



**Key-Work**

DATA DRIVEN SERVICES



# **From the Ground to the Cloud**

## Crunching Trillions of Events a Day

# Agenda

Online Advertisement

Analysing Web-Traffic

Environment On Premise

Why and How to Use Cloud Infrastructure

Migrating Existing Solution to GCP

Project Outcome and Lessons Learned



# Profitieren Sie von der 20-jährigen Marketing- und IT-Erfahrung

Key-Work ist Ihr Berater, Implementierungspartner und entwickelt eigene Software



hochmotivierte  
Experten



Ihr Partner für Data Driven  
Services seit 1999



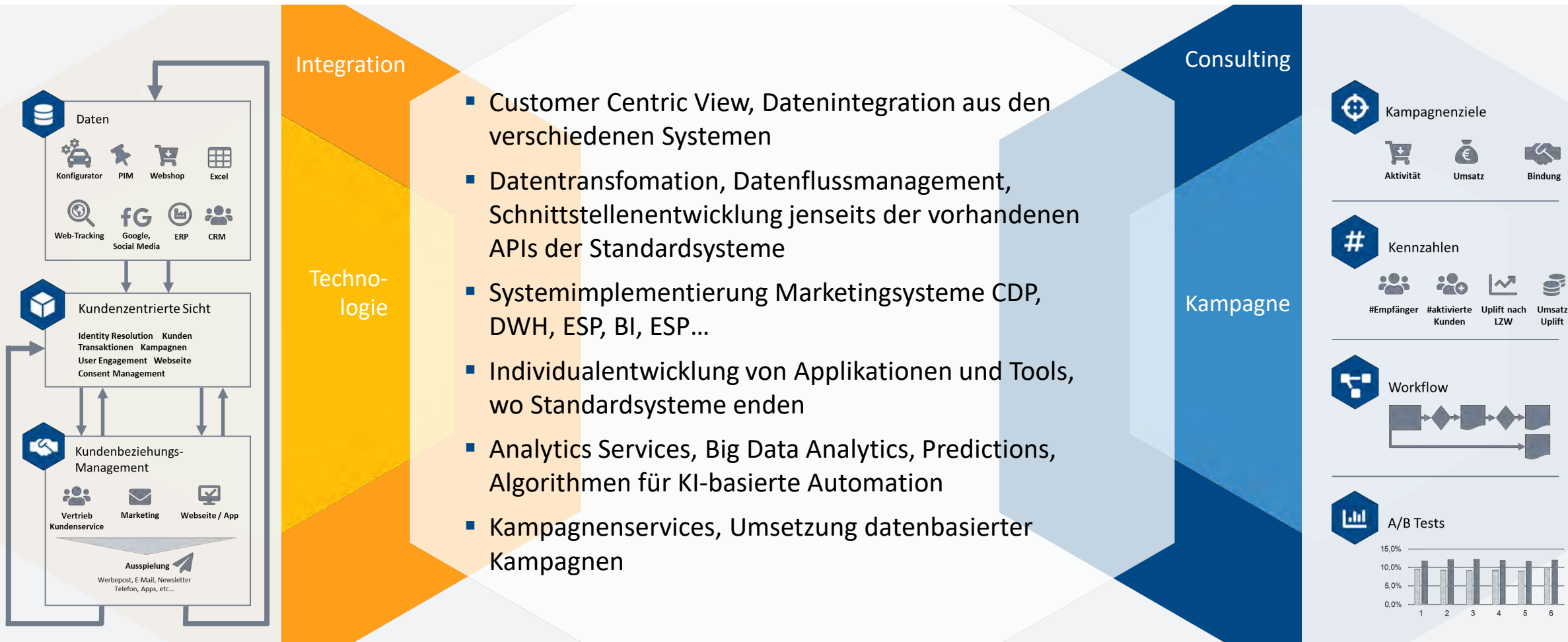
Marketing Services  
Data Platform Development  
Application Development

Key-Work in der Data Driven  
Marketing Umgebung

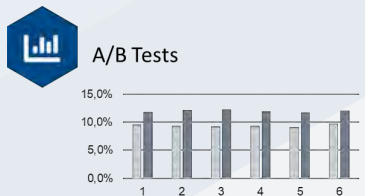
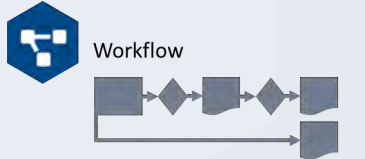


Technologie ◀ Handlungsmöglichkeiten ▶ Kampagnen

# Key-Work Leistungs-Portfolio



- Customer Centric View, Datenintegration aus den verschiedenen Systemen
- Datentransformation, Datenflussmanagement, Schnittstellenentwicklung jenseits der vorhandenen APIs der Standardsysteme
- Systemimplementierung Marketingsysteme CDP, DWH, ESP, BI, ESP...
- Individualentwicklung von Applikationen und Tools, wo Standardsysteme enden
- Analytics Services, Big Data Analytics, Predictions, Algorithmen für KI-basierte Automation
- Kampagnenservices, Umsetzung datenbasierter Kampagnen



# Agenda

## Online Advertisement

Analysing Web-Traffic

Environment On Premise

Why and How to Use Cloud Infrastructure

Migrating Existing Solution to GCP

Project Outcome and Lessons Learned



# A typical Website consists of contents and advertisements

**Das kleine Gartenhaus**  
Wetterfest, nie mehr streichen, über 100 Farben und Dekors, Modulare Größen, mehr Stauraum

**Gartenhaus selber bauen**  
von THOMAS HESS  
Bewertung: ★★★★★ (32)  
Mit dieser Schritt-für-Schritt-Anleitung können Sie ein individuelles Gartenhaus selber bauen. Außerdem geht es um Fundament und Baugenehmigung.

**Microsoft Surface**  
Elegant und mobil – Microsoft Surface Laptop 2  
★★★★★ 36 Amazon.de Kundenbewertungen  
"Schön, elegant und praktikabel"

1793  
f  
p  
t  
m

**ZEIT ONLINE**  
Suche

**RWE**  
RWE wird einer der weltweit größten Stromerzeuger aus

**Klimapolitik**  
**Wir brauchen noch viele Demonstrationen**  
Pünktlich zum Protest von Extinction Rebellion verwässert die Regierung ihre Klimapolitik – und tut noch so, als sei nichts. So zerstört man seine eigene Glaubwürdigkeit.  
Ein Kommentar von **Alexandra Endres**  
7. Oktober 2019, 17:44 Uhr / 4 Kommentare

**SPIEGEL ONLINE**  
Anmelden

Schlagzeilen | DAX 12.084,98 | Abo

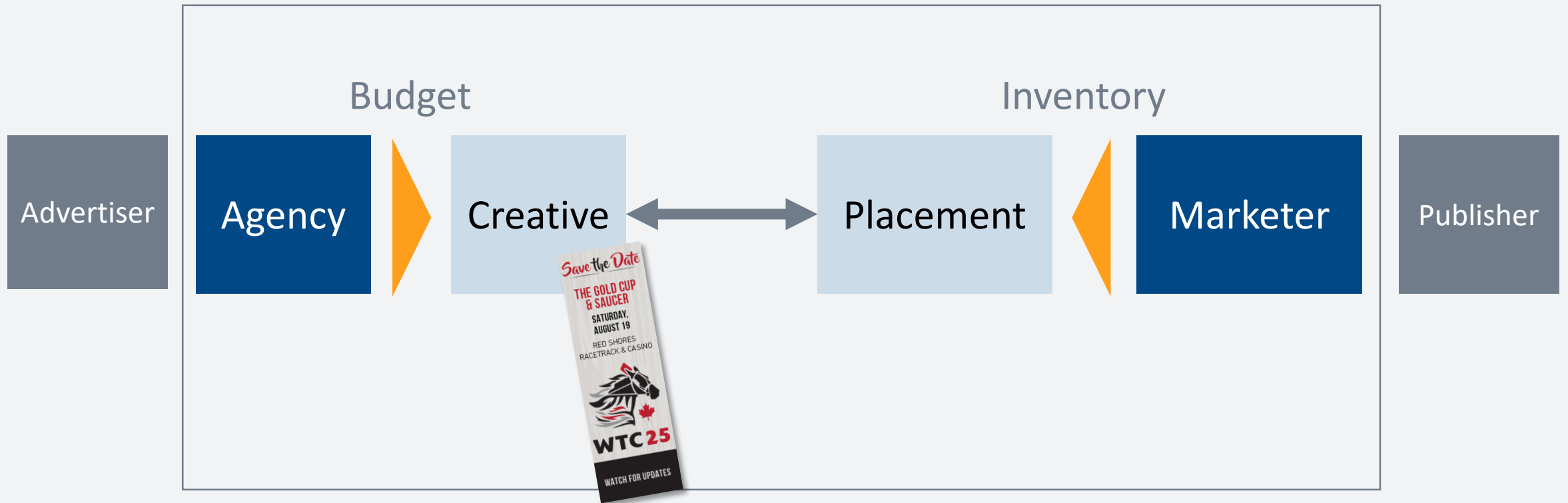
**Aufschub**  
Alle Artikel auf einen Blick...

**ANSTOSSEN STATT ANKLICKEN.**  
Beck's Pilsener

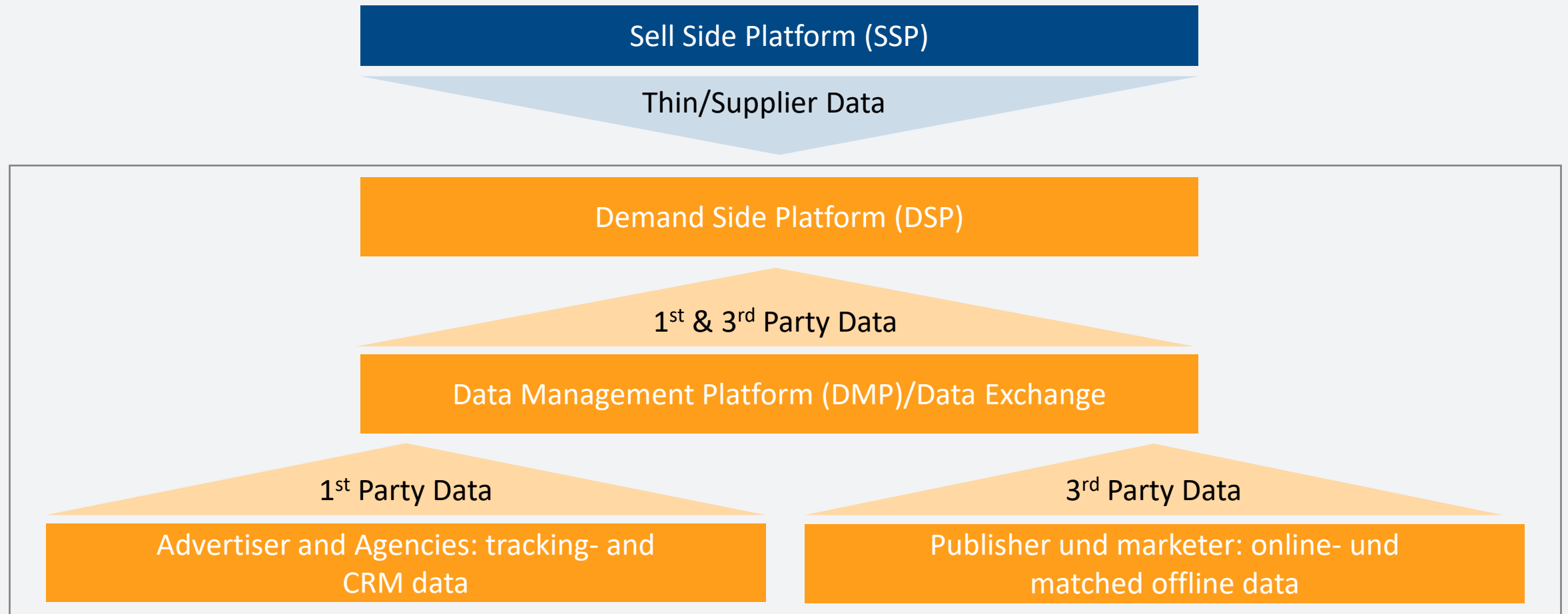
# Advertisements on websites are placed based on a contract between advertiser and publisher

Demand side –  
searches ad space and audience

Supply side –  
offers website and traffic



# Ad Placements in Online Marketing are Optimized Based on Data and Information – and Require a Specialized Tech-Stack



# Agenda

Online Advertisement

## Analysing Web-Traffic

Environment On Premise

Why and How to Use Cloud Infrastructure

Migrating Existing Solution to GCP

Project Outcome and Lessons Learned



# Masses of Data are Collected when Visitors view Website Contents and Ads are Placed on the Website

„Impressions“

„Clicks“

„Actions“

**User Data:** Cookie ID, Cross Device ID, Device Type, Operation System, conforming to GDPR!

**Geo Data:** Country, City, Geo-Location

**Campaign Data:** Advertiser, Agency, Campaign, Website, Placement, Creative

For the current project: 10.000-100.000 events per second decomposing to  
>100 Billion data sets per day to be analysed – and still scaling up

# Agenda

Online Advertisement

Analysing Web-Traffic

**Environment On Premise**

Why and How to Use Cloud Infrastructure

Migrating Existing Solution to GCP

Project Outcome and Lessons Learned



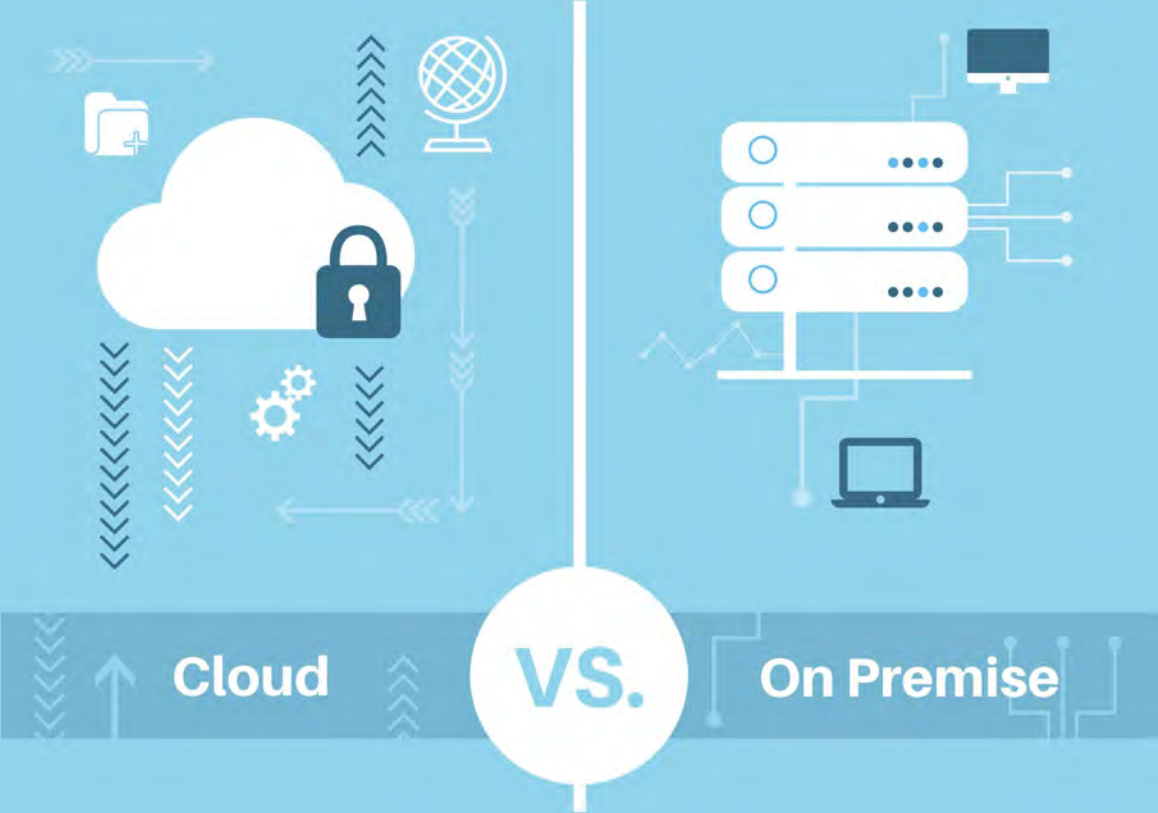
# Infrastructure Involved

## Numbers

**Applications (Wildfly, MariaDB)** Global > 200 Hosts

**Data (Vertica, HDFS, Couchbase)** Global > 300 Hosts

**Processes (Storm, Spark)** Global > 100 Hosts



# Globally Distributed Data Centers

APAC

NA

EU

CH (unchanged)



# Agenda

Online Advertisement

Analysing Web-Traffic

Environment On Premise

## Why and How to Use Cloud Infrastructure

Migrating Existing Solution to GCP

Project Outcome and Lessons Learned



# Reasons for Using Cloud

**Startups: no infrastructure needed, just scale with the business**

**Established business: IT Infrastructure is not key business area**

- small and medium size: need to concentrate on core business – don't get distracted by running IT infrastructure
- For big companies it also needs a lot of attention and resources – it could be better to concentrate on business solutions rather than infrastructure

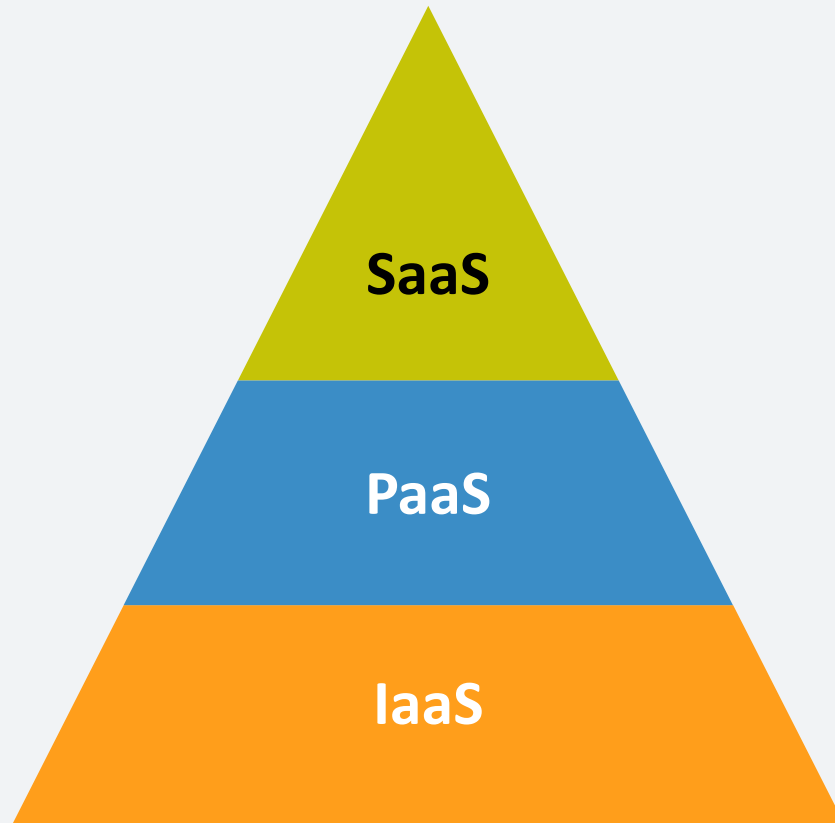
**Hybrid approach: Run Data Lake in the Cloud**

- avoid to pull large data volumes generated in the cloud into your own data center
- keep curated and raw data in data lake in the cloud
- ship extracts to on premise data center for operational reporting and analytics

**In all Scenarios: Increased Agility (!) and Potential for Reduced Cost (?)**



# Cloud Infrastructure Service Levels



Using Cloud Infrastructure allows to concentrate on core business – no need to run a data center

Depending on service model keep the freedom but also the burden of deploying any technology or use technology/ solutions provided by the Cloud Service Provider

# Benefits and Economics are Important

- Sizing and Costs -> over sizing is important on-premise – rightsizing is important in the cloud making intensive use of automation to adapt sizing dynamically
- Risk -> offload responsibility to cloud service provider; use intrinsic redundancy
- Knowledge and human resources -> concentrate on building solutions not hauling computer and IT-Infrastructure
- Use PaaS as much as possible/sensible – rightsizing is typically easier with PaaS approach, but consider tradeoff with respect to vendor lock-in



# Agenda

Online Advertisement

Analysing Web-Traffic

Environment On Premise

Why and How to Use Cloud Infrastructure

**Migrating Existing Solution to GCP**

Project Outcome and Lessons Learned



# Project Objectives and Timeline

Reduce Costs for Data Center IT Infrastructure and Concentrate on Core Business

**Approach:**

- Replace on premise data centers with cloud infrastructure
- Move complete solutions to Google Cloud Platform (GCP)



## Migration Project

# Proofs of Concept

## Define Mapping for Cloud Native Services (PaaS-Szenario)

### Proof-of-Concept

- Features supported
- Evaluate code adjustments needed
- Performance implications and scaling
- Expected costs



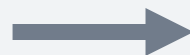
# In Order to Employ PaaS you need to find an appropriate Cloud Native or Platform Service



**Google** BigQuery



**Google** BigTable



**Google** Cloud Storage

# Agenda

Online Advertisement

Analysing Web-Traffic

Environment On Premise

Why and How to Use Cloud Infrastructure

Migrating Existing Solution to GCP

Project Outcome and Lessons Learned



# Iterations and Lessons Learned – Vertica

## Migrating Vertica functionality wasn't straight forward

- Performance
- Reformulating queries

## Switch to IaaS approach

- Build abstraction layer to allow dynamic switching between Vertica and Big Query
- Immediate impact: Doubling costs for Vertica Infrastructure

## Optimize Sizing and Footprint

- Right-sizing together with GCP consultants
- Cleaning up data
- Further effort to employ platform services do not show a promising ROI



# Iterations and Lessons Learned – HDFS

## Using HDFS API for Google Storage

- Some solutions implemented functionality beyond the standard API

## Short-term solution: use HDFS-Cluster in GCP

- Again resort to IaaS approach
- Adjust solution to use cloud platform with Google API



# Recommendations

Know the limitations of each GCP component upfront – need to find that information online to plan your implementation

Review footprint of your applications and data – optimize types of resources for use case e.g. expensive vs. cheap storage

Plan data migration early and consider volume, bandwidth and location – gather data requirements thoroughly

Script everything – no step of the initial deployment and transfer of data should be done through the command line or the web interface

Managing costs is important, monitor costs and make sure you understand how to profit from committed use discounts



# Finale Outcome



Three data centers could be replaced with common Google Cloud Platform infrastructure

Reduced effort keeping data centers running

More agility setting up and adjusting environments

Reduction in cost by about 30%

## **Vielen Dank für Ihre Aufmerksamkeit**



**Dr. Klaus Schroer-Hölzer**  
Head of Data Platform Development  
+49 (0)721 78203 272  
klaus.schroer@key-work.de



**key-work.de**