
AWS
re:Invent
NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

PRT231

SPONSORED BY TERADATA

Charter Communications' journey to a modern cloud analytics architecture

Peter Singla

Lead Database Administrator,
Enterprise Data Platforms
Charter Communications

Ashish Yajnik

Senior Vice President, VantageCloud
Teradata

Doug Mbaya

Sr. Partner Solutions Architect
AWS

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

Data and analytic modernization requirements

Charter’s mission statement

The beginning: On-premises architecture

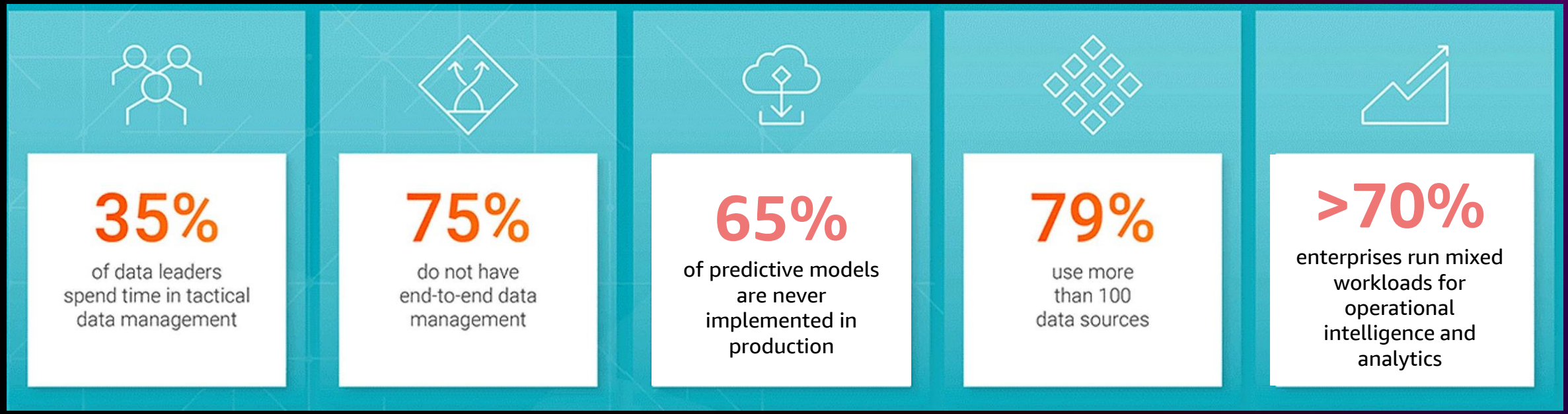
The journey to the cloud

The modern architecture

Data analytics trends on AWS

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Significant cost, agility, and data challenges prevent organizations from accelerating digital transformation



Unlock data

Activate analytics

Accelerate value

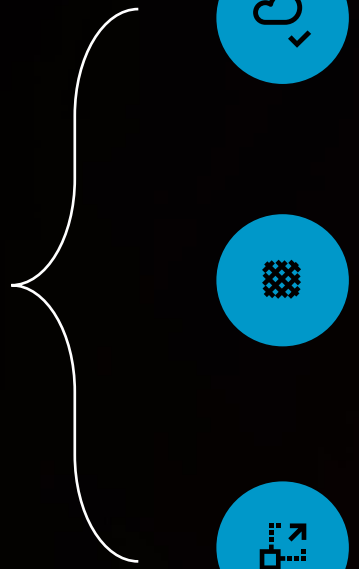
aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Customer needs for modernization

3

key areas of modernization



Customer requirements	Capabilities
Modernize cloud architecture	• Agility in the cloud • Self-service • Multidimensional scale
Modernize data management	• Comprehensive data management • Data collaboration • Secure access and data protection
Operationalize analytics at scale	• In-database analytics • Bring your own analytics • Partner integration

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Teradata VantageCloud evolving to provide more flexible support for modern data architectures



Centralized object storage

- Economical, durable, performant and ACID-compliant data storage, with time travel access
- Open data formats & support for multi-structured data
- Simplified data sharing



Multi-cluster compute with elasticity

- Simple, flexible, and automated elastic scaling
- API-driven model for infrastructure-as-code deployment
- Built-in fault tolerance
- Zero down time elasticity & upgrades



Business autonomy

- Self-service provisioning
- Separation of compute & storage
- Self-service data exploration with compute isolation
- Non-disruptive upgrades



Optimized pricing & governance

- Independent compute for simple chargeback…
- …with centralized financial governance to enable effective management of TCO, with reduced overhead



Open & connected

- Real-time ecosystem integration & connectivity
- Bring your own analytics/ML
- Cloud-native services integration – API-first design

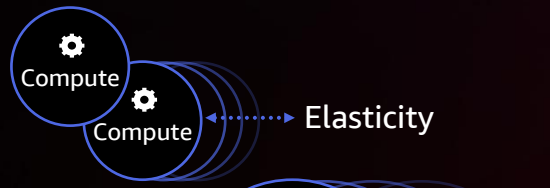
aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

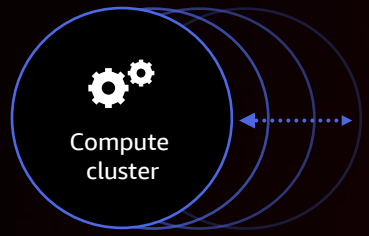
AWS services with Teradata VantageCloud lake architecture overview

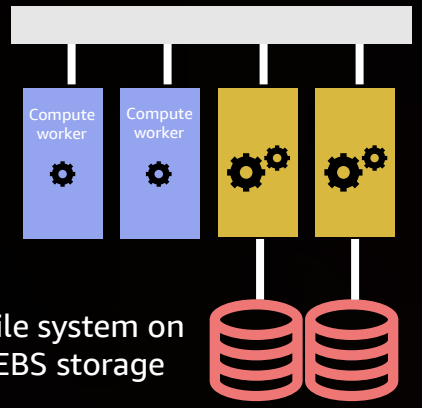


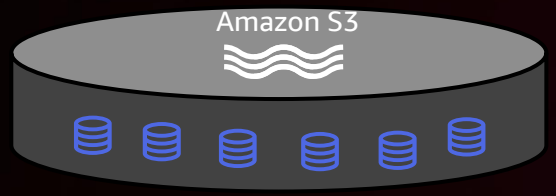
Self-service console











Block file system

Object file system

Self-service console hosted on AWS

- Web-based self-service console for system and user administration
- Web-based query editor
- Query and billing monitoring

Primary cluster with Amazon EC2 and Amazon EBS storage

- Manages query planning and the distribution of work across the environment
- Manages the high-performance enterprise block storage for select operational queries
- Expand compute-only capacity with compute worker nodes

VantageCloud compute clusters with Amazon EC2

- Stateless query processing to support isolated workloads
- Multiple clusters with access to shared object storage
- Policy driven auto-scale elasticity by cluster

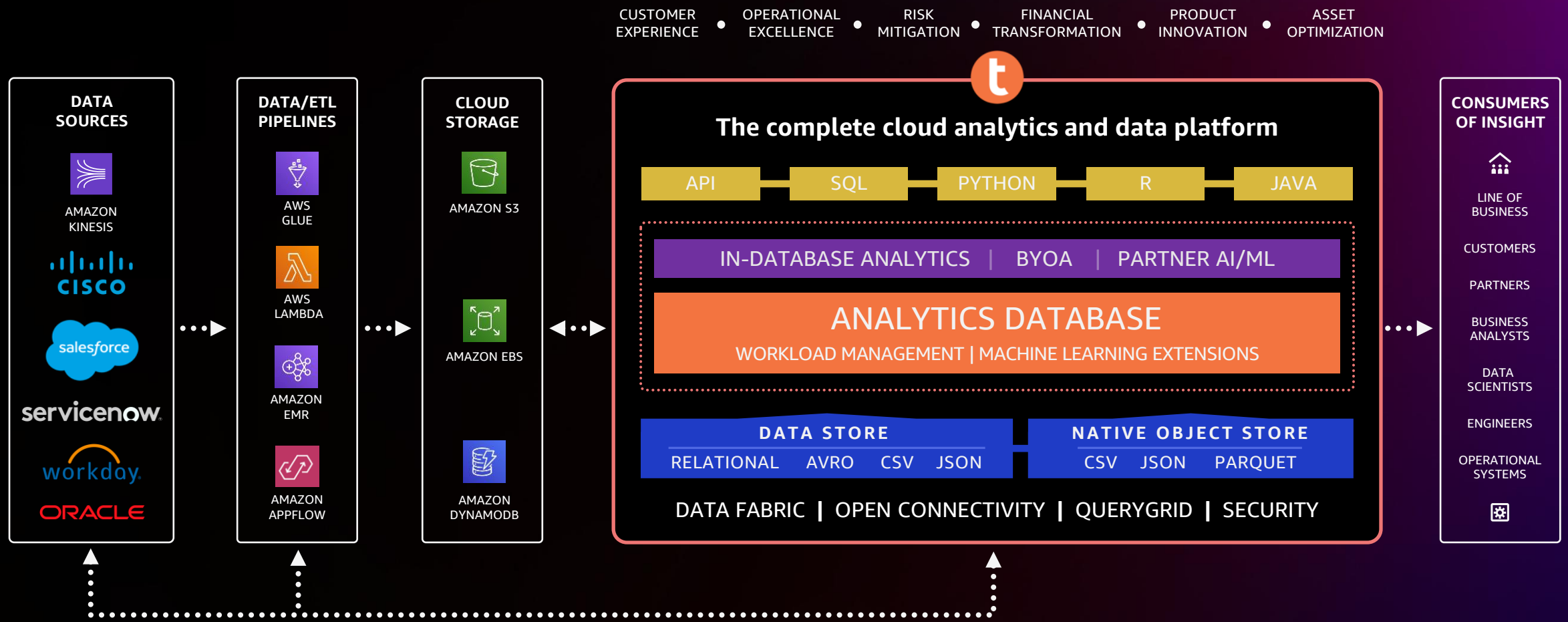
Object file system on Amazon S3

- Massively scalable low-cost storage
- Fully ACID compliant and optimized for performance including caching, optimized spool, indexed metadata

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Teradata VantageCloud on AWS

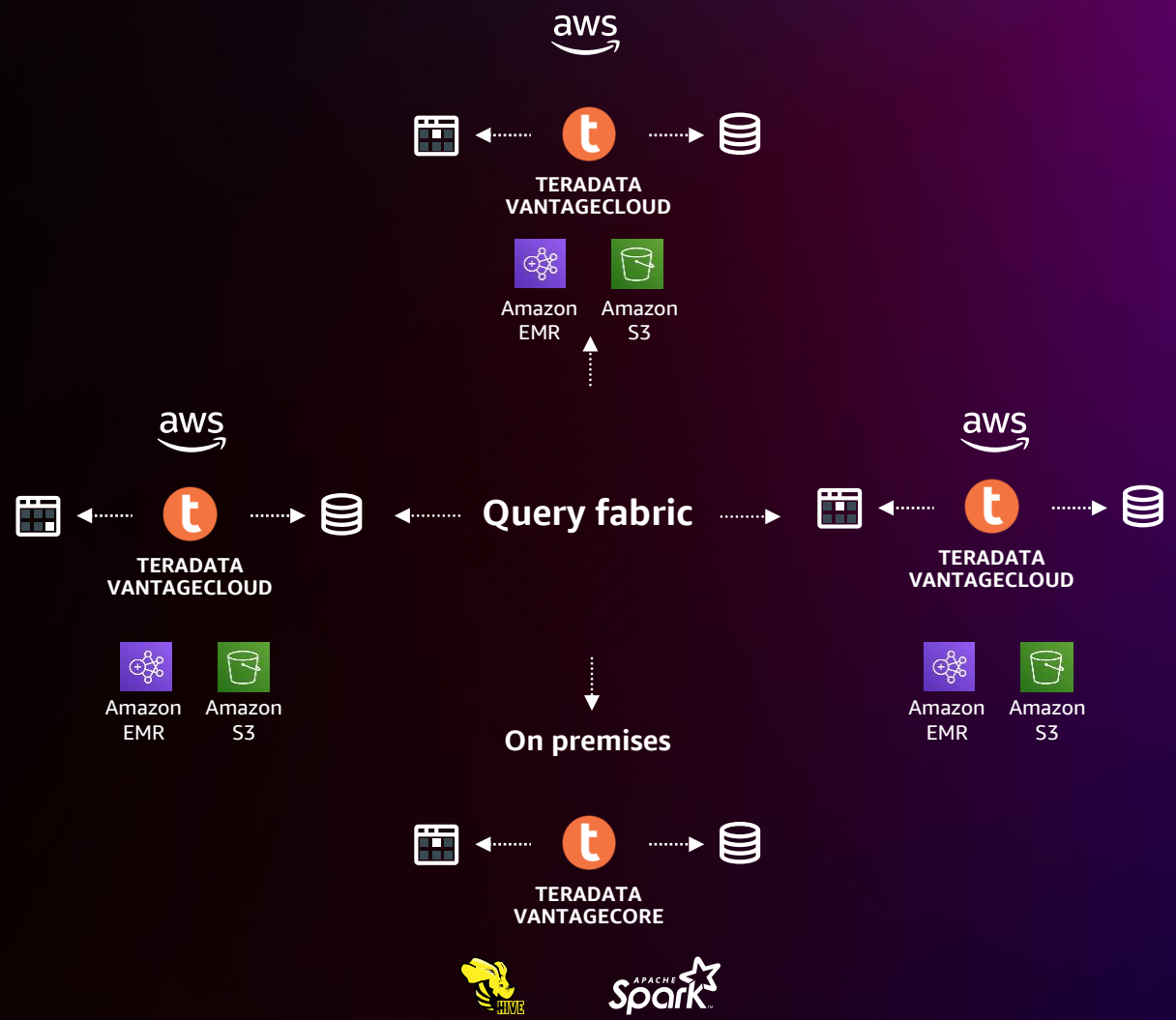


© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Hybrid query fabric

INTELLIGENT, COMPREHENSIVE QUERY FABRIC

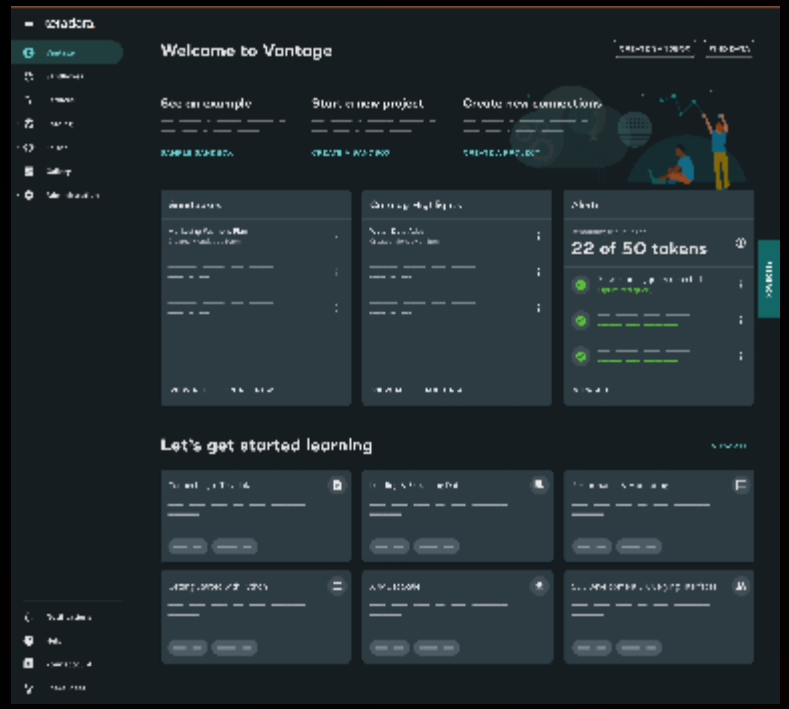
- Hybrid-ready enables pushdown processing, minimizing data transfer and egress cost
- Comprehensive set of data source connectivity
- Search, explore, and understand the context behind the data
- Sharing data via marketplaces, internal and external to the organization, with governance



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

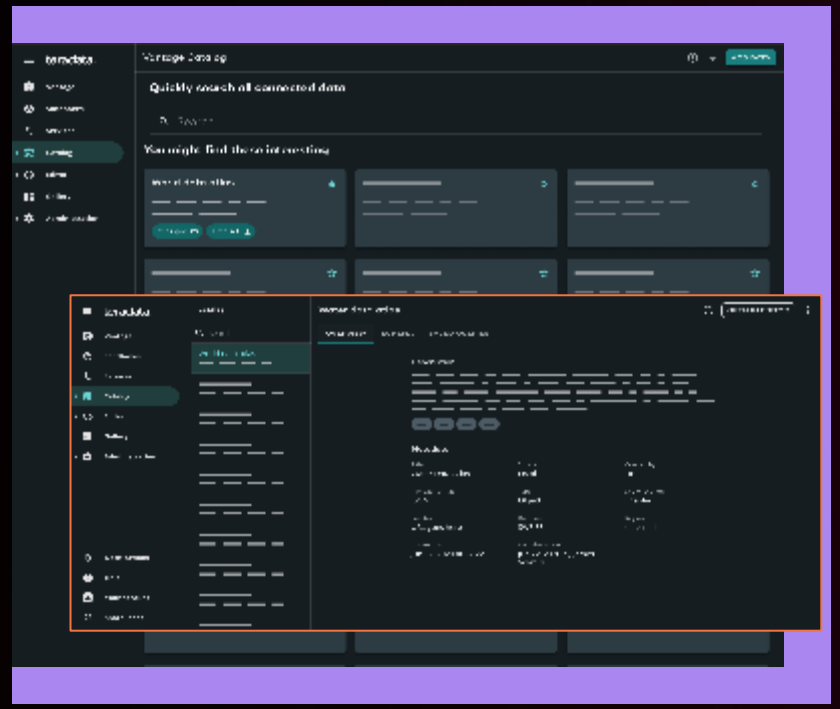
Self-service search, explore, share with AWS

1 Search & filter



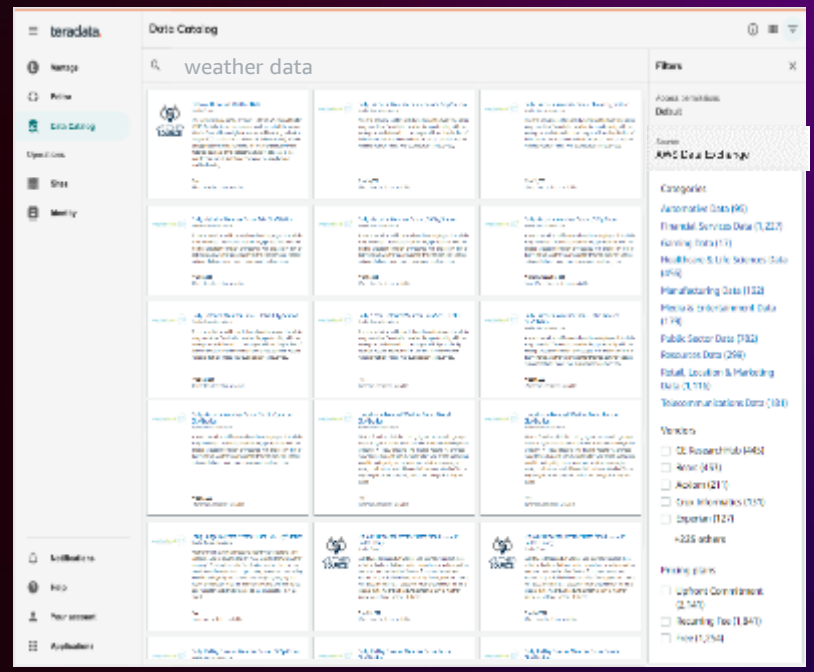
Keyword **search with AWS Glue** in unknown datasets; explore format, provenance metadata

2 Browse & select



Enable users to merge and join relational and semi-structure

3 Sharing via marketplaces



AWS Data Exchange sharing marketplace integration to add more data

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Operationalize analytics at scale inside database

IN-DATABASE ANALYTICS

Geospatial, time-series, temporal, nPath

Powerful extensions to traditional relational operations

- **Geospatial functions**
- **Time-series functions**
- **Temporal functions**
- **nPath functions**

- Implementation of standard ST_Geometry data type with rich set of spatial functions
- High-performance time-aware functions
- nPath analysis, sessionize, attribution working together are a core differentiator of Teradata

VantageCloud analytic library

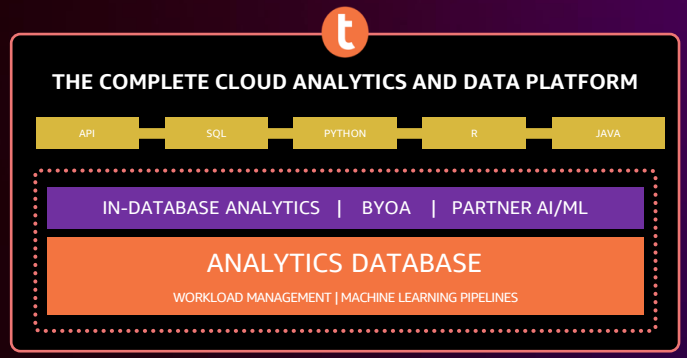
Suite of advanced analytic functions

- **Data transformation**
- **Statistical hypothesis tests**
- **Feature engineering, PCA**
- **GLM, clustering**

- Rich library of advanced functions from descriptive statistics, hypothesis testing, clustering, regression, and predictive algorithms
- Use preferred languages R, Python, or SQL
- Available on any platform (Nov. 2020)

Advanced SQL engine

Advanced SQL engine with native ML extensions arranged in analytic pipelines linked to industry models



- Highest performance possible
- New functions covering entire data analytics lifecycle
- Arranged in pipelines to support specific problems (e.g., classification, regression, time-series, segmentation, text processing)
- Linked to industry data models

aws

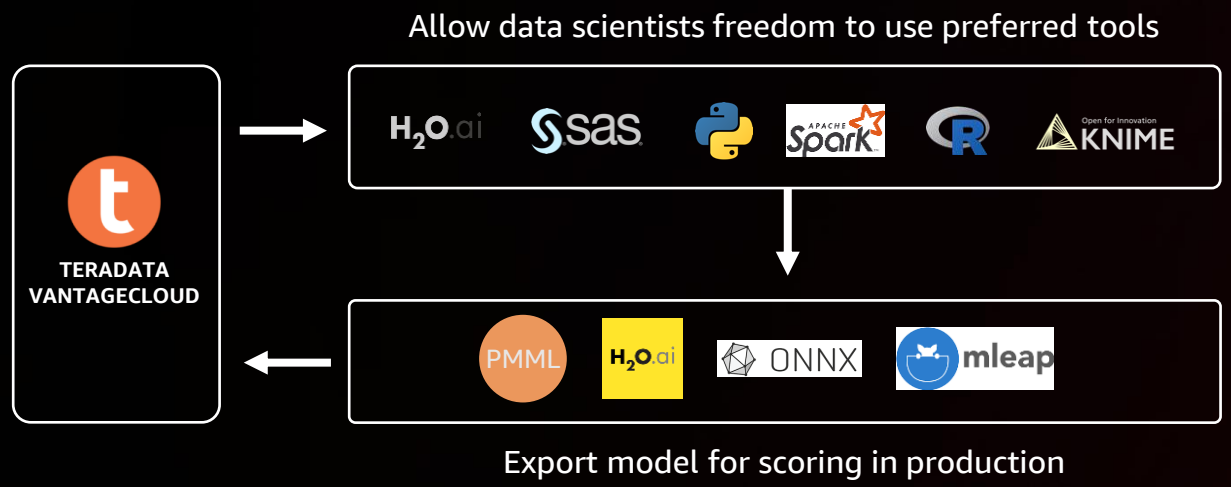
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Operationalize analytics at scale

BRING YOUR OWN ANALYTICS

Model sharing

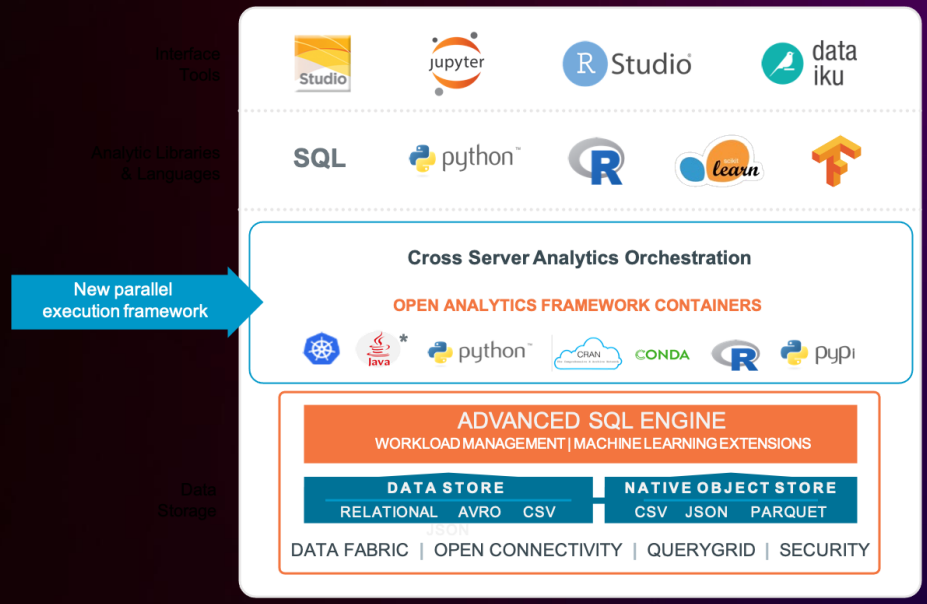
Import existing ML models into VantageCloud



- Data scientists use many different development platforms to create ML models – with model sharing, these models can be imported into VantageCloud and used to score in production workloads
- Allow data scientists freedom to use preferred tools

Open analytics framework

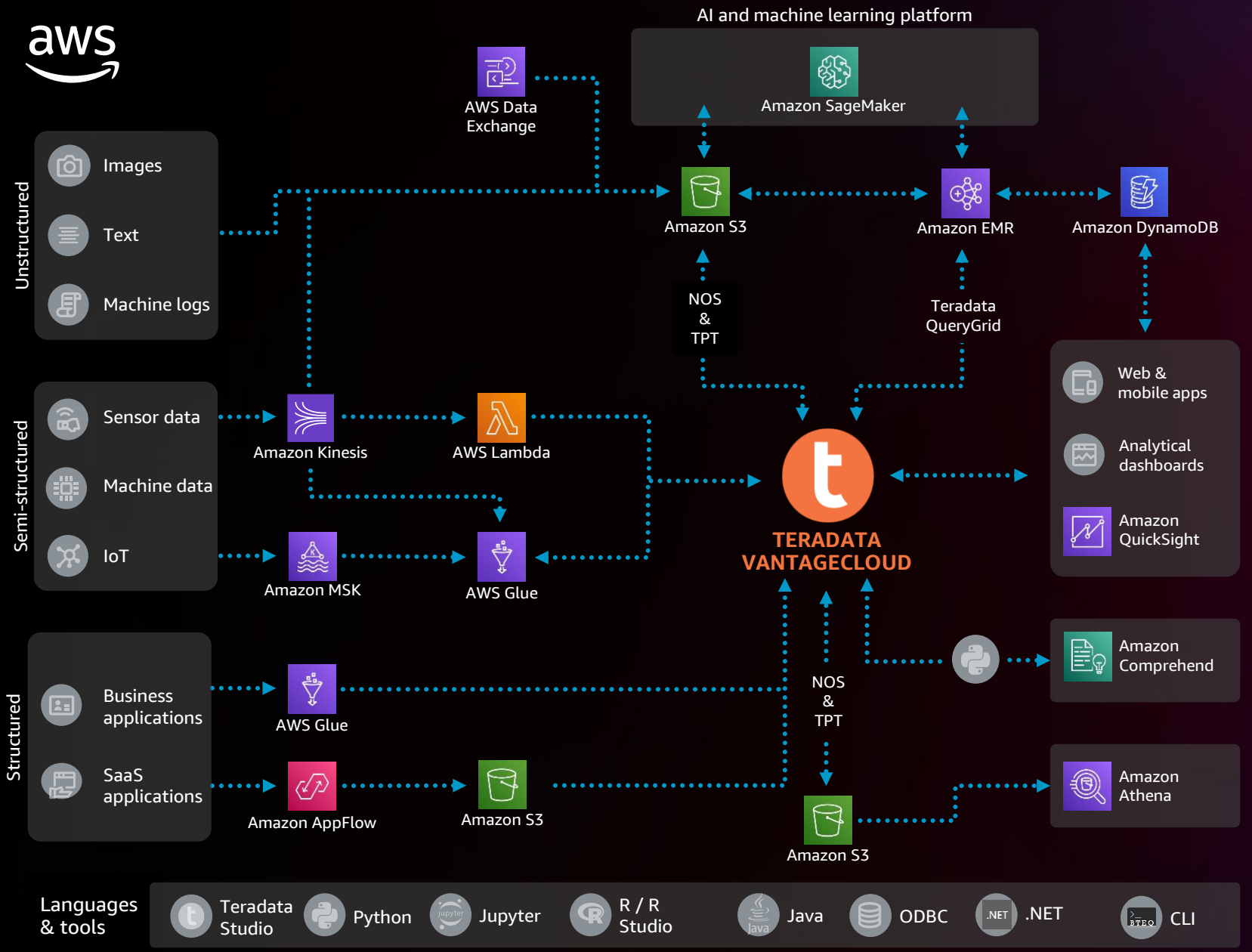
Advanced SQL engine with native ML extensions arranged in analytic pipelines linked to industry models



- Execute R/Python code with data parallel computation
- Serverless architecture enables auto scaling, with costs only for consumption

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

INTEGRATIONS



ADDITIONAL INTEGRATIONS

- AWS PrivateLink PRIVATE CONNECTIONS
- AWS Marketplace OFFERING
- AWS CloudFormation ORCHESTRATION
- AWS KMS KEY MANAGEMENT
- Amazon EKS KUBERNETES CONTAINERS
- Amazon EBS PERSISENT STORAGE
- AWS CloudTrail API MONITORING
- Amazon CloudWatch RESOURCE MONITORING
- Amazon EBS snapshots RAPID BACKUPS

Note: Not all service integrations are shown

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Capturing all workloads, driving growth in the data & analytics market requires agility and enterprise grade price performance

Central high-performance data and analytics that run the business

ENTERPRISE ANALYTICS



DEPARTMENTAL & AD HOC NEEDS

Project-based work
Business hours vs. non-business hours
Monthly/quarterly reporting



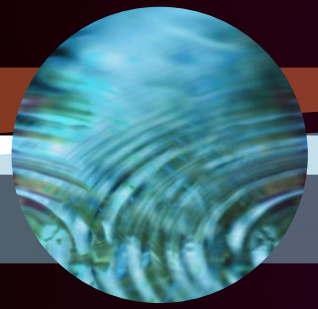
Demanding risk and regulatory requirements

COMPUTE ISOLATION



DATA LAKE EXPLORATION

Uncovering value in low-cost infinite storage



Drive down the cost and risk of experimenting with new ideas

ANALYTIC INNOVATION

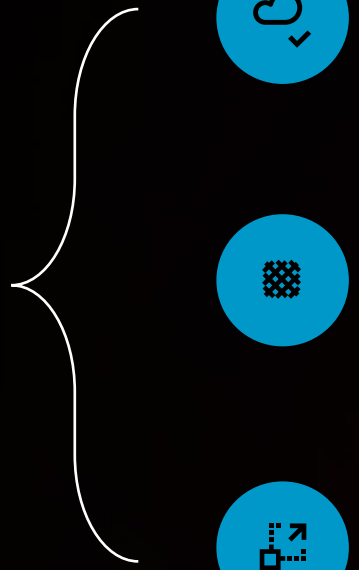


© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Customer needs for modernization

3

key areas of modernization



Customer requirements	Capabilities
Modernize cloud architecture	• Agility in the cloud • Self-service • Multidimensional scale
Modernize data management	• Comprehensive data management • Data collaboration • Secure access and data protection
Operationalize analytics at scale	• In-database analytics • Bring your own analytics • Partner integration

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Charter Communications

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Key takeaways

1

Examining the starting point architecture and on-premises ecosystem, and the direction defined by the mission to modernize

2

Evaluating the present state and analyzing the future state with predictable and manageable cost

3

Modernizing analytics in the cloud with access to modern analytic features and cloud services

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Our mission statement

Our goal is to provide an analytical platform for pervasive data intelligence, to meet business requirements through innovative and cost-effective solutions

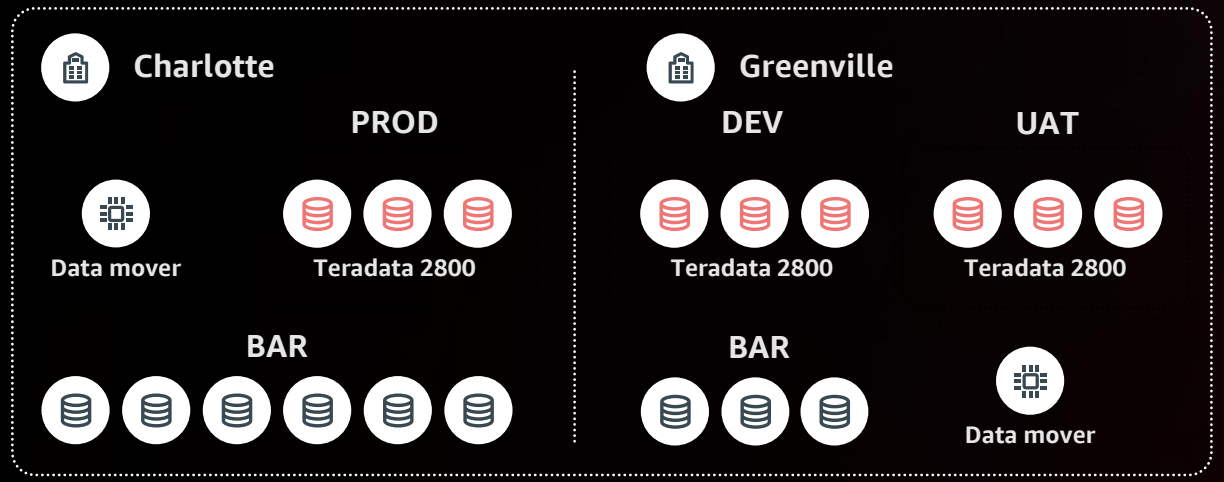
Modernizing analytics in the cloud with predictable and manageable cost

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The beginning:
On-premises architecture

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

On-premises architecture and ecosystem





Legacy architecture

On-premises capacity

- Complex capacity expansion
- Pay for capacity
- End of life

Isolated ecosystem

- Single system monolithic design
- Limited elastic performance
- Limited access to object stores

Business continuity

- RTO in weeks to months
- Constrained retention
- Expensive backup solutions

Traditional legacy workload

Limited advanced analytics

Zero self-service

Legacy BI tools

Business intelligence

Legacy ETL

Logically isolated workload

Data integration tools

TPT utilities framework

- Tbuild
- Fastload

Data replication tools

IBM Infosphere Data Replication

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The journey to the cloud

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

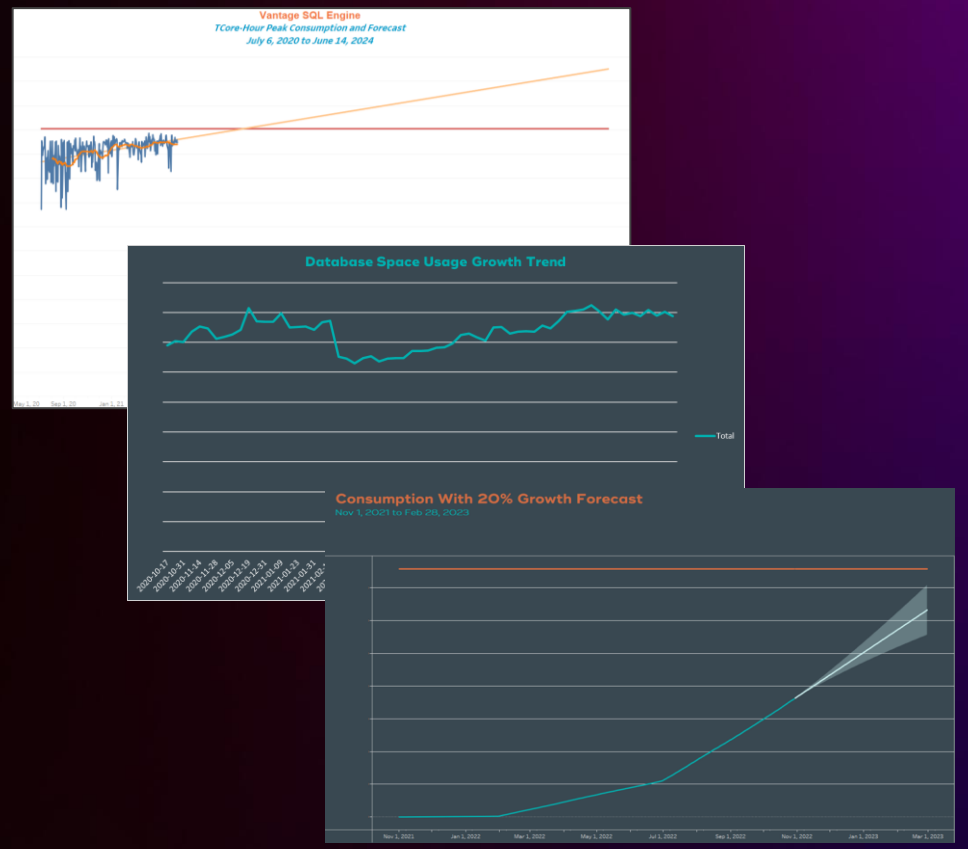
Preparing for the Journey

Usage analysis

Moving to a cloud consumption model means provisioning enough for the first year's growth and then gradually scaling as consumption trends shift

Examining the existing compute demand of workloads prepares us for the journey to the cloud

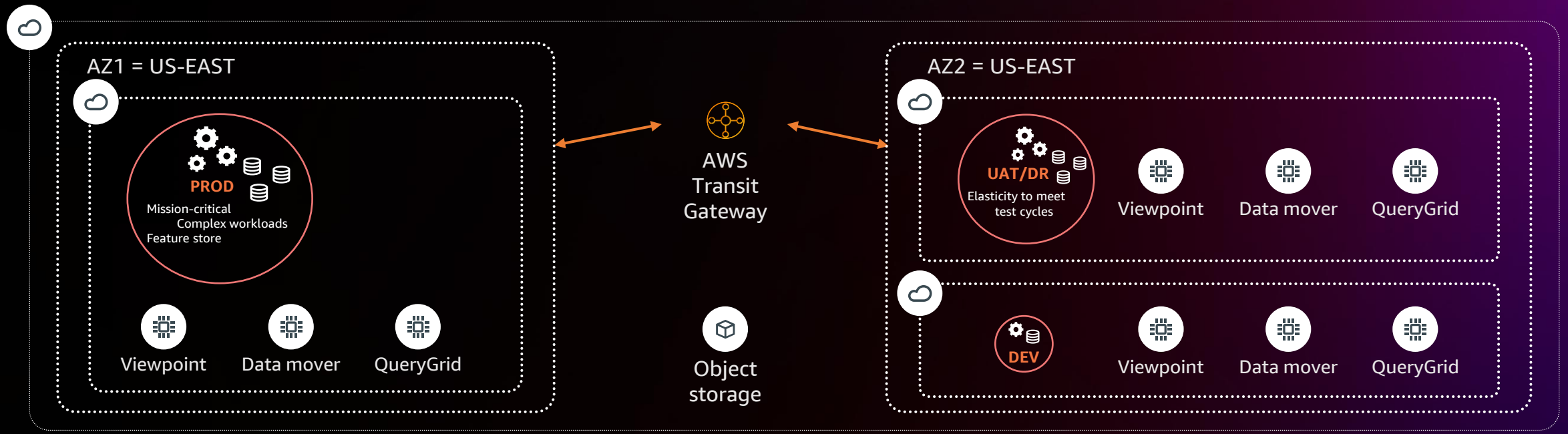
- Measuring compute demand determines the initial provisioned resources
- Measuring logical I/O estimates the Vantage Unit consumption in the cloud
- Measuring storage trends allows for minimal provisioning of storage and allows for scaling in the future
- Measuring ingress and egress to consider cloud service provider (CSP) costs when moving data out of the cloud region



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Preparing for the Journey

Defining the modern architecture



Modern architecture

Scalable consumption

- As a service
- Elasticity
- Consumption pricing and monitoring

Connected ecosystem

- Virtual fabric with QueryGrid
- Logically and physically Isolate workloads
- NOS to object store

Business continuity

- RTO in days
- Extendable retention
- Easy backups

Modern cloud workload

Democratized self-service

ML/AI

Cloud BI tools

Business intelligence and advanced analytics

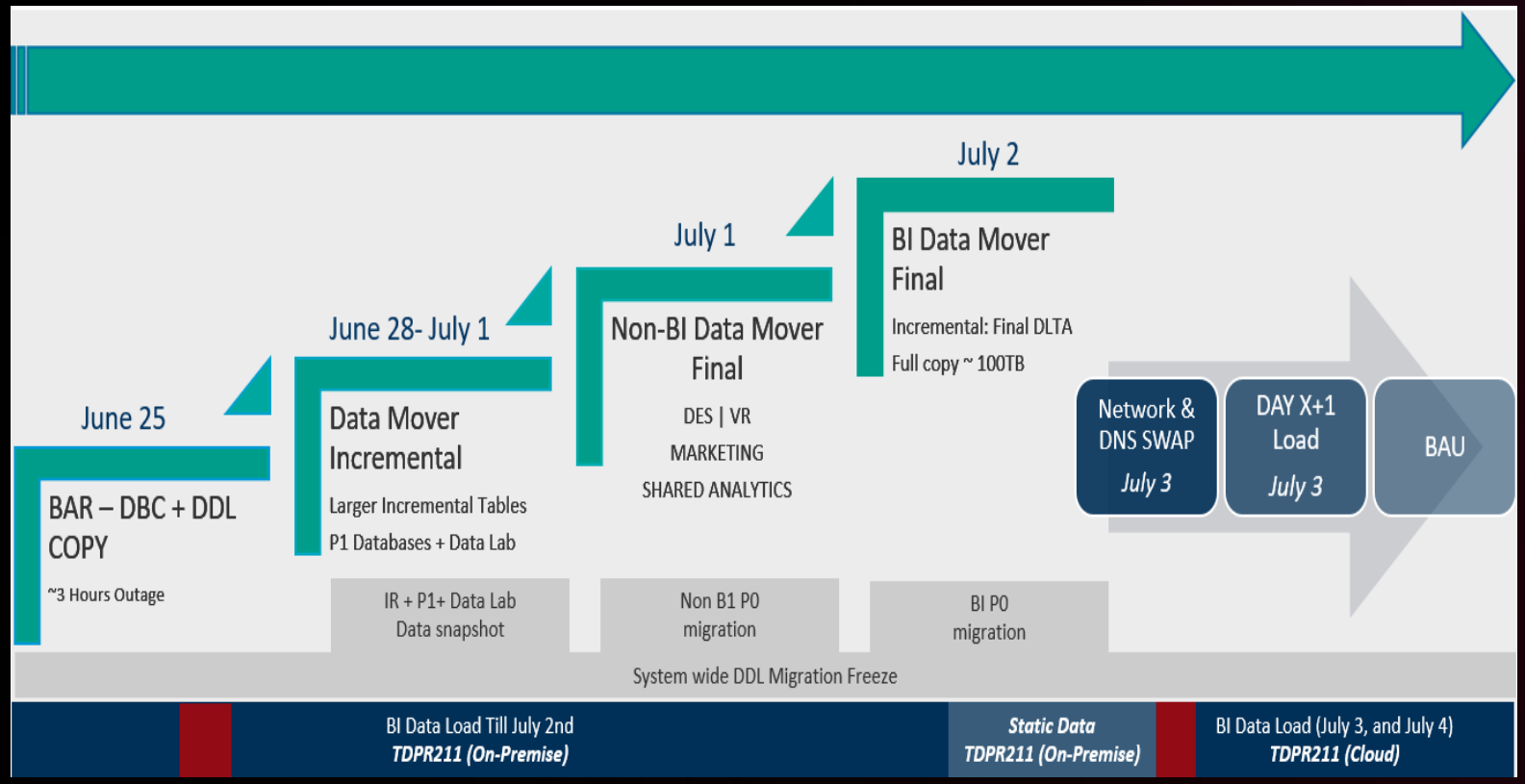
Real-time streaming

Autonomy for the business

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The journey at a glance



Migrated a total of ~310 TB across three systems using 10 Gbps AWS Direct Connect

Hybrid approach combining backup/ restore with data mover reduced total outage to less than 4 hours

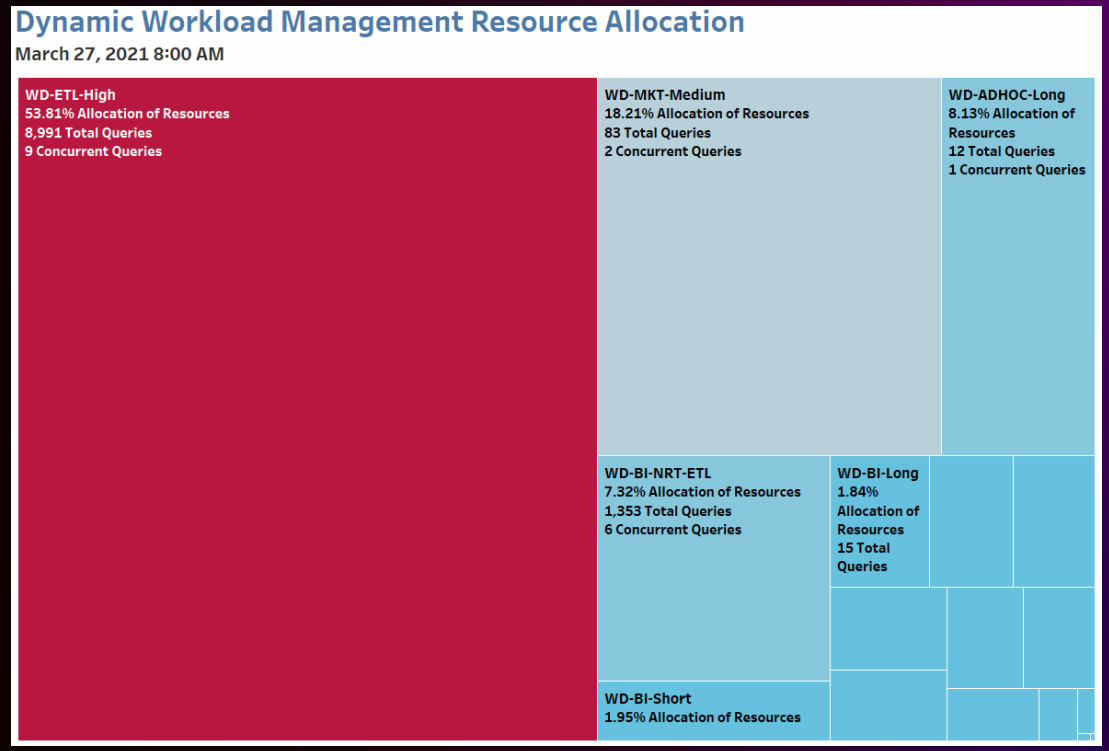
Cloud architecture and migration planning services (CAMPS) finalized network connectivity and data ingress/egress usage

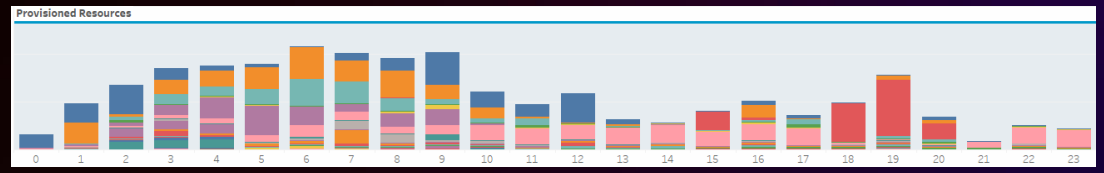
Vantage Units consumption monitoring on AWS using business utilization console (BUC) quickly revealed cost deviations

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The Journey – Life in the cloud

- Provisioned enough resources for growth in the first year
- Credits charged are based on compute resources consumed, not what is provisioned or run time
- Workload management empowers us to allocate more resources to workloads needing improved response time and concurrency
- Workload queries with more allocated resources consume the same amount of resources faster to improve response time, but the cost does not change
- Remaining resource allocation will elastically expand and contract based on resource demand
- Query optimization features help to improve performance by reducing resource consumption and therefore cost
- Scaling the system up or out does not occur until current workloads or new workloads outgrow the current provisioning





© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Lessons learned

- Teradata VantageCloud unit consumption needs to be considered due to default fallback enabled in the cloud when migrating from non-fallback systems

- Database permanent space was doubled due to default fallback enabled in cloud vs. on premises (no fallback)

- Teradata Parallel Transport (TPT) utility performance from on-premises to cloud system – Updated TPT buffer size to 16 MB

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The modern architecture

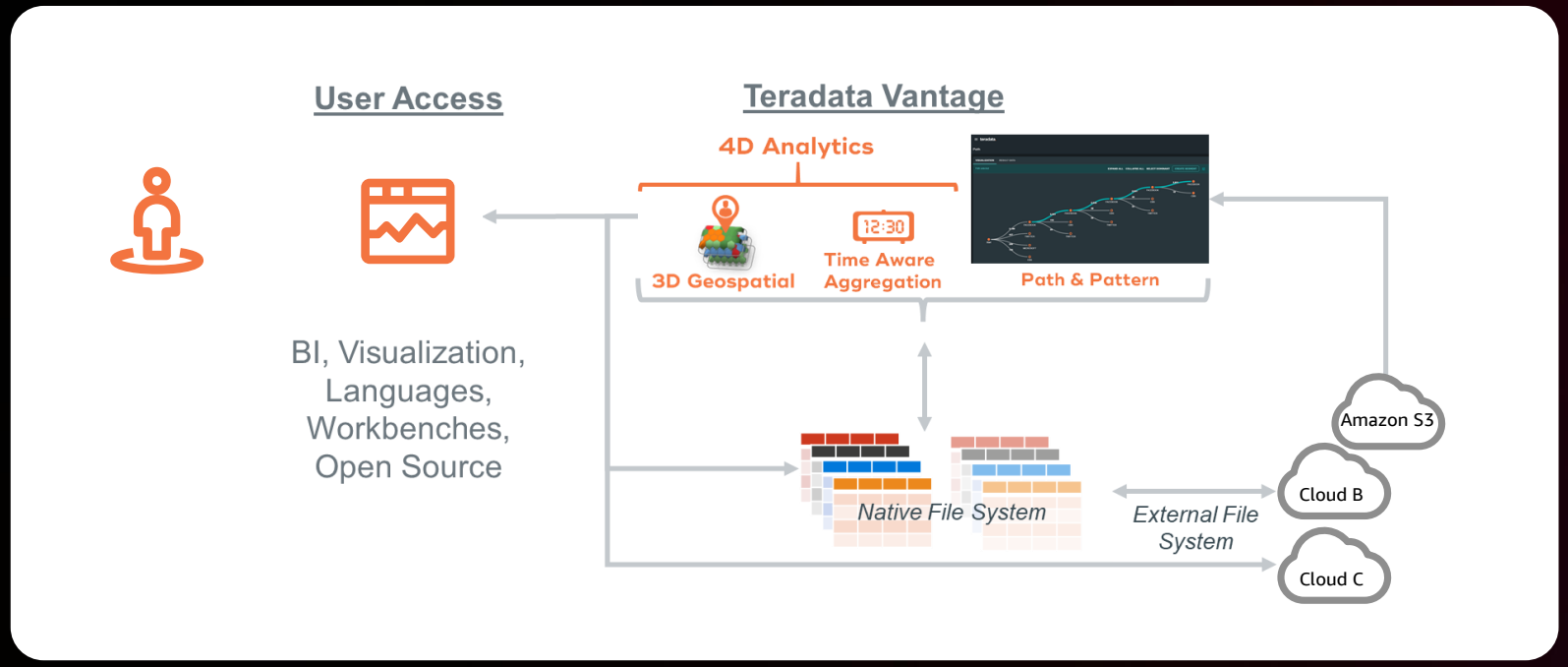
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The modern architecture

Leveraging native object store

Native object store (NOS) makes it easy for users to access datasets located on external object stores like Amazon S3 using standard SQL

NOS also allows users to share data from the VantageCloud relation tables to an external object store location as Parquet objects using Write_NOS



Key points

- Highly scalable
- Low cost per terabyte
- Structured, semi-structured, and unstructured data
- Support for open formats
  - Parquet
  - JSON
  - CSV

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The modern architecture

TIME-SERIES FORECASTING PIPELINE

Time-series ML library

Current

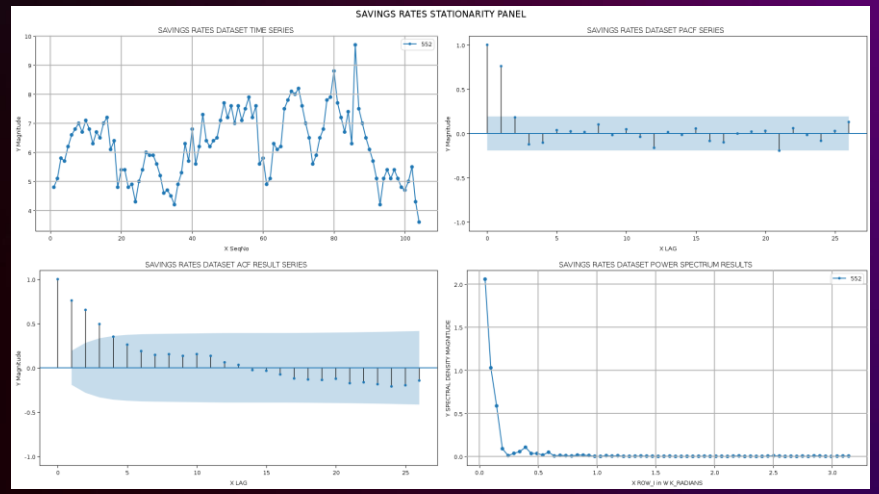
- nPath
- Statistical hypothesis tests
- Linear regression
- Feature engineering and transform

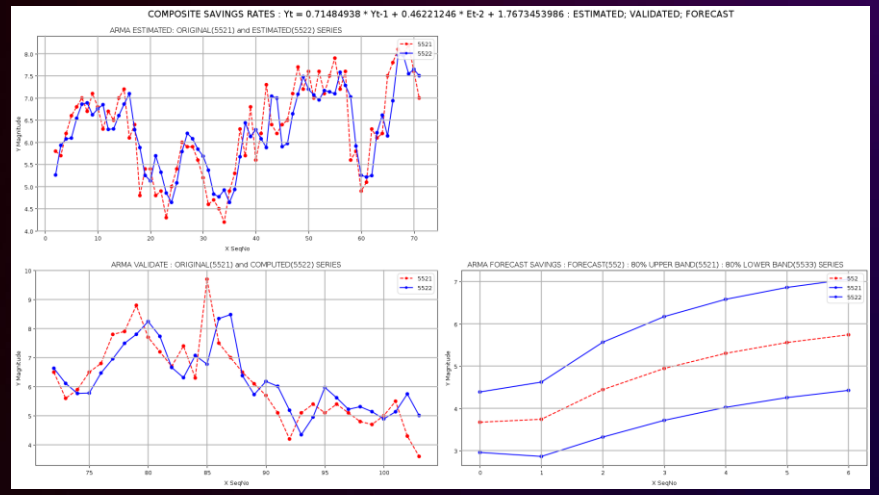
Future

- Time-series resampling
- Dynamics time warping
- ARIMA modeling and forecasting

Benefits

- Path and pattern recognition of segmented subscriber journey and experience
- Real-time anomaly detection with triggers for real-time alerting and launching actions
- Forecasting impacts to the subscriber experience in repair calls, self-install, outage events, etc.
- Predicting subscriber cancellations and downgrades
- Identifying golden path to subscriber upgrades and new services





© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The modern architecture

LOCATION INTELLIGENCE

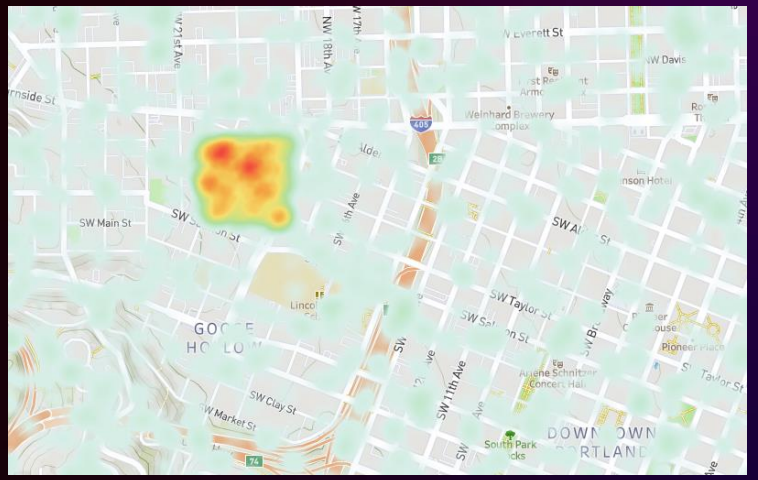
3D Geospatial + Time series = 4D

- Geometry data type
- Temporal and spatial functions
- Support for over 3K spatial reference systems (SRS)
- Visualization and location decoding with Google Maps API
- Geospatial indexing, statistics, and metadata
- Geospatial tessellation
- Import and export ESRI compatible GeoJSON, Well-Known-Text (WKT), Well-Known-Binary (WKB), and Shape files

Benefits

- Weather forecasting and outage prediction
- Overflowing river, lake, and other water intrusion risks
- Service health anomaly detection
- Peak period geospatial impact
- Service/repair incident spatial path and pattern recognition
- Identify fixed wireless access residential candidates



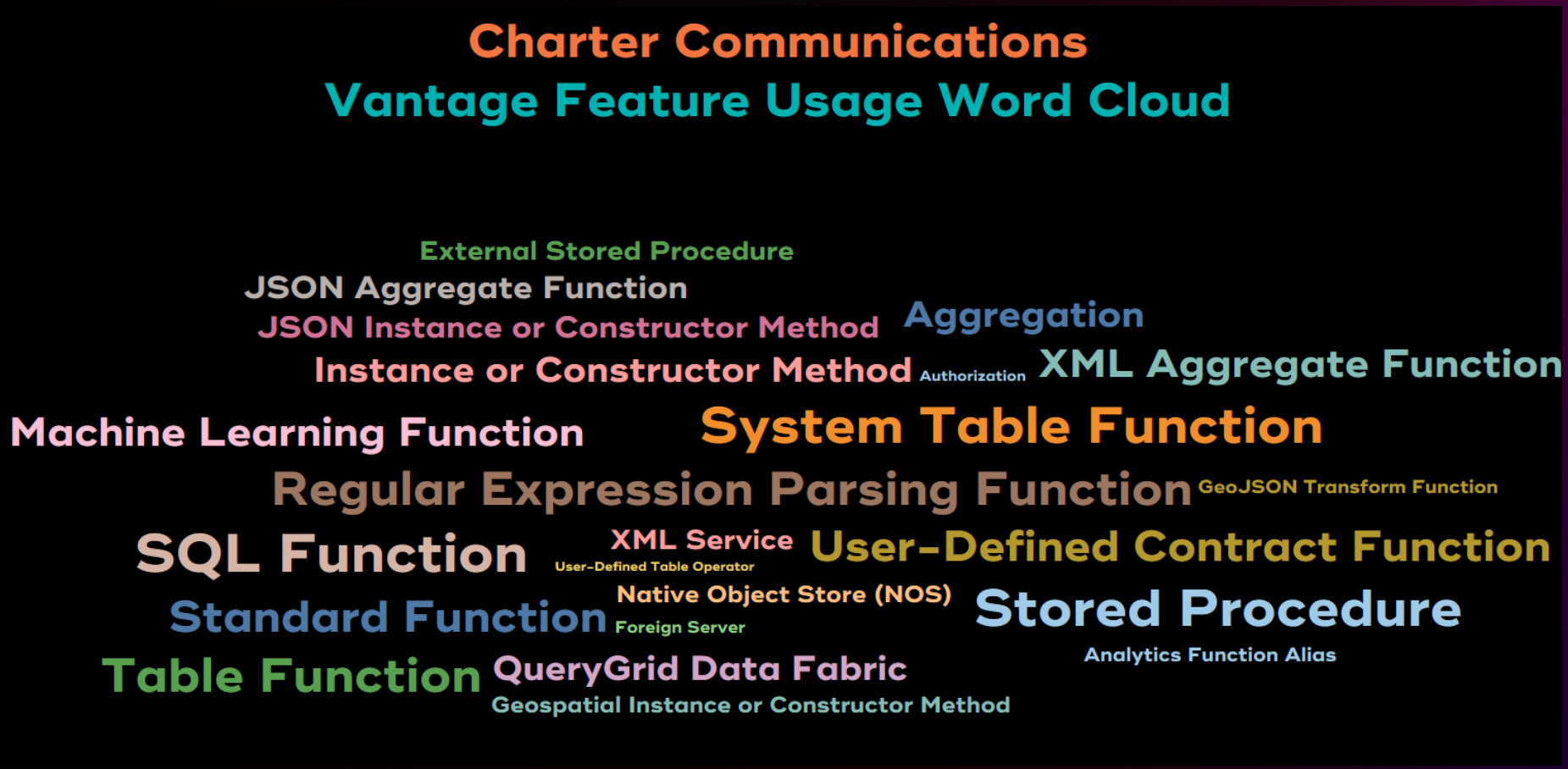


aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The modern architecture

ANALYTICAL FUNCTIONS



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Key takeaways

1

We started with an on-premises ecosystem with the direction to journey to the cloud and modernize the architecture

2

We visualized the future architecture and ecosystem with predictable and manageable cost

3

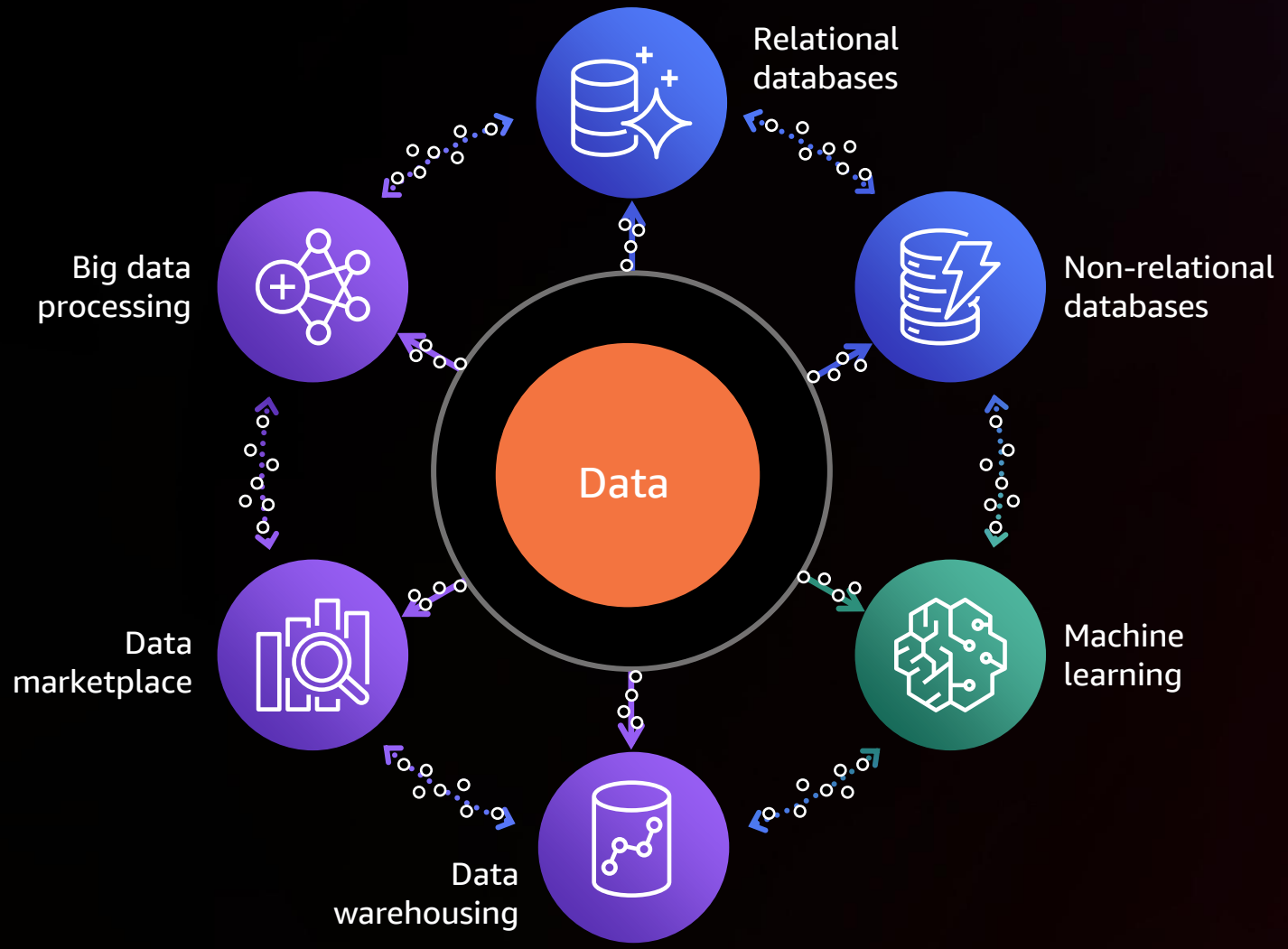
The modern analytic features and cloud services are now at our disposable to progress into the future

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data analytics trends on AWS: Data gravity time to value

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data gravity on AWS



- Scalable data lake
- Purpose-built data services
- Limited data movement
- Unified governance
- Performant and cost effective

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Purpose-built data services

Variety of sources and data types

Diverse analytics needs

Performance

Inflexible tools



Data scientists



Data engineers



Data analysts



Data owners

Choose the right service

- Never compromise on performance, scale, or cost

Best of breed

- Specialized, best-of-breed analytics services are designed to be the best at specific tasks

Democratization

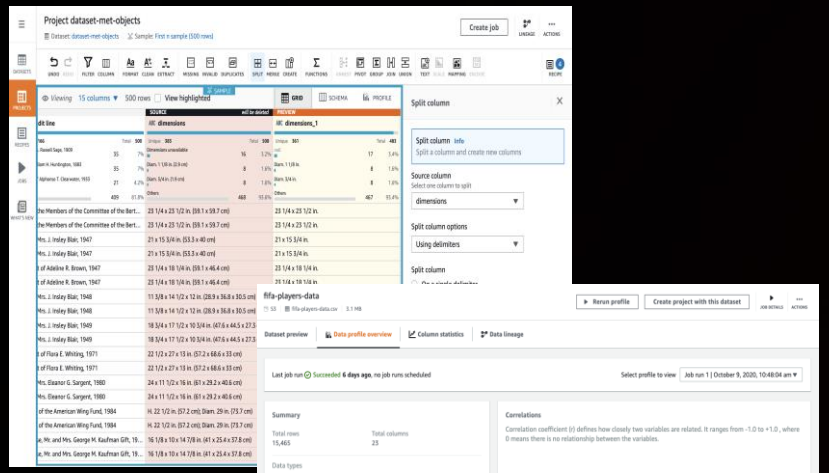
- Allow different personas to use the right tool for the job
- Remove gate keepers

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Glue DataBrew

VISUAL DATA PREPARATION FOR ANALYTICS AND MACHINE LEARNING



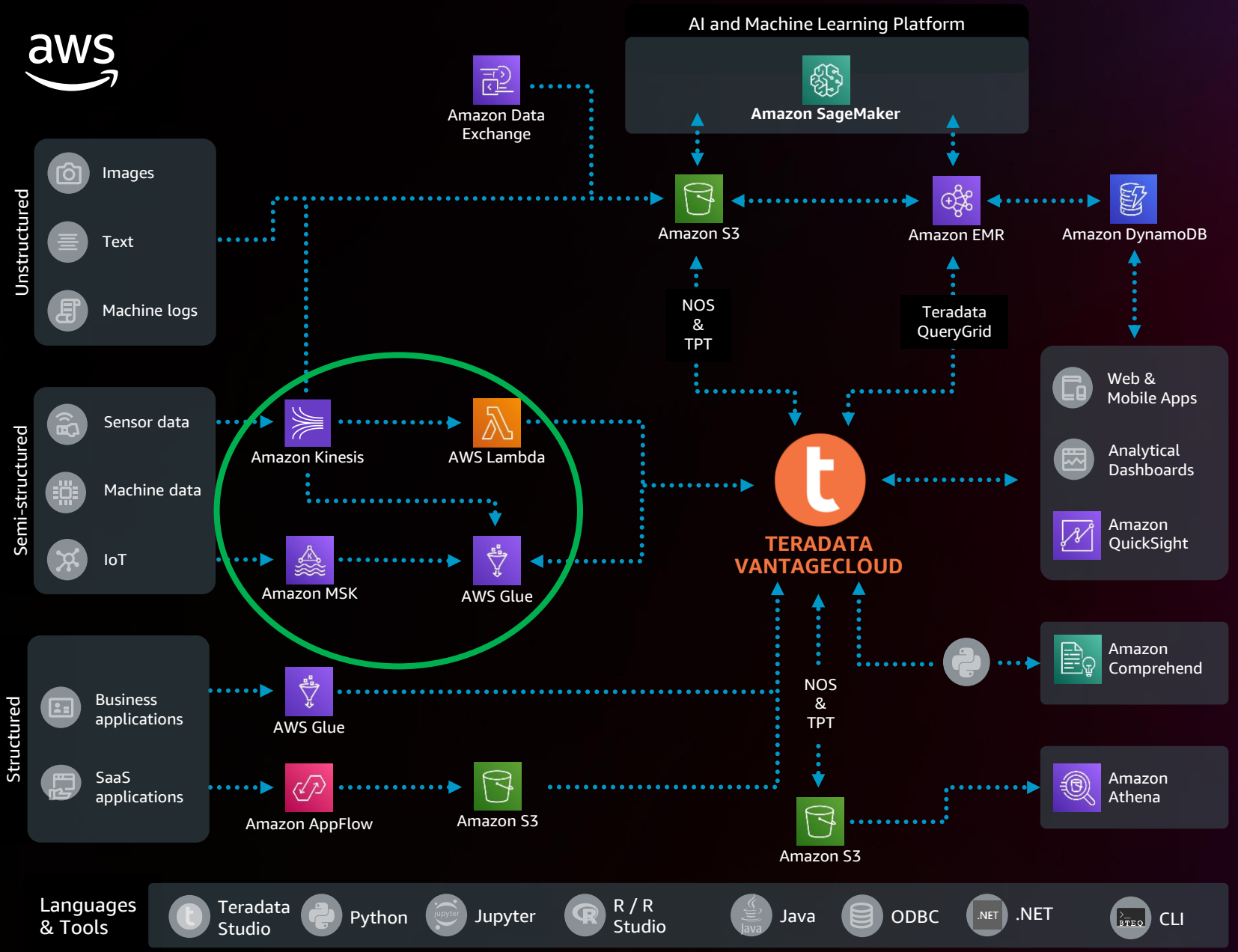
- Clean and normalize data with a visual interface (no code)
- 250+ built-in transformations without writing code
- Profile data to understand data patterns and anomalies
- Work on large datasets at scale

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Time to value: From data to intelligence

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Integrations



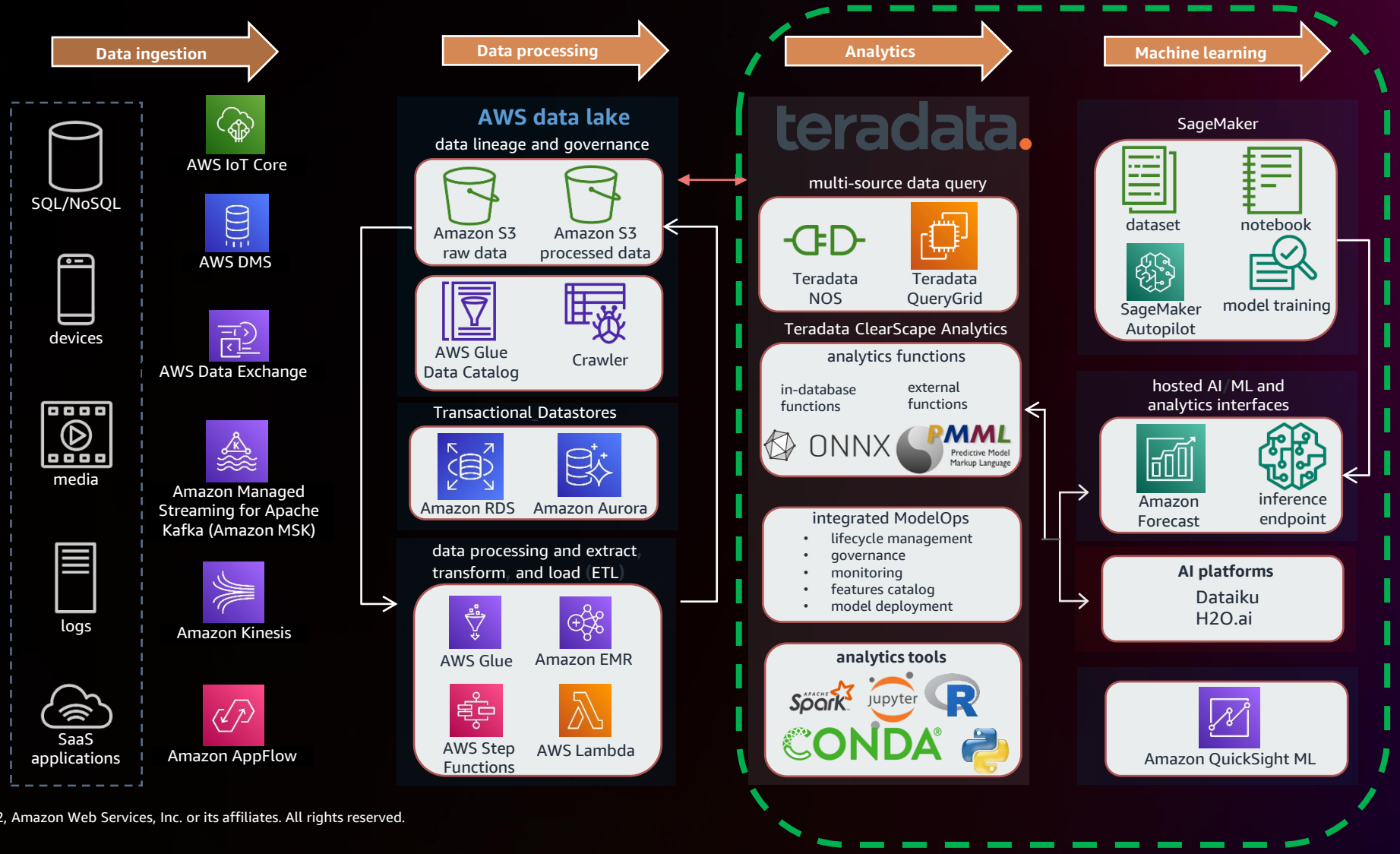
ADDITIONAL INTEGRATIONS

- AWS PrivateLink
  Private connections
- AWS Marketplace
  Offering
- AWS CloudFormation
  Orchestration
- AWS Key Management Svcs
  Key management
- Amazon EKS
  Kubernetes containers
- Amazon EBS
  Persisent storage
- Amazon CloudTrail
  API monitoring
- Amazon CloudWatch
  Resource monitoring
- Amazon EBS Snapshots
  Rapid backups

Note: Not all service integrations are shown.

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Modern data pipeline using AWS and Teradata



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Ashish Yajnik
ashish.yajnik@teradata.com

Peter Singla
peter.singla@charter.com

Doug Mbaya
mbay@amazon.com



Please complete the session survey in the **mobile app**

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.