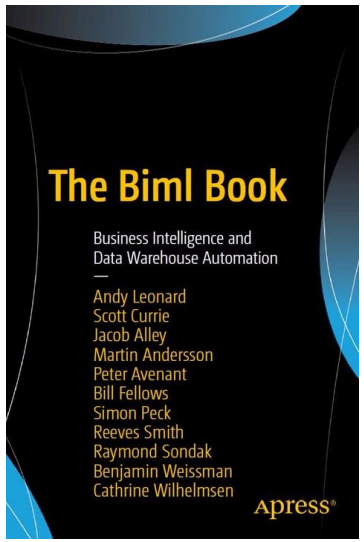
Ben Weissman

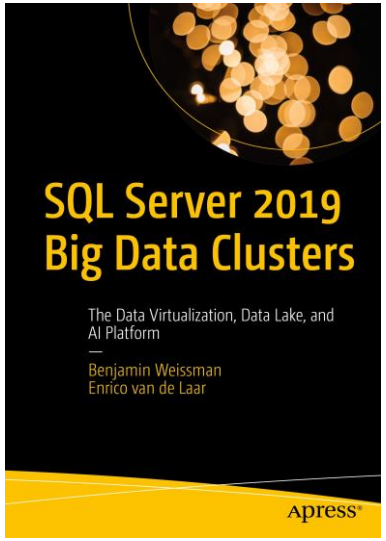
@bweissman
b.weissman@solisyon.de

SQL Server 2019 Big Data Clusters

Who am I?

- Ben Weissman, Solisyon, Germany
- @bweissman
- b.weissman@solisyon.de
- SQL Server since 6.5
- Data Passionist





Please Talk Data To Me

BimlHero CERTIFIED EXPERT

GENESEE ACADEMY World-Class Training — Certified Data Vault Modeler

Microsoft CERTIFIED Solutions Associate — Machine Learning

Microsoft CERTIFIED Solutions Expert — Data Management and Analytics

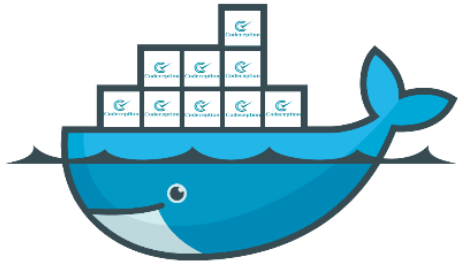
Microsoft Most Valuable Professional

MICROSOFT PROFESSIONAL PROGRAM — Data Science, Big Data, Artificial Intelligence, Data Analysis

FRIEND OF redgate 2020

What to look at before getting started…

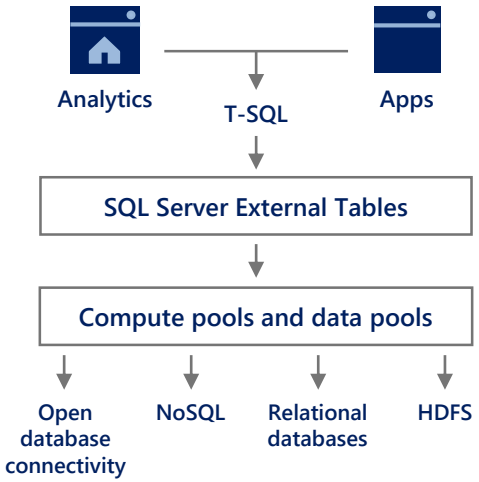
kubernetes



docker

So, what is a SQL 2019 Big Data Cluster?

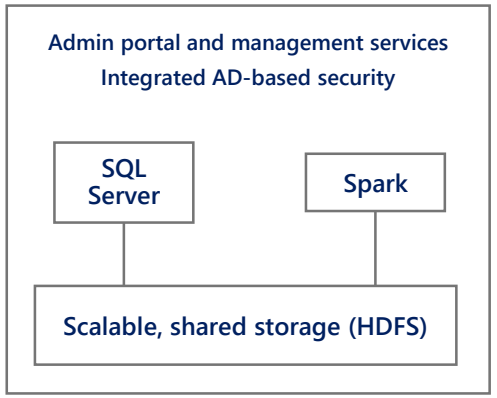
Data Virtualization



Combine data from many sources without moving or replicating it

Scale out compute and caching to boost performance

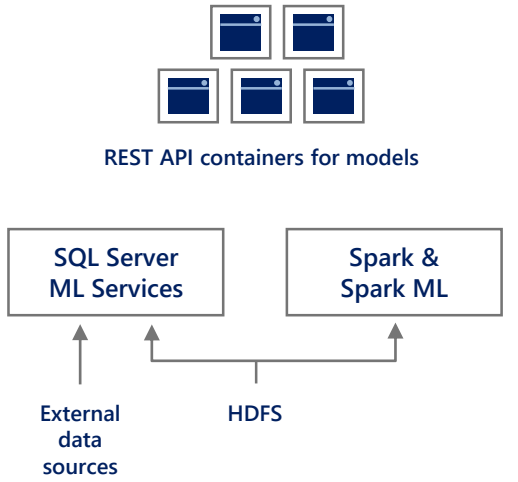
Managed SQL Server, Spark and Data Lake



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

AI/ML 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

This slide: © by Microsoft

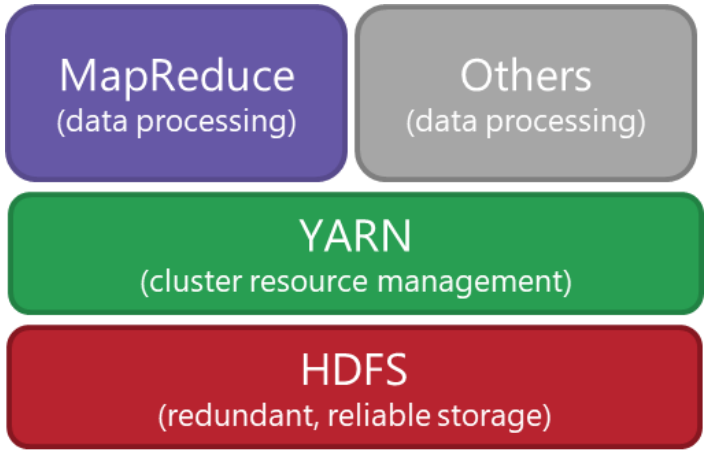
SOLISYON

A little primer…

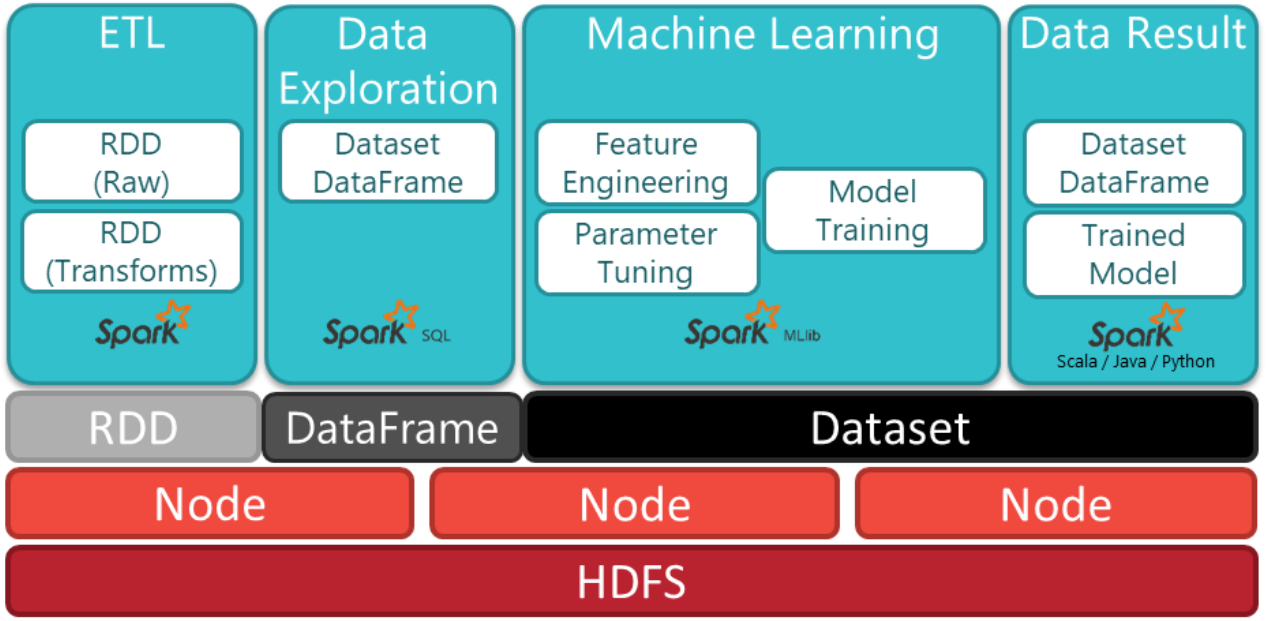
Scaled Processing and Scaled Storage

The foundation of scale

Hadoop



Spark

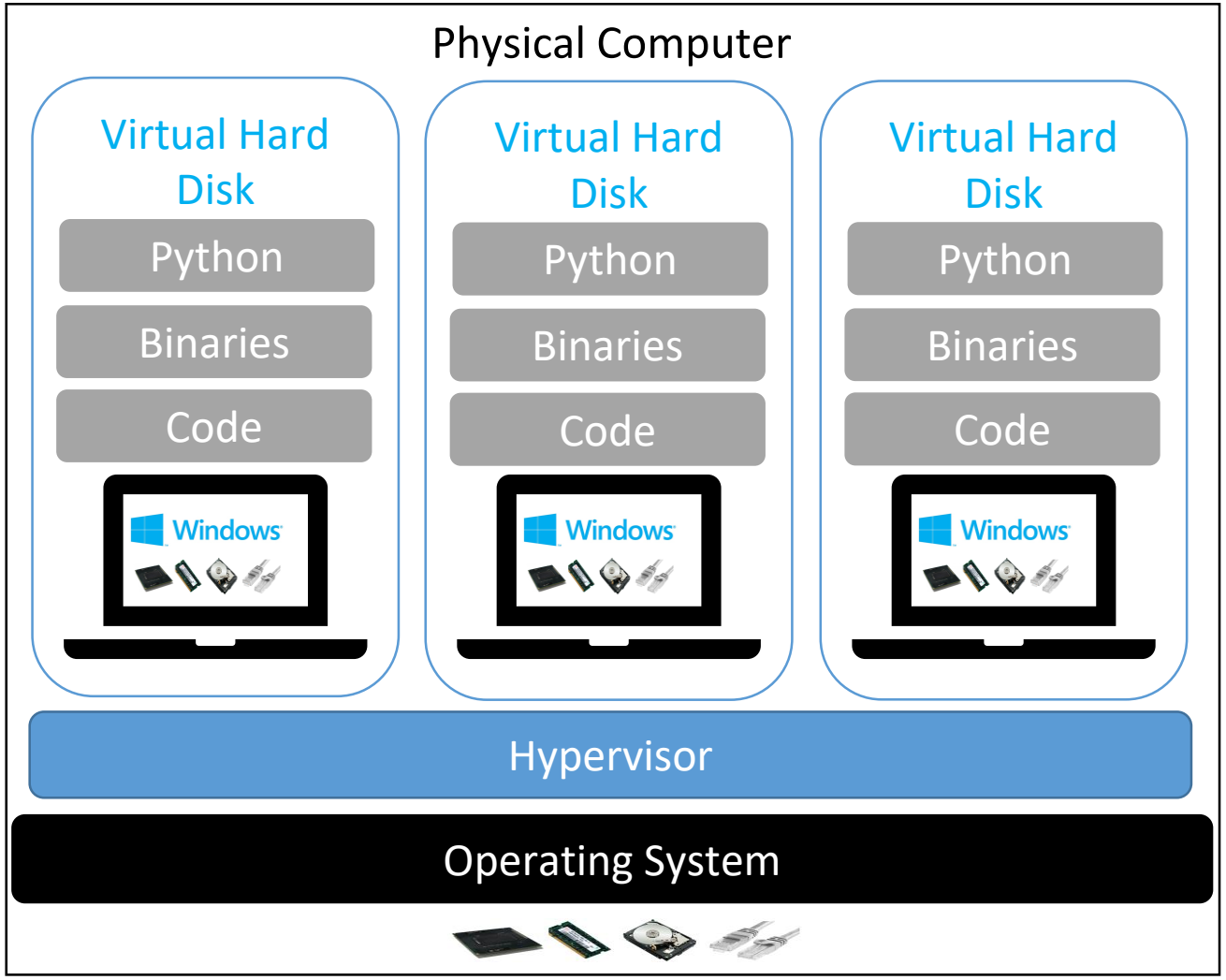


Hardware Abstraction

Hyper-V

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

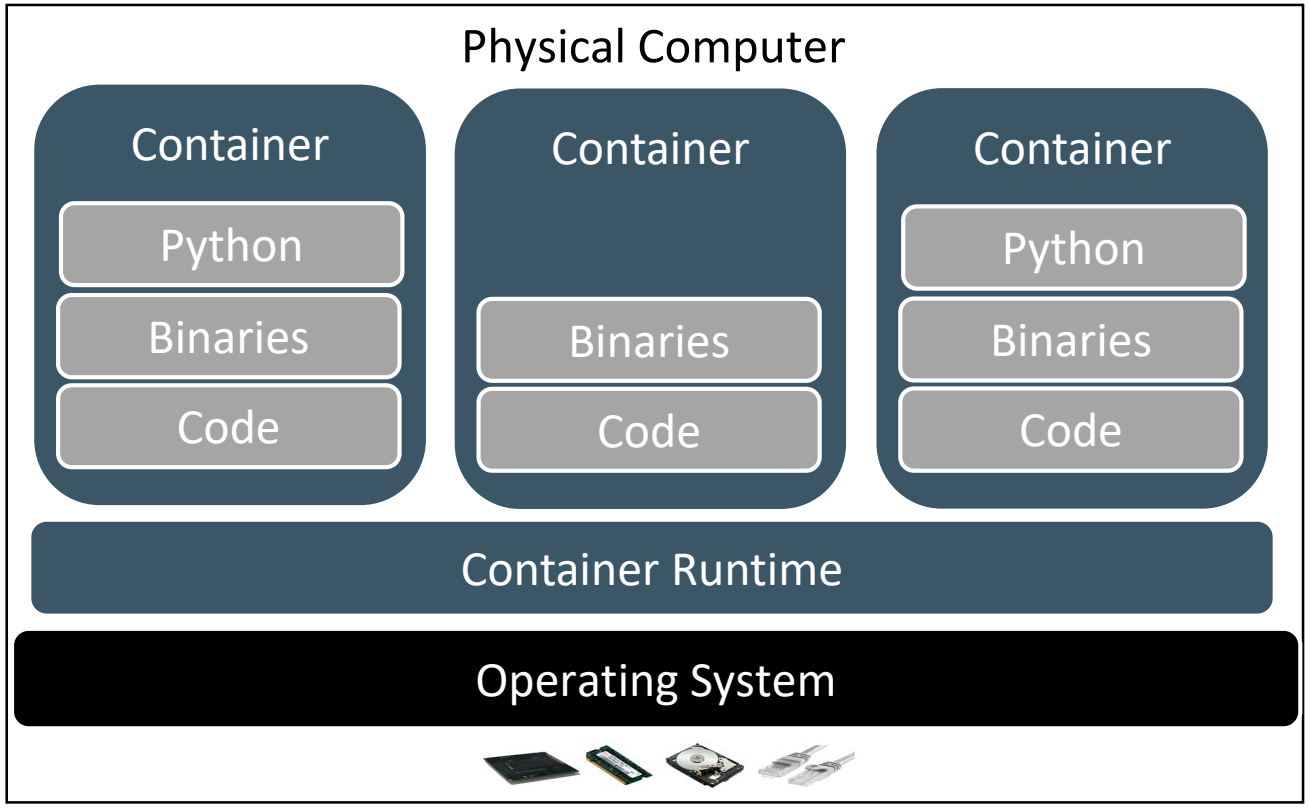


Abstracting the OS, allowing complete portability

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

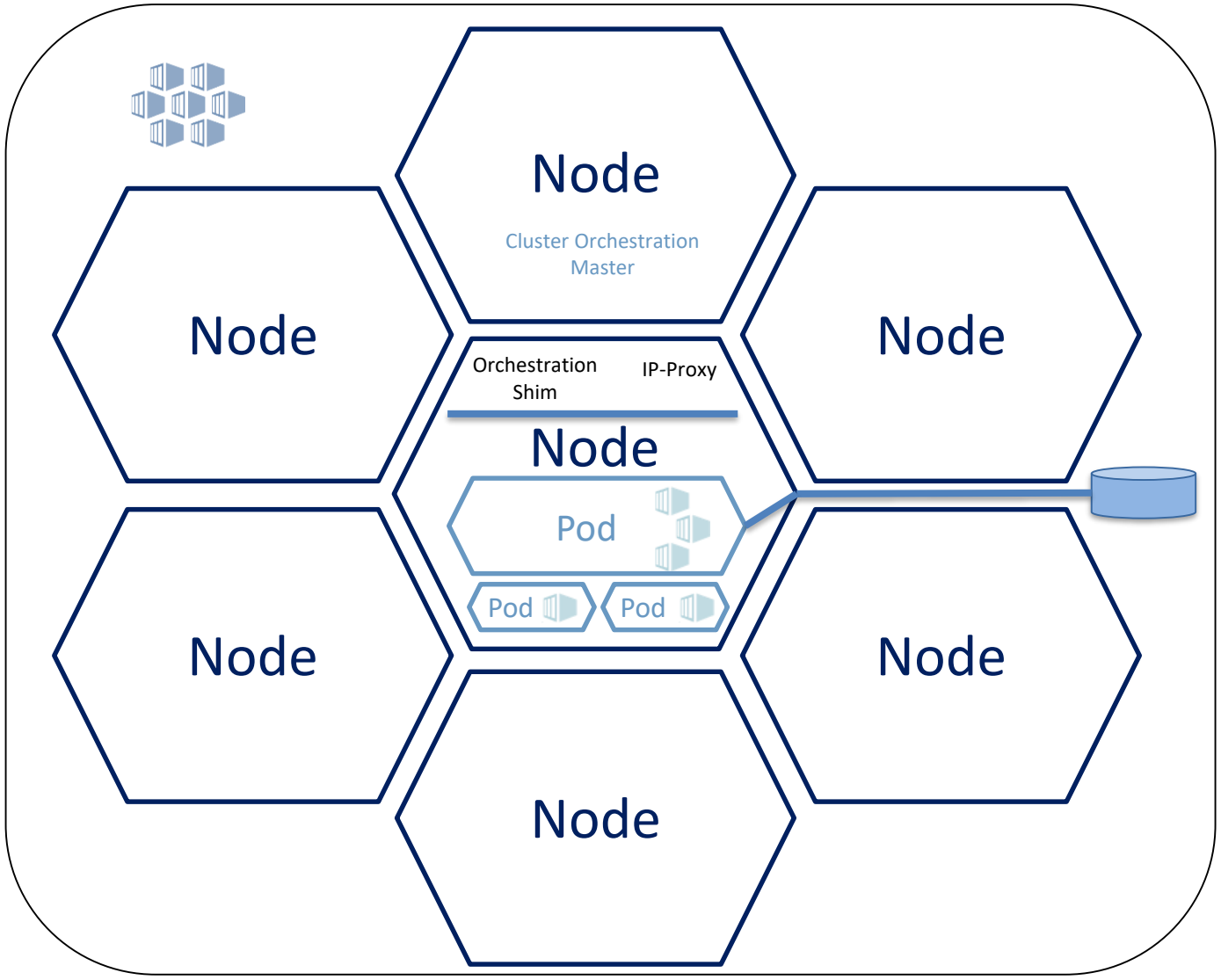
Storage and networking are a consideration

Scale is achieved through multiple containers



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*



Want to learn more…

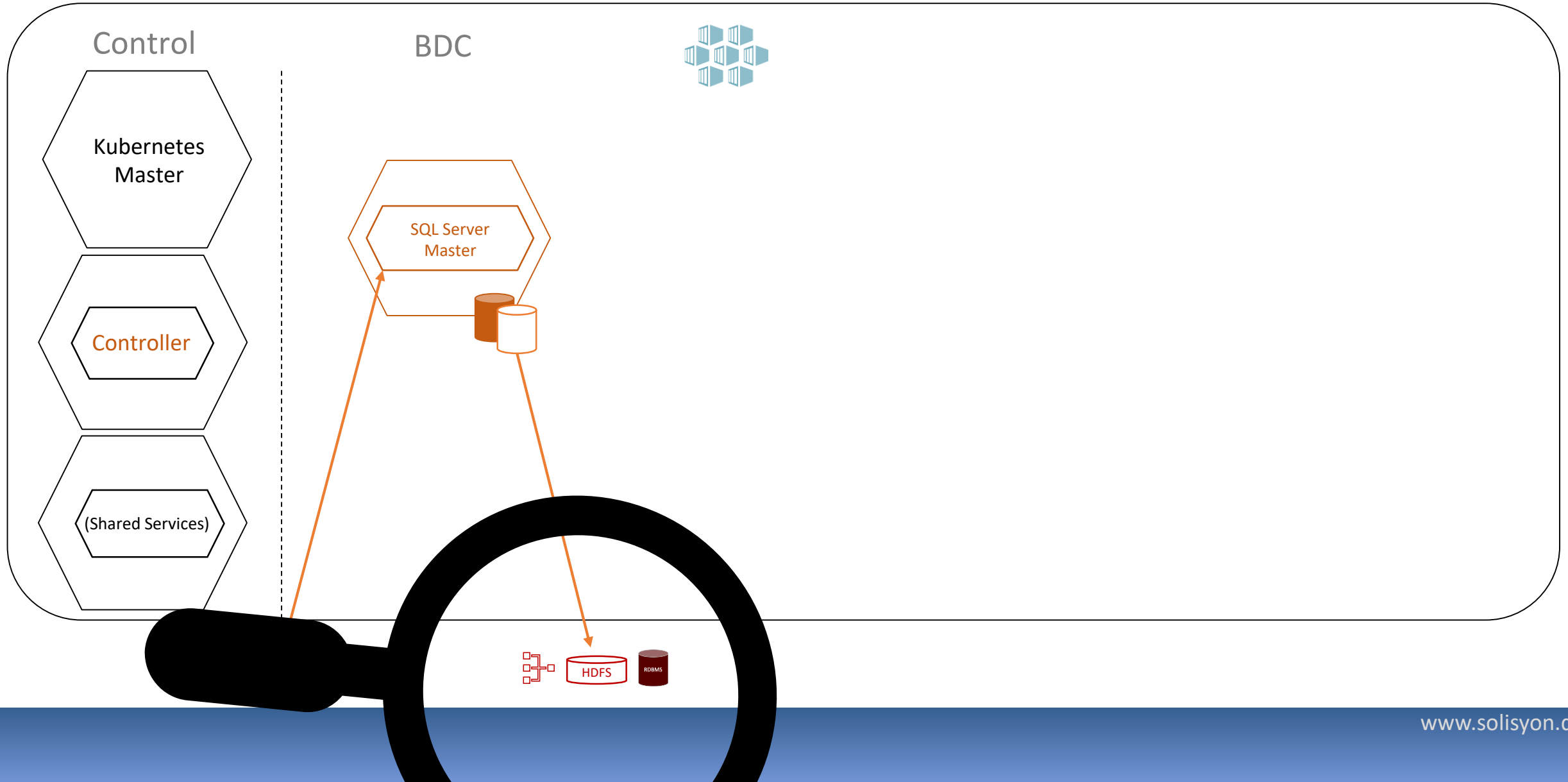
…without all that tech stuff?

https://www.cncf.io/wp-content/uploads/2019/07/The-Illustrated-Childrens-Guide-to-Kubernetes.pdf

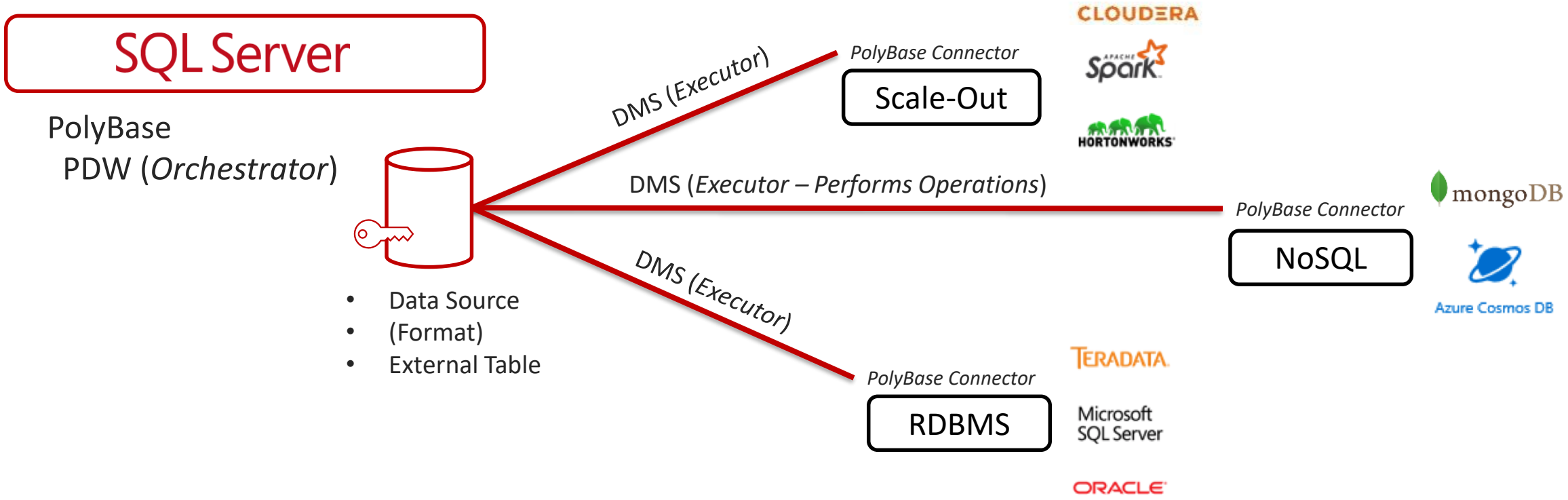
Or talk to this guy...

@nocentino

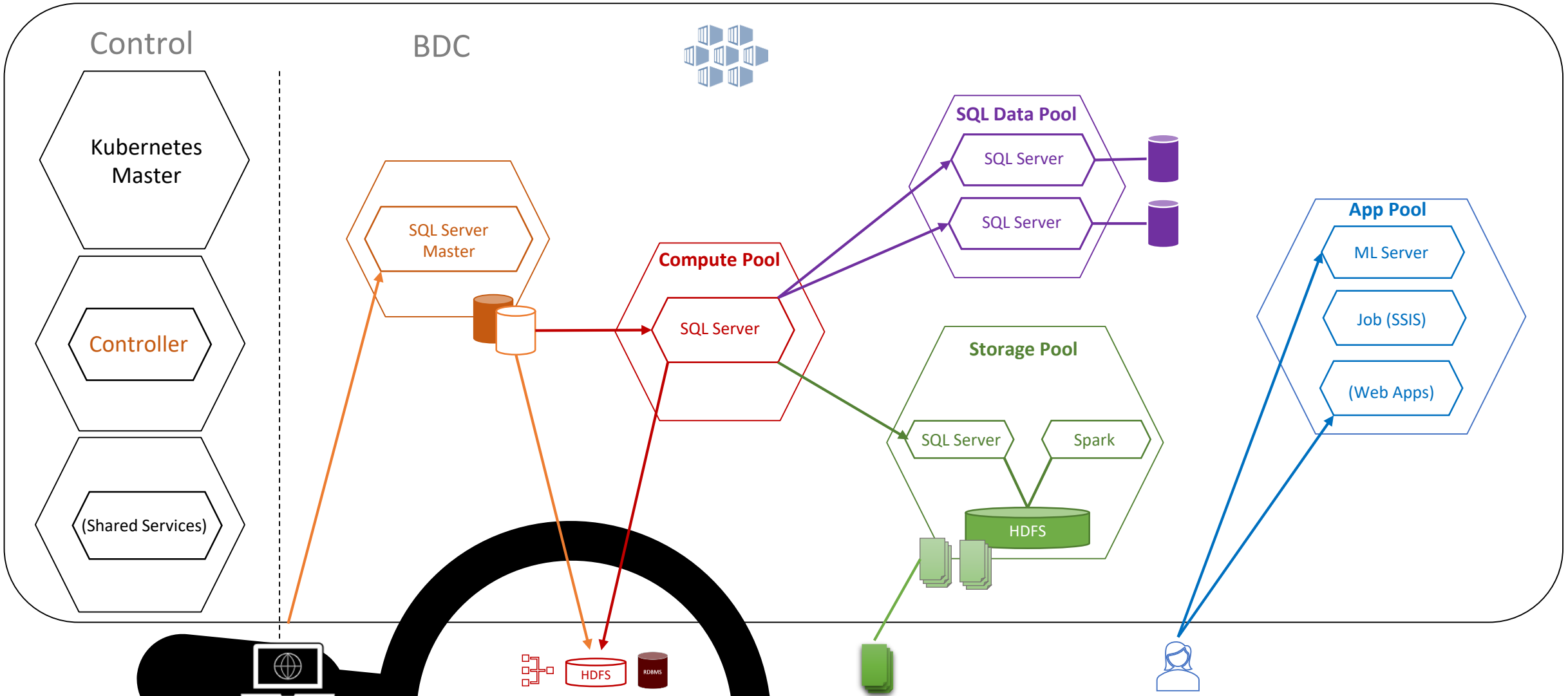
OLTP, Data Virtualization, Data Mart and Big Data



Data Virtualization



OLTP, Data Virtualization, Data Mart and Big Data



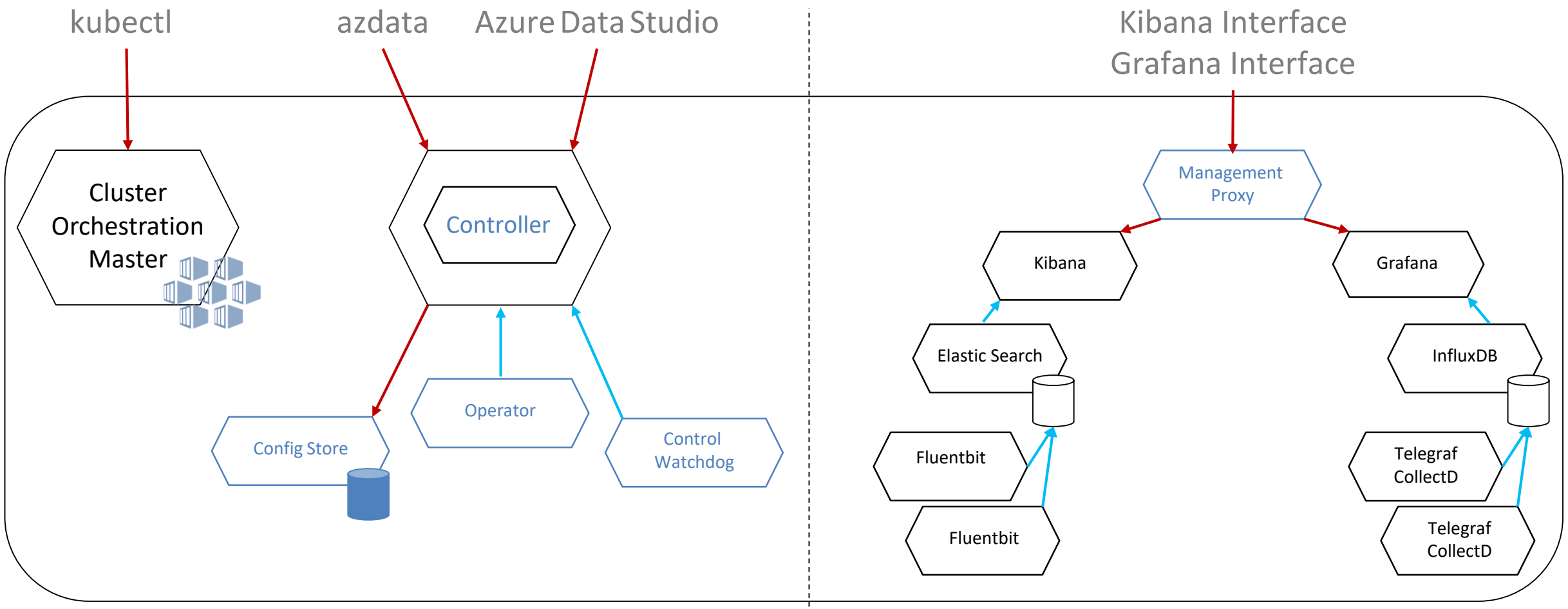
Demo

- PolyBase in SQL 2019
- Storage Pool
- Data Pool
- Spark Queries

SOLISYON

Tools, Management & Monitoring

Managing the Big Data Cluster



SOLISYON

Deployment

How can I get it installed?

PolyBase only

- Get SQL 2019 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 and Data Virtualization Extension

How can I get it installed?

The full package

- Decide on a Kubernetes environment (AKS, kubeadm, …)
- Install Azure Data Studio and Data Virtualization Extension
- Install Prerequisites*
- Deploy the cluster using azdata/Azure Data Studio

* Prerequisites

```
Set-ExecutionPolicy Bypass -Scope Process -Force; iex ((New-Object System.Net.WebClient).DownloadString('https://chocolatey.org/install.ps1'))
choco install notepadplusplus -y
choco install 7zip -y
choco install curl -y
choco install sqlserver-cmdlineutils -y
choco install azure-cli -y
choco install azure-data-studio -y
choco install python3 -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 kubernetes-cli -y
pip3 install Kubernetes
choco install visualcpp-build-tools -y
pip3 install -r https://aka.ms/azdata
```

Install Sample Data

.\bootstrap-sample-db.cmd

USAGE: .\bootstrap-sample-db.cmd <CLUSTER_NAMESPACE> <SQL_MASTER_IP> <SQL_MASTER_SA_PASSWORD>

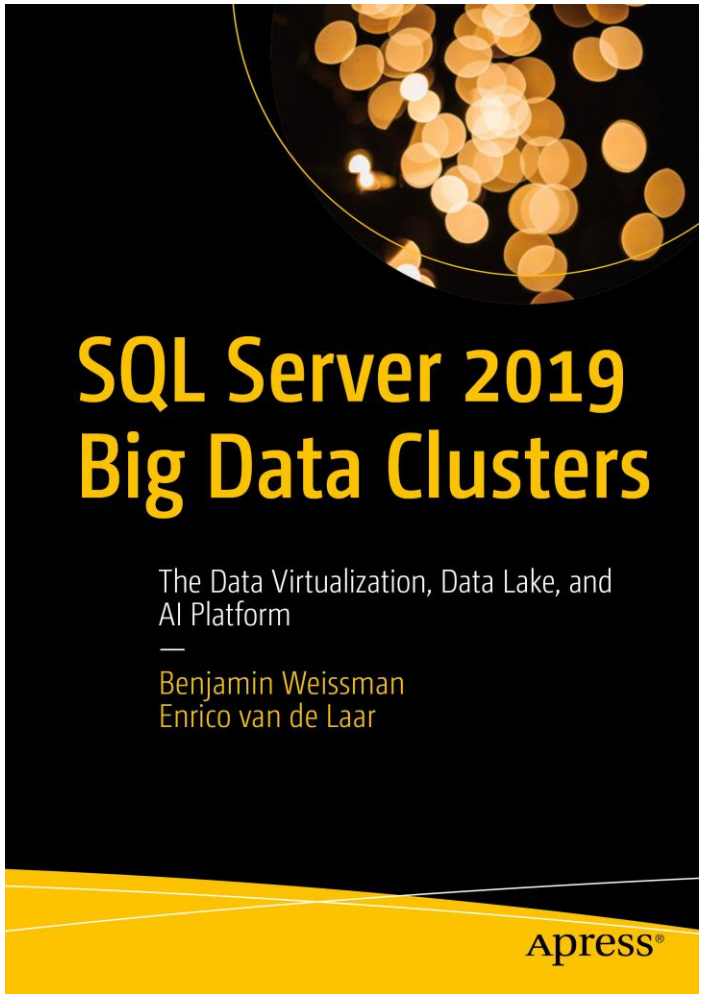
<BACKUP_FILE_PATH> <KNOX_IP> [<KNOX_PASSWORD>]

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

Ben Weissman

@bweissman
b.weissman@solisyon.de

linkedin.com/in/weissmanben/

Thank You For Your Time!