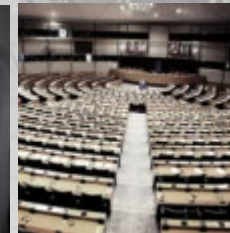


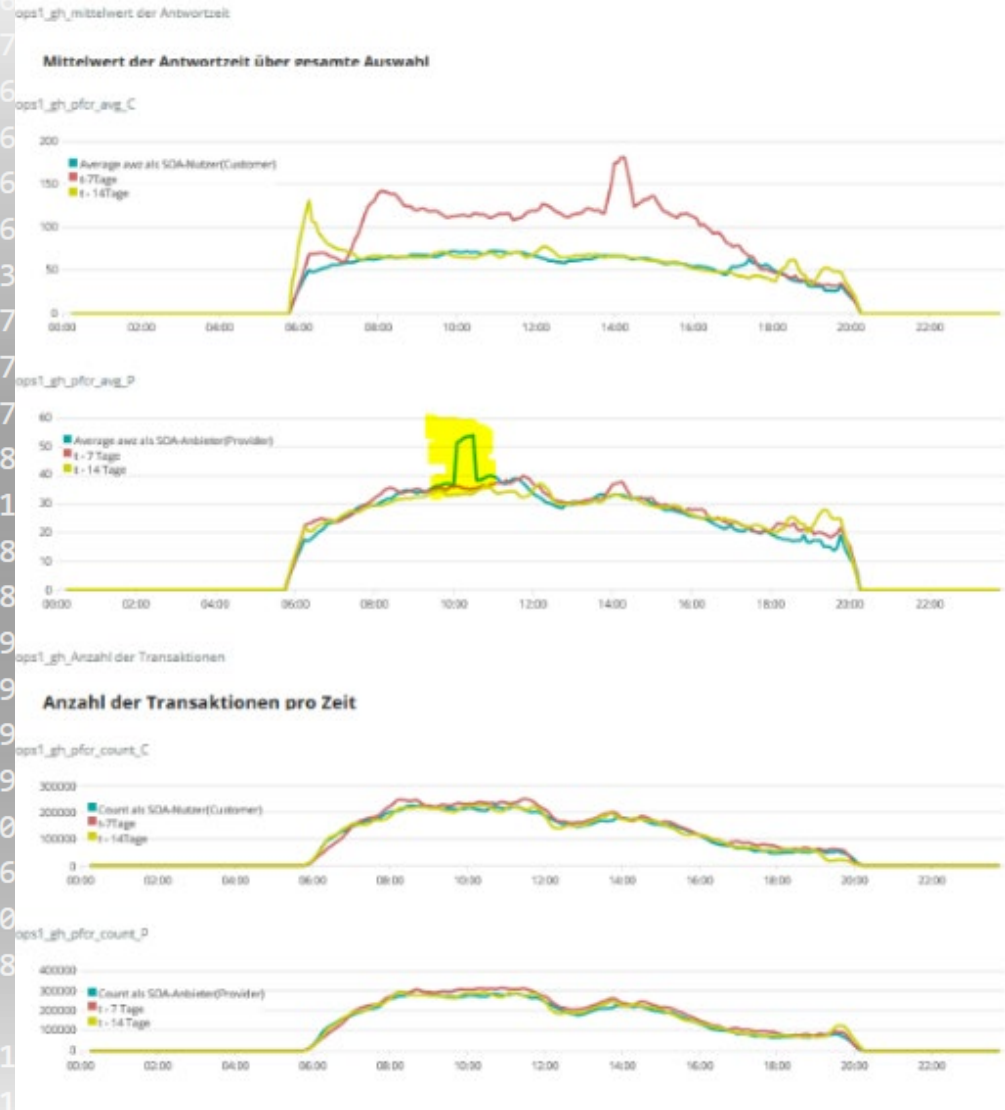
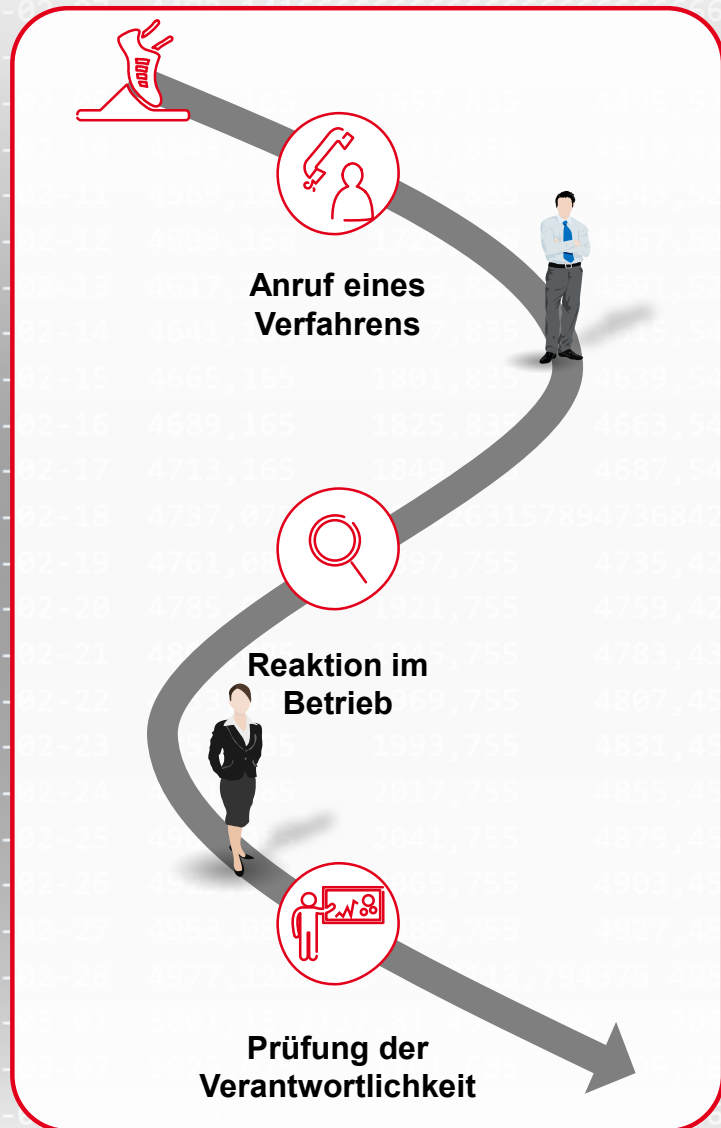
Eine Data Science-Architektur für die Analyse und Visualisierung von Rechenzentrumsdaten der Bundesagentur für Arbeit

The Architecture Gathering 2019

München, 17. Oktober 2019, Matthias Seßler, Eldar Sultanow, Frank Pelzel, Oliver Weiß



Alltag im Betrieb von IT-Infrastruktur: Die Prüfung von Systemausfällen ist zeitintensiv und nicht verlässlich vorhersagbar

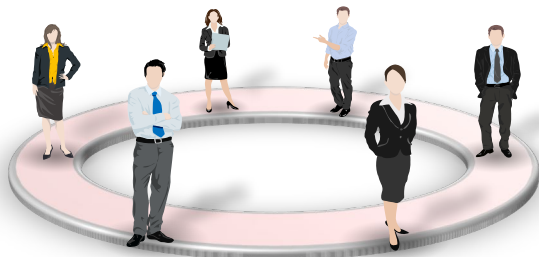


Die Infrastruktur eines der größten IT-Entwicklers und -Betreibers Deutschlands ist das Herzstück zum Betrieb der Kundenservices



Bundesagentur für Arbeit

- **Zentraler Dienstleister** am Arbeitsmarkt
- **Körperschaft des öffentlichen Rechts** mit **Selbstverwaltung**
- Ca. **95.000 Beschäftigte** im Bundesgebiet
- **BA** mit:
 - Zentrale
 - 10 Regionaldirektionen
 - >150 Agenturen für Arbeit
 - Ca. 600 Geschäftsstellen
 - 303 Jobcenter (gE)
 - 12 Bildungs- und Tagesstätten (BTS)
 - 7 besondere Dienststellen in
 - 1.600 Liegenschaften



Bundesagentur
für Arbeit



BA-Informationstechnik

Kurzprofil

- Hauptsitz: Nürnberg
- IT-Mitarbeiter: 2.000
- Vernetzte PC: 170.000
- Server: 10.000



Systemlandschaft

- Über 100 eigene IT-Verfahren
- Drei hochverfügbare zentrale Rechenzentren
- 17 regionale IT-Stützpunkte

Output (monatlich)

- E-Mail-Volumen: 49 Mio. E-Mails
- Überweisungen: 17 Mio. Überweisungen (8 Mrd. Euro)
- Postsendungen: 12 Mio. Sendungen
- Druckseiten: 58 Mio. Seiten

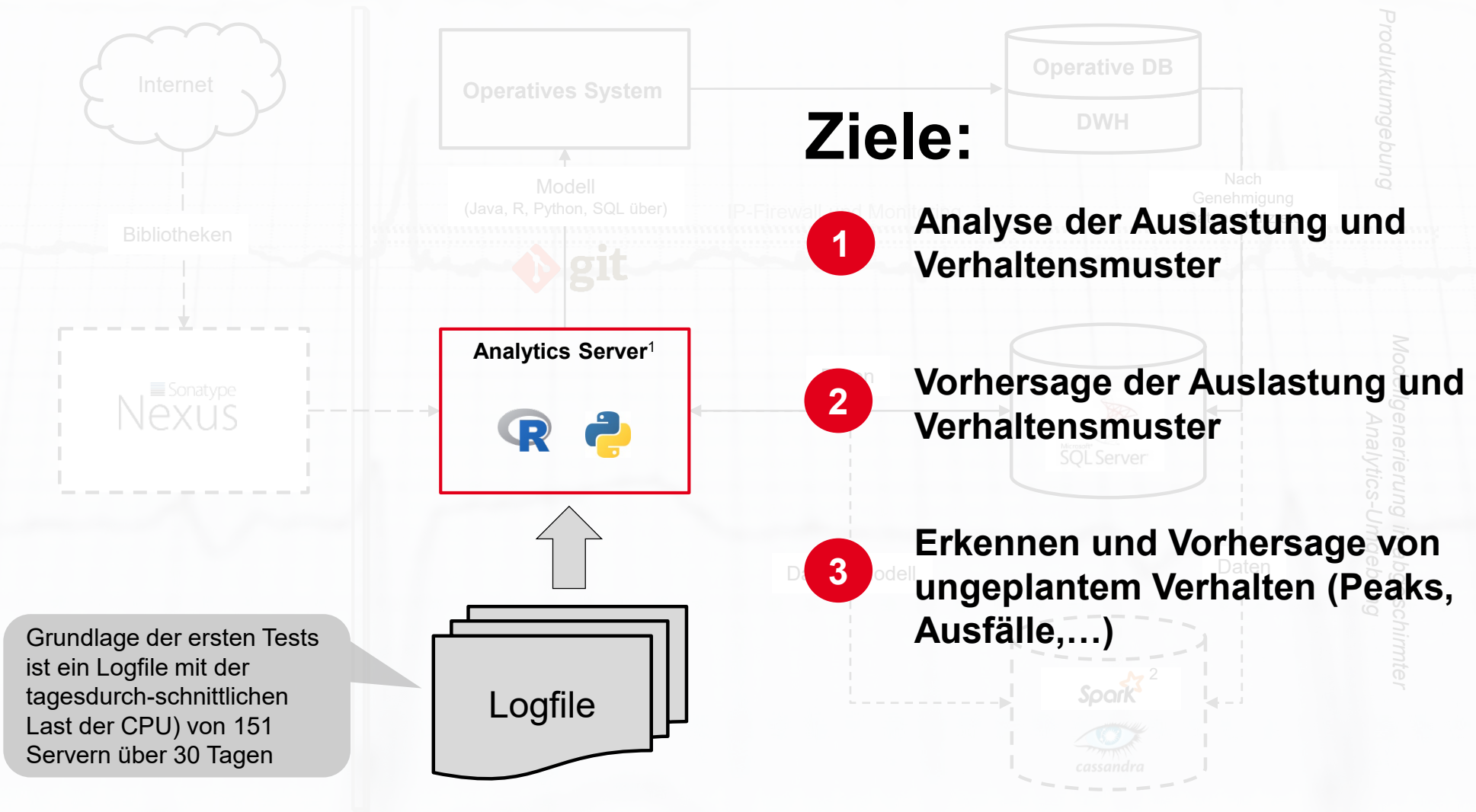


IT-Systemhaus

Agenda

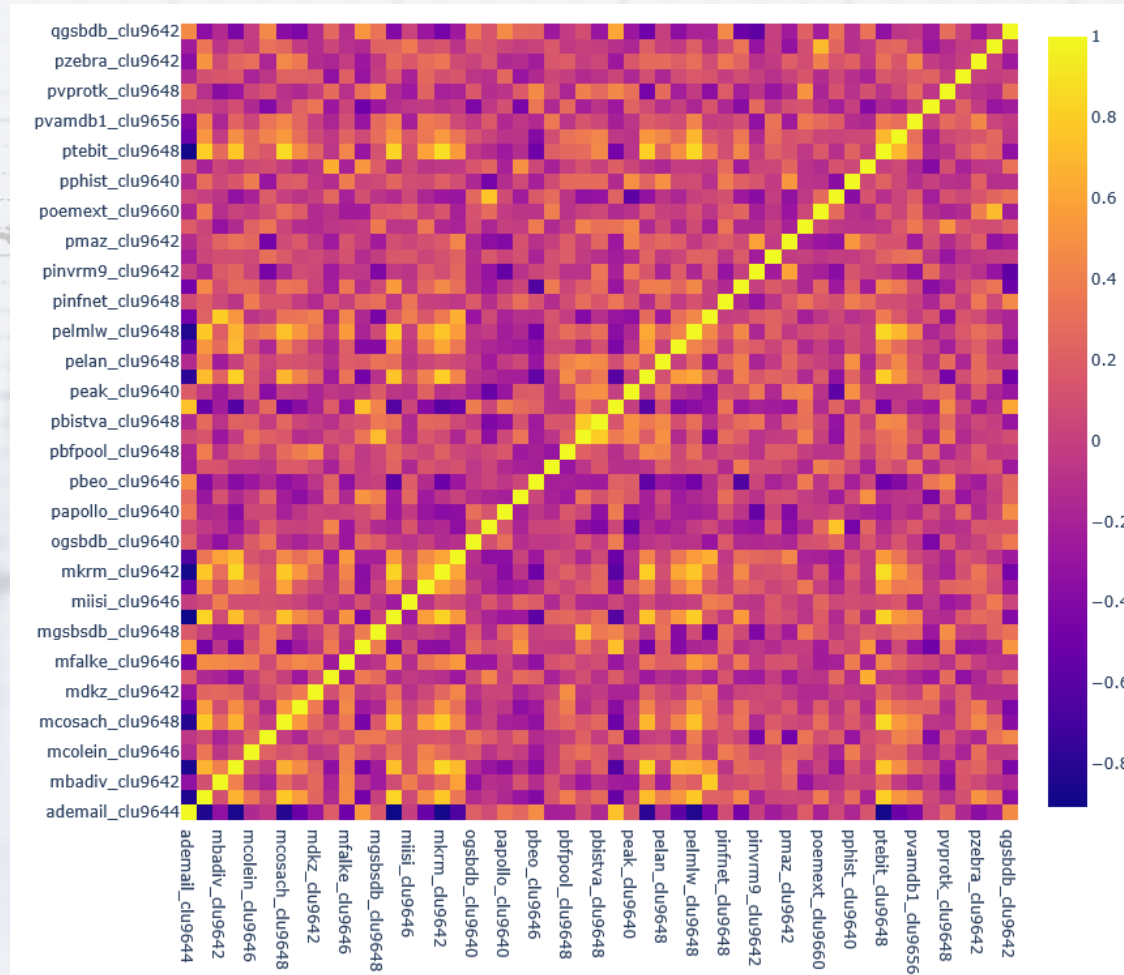
- **Zeitreihenanalyse von Rechenzentrums-Ressourcen**
- Prototypische Umsetzung einer Visualisierung
- Referenzarchitektur
- Live-Demo

Die Zeitreihenanalyse findet in einer bewährten Analytics-Umgebung statt



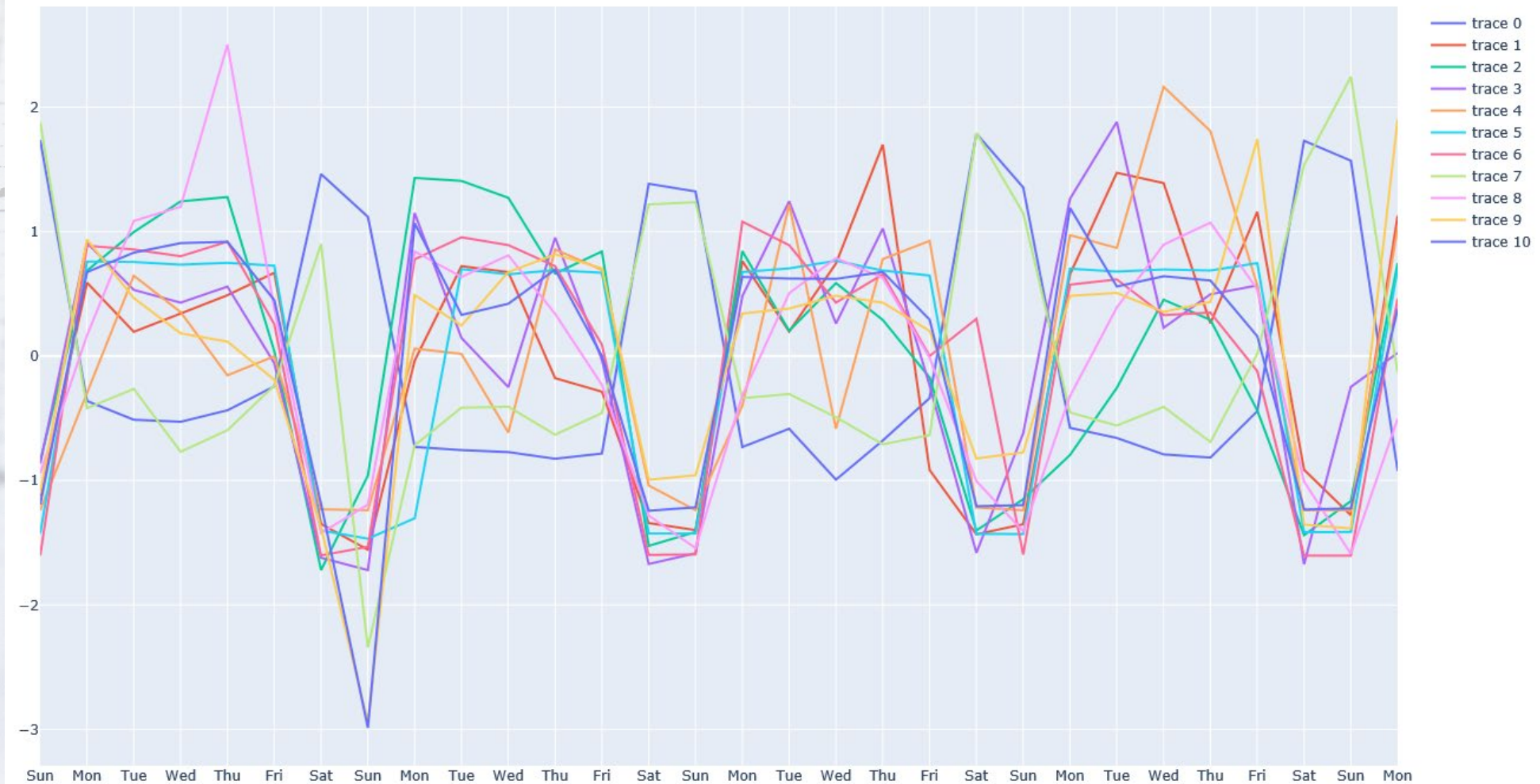
1. Ziel: Analyse der Auslastung und Verhaltensmuster von Infrastrukturen

Korrelationsmatrix des Verlaufs der CPU Auslastung für ausgewählte Server:



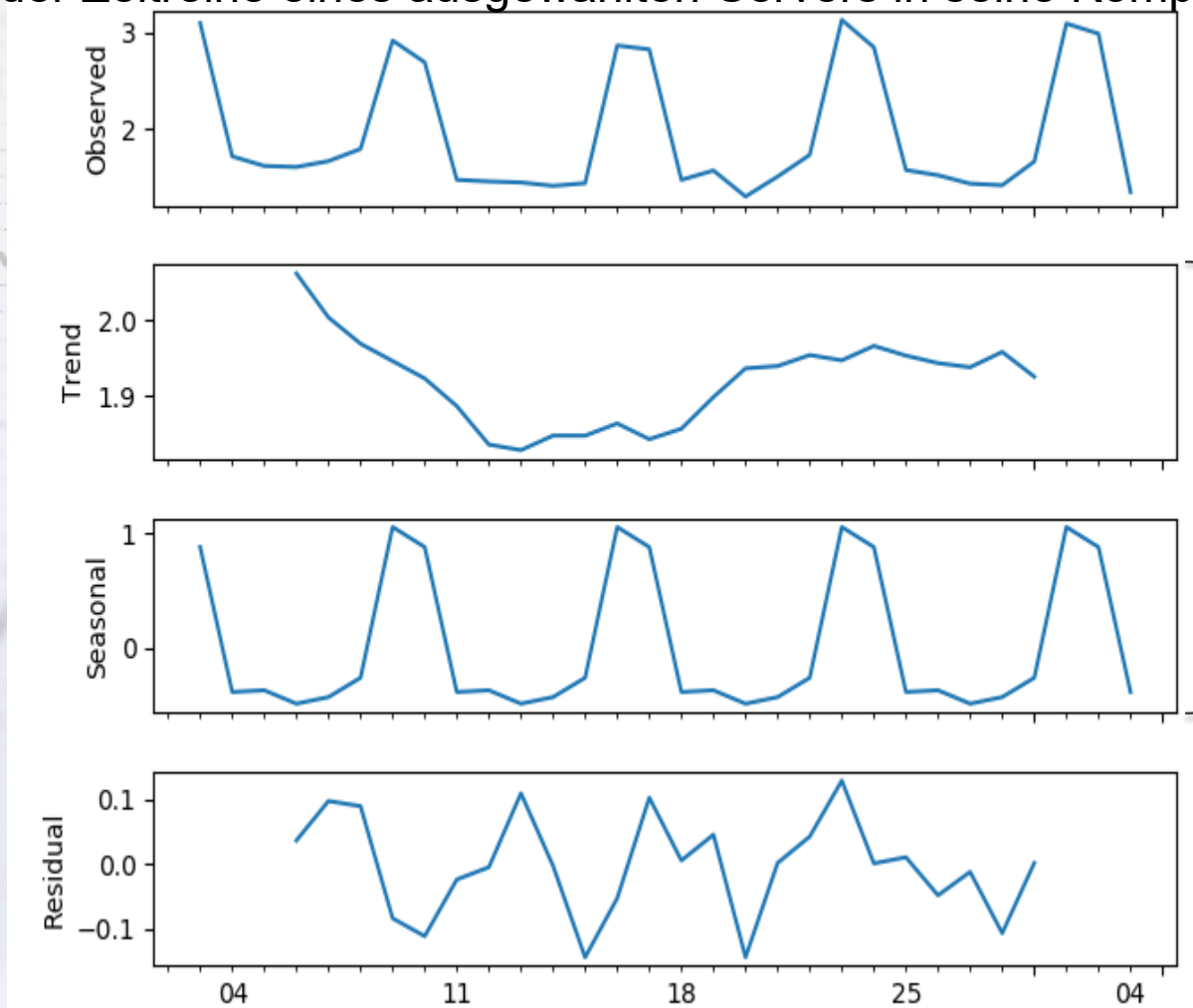
1. Ziel: Analyse der Auslastung und Verhaltensmuster von Infrastrukturen

Stark korrelierte, standardisierte Auslastungsverläufe mit ausgewähltem Server:



2. Ziel: Vorhersage der Auslastung und Verhaltensmuster von Infrastrukturen

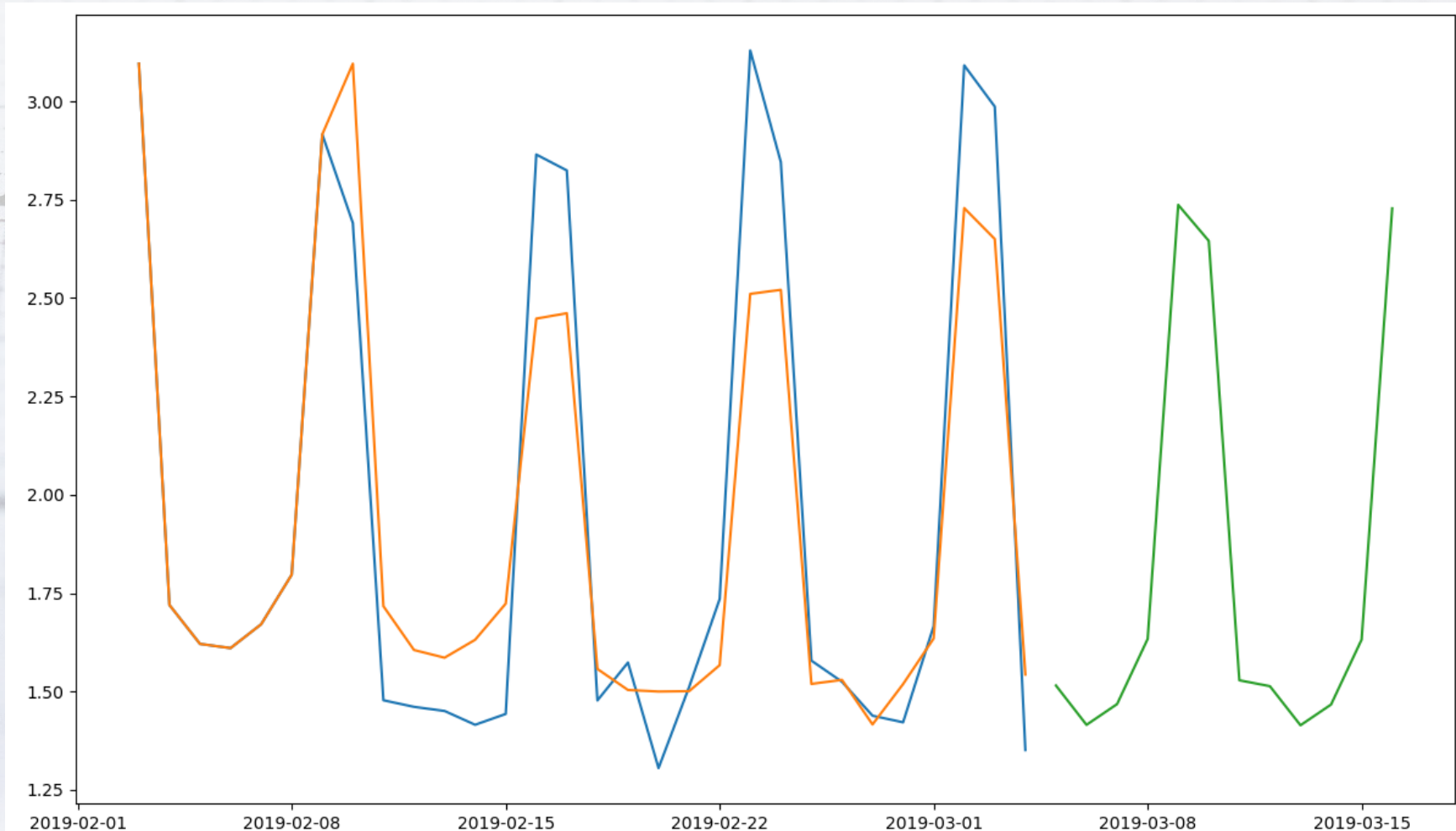
Zerlegung der Zeitreihe eines ausgewählten Servers in seine Komponenten:



Langzeit-
verhalten

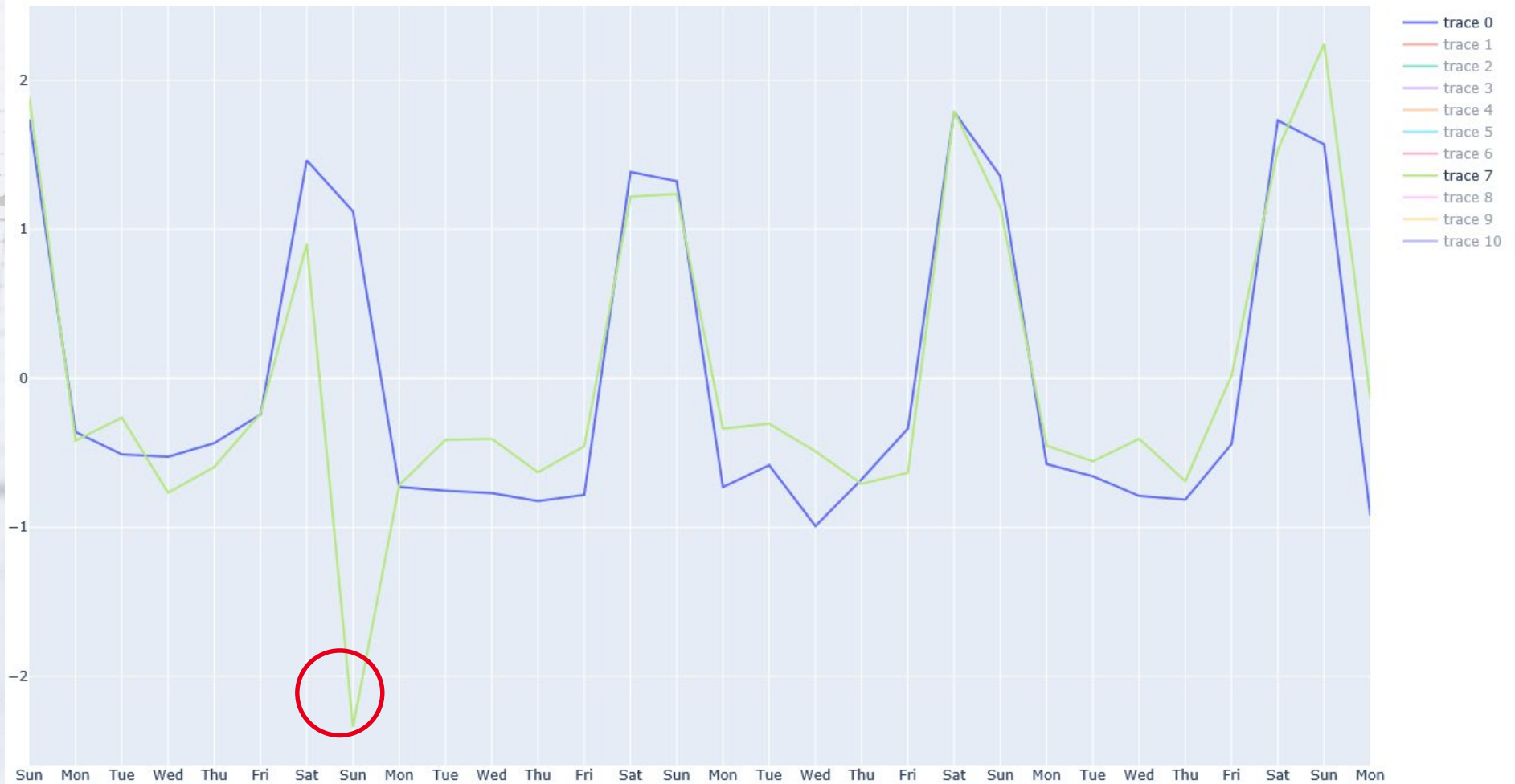
2. Ziel: Vorhersage der Auslastung und Verhaltensmuster von Infrastrukturen

Vorhersage der Auslastung:



3. Ziel: Erkennen und Vorhersage von ungeplantem Verhalten (Peaks, Ausfälle,...)

Zusätzliche Informationen können ungeplantes Verhalten sichtbar/vorhersagbar machen:

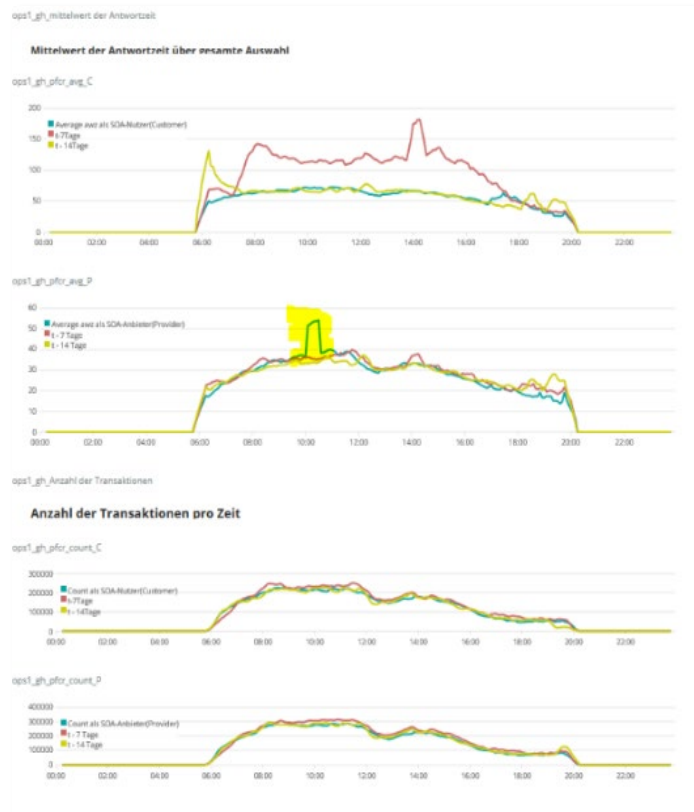


Agenda

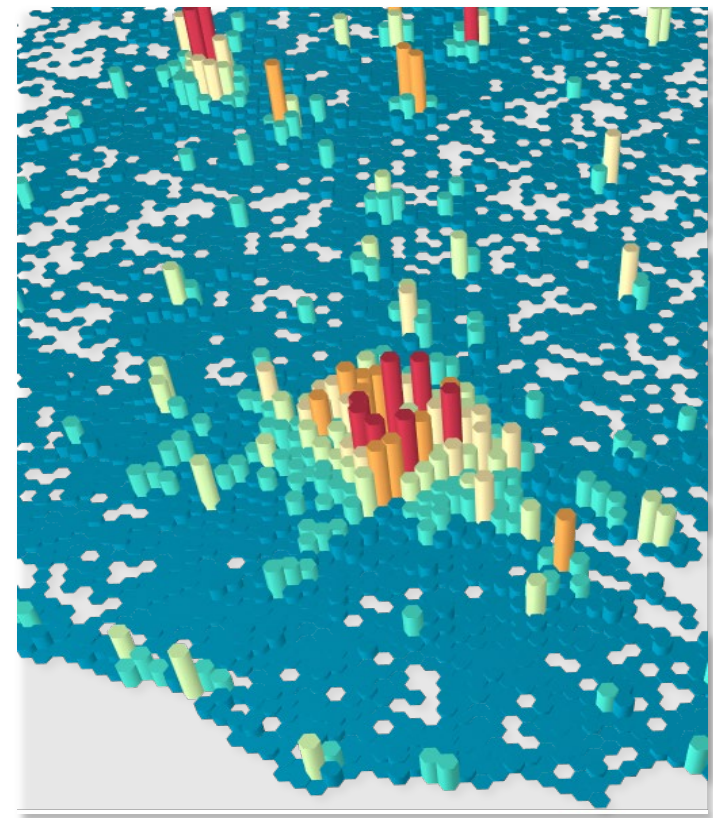
- Zeitreihenanalyse von Rechenzentrums-Ressourcen
- **Prototypische Umsetzung einer Visualisierung**
- Referenzarchitektur
- Live-Demo

Ziel des Prototyps ist ein schneller Überblick über den Betriebszustand eines Datacenters

Bisher: Informationen verteilt auf unzählige Dataplots



Prototyp: Aufbereitung der Daten in 3D, Nutzung Geo-/Stadtmetapher



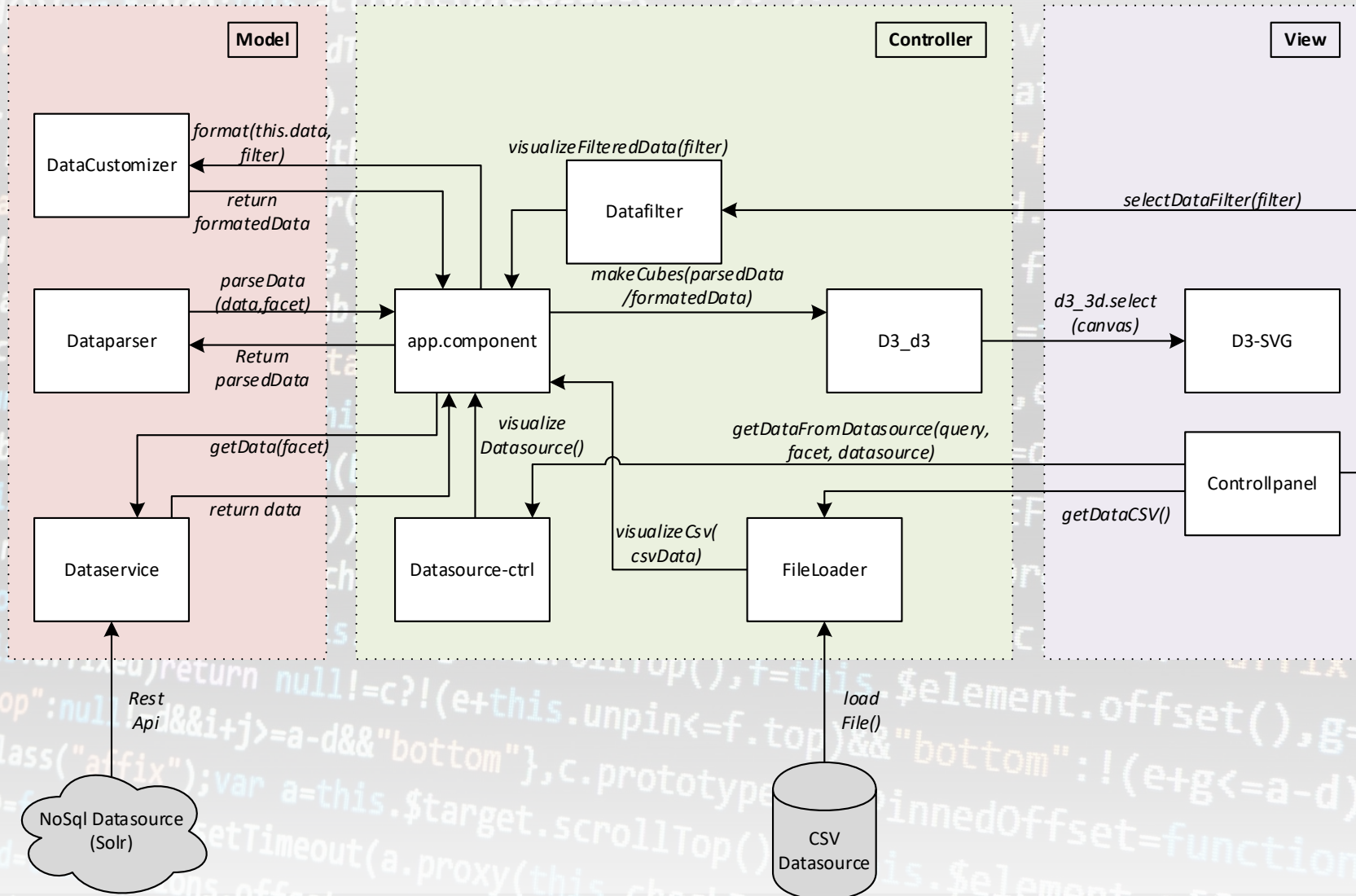
UBER, deck.gl examples, <https://deck.gl/#/examples/core-layers/hexagon-layer> [02.10.2019]

<https://pixabay.com/de/photos/programmierung-computer-umwelt-1857236/>

Die Visualisierung von Zeitreihen zu Rechenzentrums-Ressourcen bilden den Kern der Anforderungen

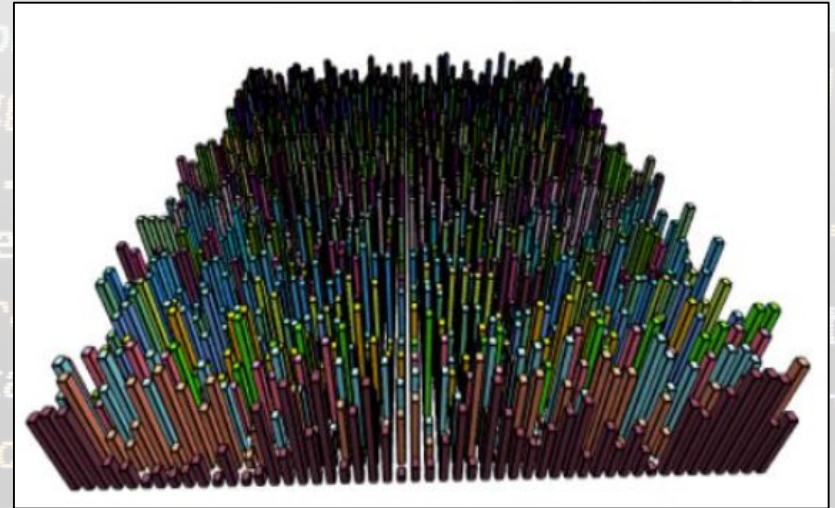
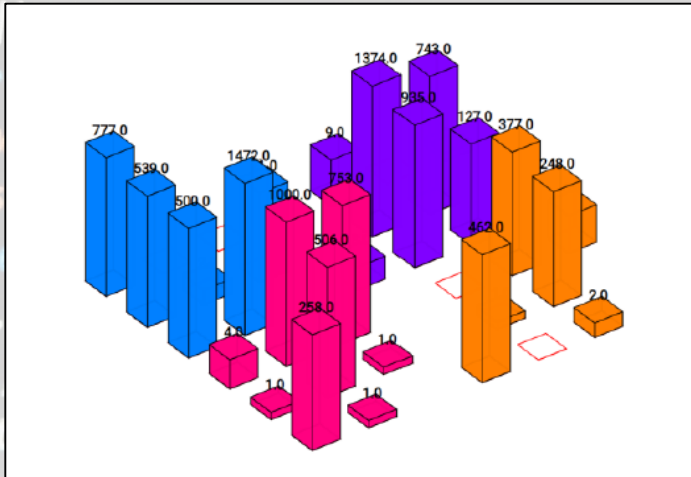
Use Case	Funktionale Anforderung	Qualitätsanforderung
Datenquelle ändern	Freie Wahl aus mehreren Quellen	Portabilität Datenanbindung
Visualisierung anpassen	Parameter wie z.B. Größe der Blöcke einstellbar	Portabilität Visualisierung, Performance Darstellung
Zeitpunkt wählen	Freie Wahl des Zeitpunkts, Veränderung über die Zeit sichtbar machen	Benutzbarkeit Oberflächen, Performance Darstellung
Filter anwenden	Einschränken der darzustellenden Informationen	Benutzbarkeit Oberflächen, Performance Darstellung

Ein erster Prototyp verwendet die Technologien Angular, Typescript, Material, Node.js, D3_3d.js und Solr



Die Nachteile des Prototyps sind schlechte Skalierbarkeit und fehlende Integrationsoptionen in Bestandsprodukte

Skalierbarkeit



Dashboard-Integration



Die Evolution der Architektur zur Erfüllung der Qualitätsanforderungen ist mit einigen Hindernissen versehen

Vega

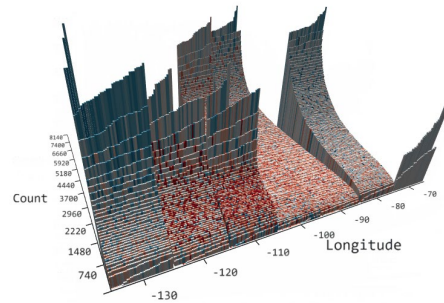


- 😊 Kibana Integration
- 😊 Modulare Datenquellen

- 😞 Keine z-Achse / kein 3D
- 😞 Zukünftige Unterstützung offen

UW Interactive Data Lab, Vega, <http://vega.github.io> [02.10.2019]

Sanddance



- 😊 Web.gl
- 😊 Integration +Standalone

- 😞 Kaum Doku vorhanden
- 😞 Zukünftige Unterstützung offen
- 😞 Baut auf sehr alter Vega Version auf

Microsoft, SandDance, <https://github.com/microsoft/SandDance> [02.10.2019]

Three.js



- 😊 Web.gl
- 😊 Integration +Standalone
- 😊 Viele Nutzer + Contrib.
- 😊 Ausführliche Doku

- 😞 Aufwendiger Umbau auf React.js

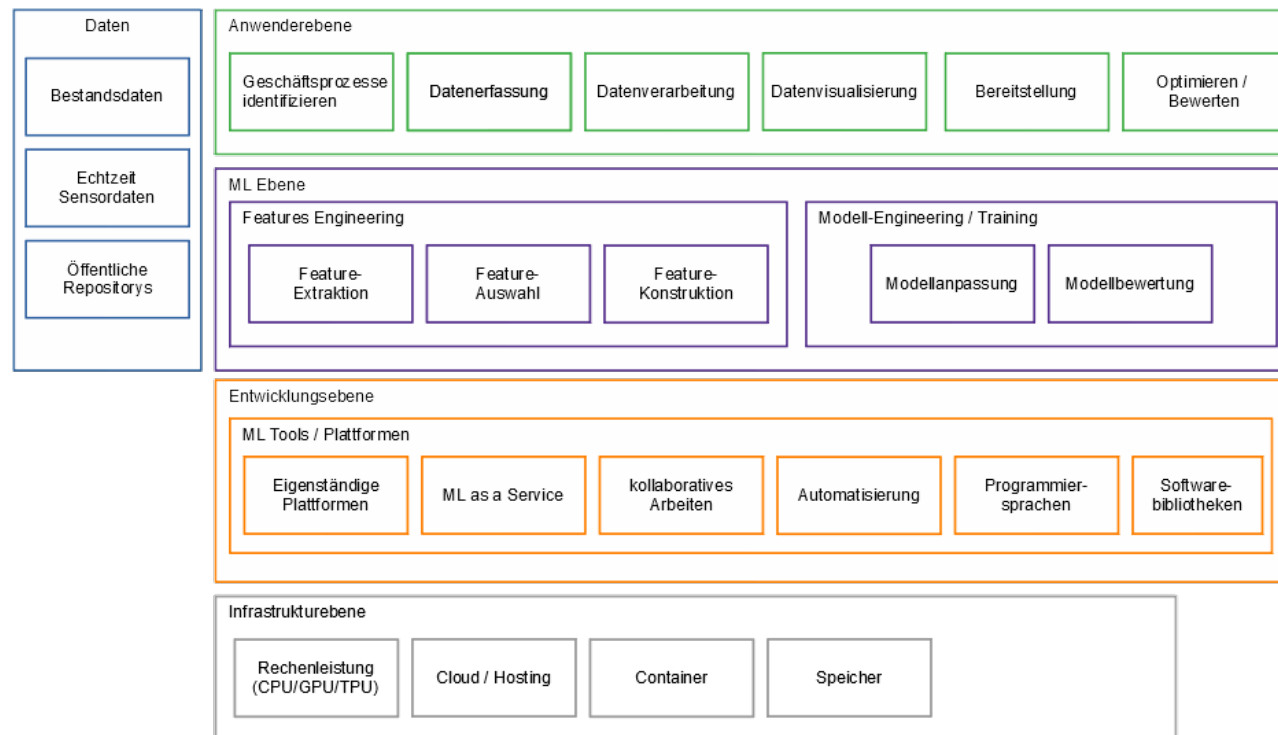
Ricardo Cabello, Three.js, <https://en.wikipedia.org/wiki/Three.js> [02.10.2019]

Agenda

- Zeitreihenanalyse von Rechenzentrums-Ressourcen
- Prototypische Umsetzung einer Visualisierung
- **Referenzarchitektur**
- Live-Demo

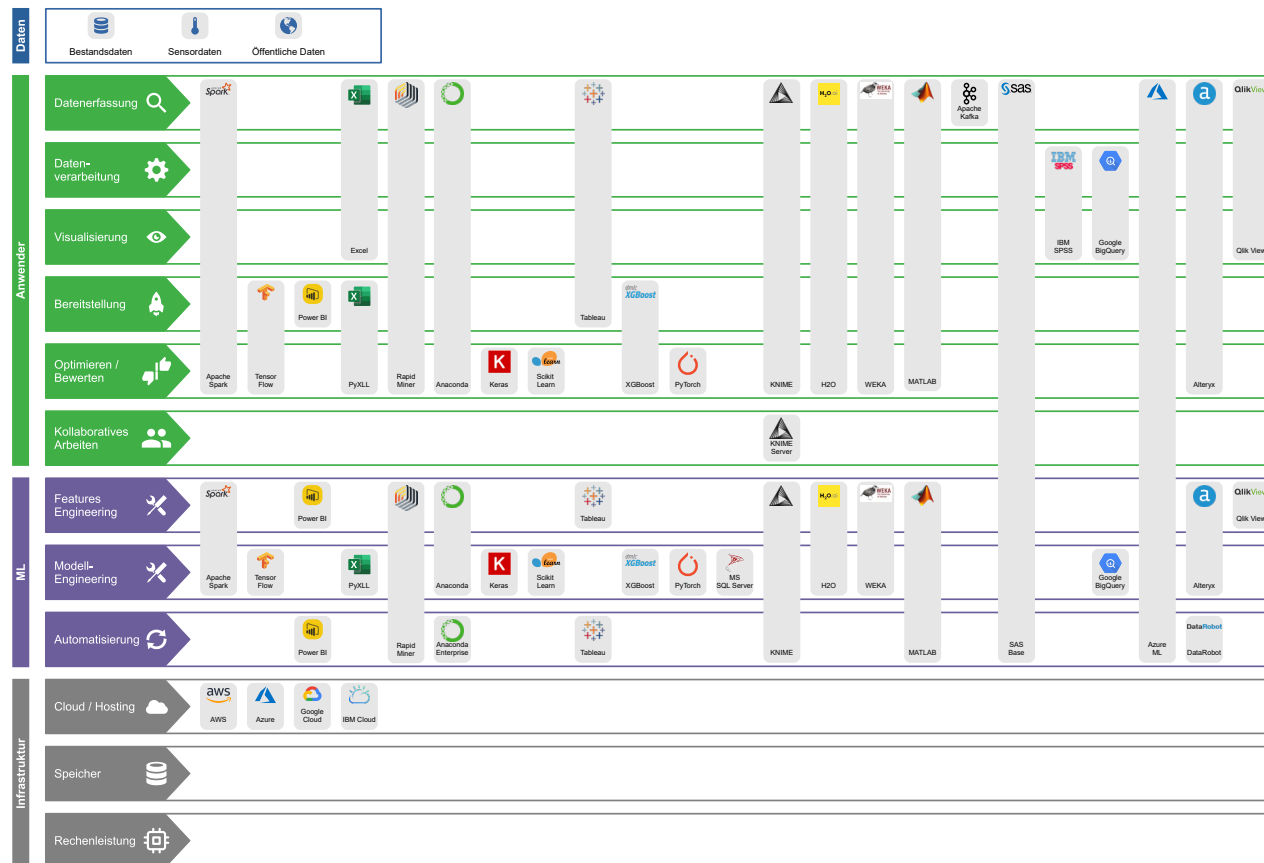
Es gibt eine unübersichtliche Vielzahl an Data Science Tools und Frameworks

Referenzarchitektur: Data Science Kategorien

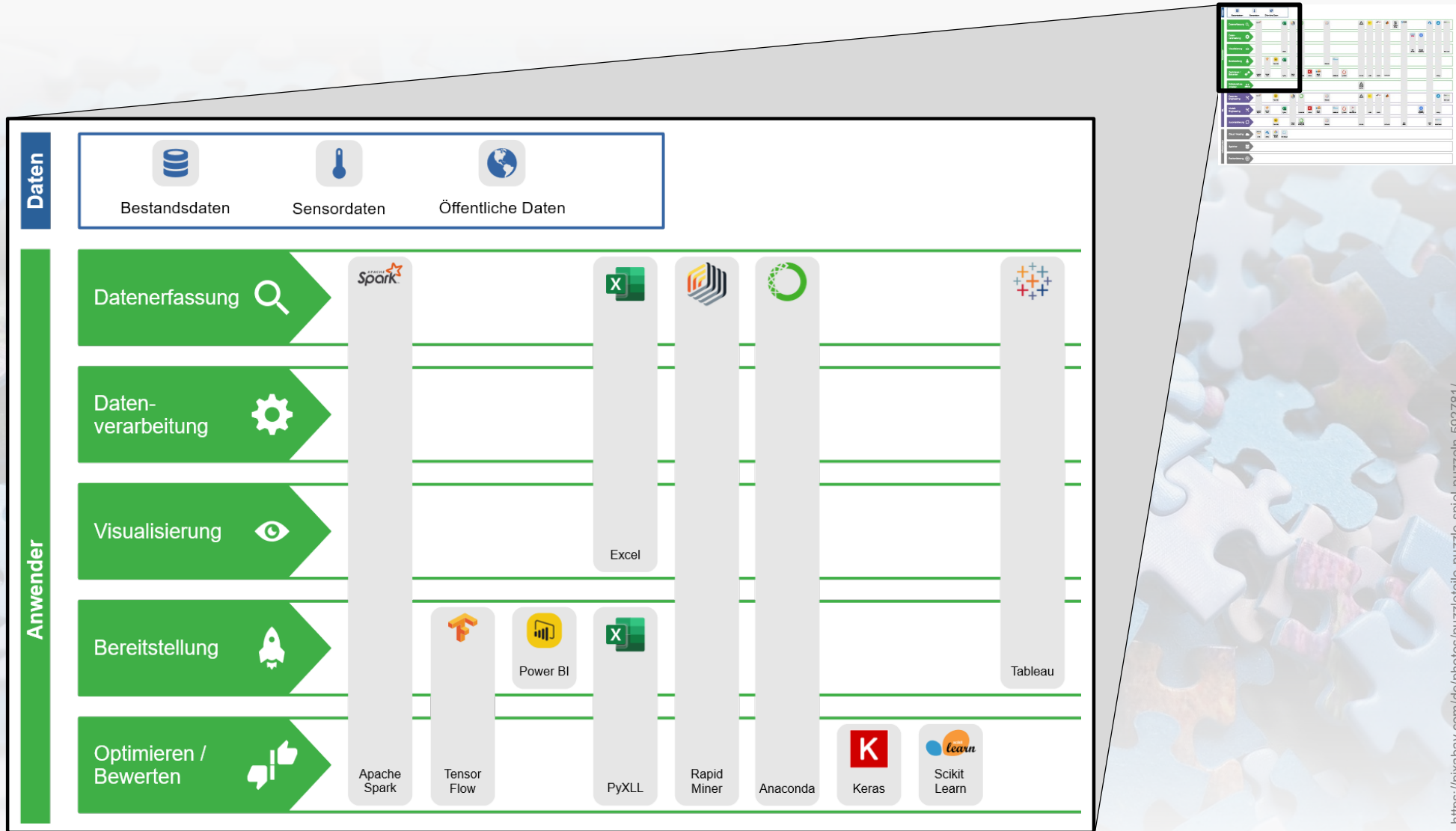


Es gibt eine unübersichtliche Vielzahl an Data Science Tools und Frameworks

Referenzarchitektur: Feingliederung



Exemplarischer Ausschnitt des Anwender-Layers



Agenda

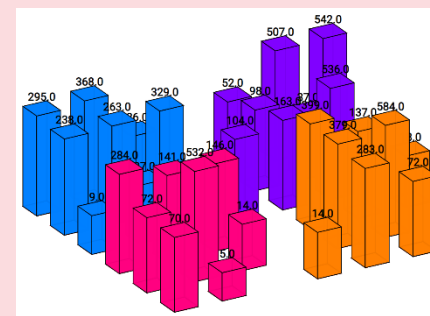
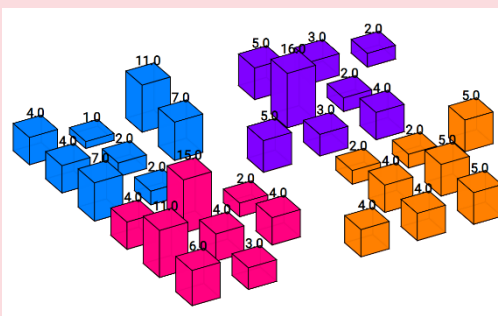
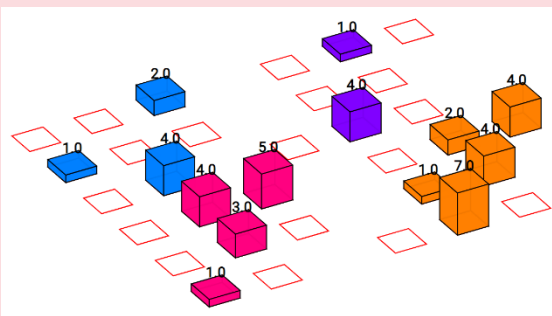
- Zeitreihenanalyse von Rechenzentrums-Ressourcen
- Prototypische Umsetzung einer Visualisierung
- Referenzarchitektur
- **Live-Demo**

Live Demo: Die 3D-Visualisierungsmethodik erlaubt die Ableitung übergreifender Erkenntnisse zum Zustand der IT-Infrastruktur



Vorführung der Visualisierung

- Die Visualisierung bildet den **Zustand** der **Rechenzentren** ab (hier 2 RZ mit jeweils 2 Clustern)
- Jeder **Balken** repräsentiert eine **Serverinstanz** und die **Höhe des Balkens** repräsentiert die **Auslastung**
- Die **interaktive GUI** ermöglicht **Zeitreisen**: Mit dem Slider kann man die Auslastungssituationen vor- und zurückspulen



Vielen Dank für Ihre Aufmerksamkeit. Für Fragen stehen wir Ihnen jederzeit zur Verfügung!

Matthias Seßler

**Bundesagentur für Arbeit
IT-Systemhaus**



Bundesagentur
für Arbeit

Tafelhofstraße 4
90443 Nürnberg, Germany

E: matthias.sessler@arbeitsagentur.de

Eldar Sultanow

Capgemini



Bahnhofstraße 11C
90402 Nürnberg, Germany

E: eldar.sultanow@capgemini.com

Oliver Weiß

Capgemini



Mainzer Landstraße 178-190
60327 Frankfurt, Germany

E: oliver.weiss@capgemini.com

Frank Pelzel

**Bundesagentur für Arbeit
IT-Systemhaus**



Bundesagentur
für Arbeit

Südwestpark 26
90449 Nürnberg, Germany

E: frank.pelzel@arbeitsagentur.de

Martina Hofmann

**Bundesagentur für Arbeit
IT-Systemhaus**



Bundesagentur
für Arbeit

Südwestpark 26
90449 Nürnberg, Germany

E: martina.hofmann4@arbeitsagentur.de