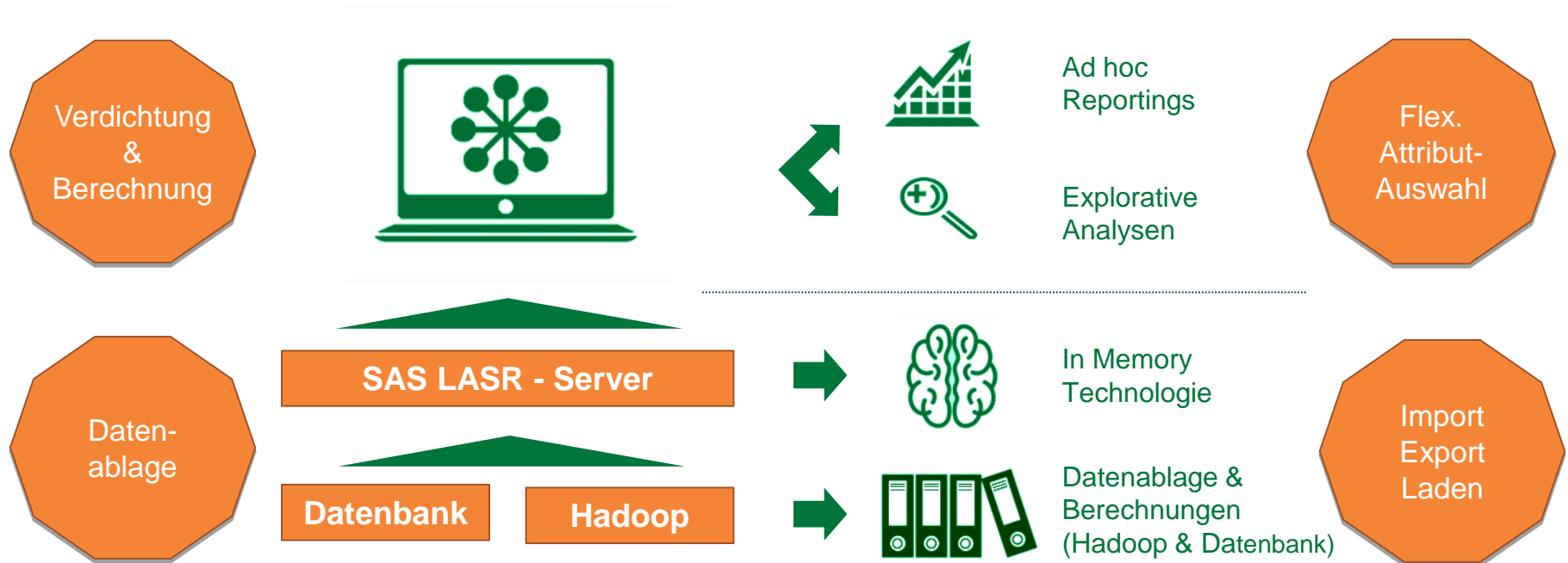


BigData mit SAS VA in der Virtual Private Cloud

Warum Prototypen ein guter Einstieg sind und Hardware-Engpässe überbrücken.
Und: Kann die Cloud die eigene Serverfarm dauerhaft und kostengünstig ersetzen?

Die Vision: Die Revolutionierung des analytischen Tages- und Projektgeschäfts auf kleinen und riesigen Datenmengen.

- Schaffung einer bisher nicht erreichten Flexibilität UND Geschwindigkeit vom Datenaufbereitungsprozess bis zur Nutzung der analytischen Frontends
- für die komplette Vielfalt des analytischen Tages- und Projektgeschäfts,
- damit alle analytisch arbeitenden Mitarbeiter leicht selbst kleine bis riesige Analysetabellen erzeugen,
- um dadurch End-to-End allein den Analyseprozess durchführen zu können.



Die Plattform: klassisch, effizient oder virtuell?

Gesucht: Plattform ca. 3,6 Mrd. Zustandskombinationen und geschätzt 3,5 TB RAM.

VA & Hadoop Distributed (MPP)

SAS VA auf 12-48 Blades
192-768 Cores
3-12TB RAM

Hadoop auf 4 DL380
64 Cores
1TB RAM



500.000-700.000 €
up&running 6-9 Monate

VA & Hadoop Co-Located 16-20 HP DL 360

2x SAS-Master
2x Hadoop-Master

16x SAS-/
Hadoop-Worker
256 Cores
4TB RAM
28TB HDD



350.000 €
up&running 3-5 Monate

Amazon MPP Installation Prototyp / Alternative

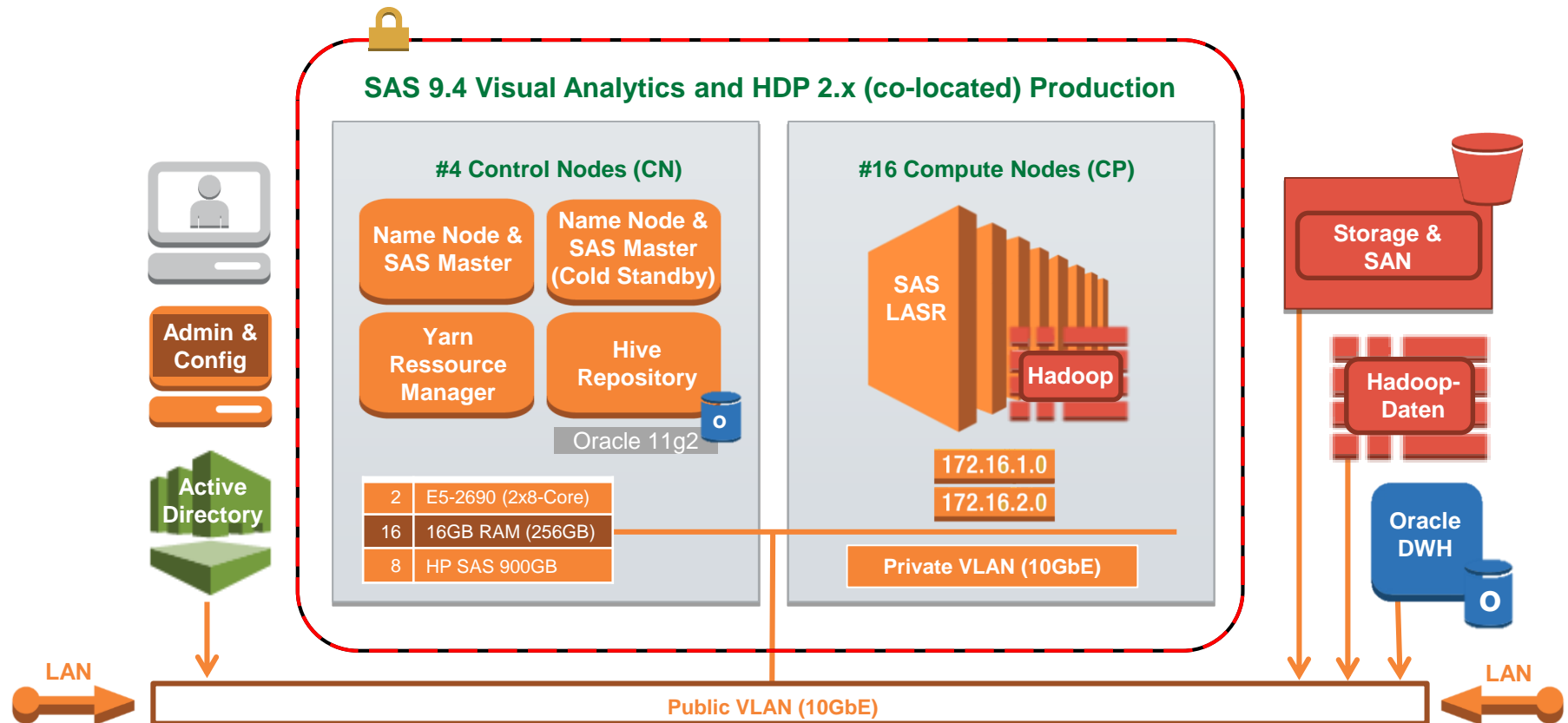
1x SAS-Master
i2.2xlarge
8 Core
61GB RAM

3x SAS-/
Hadoop-Worker
24 Cores
256 GB RAM



60.000 € / 2 Jahre
up&running 1 Monat

Das Zieldesign der co-located Installation wurde mit SAS und Hortonworks erarbeitet und abgestimmt.



Anforderungen

Integration in bestehende IT-Infrastruktur.

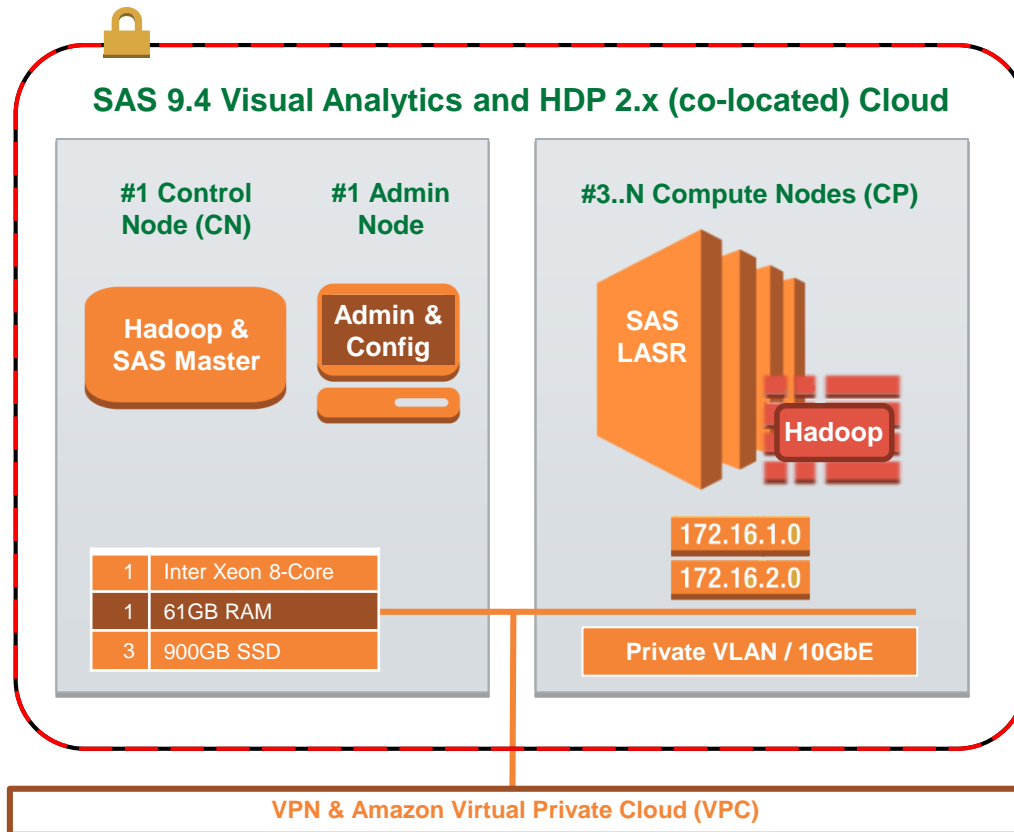
Aufbereitung der Analyse-Daten mit **BigData-Technologien** erforderlich.

Daten HDFS abgelegt & über **SAS Embedded Process** angebunden.

10GbE Interconnect für **InMemory Technologien** erforderlich.

Flexibler Zugriff auf **alle Datenquellen**.

Der Prototyp für die Amazon Cloud wurde konzipiert, um die angestrebte Realisierungsplattform zu testen.



Ziele

- Prototyp simuliert Zieldesign im Faktor 1:5.
- Ziel ist das **Sammeln von Erfahrungen** für die finale Realisierungsplattform.
- Prototyp muss **flexibel** und **on-Demand horizontal & vertikal skaliert** werden können.
- Ressourcen-Bereitstellung, -Verwaltung und **Anwenderzugriff** müssen **über Web** realisiert werden können.
- Unterschiedliche **Hardware- & Software-Konfigurationen** müssen getestet werden können.
- Prüfung der Cloud als dauerhafte Alternative zum Rechenzentrum.

Anforderungen

Anbindung der Cloud an das Intranet (**VPN**).

Parallelbetrieb der SAS und Hadoop Verwaltungskomponenten.

Auftragsdatenverarbeitung von **Kundendaten** in der Cloud (ADV).

On-Demand Anpassung von Arbeits- und Festplattenspeicher.

Flexibler Zugriff auf **alle Datenquellen**.

Amazon Web Services

Compute



EC2
Virtual Servers in the Cloud



Lambda
Run Code in Response to Events



EC2 Container Service
Run and Manage Docker Containers

Storage & Content Delivery



S3
Scalable Storage in the Cloud



Elastic File System PREVIEW
Fully Managed File System for EC2



Storage Gateway
Integrates On-Premises IT Environments with Cloud Storage



Glacier
Archive Storage in the Cloud



CloudFront
Global Content Delivery Network

Database



RDS
MySQL, Postgres, Oracle, SQL Server, and Amazon Aurora



DynamoDB
Predictable and Scalable NoSQL Data Store



ElastiCache
In-Memory Cache



Redshift
Managed Petabyte-Scale Data Warehouse Service

Networking



VPC
Isolated Cloud Resources



Direct Connect
Dedicated Network Connection to AWS



Route 53
Scalable DNS and Domain Name Registration

Administration & Security



Directory Service
Managed Directories in the Cloud



Identity & Access Management
Access Control and Key Management



Trusted Advisor
AWS Cloud Optimization Expert



CloudTrail
User Activity and Change Tracking



Config
Resource Configurations and Inventory



CloudWatch
Resource and Application Monitoring

Deployment & Management



Elastic Beanstalk
AWS Application Container



OpsWorks
DevOps Application Management Service



CloudFormation
Templated AWS Resource Creation



CodeDeploy
Automated Deployments

Analytics



EMR
Managed Hadoop Framework



Kinesis
Real-time Processing of Streaming Big Data



Data Pipeline
Orchestration for Data-Driven Workflows



Machine Learning
Build Smart Applications Quickly and Easily

Application Services



SQS
Message Queue Service



SWF
Workflow Service for Coordinating Application Components



AppStream
Low Latency Application Streaming



Elastic Transcoder
Easy-to-use Scalable Media Transcoding



SES
Email Sending Service



CloudSearch
Managed Search Service

Mobile Services



Cognito
User Identity and App Data Synchronization



Mobile Analytics
Understand App Usage Data at Scale



SNS
Push Notification Service

Enterprise Applications



WorkSpaces
Desktops in the Cloud

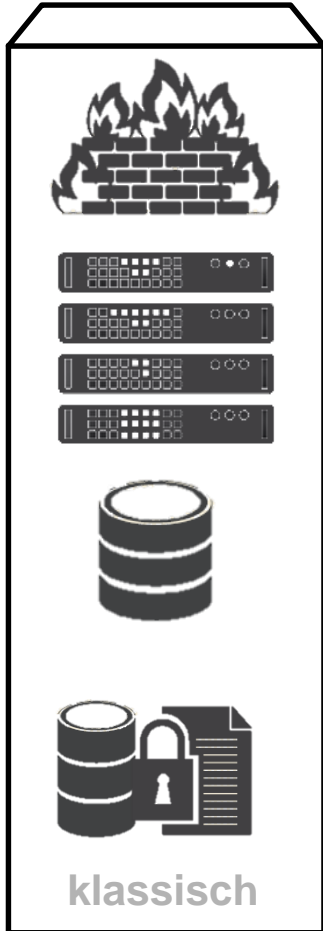


WorkDocs
Secure Enterprise Storage and Sharing Service



WorkMail PREVIEW
Secure Email and Calendaring Service

Die Amazon Cloud als Rechenzentrum



klassisch



IAM

IAM - AWS Identity and Access Management

- Verwaltung und Konfiguration von Benutzern und Rechten.
- Trennung der Maschinen-Accounts und Benutzer-Accounts.
- Anbindung eines unternehmensinternen MS Active Directory möglich.



Amazon VPC

VPC - Amazon Virtual Private Cloud

- Voller Zugriff auf alle Netzwerk-Parameter, wie IP-Adressen, Subnetze, Firewalls, Gateways, VPN, Routen etc.
- Einfache .. komplexe Konfigurationen, variable Leistungsklassen.



Amazon EC2

EC2 - Amazon Elastic Compute Cloud

- Root-Zugriff auf vordefinierte Linux- & Windows-Images.
- Breites Leistungsspektrum von Hardware-Konfigurationen.
- 24/7-Betrieb oder On-Demand-Betrieb.



Amazon RDS

RDS - Amazon Relational Database Service

- Managed, on-Demand MySQL-, Oracle-, MS SQL-, PostgreSQL-Instanzen.
- "License Included"- und "Bring-Your-Own-License (BYOL)"-Modelle.
- Snapshots, Replikation, Backups, Patch-Management eingeschlossen.



Amazon S3

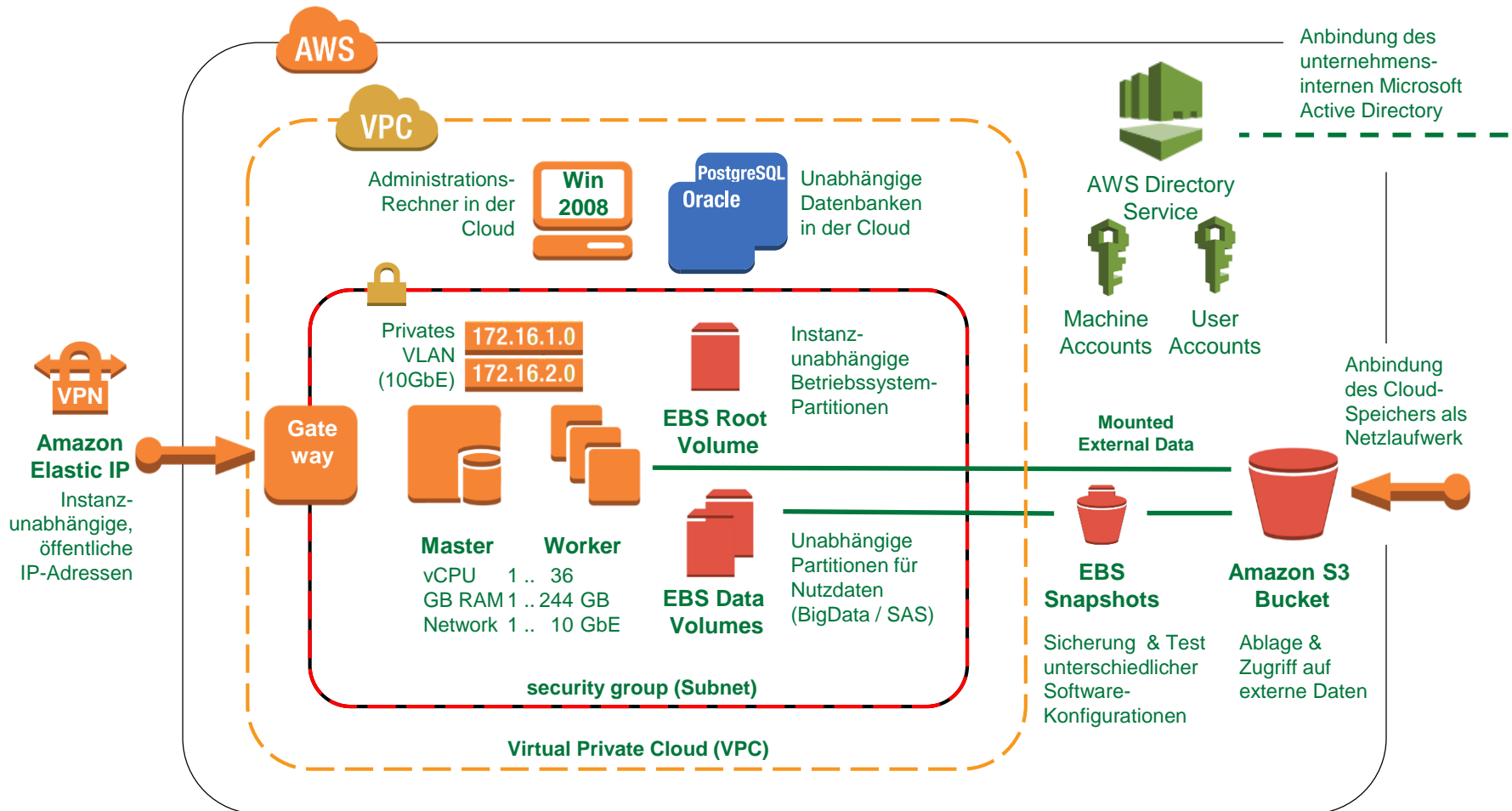
S3 - Amazon Simple Storage Service

- Skalierbarer Online-Speicherplatz unterschiedlicher Güteklassen.
- Kann mit allen Amazon-Diensten gekoppelt und auch als Netzlaufwerk angebunden werden.
- Geeignet zur Datenablage, Archivierung, Backup etc.

Die Amazon Cloud als Alternative zur eigenen Serverfarm



Die Cloud kann auch bei komplexen Szenarien professionell an bestehende IT-Infrastruktur angebunden werden und eignet sich als Alternative.



Lessons Learned: Die Cloud ist eine kostengünstige Alternative und bietet neue Flexibilität, ... wenn man sich traut ...

- **Parallel-Betrieb bzw. Co-located-Installation ist möglich und optimiert Nutzen.**
Volle Leistung der Plattform steht täglich von 9-17 Uhr für SAS VA zur Verfügung; nachts werden Daten aufbereitet und die volle Performance des Clusters für Hadoop genutzt.
- **Hardwarekonfiguration ist flexibel und ermöglicht mehr Freiheitsgrade als ein klassisches Rechenzentrum.**
Hardware- und Netzwerkkonfigurationen sind flexibel, so dass **horizontale & vertikale** Anpassungen auf Basis nachfolgender Fachabstimmungen und auch Lizenzverhandlungen möglich sind.
- **High-Performance und on-Demand-Provisioning**
Realisiert werden konnte eine real-time-fähige High-Performance Plattform zur Analyse großer Datenmengen. Die Bereitstellung erfolgt nach Anforderung und Anwenderberechtigung. Die Geschwindigkeit ist nur durch die Netzwerkanbindung des eigenen Unternehmens beschränkt.
- **Installation und Support durch SAS wie im lokalen Netzwerk**
Root-Zugriff erfolgt via SSH-Konsole, VPN oder auch Windows-Plattformen in der Cloud.
- **Datenhaltung in der Cloud bedarf rechtlicher Abstimmung und viel Überzeugungsarbeit**
Datenhaltung gemäß Datenschutzbestimmungen (ADV) ist in der Regel möglich. Die emotionalen Hürden sind erfahrungsgemäß größer als die rechtlichen. Professionelle Partner können bei der Abstimmung behilflich sein.

Ende

Vielen Dank für Ihre Aufmerksamkeit!

