

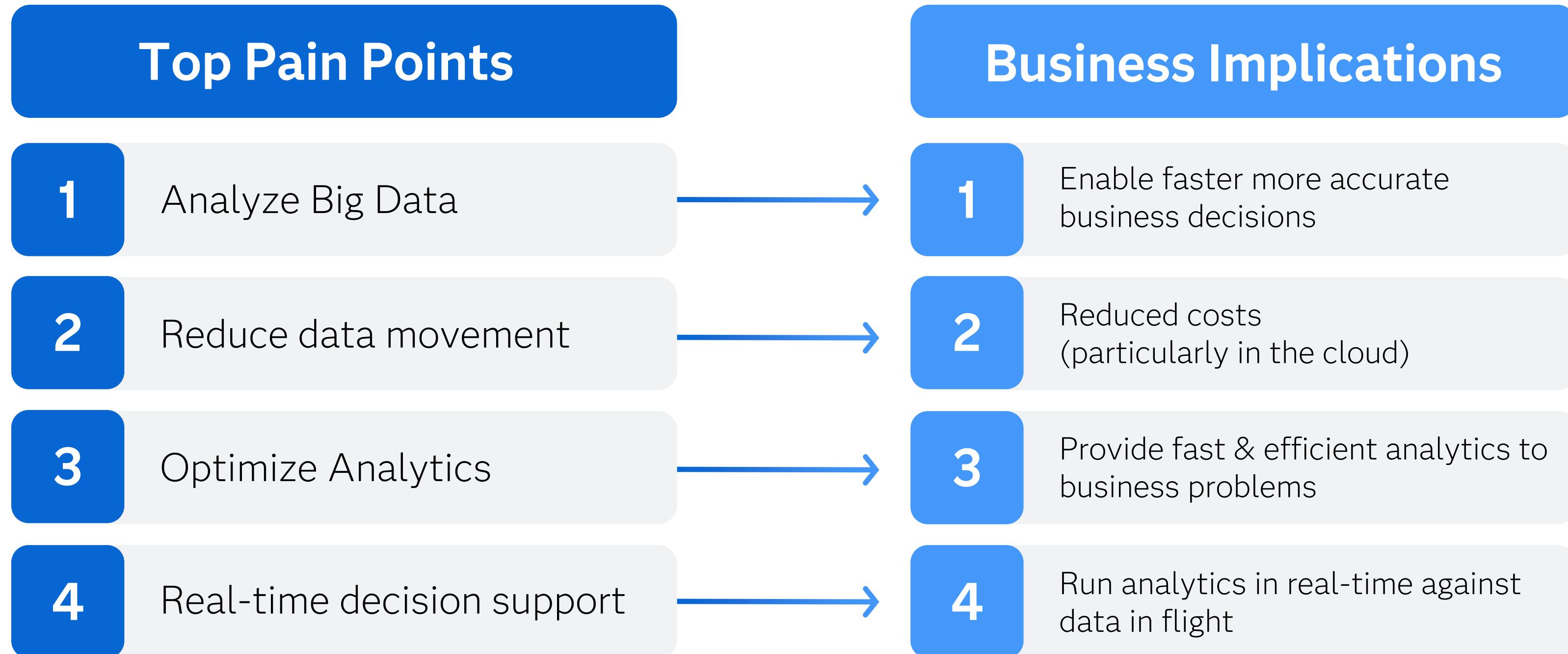
# Mehrwerte der SAS Viya with SingleStore Solution

## 33. Plattform Netzwerk Treffen

Benjamin Walther – Senior Technical Architect  
15.05.2025



# Pain Points & Business Implications



# I asked ChatGPT

## I need help with finding cost pain points in cloud computing for analytical workload

Certainly! Here's a summarized list of key cloud cost pain points for analytical workloads:

### ◆ Top Cloud Cost Pain Points for Analytics

- Overprovisioned Compute

- Idle or oversized VMs/containers lead to wasted resources.

- Excessive Storage Costs

- Unused data and lack of tiering/lifecycle policies inflate storage bills.

- Data Transfer Charges

- Moving data across regions or clouds incurs high egress fees.

- Inefficient Jobs & Pipelines

- Poorly optimized queries and long-running ETL jobs consume more compute.

- Idle Interactive Environments

- Notebooks and tools left running when not in use cost money.

- Uncontrolled Use of Premium Services

- High-cost analytics platforms (e.g., BigQuery, Snowflake) without usage limits.

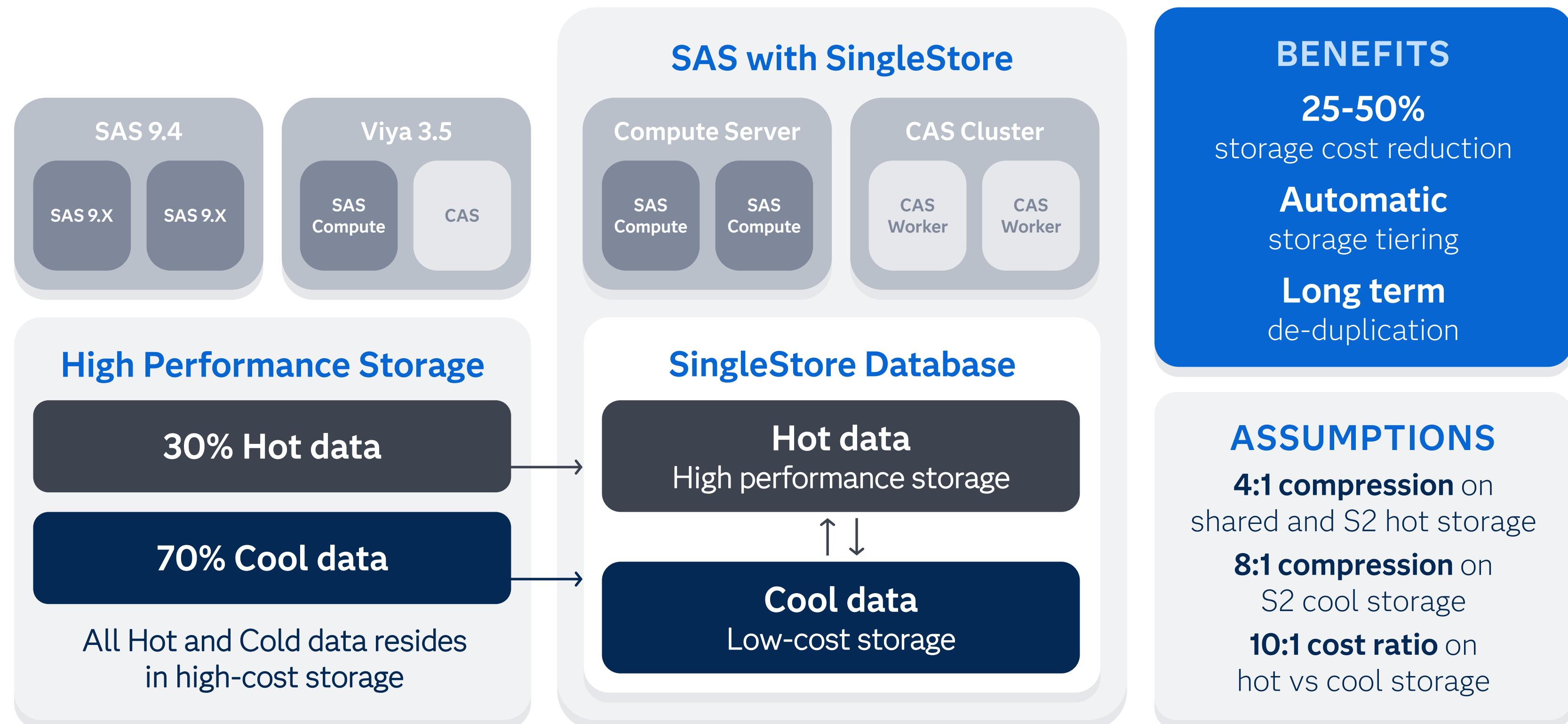
- Limited Cost Visibility

- Lack of dashboards or tagging makes it hard to track and manage spend.

# SAS with SingleStore



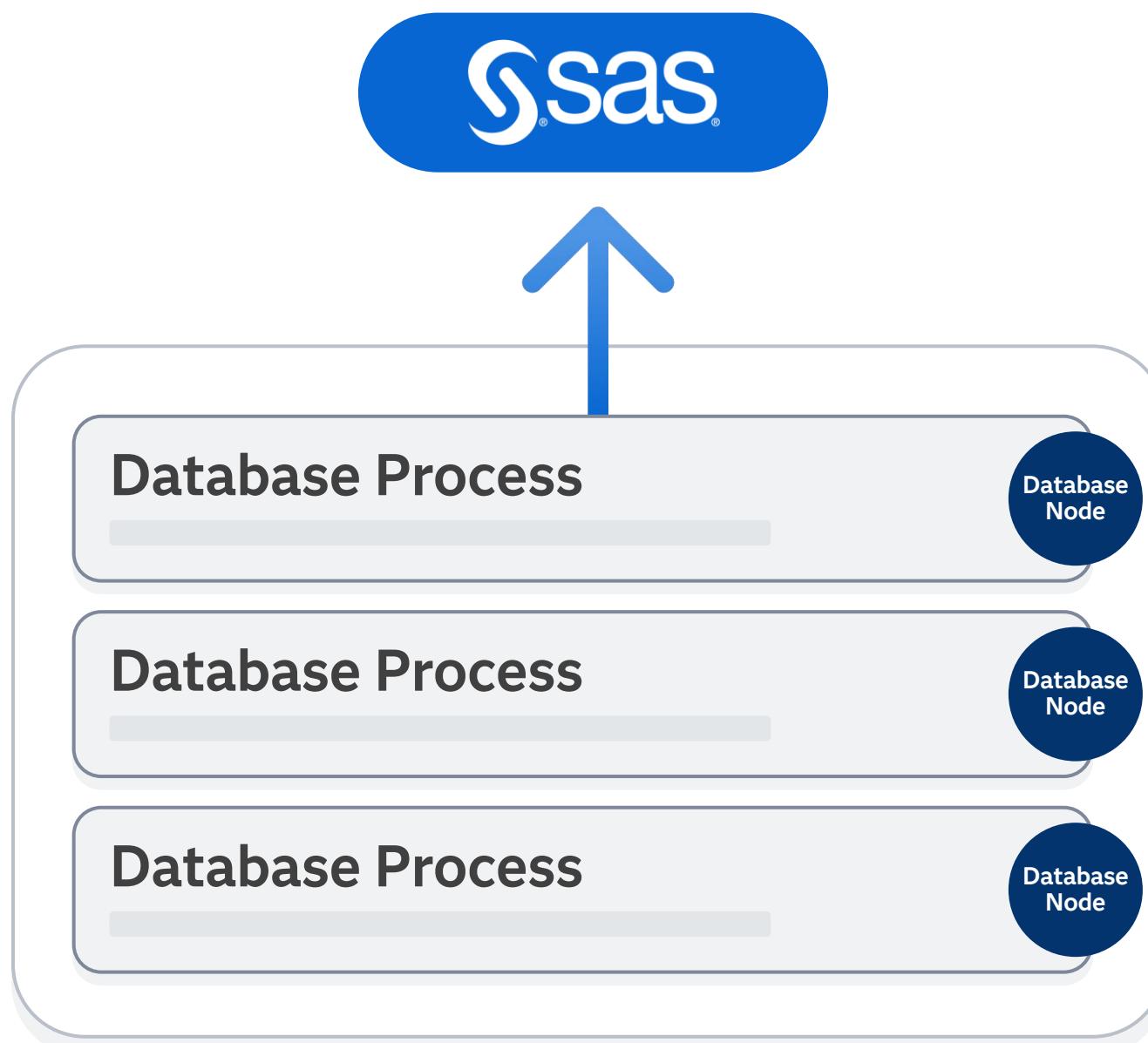
# SAS with SingleStore Data Storage Savings



# What is SAS In-Database (InDb)

## SAS Access

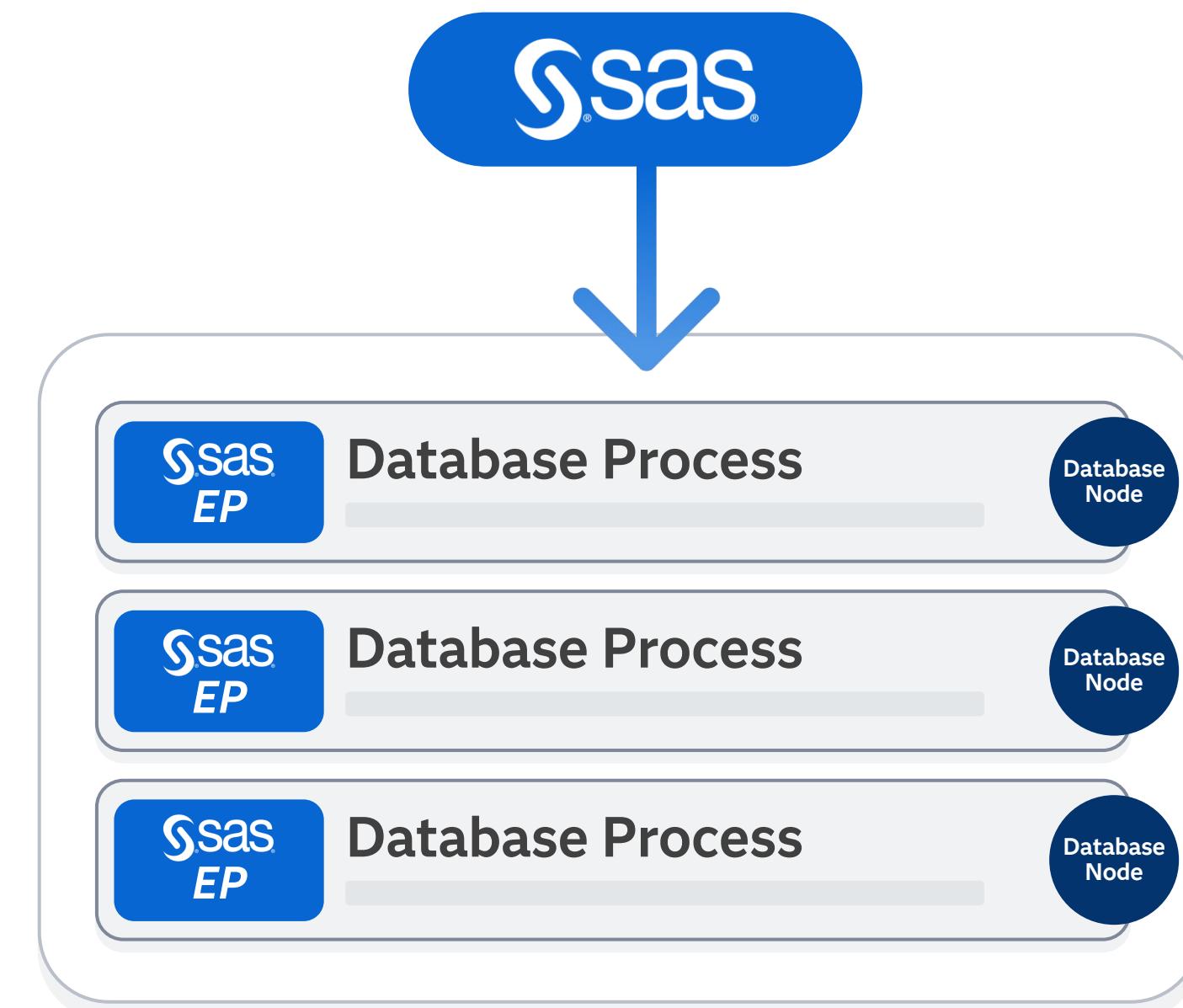
Move data to computation



## SAS InDb

Move computation to the data

Inject the processing power of SAS where the data lives



# SAS Viya Modern Storage Features

Our most advanced data suite



## Advanced Capabilities

### Available:

- Parallel Read/Write
- Scoring Accelerator
- Computed Columns
- Where Processing
- SAS Formats
- Streaming Data
- Data Step Acceleration
- simple.summary

### Next:

- simple.distinct
- simple.orderby
- Publishing Destination for Model Manager
- SCR connection to SingleStore

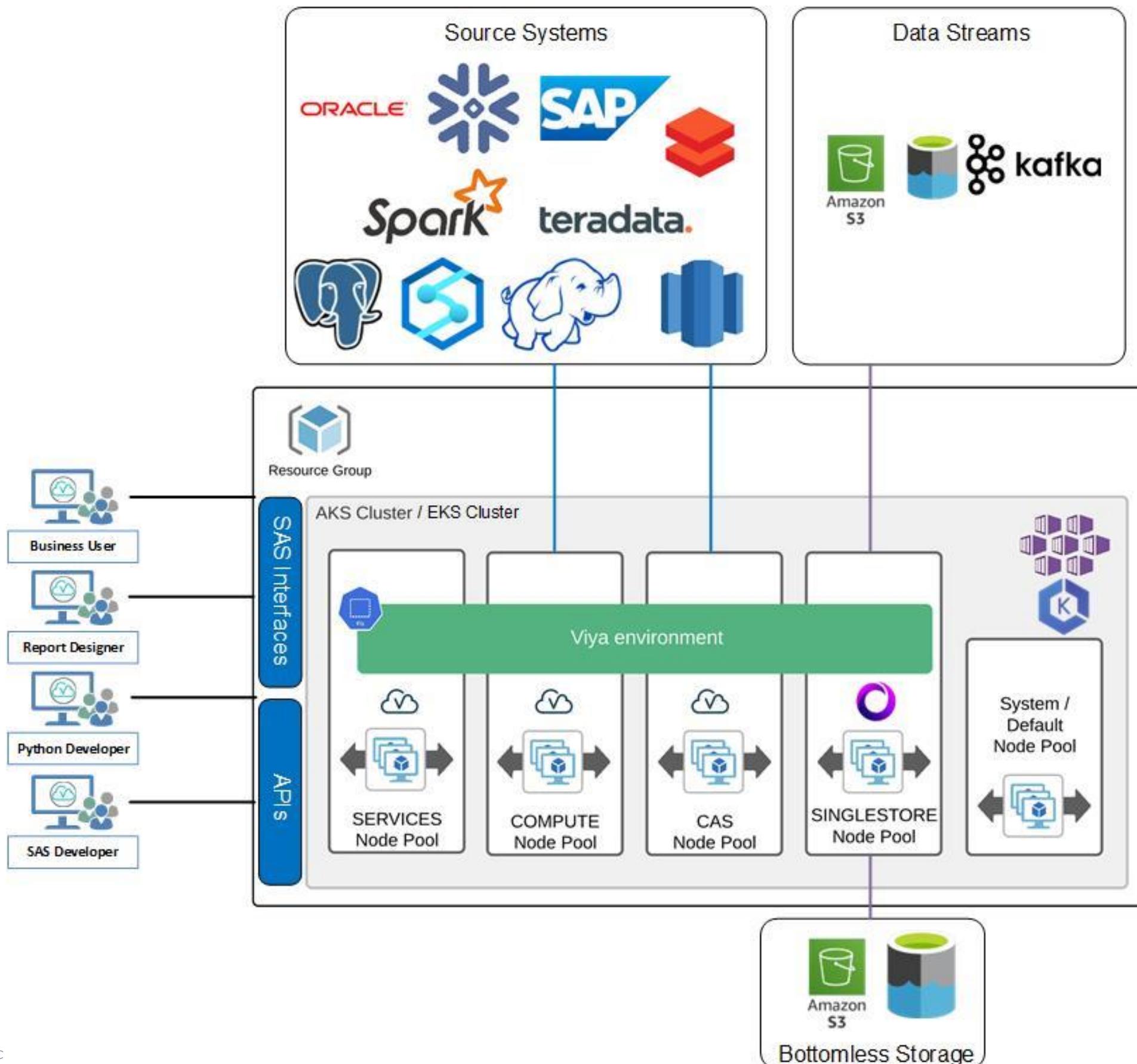


Accelerate Decision Making  
and Reduce Costs

SAS streamlines data access, providing  
flexible and open access to data

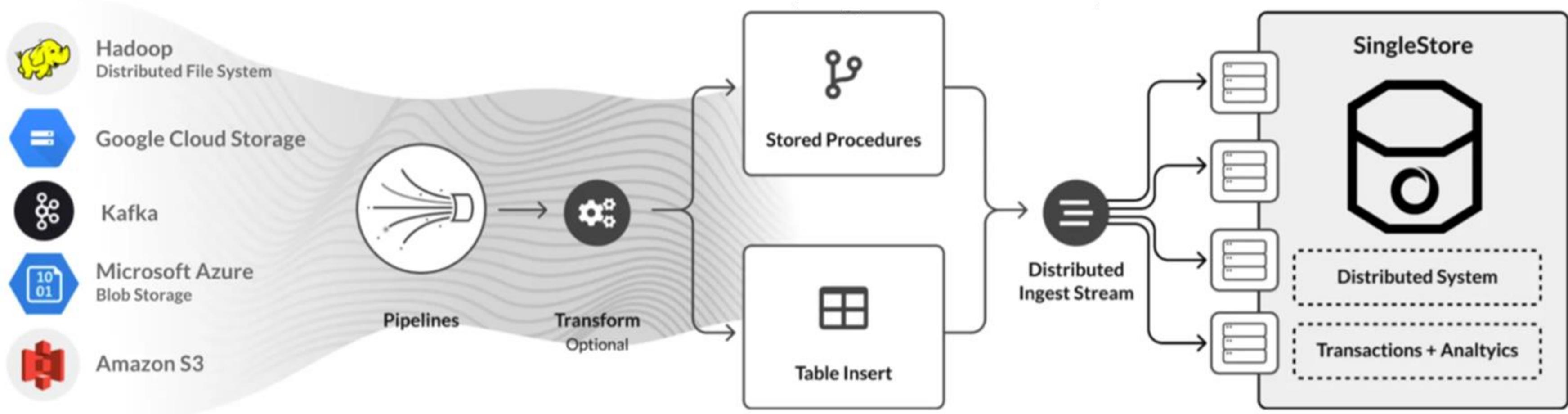
SAS modern data compression,  
automatic object storage tiering and  
analytic performance, can reduce time  
and cost to Analytical insights

# SAS Viya with SingleStore Ecosystem



# Reporting & Real-Time Data Architecture with SingleStore

# SingleStore Pipelines



# SingleStore Pipeline Example

SingleStore Pipelines is a feature that continuously loads data as it arrives from external sources.

- Create Pipeline

```
CREATE PIPELINE pipeline_without_stored_proc  
AS LOAD DATA FS '/data.txt'  
INTO TABLE t;
```

- Create Pipeline ... into Procedure

```
CREATE OR REPLACE PROCEDURE proc(batch QUERY(a INT))  
AS
```

```
BEGIN
```

```
INSERT INTO t(a) SELECT * FROM batch;
```

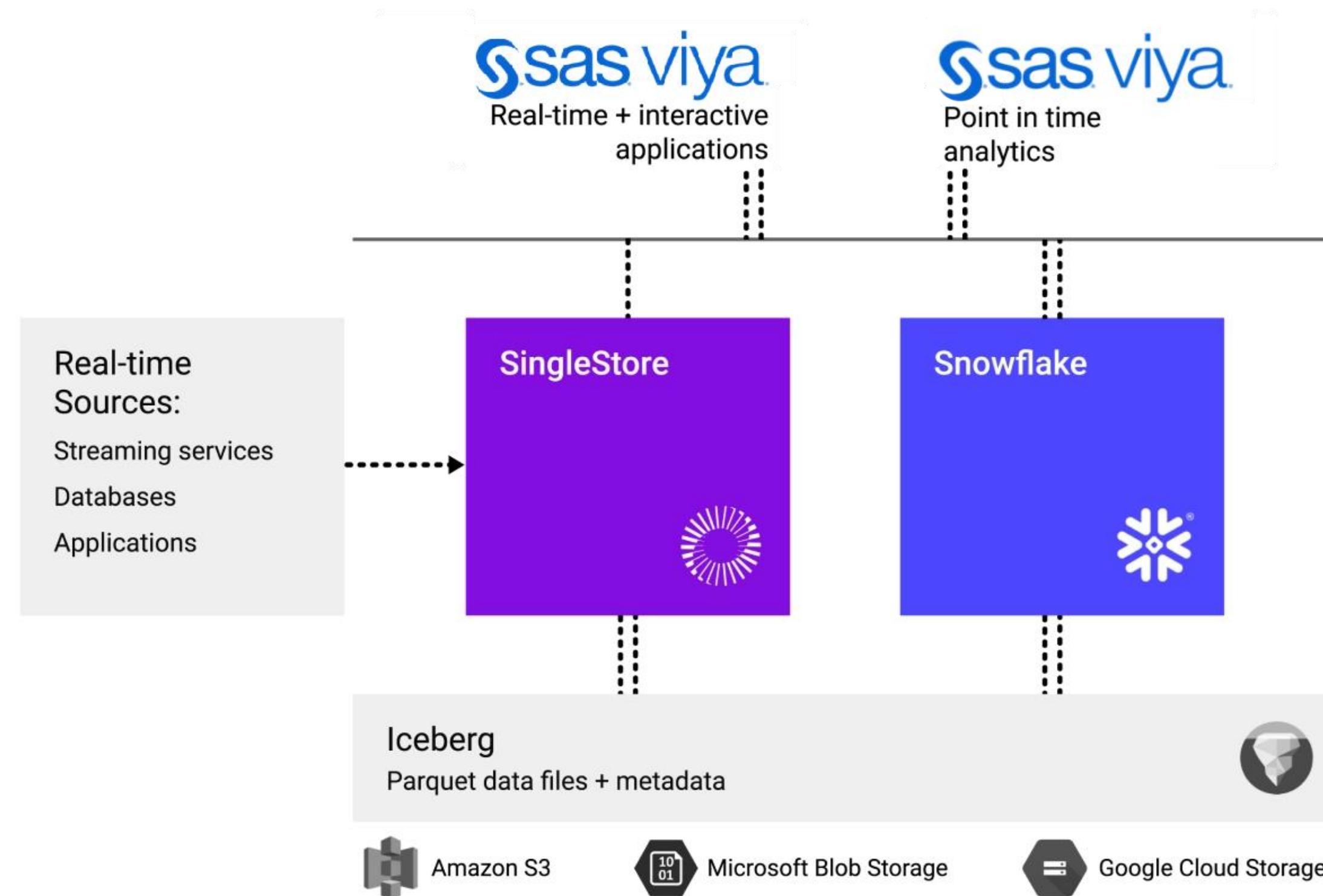
```
END //
```

```
DELIMITER ;
```

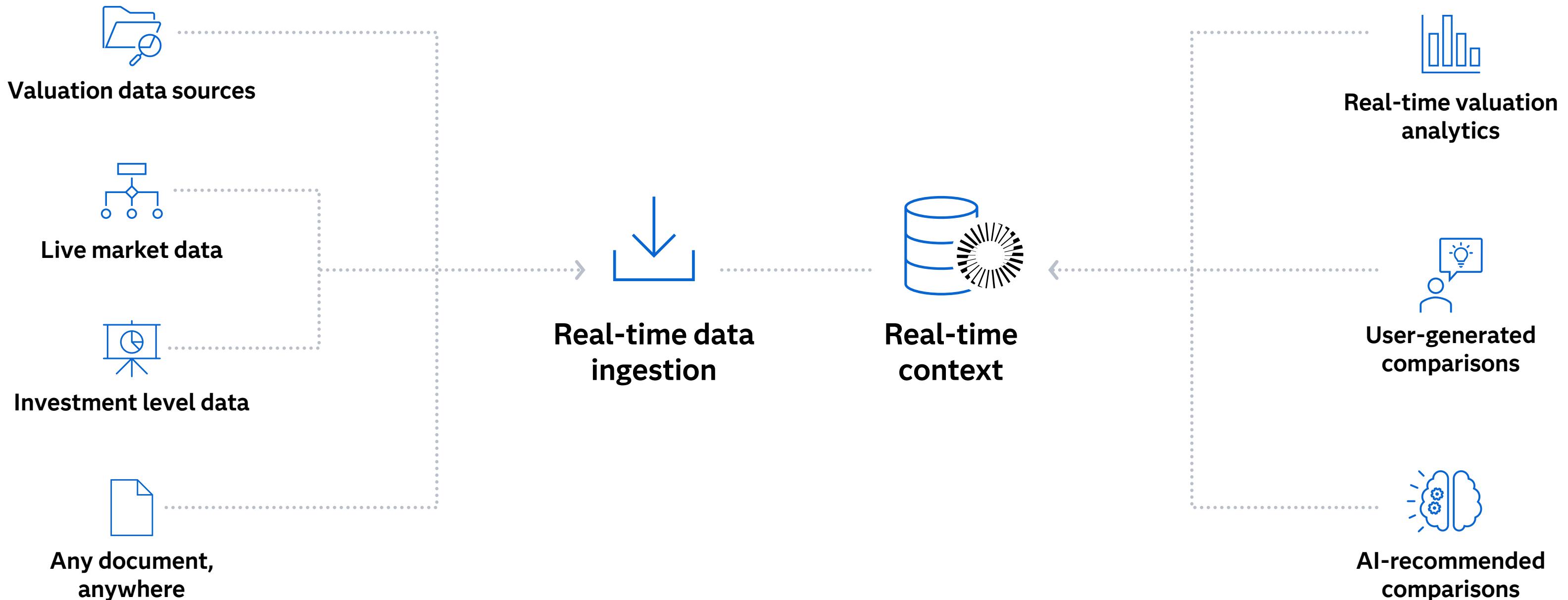
```
CREATE PIPELINE pipeline_using_stored_proc  
AS LOAD DATA FS '/data.txt'  
INTO PROCEDURE proc;
```

# SingleStore Iceberg Ingest

The benefits of adopting SingleStore's zero ETL solution for Iceberg extend its technical efficiencies.

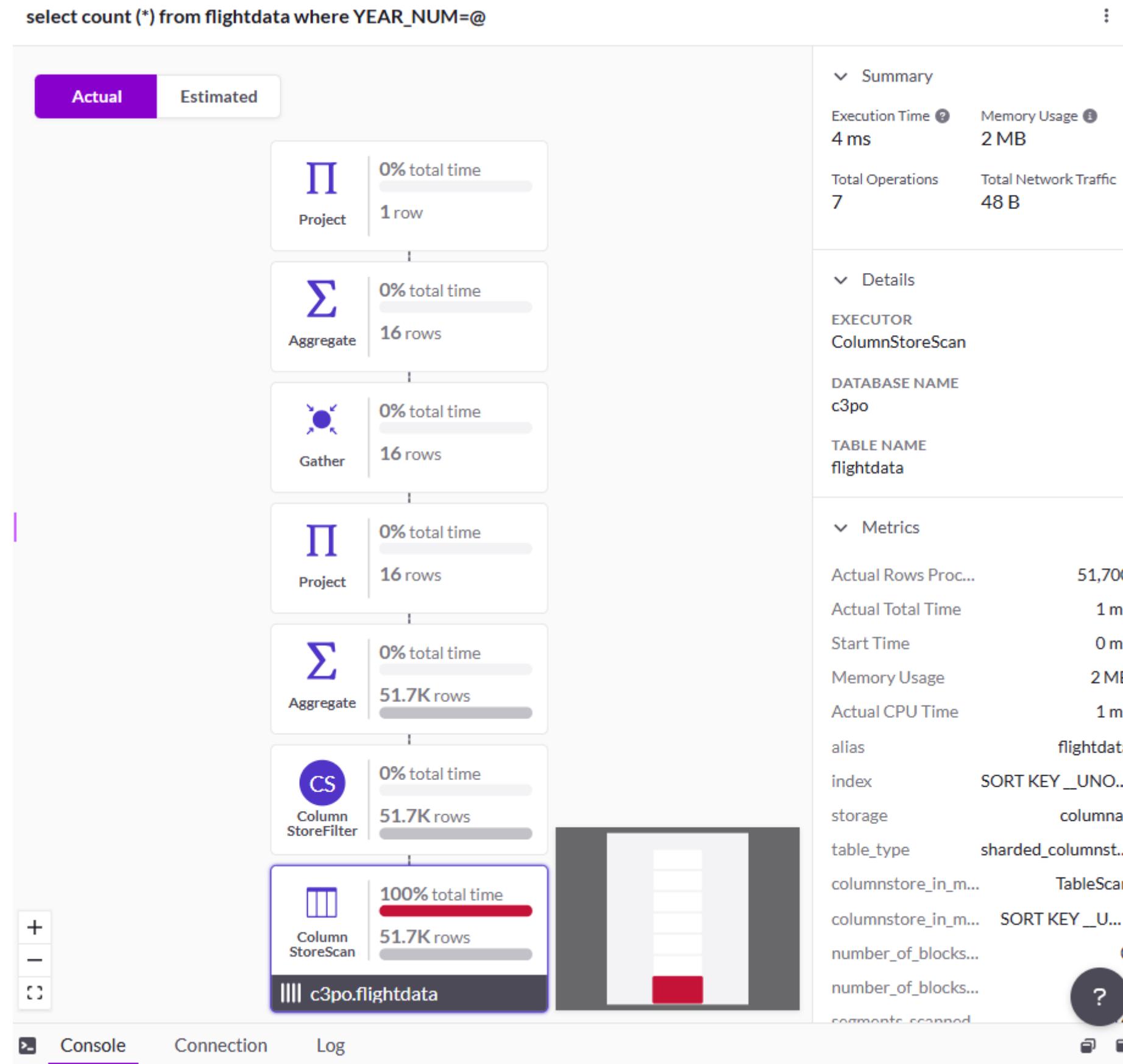


# Real-Time Data Architecture with SingleStore Pipelines

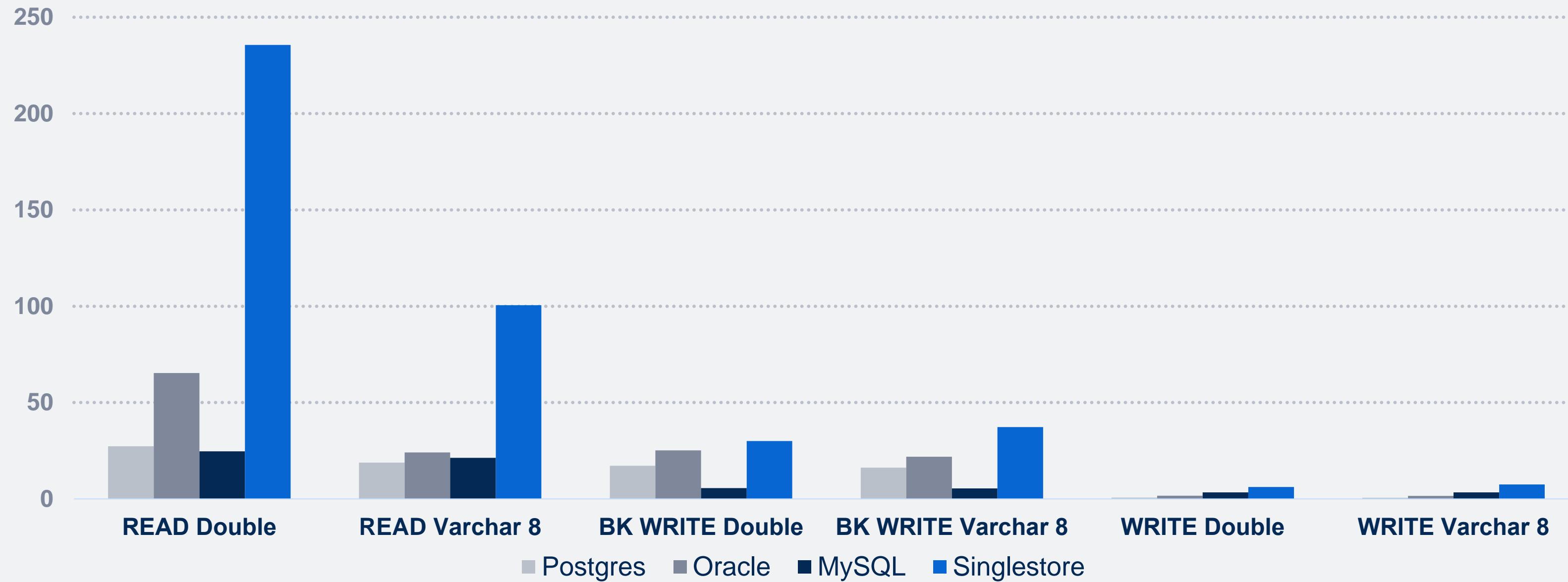


# Performance and efficiency

# Query profile and visual explain for optimization



## Performance in MBps



## SAS Viya Performance

READ 10 million rows

BULKLOAD WRITE 2.5 million rows

WRITE 100k rows

# Optimizing Table Data Structures

## Sharding & Sort Keys

people_1			people_2			people_3			people_4			people		
user	first	last	user	first	last	user	first	last	user	first	last	user	first	last
areece	alex	reece	jdoe	john	doe	jdoe	john	doe	amons	amanda	monson			
anick	alex	nick	jsmith	john	smith									
amace	alex	mace												
thom	tom	holmes												

Shard Key set to column “first”.

They are responsible for distribution of data across partitions. Shard key columns should be as unique as possible.

Sort Key on “Column A” to arrange data in segments.

When filtering on the sort key, the amount of scanning the engine must perform is decreased.

	Column A	Column B
SEGMENT	1	9
	1	5
	1	2
	1	8
	1	1
	1	8
	1	7
	1	4
	1	10
	1	1
SEGMENT	2	3
	2	1
	2	3
	2	10
	2	4
	2	2
	2	9
	2	4
	2	6
	2	5
SEGMENT	3	4
	3	10
	3	1
	3	5
	3	4
	3	8
	3	2
	3	9
	3	6
	3	4

# SAS Viya with SingleStore

## Load and performance test an existing insurance customer



### Current Situation

- An existing, complex report with a large amount of data (wide ABT, ~200M records), which is used interactively by more than 5000 users, is experiencing performance and stability problems.
- The average call-up time of the report is too slow at approx. 60 seconds.
- The controller of the inventory system was overloaded under the load of approx. 260 simultaneous users after the monthly loading of the current data. This behavior was eliminated by a significant increase in hardware.
- As the number of users and data volumes will continue to grow each month, a future-proof, high-performance solution must be found.

### Conclusion

- **SAS clearly recommends** the use of “**SAS Viya with SingleStore**” based on the findings of the load test carried out.
- The average reporting runtimes of the VA report could be **significantly reduced** in the optimization test carried out by SAS with Viya 4 (net report loading time for 300 concurrent users: **<10sec**).
- In the scenarios tested (CAS only, SingleStore only, CAS-SingleStore Mix), the CAS-SingleStore Mix proved to be the best performing and most stable variant across the board.



### Results

- In the low-concurrency test, report runtimes of consistently less than 5 seconds were measured for several configurations (average user).
- In **the heavy load test with 300 concurrent users** working interactively for several minutes, the CAS SingleStore Mix was able to achieve average total **loading times of <20sec** (incl. login + opening the app) and response times <5sec
- Scaling the performance values in the high-concurrency scenarios was only possible by using SingleStore.
- In the course of the tests, additional tuning options were identified which have not yet been implemented due to a lack of comparability, e.g:
  - Activation of the VA Report Cache (faster loading with repeated access)
  - Further reduction of reporting objects
  - More efficient use of lookup tables, use of views (star schema in SingleStore)
  - SingleStore query profiling: optimized shard and sort keys as well as data projection.

# Thank you!

benjamin.walther@sas.com

