

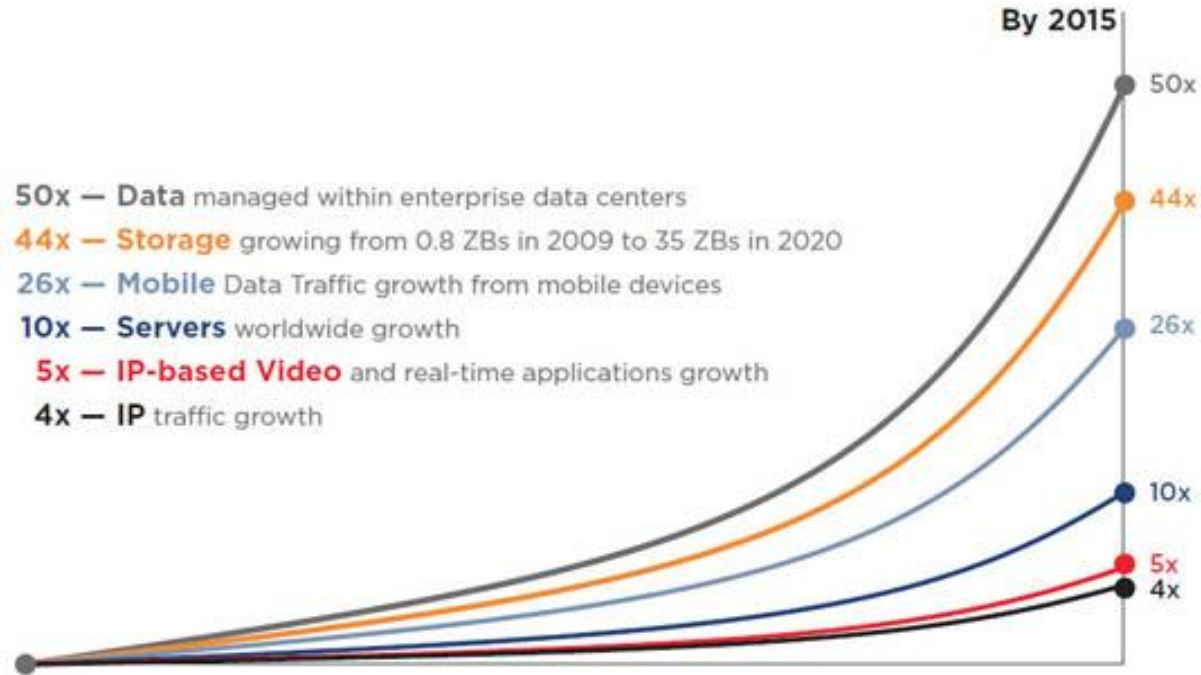
WHAT'S THE BIG DEAL WITH BIG DATA?

DAVID SHAMLIN
SENIOR R&D DIRECTOR
SAS INSTITUTE



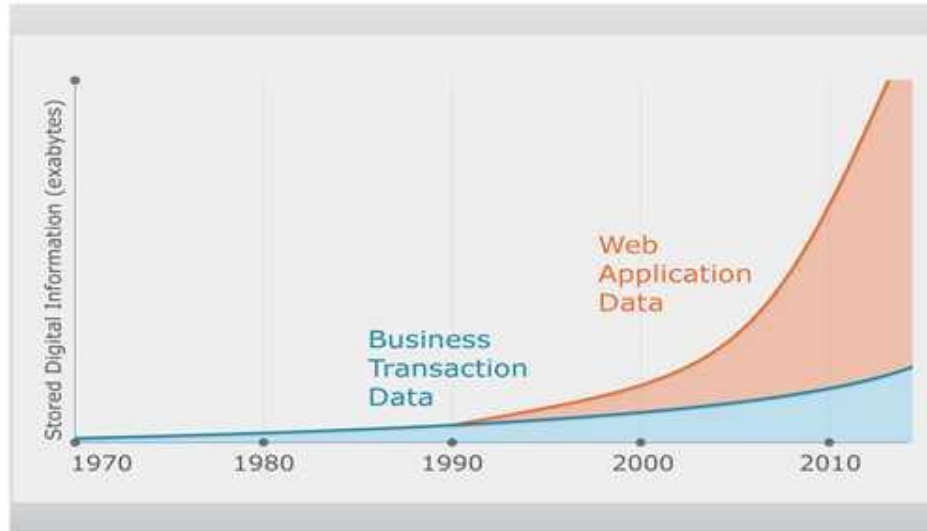
- What is Big Data?
- Compute patterns?
- Technologies

DATA IS GROWING



Source: Cisco VNI, June 2011; Gartner, 2009 & 2011; IDC, 2011

DATA IS CHANGING



Complex, Unstructured

- Text
- Images
- Audio
- Video ...

Relational

BIG DATA | WHAT IS IT?

- Volume → scale & rate of growth
- Variety → diverse & disparate
- Velocity → fluid & fast changing
- Value → descriptive & predictive analytics

ANALYTICS

FORECASTING

Leveraging historical data to drive better insight into decision-making for the future

DATA MINING

Mine transaction databases for data of spending patterns that indicate a stolen card

TEXT ANALYTICS

Finding treasures in unstructured data like social media or survey tools that could uncover insights about consumer sentiment

OPTIMIZATION

Analyze massive amounts of data in order to accurately identify areas likely to produce the most profitable results



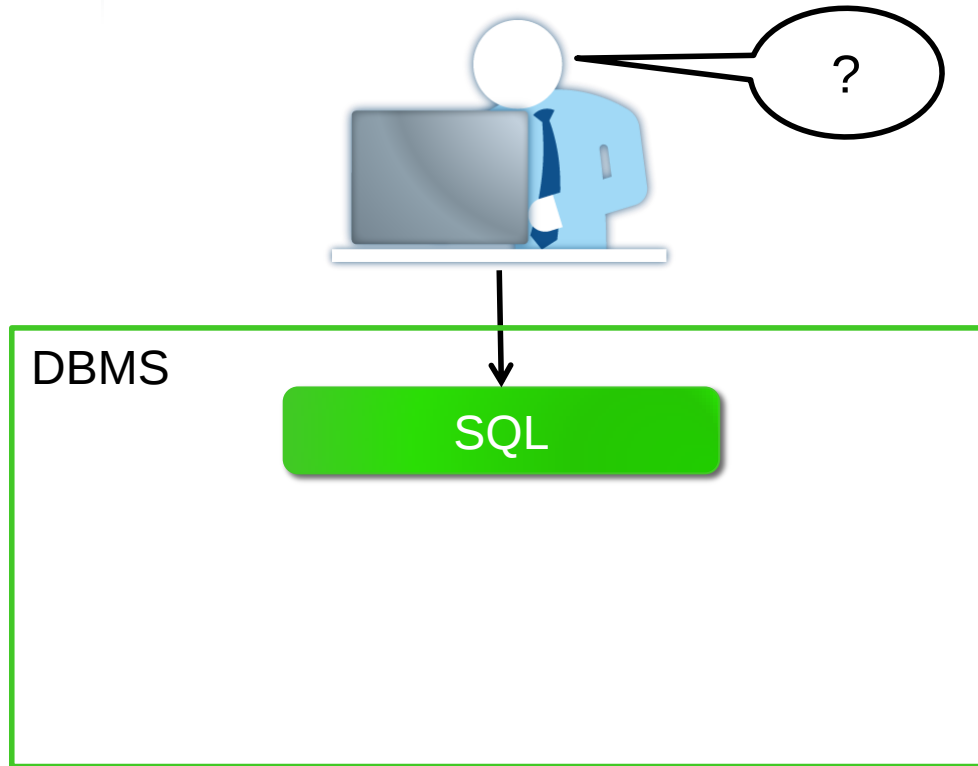
BIG DATA | WHAT'S IN IT FOR ME?

- Many larger volumes of information
- Increase cycle time on your existing data sets
- Use your existing data in more complex ways
- Capture and process new data streams
- Use all of your data

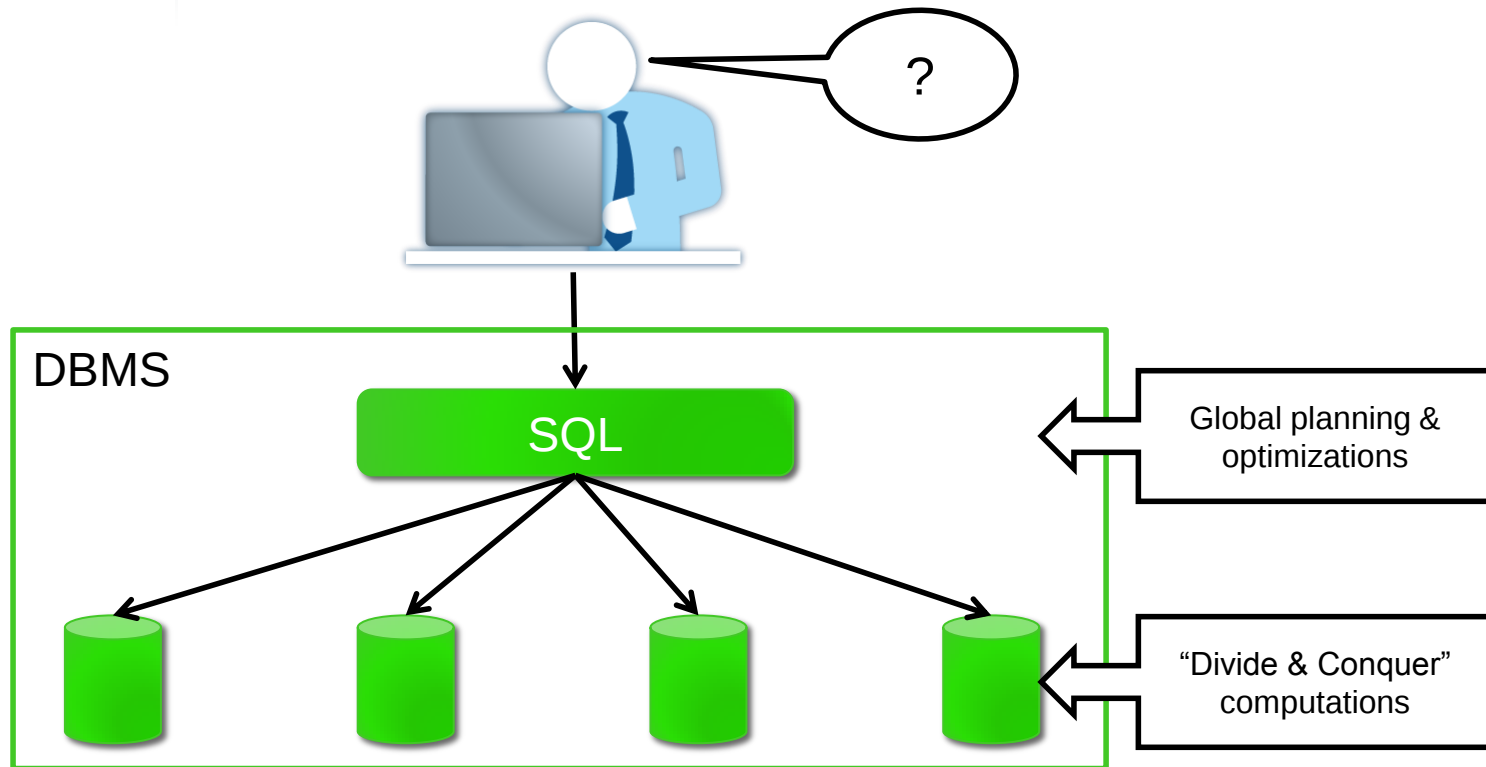
PATTERNS



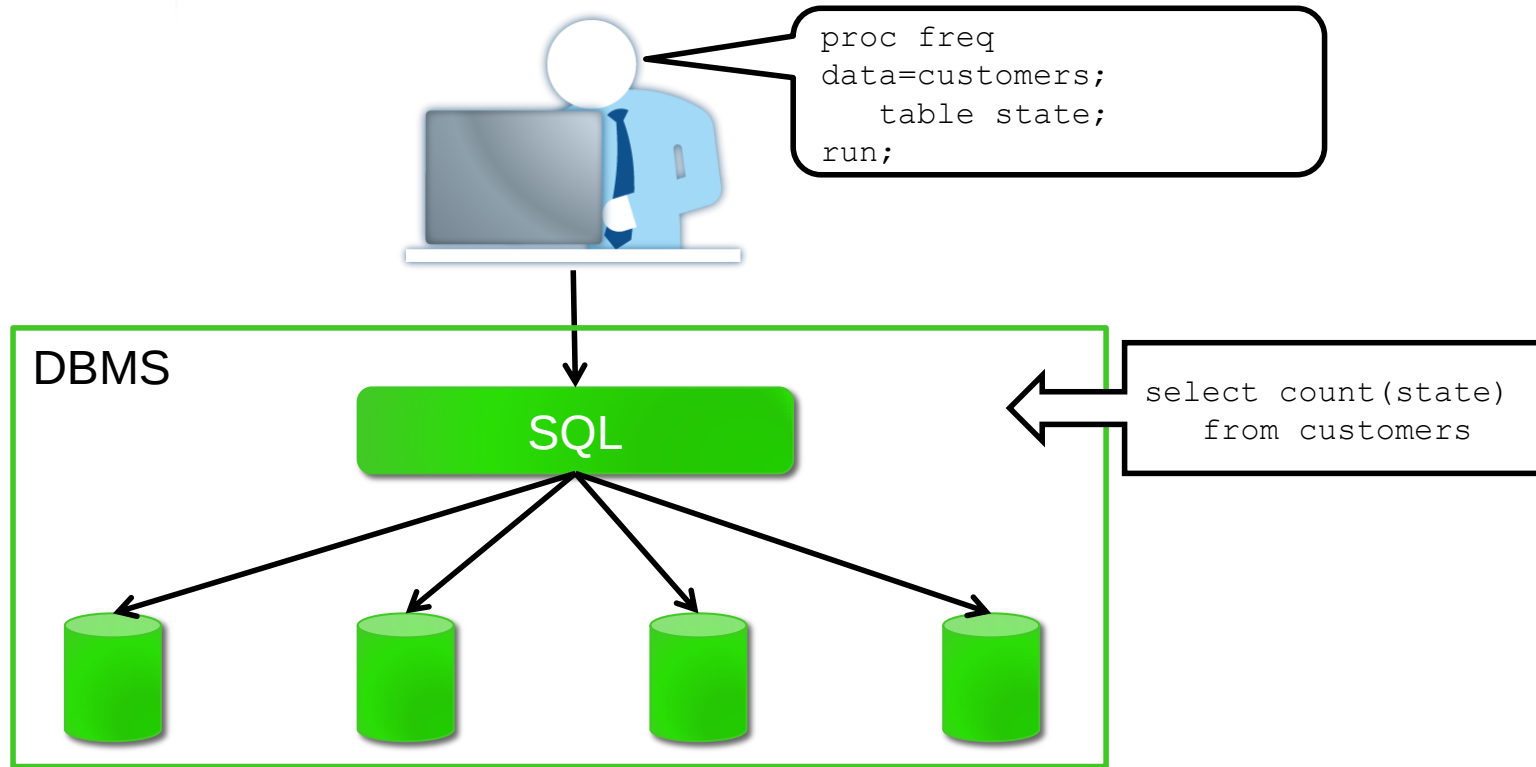
PATTERN SAS/ACCESS



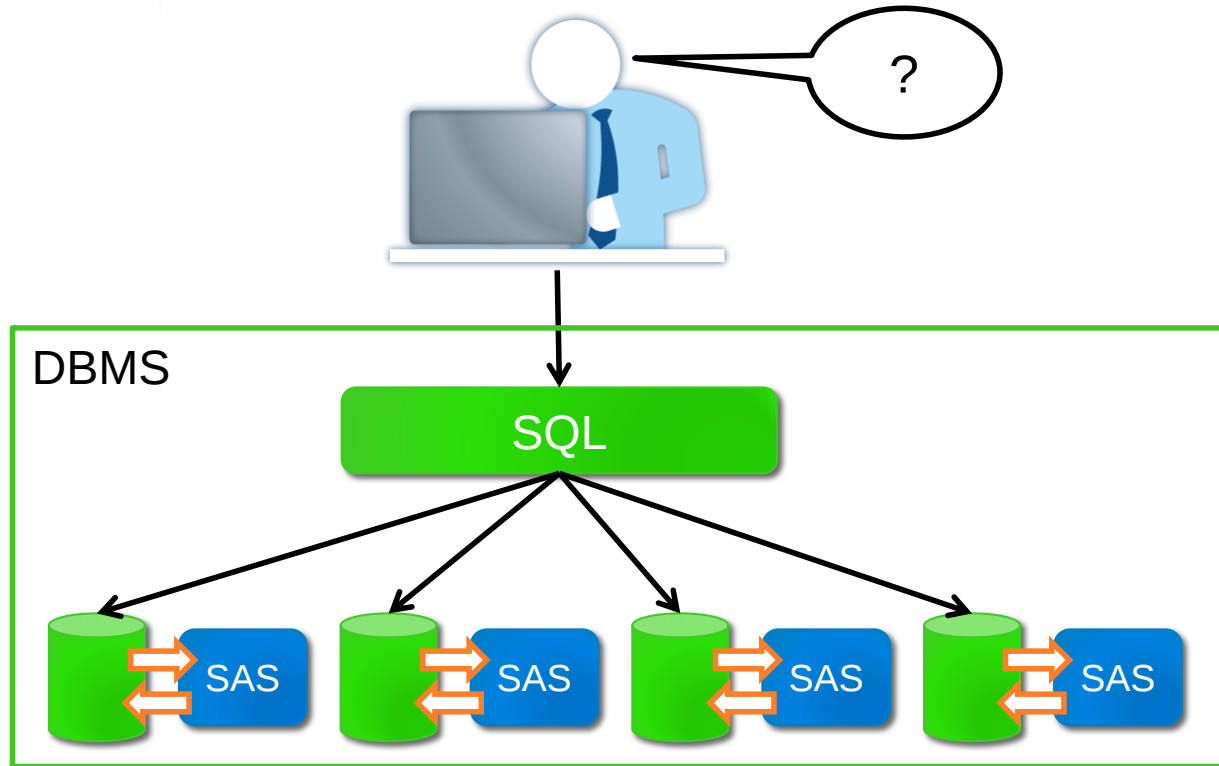
PATTERN “SHARED NOTHING” (MPP) ARCHITECTURES



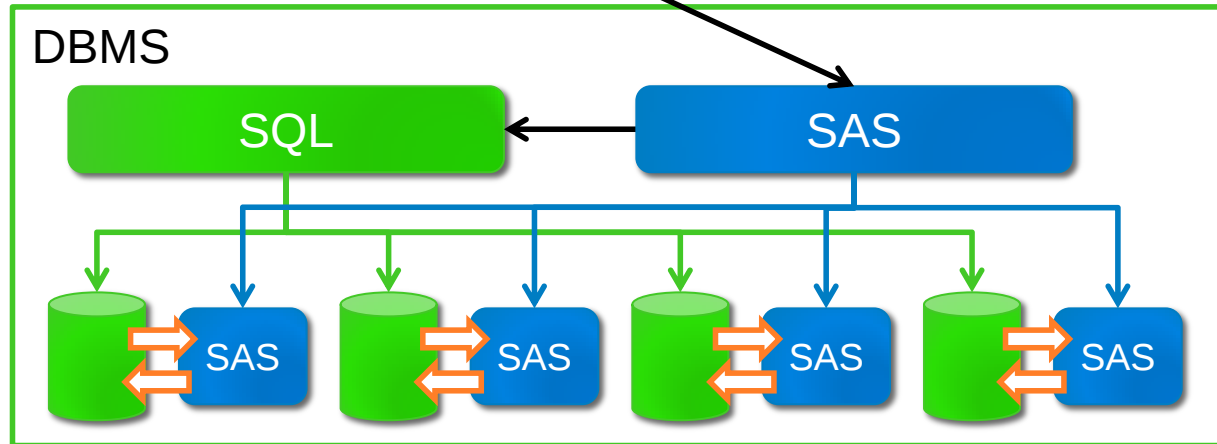
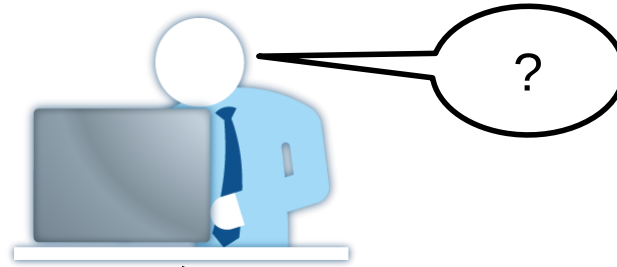
PATTERN PUSH DOWN*



PATTERN IN-DATABASE*



PATTERN IN-MEMORY*





SAS TECHNOLOGY



- SAS/ACCESS
 - SQL Implicit Passthrough
 - In-Database Formats
 - In-Database Base Procs
 - DS1 to LASR
 - DS1 to Hadoop
-  New in 9.4M1
- SAS Grid Computing
 - SAS Scoring Accelerator
 - SAS Code Accelerator
 - SAS Data Quality Accelerator
 - SAS Analytics Accelerator for Teradata
 - SAS High Performance Analytics
 - SAS Visual Analytics (LASR)

- In-database
- In-memory
- Symmetric
- Asymmetric
- Push down
- SAS Embedded Process (EP)
- In-database Base PROCs
- High Performance Analytics (HPA)
- Visual Analytics (VA)
- LASR
- Scoring Accelerator
- Code Accelerator
- Data Quality (DQ) Accelerator

PRODUCT SAS® GRID COMPUTING

Capability

Workload Management



Effectively manage jobs and users

High Availability



Avoid user or service disruption

Distributed Processing



Improved Performance

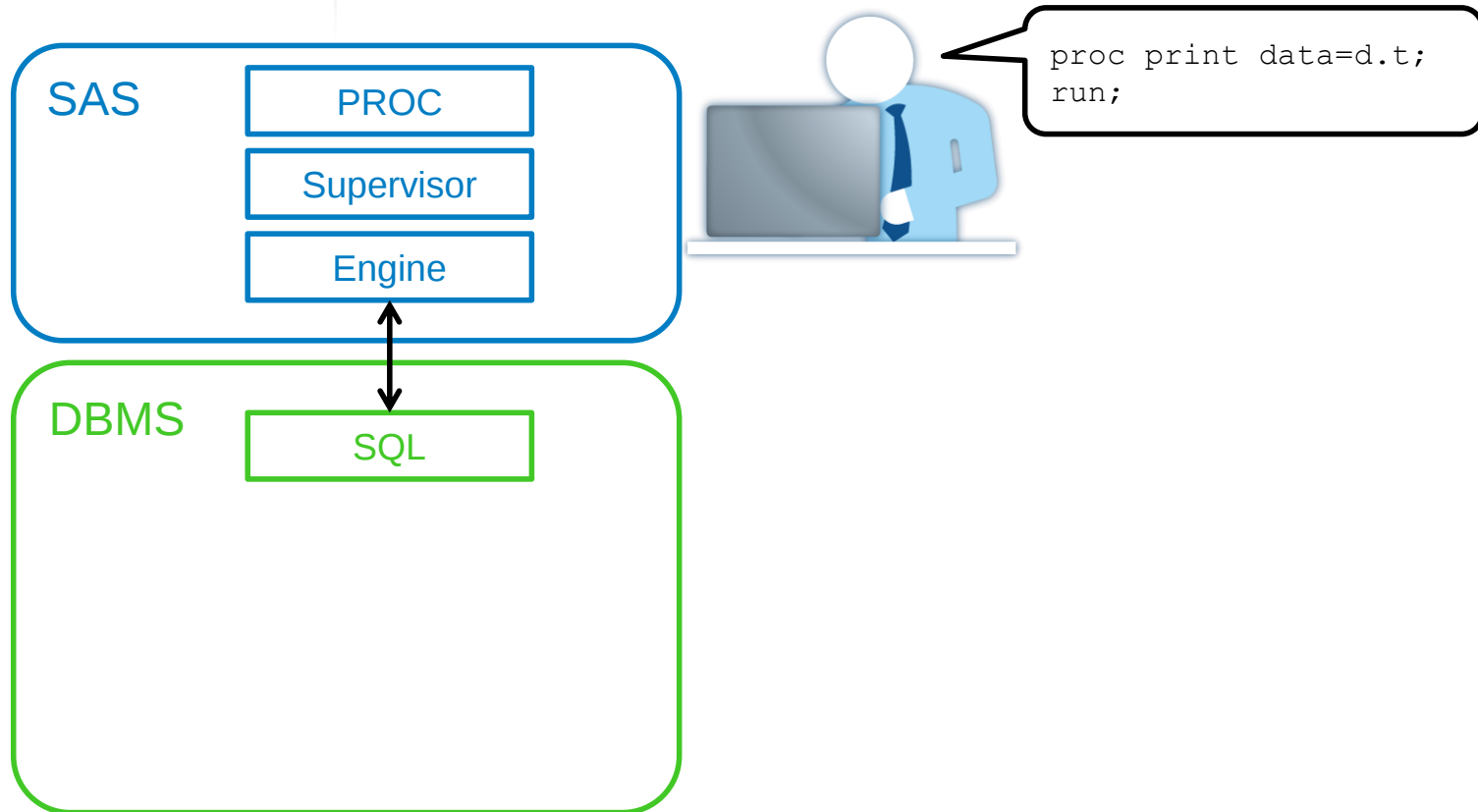
Leverage Commodity Hardware



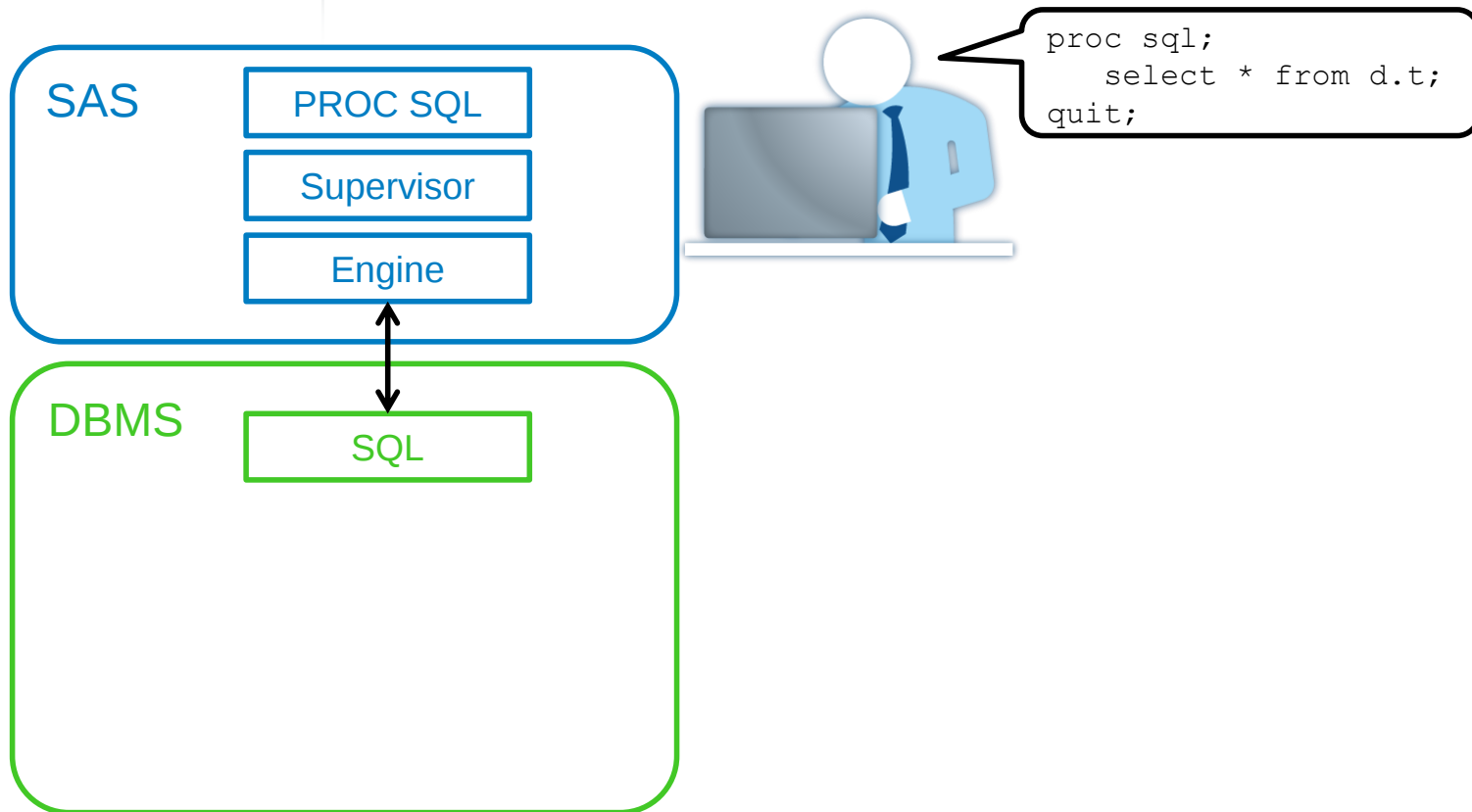
Reduce Costs

Why it Matters

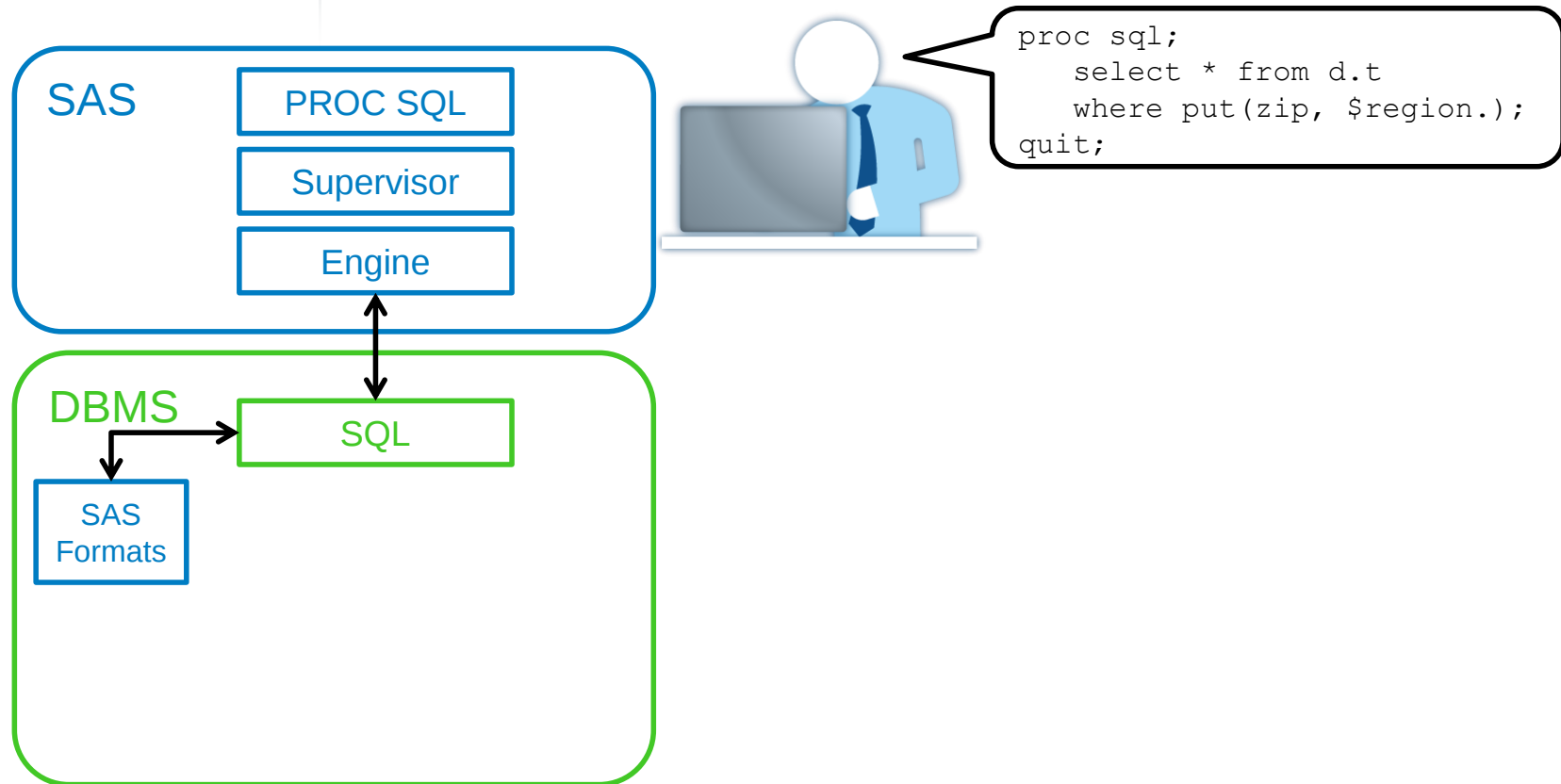
PRODUCT SAS/ACCESS



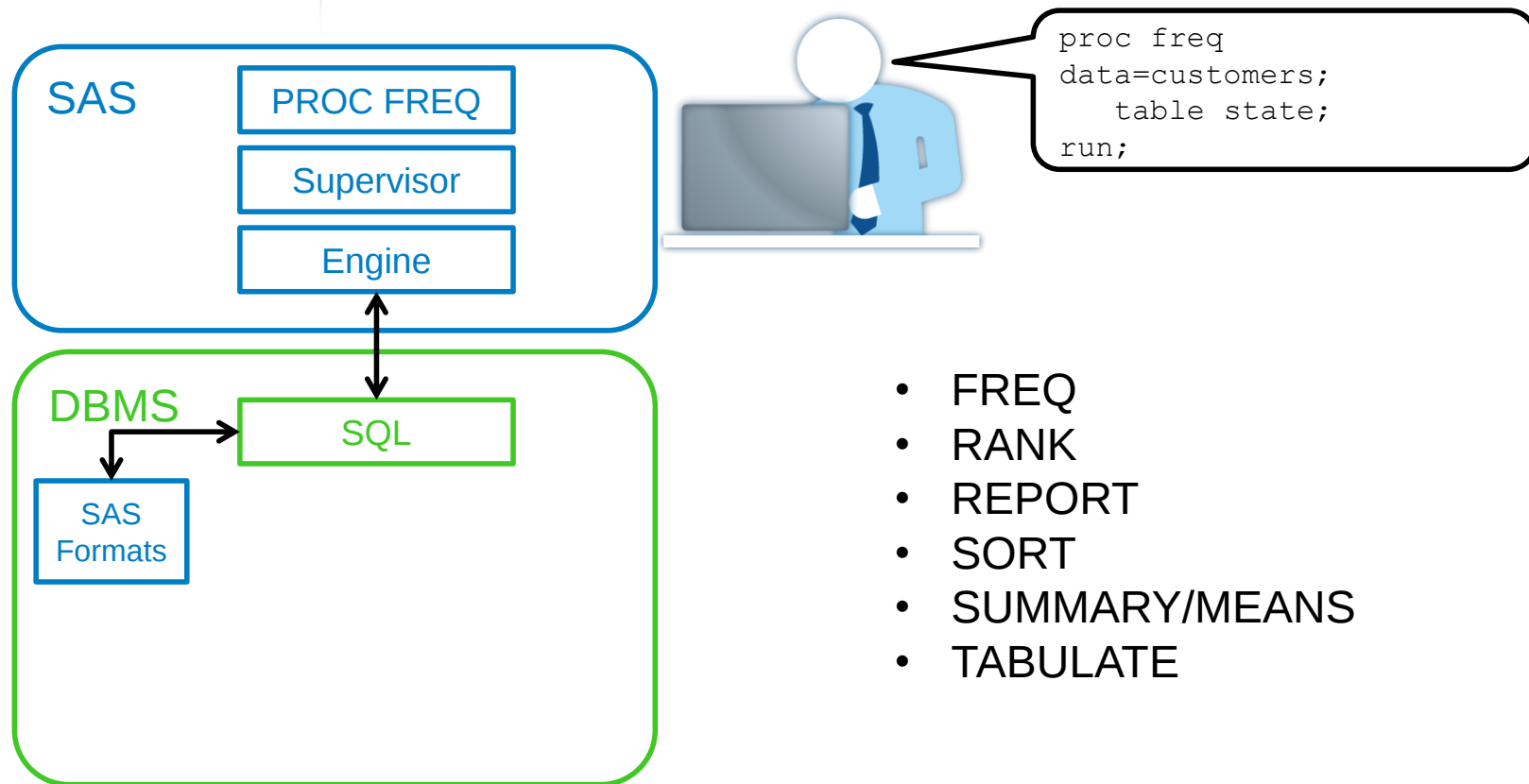
COMPONENT SQL IMPLICIT PASSTHROUGH (IP)



COMPONENT IN-DATABASE FORMATS

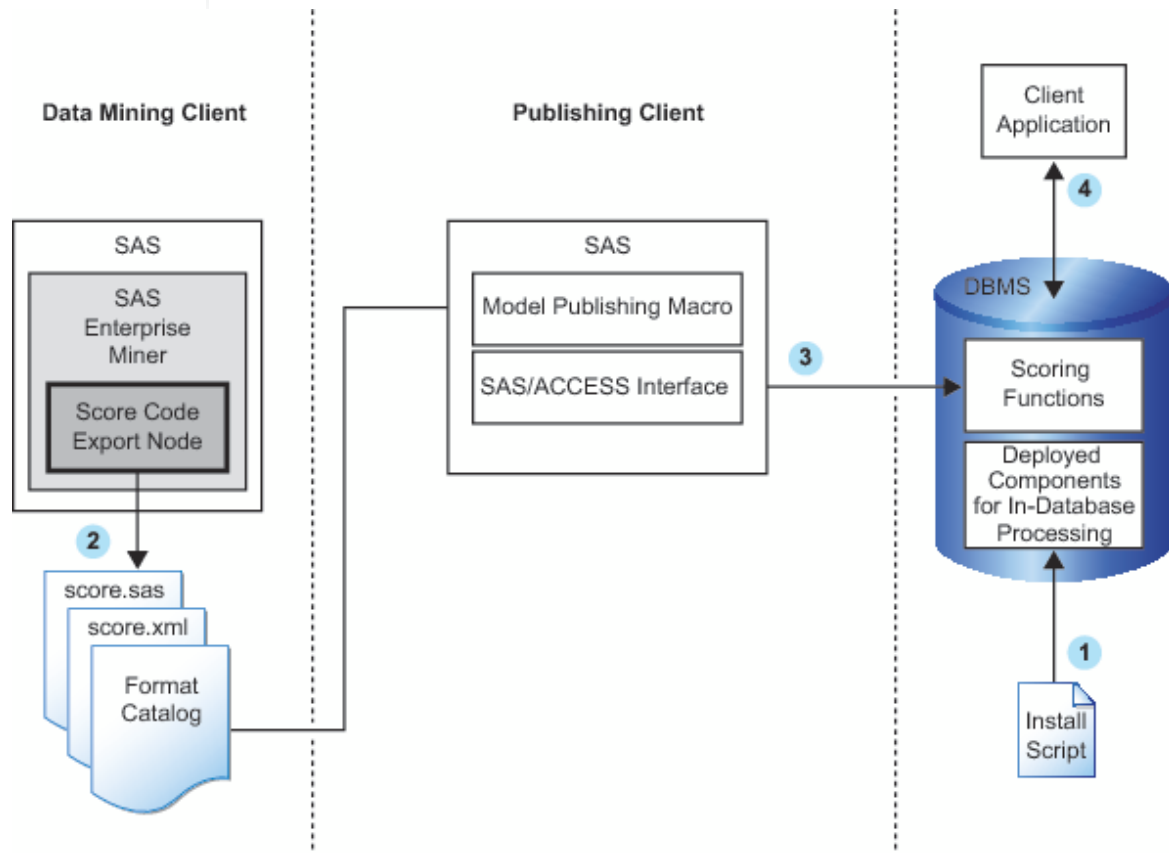


COMPONENT IN-DATABASE BASE PROCS

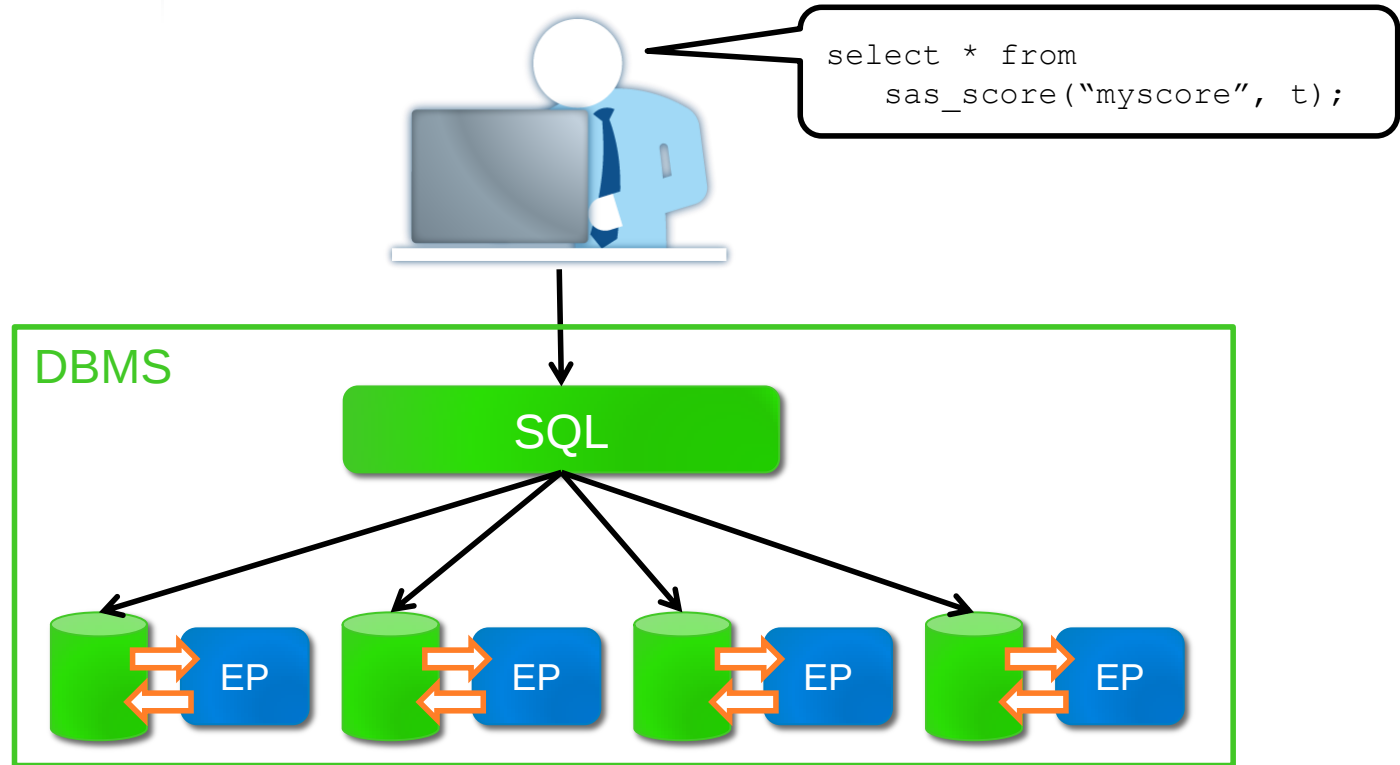


- FREQ
- RANK
- REPORT
- SORT
- SUMMARY/MEANS
- TABULATE

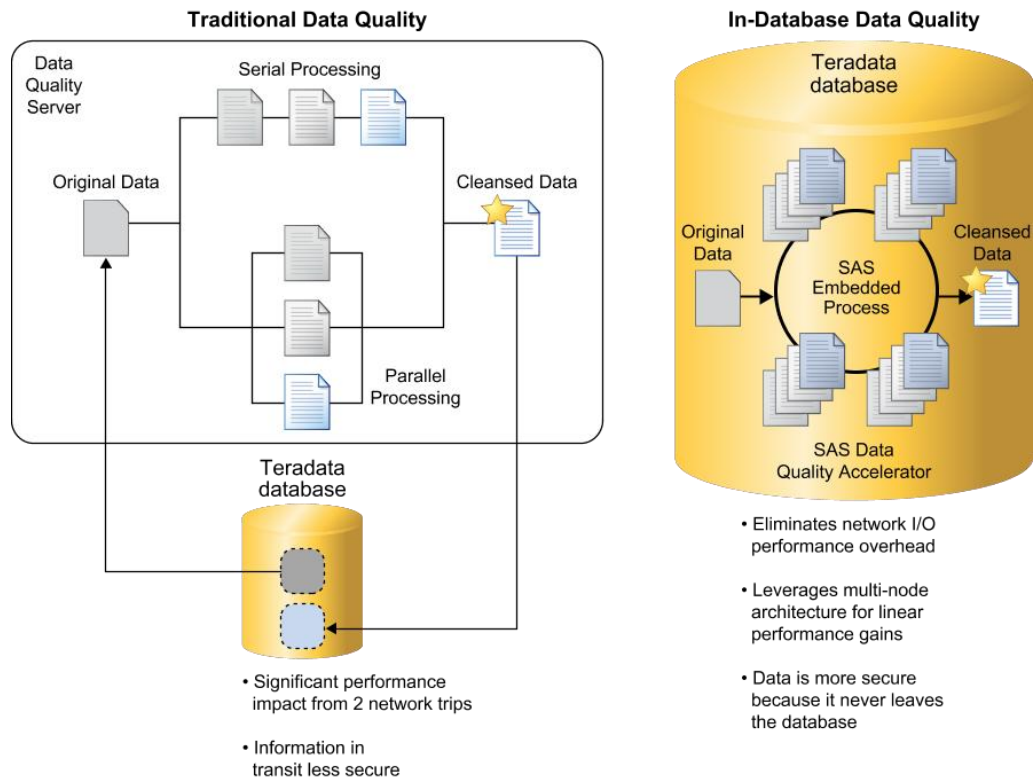
PRODUCT SAS® SCORING ACCELERATOR



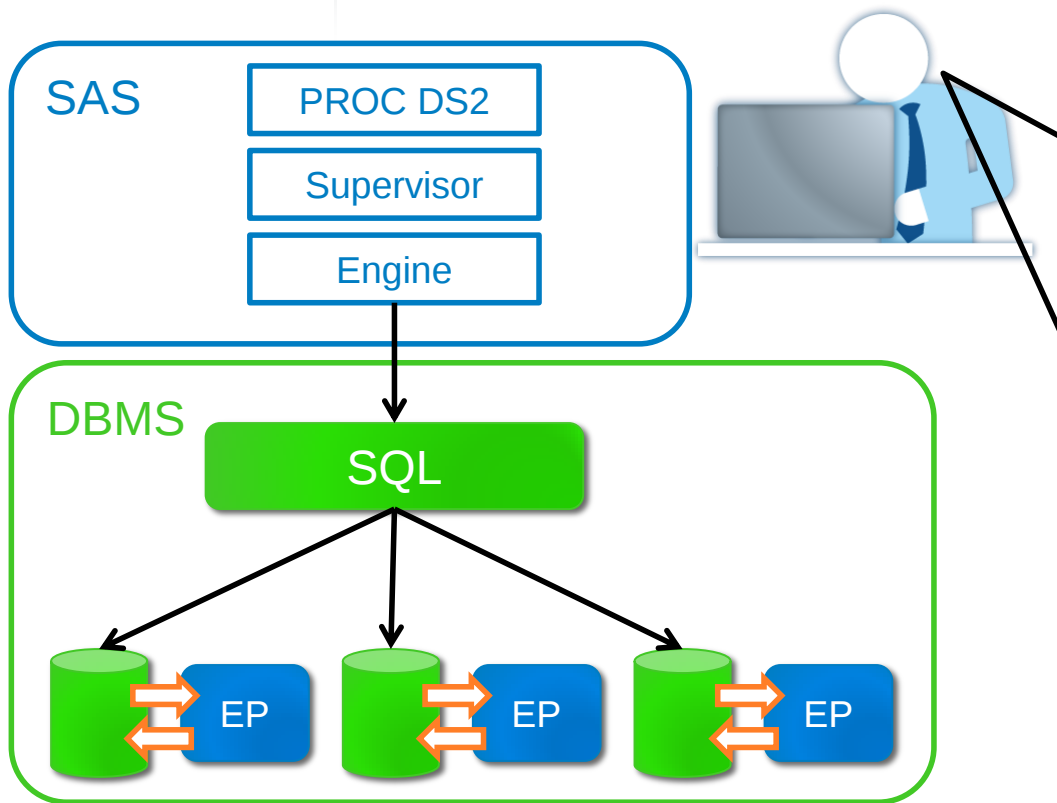
SAS EMBEDDED PROCESS (EP)



PRODUCT SAS® DATA QUALITY ACCELERATOR



PRODUCT SAS® CODE ACCELERATOR



```
proc ds2;
  thread th_pgm;
    dcl double s;
    method run();
      set d.t;
      s = s + 1;
    end;
    method term();
      output;
    end;
  endthread;
run;

data _null_;
  dcl double tot;
  dcl thread th_pgm s_inst;
  method run();
    set from s_inst
    thread=1;
    tot = tot + s;
  end;
enddata;
run;
```

WHAT | SAS® IN-DATABASE

Capability

Process SAS functions
inside the database

Streamline model development
and deployment life-cycle

Leverage existing database
architecture

Run existing SAS code
without modifications



Why it Matters

Better data governance

Faster time-to-results

Improved utilization of IT infrastructure

Gain efficiency

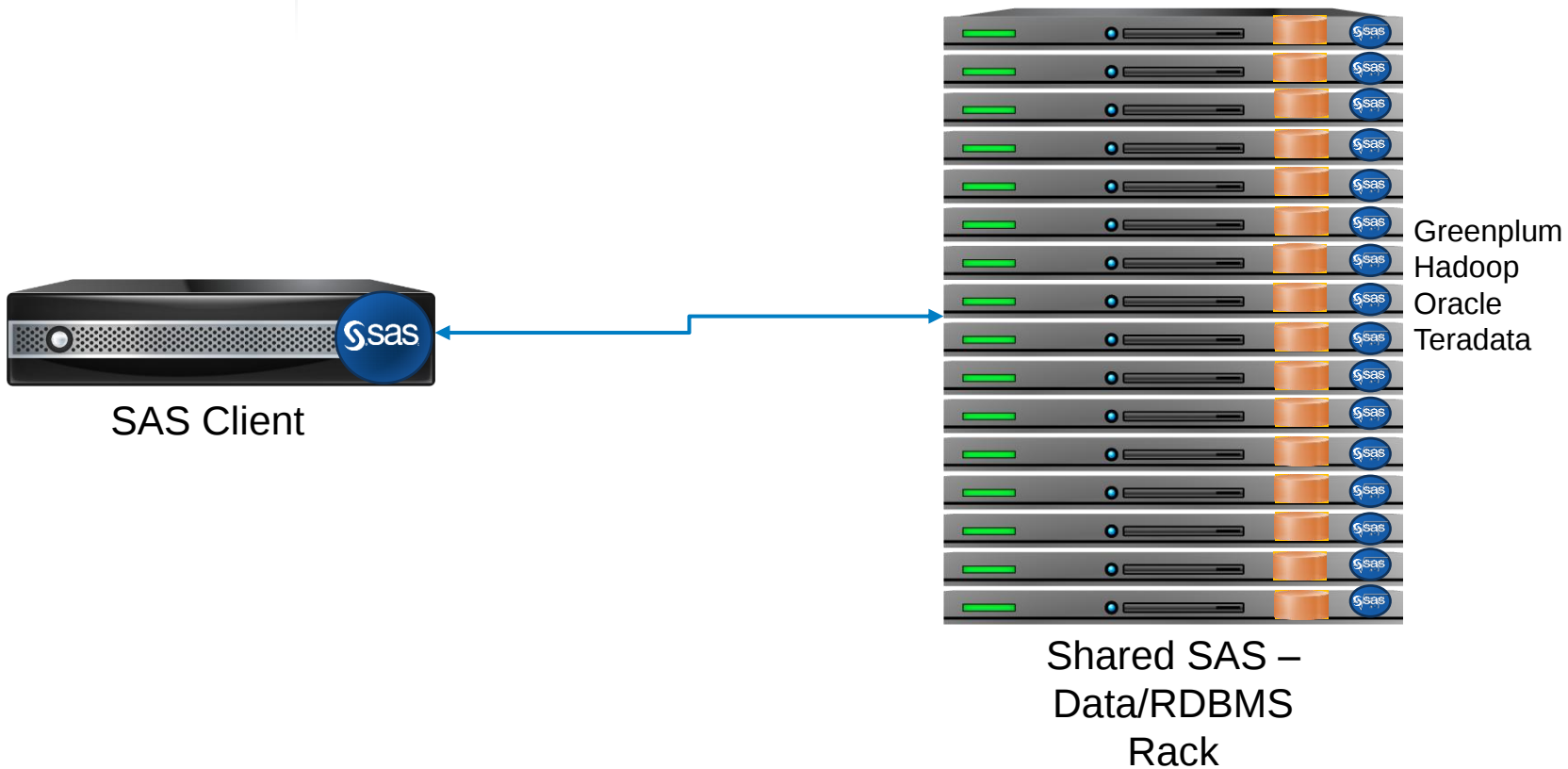
PRODUCT

SAS® IN-MEMORY ANALYTICS

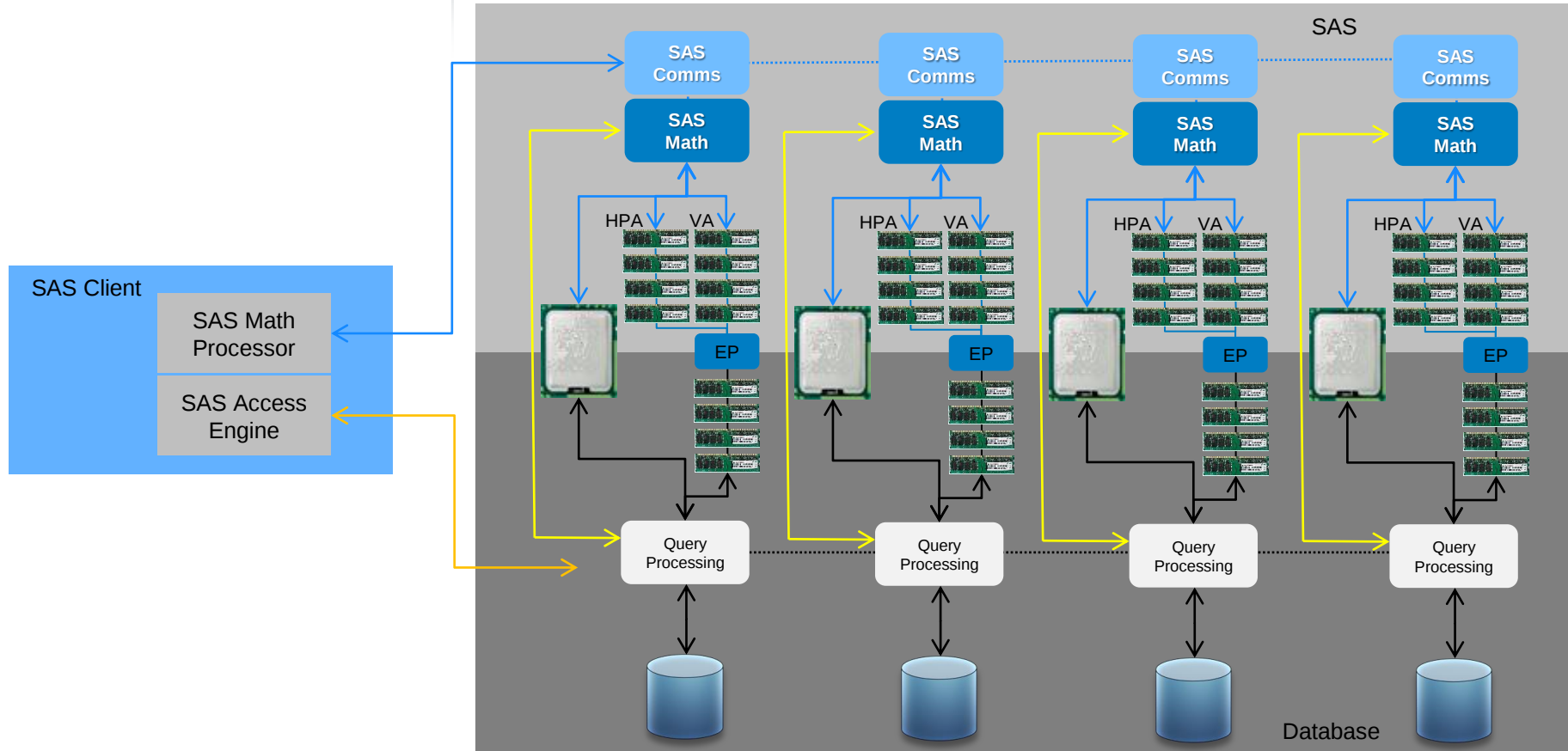
SAS® High-Performance Statistics	SAS® High-Performance Econometrics	SAS® High-Performance Optimization	SAS® High-Performance Data Mining ¹	SAS® High-Performance Text Mining	SAS® High-Performance Forecasting ²
HPLOGISTIC HPREG HPLMIXED HPNLMOD HPSPLIT HPGENSELECT	HPCOUNTREG HPSEVERITY HPQLIM	HPLSO Select features in OPTMILP OPTLP OPTMODEL	HPREDUCE HPNEURAL HPFOREST HP4SCORE HPDECIDE	HPTMINE HPTMScore	HPFORECAST
Common Set (HPDS2, HPDMDb, HPSAMPLE, HPSUMMARY, HPIMPUTE, HPBIN, HPCORR)					

- Common set of HP procedures will be included in each of the individual SAS HP “Analytics” products

SYMMETRIC HPA “SHARED-RACK” ENVIRONMENT

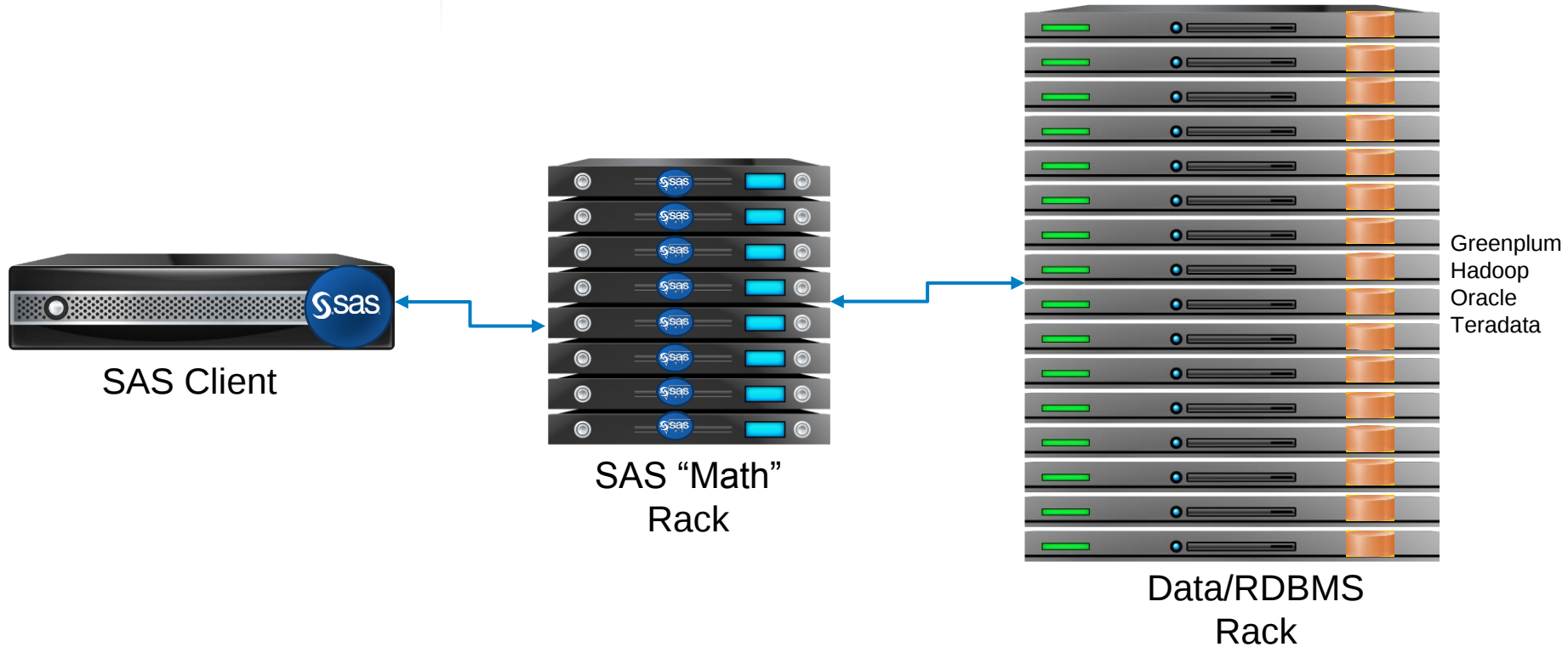


CONCEPT IN-MEMORY INFRASTRUCTURE



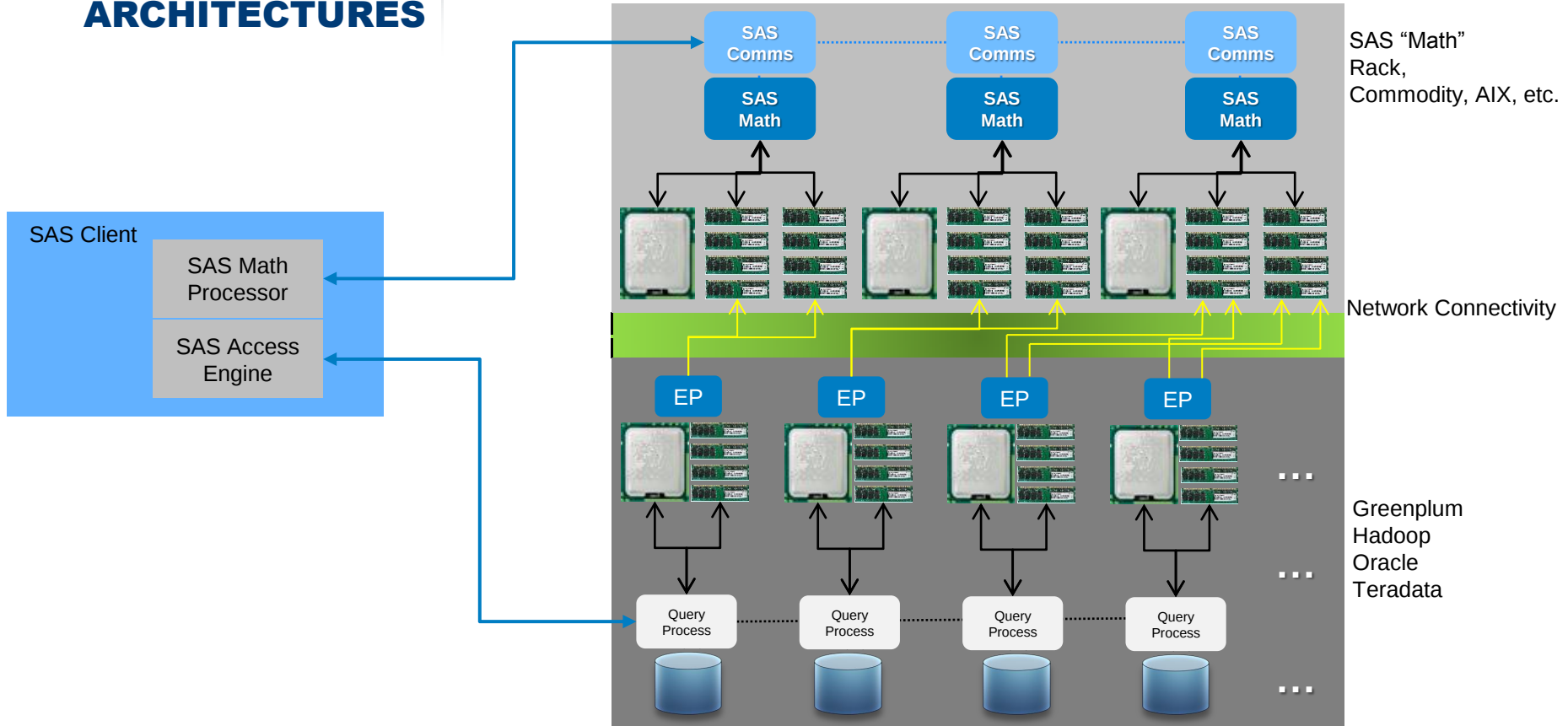
ASYMMETRIC HPA

“SPLIT RACK” ENVIRONMENT



HIGH-PERFORMANCE ARCHITECTURES

ASYMMETRIC – SAS RACK



WHAT

SAS® IN-MEMORY ANALYTICS

Capability

In-memory architecture for data and analytic processing



High-performance analytic capabilities within select SAS products and solutions



Distributed environment form factor (Hadoop or Relational databases)



Why it Matters

Solve your most complex problems in near-real time

Derive highly accurate results through improved modeling

Scalable and reliable analytic infrastructure

HIGH PERFORMANCE VISUALIZATION



Scan rate:

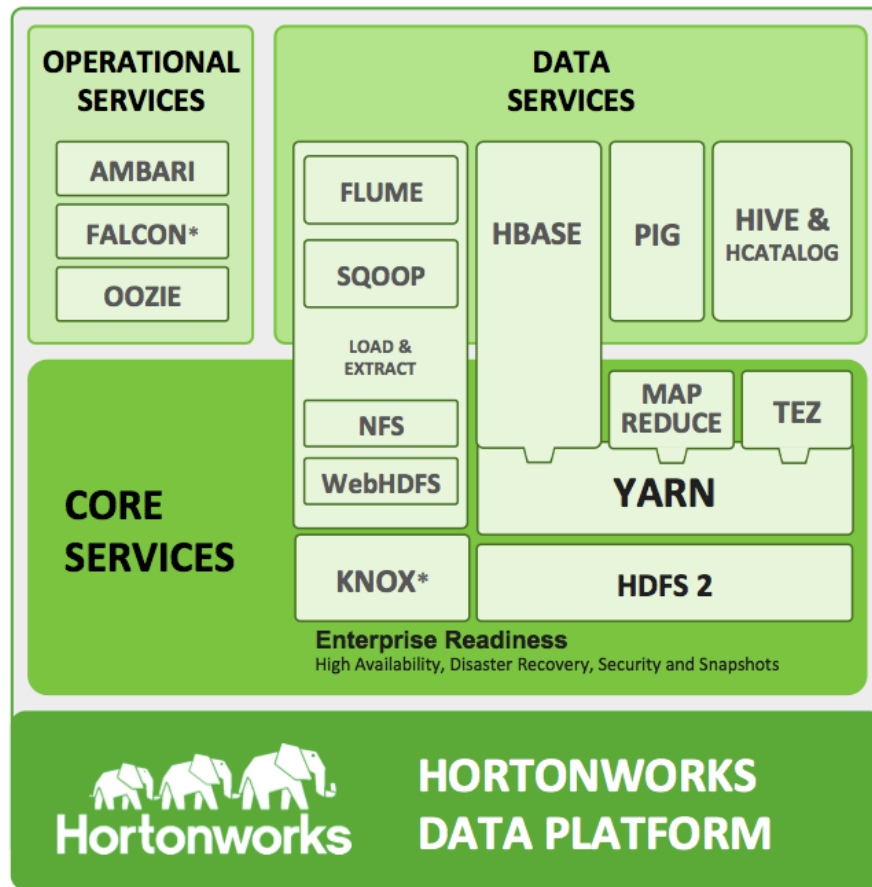
- 1 billion records per second

Analytics:

- Summarization of 1 billion records 0.2 seconds
- 45 simultaneous pairs of correlations on 1 billion records in ~ 5 seconds

“Billion is the new million”

HADOOP





SAS & HADOOP

SAS® WITHIN THE HADOOP ECOSYSTEM



Next-
Generation
SAS® User

SAS® User
User Interface

SAS® Display Manager

SAS® Enterprise Guide®

SAS® Data Integration

SAS® Enterprise Miner™

SAS® Visual Analytics

Metadata

SAS Metadata

Data Access

Base SAS & SAS/ACCESS® Interface to Hadoop™

In-Memory Data Access

Data Processing

Pig

Hive

HBASE

Impala

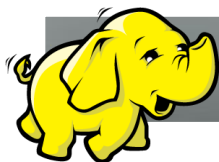
SAS Embedded Process DS2 Accelerators

SAS® High-Performance Analytic Procedures

SAS® LASR™ Analytic Server

File System

HDFS



- In-database
- In-memory
- Symmetric
- Asymmetric
- Push down
- SAS Embedded Process (EP)
- In-database Base PROCs
- High Performance Analytics (HPA)
- Visual Analytics (VA)
- LASR
- Scoring Accelerator
- Code Accelerator
- Data Quality (DQ) Accelerator

SAS PRODUCTS FOR COMMERCIAL DATABASE PRODUCTS

SAS® High-Performance Analytics

www.sas.com/software/high-performance-analytics/index.html

Best starting point. From this page, you can drill down into more specific product areas to get to additional pages for individual products/components. System requirements are listed at the product/component page level. Some DBMS products may be listed.

In-Database Processing

www.sas.com/software/high-performance-analytics/in-database-processing/index.html

Identifies DBMS products supported.

BIG DATA SHARE YOUR FEEDBACK

TO: david.shamlin@sas.com

SUBJ: BASUG: ...

THANK YOU!



**THE
POWER
TO KNOW®**