De-identifying Cloud Data Pipelines Webinar

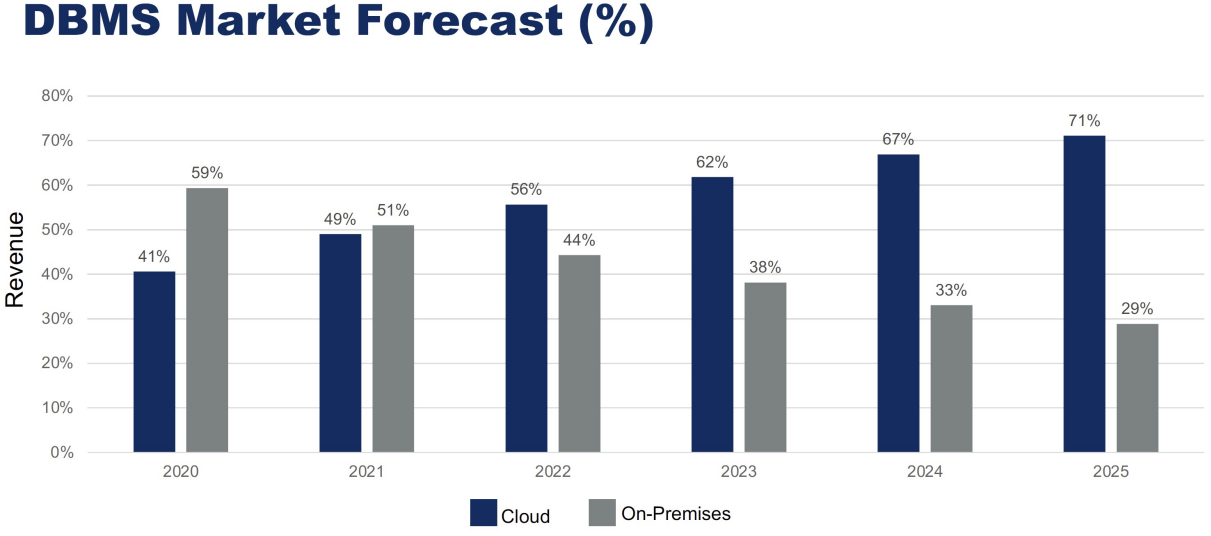
March 17, 2022

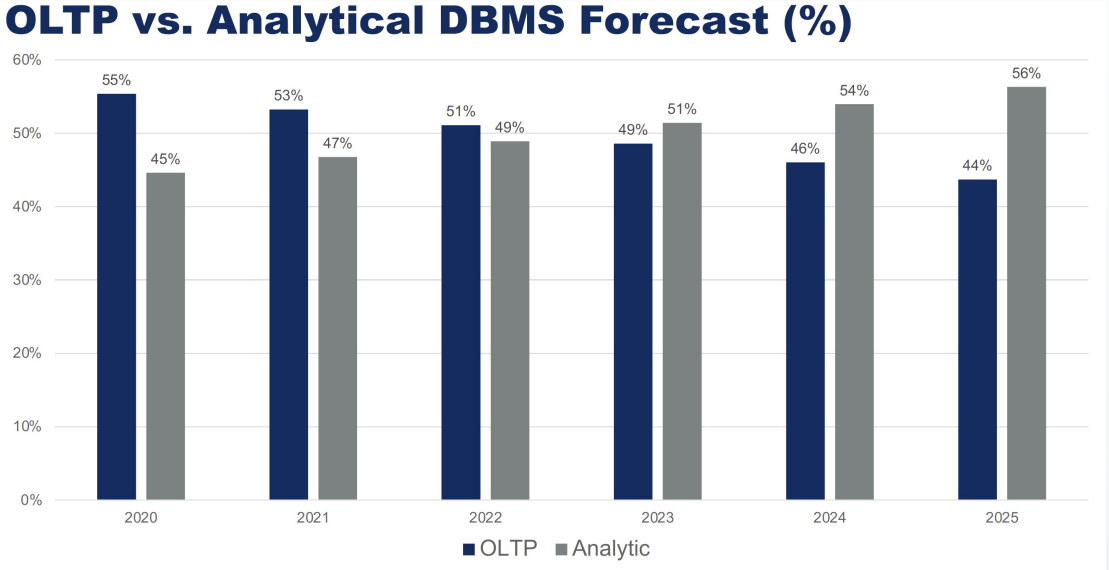
Ameesh Divatia
CEO & Co-founder

Sylvain Yelle
VP of Sales

Market Dynamics

- Growth of Cloud
  - dbPaaS > On-Premise
    - ~70% of DBMS by 2025
- Driven by:
  - Analytics > OLTP workloads
    - ~56% of workloads BY 2025

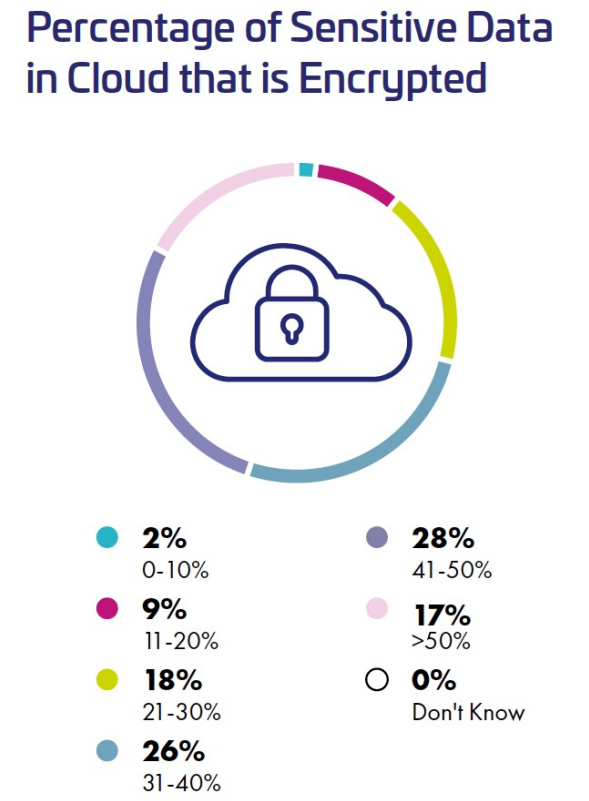




Source: Gartner, Q4 2021

Data Vulnerability

- 83% of sensitive data in the cloud is NOT protected



Source: Thales e-Security, Q4 2021

- 75% of the world's population will have its personal data protected by modern privacy regulations by the end of 2023

Source: Gartner, Q4 2021

Business Needs for Data De-Identification

- Migration to cloud
  - Digital transformation initiatives are accelerating the growth of cloud workloads
- Data analytics
  - Availability of sophisticated cloud-based analytics tools is the primary driver now
- Secure data sharing
  - Organizations are looking to collaborate on shared data sets without compromising data privacy

Methods of Securing Data

Parameter	Legacy Tokenization	Transparent Data Encryption (TDE)	Format Preserving Encryption (FPE)	Dynamic Masking
Operation	**Generate** token value based on a vault that contains all possible values -> doubles storage	**Encrypts** data at rest in storage. Data is decrypted when retrieved and is in the clear in the database server.	**Encrypts** using the AES algorithm to replace a sensitive data field with an encrypted value while preserving its size and data type.	**Obscures** sensitive data is either fully or partially based on personas (eg. last 4 digits of a credit card)
Performance	Degrades with data set size	Good performance	Hardware accelerated performance	Good performance
Protection	Secure unless there is a cardinality issue	Weak because it only protects physical disks	Very secure due to the robustness of the AES algorithm	Very Secure as original data is not exposed
Cloud Awareness	Not compatible because it requires installation of an agent on the database	Yes	Very cloud friendly with availability of key stores	Yes
Privacy Enhanced Computation	Not available	Not applicable as data is decrypted as soon as it is accessed	Yes, with external function support	Not possible

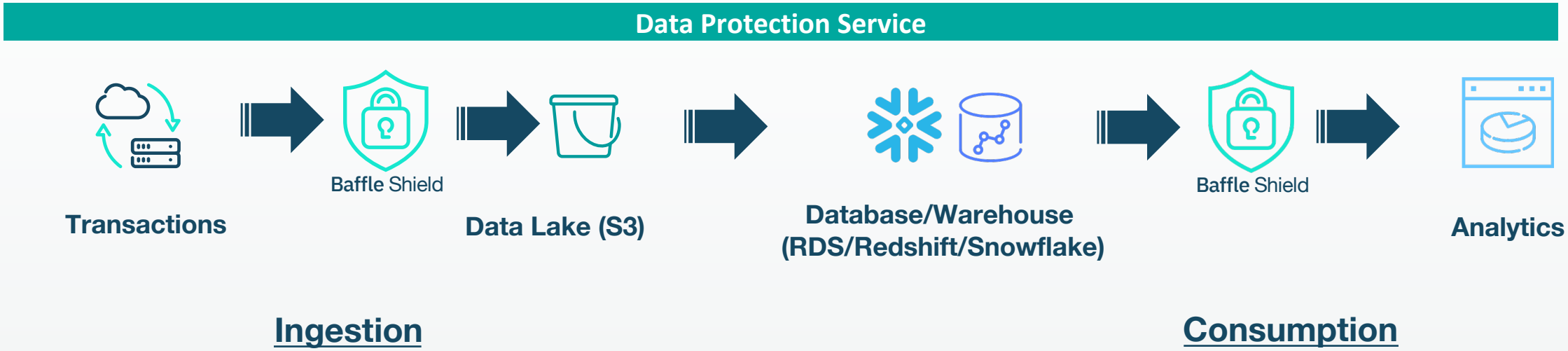
Next-Gen Data Protection Approaches

- Data Security Platform (DSP)[1]
  - Identified as a new Gartner category for data management
  - Consolidates various data discovery, transformation and monitoring capabilities into a unified service offering
- Privacy Enhanced Computation (PEC)[2]
  - Top 10 Gartner Strategic Technology Trend in 2022
  - Defined as the ability to process protected data ‘in use’
- Role-Based Access Control (RBAC)
  - Data transformation requires access control to be effective
  - Integration into existing IAM frameworks essential in rolling out an effective cloud-based data protection service

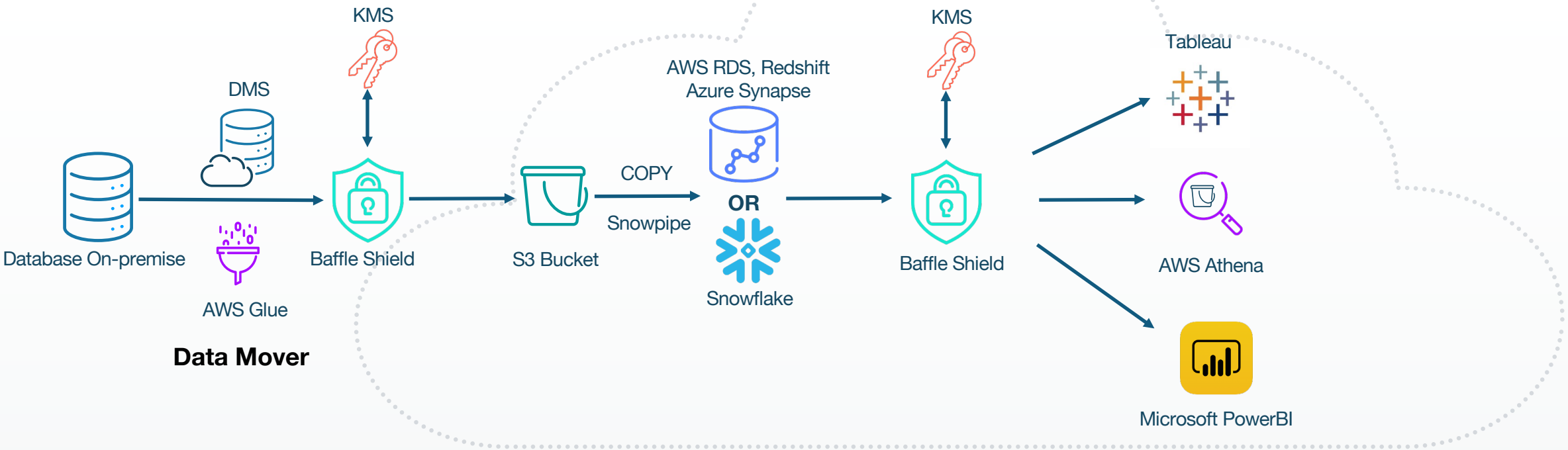
[1] Gartner® Report 2022 Strategic Roadmap for Data Security Platform Convergence
[2] Gartner® Top Strategic Technology Trends for 2022

Securing the Analytics Pipeline

Transforms data on the fly as it moves into the pipeline and
Control, who can access and use that data as it is consumed by the business



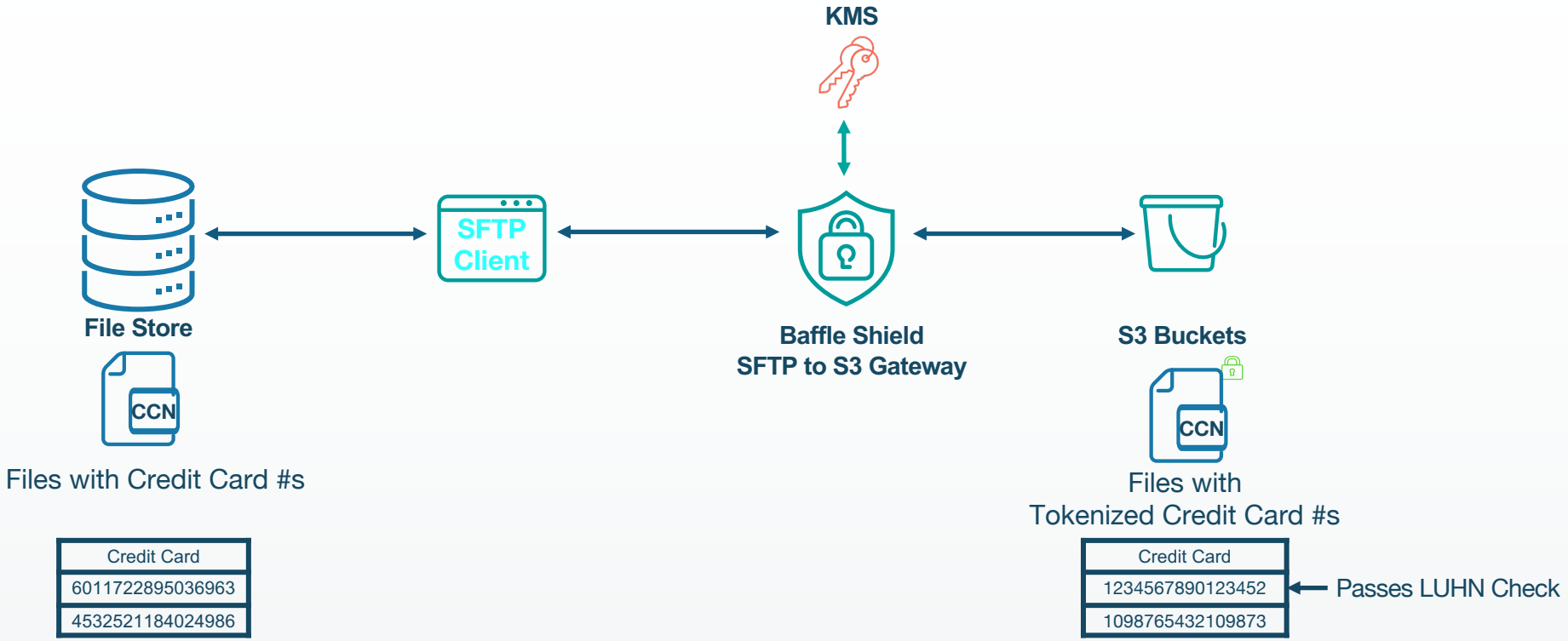
Data Protection Service



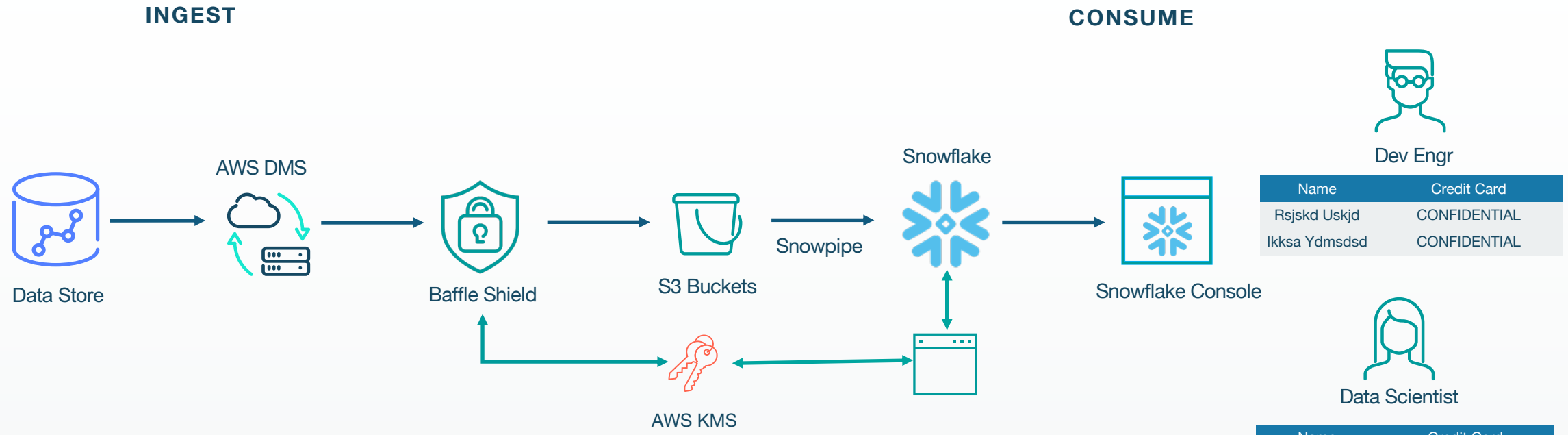
- Installed inline during data migration or extraction
- Protects sensitive data at the field-level
- No application changes required for consumption

Use Cases

File-level Protection – SFTP Scenario



Field-level Protection – Data Pipeline



Name	Credit Card
Michael Grace	6011722895036963
Steven Roelle	4532521184024986

Table with PII/PHI

Name	Credit Card
Rsjskd Uskjd	1234567890123452
Ikksa Ydmsdsd	1098765432109873

Table with tokenized PII/PHI

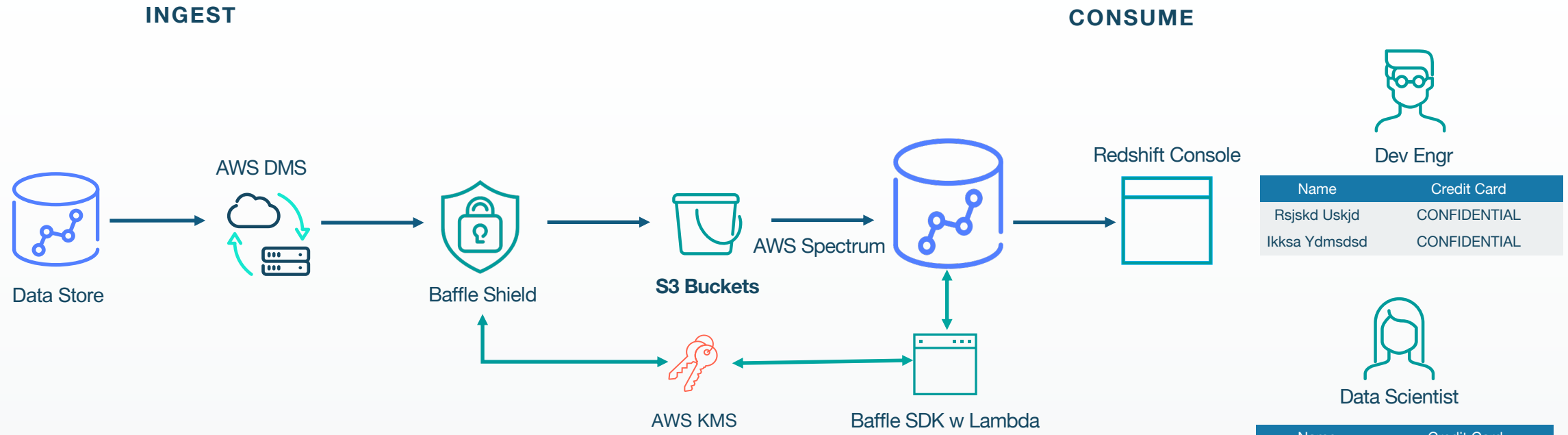
Baffle SDK w Lambda

Name	Credit Card
Rsjskd Uskjd	1234567890123452
Ikksa Ydmsdsd	1098765432109873

Table with tokenized PII/PHI

Persona-based Dynamic Masking

Field-level Protection – Data Pipeline



Name	Credit Card
Michael Grace	6011722895036963
Steven Roelle	4532521184024986

Table with PII/PHI

Name	Credit Card
Rsjskd Uskjd	1234567890123452
Ikksa Ydmsdsd	1098765432109873

Table with tokenized PII/PHI

Name	Credit Card
Rsjskd Uskjd	1234567890123452
Ikksa Ydmsdsd	1098765432109873

Table with tokenized PII/PHI

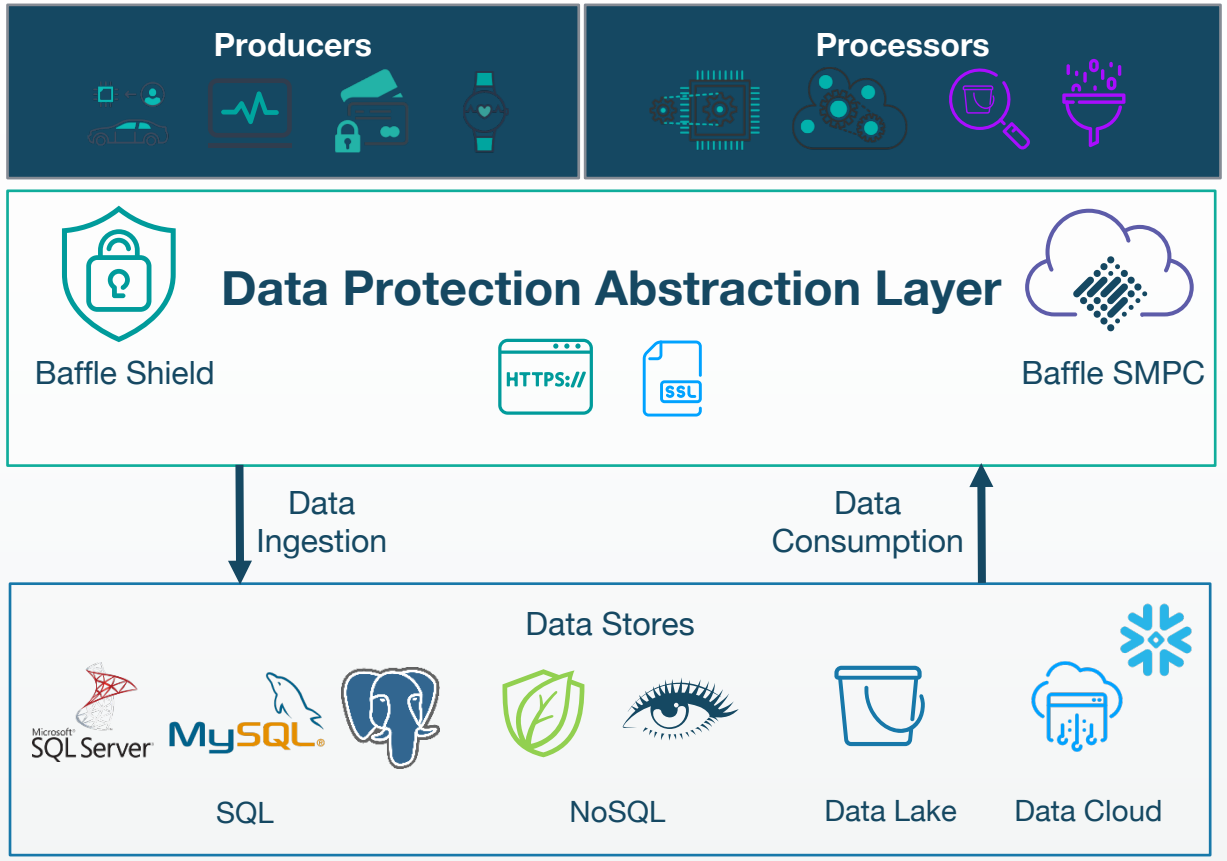
Persona-based Dynamic Masking

Demo

Benefits of a Data Protection Service

- **Ingest** sensitive data directly in a de-identified state
  - Field-level tokenization or encryption without any application changes
  - No visibility of sensitive data to the cloud administrator
- **Consume** sensitive data with:
  - Analytics applications processing sensitive data including mathematical operations on de-identified fields
  - Policy based field-level access control by persona
  - Comprehensive key management including rotation and retirement

Summary



Data producers and processors working together without compromising privacy

Q & A

Contact us at:	**info@baffle.io**
Request a Demo at:	**https://baffle.io/request-a-demo/**
Download the Gartner Report at:	**https://tinyurl.com/2p97hv2d**

Thank You