aws SUMMIT

INDIA | MAY 25, 2023

Sony Pictures transforming business using AWS Analytics and AI/ML services

Mahesh Patil
Head IT - Infrastructure, Analytics & Data Lake
Sony Pictures Network India

Rahul Sureka
Enterprise Solutions Architect
AWS India

aws

Agenda

- AWS modern data strategy for M&E

- SPNI business transformation journey

- About Sony Pictures & Vision

- SPNI data lake architecture

- MLOps platform

- Building on modern data platform

- Benefits of building on AWS

Key media and entertainment business drivers

Do you have all the insights needed to drive your business?

Drive advertising revenue & reduce content costs

Increase lifetime value of user

Attract new customers & reduce churn

Innovate & Accelerate outcomes

Common analytics barriers
to driving business outcomes

- Limited insight into customer experiences and operations

- Need to make quicker decisions

- Enabling innovation with analytics & machine learning

- Technical staff & cost to scale self-managed infrastructures

THE ASPIRATION
Becoming a data-driven organization

"An organization that harnesses data as an asset, to drive sustained innovation and create actionable insights to supercharge the experience for their customers so they demand more…

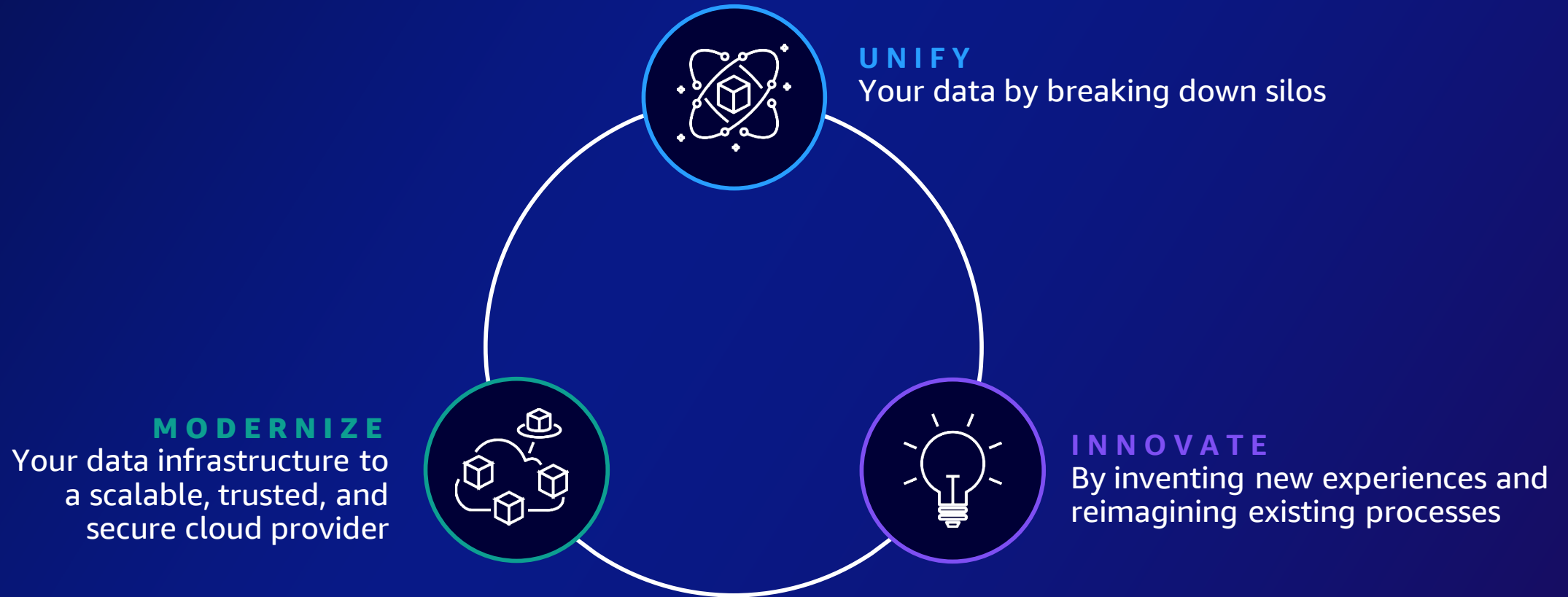**85%** of businesses want to be data driven

37% have been successful

Source: Forbes Online; New Vantage Partners Survey

Modern data strategy for better business outcomes
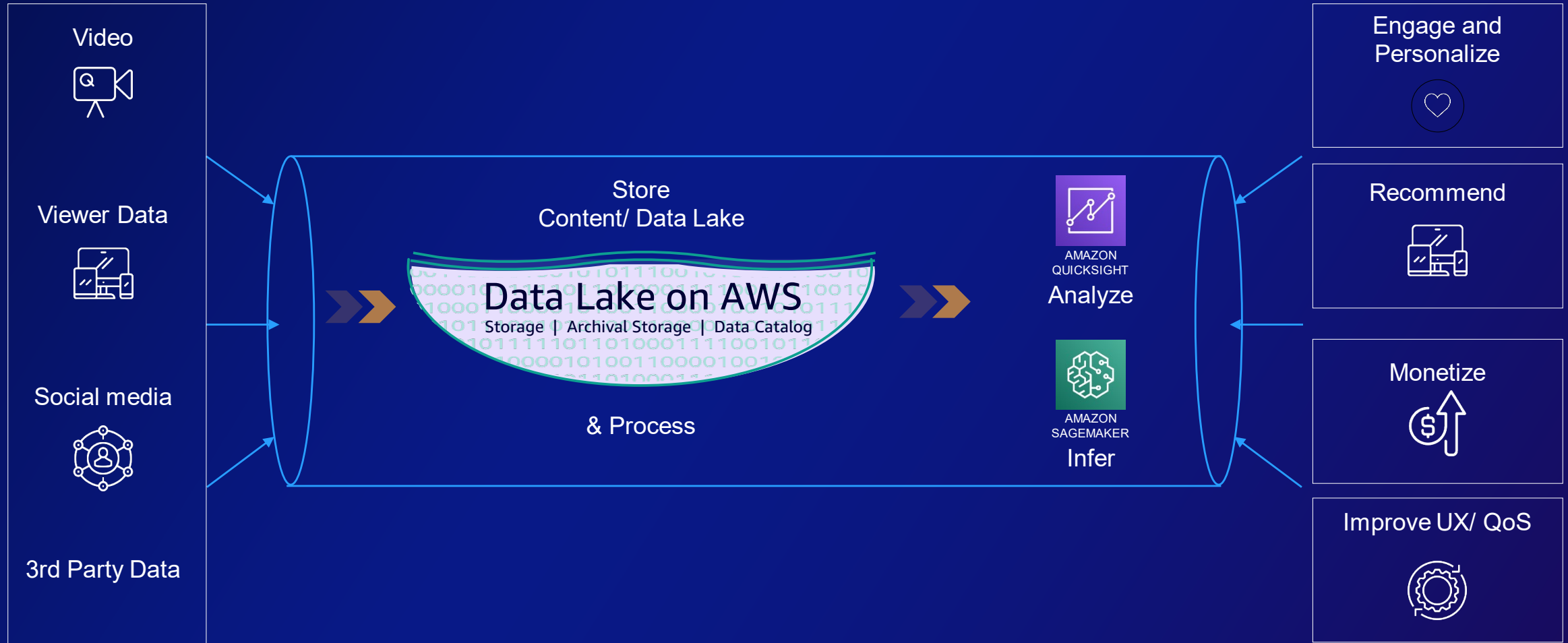
START ANYWHERE

UNIFY
Your data by breaking down silos

INNOVATE
By inventing new experiences and reimagining existing processes

MODERNIZE
Your data infrastructure to a scalable, trusted, and secure cloud provider

The AWS AI/ML stack

BROADEST AND MOST COMPLETE SET OF MACHINE LEARNING CAPABILITIES

AI SERVICES

SPECIALIZED

BUSINESS PROCESSES
Amazon Personalize
Amazon Forecast
Amazon Fraud Detector
Amazon Lookout for Metrics

SEARCH
Amazon Kendra

CONVERSATION
Amazon Lex
Amazon Transcribe Call Analytics
Contact Lens
Voice ID

CODE + DEVOPS
Amazon CodeGuru
Amazon CodeWhisperer
Amazon DevOps Guru

INDUSTRIAL
Amazon Monitron
Amazon Lookout for Equipment
Amazon Lookout for Vision

HEALTH
Amazon HealthLake
Amazon Comprehend Medical
Amazon Transcribe Medical
Amazon Omics

CORE

TEXT
Amazon Translate   Amazon Comprehend

SPEECH
Amazon Polly   Amazon Transcribe

VISION
Amazon Textract   Amazon Rekognition   AWS Panorama

AMAZON SAGEMAKER

CANVAS
No-code ML for business analysts

STUDIO LAB
Learn ML

GROUND TRUTH
Label data

STUDIO IDE

Prepare data

Store features

Geospatial ML

Build with notebooks

Train models

Tune parameters

Deploy in production

Manage and monitor

EDGE MANAGER
Manage edge devices

------------- CI/CD | GOVERNANCE | RESPONSIBLE ML -------------

ML FRAMEWORKS & INFRASTRUCTURE

PyTorch, Apache MXNet, TensorFlow | Amazon EC2 | CPUs | GPUs | AWS Inferentia | AWS Trainium | Habana Gaudi | FPGA

Modern data architecture for media enrichment

INSIGHTS TO ENHANCE VIEWER ENGAGEMENT, PERSONALIZATION, MONETIZATION

Video

Viewer Data

Social media

3rd Party Data

Store
Content/ Data Lake

Data Lake on AWS
Storage | Archival Storage | Data Catalog

& Process

AMAZON QUICKSIGHT
Analyze

AMAZON SAGEMAKER
Infer

Engage and Personalize

Recommend

Monetize

Improve UX/ QoS

Sony Pictures – Modern data strategy

Mahesh Patil
Head (Infrastructure and Analytics)
Sony Picture Network India

aws

We are a leading media company in India with meaningful presence in the TV and Digital segments

SONY PICTURES NETWORKS

.......and **A LOT MORE !**

TV Broadcasting (est. 1995)
A robust portfolio of 26 channels with market leading presence in Hindi entertainment and Sports

Digital platform (est. 2013)
Sony LIV – one of the most-watched OTT video platform in the country

Content production
SPN Productions - Movie production (est. 2013)*;
Studio NEXT – Content production (est. 2018)

*Now falls under SPE and not SPNI

SPNI | Challenges at scale

More data than ever was being generated with time. **Speed** and scaling was a concern

Data of all types was stored in **silos** across multiple data stores. No **single source of truth** for data
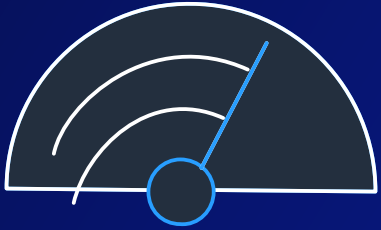
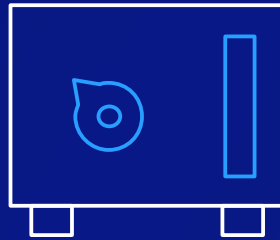Machine learning and real time cases adoption was challenged **legacy** designs

Demand for better **security**, privacy, and compliance regulations

SPNI | Our vision

More agility, and flexibility with dynamic scale

Comprehensive data analytics platform for all business units

Single source of truth

Faster time to market with lower TCO

Automated CI-CD and Better governance

Future ready to innovate faster
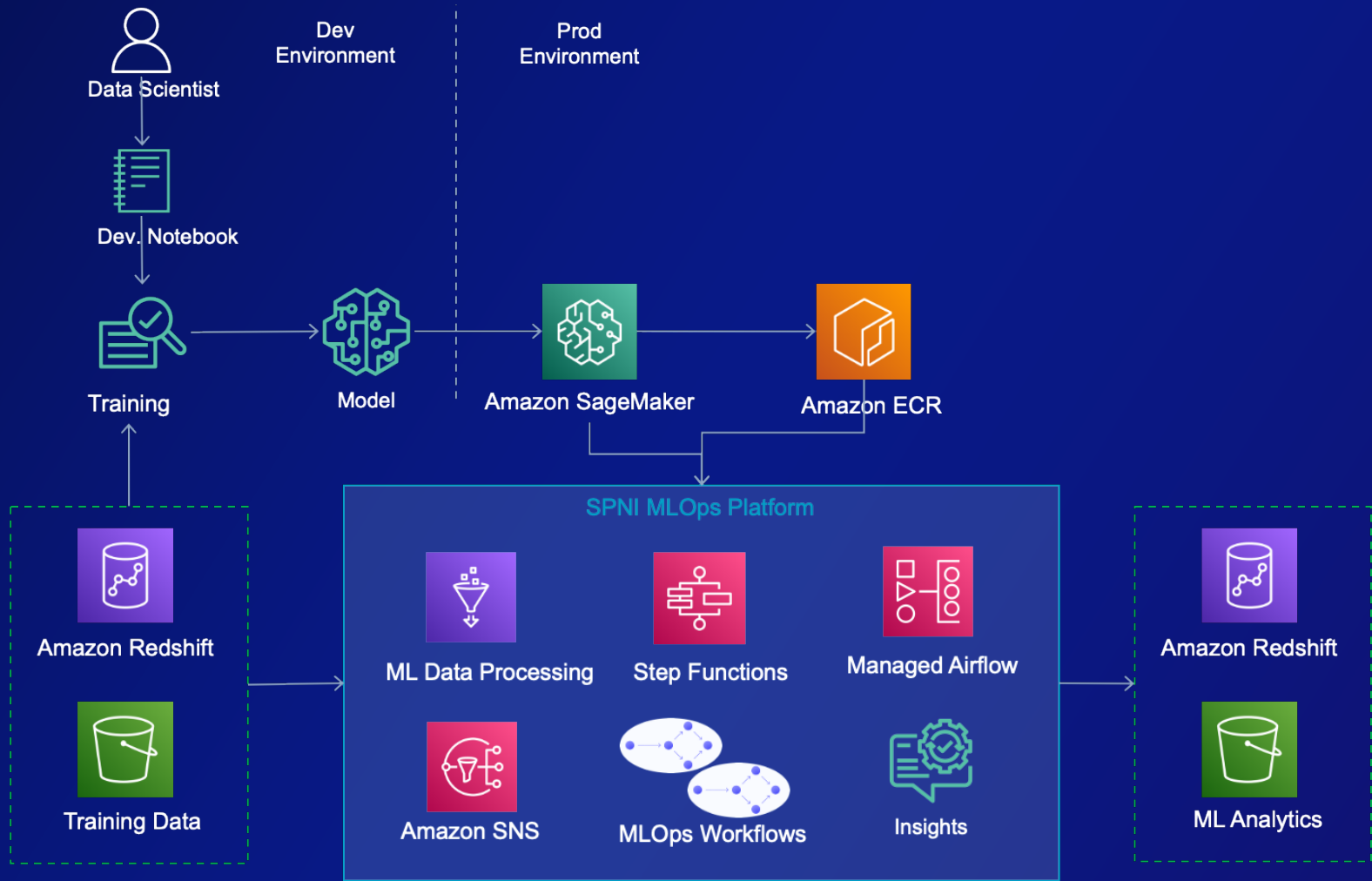
Enterprise data lake architecture

Multiple Source Systems	Data Ingestion Layer	Logically Grouped and Cleansed Datasets with Metadata	Processed and Ready to Use Data Sets	Consume

Multiple Source Systems
- Sales Distribution
- Content, Production & Ops
- Broadcasting
- Enterprise Application & ERP
- Digital Publishing

Logically Grouped and Cleansed Datasets with Metadata

Finance
E.g. Balance Sheet, Cost Objects

Sales & Marketing
E.g. Account, Opportunities, Deals, Advertisers

Content
E.g. Assets, Schedules

Commercial
E.g. Commercials, Rights

Distribution
E.g. Operators

Digital
E.g. Viewership, Subscribers

Other Data
E.g. HR, Legal, Market Intelligence

Master/Reference/Hierarchies
E.g. Customer, Titles

Unstructured, Semi-structured Datasets
E.g. Agreements, Logs, Events, Video, Audio

Processed and Ready to Use Data Sets
- Consumer Segmentation
- Inventory Analytics
- Consolidated Financial
- Personalized Recommendations
- Contracts management
- Pricing Optimization
- Market Performance
- Competitive Insights
- Competitive Insights

Consume
- Business Intelligence
- Reports
- Data Science
- API

23+ Data Source, 690+Data Sets, 1700+ Database Tables

SPNI – Data lake architecture & Use cases

SPNI future ready data lake architecture

SPNI ML Ops Architecture

Data Scientist

Dev Environment

Prod Environment

Dev. Notebook

Training → Model → Amazon SageMaker → Amazon ECR

SPNI MLOps Platform

ML Data Processing   Step Functions   Managed Airflow

Amazon SNS   MLOps Workflows   Insights

Amazon Redshift
Training Data

Amazon Redshift
ML Analytics

Outcome

Modularity

Faster Innovation, Future Ready

Scalable

Deeper Automation, Lower TCO

Multi Environment Deployment

High Availability, Data Reliability

FPC optimal scheduling

Objective

Automated FPC to define scheduling of different episode to improve market share

Solution

- Developed a machine learning algorithm(ML) to prepare FPC weekly plan with an aim to improve the market share of Sony channels

- ML model predicts the viewership share
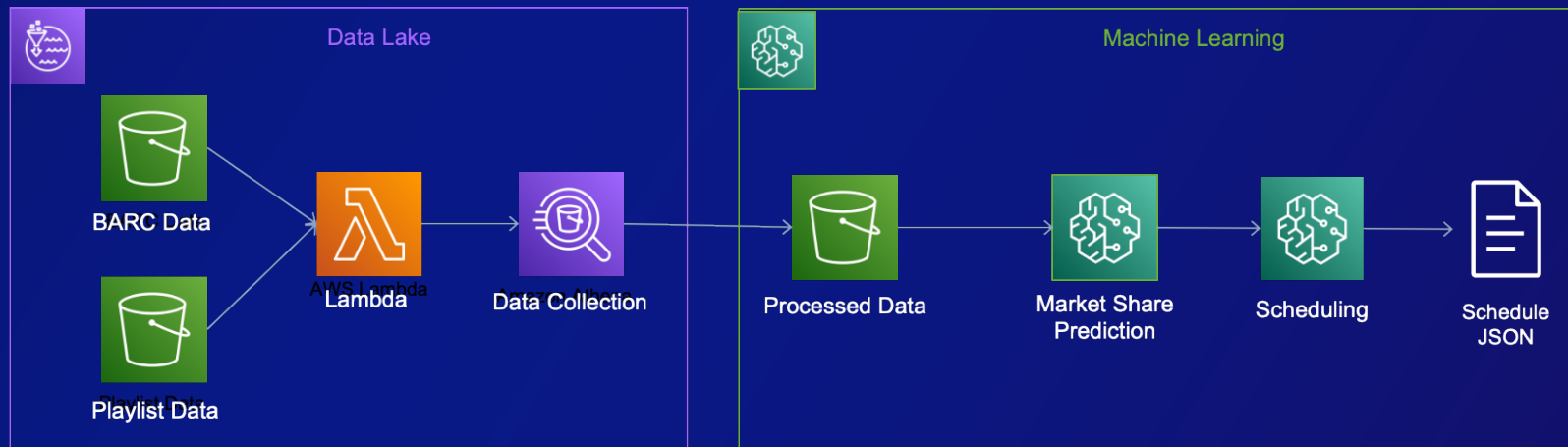
Content plagiarism

Objective

To intelligently compare and detect content plagiarism in a submitted story against the stories present in the database.

Solution

- Translation of input text into a common language for comparison

- Generates a sentence wise comparison analysis for easier detection of plagiarism

- Identifies the most similar stories and flag the suspicious cases for review
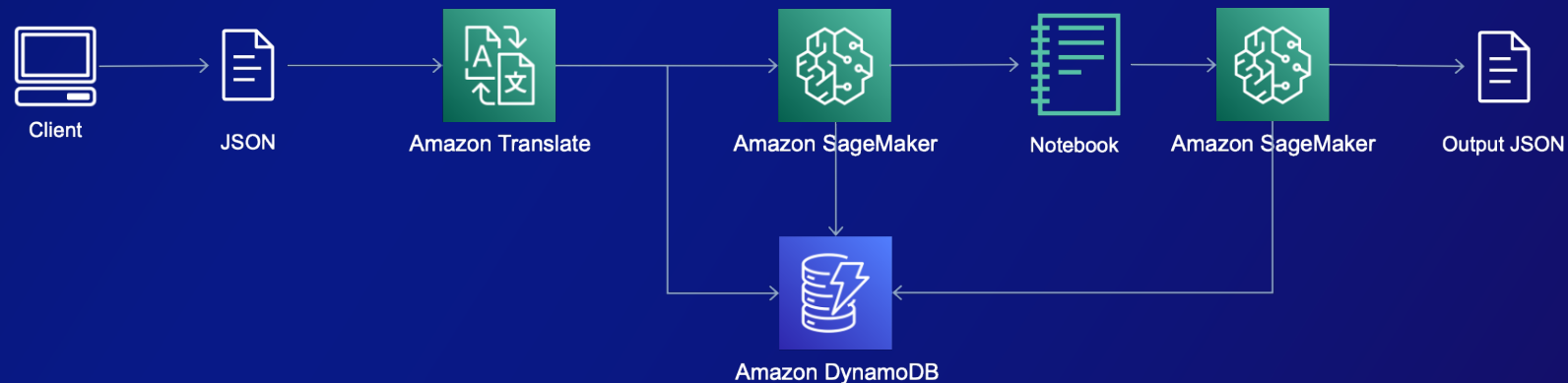
# Viewership analytics

Objective

The aim is to build an Image Processing system to extract textual and tabular information for processing monthly subscription reports using the Robotics Process Automation(RPA) tools for viewership analytics.

Solution

- Image pre-processing techniques to improve image quality

- Extracts tabular and textual information from documents (pdf, jpg)

MSR Documents → Amazon S3 → Amazon Textract → csv Files → RPA

Benefits of building on AWS
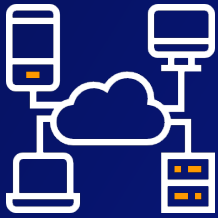
**Single source of truth –
One lake**

**Faster innovation,
Future ready**

**Deeper automation,
Lower TCO**

**High speed ETLs and
Dynamic scale**

**High availability,
Data reliability**

High skilled specialists

skillbuilder.aws

Your time is now
Build in-demand cloud skills your way

Thank you!

Please complete the session survey

Mahesh Patil

Head IT - Infrastructure, Analytics & Data Lake

Sony Pictures Network India

Rahul Sureka

Enterprise Solutions Architect

AWS India