AWS
re:Invent

MKT212

Deep dive on using AWS Data Exchange for ML and data analytics

Ravi Turlapati

Senior Technical Product Manager
AWS Data Exchange
Amazon Web Services

Kanchan Waikar

Senior Partner Solutions Architect
AWS Marketplace
Amazon Web Services

Kartik Iyengar

Chief Technology Officer
Life Sciences Practice
Virtusa

AWS re:Invent

aws

Realizing the potential of data science

- What is the probability of a patient being diagnosed with chronic kidney disease within 3 years of being diagnosed with type 2 diabetes?

- What is the probability of a patient being diagnosed with stage 4 lung cancer provided that a given set of genomic markers are present?

- What medication should be prescribed to a diabetes patient, provided we have a complete history of glucometer readings?

Agenda

How is data analysis powering better decision-making?
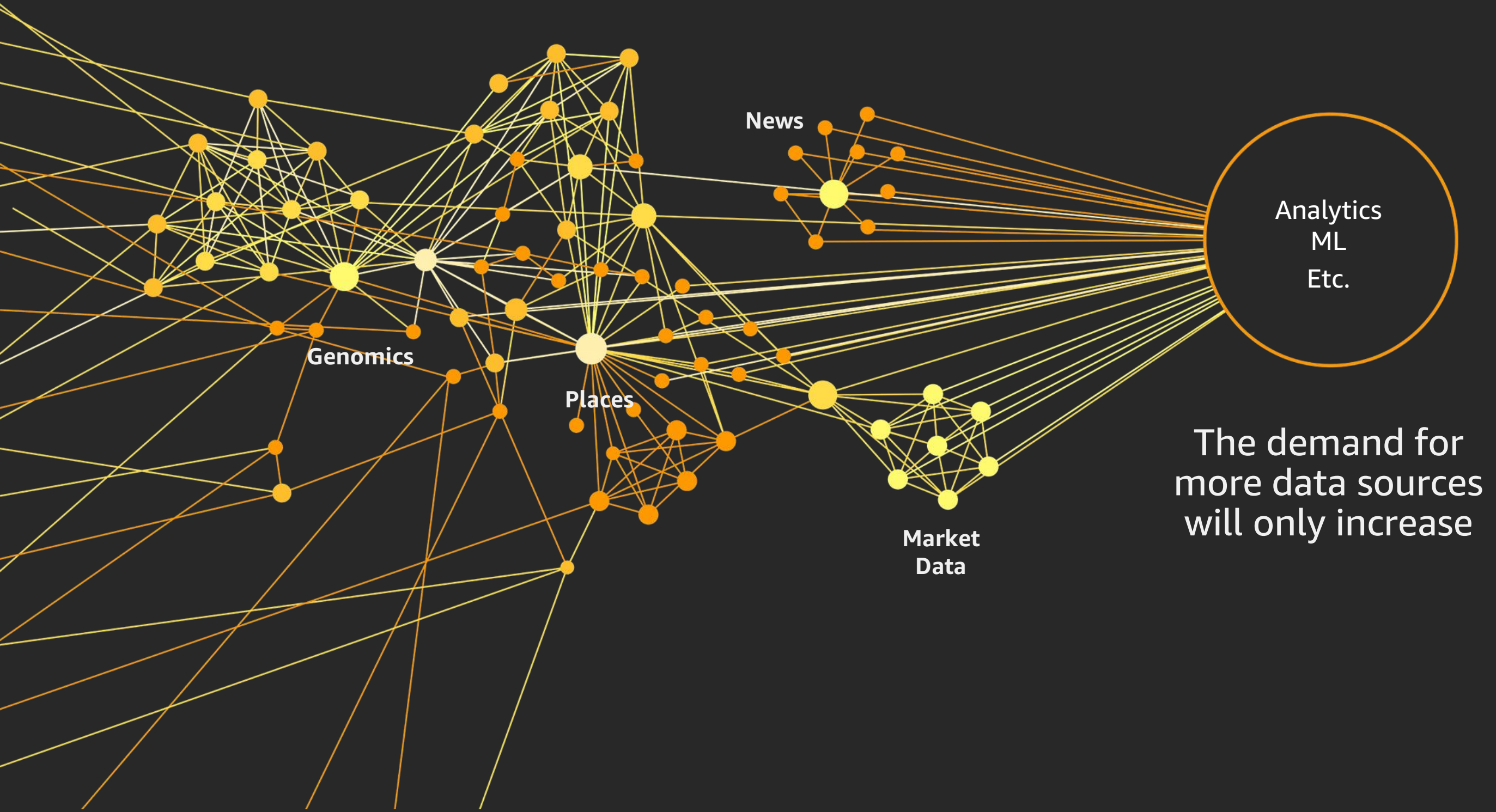

Introducing AWS Data Exchange

Demo—Using AWS Data Exchange as a provider and a subscriber

Demo—training machine learning (ML) models in Amazon SageMaker using third-party data and algorithms from AWS Marketplace

---

vLife's experience using AWS Data Exchange to distribute synthetic data sets and power data science

---

Q&A

Data is transforming the way businesses innovate

News

Analytics
ML
Etc.

Genomics

Places

Market
Data

The demand for
more data sources
will only increase

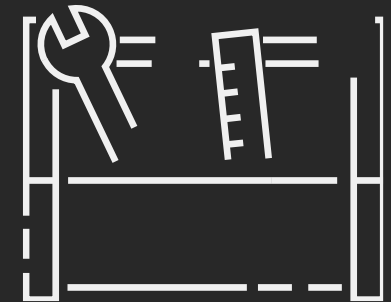# Customer expectations have been transformed, too

Everything easy to use

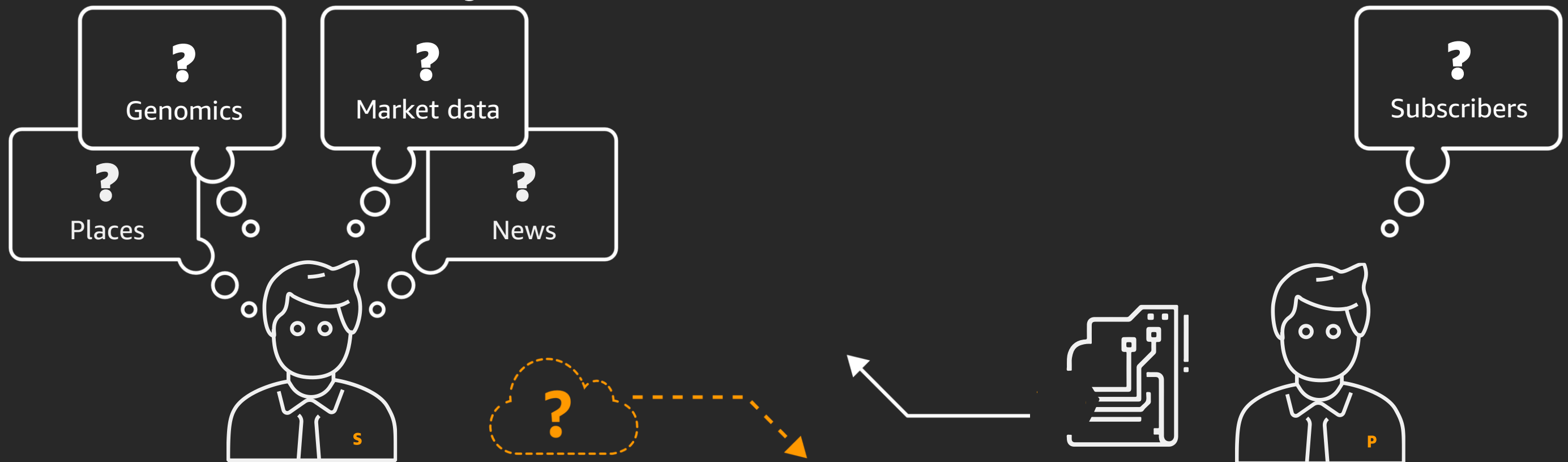**Everything cloud native**

Seamless experience

Self-service preference

Innovation is limited when you can't find the data you need, even when there is a data provider that wants to share it with you
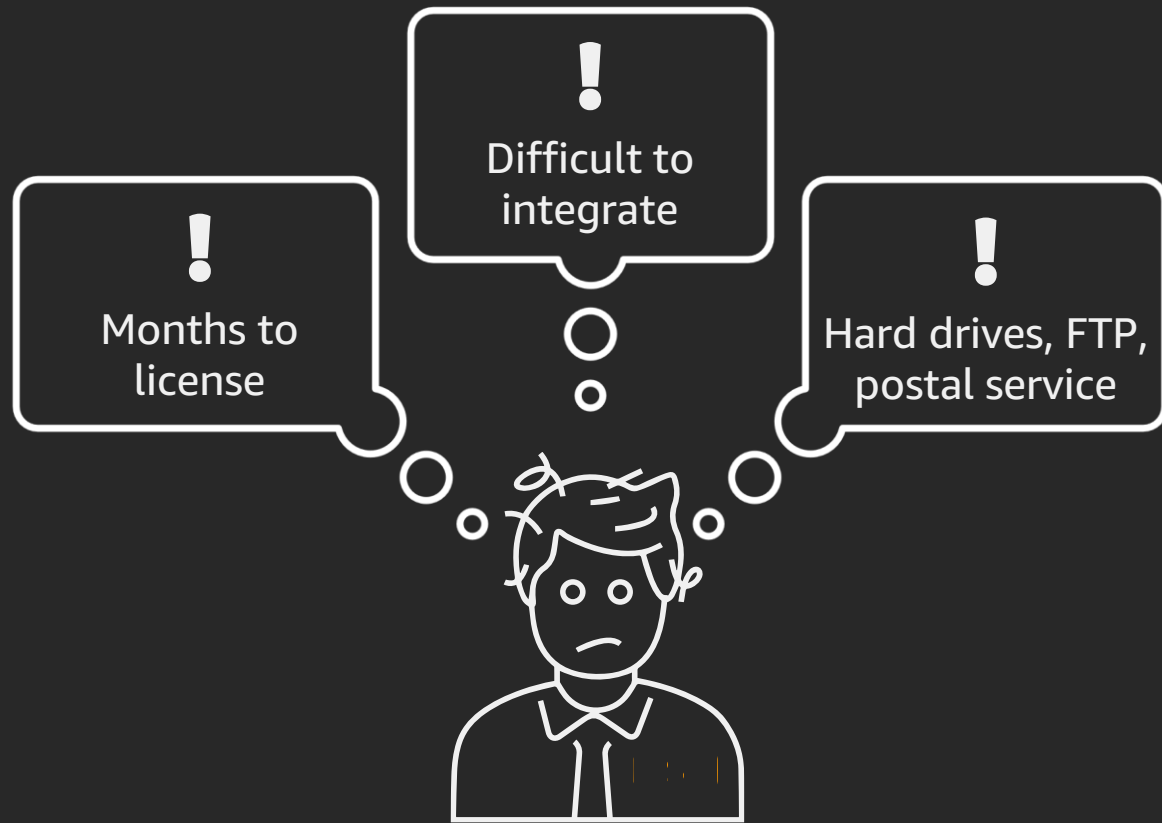
It's simply too hard for organizations to exchange data today

Difficult to integrate

Months to license

Hard drives, FTP, postal service

Subscriber

Complicated contracts

Inconsistent payments

Customer acquisition cost
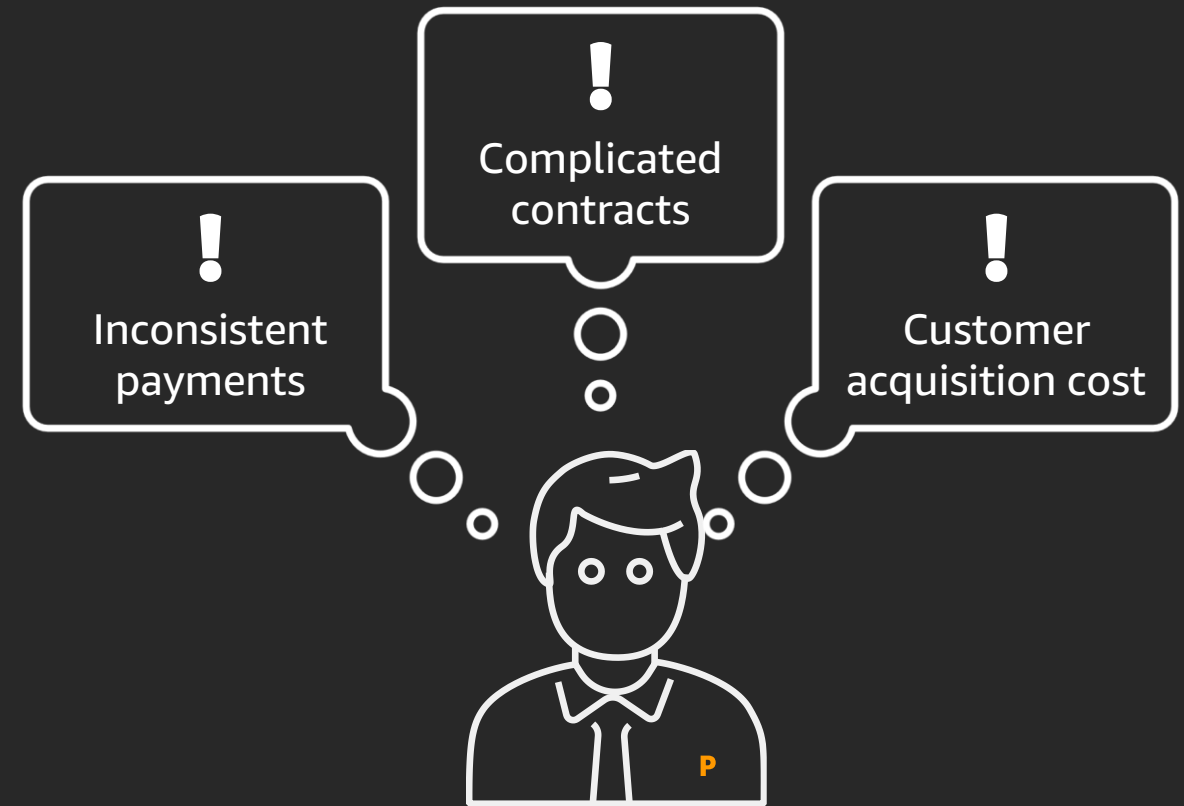
Provider

It's simply too hard for organizations to exchange data today

Months to license

Difficult to integrate

Hard drives, FTP, Postal service

Frustrated subscribers

Subscriber

Fewer data products in market

Inconsistent payments

Complicated contracts

Customer acquisition cost

Provider
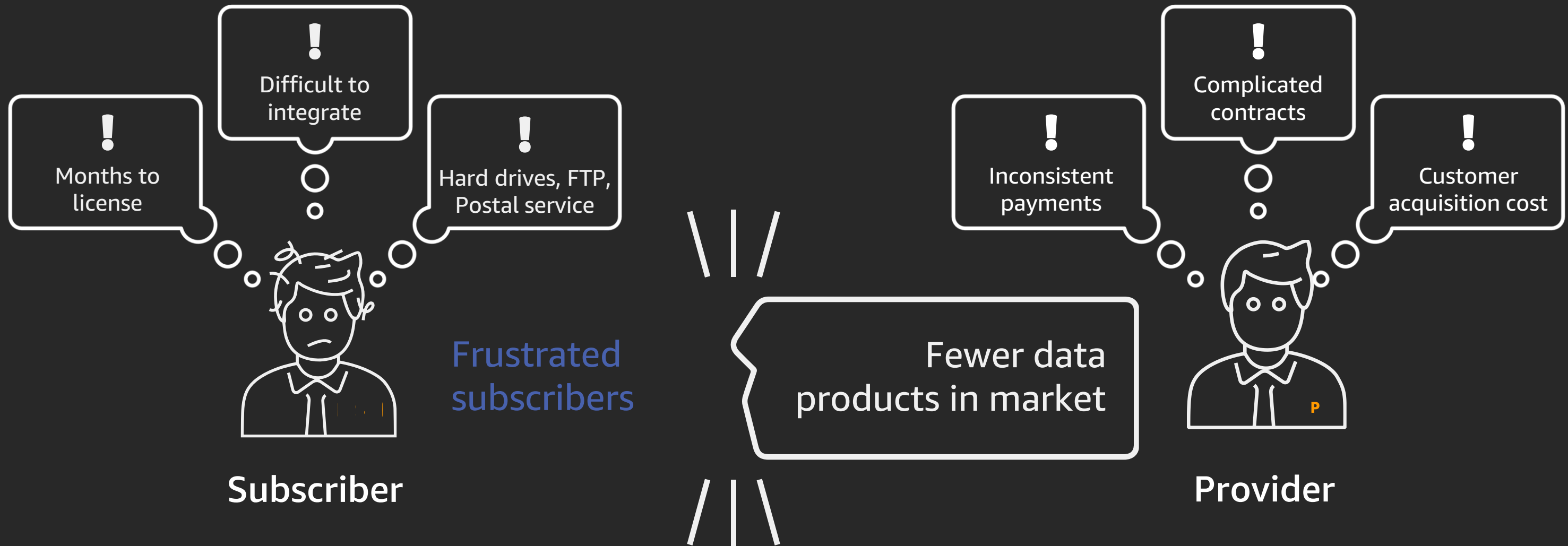
AWS Data Exchange makes it easy for AWS customers to find, subscribe to, and use third-party data in the cloud
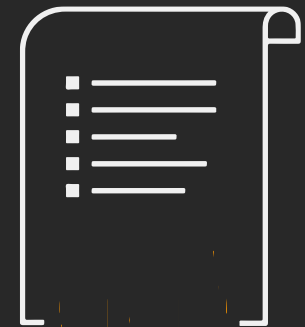
One place to exchange all kinds of data
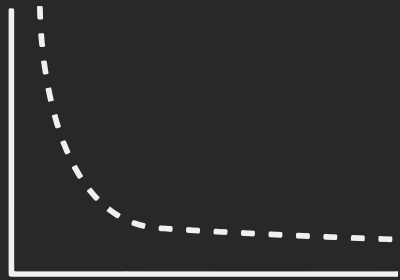
Natively integrate data into AWS

Easy subscription management

Simplified and secure billing and provisioning

AWS Data Exchange makes it easy for data providers to reach customers, too!
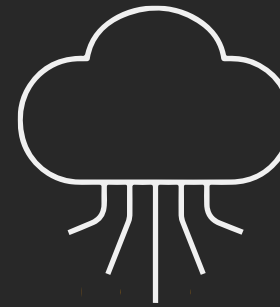
Efficiently reach the long-tail of subscribers on AWS

AWS Data Exchange

Join a growing breadth of providers on AWS
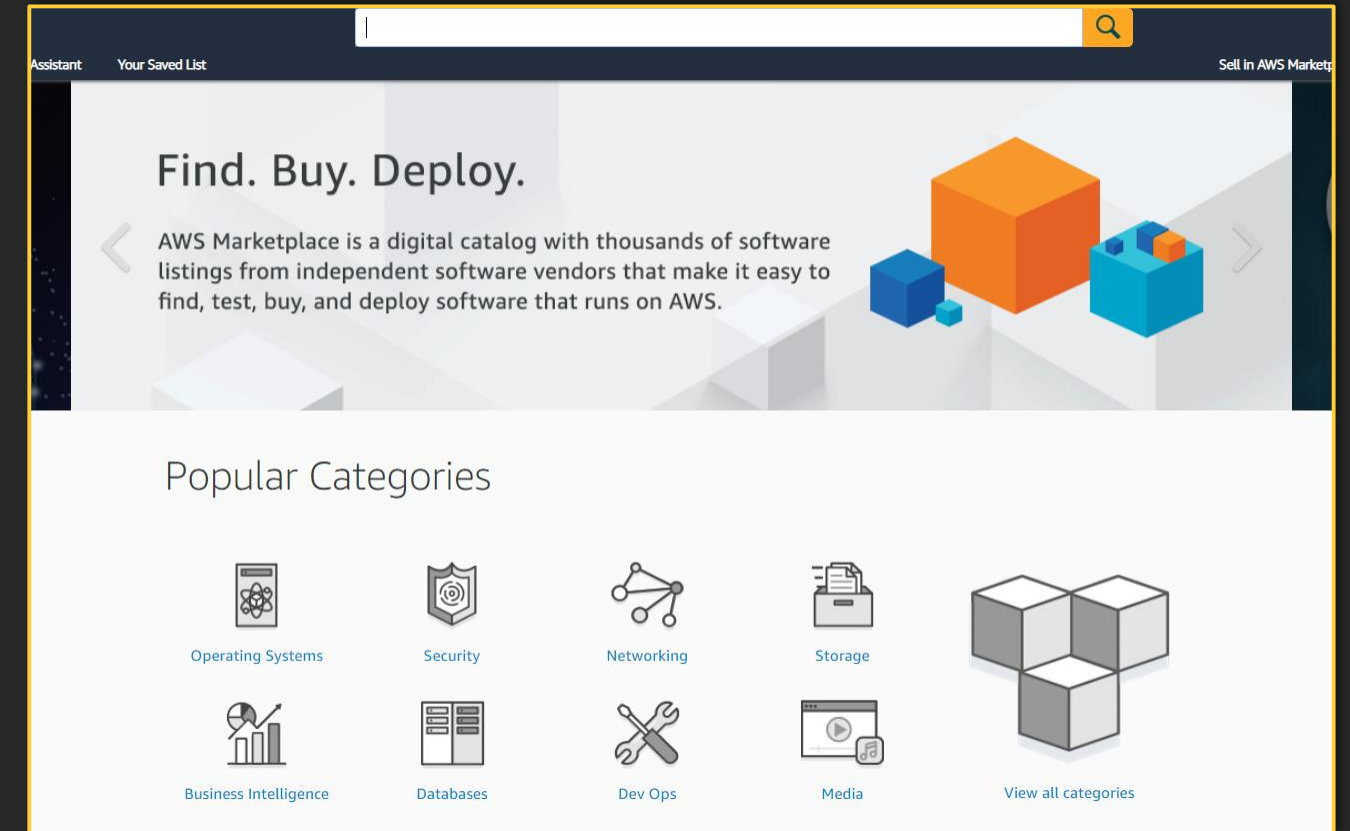
Self-service, create data sets, in a cloud native way

No longer need to build and maintain data storage, delivery, billing, or entitling technology
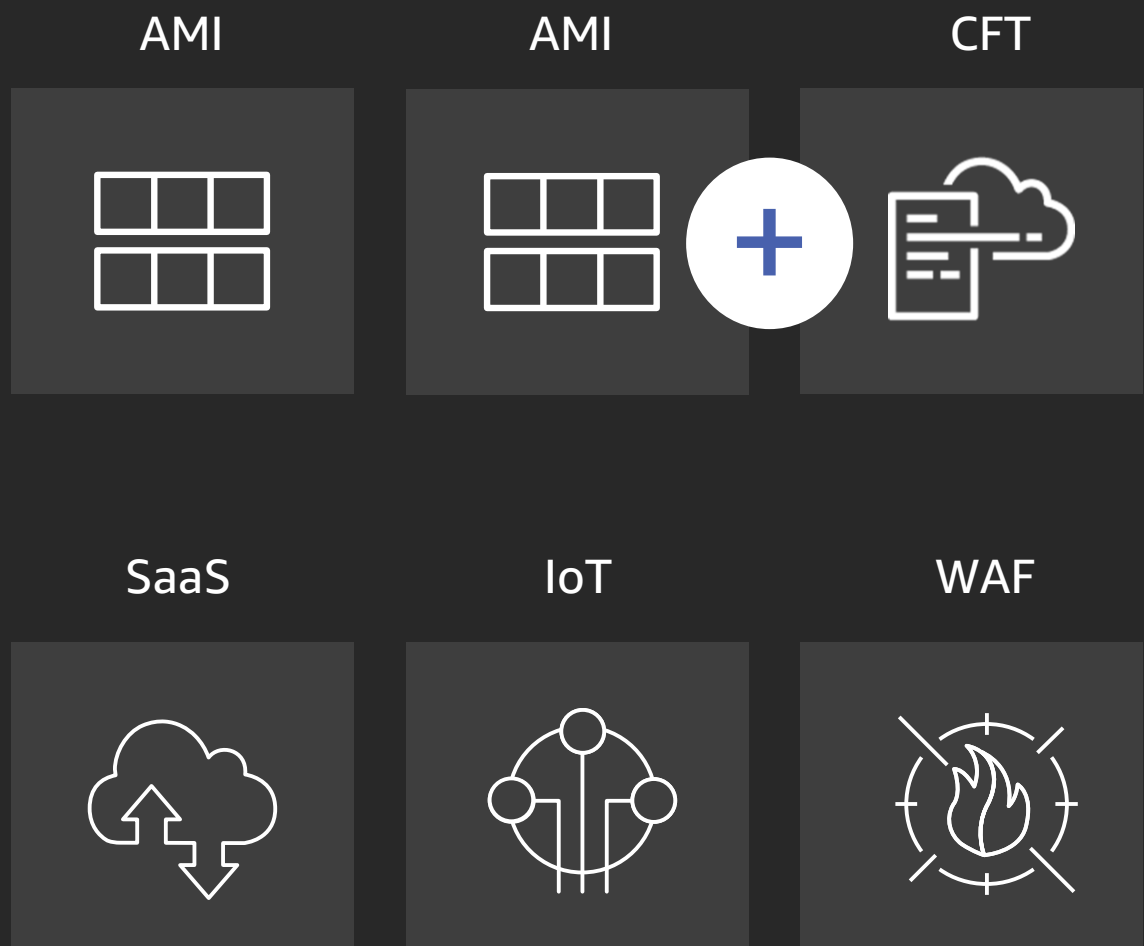
A growing digital software catalog

aws marketplace

- Deploy software on demand
- **1,500+** ISVs
- Over **7,000** product listings
- **260,000+** active customers
- Over **850 million** hours of Amazon EC2 deployed monthly
- Deployed in **20 Regions**
- Offers **39 categories**

- Flexible consumption and contract models
- Easy and secure deployment, almost instantly
- One consolidated bill
- Always evolving

Assistant   Your Saved List   Sell in AWS Marketp

Find. Buy. Deploy.

AWS Marketplace is a digital catalog with thousands of software listings from independent software vendors that make it easy to find, test, buy, and deploy software that runs on AWS.

Popular Categories

Operating Systems	Security	Networking	Storage
Business Intelligence	Databases	Dev Ops	Media

View all categories

Last year's re:Invent launch: Machine learning

AMI

AMI

CFT

SaaS

IoT

WAF

2018
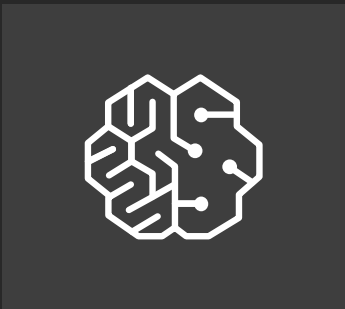
Amazon SageMaker

Containers

New: Data products from AWS Data Exchange

AMI

AMI **+** CFT

Amazon
SageMaker

SaaS

IoT

WAF

Containers
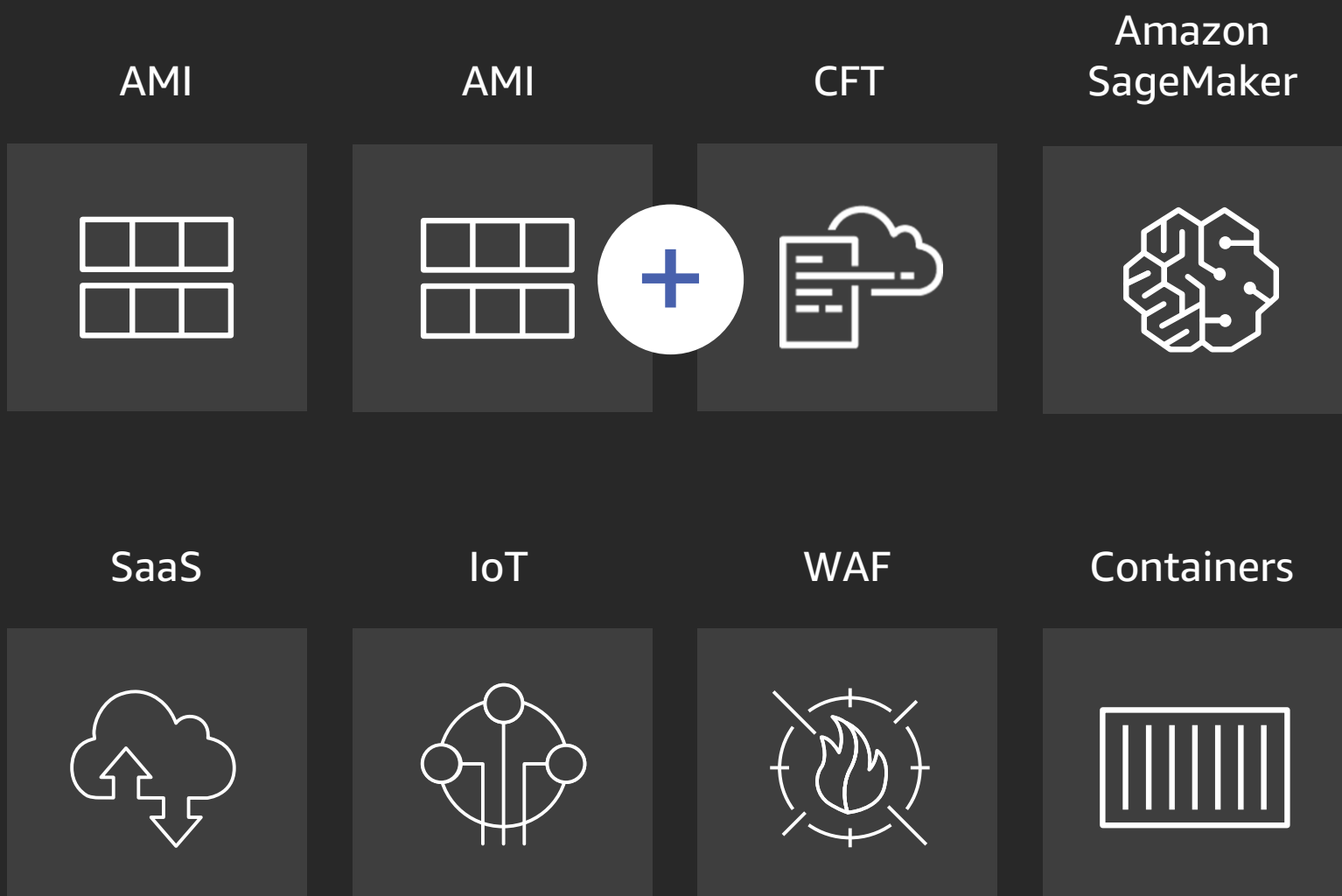
NEW

+

AWS Data Exchange products in
AWS Marketplace

Natively integrated with AWS

Global distribution through AWS

Upload, package, and curate data sets directly from Amazon S3

S3

Easy-to-use API and console experience

Analyze AWS Data Exchange catalog with Amazon Athena, Amazon Redshift and AWS Glue

Consolidated AWS billing; pay only for the data licensed

AWS Data Exchange is seamlessly integrated with downstream AWS analytics and ML services
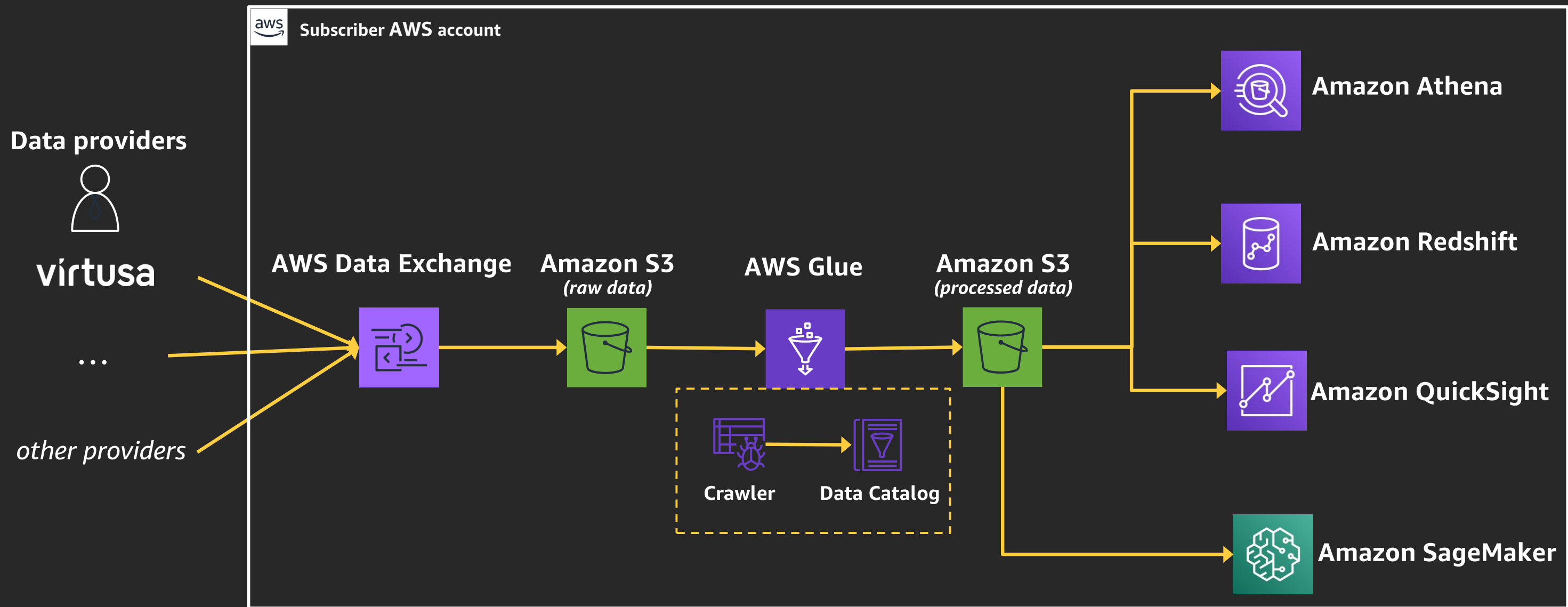
Automatically copy your AWS Data Exchange data sets to your AWS data lake on Amazon S3

AWS Data Exchange

New revision

New dataset revision

Subscriber AWS account

AWS Data Exchange
Subscription

12 month subscription

CloudWatch Event

Amazon CloudWatch
Event Rule

Source:
AWS.dataexchange

Detail-type:
Revision published to dataset

Event trigger

AWS Lambda
Function

Create and start export job
Up to 100 assets per job

Export assets

Amazon S3
Bucket
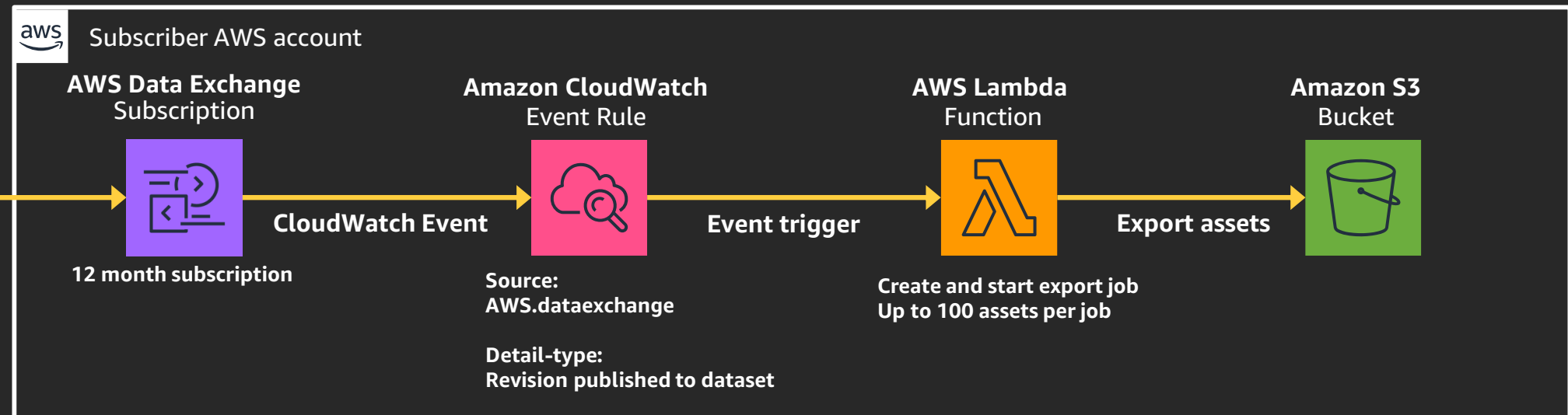
Easy subscription management

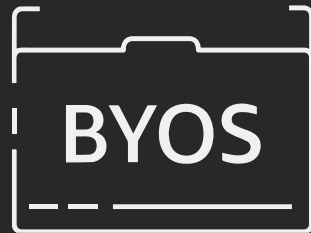**Bring your own subscription terms**

Automatically consume new data/updates via CloudWatch Events

Manage all your data in one place
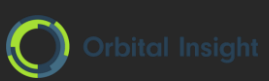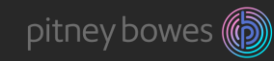
Built-in security and compliance controls

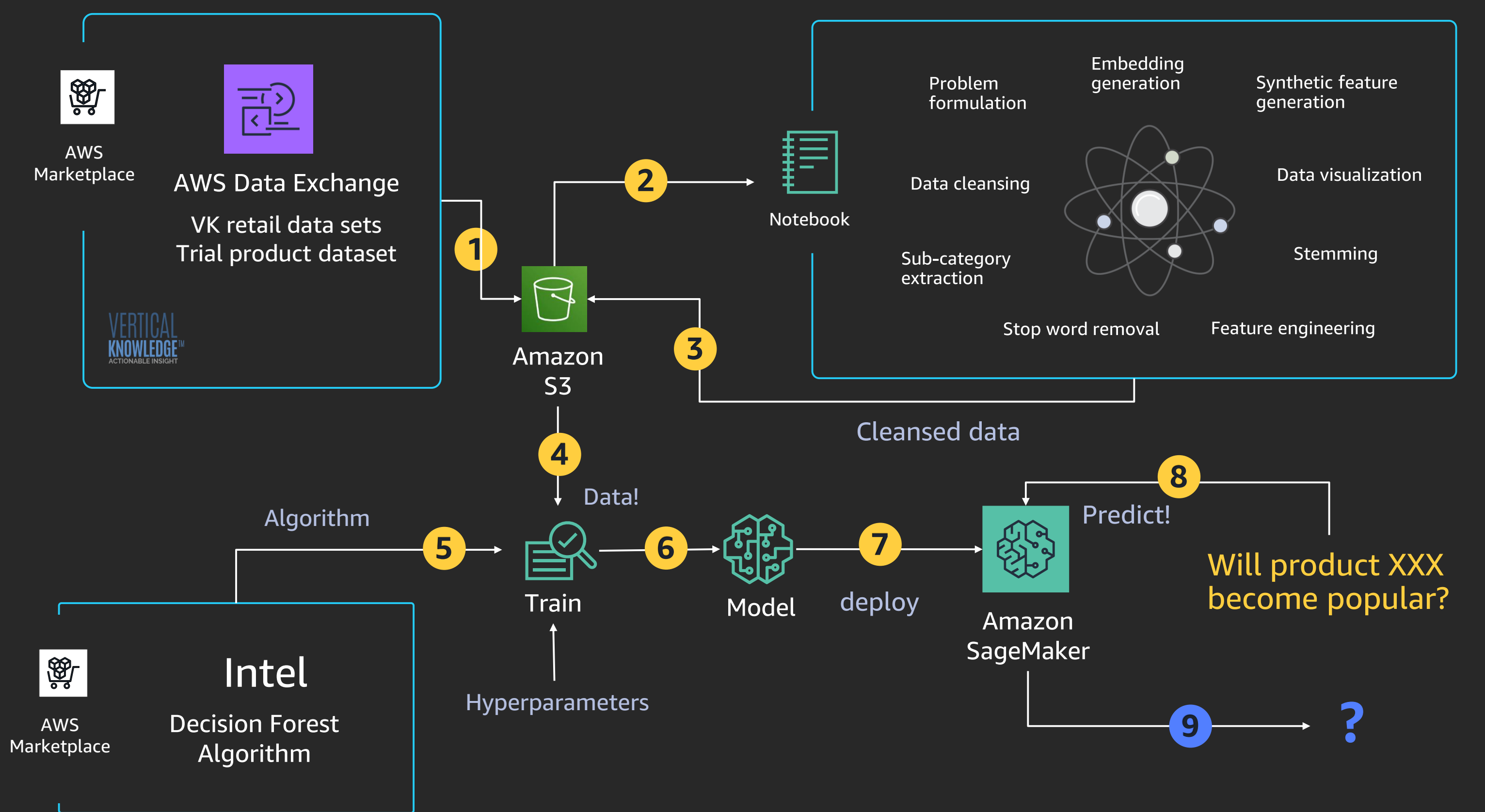**Encrypted at rest and in transit**

Subscription verification, compliant with GDPR, and more attestations on the way

Integrated with AWS Identity and Access Management (IAM)
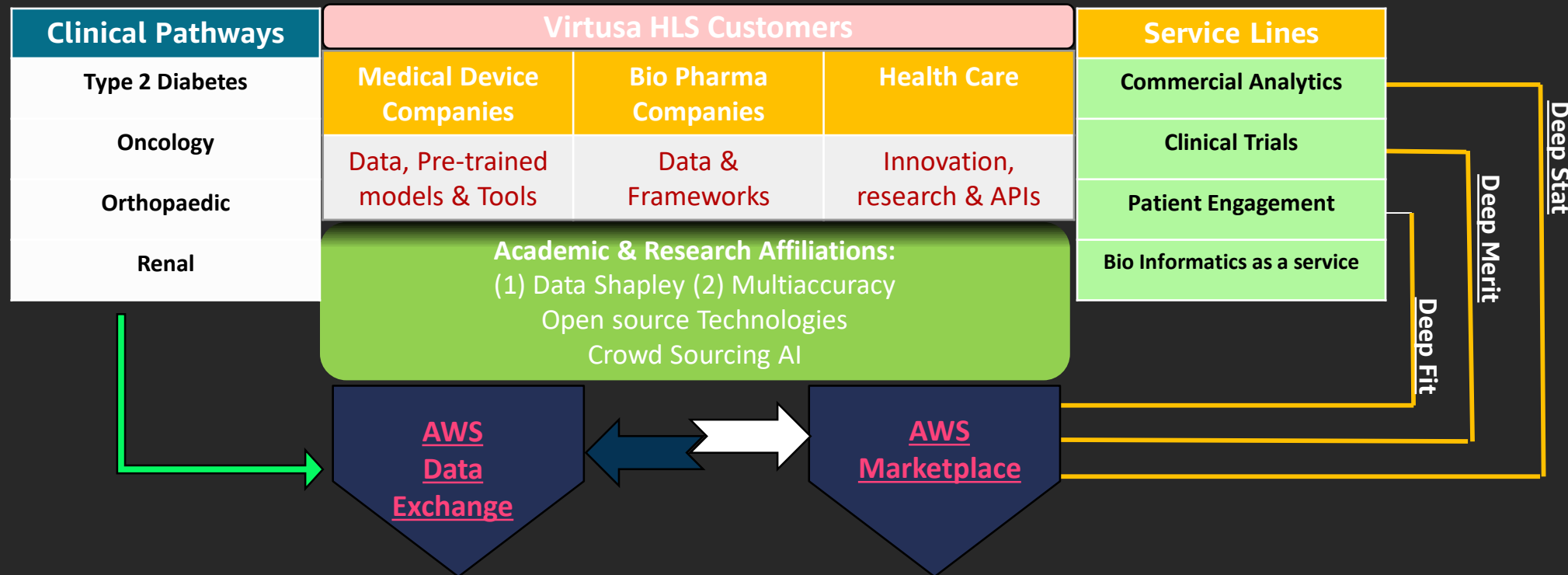
AWS Data Exchange demo

Virtusa

aws

Introduction to vLife

- A cloud-based innovation platform built by Virtusa to solve problems related to data in the life sciences domain

- A HIPAA-compliant data lake containing 9 large datasets (EMR/Human Genome/Synthetic/RWD/Image/Device etc.)

- A platform to help our customers move from data to data Science

Clinical Pathways
Type 2 Diabetes
Oncology
Orthopaedic
Renal

Virtusa HLS Customers		
Medical Device Companies	Bio Pharma Companies	Health Care
Data, Pre-trained models & Tools	Data & Frameworks	Innovation, research & APIs

Academic & Research Affiliations:
(1) Data Shapley (2) Multiaccuracy
Open source Technologies
Crowd Sourcing AI

Service Lines
Commercial Analytics
Clinical Trials
Patient Engagement
Bio Informatics as a service

AWS Data Exchange

AWS Marketplace

Deep Stat

Deep Merit

Deep Fit

virtusa

vLife

vLife's journey on the AWS Marketplace

5 listed ML-related product offerings on the AWS Marketplace

Published 35 data products containing electronic health records using AWS Data Exchange
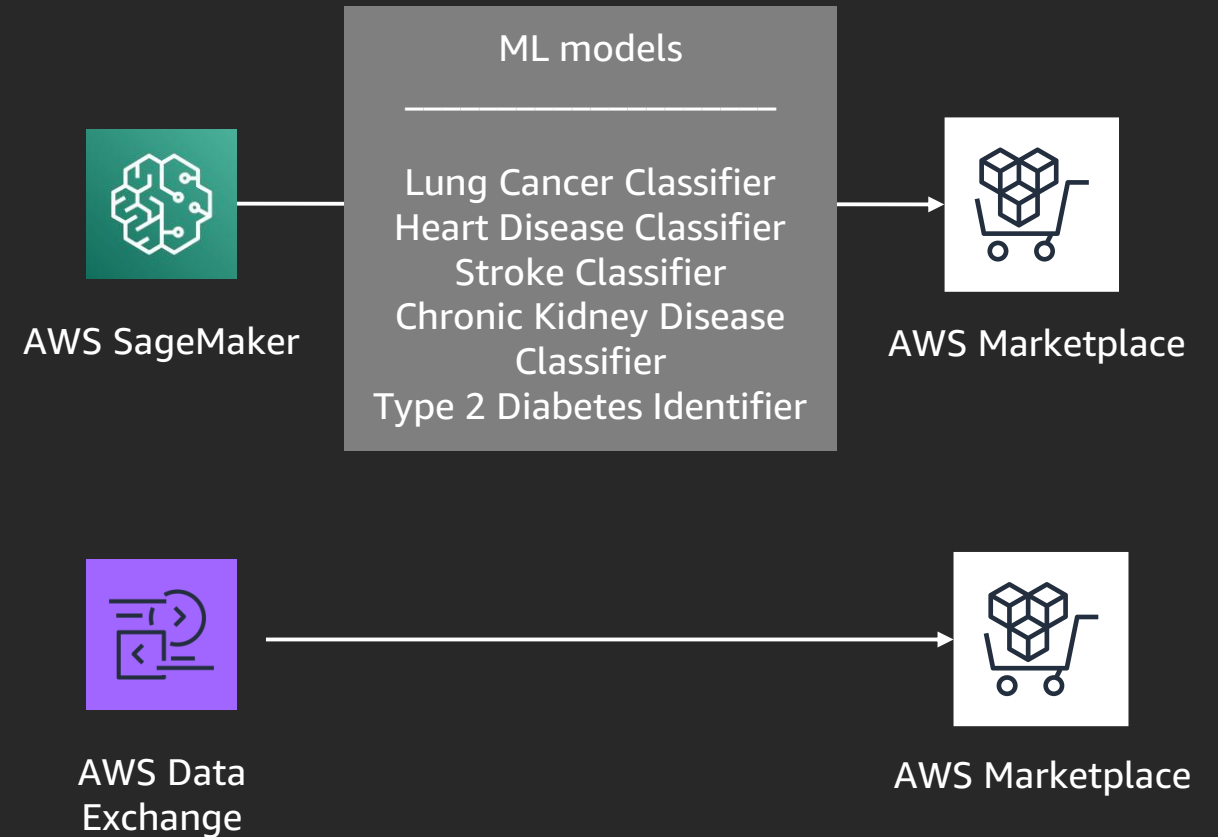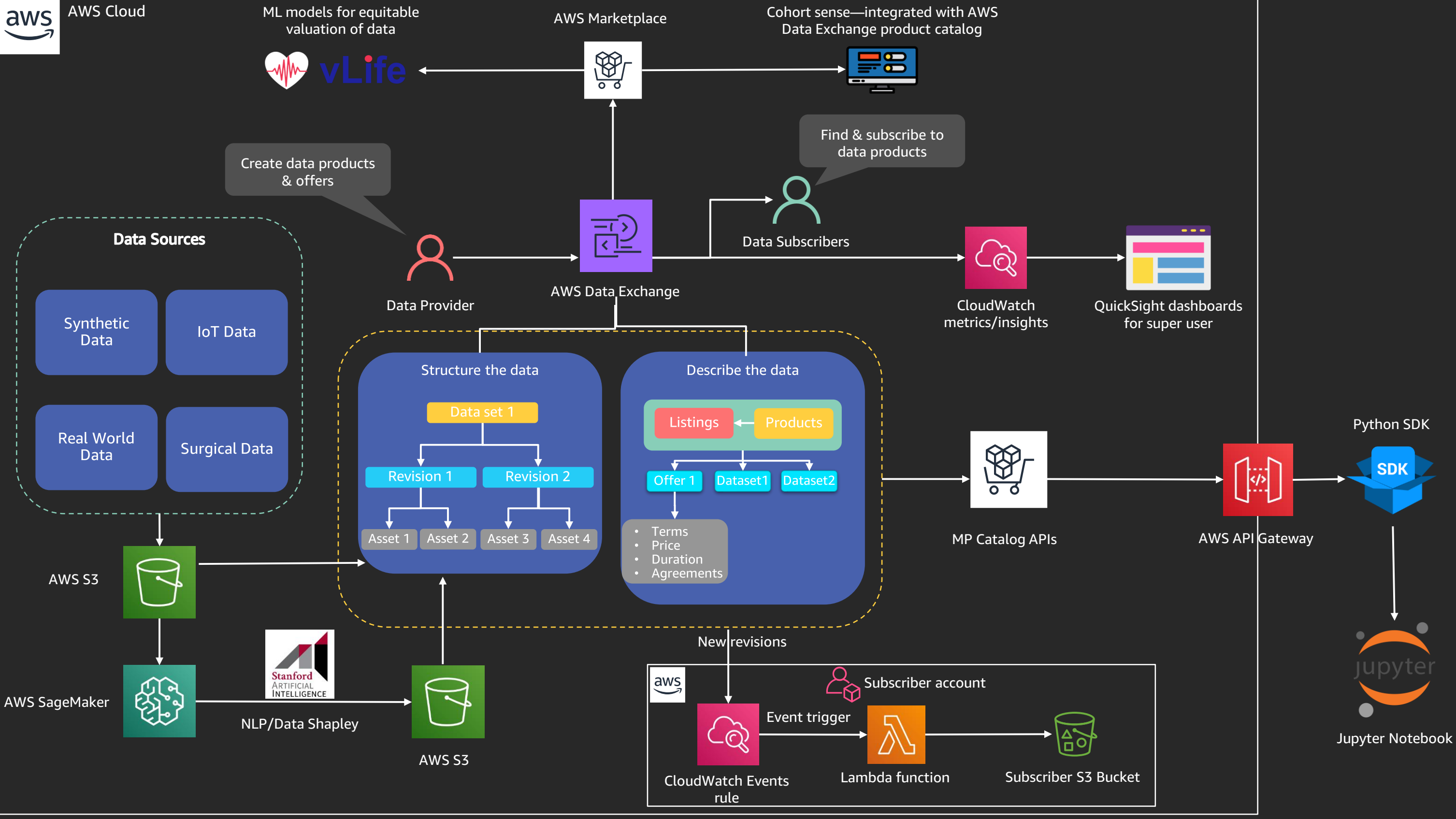

Simulated US population data across 35 ailments

364 million virtual lives

815 billion records

96 TB total size

Offer RWD/simulated and dark data via vLife

ML models
_____

Lung Cancer Classifier
Heart Disease Classifier
Stroke Classifier
Chronic Kidney Disease Classifier
Type 2 Diabetes Identifier

AWS SageMaker

AWS Marketplace

AWS Data Exchange

AWS Marketplace

Distributing synthetic data sets using AWS Data Exchange to aid data science

AWS Cloud

AWS EC2 synthetic data generator

Synthetic data
_____

Allergic rhinitis
.
35 ailments
.
Veteran substance abuse

Patients
Medication Requests
Conditions
Encounters
Procedures
Observations
Attributes
Quality of Life
Providers

AWS S3

1

2

AWS Data Exchange

Datasets
_____

Synthetic data for
allergic rhinitis
.
35 datasets
.
Synthetic data for veteran
substance abuse

3

AWS Marketplace

Synthetic data
products using 1
or more datasets

Data science use cases enabled via AWS Data Exchange

Surgical data (Smith & Nephew)

Synthetic data (Cardinal Health)

Real-world data (Foodbuy LLC)

Dark data (Foodbuy LLC)

IoT data (Modjoul)

Test data (GenRocket)

Impact: Questions we can now answer through data science

- What is the probability of a patient being diagnosed with chronic kidney disease within 3 years of being diagnosed with type 2 diabetes?

- What is the probability of a patient being diagnosed with stage 4 lung cancer provided that a given set of genomic markers are present?

- What medication should be prescribed to a diabetes patient, provided we have a complete history of glucometer readings?

Thank you!

Ravi Turlapati

rtturlap@amazon.com

AWS re:Invent

aws

Please complete the session survey in the mobile app.