aws SUMMIT

INDIA | MAY 25, 2023

ANA004

Dream11's journey of building their data platform on AWS

Vatsal Shah
Senior Solutions Architect
AWS India

Vikas Gite
Principal Engineer
Dream11

Agenda

- Why modern data architecture?
- Modern data strategy
- Building modern data architectures on AWS
- Dream 11’s data platform
- Journey through time – The past, the present & the future
- Benefits of running on AWS and key takeaways

The business value of a modern data strategy

$65 million
increase in net income.[1]

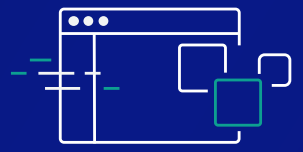
415%
five-year ROI[2]

48%
reduced total cost of operations[2]

[1]Read the article on Forbes.com | [2]Read the report

Challenges of data analytics at scale

Variety of sources and data types



Performance

Increasing and unpredictable cost

Diverse analytics needs

Manageability

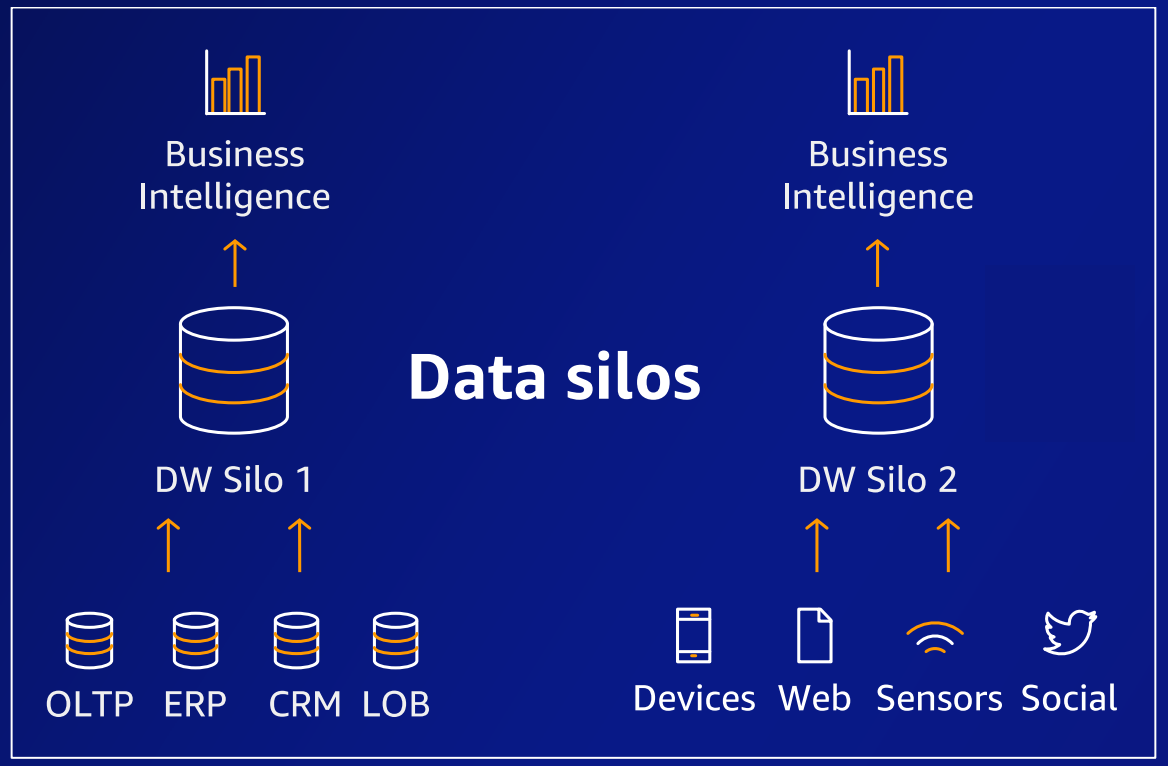
Inflexible tools

Data volume and velocity

Scalability

Security and Compliance

Challenges in data silos



- Relational vs Non-Relational
- TBs-PBs Scale
- Schema Defined Prior to Data Load
- Operational and Ad Hoc Reporting
- Large Initial Capex + $$K / TB/ Year

The five pillars of a modern data architecture

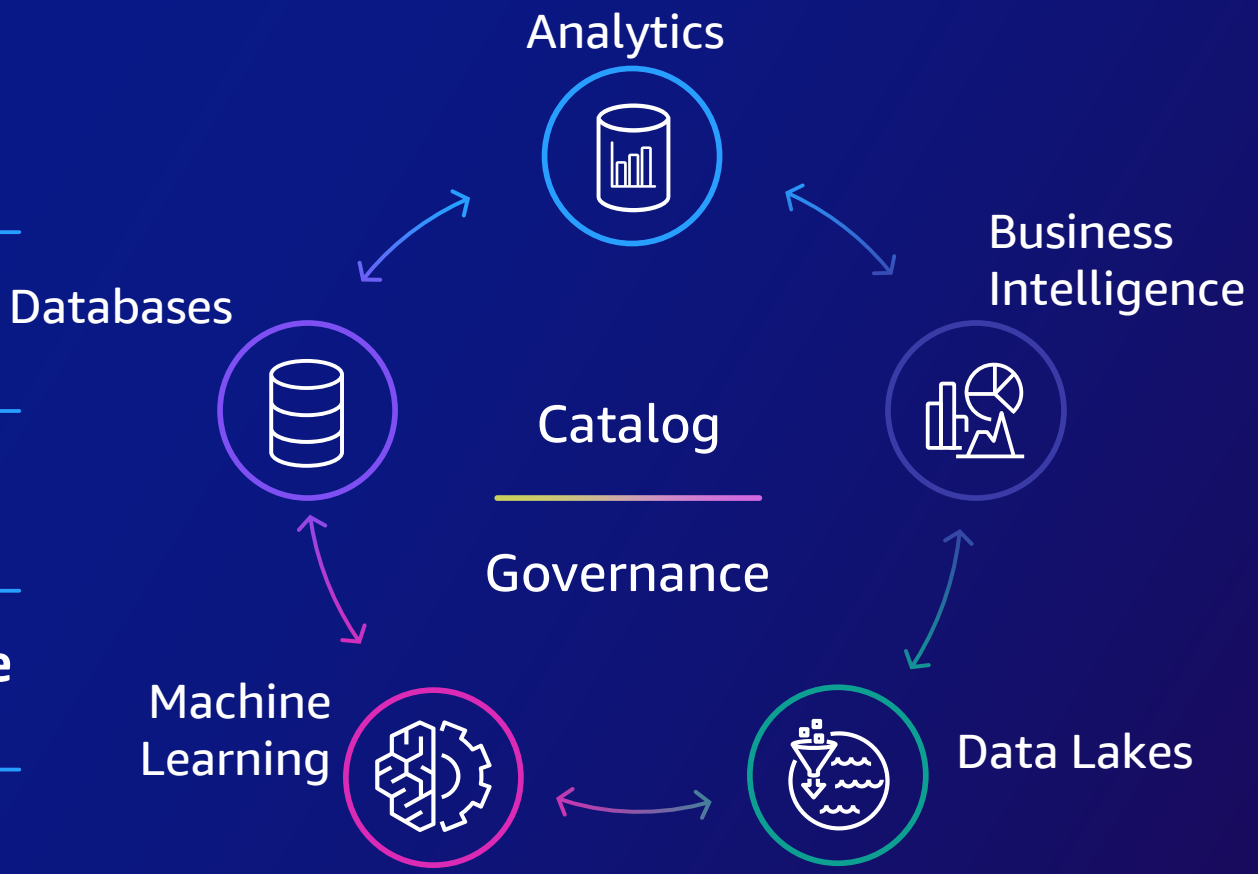
Unified analytics

Highest performance at the lowest cost

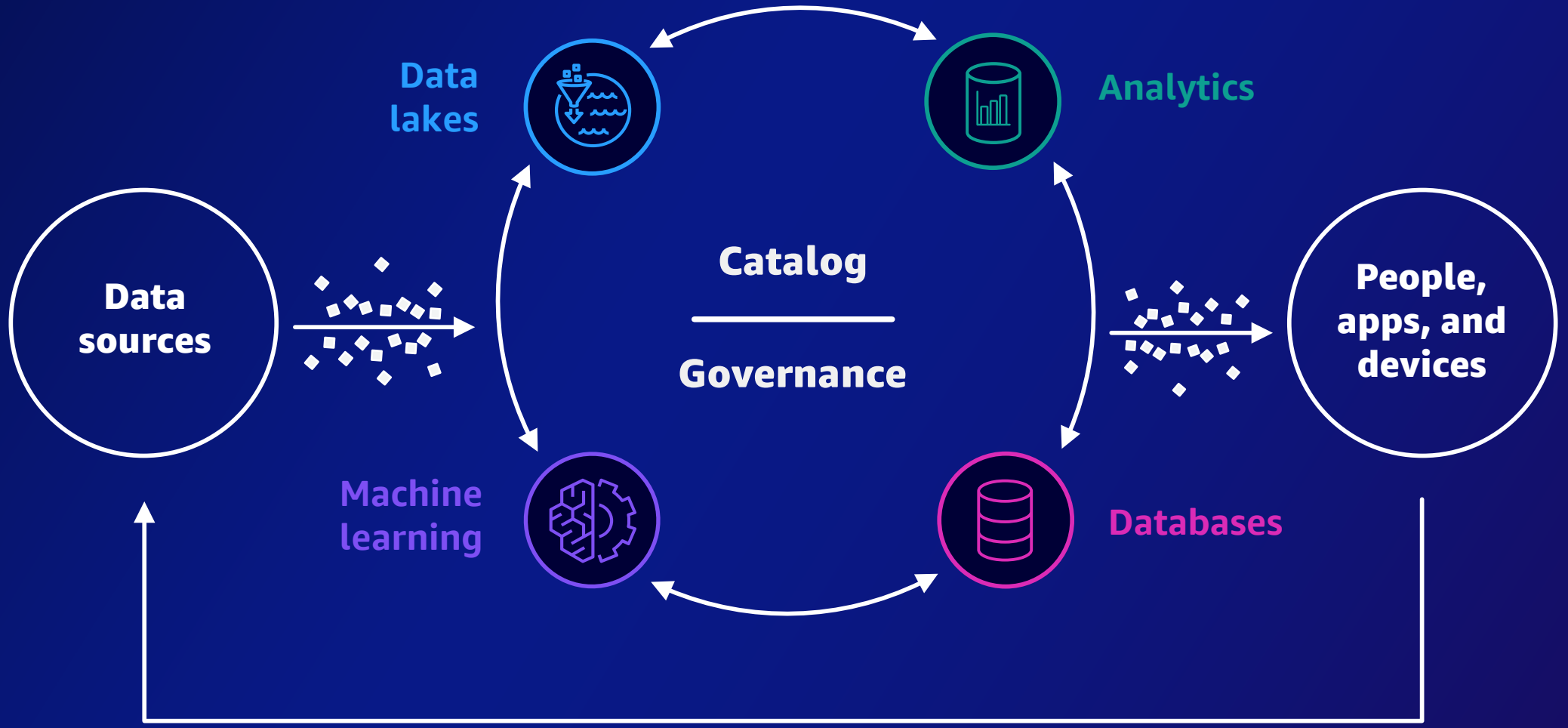
Machine learning integration

Unified data access, security and governance

Insights for everyone



Modern data architecture



Broadest and most cost-effective set of analytics services

Interactive query
Amazon **Athena**

Big data processing
Amazon **EMR**

Operational and log analytics
Amazon **OpenSearch Service**

Real-time analytics
Amazon **Kinesis** and Amazon **MSK**

Business Intelligence
Amazon **QuickSight**

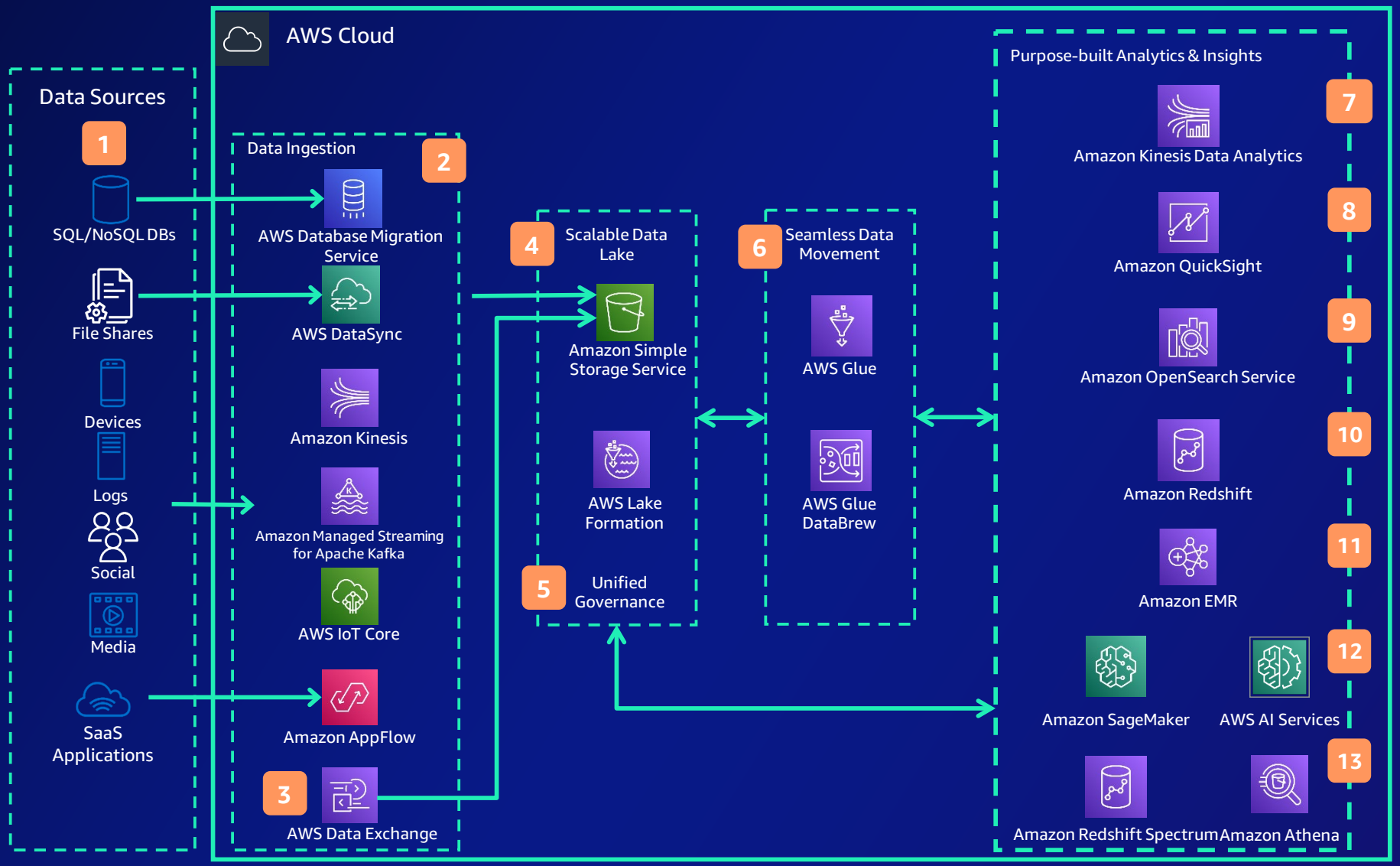
Data warehouse
Amazon **Redshift**

Centralize your data at scale
AWS-powered data lakes

Internal and External Data Integration
AWS Glue and AWS Data Exchange

Modern data analytics reference architecture on AWS

BUILD DATA ANALYTICS PIPELINES USING MODERN DATA ANALYTICS APPROACH TO DERIVE INSIGHTS FROM THE DATA



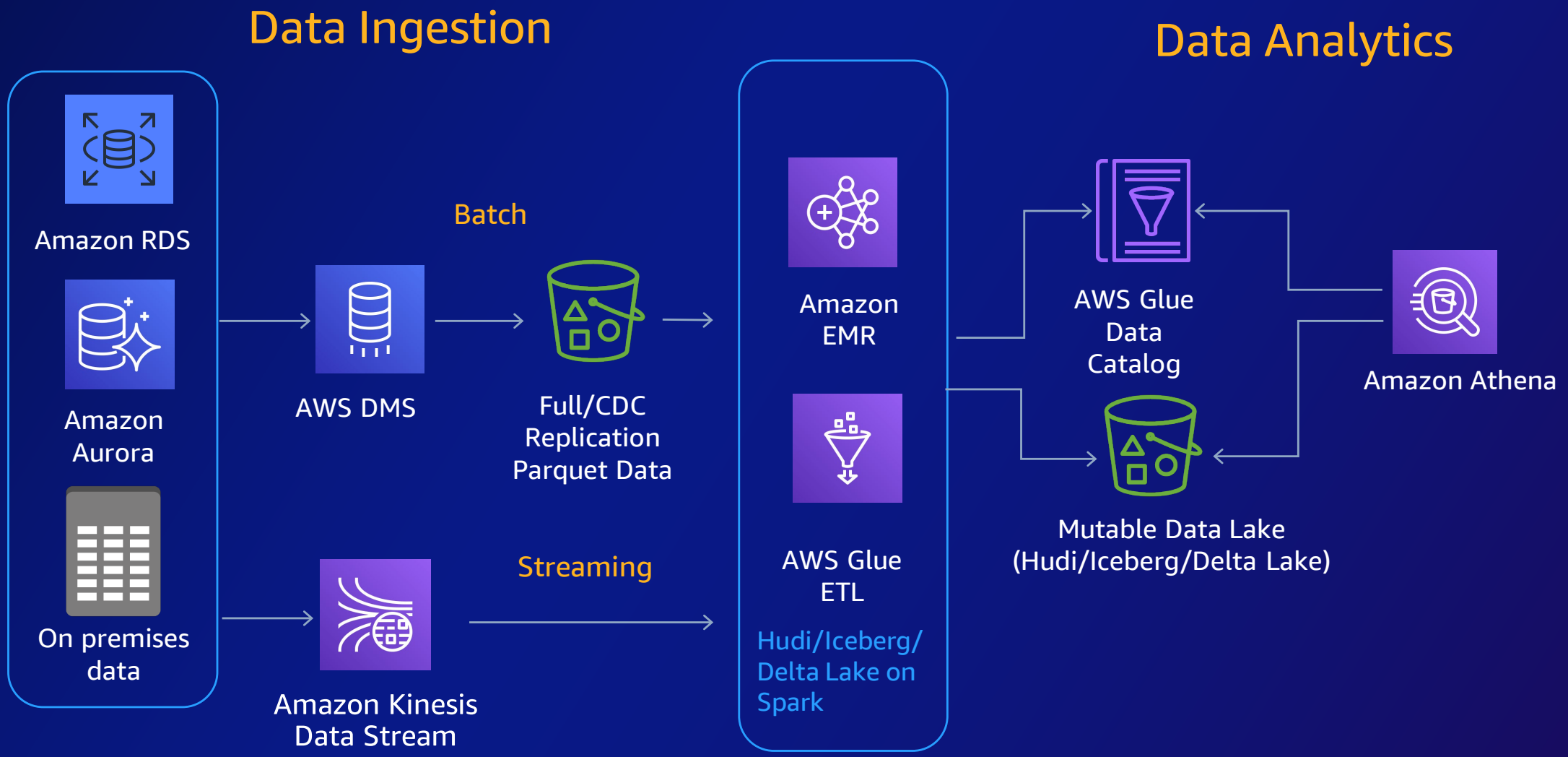
1. Data is collected from multiple data sources across the enterprise, SaaS applications, edge devices, logs, streaming media, and social network
2. Based on the type of the data source, **AWS Database Migration Service, AWS DataSync, Amazon Kinesis, Amazon Managed Streaming for Apache Kafka, AWS IoT Core**, and **Amazon AppFlow** are used to ingest the data into a Data Lake in AWS
3. **AWS Data Exchange** is used for integrating third-party data into the Data Lake
4. **AWS Lake Formation** is used to build the scalable data lake, and **Amazon S3** is used as the data lake storage
5. **AWS Lake Formation** is also used to enable unified governance to centrally manage the security, access control, and audit trails
6. **AWS Glue, AWS Glue DataBrew,** and **AWS Glue Elastic Views** are used to transform, enrich, move, and replicate data across multiple data stores and the data lake
7. **Amazon Kinesis Data Analytics** is used to transform and analyze streaming data in real time
8. **Amazon QuickSight** provides machine learning-powered business intelligence
9. **Amazon OpenSearch Service** can be used for operational analytics
10. **Amazon Redshift** is used as a Cloud Data Warehouse
11. **Amazon EMR** provides the cloud big data platform for processing vast amounts of data using open source tools
12. **Amazon SageMaker** and **AWS AI services** can be used to build, train, and deploy machine learning models, and add intelligence to your applications
13. **Amazon Redshift Spectrum** and **Amazon Athena** enable interactive querying, analyzing, and processing capabilities

AWS Reference Architecture

http://bit.ly/3VoIVQE

Architecture of transactional Data Lake

TRANSACTIONS ENABLED BY HUDI, ICEBERG, DELTA LAKE

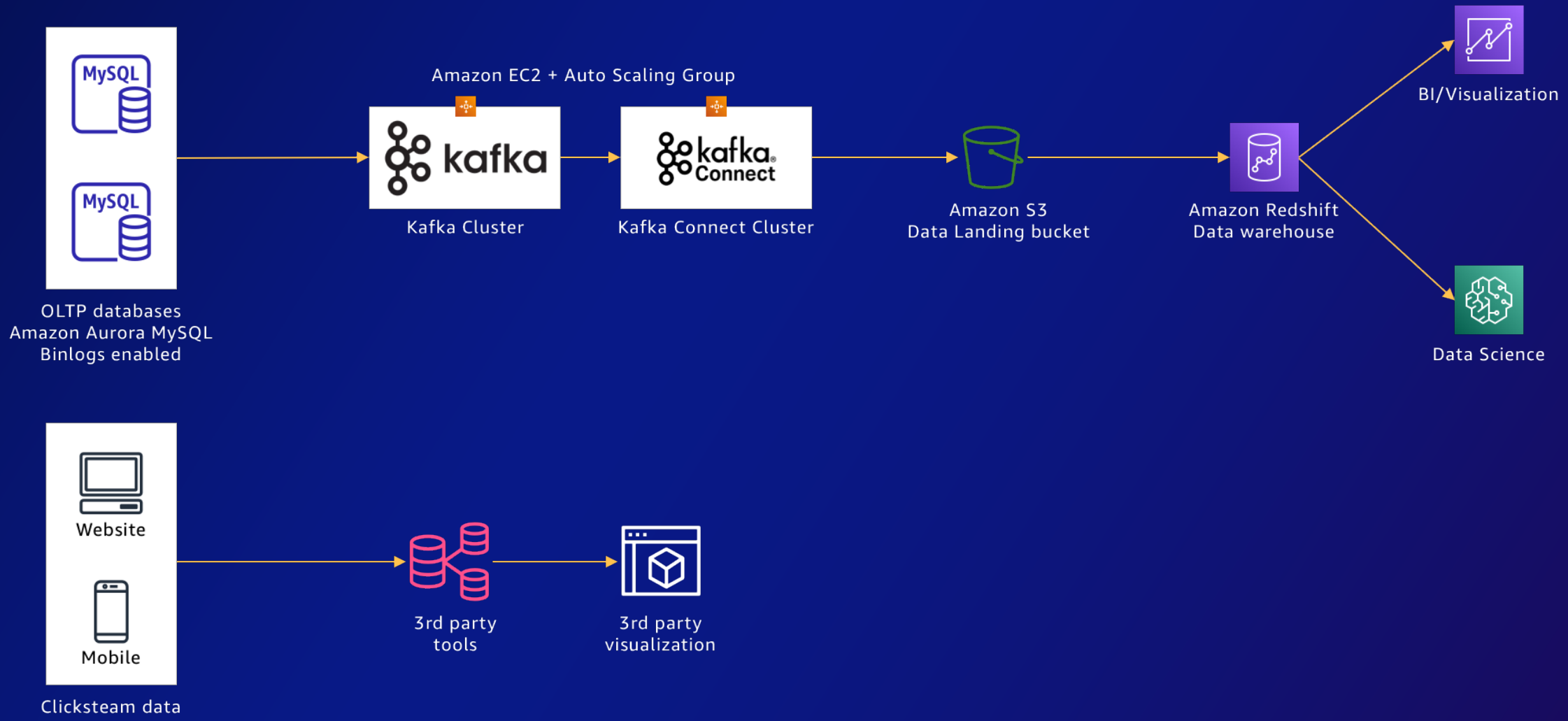


Dream Sports

- 150M users
- 10.56M concurrency
- Billions of events / day
- Petabyte scale data platform
- Hundreds of microservices
- ML driven personalisation, scaling etc.

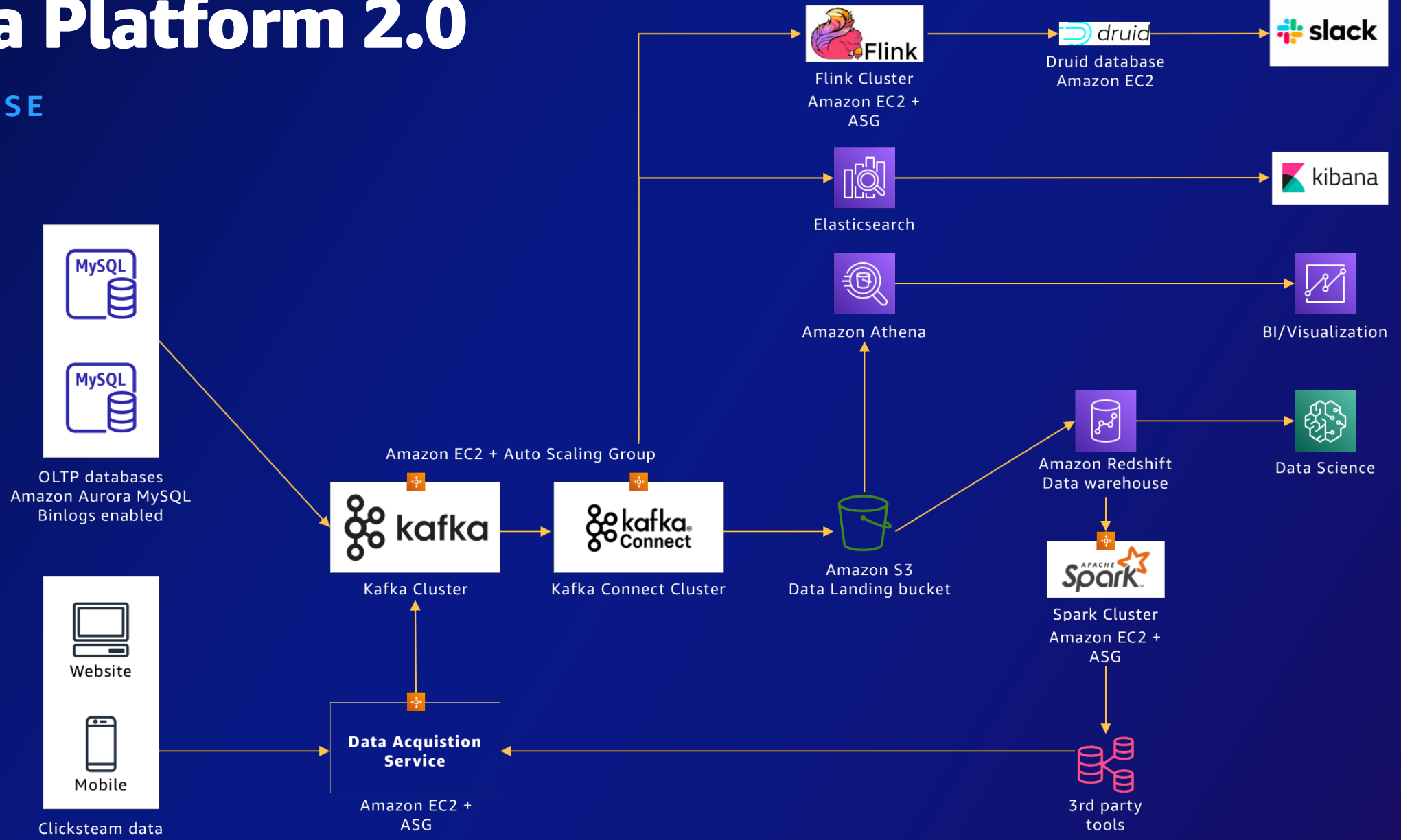
Data Platform 1.0

DAY 0

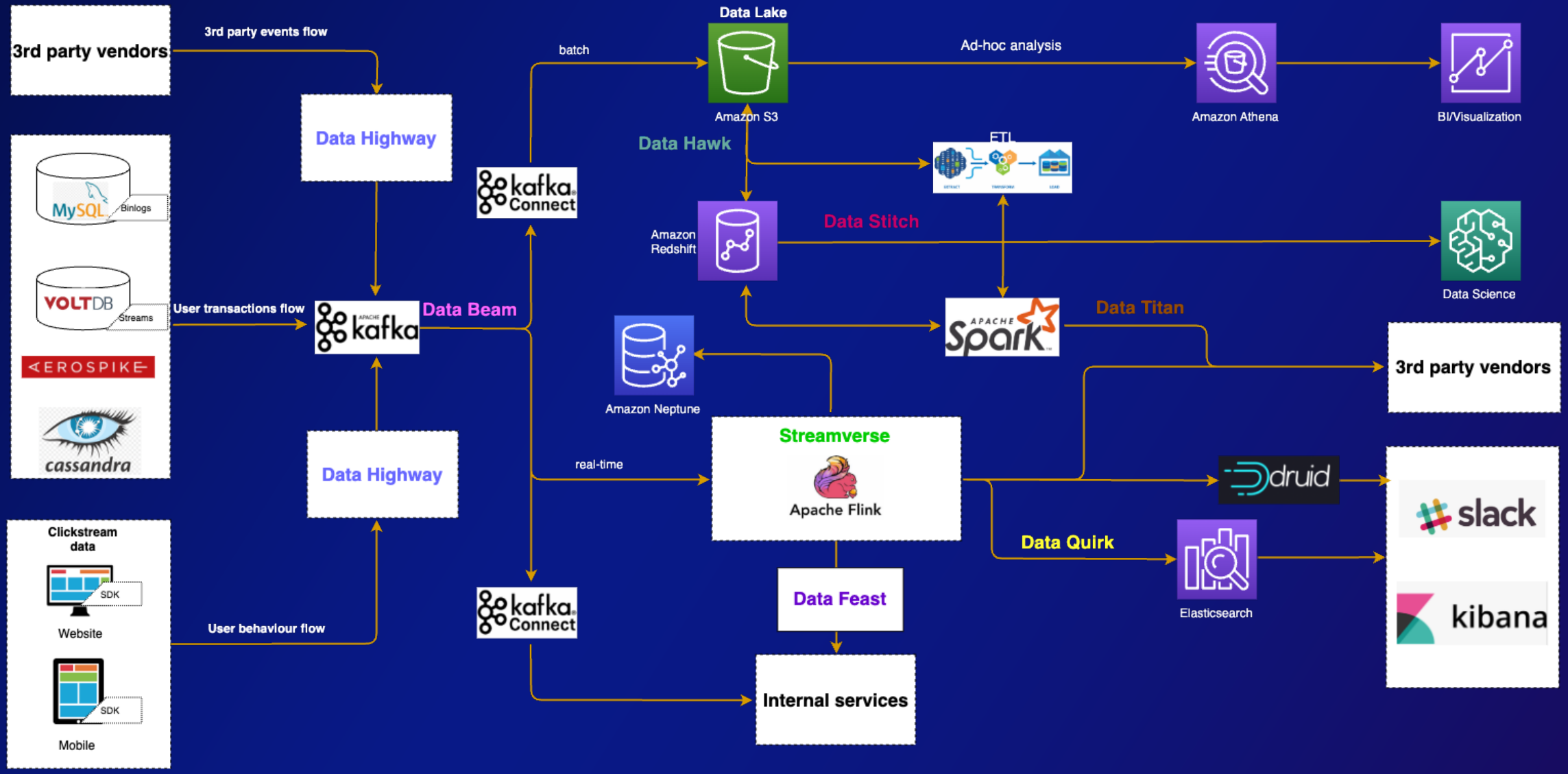


Data Platform 2.0

IN-HOUSE



Current architecture



Key takeaways and benefits of running it on AWS

- Start simple; evolve with scale and complexity
- Serverless for unknown traffic patterns
- Decouple fault planes
- Managed services for Spark, Presto, etc.
- Native support for Apache Iceberg, Apache Hudi, etc.
- Instant scalability with high reliability

Important resources

- Modern data analytics reference architecture - http://bit.ly/3VoIVQE
- Transactional & mutable data lakes workshop - https://catalog.us-east-1.prod.workshops.aws/workshops/520e974c-0fee-4585-9601-9af535d4d908/en-US
- Dream 11 tech blog - https://tech.dream11.in/blog
- AWS Analytics blogs - https://aws.amazon.com/blogs/big-data/category/analytics/

skillbuilder.aws

Your time is now

Build in-demand cloud skills your way

Thank you!

Please complete the session survey

Vatsal Shah
Senior Solutions Architect
AWS India

Vikas Gite
Principal Engineer
Dream11