TIME SERIES SERVICE INFRASTRUCTURE AND ANOMALY DETECTION USE CASE

Big Data All-Hands Meeting 2017

Jan Frenzel

Dr. René Jäkel

Dr. Christoph Lehmann

Yedhu Sastri

www.scads.de

ScaDS DRESDEN LEIPZIG

AGENDA

- ScaDS introduction
- Motivation and challenges
- Our solution
- Implementation
- Use case: cooling system analysis
- Conclusion and future work

ScaDS
DRESDEN LEIPZIG

SCADS INTRODUCTION

ScaDS DRESDEN LEIPZIG

Competence Center for Scalable Data Services and Solutions

- One of two BMBF-funded competence centers for Big Data
- Big Data research project
  - Various research domains
  - Infrastructures for Big Data

Center for Information Services & High Performance Computing

- HPC provider for TU Dresden and Saxony
- Service provider for TU Dresden and others

STRUCTURE OF THE COMPETENCE CENTER

- Life Sciences
- Material Sciences
- Environmental and Traffic Sciences
- Digital Humanities
- Business Data

Service Center

Big Data Life Cycle Management und Workflows

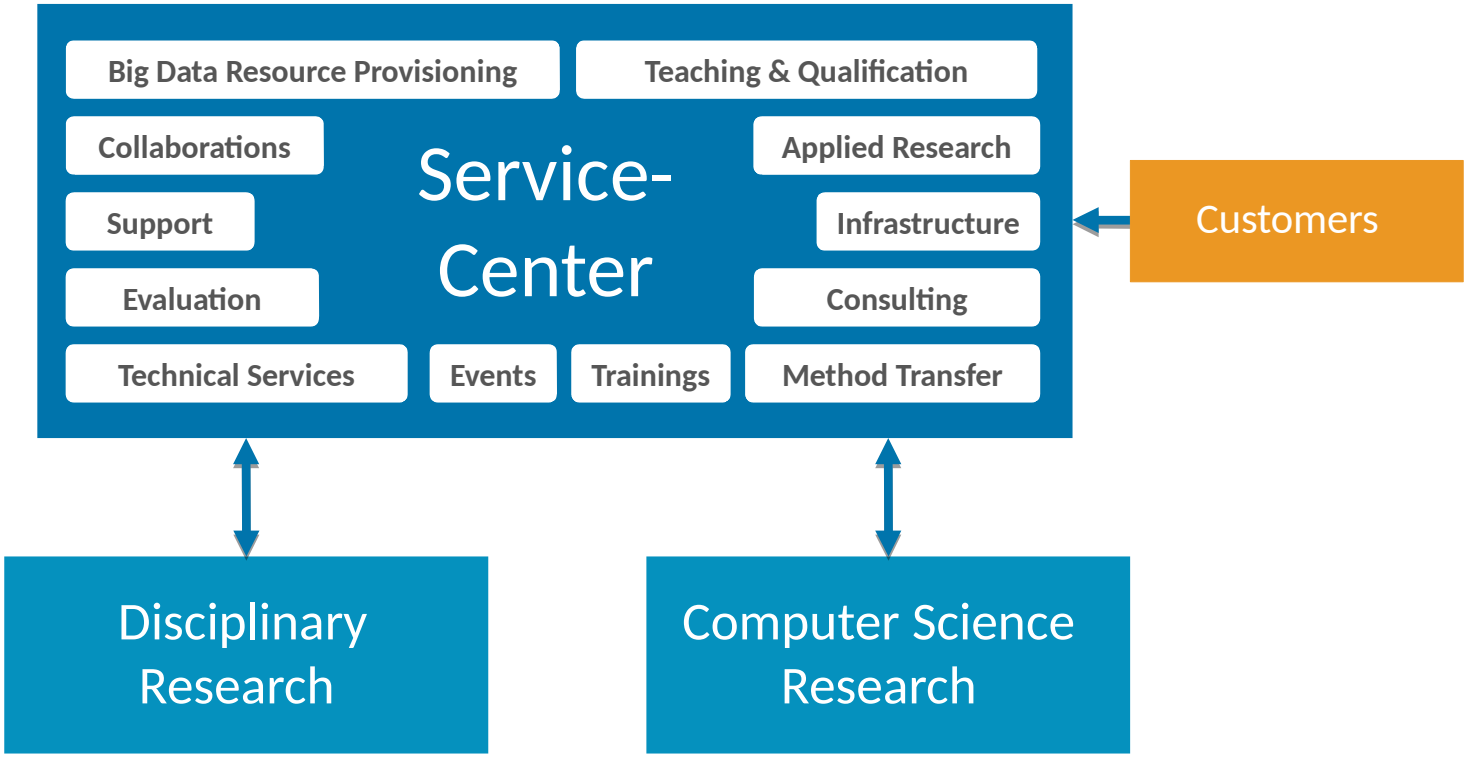
Data Quality/
Data Integration

Knowledge Extraction

Visual Analysis

Efficient Big Data Architectures

STRUCTURE OF THE COMPETENCE CENTER



ScaDS
DRESDEN LEIPZIG

MOTIVATION AND CHALLENGES

MOTIVATION AND CHALLENGES

- Data is everywhere!
- Data importance
  - IoT
  - Industry 4.0
  - Home automation
  - Environmental sensors
  - Software monitoring
  - …

MOTIVATION AND CHALLENGES

Common tasks:

- Track change of data
- Compare data changes
- Find patterns
- Find anomalies, outliers
- Predict future values

Aim: Set up service/infrastructure to support solving related tasks!

MOTIVATION AND CHALLENGES

Basic requirements/challenges for our service/infrastructure:

- Persistent data (storage, reference, data safety, data security)
- Description of data (metadata)
- Easy access to data (collaboration, permissions, API, frontend)

ScaDS
DRESDEN LEIPZIG

OUR SOLUTION

OUR SOLUTION: WEB SERVICE INFRASTRUCTURE

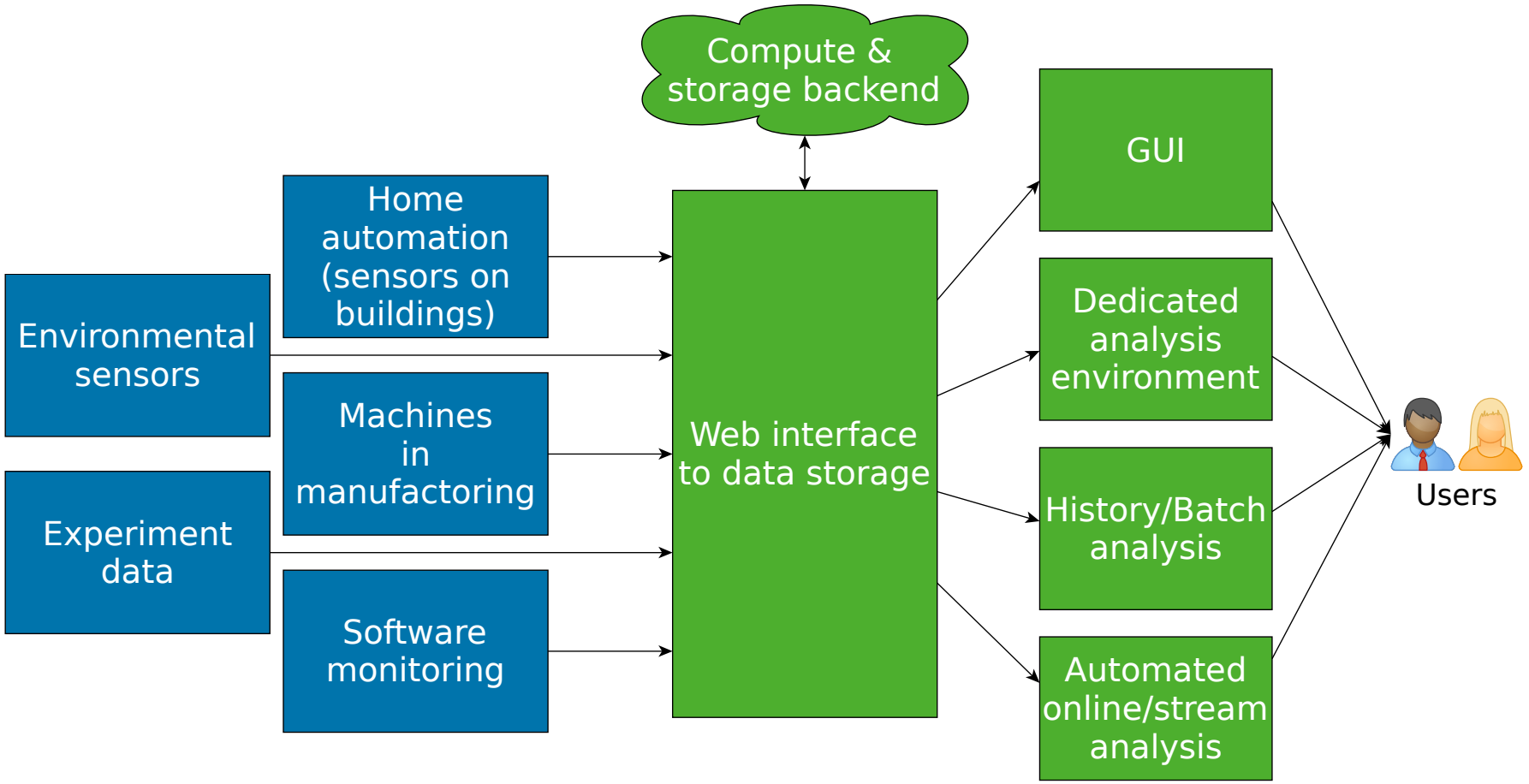
Provide a web service for storage and analysis of time series which can be used to

- reference,
- annotate,
- query, and
- collaboratively analyze

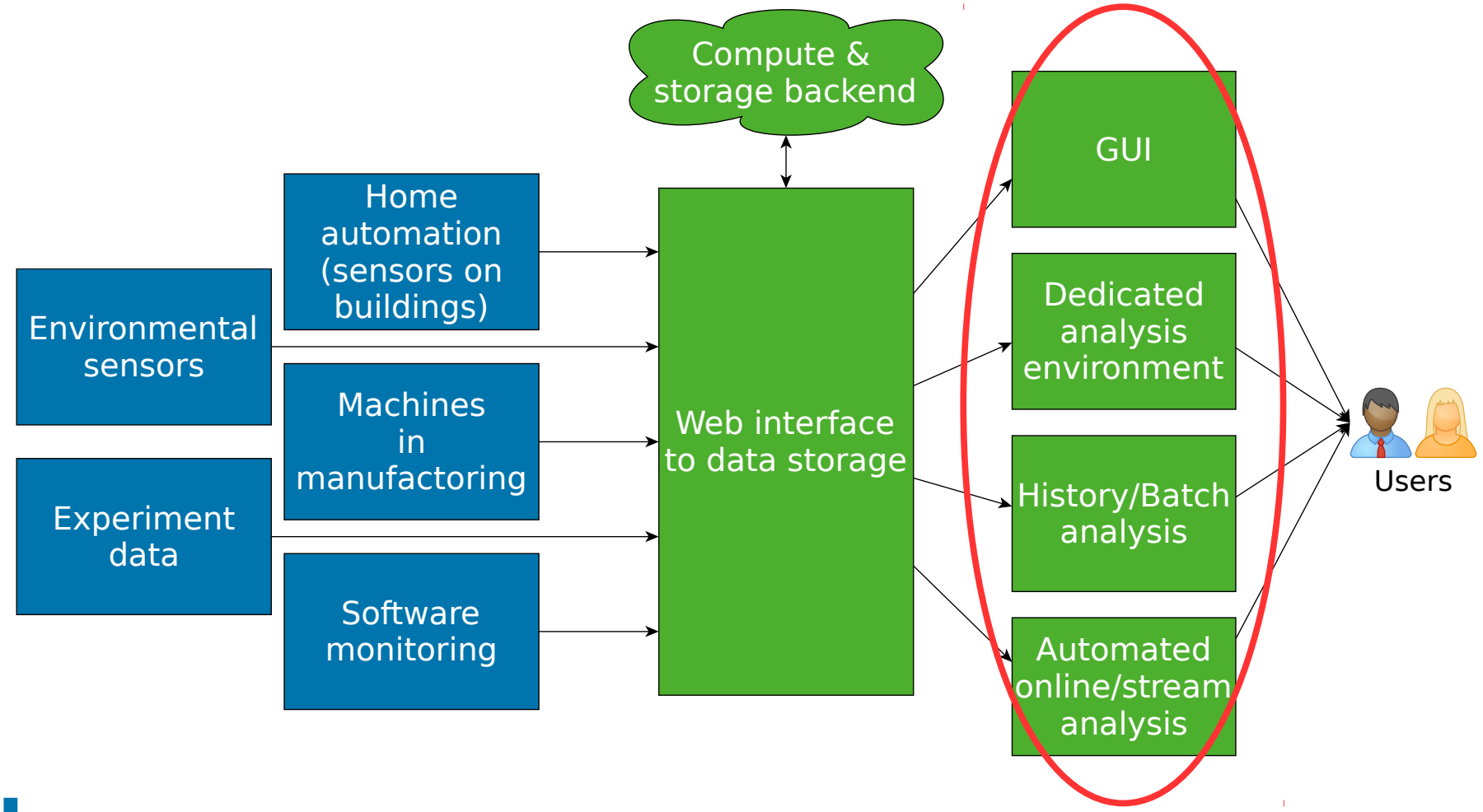
time series.

This matches the requirements.

OUR SOLUTION: WEB SERVICE ARCHITECTURE



OUR SOLUTION: WEB SERVICE ARCHITECTURE

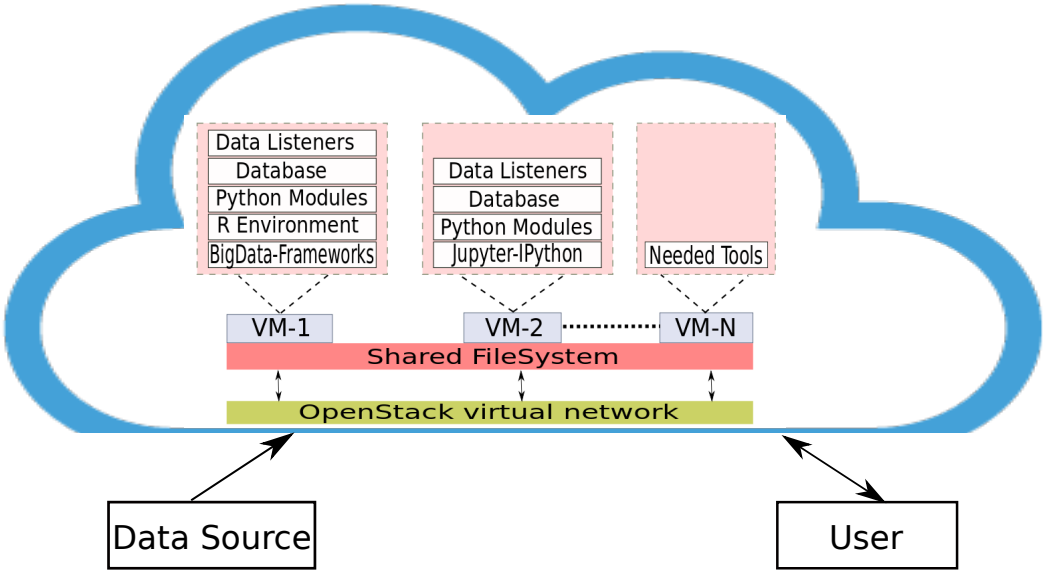


ScaDS
DRESDEN LEIPZIG

IMPLEMENTATION

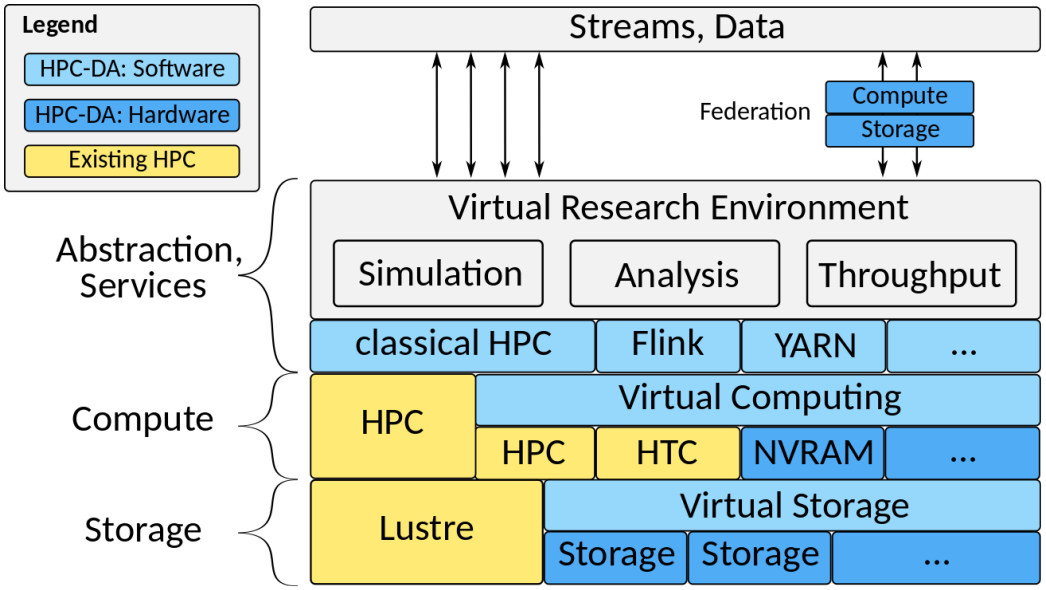
IMPLEMENTATION

- Use case was run in private cloud
  - Possibility to create particular analysis environment
  - Data transfer to user environment
  - Flexibility of software tools

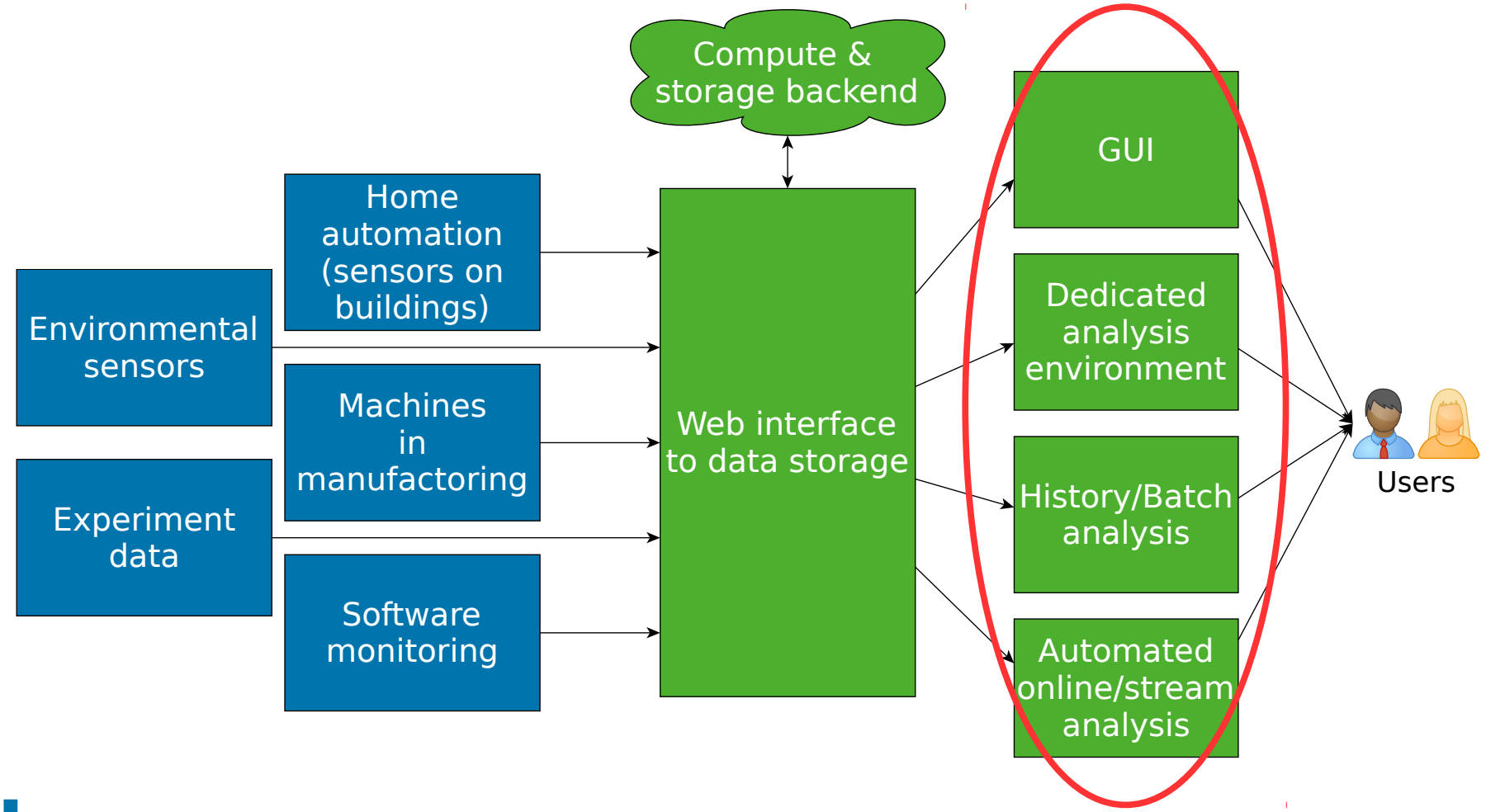


IMPLEMENTATION

- HPC environment also possible
  - For beginners: not as easy as cloud
  - Module-based software selection
  - Made for compute-intensive tasks



OUR SOLUTION: WEB SERVICE ARCHITECTURE



ScaDS
DRESDEN LEIPZIG

USE CASE:
COOLING SYSTEM ANALYSIS

USE CASE: COOLING SYSTEM ANALYSIS

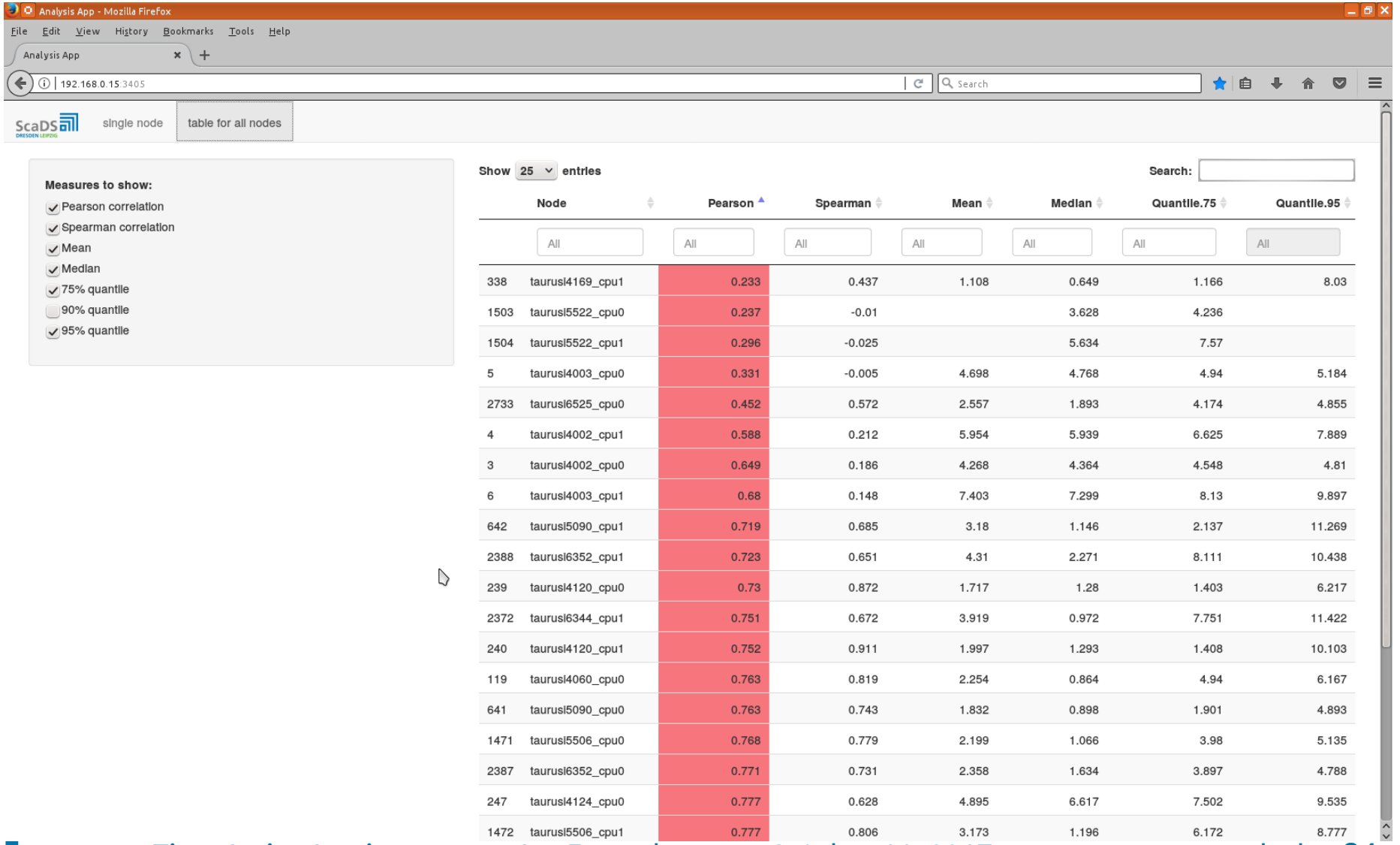
- Objective: Identify problems of the cooling system of an HPC machine.
- What is a “normal” interplay of temperature and power?
- Identification of anomalies
- Sensor data of approx. 3000 CPUs for power consumption [W] and temperature [°C], i. e. 6000 single data files/streams
- Time series data (intervals of 2 seconds and 10 seconds)

USE CASE: COOLING SYSTEM ANALYSIS

- Tasks:
  - Data preparation/aggregation for all CPUs (checking for errors/duplicates, merging different timestamps, etc.)
  - Data analysis for anomaly detection (construction of appropriate measures that characterize the behaviour of the cooling system and indicate anomalies)
  - Data visualization (user-friendly, platform independent)
- All tasks should be done with both, historic and streaming data
- Most important results:
  - Using correlation for indication of anomalies
  - Cluster analysis for pattern recognition of cooling behaviour

USE CASE: COOLING SYSTEM ANALYSIS

USE CASE: COOLING SYSTEM ANALYSIS



	Node	Pearson	Spearman	Mean	Median	Quantile.75	Quantile.95
338	taurusl4169_cpu1	0.233	0.437	1.108	0.649	1.166	8.03
1503	taurusl5522_cpu0	0.237	-0.01		3.628	4.236	
1504	taurusl5522_cpu1	0.296	-0.025		5.634	7.57	
5	taurusl4003_cpu0	0.331	-0.005	4.698	4.768	4.94	5.184
2733	taurusl6525_cpu0	0.452	0.572	2.557	1.893	4.174	4.855
4	taurusl4002_cpu1	0.588	0.212	5.954	5.939	6.625	7.889
3	taurusl4002_cpu0	0.649	0.186	4.268	4.364	4.548	4.81
6	taurusl4003_cpu1	0.68	0.148	7.403	7.299	8.13	9.897
642	taurusl5090_cpu1	0.719	0.685	3.18	1.146	2.137	11.269
2388	taurusl6352_cpu1	0.723	0.651	4.31	2.271	8.111	10.438
239	taurusl4120_cpu0	0.73	0.872	1.717	1.28	1.403	6.217
2372	taurusl6344_cpu1	0.751	0.672	3.919	0.972	7.751	11.422
240	taurusl4120_cpu1	0.752	0.911	1.997	1.293	1.408	10.103
119	taurusl4060_cpu0	0.763	0.819	2.254	0.864	4.94	6.167
641	taurusl5090_cpu0	0.763	0.743	1.832	0.898	1.901	4.893
1471	taurusl5506_cpu0	0.768	0.779	2.199	1.066	3.98	5.135
2387	taurusl6352_cpu0	0.771	0.731	2.358	1.634	3.897	4.788
247	taurusl4124_cpu0	0.777	0.628	4.895	6.617	7.502	9.535
1472	taurusl5506_cpu1	0.777	0.806	3.173	1.196	6.172	8.777

ScaDS
DRESDEN LEIPZIG

CONCLUSION AND FUTURE WORK

CONCLUSION

- Virtual cloud infrastructure is
  - Time-saving
  - Cost-efficient (i. e. setup and maintenance)
  - Highly flexible
  - Platform-independent (web application, notebooks)
  - User-specific, user-friendly
- HPC resources available for compute-intensive tasks
- Feasibility proven: Use-case

FUTURE WORK

- Future work:
  - User + permission management
  - Resource provisioning (templates)
  - High availability operation mode
  - Publicly available web service + visualization
  - Workflow management

ScaDS DRESDEN LEIPZIG

THANK YOUR FOR YOUR ATTENTION!

CONTACT

Jan Frenzel
Technische Universität Dresden
Center for Information Services and High Performance Computing (ZIH)
E-mail: jan . frenzel @ tu – dresden . de

GEFÖRDERT VOM

Bundesministerium
für Bildung
und Forschung

DLR Projektträger

ScaDS DRESDEN LEIPZIG

CONTACT

SCIENTIFIC COORDINATOR

Prof. Dr. Wolfgang E. Nagel

Technische Universität Dresden
Zentrum für Informationsdienste und Hochleistungsrechnen
01062 Dresden

Telefon: +49 351 463-35450
E-Mail: wolfgang.nagel@tu-dresden.de

DEPUTY SCIENTIFIC COORDINATOR

Prof. Dr. Erhard Rahm

Universität Leipzig
Fakultät für Mathematik und Informatik
Augustusplatz 10
04109 Leipzig

Telefon: +49 341 97-32221
E-Mail: rahm@informatik.uni-leipzig.de

MANAGEMENT DIRECTOR

Dr. René Jäkel

Technische Universität Dresden
Zentrum für Informationsdienste und Hochleistungsrechnen
01062 Dresden

Telefon: +49 351 463-42331
E-Mail: rene.jaekel@tu-dresden.de