
AWS
re:Invent

ARC202

Nationwide's journey to a governed data lake on AWS

Jeff Anderson
Enterprise Data Architect
Nationwide

Viral Desai
Principal Solutions Architect
AWS

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

About Nationwide



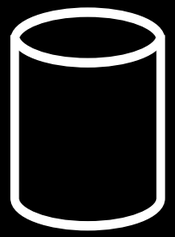
Nationwide®

is on your side

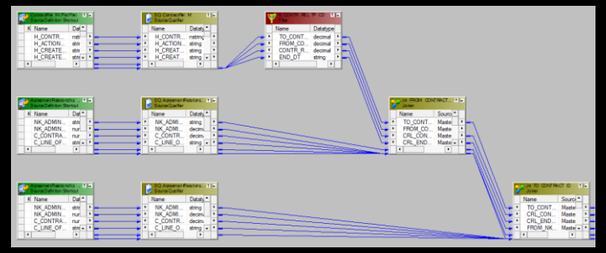
Nationwide, a Fortune 100 company based in Columbus, Ohio, is one of the largest and strongest diversified insurance and financial services organizations in the United States. Nationwide is rated A+ by both AM Best and Standard & Poor's. An industry leader in driving customer-focused innovation, Nationwide provides a full range of insurance and financial services products, including auto, business, homeowners, farm, and life insurance; public- and private-sector retirement plans, annuities, and mutual funds; excess and surplus, specialty, and surety; and pet, motorcycle, and boat insurance.

For more information, visit nationwide.com.

Big enterprise, big complexity



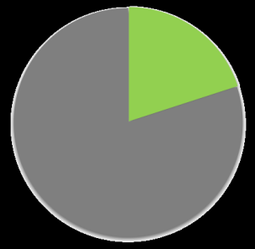
#,000s of databases & warehouses



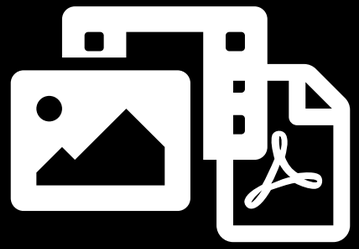
###,000s of ETL jobs



Proprietary platforms



80% of data going unused



Influx of new data types

People who work with data



Data scientists



Data engineers



Data analysts



Information workers



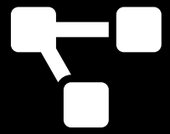
Administration



Build and train ML models



Data discovery and modeling



Profile, classify, data quality



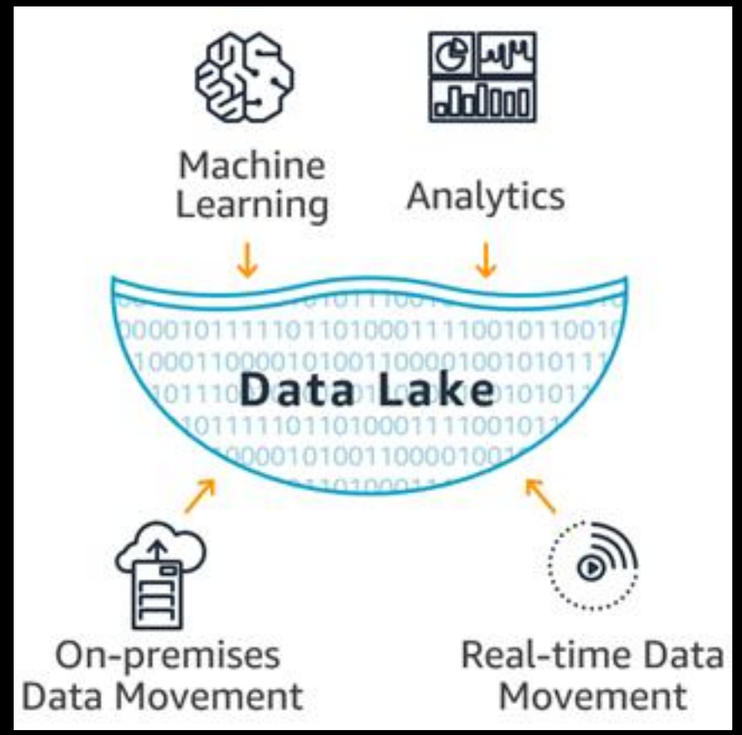
Ad hoc reporting, analysis



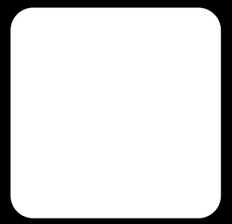
Sharing, archival and purpose-built engines

Value proposition

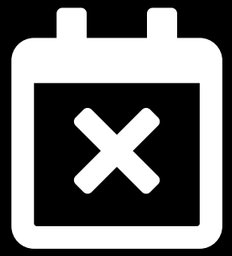
- Gain deeper insights to drive innovation
  - Investment portfolios
  - Customer profiles and activity
- Innovate
  - New opportunities
  - Enhance customer experience
- Agility, experimentation, and depth of analysis



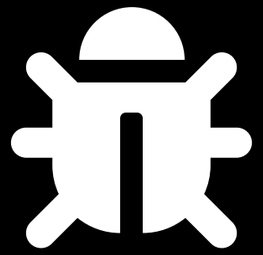
On-premises data lake pain points



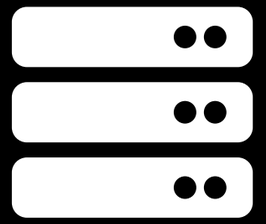
Server patching



Frequent downtime for maintenance



Bugs in vendor software



Coupled storage and compute



Large capital expense and maintenance contracts

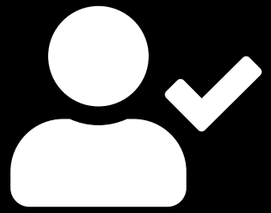


Big shared environment



Complex security

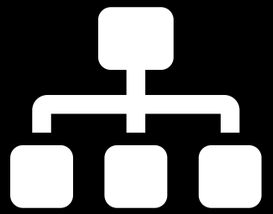
What is a governed data lake?



Data must be profiled, analyzed, and classified before it can be added



Data must be understandable

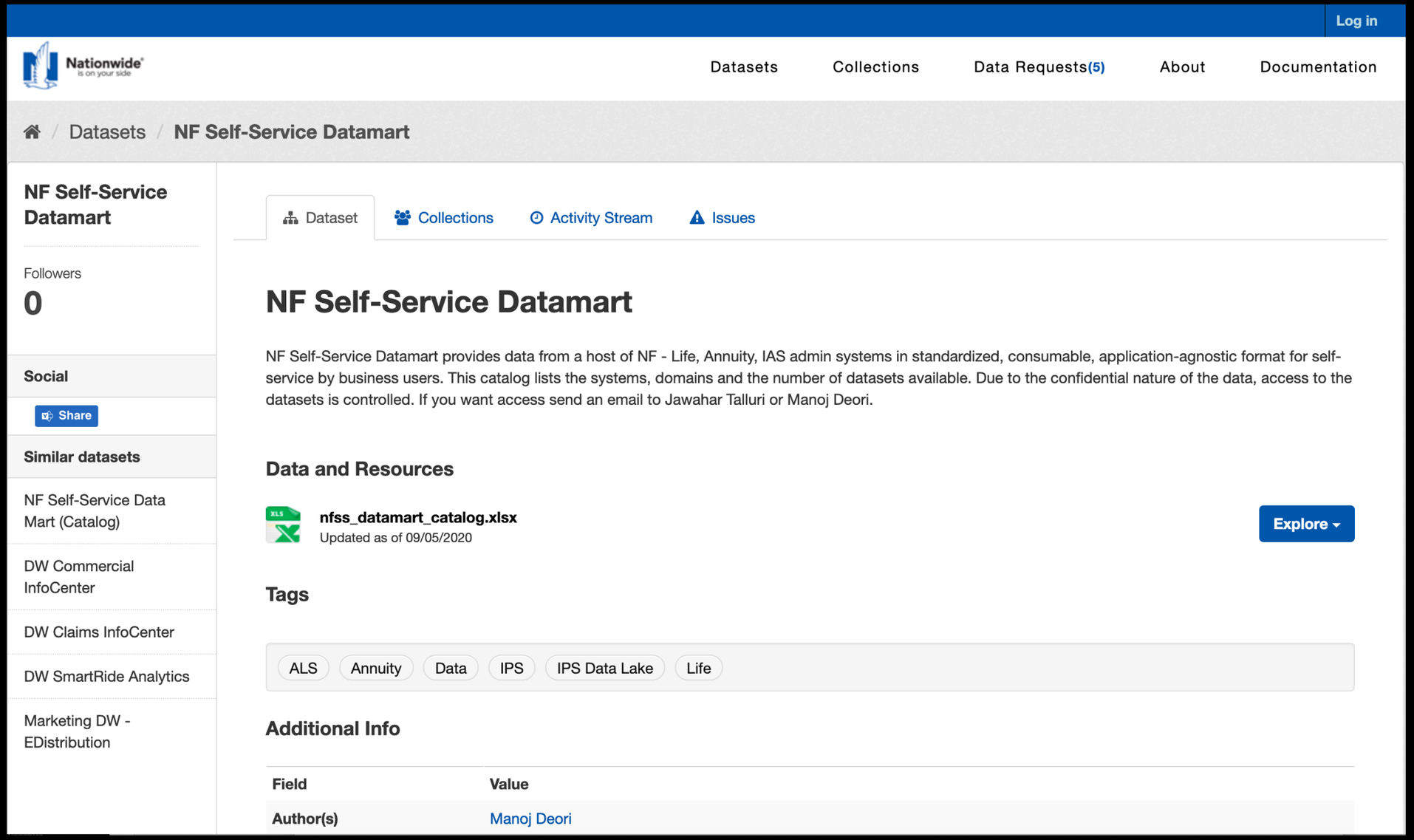


Data lineage must be known

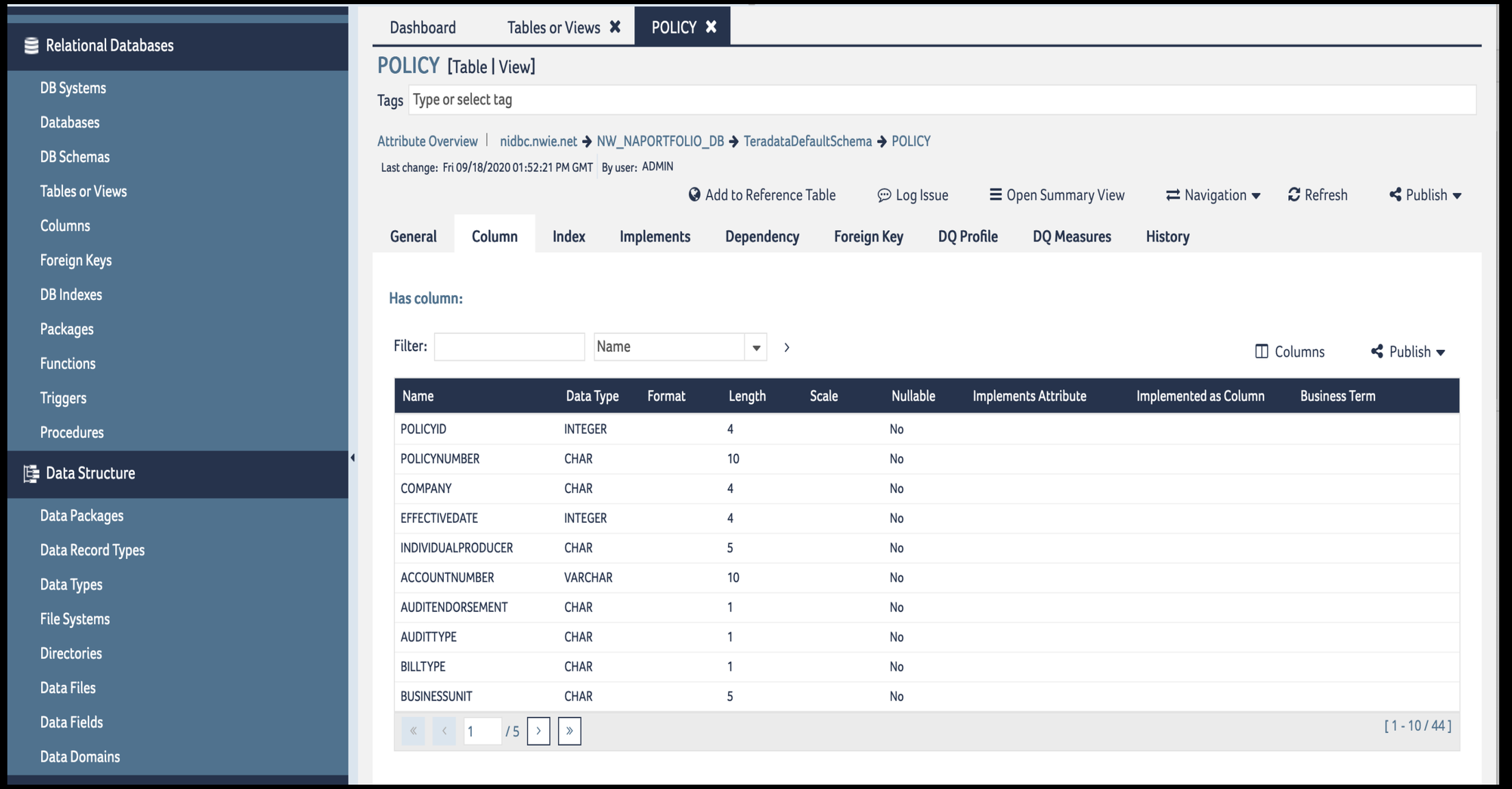


Data must be discoverable and accessible

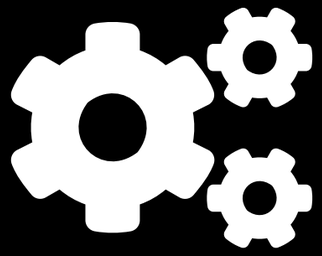
The Nationwide Data Marketplace



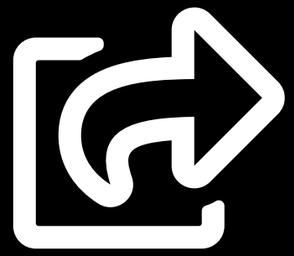
The Nationwide Governance Catalog



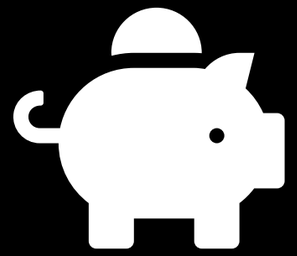
Additional requirements



Configurable and fully automated



Event driven



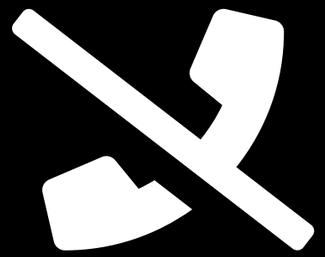
Low cost



Secure and auditable



Query optimized

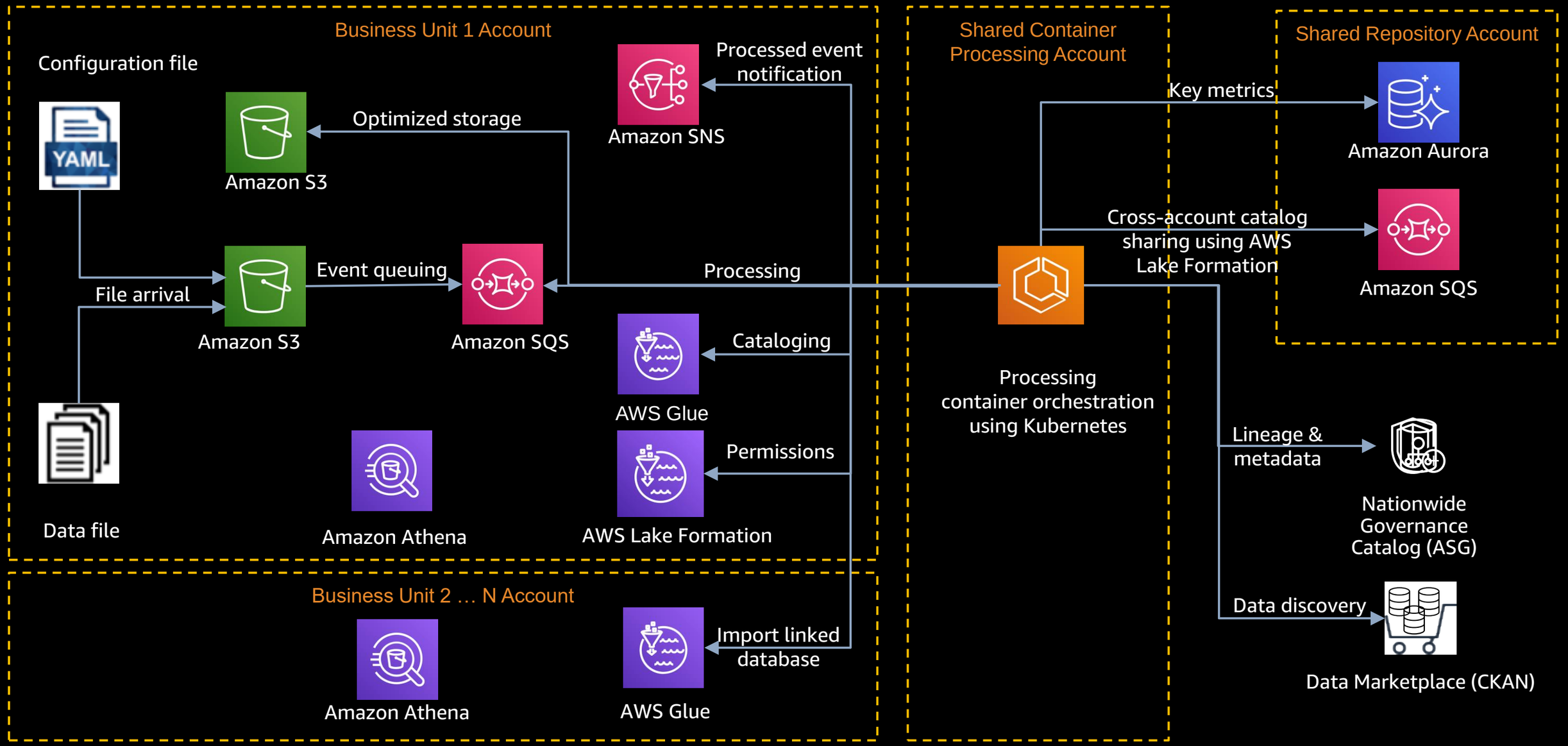


Needs minimal support

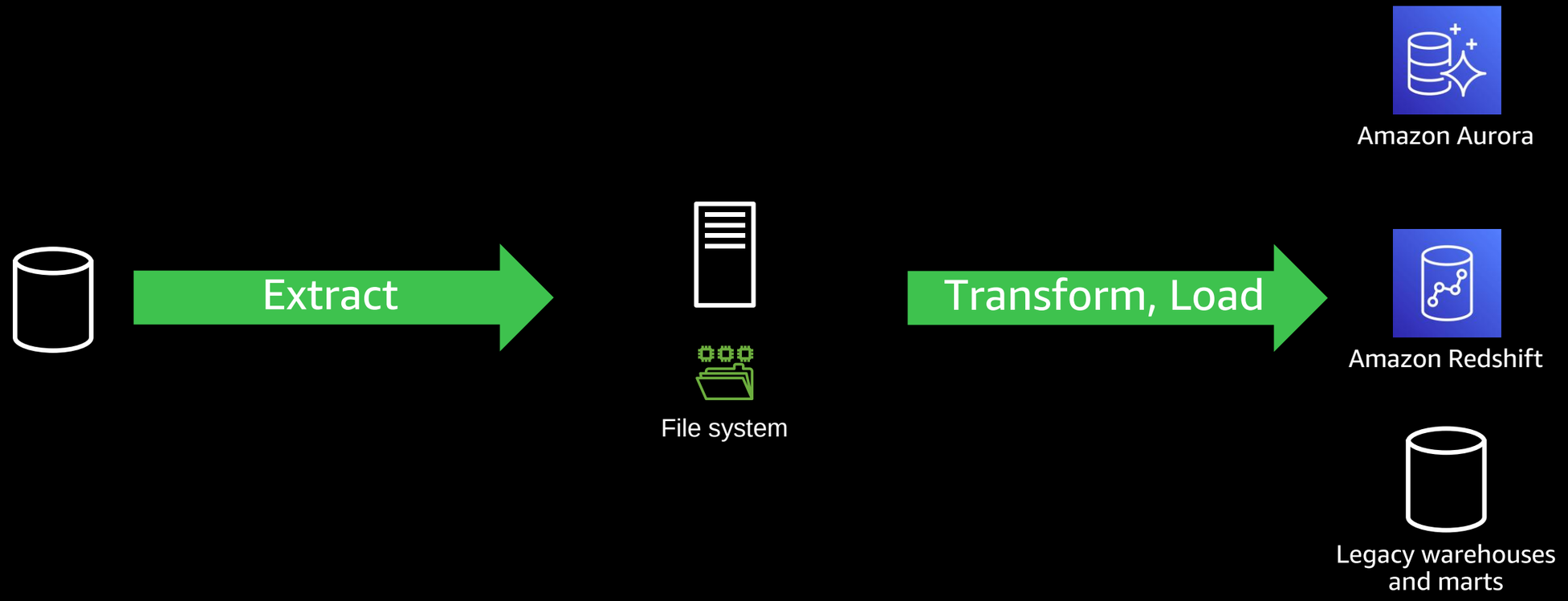


Self-service

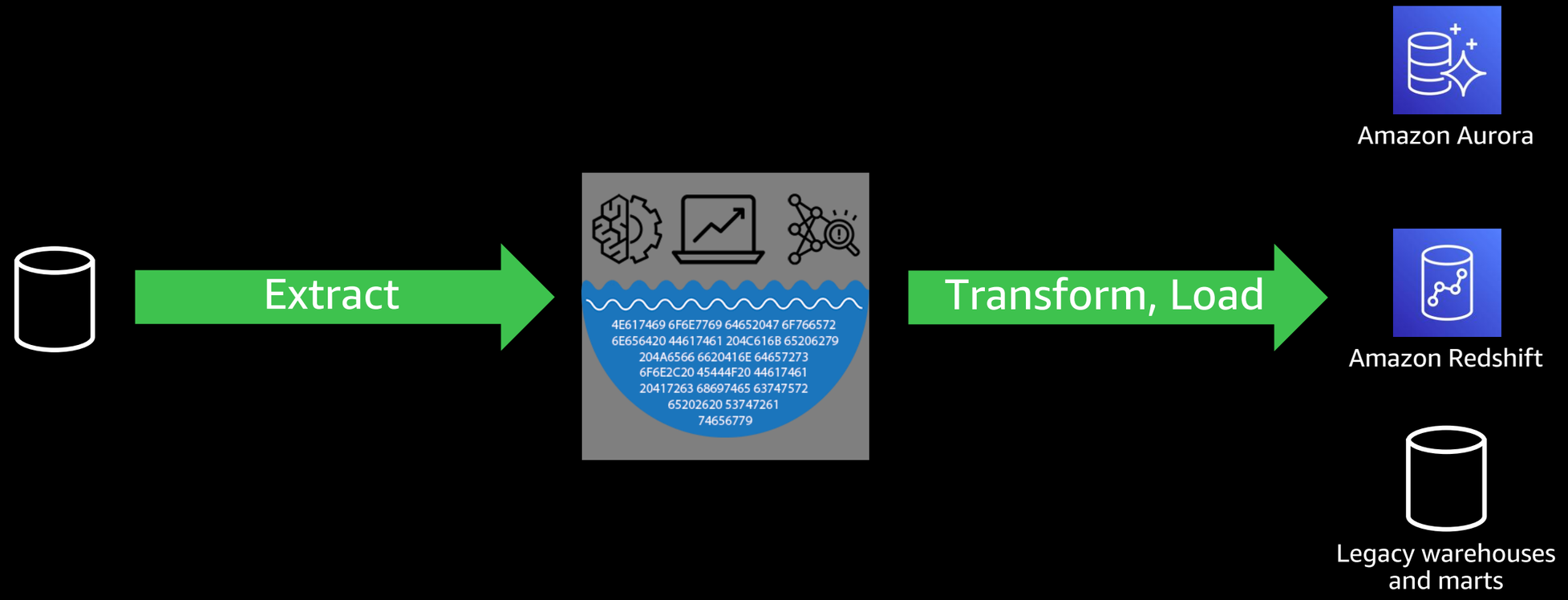
Architecture



Hydrating the governed data lake



Hydrating the governed data lake



Source

Governed data lake

Target

Delivering business outcomes

Nationwide®
is on your side



Providing better marketing campaigns for our financial advisors



Developing ML models for fast, accurate insurance underwriting



Maximizing safe driving discounts for our members with SmartMiles®

Lessons learned



1. Optimize processing with proper file sizes
2. Use the right tools for the job
3. Think about the scale
4. Loose coupling
5. Partner with your AWS team

Road ahead



- Create a secure collaborative platform to spur innovation
  - Expand data-sharing capability
  - Automatically create non-sensitive data views
  - Request access from the Nationwide Data Marketplace
  - Identify and lock down misclassified data

- Build additional capabilities
  - Support for streaming
  - Change Data Capture (CDC)
  - Complete NGC integration

Thank you!



AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Please complete the session survey

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.