Preserving Databases

An emulation-based approach

Dr. Klaus Rechert (University of Freiburg)

@kurau5u

Databases for 2080 – Oct 5 2021

Preservation Challenge(s)

- Is it possible to access today‘s databases in 60 years
  - Mindprovoking *theoretical* question

- Database is a weakly defined term
  - SQL, NoSQL, etc… ?
  - Which data format ?
  - How much context is required ?

Can we plan for 60 years?

What is the resulting risk profile?

Choosing a strategy

- Data-driven
  - Extract data (e.g. SQL dump)
  - Migrate / generalize data (CSV, SIARD, etc)
  - Reconstruction in the future
  - Rewrite queries, replace / substitute clients

- Software-driven
  - Preserve the software
  - Describe and manage a technical stack
    - → Emulation strategy
  - Access through „native“ interfaces e.g. UI, ODBC over TCP/IP, original software client etc.

Data
Structural information

Database software

Choosing a (lazy) strategy

- Preserving Database Software
  - Common, widely used software?
  - Standard / simple setup, with little manual customization?
  - Limited interaction with external systems?

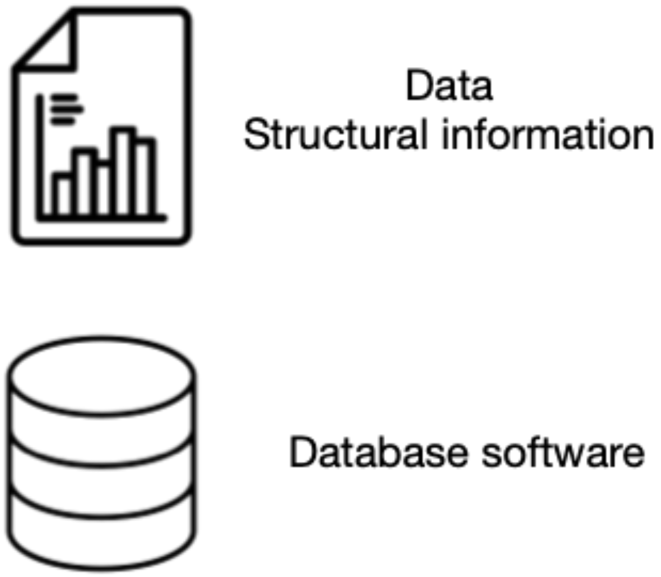
→ Low technical complexity

→ Able to bear some risks ?

→ Outsource software preservation

→ Create some metadata on the current setup

→ Take the free ride



Data
Structural information

Database software

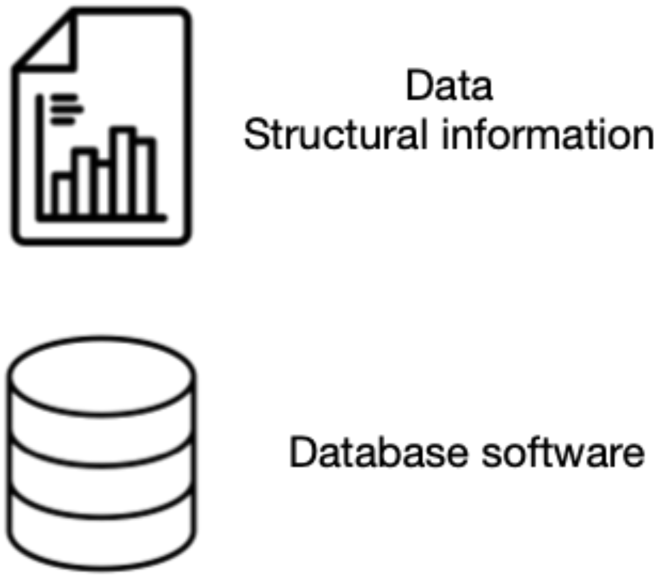
Choosing a (less lazy) strategy

- Preserving Database Software
  - Common, widely used software?
  - Standard / simple setup, with little manual customization?
  - Limited interaction with external systems?

→Use a contemporary emulation framework

→Create an installation of the database and deploy the data

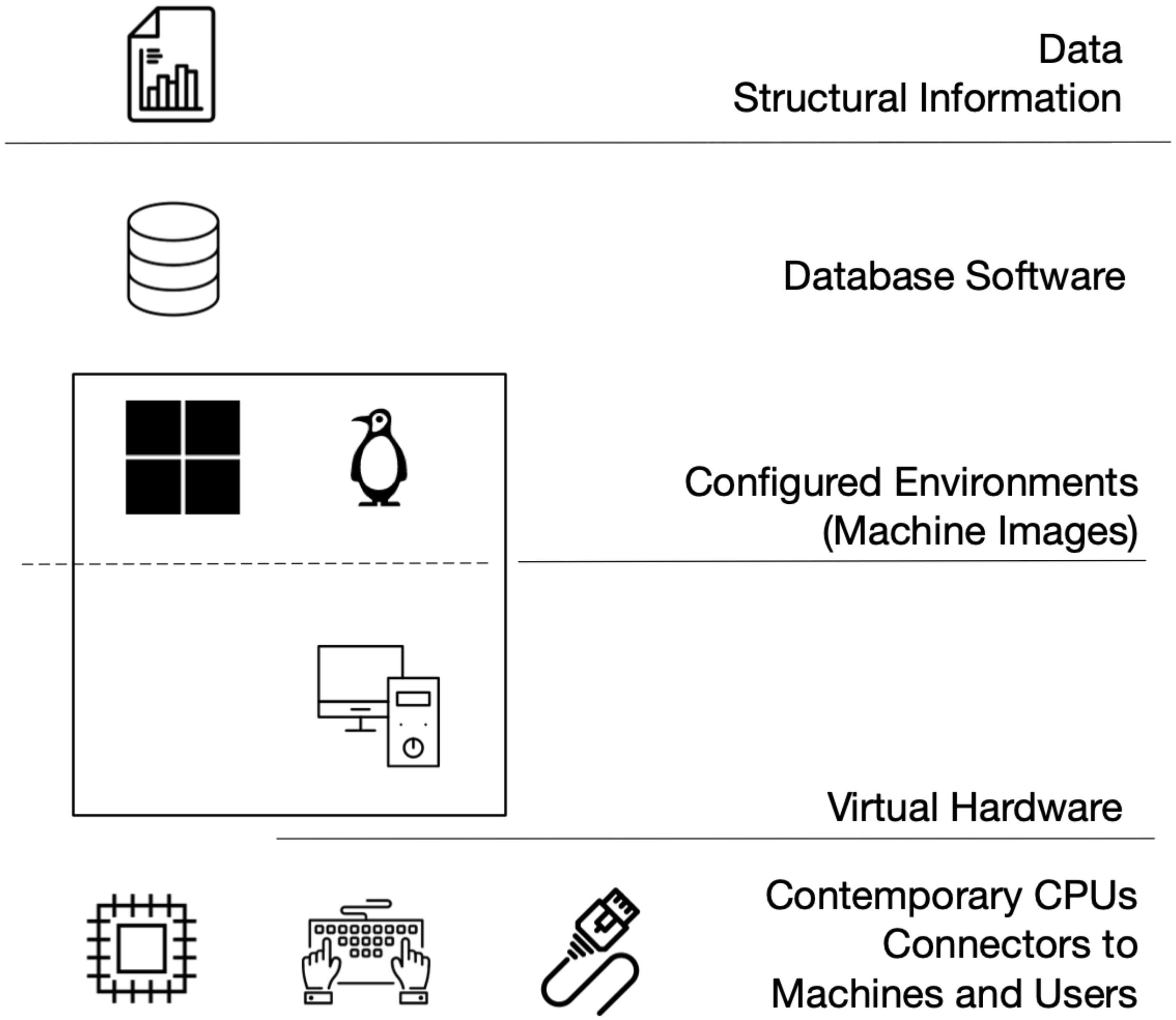
→Verify your result



Data
Structural information

Database software

Technical Stack



Data
Structural Information

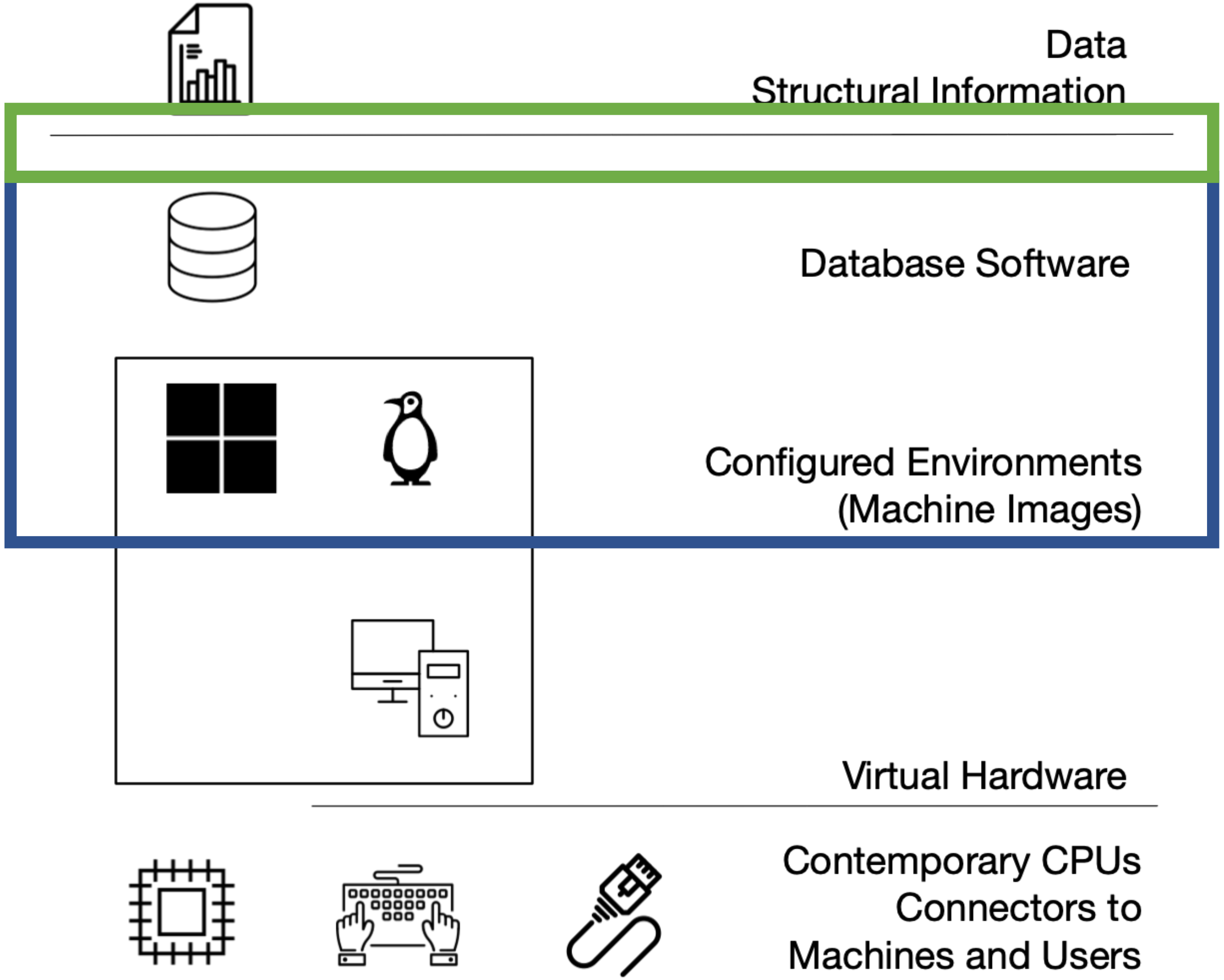
Database Software

Configured Environments
(Machine Images)

Virtual Hardware

Contemporary CPUs
Connectors to
Machines and Users

Implementing the Technical Stack

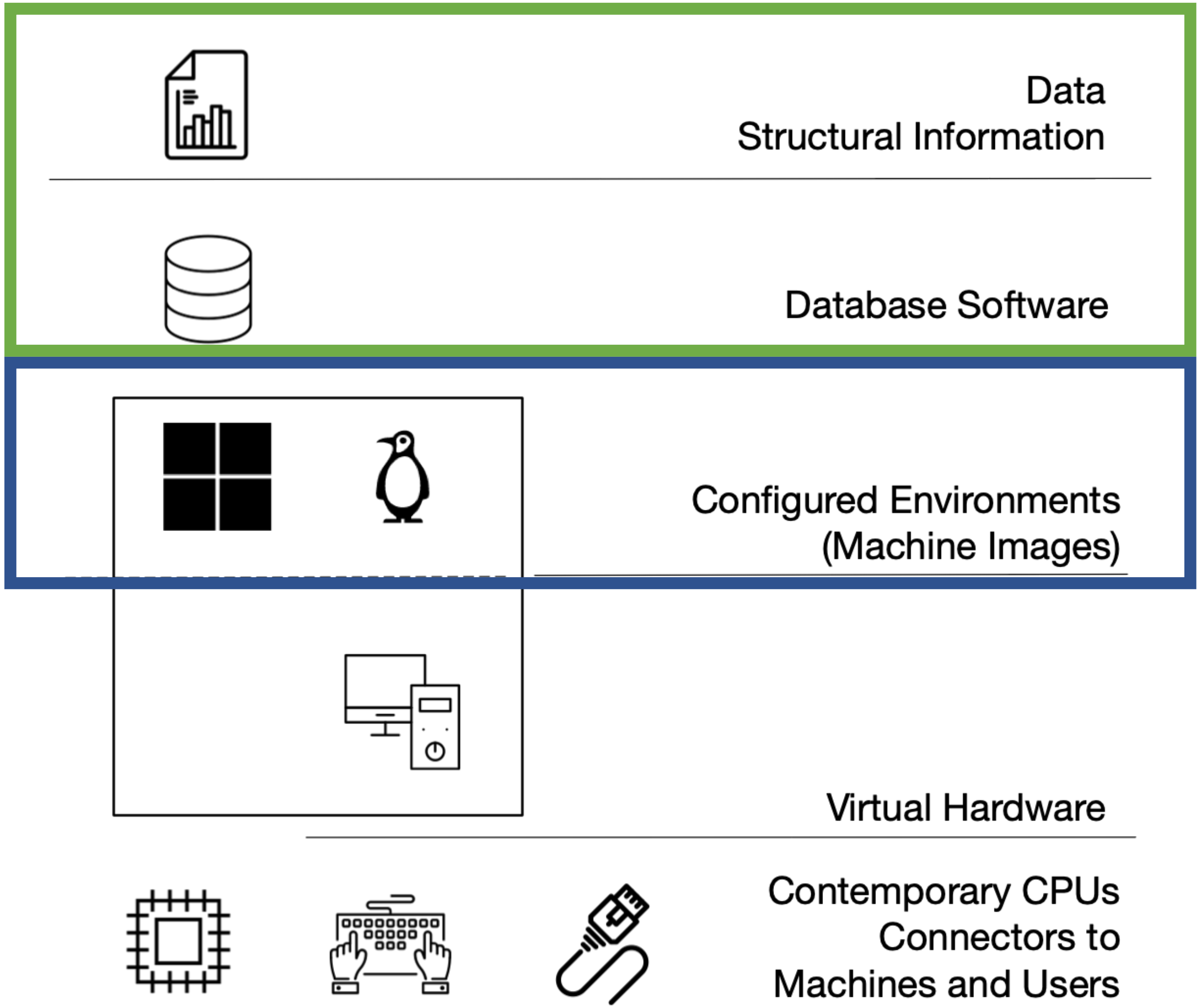


Create installation/configuration templates for common databases (MySQL/MariaDB, Postgres, etc…)

- Standardize and simplify common installations
- Preserve (encoded) operation knowledge
- Start deployment from file-system backup (data)
- Maintain stack/template independent from data

→ Reusable for many artifacts
→ Can be highly automated
→ Cooperate and share

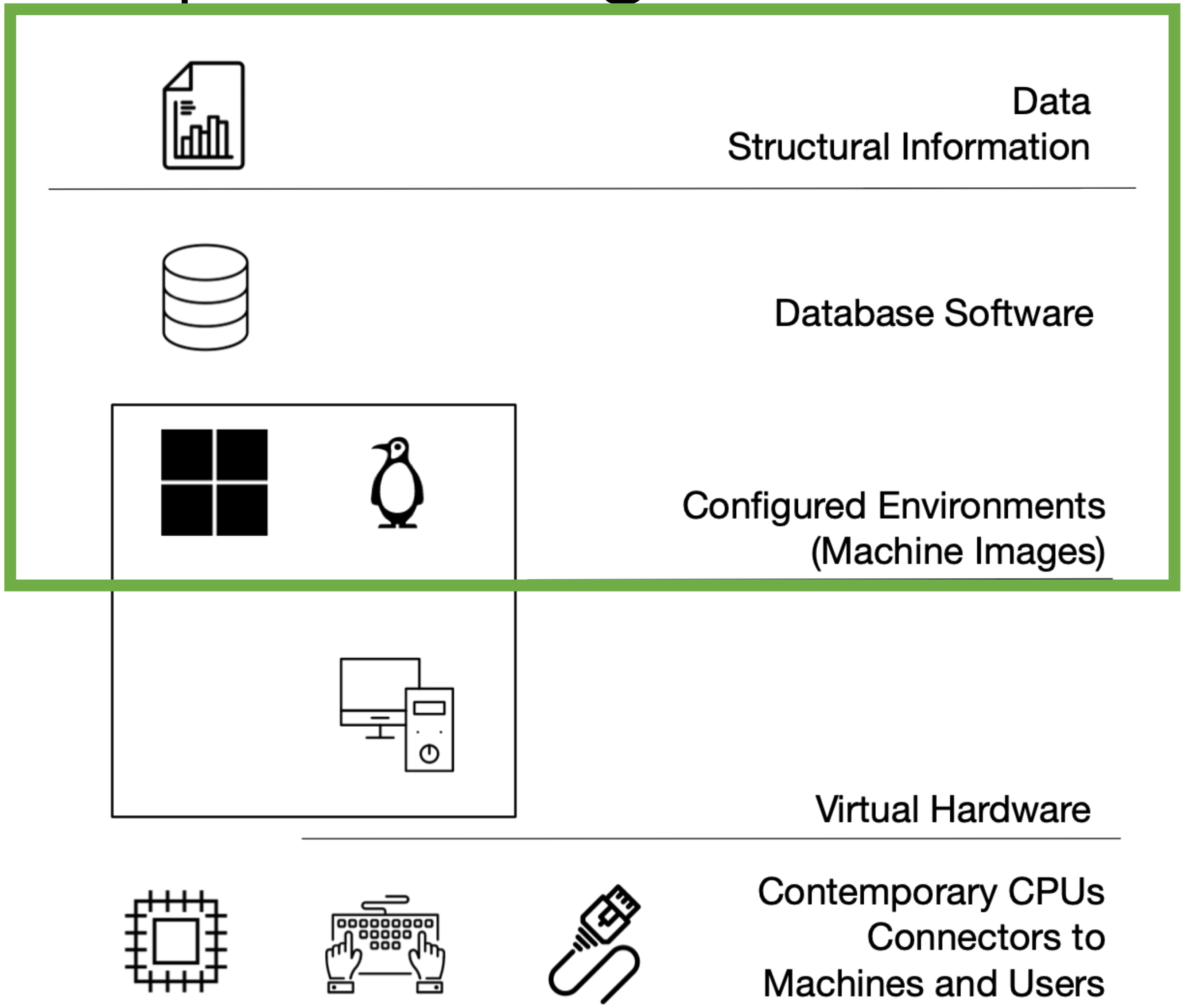
Implementing the Technical Stack



Complex setups may require manual or installation steps.
- build on top standardized software stacks

→ Reusable for many artifacts
→ Can automated
→ Cooperate and share

Implementing the Technical Stack



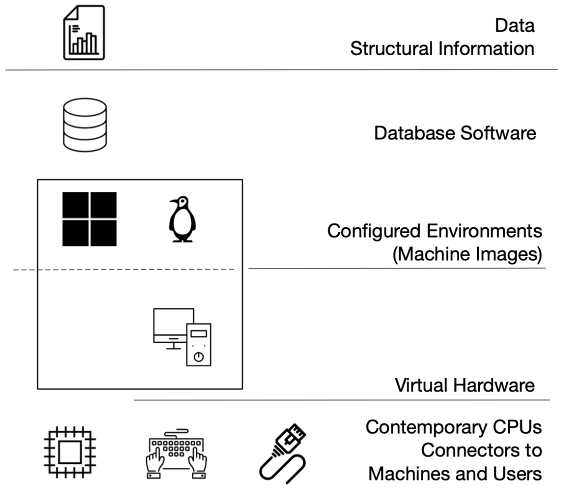
Complex setups may not be trivially rebuildable.

- Preserve full VM / disk image
  - generalize hardware dependencies!
- Rebuild from a machine from filesystem backup (e.g. tape)
  - E.g. use container

→ Reusable for many artifacts
→ Can be highly automated
→ Cooperate and share

Emulation

- Emulation as Conceptual Framework
  - Different options to preserve a DB instance
    - Complexity matters
    - Long-term risk profile

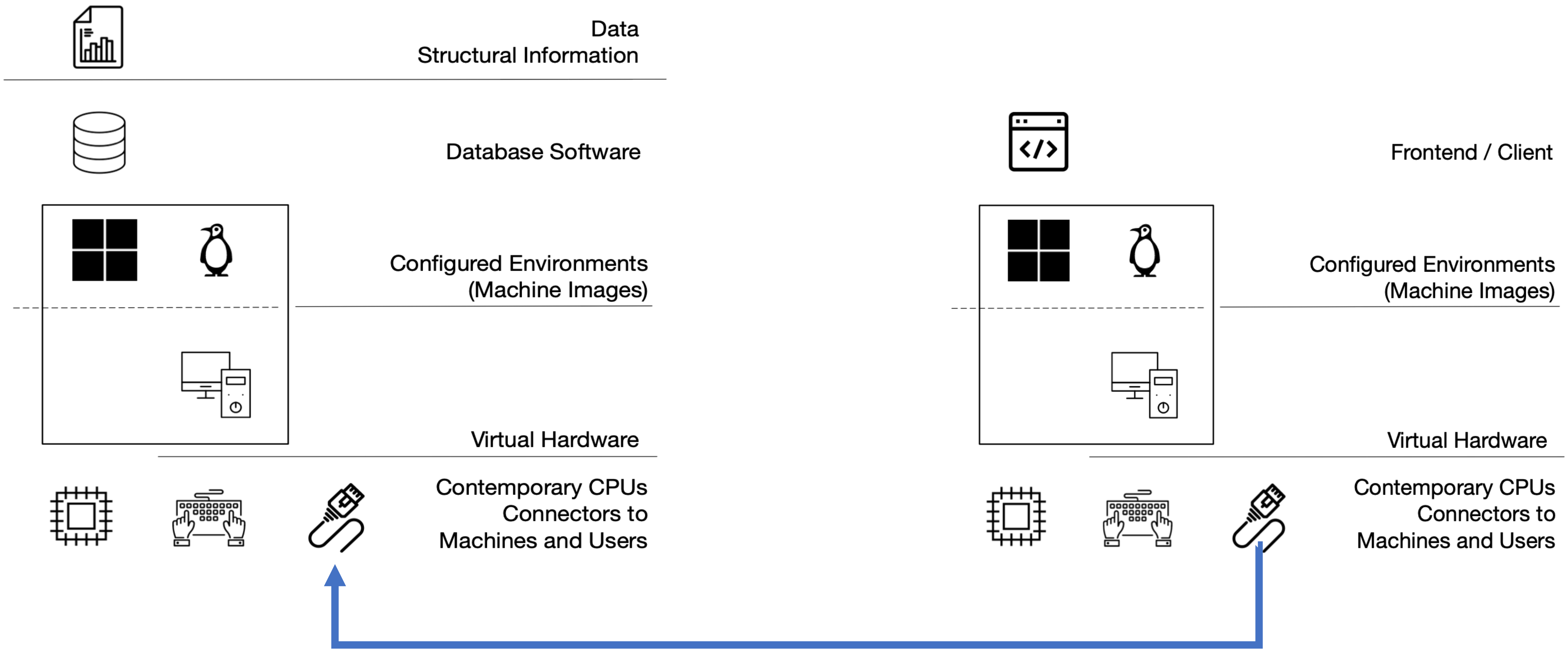


  - Emulation is software-based
    - Preservation planning: prepare for obsolescence
  - Scaling emulation
    - Emulation scales well with users! Cooperate!
    - Plenty of automation options

Emulation

- Use Emulation as Access Technology
  - To contextualize the software stack
    - Provide interfaces and interaction to enable data access
    - Rebuild administrative / business processes
  - Prolonged sunset phase of obsolete services – continuous access

Contextualize & Connect



Data
Structural Information

Database Software

Configured Environments
(Machine Images)

Virtual Hardware

Contemporary CPUs
Connectors to
Machines and Users

Frontend / Client

Configured Environments
(Machine Images)

Virtual Hardware

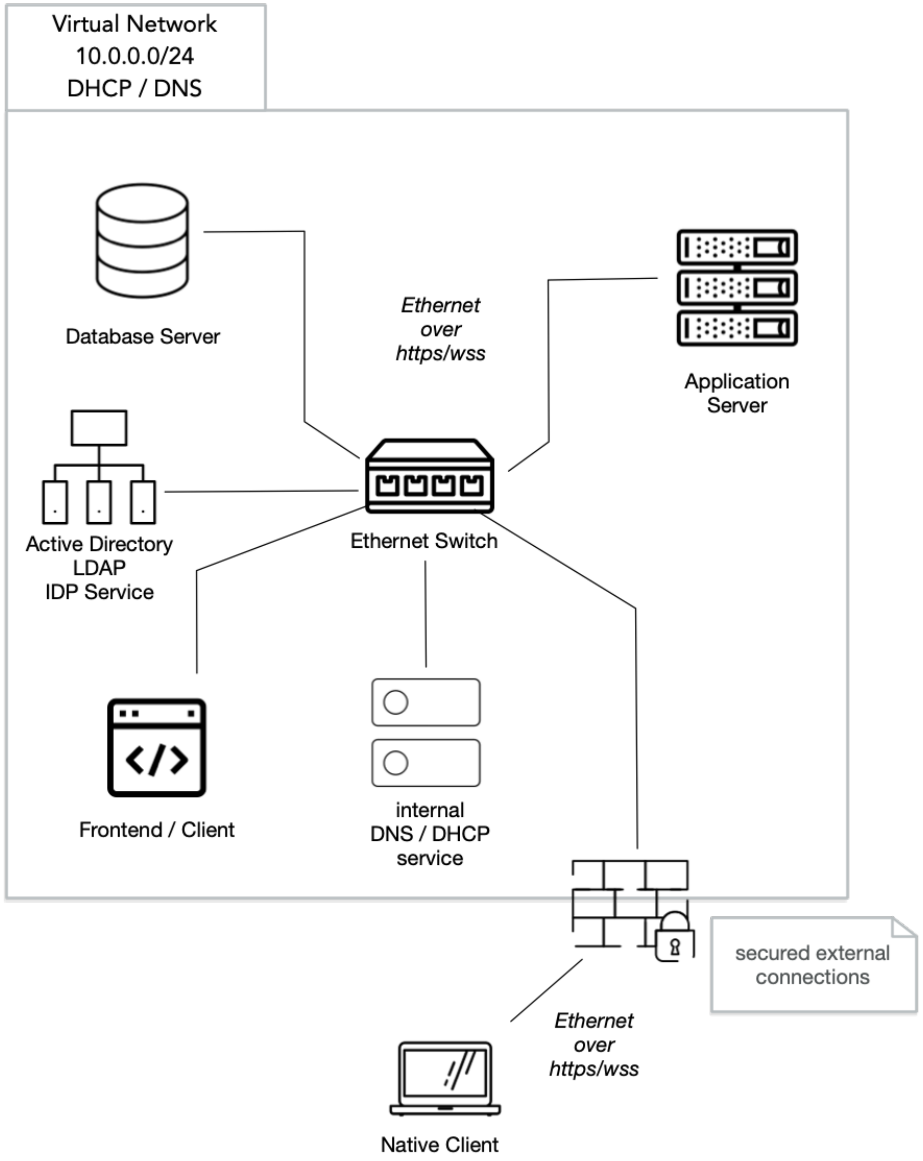
Contemporary CPUs
Connectors to
Machines and Users

Emulating a networked environment

Emulating Networks

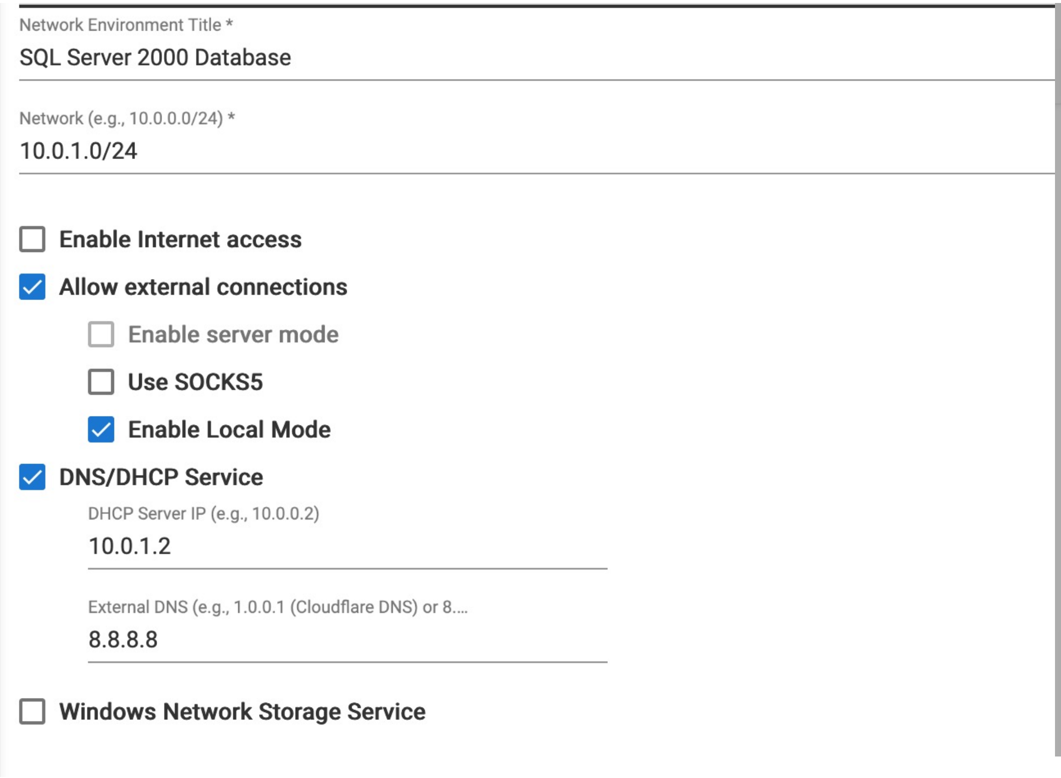
- Extending the scope of emulation
  - Interconnected instances
- Emulate *cable, copper* and Ethernet components
  - Simple (re-)implement
  - Any higher protocol is supported
- Isolation / Security
  - Full isolation
  - Integrated network services

- Networked integration options



Emulating Networks

1. Create a new network



Network Environment Title *
SQL Server 2000 Database

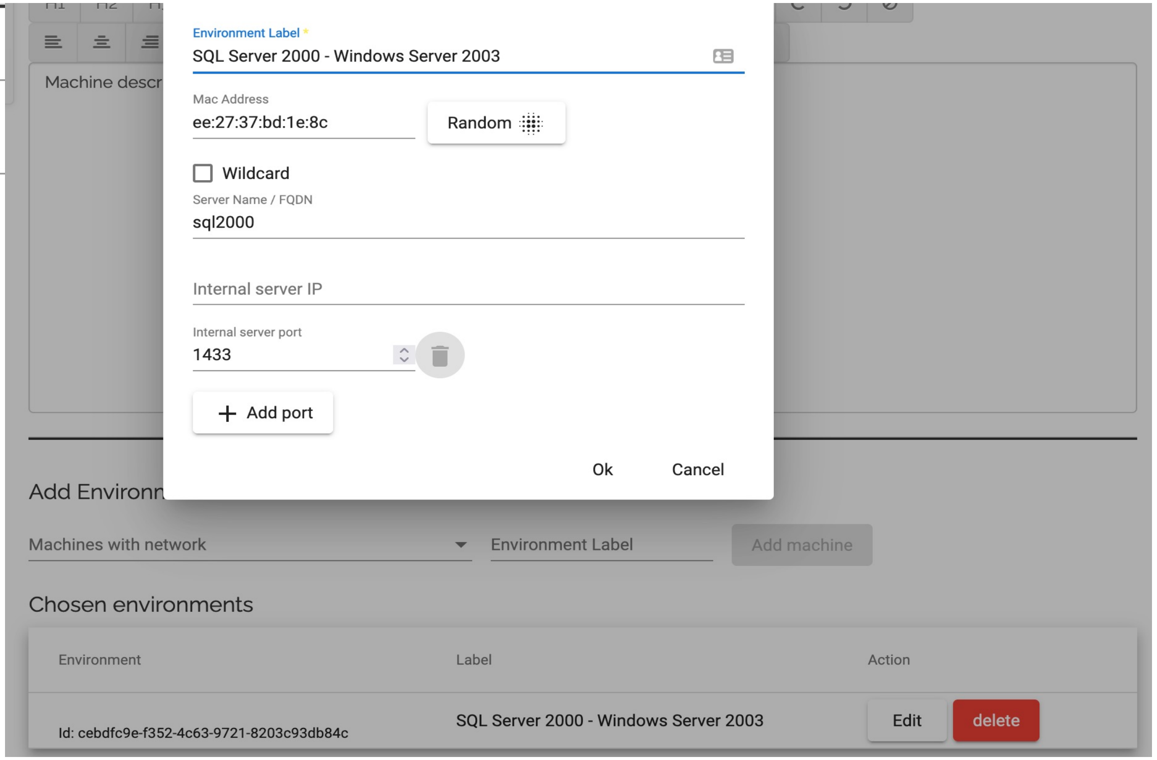
Network (e.g., 10.0.0.0/24) *
10.0.1.0/24

- ☐ Enable Internet access
- ☑ Allow external connections
  - ☐ Enable server mode
  - ☐ Use SOCKS5
  - ☑ Enable Local Mode
- ☑ DNS/DHCP Service

  DHCP Server IP (e.g., 10.0.0.2)
  10.0.1.2

  External DNS (e.g., 1.0.0.1 (Cloudflare DNS) or 8....
  8.8.8.8
- ☐ Windows Network Storage Service

2. Add machines to he network



Machine descr

Environment Label *
SQL Server 2000 - Windows Server 2003

Mac Address
ee:27:37:bd:1e:8c

- ☐ Wildcard

Server Name / FQDN
sql2000

Internal server IP

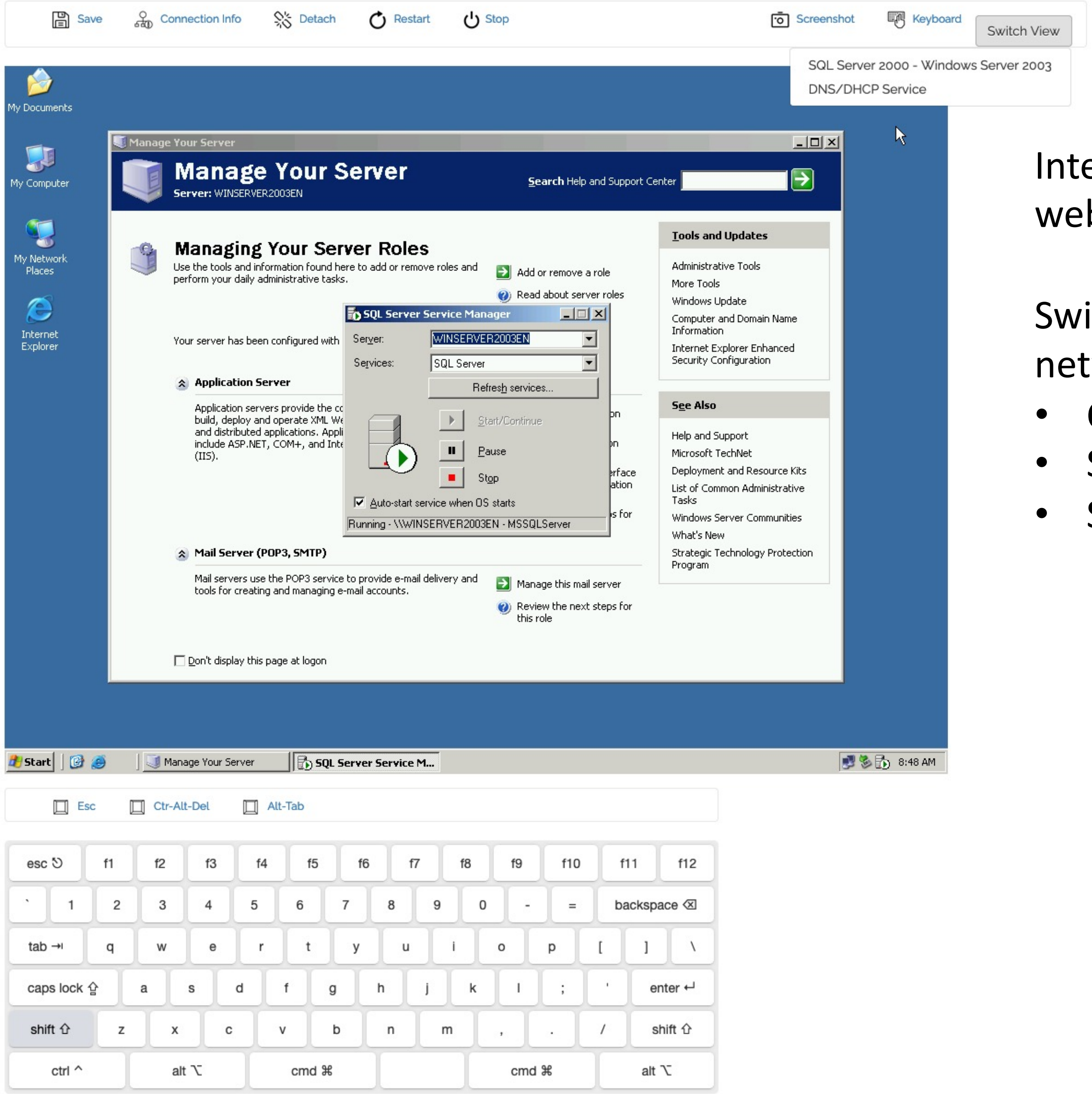
Internal server port
1433

Add Environn

Machines with network
Environment Label

Chosen environments

Environment	Label	Action
Id: cebdfc9e-f352-4c63-9721-8203c93db84c	SQL Server 2000 - Windows Server 2003	Edit delete


Save
Connection Info
Detach
Restart
Stop
Screenshot
Keyboard
Switch View
SQL Server 2000 - Windows Server 2003
DNS/DHCP Service
My Documents
My Computer
My Network Places
Internet Explorer
Manage Your Server
Manage Your Server
Search Help and Support Center
Managing Your Server Roles
Add or remove a role
Read about server roles
Application Server
Mail Server (POP3, SMTP)
Manage this mail server
Review the next steps for this role
Tools and Updates
Administrative Tools
More Tools
Windows Update
Computer and Domain Name Information
Internet Explorer Enhanced Security Configuration
See Also
Help and Support
Microsoft TechNet
Deployment and Resource Kits
List of Common Administrative Tasks
Windows Server Communities
What's New
Strategic Technology Protection Program
SQL Server Service Manager
Refresh services...
Start/Continue
Pause
Stop
Start
Manage Your Server
SQL Server Service M...
8:48 AM
Esc
Ctr-Alt-Del
Alt-Tab
esc f1 f2 f3 f4 f5 f6 f7 f8 f9 f10 f11 f12
` 1 2 3 4 5 6 7 8 9 0 - = backspace
tab q w e r t y u i o p [ ] \
caps lock a s d f g h j k l ; ' enter
shift z x c v b n m , . / shift
ctrl alt cmd ⌘ cmd ⌘ alt

Server: WINSERVER2003EN

Use the tools and information found here to add or remove roles and perform your daily administrative tasks.

Your server has been configured with

Application servers provide the co… build, deploy and operate XML We… and distributed applications. Appli… include ASP.NET, COM+, and Inte… (IIS).

Mail servers use the POP3 service to provide e-mail delivery and tools for creating and managing e-mail accounts.

□ Don't display this page at logon

Server: WINSERVER2003EN

Services: SQL Server

☑ Auto-start service when OS starts

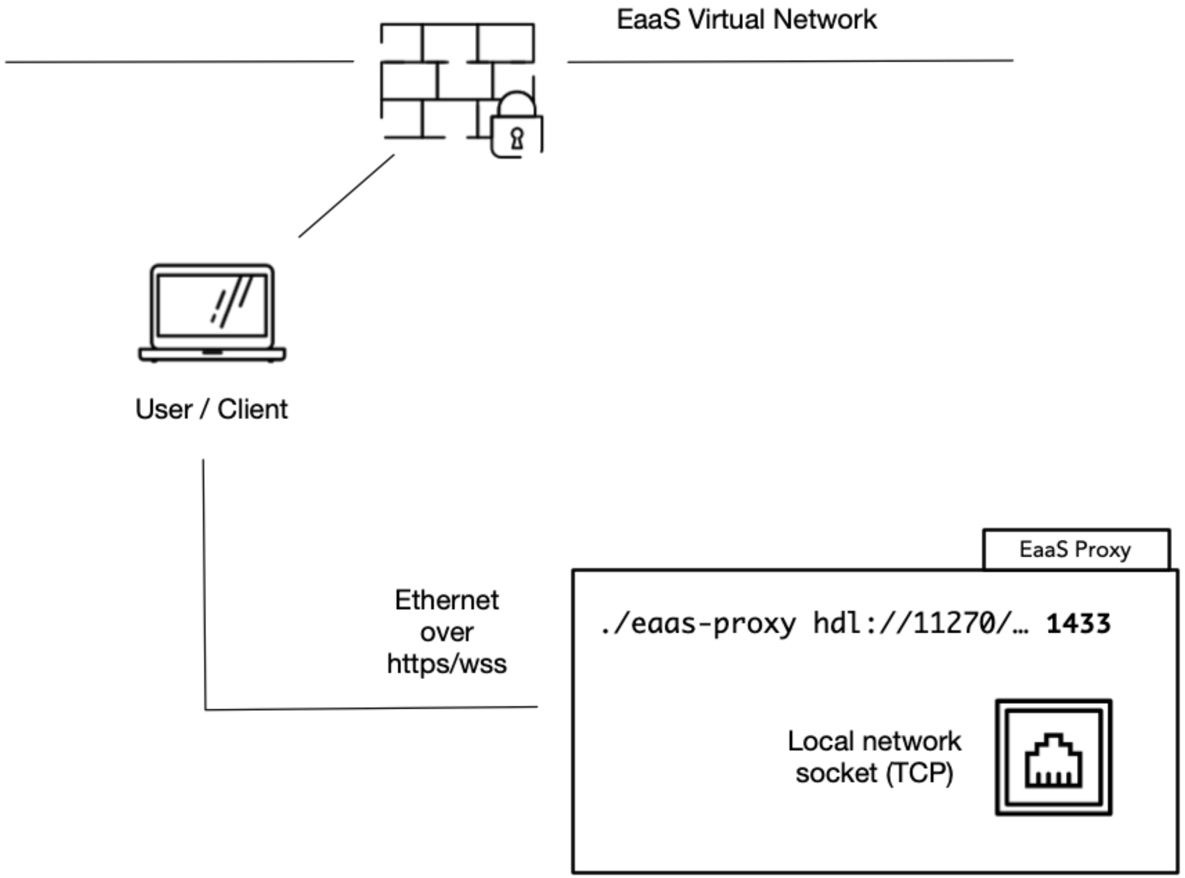
Running - \\WINSERVER2003EN - MSSQLServer

Interactive access, e.g. through web browsers.

Switch between machines in the network, e.g.

- Clients
- Server
- Service components

Networked Integration / Local Access



EaaS Virtual Network

User / Client

Ethernet over https/wss

EaaS Proxy

```
./eaas-proxy hdl://11270/… 1433
```

Local network socket (TCP)

Connect contemporary machines / clients / software.

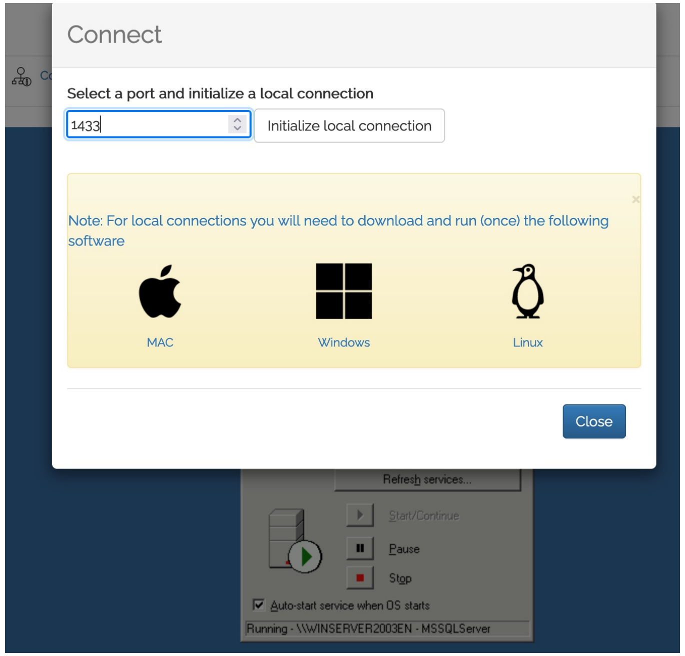
Can be started automated / from command line.

Is able to startup a emulated network (session management).

More *emulation* options here on the protocol level.

Networked Integration / Local Access

Get connection individualized connections details (e.g. as URL)



Connect

Select a port and initialize a local connection

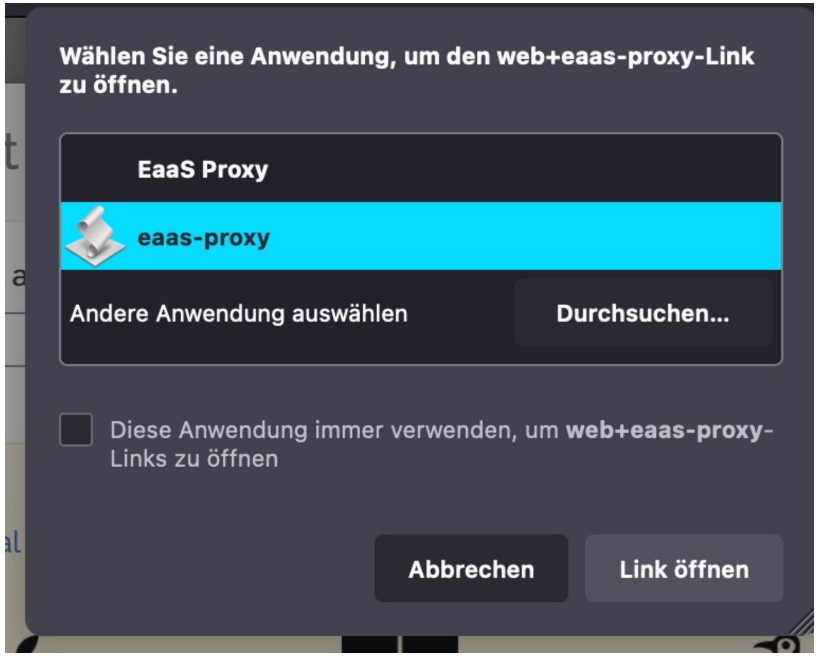
1433

Note: For local connections you will need to download and run (once) the following software

☑ Auto-start service when OS starts

Running - \\WINSERVER2003EN - MSSQLServer

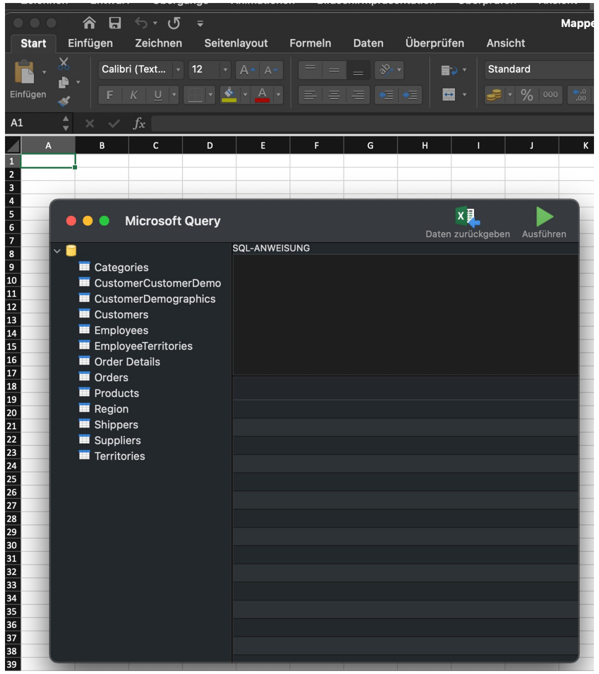
Open local connection



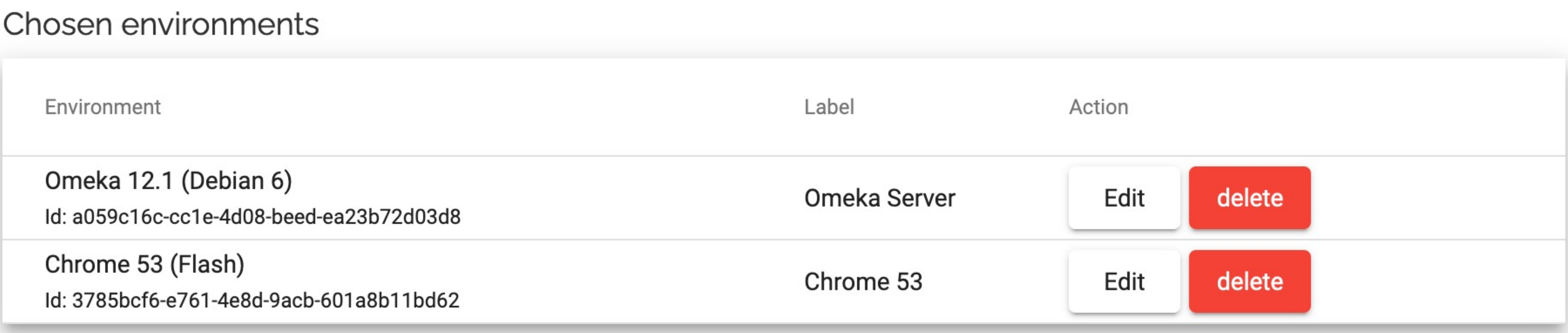
Wählen Sie eine Anwendung, um den web+eaas-proxy-Link zu öffnen.

□ Diese Anwendung immer verwenden, um **web+eaas-proxy**-Links zu öffnen

Example: connect Excel via ODBC data source



- Categories
- CustomerCustomerDemo
- CustomerDemographics
- Customers
- Employees
- EmployeeTerritories
- Order Details
- Orders
- Products
- Region
- Shippers
- Suppliers
- Territories

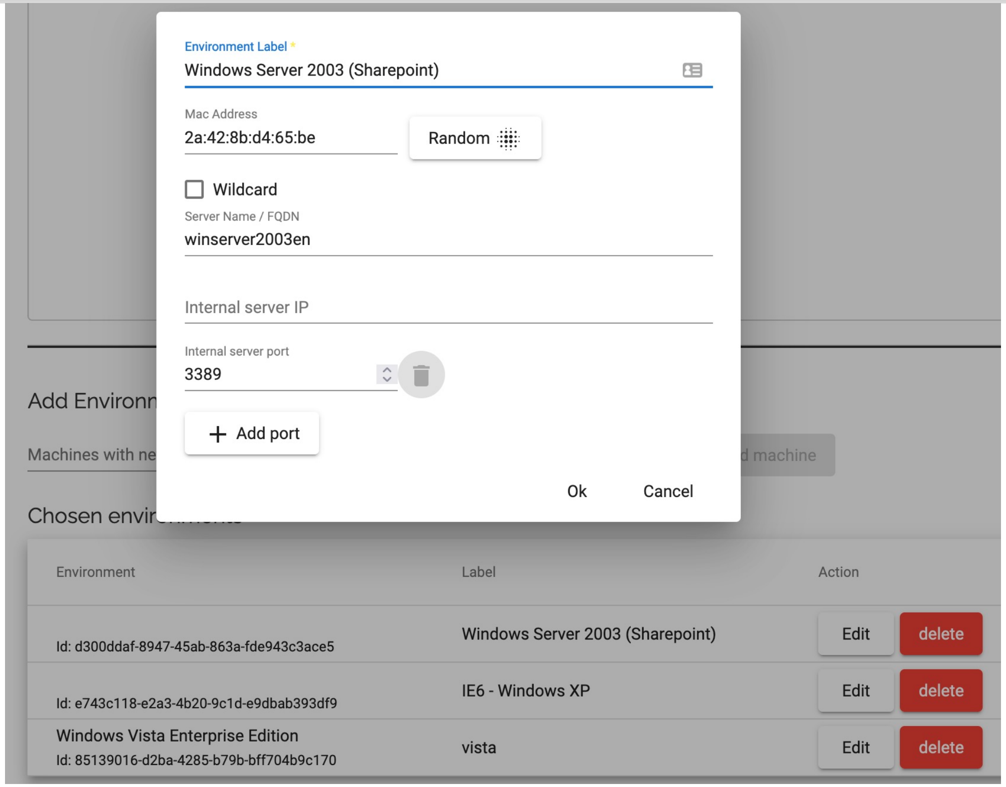


Chosen environments

Environment	Label	Action
Omeka 12.1 (Debian 6) Id: a059c16c-cc1e-4d08-beed-ea23b72d03d8	Omeka Server	Edit delete
Chrome 53 (Flash) Id: 3785bcf6-e761-4e8d-9acb-601a8b11bd62	Chrome 53	Edit delete

Further examples:

- Native client within the network
- SSH connections
- RDP connection
- Web frontends e.g. phpmyadmin
- ERM/CMS support e.g. Sharepoint
- Flash enabled remote browser



Environment Label *: Windows Server 2003 (Sharepoint)

Mac Address: 2a:42:8b:d4:65:be

□ Wildcard

Server Name / FQDN: winserver2003en

Internal server IP

Internal server port: 3389

Chosen envir

Environment	Label	Action
Id: d300ddaf-8947-45ab-863a-fde943c3ace5	Windows Server 2003 (Sharepoint)	Edit delete
Id: e743c118-e2a3-4b20-9c1d-e9dbab393df9	IE6 - Windows XP	Edit delete
Windows Vista Enterprise Edition Id: 85139016-d2ba-4285-b79b-bff704b9c170	vista	Edit delete

Summary

- Emulation is a software focused strategy
  - Implementation depends on complexity of the setup and anticipated risk profile
    - Relies on the future availability of a hardware emulator
  - Scales with availability of pre-configured stacks
    - Plenty of options for cooperation and automation

- Emulation is a tool to offer a prolonged (endless) sunset phase of obsolete services
  - Provide continuous access – adapt front-end access over time
  - Scales with the number of (cold) instances

- Operating obsolete software systems remains a huge non-technical problem
  - Some technical solutions are possible