Data Fabrics

Wessel de Meyer
Principal Systems Engineer, SAS NZ
The Evolution of the Data Landscape
An evolving Data Landscape introduces challenges

- Skill shortage
- Data inconsistencies
- Duplicate Data
- Data Latency
- Data Protection/Security
- Data Silos
- Moving Data to the Cloud
- Data Wrangling
What is a Data Fabric?

“Data Fabric is an architecture and set of data services that provide consistent capabilities across a choice of endpoints spanning on-premises and multiple cloud environments”

- NetApp
What is a Data Fabric?

“Data fabric is a combination of architecture and technology that is designed to ease the complexities of managing many different kinds of data, using multiple database management systems, and deployed across a variety of platforms”

- Eckerson Group
What is a Big Data Fabric?

“Big data fabric is an emerging platform which accelerates business insights “by automating ingestion, curation, discovery, preparation and integration from data silos”

- Forrester Research
Why Data “Fabric”? Fabrics are interconnected structures where multiple nodes appear as a single logical unit.

What makes up a Data Fabric?
1. Technology / Data Services
2. Architecture
Data Fabric

Data Access & Federation

Data Catalog & Discovery

Data Quality & Governance

Data Integration

Data Preparation

Data Streams & Analytics

Streaming Data

Data Lake

EDW

On-Premises

New Data

Cloud DW

Off-Premises

Cloud Applications

Data Management

Copyright © SAS Institute Inc. All rights reserved.
Forrester Data Fabric Reference Architecture

Data Management

- Data Security
- Governance
- Metadata
- Search
- Data Quality
- Lineage

Caching, In-Memory, Embedded, Self-Service

Data modeling, Preparation, Curation, Virtualization

Transformation, Integration, Cleansing

Data access
- Data discovery
- Orchestration
- Processing and persistence
- Data Ingestion

Ingestion, machine learning, streaming, data movement

On-premises sources
- Cloud sources

NoSQL
- RDBMS
- Hadoop
Data Scientists

Challenges

- Understanding what data is available
- Getting access
- Putting the data to use
  - Data preparation
  - Analytics
  - Decisioning
- Connected, always-on data

“Is this the best data available?”
Data Engineers

Challenges

• Modernize & simplify data infrastructure
• Optimize use of storage & compute
• Leverage new technologies
• Safe movement of data to the cloud

“We can make this run faster”
CIO and CDO

Challenges

• Simplify and Modernize Data infrastructure
• Protect the data
• Compliance
• Data Governance
• Reduce Infrastructure Costs

“Data is secure and protected”
Thank You