# FANS
# Ulike dataplattformer for SAS Viya

Hvilke muligheter finnes i SAS Viya og hva kommer?

Jonas Lie-Nielsen

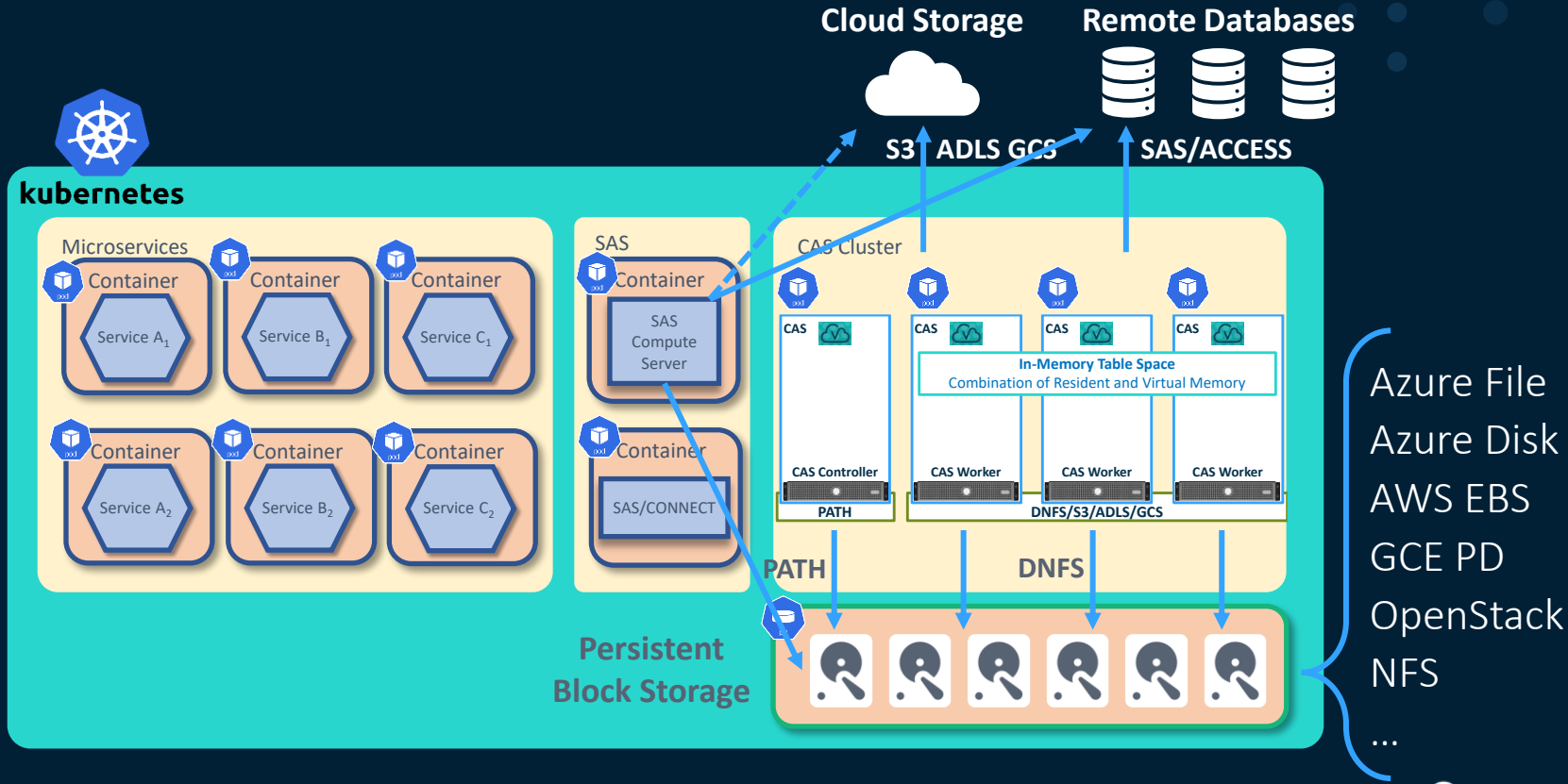
# How can SAS viya enable the journey into the cloud for your data and analytics portifolio?

# SAS Viya in the cloud data landscape

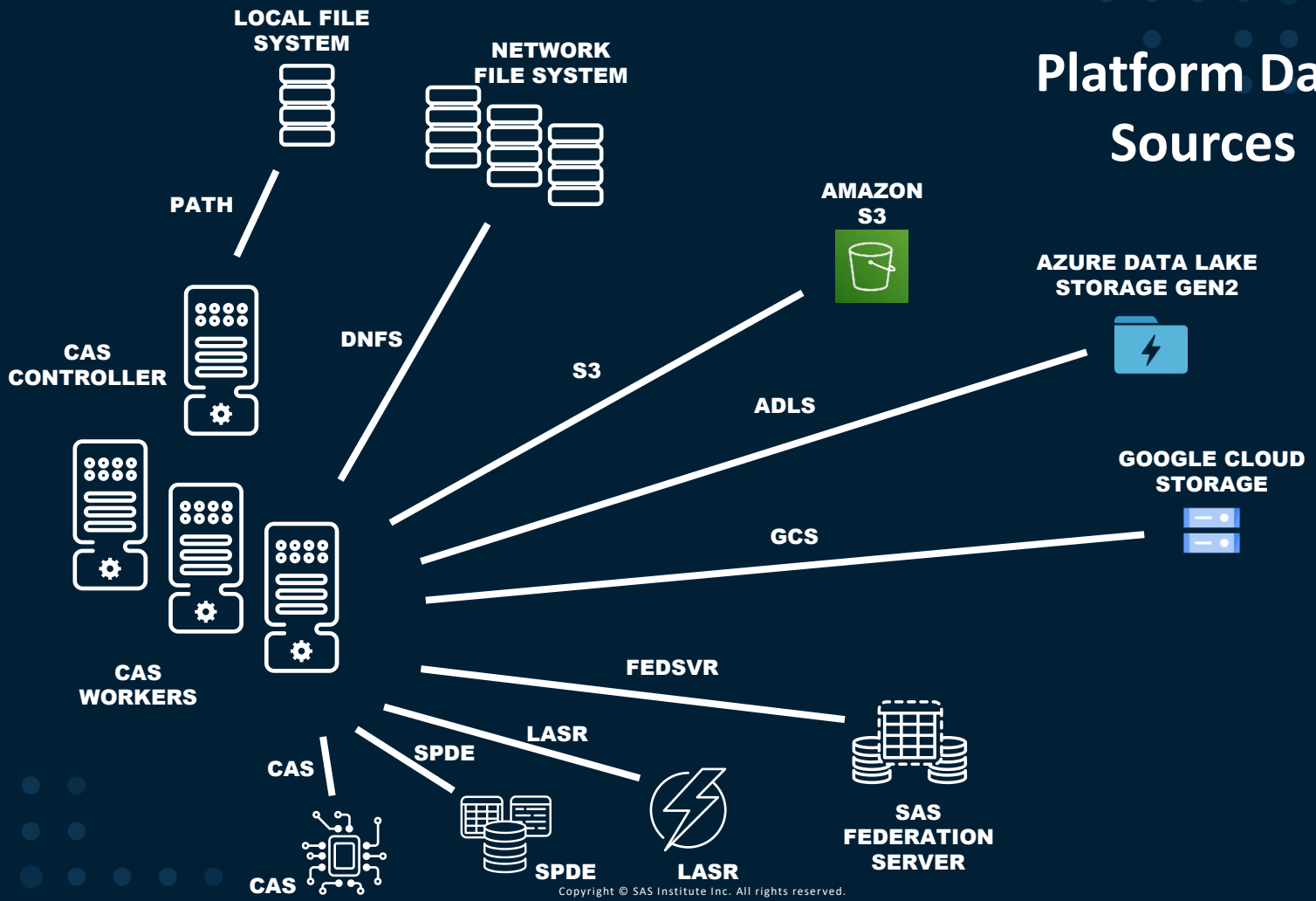
## Type of data connections

- Platform Data source
  - A file system utilized by Viya as its internal data storage
  - Can be configured both in parallel to all workers and singlethreaded to the controller
  - DNFS, S3, ADSL, GCS, CAS ++
- CAS data connector
  - Database connectors working between CAS and the db
  - Three modes
    1. Single-threaded
    2. Multi-node
    3. Parallel
- SAS Compute libname engine – access as for sas9

# CAS Data Architecture in Containers/Kubernetes



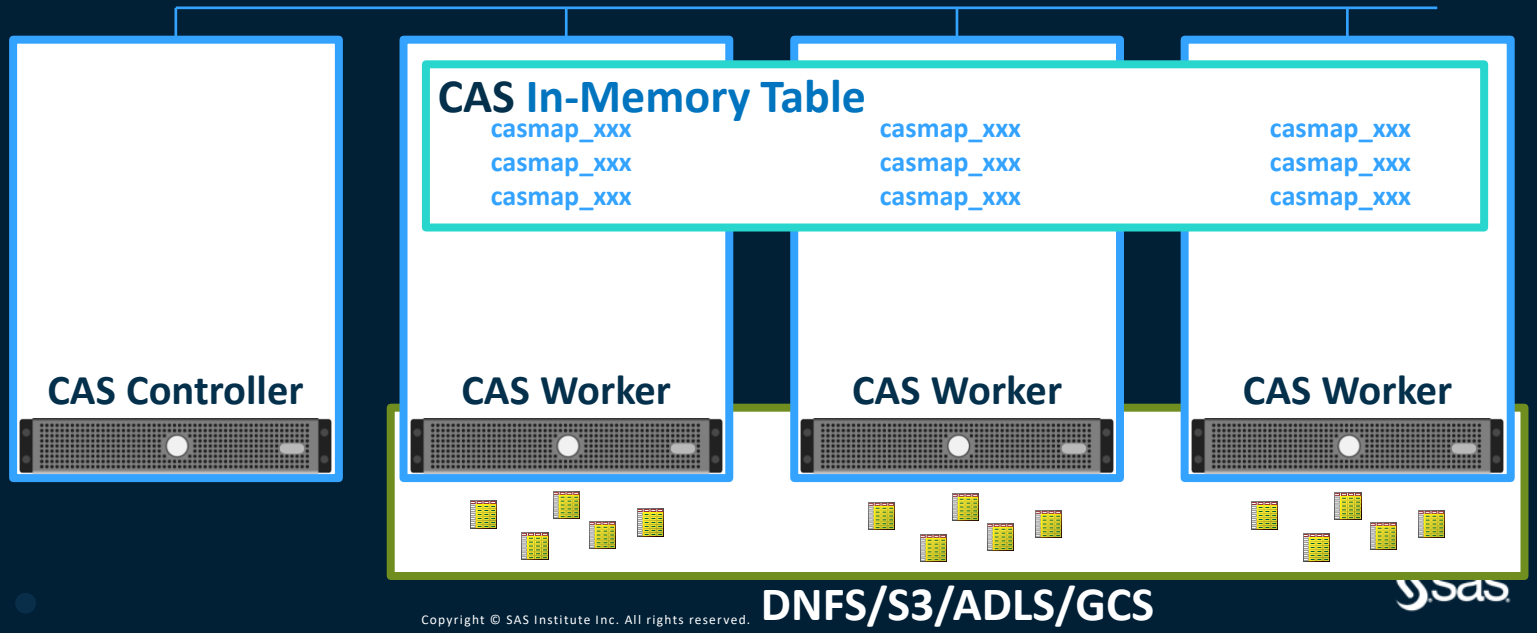
# Platform Data Sources



# Server Side – Parallel Load – Platform Data Source

## DNFS / S3 / ADLS / GCS

```
caslib casdnfs path="/mnt/dfs/" type=dnfs ;
caslib cass3 datasource=(srctype="s3",bucket="sas-gle",...) ;
caslib myadls datasource=(srctype="adls",accountname="az",filesystem="/data",...) ;
```



# CAS Platform Data Sources / File Types / Loading

| File Type / CASLIB Type | PATH | DNFS | S3 | ADLS | GCS |
|---|---|---|---|---|---|
| SASHDAT | S | P | P | | |
| SAS7BDAT | S[1] | | | | |
| CSV[2] | S[3,6] | P[3,6] | P[3,6] | P[3] | P |
| Parquet | S | P | P | P | P |
| ORC | S | | | S | |
| Image | S | P[4] | P[4] | | |
| Document | S | P[4] | P[4] | | |
| Audio | S | P[4] | P[4] | | |
| Video | S | P[4] | P[4] | | |
| Others | S | | S[5] | | |

S serial

P parallel



*See notes for details*

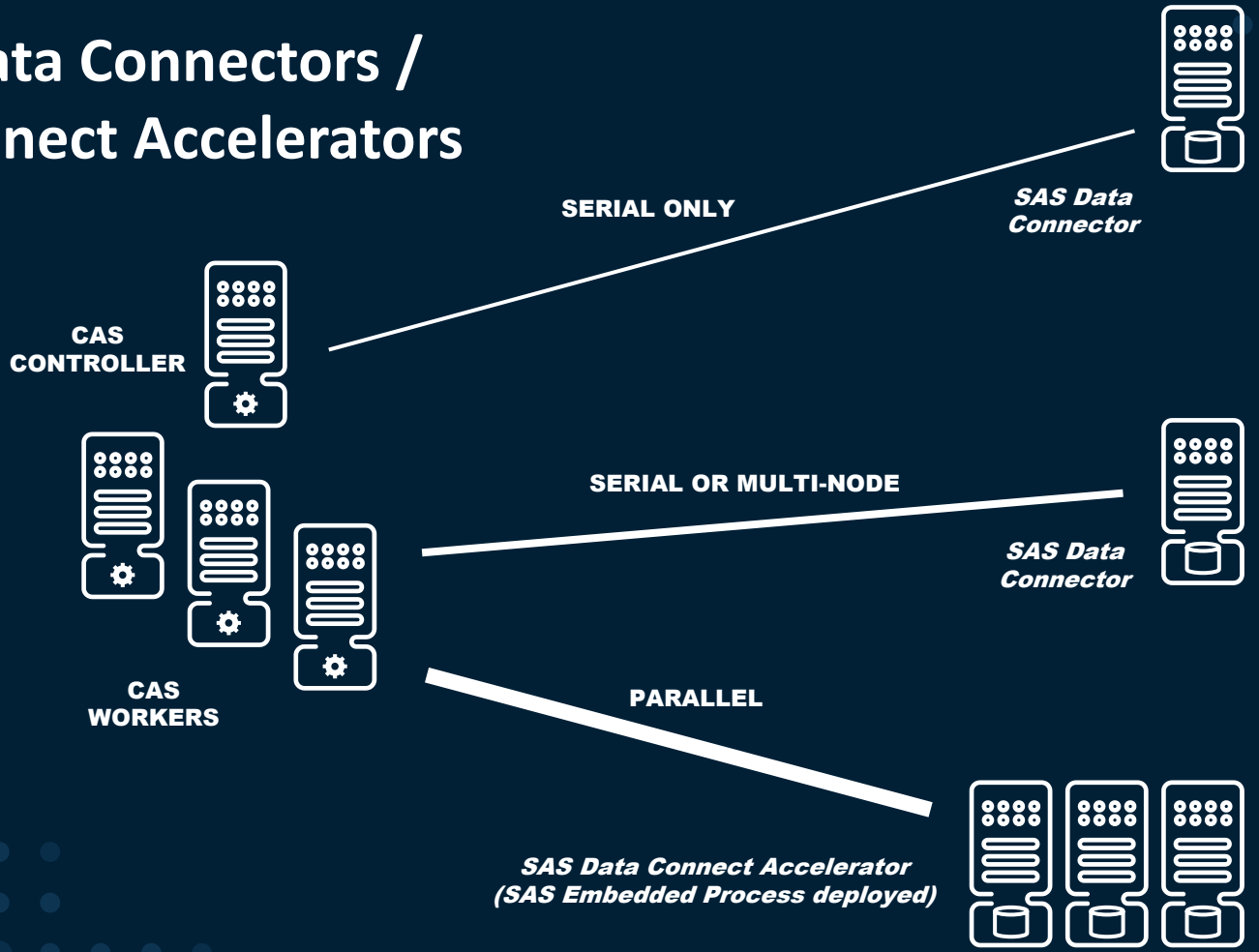
# CAS Platform Data Sources / File Types / Saving

| File Type / CASLIB Type | PATH | DNFS | S3 | ADLS | GCS |
|---|---|---|---|---|---|
| SASHDAT | S | P | P | | |
| SAS7BDAT | S | | | | |
| CSV | S | P | P | S | S |
| Parquet | S | P | P | P | P |
| ORC | S | | | S | |
| Others | S | | S[1] | | |

S serial

P parallel

# Data Connectors / Connect Accelerators




PC FILES

AMAZON REDSHIFT
DB2
GOOGLE BIGQUERY
GREENPLUM
HADOOP HIVE
IMPALA
JDBC
MS SQL SERVER
MONGODB
MYSQL
NETEZZA
ODBC
ORACLE
POSTGRESQL
SALESFORCE
SAP HANA
SNOWFLAKE
SPARK
TERADATA
VERTICA
YELLOWBRICK

HADOOP HIVE
SINGLESTORE*
SPARK
TERADATA

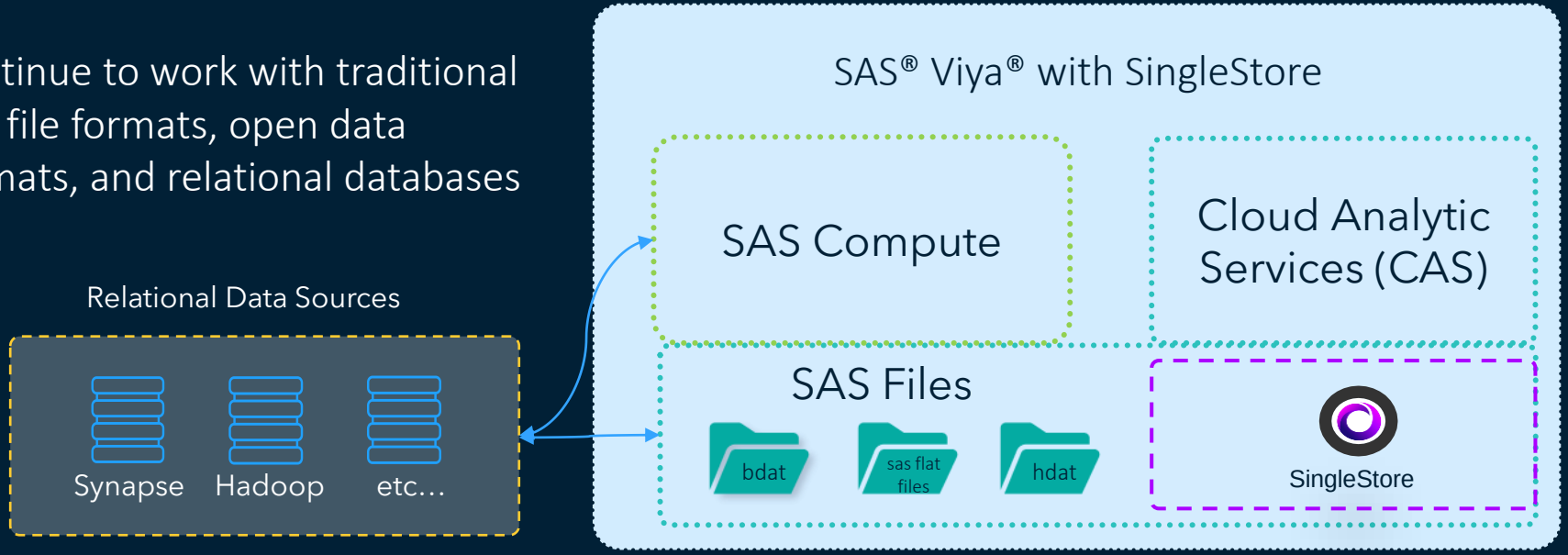
*SAS VIYA WITH SINGLESTORE

# Cloud Platforms Connectivity Support


- Oracle Cloud Platform – Oracle, Oracle MySQL
- Teradata Cloud
- Vmware – Teradata Vantage

# SAS® Viya® with SingleStore

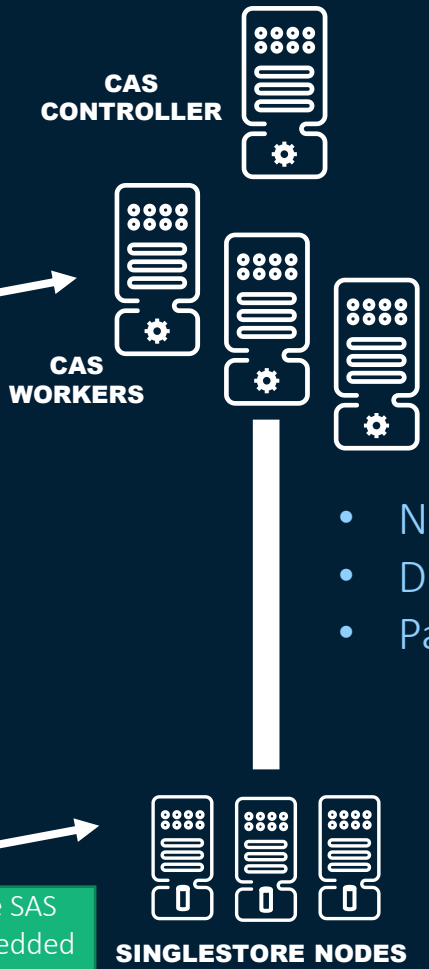
Continue to work with traditional SAS file formats, open data formats, and relational databases




- CAS Integrated with SingleStore via Embedded process
- Pushes analytical actions down to the data

# CAS S2 Data Access

CAS acts as the computation engine

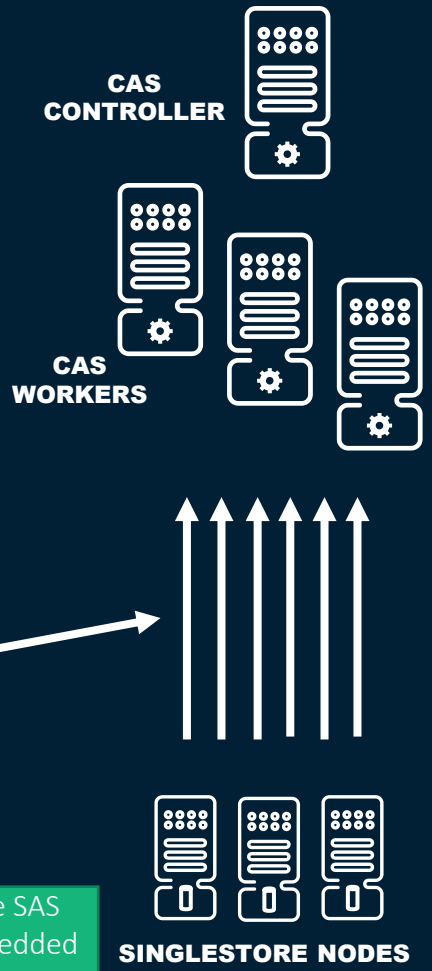



S2 acts as the data server

The SAS Embedded Process (EP)

- Native Data Access
- Direct Reads – No Staging
- Pass-Through:
  - WHERE Clause
  - Column Subsetting
  - Computed Columns
  - Formats
  - *More on the way

# CAS S2 Data Access




SingleStore can also act as a traditional SAS Viya massively parallel load source

The SAS Embedded Process (EP)

This is also how all other parlell data connectors work

# SAS In-Database Technologies

- All In-Database Technology products are included in SAS Viya Programming and SAS Viya Enterprise
- Available as optional add-ons in the other Viya 4 Packages (see packaging docs for details)
- SAS Viya Programming and SAS Viya Enterprise will keep on including **ALL** SAS In-Database Technologies

| | From Q2 2022 |
|---|---|
| SAS In-Database Technologies | SAS In-Database Technologies for Teradata |
| | SAS In-Database Technologies for Cloudera Data Platform |
| | SAS In-Database Technologies for Hadoop Cloud Services |
| | SAS In-Database Technologies for Azure Synapse |
| | SAS In-Database Technologies for Databricks |

# SAS/ACCESS Interface to Snowflake

## Supported Features

| Category | Feature | Supported |
|---|---|---|
| In-Database Processing | SQL Passthrough (PROC SQL) | Yes |
| | Procedures | Yes |
| | Processing with PROC FEDSQL and PROC DS2 | Yes |
| | SQL Functions | Yes |
| Performance | Bulk load data to Snowflake (from SAS Compute Server) | Yes |
| | Bulk unload data from Snowflake (to SAS Compute Server) | Yes |
| CAS Related | Serial Data Transfer (including multi-node support) | Yes |
| | Parallel Data Transfer | No |

# Connectivity Domain

## Connect Everywhere

# Released

**2021.2.6**
**SAS/ACCESS-Data Connectors**
- Hadoop – Validated support for AWS EMR 6.3

**In-Database technologies**
- Following now available as add-ons
  - SAS In-Database Technologies for Databricks
  - SAS In-Database Technologies for Azure Synapse Analytics
  - SAS In-Database Technologies for Cloudera
  - SAS In-Database Technologies for Hadoop Cloud Services
  - SAS In-Database Technologies for Teradata

**2022.1.1**
**SAS/ACCESS-Data Connectors**
- **Databricks on Azure – Improved bulk loading**

**2022.1.3**
**Singlestore**
- **Support for Auth Domain and CAS attribute**

**Catalog**
- **Support for crawling Created/Modified Date**

**2022.1.4**
**SAS/ACCESS-Data Connectors**
- **CData Drivers for Spark and Databricks included by default in SAS/ACCESS Spark**
- **CData driver for Hadoop included by default in SAS/ACCESS Hadoop**
- **Support for GCS storage added for Parquet compute engine**

**2022.09**
**SAS/ACCESS-Data Connectors**
- Support for AWS S3 storage added for Parquet compute engine

**2022.10**
**SAS/ACCESS-Data Connectors**
- Spark - Support for AWS EMR 6.4
- 9.4M8 – Validation work
- Parquet – Compression types support on compute.

**2022.11**
**SAS/ACCESS-Data Connectors**
- Hadoop – Support for Azure HDInsight 5
- Hadoop – Support for Google DataProc
- Spark – Support for HDInsight 5 Spark

# Connectivity Domain

## Connect Everywhere

## Now

**In-Database technologies**
- SAS 9.4M8 Testing
- In-DB Synapse – Synapse AAD Support
- In-DB Cloud Hadoop Services – HDInsight

**Data Connectors & Open File Formats**
- CosmosDB – SQL API - 2023.06
- SAP IQ – SAS/ACCESS – 2023.02 – DC 2023.03
- Informix - SAS/ACCESS & Data Connectors 2023.03
- SingleStore – SAS/ACCESS – 2023.03
- Add DBTYPE to all Data Connectors – 2023.02
- Google Big Query – Performance Improvements – 2023.03
- Compute - Parquet files - Logical Data Types Support - 2023.03
- S/A&DC Teradata - Implement Single Sign On – 2023.04

**Hybrid Cloud Management**
- Cloud Data Exchange on Viya 4 – 2023.03
- Common Connectivity Dialog (C3) Release 1 - 23Q1

## Next

**In-Database technologies**
- In-DB Cloud Hadoop Services – Support AWS EMR 5/6, Google DataProc
- Parallel Load Azure Synapse Analytics
- In-DB - Investigate potential options for In-DB functionality in Snowflake

**Data Connectors & Open File Formats**
- Spark – Databricks AAD Support
- SAP IQ – Data Connector & Bulk Load support
- SingleStore – SAS/ACCESS – Performance Enhancements
- Compute - Parquet Files - Support for ADLS2 -2023.05

**Hybrid Cloud Management**
- Cloud Data Exchange Release 2
- Common Connectivity Dialog (C3) Release 2

## Future

**In-Database technologies**
- Kerberos Support CDP7.2 Public Cloud & CDP7.1 Private
- Code & DQ Accelerators – All –In-DB Products

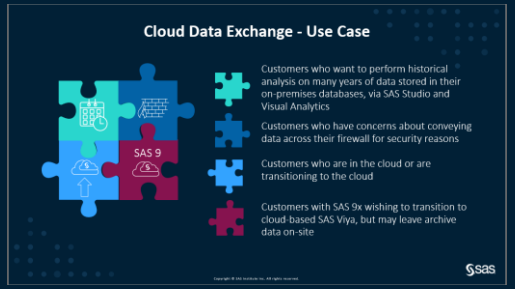
**Data Connectors & Open File Formats**
- Parquet files support for partitioned data. – 2023.08
- Spark – Support for EMR, Dataproc
- GCP – Support native Authentication to supported GCP Data Stores
- Vertica – Update with Vendor ODBC driver
- MongoDB – Additional features
- Salesforce – Support Update/Delete

**Hybrid Cloud Management**
- Cloud Data Exchange Release 3
- Common Connectivity Dialog (C3) Release 3

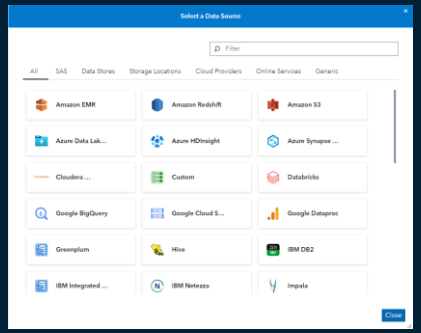
# Connectivity Domain

## CDE/C3 Updates

- Cloud Data Exchange
  - On Track for 2023.03 delivery
  - GTP (Ivor Moan) demo & Presentation




- C3
  - Initial delivery for 2023.03
  - Shared configuration of caslibs and libnames
  - Will be integrated into Studio and Data Explorer
    - REACT versions:
      - Data Explorer Mid 2023
      - Studio End 2023



# Main Capabilities

## Cloud Data Exchange


### Local Data Access from Cloud Deployment

Provides the means to expose a customer's local data to a community of remote data engineers, data analysts, and data scientists

### On Premise Firewall Friendly

Negotiates the customer's on-premise firewall securely and responsibly

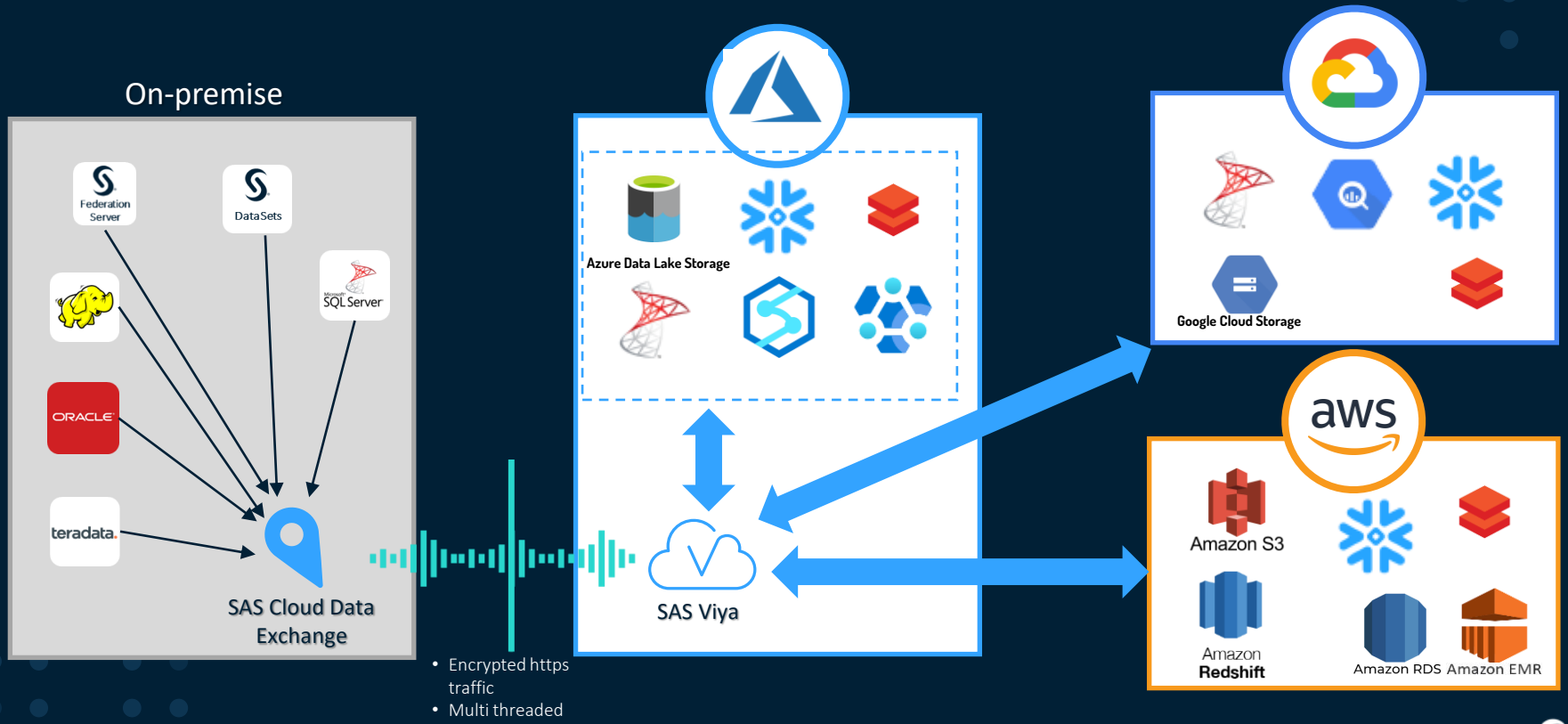
### Support for High-Volume Data Transfers

Performs high-volume data transfers from on-premises to SAS Viya to support big data

### SAS/ACCESS Powered

Leverages many of the SAS/ACCESS data source connectors used by SAS Viya

# Using SAS Viya as a engine for cloud migration




- Encrypted https traffic
- Multi threaded