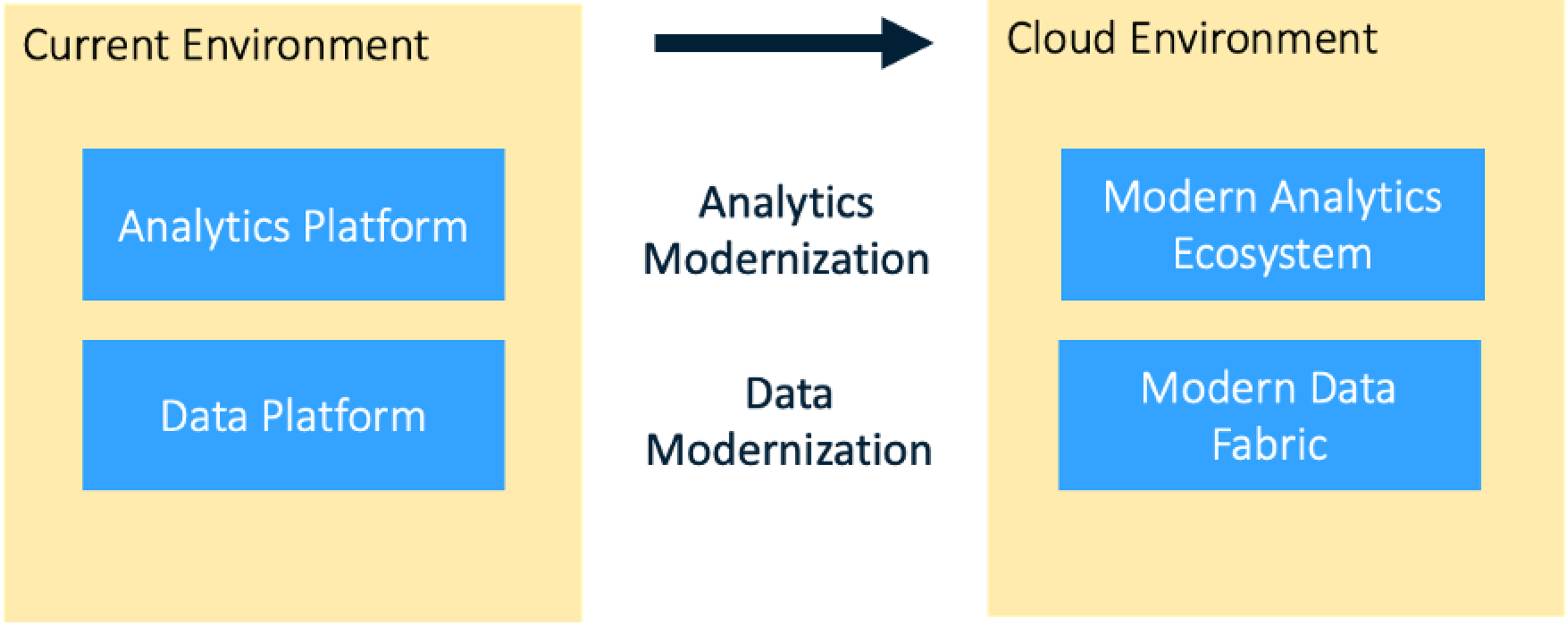
Migrating DI Studio to SAS Studio flows with Data to the cloud

FANS Kristiansand

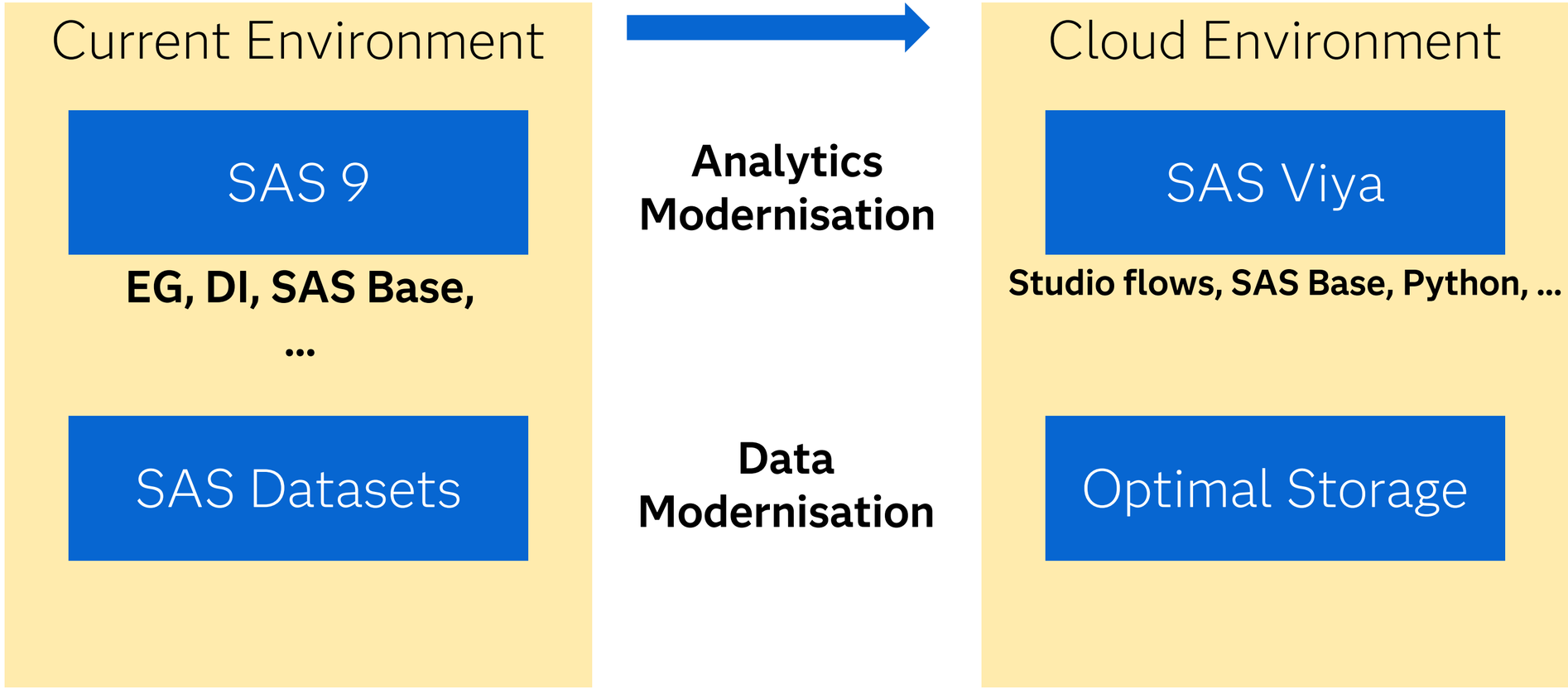
Lars Arne Skår, solution architect

Moving to Cloud

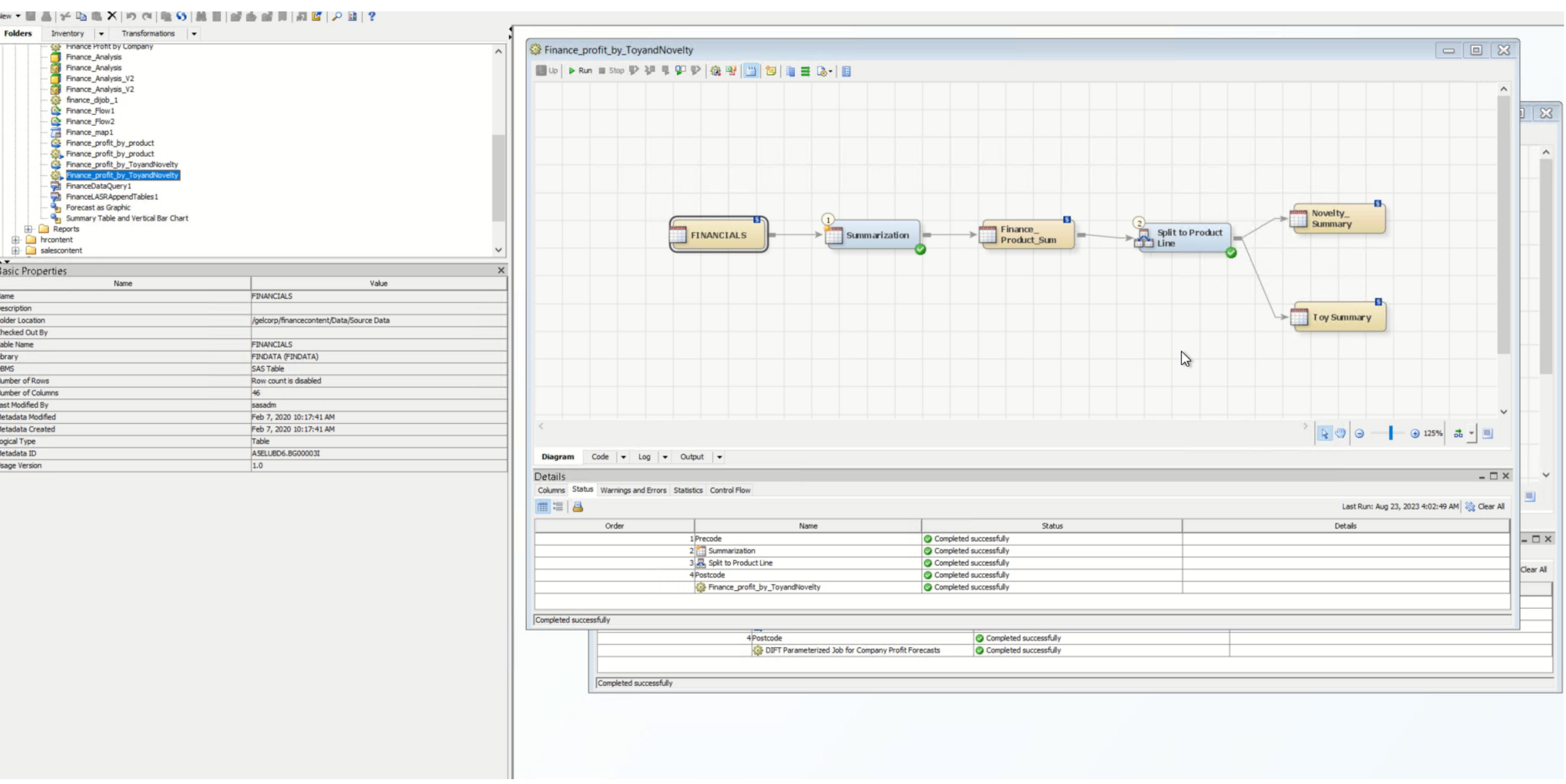


Source: Jeff Stander @ SAS

Moving SAS to the cloud

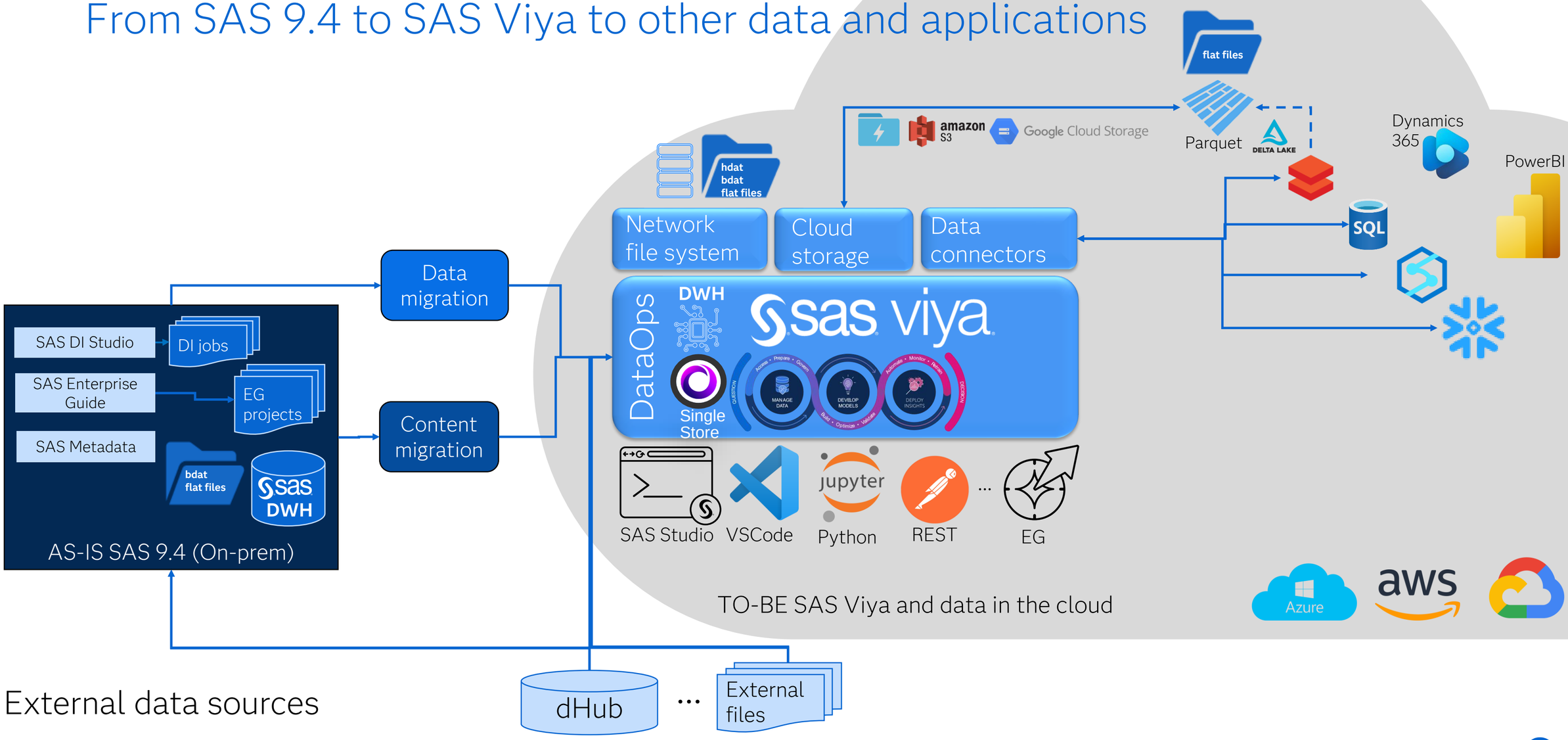


Source: Jeff Stander @ SAS



Data, analytics and AI architecture to the cloud

From SAS 9.4 to SAS Viya to other data and applications



Migration demo

- Almost as-is, use netapp – fast network filesystem mounted to SAS Viya
- Change target to CAS to support fast analytics
  - A nice side effect – save to Parquet files in cloud blob storage for bottomless cost-effective storage
- Change target to Singlestore for a fast cost-effective storage while making data available for others
- Change target to Snowflake for storage at an external cloud-based datawarehouse (efficient, expensive with compute, scale to save cost)
- Change target to Databricks for storage at an external cloud-based data lakehouse (based on Parquet, deltalake and spark)
  - Also comes with a cost for compute – scale down or shutdown to save cost
- DuckDB?

Alternatives?

- Reenginer / recode existing ETL with other tools / languages
- Use scripts or other technologies to support automation of reengineering where possible
- Emulate missing capabilities in the target platform through code or rewrite

- OR
  - Simply migrate with built-in migration support inside SAS Viya to a cloud-native optimized platform for Data & AI
  - Reuse existing IP and code – less risk, save time and effort
  - As an agnostic Data & AI platform you have full flexibility on data platforms and can easily make data available outside of the platform – Data Mesh

Why?

Rational behind this approach

Faster

Shorter time to value by less work compared to rewriting

Small changes (refactor) in order to optimize for the cloud from existing and proven data pipelines

Cheaper

5-10 times less work by reusing existing and proven data pipelines in DI Studio means less cost

Automating testing and delivery through CICD pipeline cost less than manual work

Better

Improved accessibility to analytical data by writing them to accessible storage

Reduced risk and complexity when working from proven data pipelines in a productive environment

Free up resources to business dev