Ben Weissman

SQL Server 2019 Big Data Clusters

Sponsors

SOLISYON

DYMATRIX
we know your customers.

melissa
GLOBAL INTELLIGENCE

adesso
business.
people.
technology.

IT-xPerts.GmbH
IT Compliance built in

SIOS

GDS
BUSINESS INTELLIGENCE
GmbH

COZYROC

SentryOne

redgate

dbWatch
DATABASE CONTROL

inovex

WÜRTHPHOENIX

vmware

PASS
DEUTSCHLAND e.V.
Microsoft
Data Platform
Community

You Rock!

Gold

Silver

Bronze

Many thanks to our sponsors, without whom such an event would not be possible.

Thank you

Microsoft

For the Venue

Huge “THANK YOU!” to Buck Woody

Who am I?

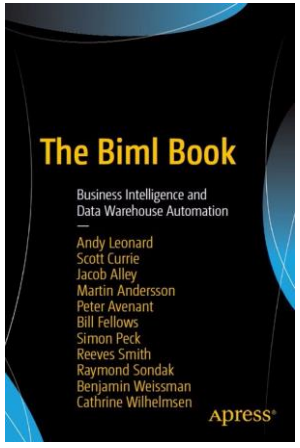
Ben Weissman, Solisyon, Nuernberg/Germany

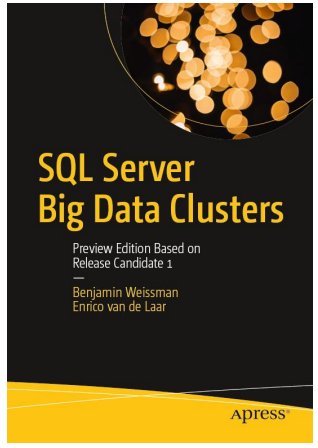
@bweissman

b.weissman@solisyon.de

SQL Server since 6.5

Data Passionist





Please Talk Data To Me

BimlHero CERTIFIED EXPERT

Microsoft CERTIFIED Solutions Associate — Machine Learning

Microsoft CERTIFIED Solutions Expert — Data Management and Analytics

MVP Microsoft® Most Valuable Professional

MICROSOFT PROFESSIONAL PROGRAM

- Data Science
- Big Data
- Artificial Intelligence
- Data Analysis

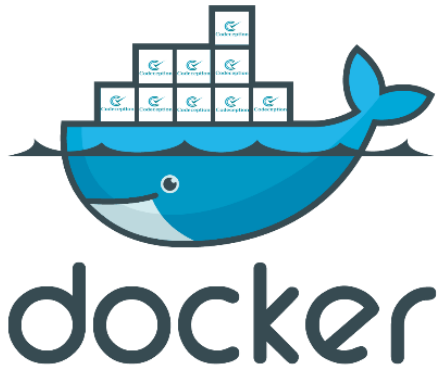
FRIEND OF redgate 2019

GENESEE ACADEMY World-Class Training

Certified Data Vault Modeler

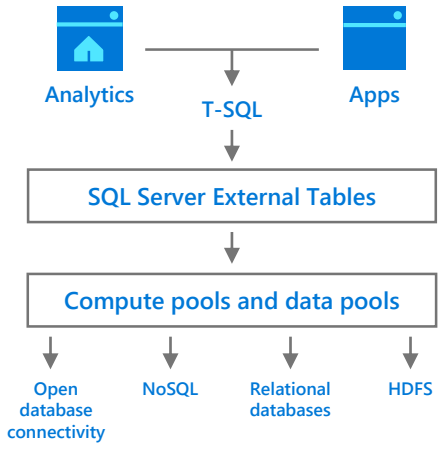
What to look at before getting started…

kubernetes



So what is a Big Data Cluster in SQL 2019?!

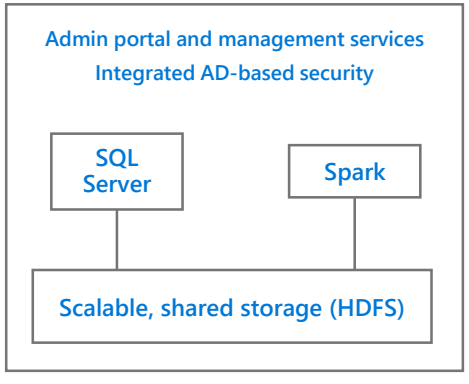
Data Virtualization



Combine data from many sources without moving or replicating it

Scale out compute and caching to boost performance

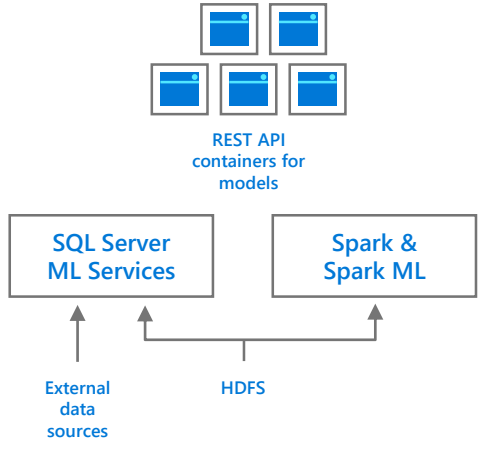
Managed SQL Server, Spark, and Data Lake



Store high volume data in a data lake and access it easily using either SQL or Spark

Management services, admin portal, and integrated security make it all easy to manage

Complete AI platform



Easily feed integrated data from many sources to your model training

Ingest and prep data and then train, store, and operationalize your models all in one system

Scale-Out Processing

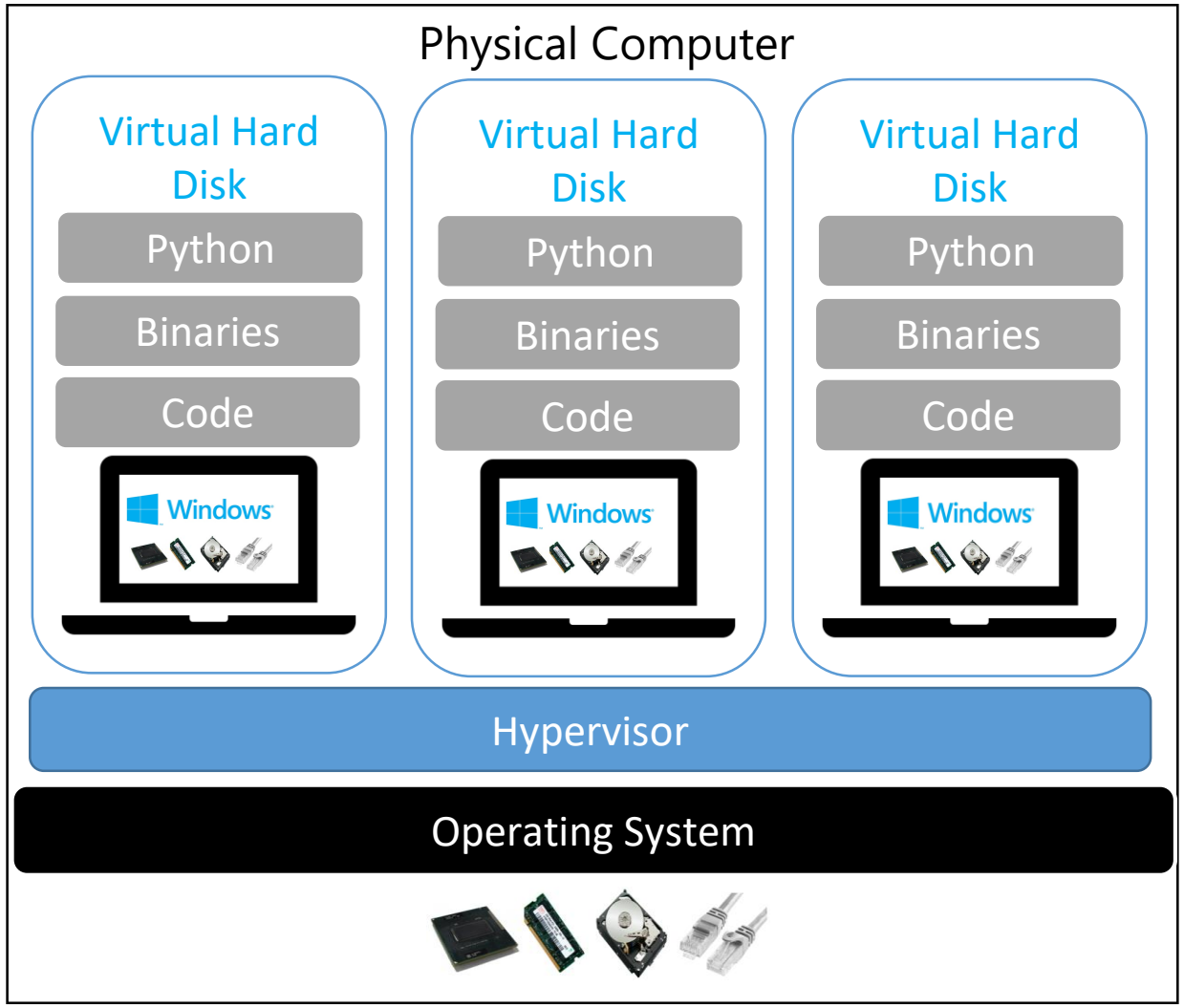
Virtualization

Hyper-V

Hardware Abstraction

Building on hardware, you can create a complete “PC” on top of a Hypervisor layer, which abstracts out the hardware. You still own the Operating System and up

This allows for scale by ring-fencing OS-level dependencies



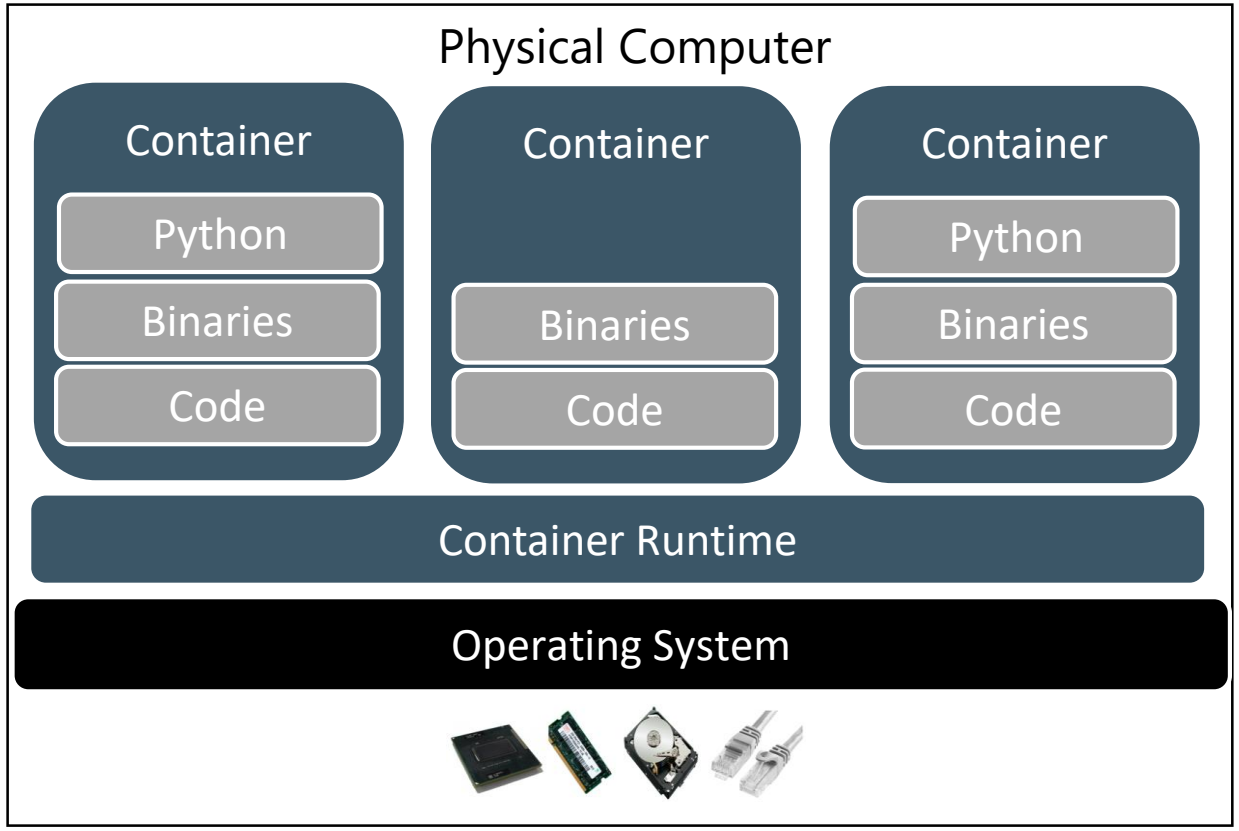
Containers

Abstracting the OS, Allowing complete portability

Containers go one level further than the Hypervisor, and focusing on binaries and applications

Storage and networking are a consideration

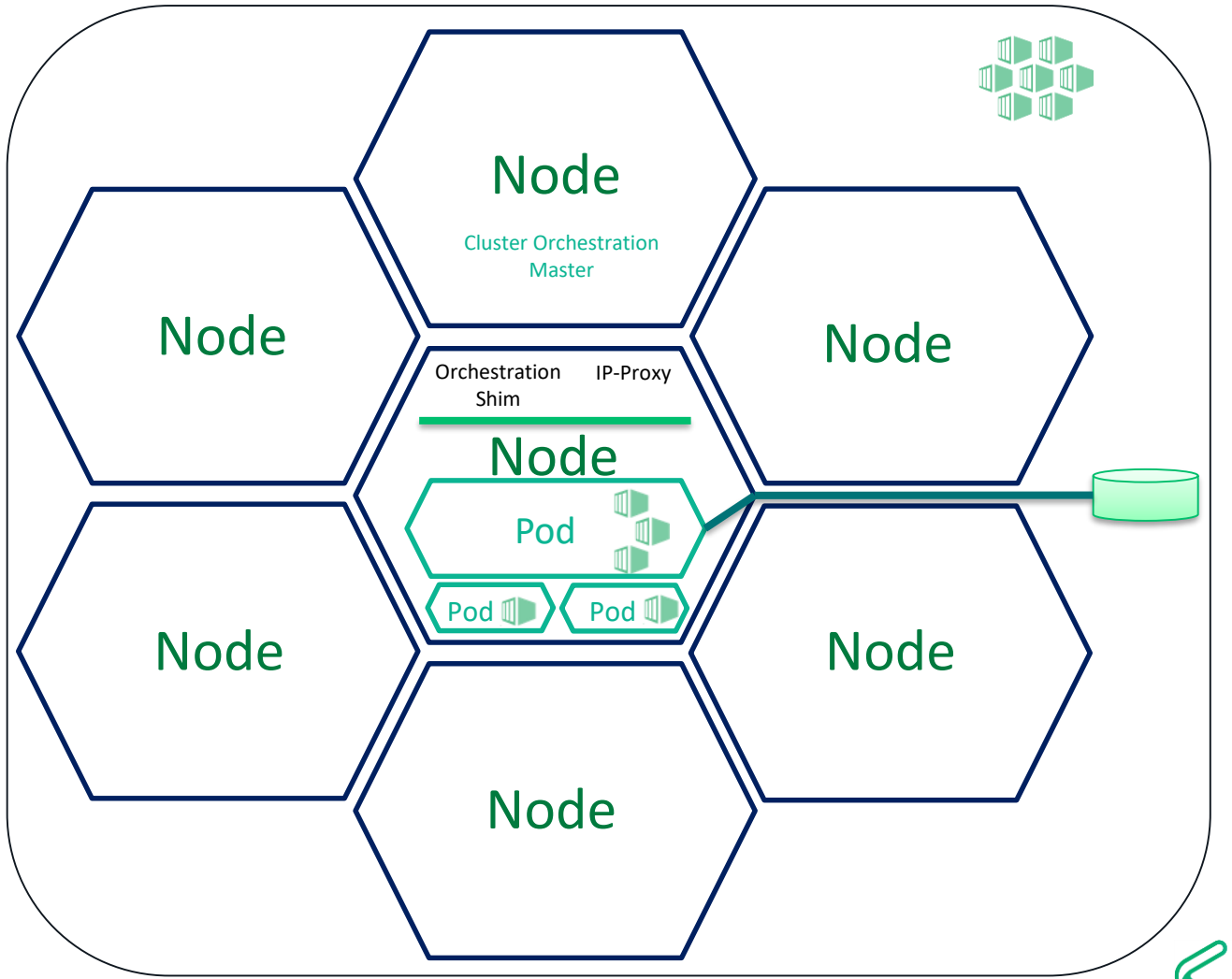
Scale is achieved through multiple containers



Container Orchestration

Containers at Scale

- **Container**(s) live in *Pods*
- **Pod**(s) are abstractions within *Nodes*
- **Node**(s) are PC's or VM's
- **Cluster**(s) are groups of *Nodes*
- Storage is by means of **Volume**(s) mounted through a *Claim*



Generic Cluster

Scale by Purpose

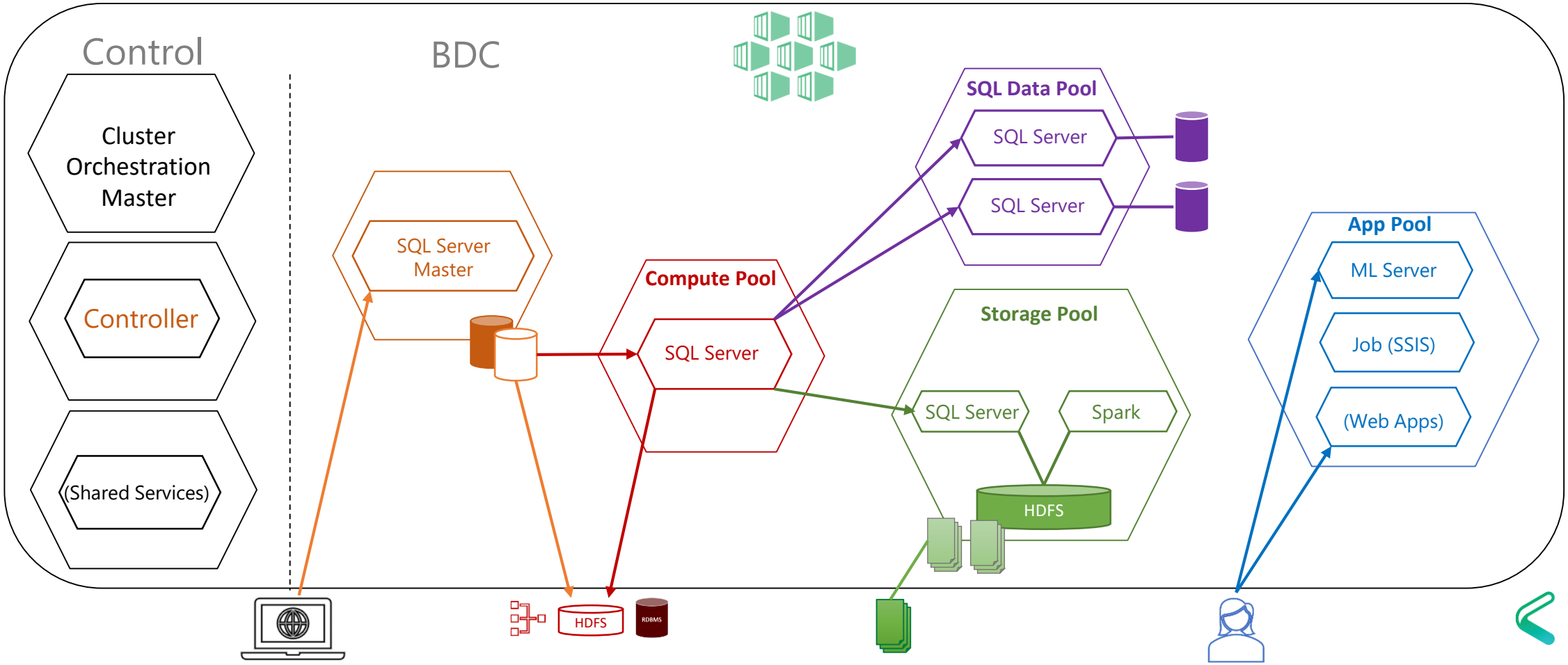
Want to learn more...

...without all that tech stuff?

Complete Architecture

SQL Server 2019 and Big Data (CTP 3.2)

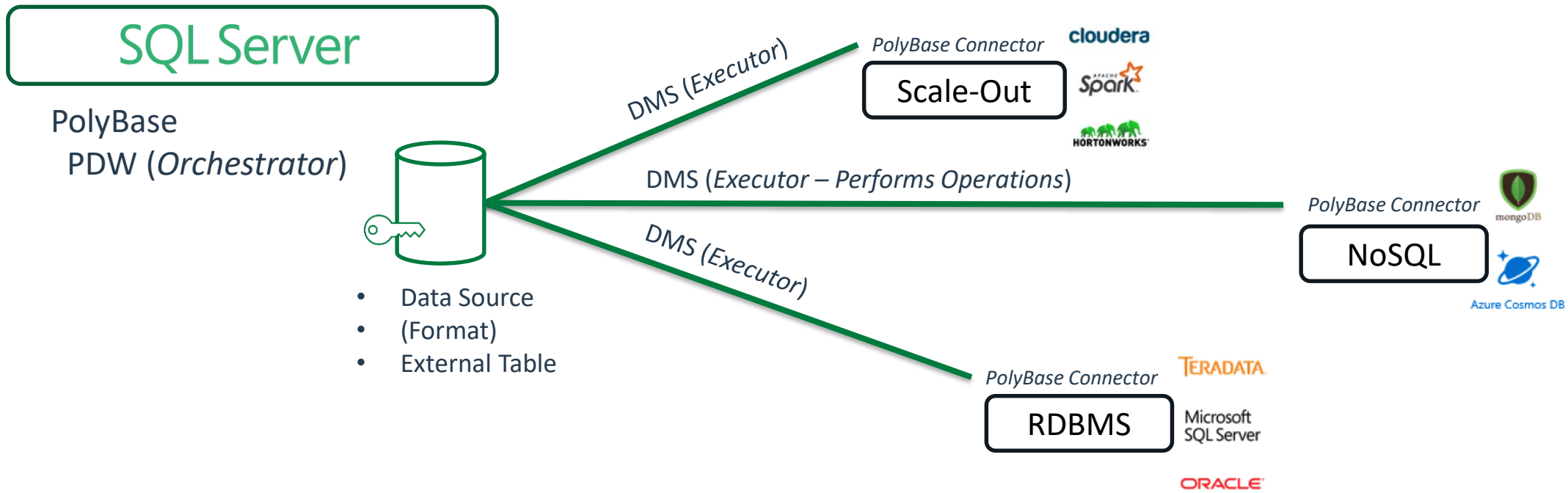
OLTP, Data Virtualization, Data Mart and Big Data



Data Virtualization

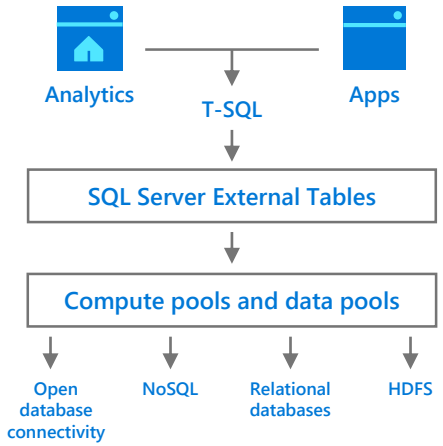
SQL Server 2019 and Big Data

Data Virtualization



Data Virtualization – So it's a linked server?

Data Virtualization



Combine data from many sources without moving or replicating it

Scale out compute and caching to boost performance

Linked Servers

- Instance scoped
- OLEDB providers
- read/write & pass-through statements
- single-threaded
- Separate configuration needed for each instance in Always On Availability Group
- Basic & Integrated authentication

PolyBase External tables

- Database scoped
- ODBC drivers
- For now: read-only operations
- Queries can be scaled out
- No separate configuration needed for Always On Availability Group
- For now: Basic authentication only

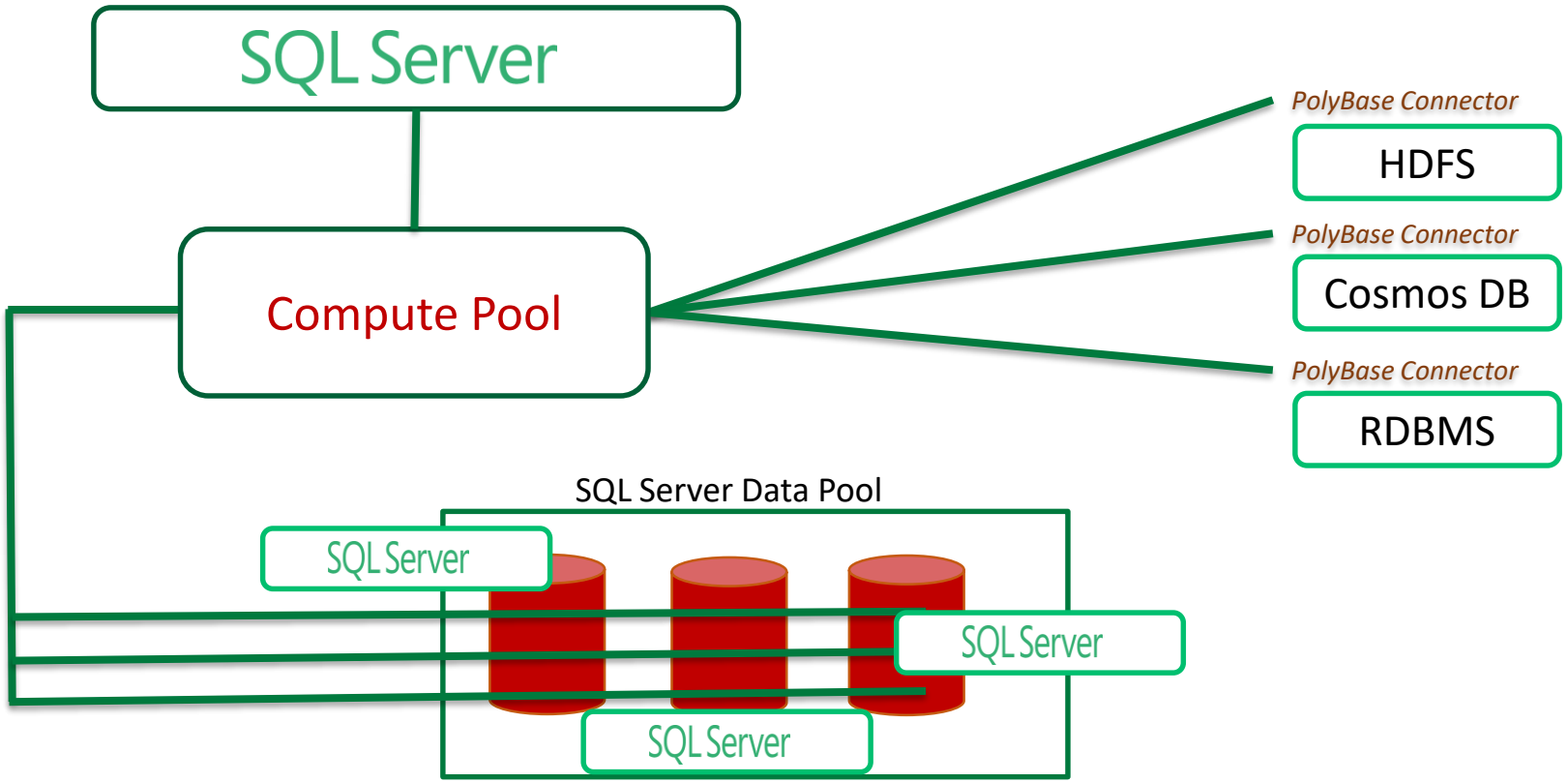
DEMO

Add external table from Azure SQL DB
Query in ADS
Automate with Biml

Data Mart

SQL Server 2019 and Big Data

Data Mart



DEMO

Add Data to HDFS

Query HDFS

Join HDFS against local SQL Table

Insert/query data from SQL Data Pool

Tools, Management and Monitoring

Deployment

How can I get it installed? – PolyBase only

- Get the latest CTP from http://microsoft.com/sql
- Install SQL Server on Windows or Linux including PolyBase
- Enable PolyBase after installation:

```
exec sp_configure @configname = 'polybase enabled', @configvalue = 1;
RECONFIGURE
```

- Restart SQL Server
- Install Azure Data Studio
- Install the vNext Extension for Azure Data Studio

How can I get it installed? – The full package

- Install Kubernetes-CLI, azdata, Python, azure-cli, curl*
- Install Azure Data Studio
  - Add vNext Extension
- Decide on a Kubernetes environment
  - Minikube, AKS, kubeadm, …
- Set environment variables**
- Deploy the cluster using azdata
- When using AKS, consider the Wizard in Azure Data Studio
- When using kubeadm, consider these scripts:
  https://github.com/Microsoft/sql-server-samples/tree/master/samples/features/sql-big-data-cluster/deployment

* Prerequisits

```
Set-ExecutionPolicy Bypass -Scope Process -Force; iex ((New-Object System.Net.WebClient).DownloadString('https://chocolatey.org/install.ps1'))
choco install azure-cli -y
choco install azure-data-studio -y
choco install python3 -y
choco install notepadplusplus -y
$env:Path = [System.Environment]::GetEnvironmentVariable("Path","Machine") + ";" +
[System.Environment]::GetEnvironmentVariable("Path","User")
python -m pip install --upgrade pip
python -m pip install requests
python -m pip install requests --upgrade
choco install curl -y
choco install 7zip -y
choco install kubernetes-cli -y
pip3 install kubernetes
pip3 install -r https://aka.ms/azdata

azuredatastudio --install-extension sql-vnext-0.15.0-win-x64.vsix
```

** Environment Variables

SET CONTROLLER_USERNAME=<controller_admin_name - can be anything>

SET CONTROLLER_PASSWORD=<controller_admin_password - can be anything, password complexity compliant>

SET KNOX_PASSWORD=<knox_password - can be anything, password complexity compliant>

SET MSSQL_SA_PASSWORD=<sa_password_of_master_sql_instance - can be anything, password complexity compliant>

Start actual deployment

azdata bdc create --accept-eula yes --config-profile aks-dev-test

https://docs.microsoft.com/en-us/sql/big-data-cluster/reference-azdata-bdc?view=sqlallproducts-allversions

Install Sample Data

.\bootstrap-sample-db.cmd

USAGE: .\bootstrap-sample-db.cmd <CLUSTER_NAMESPACE> <SQL_MASTER_IP> <SQL_MASTER_SA_PASSWORD> <BACKUP_FILE_PATH> <KNOX_IP> [<KNOX_PASSWORD>]

Default ports are assumed for SQL Master instance & Knox gateway.

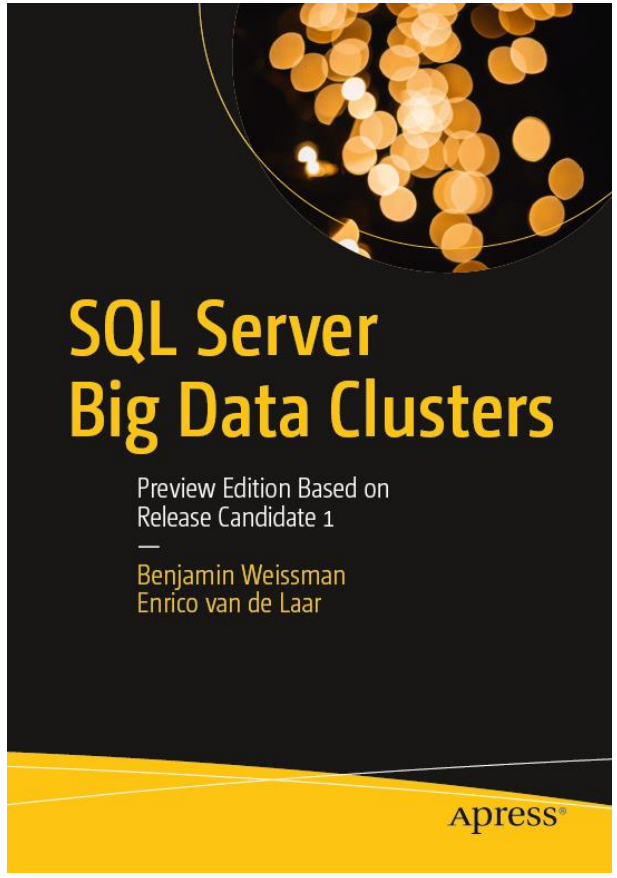
https://github.com/Microsoft/sql-server-samples/tree/master/samples/features/sql-big-data-cluster

Questions?

Ben Weissman

@bweissman

linkedin.com/in/weissmanben/



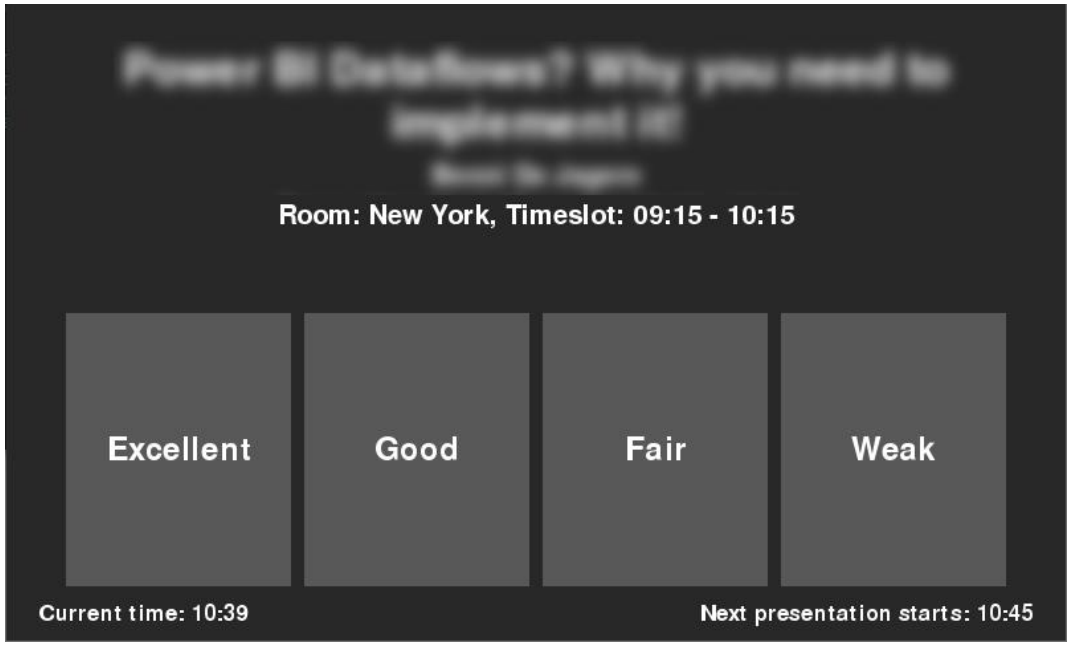
Thank you for your time!

Evaluations

Hardware provided by:

DYMATRIX
we know your customers.

Please rate this session!



Sponsors

SOLISYON

DYMATRIX
we know your customers.

melissa
GLOBAL INTELLIGENCE

adesso
business.
people.
technology.

IT-xPerts.GmbH
IT Compliance built in

SIOS

GDS
BUSINESS INTELLIGENCE
GmbH

COZYROC

SentryOne

redgate

dbWatch
DATABASE CONTROL

inovex

WÜRTHPHOENIX

vmware

PASS
DEUTSCHLAND e.V.
Microsoft
Data Platform
Community

You Rock!

Gold

Silver

Bronze

Many thanks to our sponsors, without whom such an event would not be possible.

PASS Deutschland e.V.

For further information about future events, visit our PASS Deutschland e.V. booth in the exhibitor area.