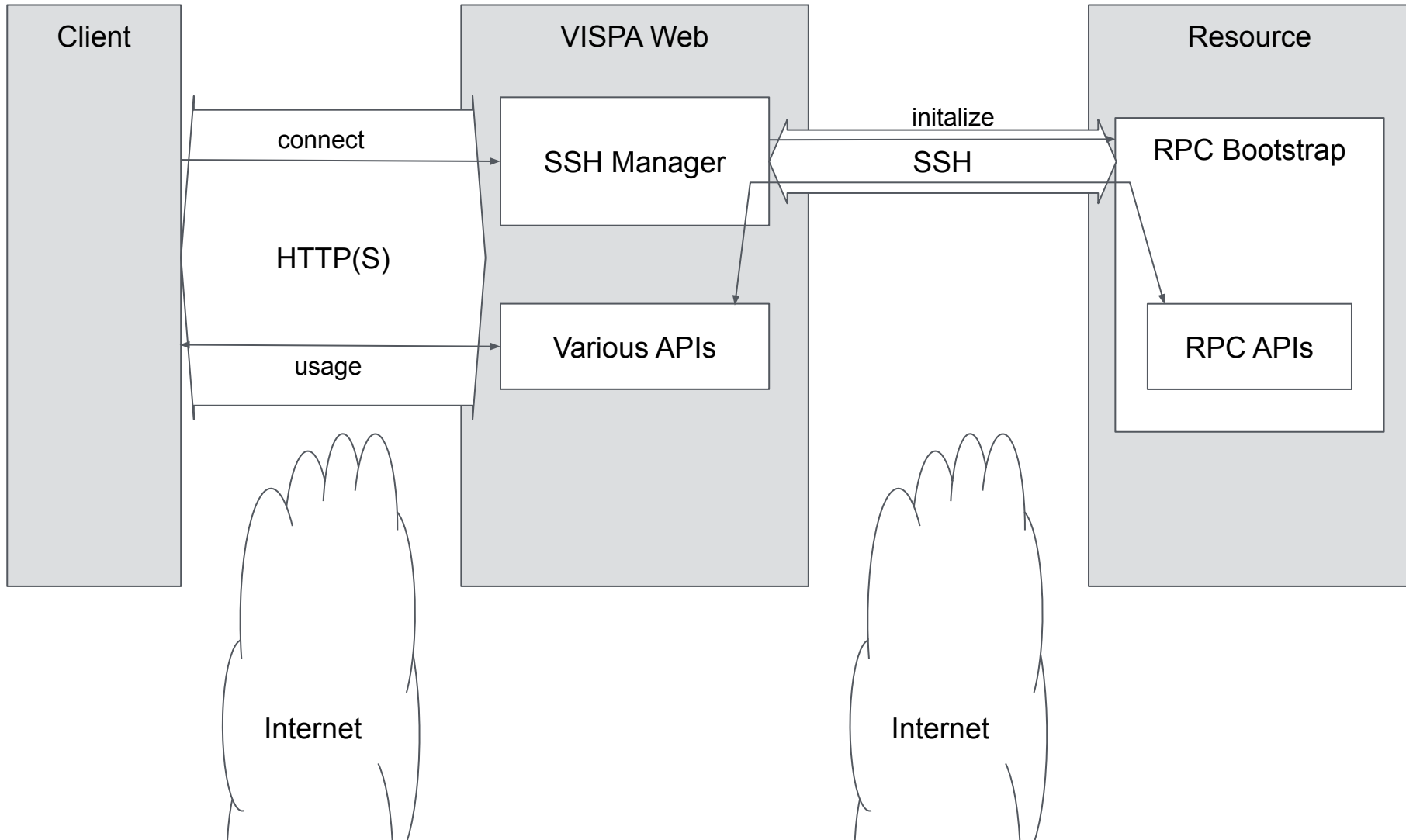
DEMO

VISPA Web

- User Interface
  - modularized, reusable OOP ode base
  - mixed technologies: jQuery, Vue.js, Lesscss, …
  - uses external projects: Ace (powerful code editor), xterm.js, …
- Server
  - CherryPy based web server
  - SqlAlchemy backed storage
  - Resource attached via RPyC & Paramiko
  - User-Management system w/ User, Group and Project
- Resource
  - python based in-memory bootstrapping
  - RPC code delivered from Server

- Extension Mechanism
  - bundles: UI & Server & RPC code
 - chosen by Server operator (not User)

Generalized paradigm

integrate existing tools with web-based UI i.e. JupyterLab

➔   HTTP API for (complex) connection brokering

Central aspects:
- highly modularized: pluggable authentication & *pipelets*
- more flexibility for the user: SSH-tunneling, Remote software
- leverages excellent work of community: JupyterLab
- custom tools mostly implemented as JupyterLab extension
    - ➔   better accessibility & community (code-)feedback
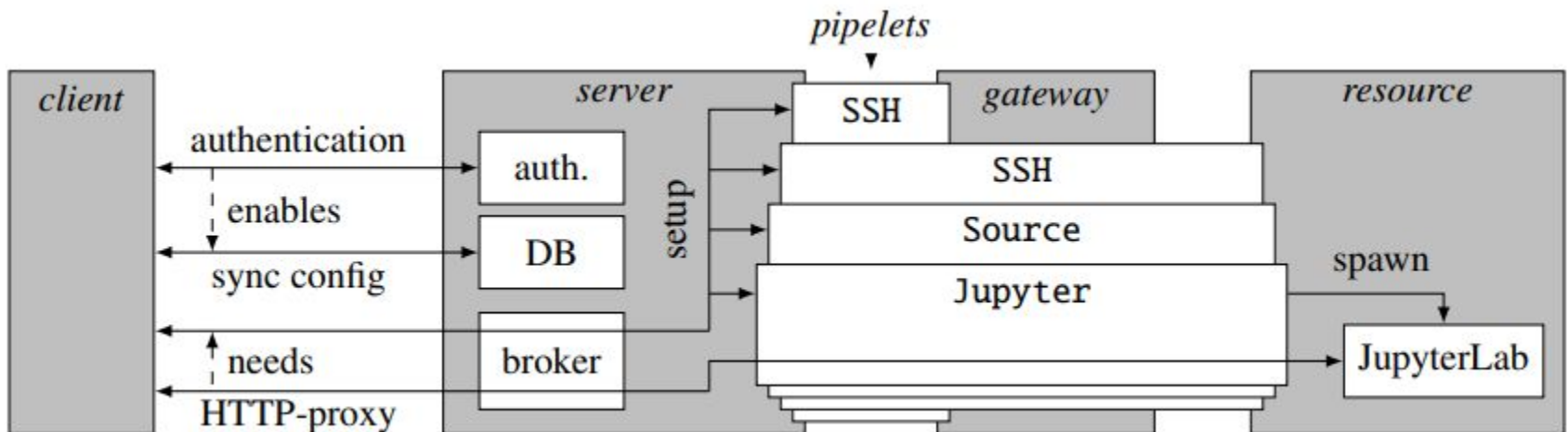
DEMO
New VISPA

- User Interface
  - Nuxt.js based
  - remaining majority handled by external projects i.e. JupyterLab
- Server
  - Node.js
  - authentication: passport.js modules
  - database: sequelize (SQL)
  - SSH connections: ssh2
- Resource-site RPC
  - handled by external projects i.e. JupyterLab

Extension Mechanism(s):
- Authentication: passport.js
- Connection type: *pipelets*
- User Interface & Tools:
  - JupyterLab Extension
  - anything with WebUI i.e. TensorBoard

- Google Colab — service
  - based on Jupyter Notebooks, integrated with Google computing resources
- Swan (CERN) — service
  - based on Jupyter Notebooks, integrated with CERN computing resources
- JupyterHub — software
  - used with Jupyter Notebooks or JupyterLab
  - configured by Administrator:
    - sole control over resources & their allocation
    - variety of integrations: Kubernetes, HTCondor, OpenStack, SSH …
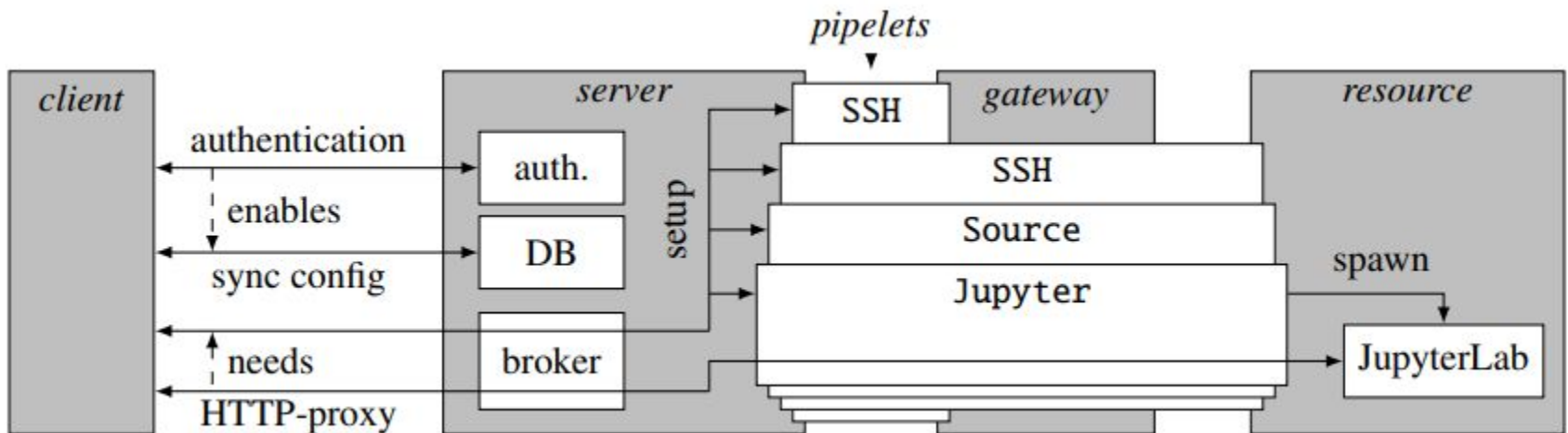 - connects to the Resource through the Internet

VISPA has

- flexibility for user
- more than Notebooks (JupyterLab does this too)
- can integrate custom Web-Frontends, e.g. TensorBoard (new VISPA)

User   **VISPA**   **Resource**

VISPA (VISual Physics Analysis)

- web-platform for data analysis
- comprehensive, browser-based, and extensible work environment
- use any external SSH resource
- configurable and integratable
- allows for specialized tools

Backup