
AWS
re:Invent

ANT309

How Disney+ uses fast data ubiquity to improve the customer experience

Martin Zapletal
Director, Software Engineering
Disney+

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

Streaming data and Disney+

Before – silos

Evolution – streaming silos

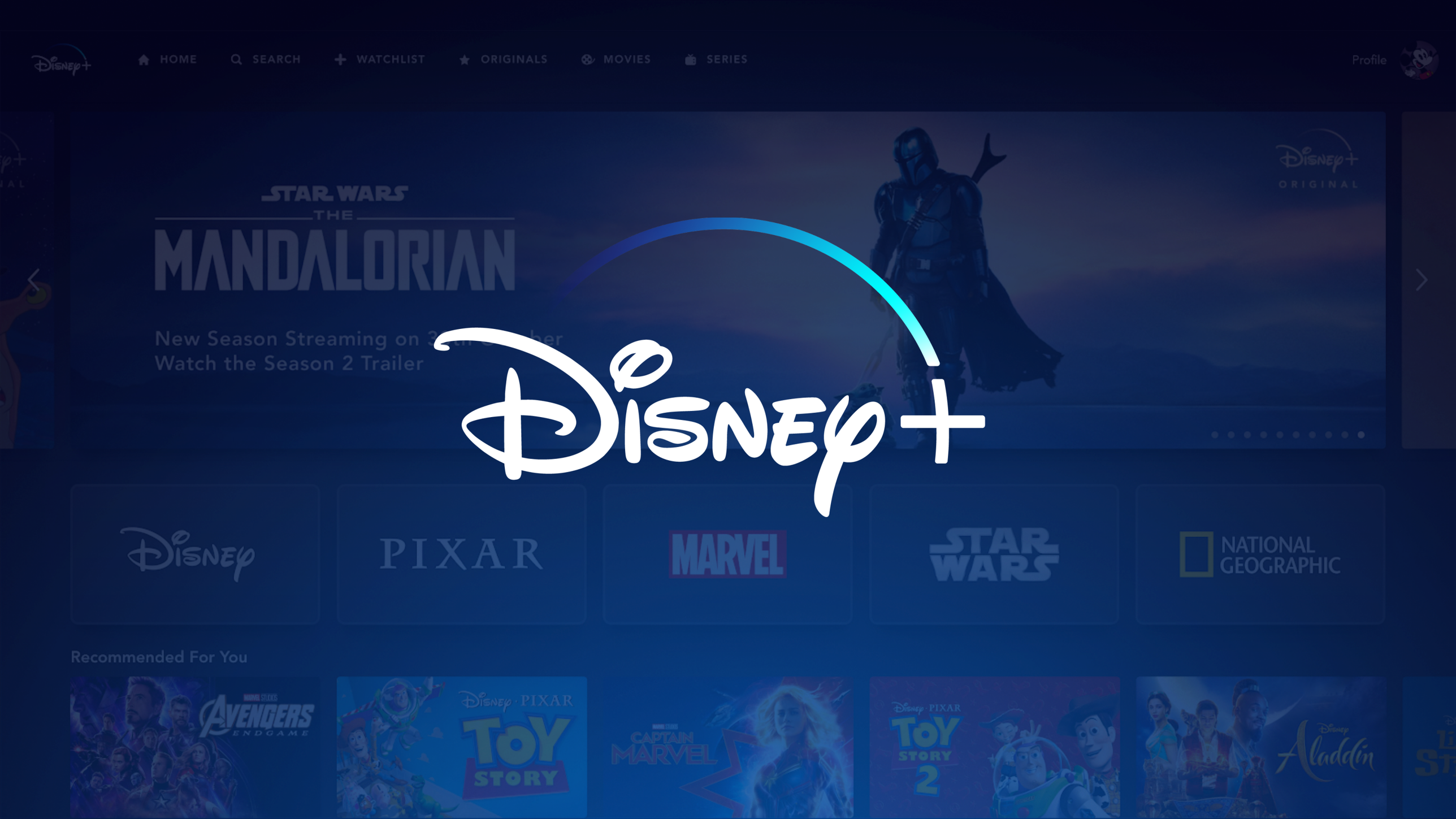
Now – data driven

Streaming Data Platform

Examples – ubiquity, platform, culture



© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.


HOME
SEARCH
WATCHLIST
ORIGINALS
MOVIES
SERIES
Profile
STAR WARS
THE
MANDALORIAN
New Season Streaming on 3
Watch the Season 2 Trailer
Disney+
ORIGINAL
Disney
PIXAR
MARVEL
STAR WARS
NATIONAL GEOGRAPHIC
Recommended For You
AVENGERS
ENDGAME
Disney · PIXAR
TOY STORY
CAPTAIN MARVEL
Disney · PIXAR
TOY STORY 2
Aladdin

Streaming data

- Variety of streaming use cases with varying needs
- Dozens of millions of users
- Hundreds of Amazon Kinesis data streams
- Thousands of shards
- Multiple regions
- Billions of events
- Terabytes of data

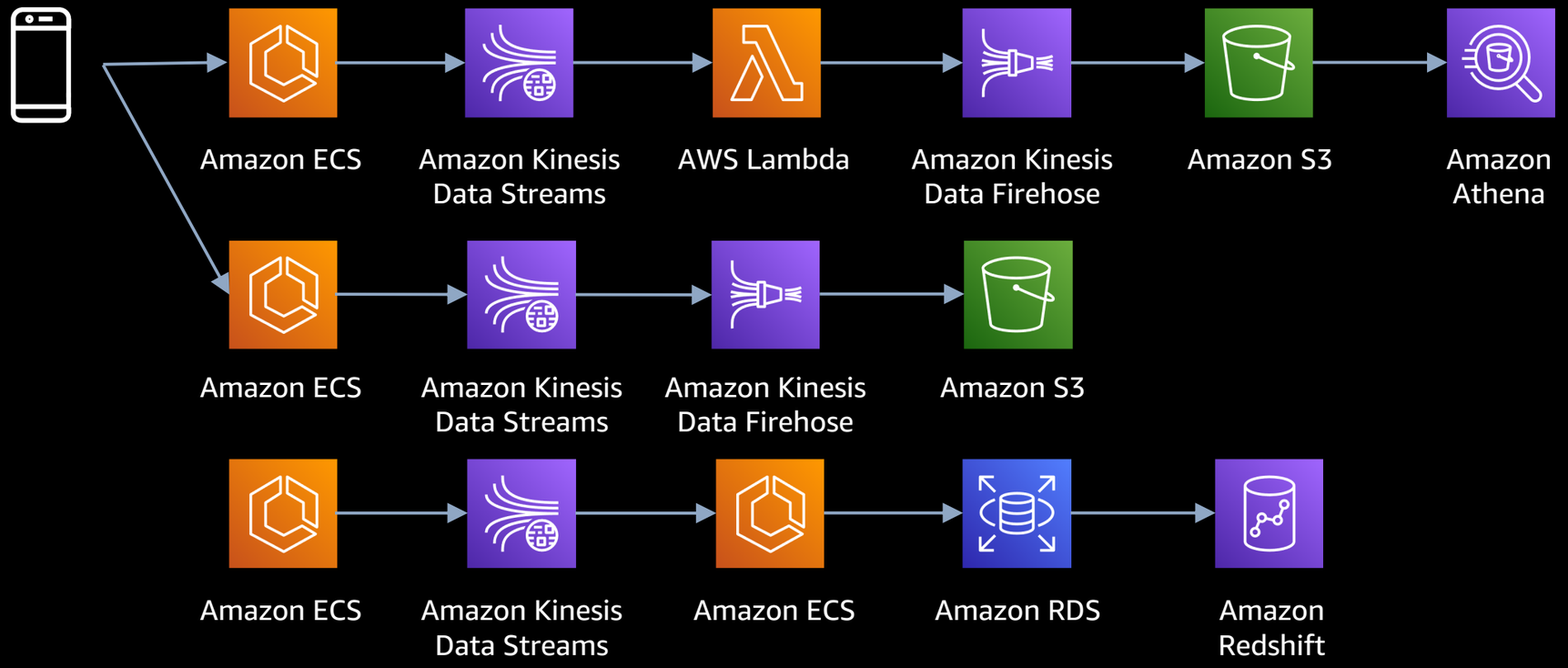
Before: Silos

- Microservices
- Databases and data warehouses
- Batch processing
- Slow and limited insights
- Silos



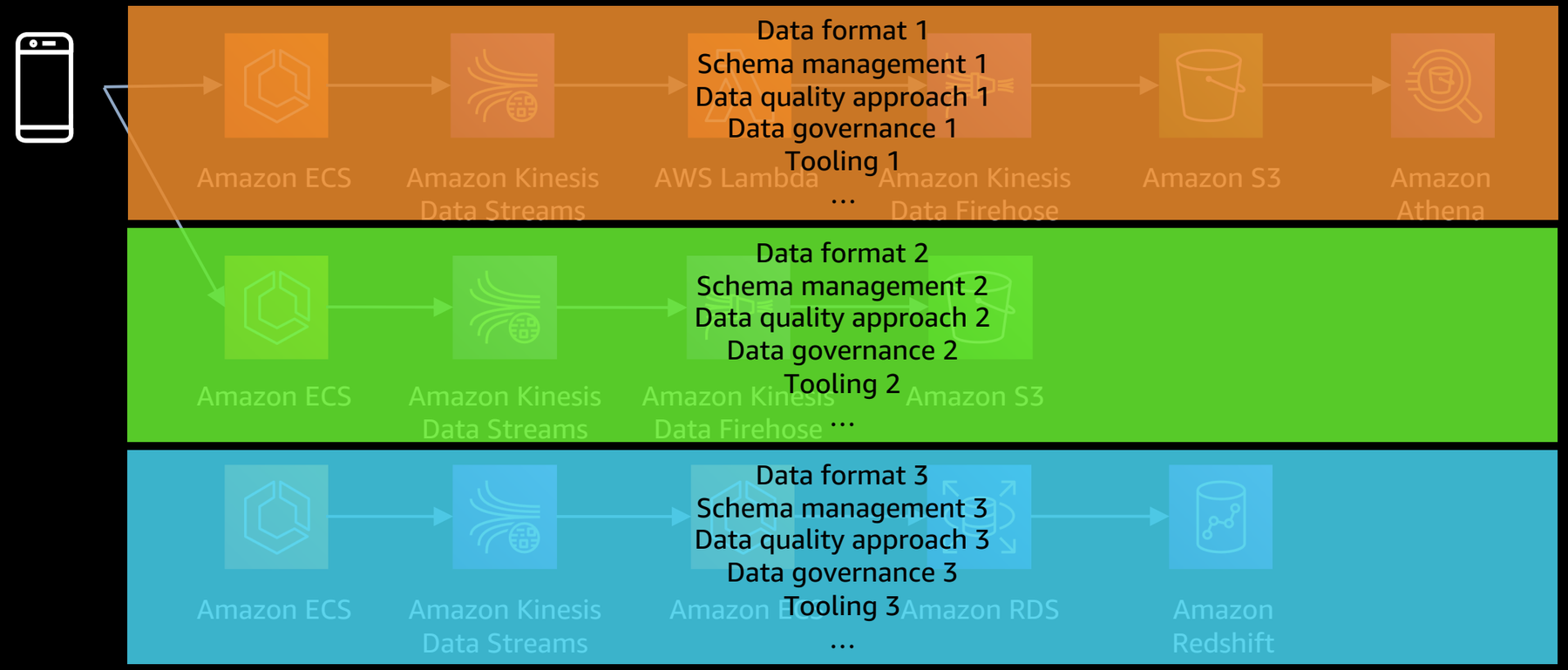
Evolution: Streaming silos

- Streaming, event driven, asynchronous
- Custom, unique integrations and data warehouses



Evolution: Streaming silos

- Streaming, event driven, asynchronous
- Custom, unique integrations and data warehouses



Now: Data driven

- (Fast) data democracy
- Real-time data, insights, ML
- Experimentation
- First-class consideration
- Culture

"Data / insights they need available at the time they need it"



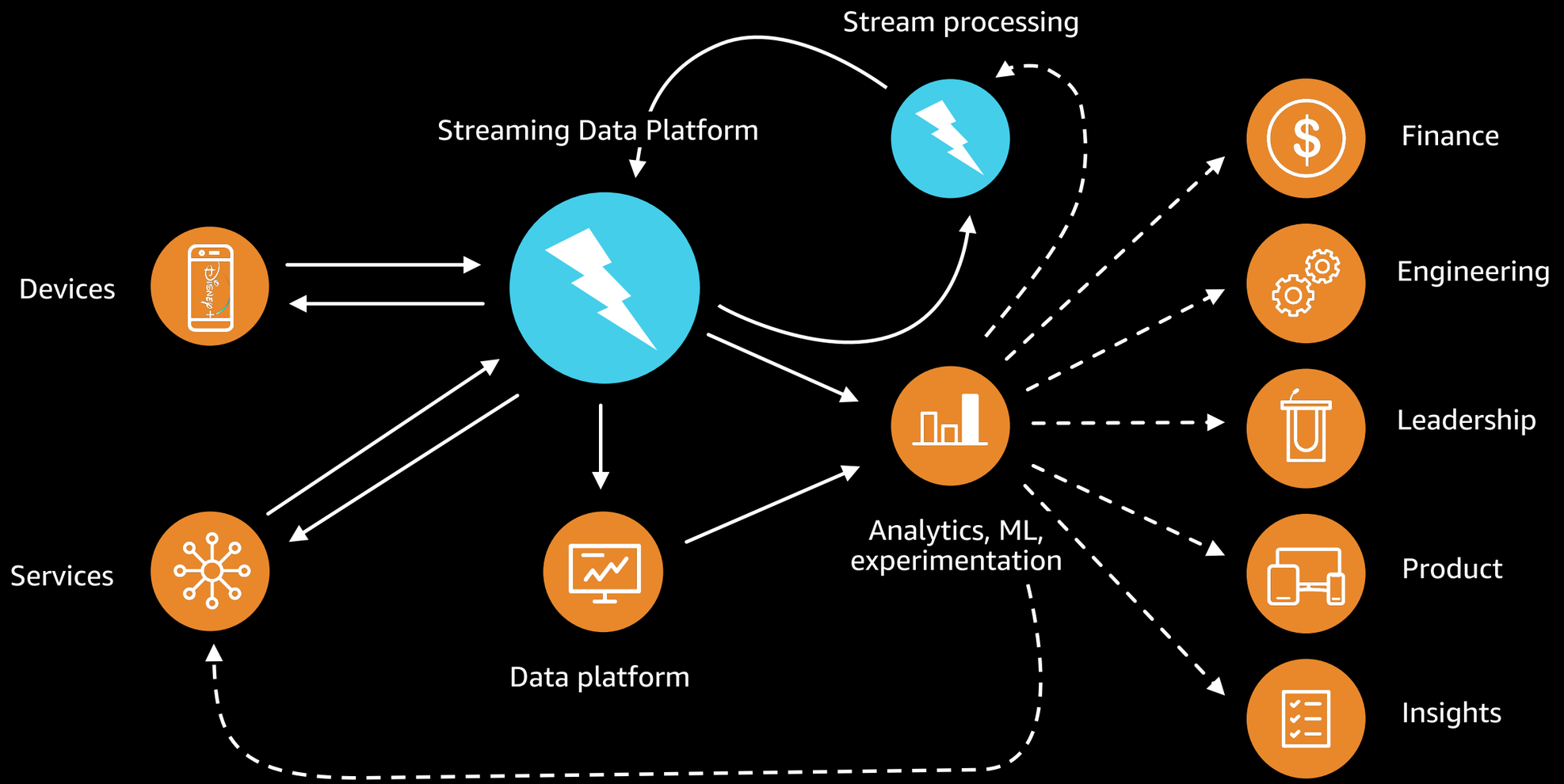
Streaming Data Platform



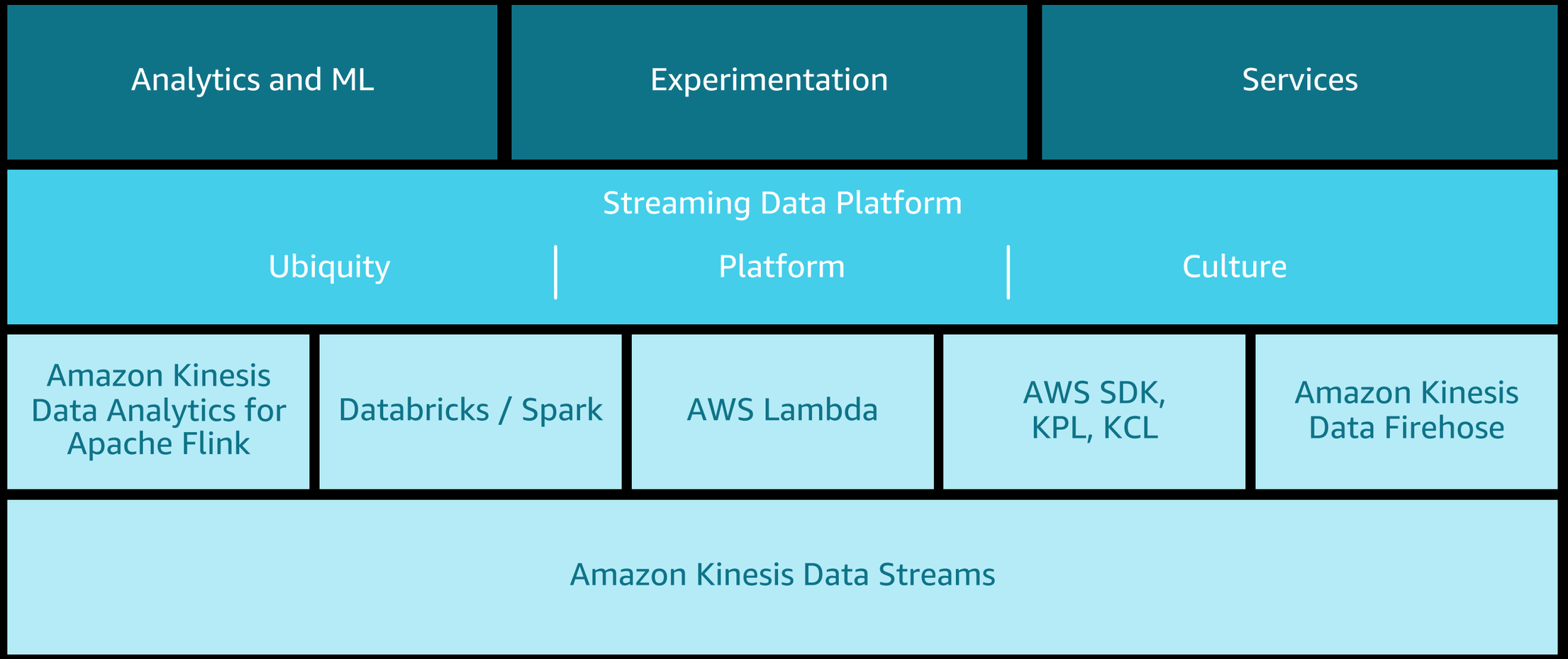
AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Streaming Data Platform



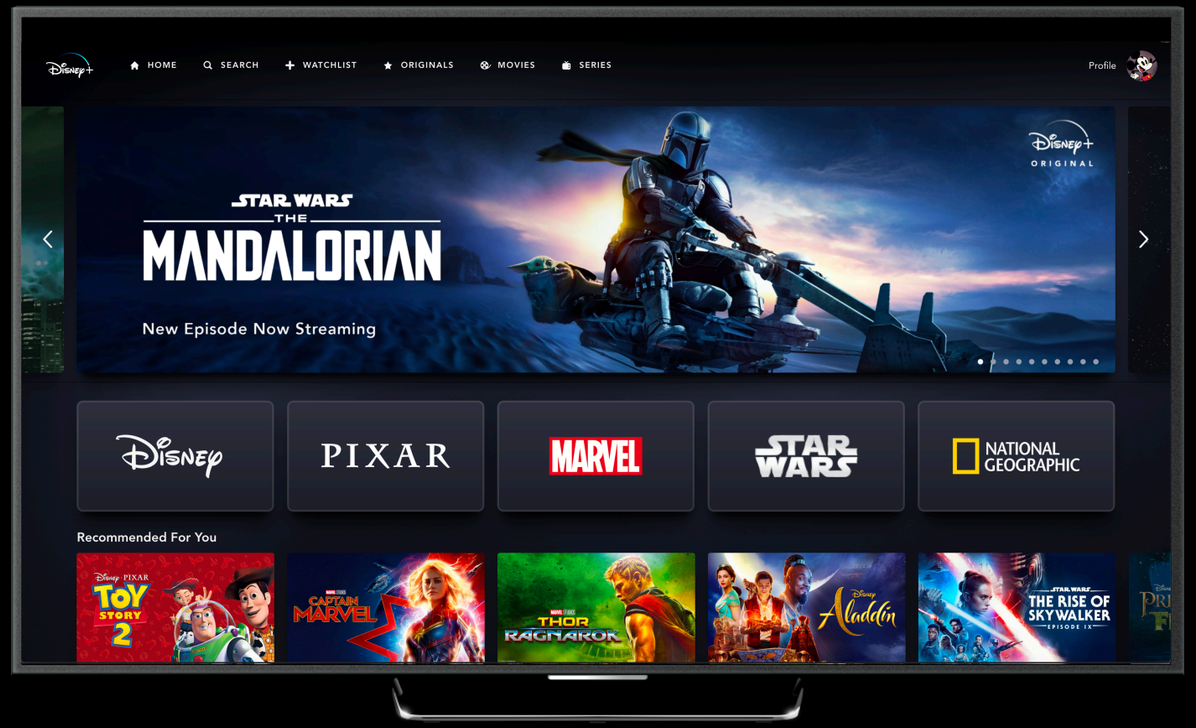
Streaming Data Platform



Kinesis

- Need a reliable, performant, cost-efficient event log
- Kinesis, Kafka, Pulsar, and others
- Amazon Kinesis Data Streams
  - Replicated, partitioned, ordered, distributed log
  - Managed
  - Replication to 3 AZs
  - Integration with other AWS services
  - Near real time
  - Scalability
  - Elasticity

Examples


HOME
SEARCH
WATCHLIST
ORIGINALS
MOVIES
SERIES
Profile
STAR WARS
THE
MANDALORIAN
New Episode Now Streaming
Disney+
ORIGINAL
Disney
PIXAR
MARVEL
STAR WARS
NATIONAL GEOGRAPHIC
Recommended For You
TOY STORY 2
CAPTAIN MARVEL
THOR RAGNAROK
Aladdin
THE RISE OF SKYWALKER


Toy Story
:18:11 / 01:04:41

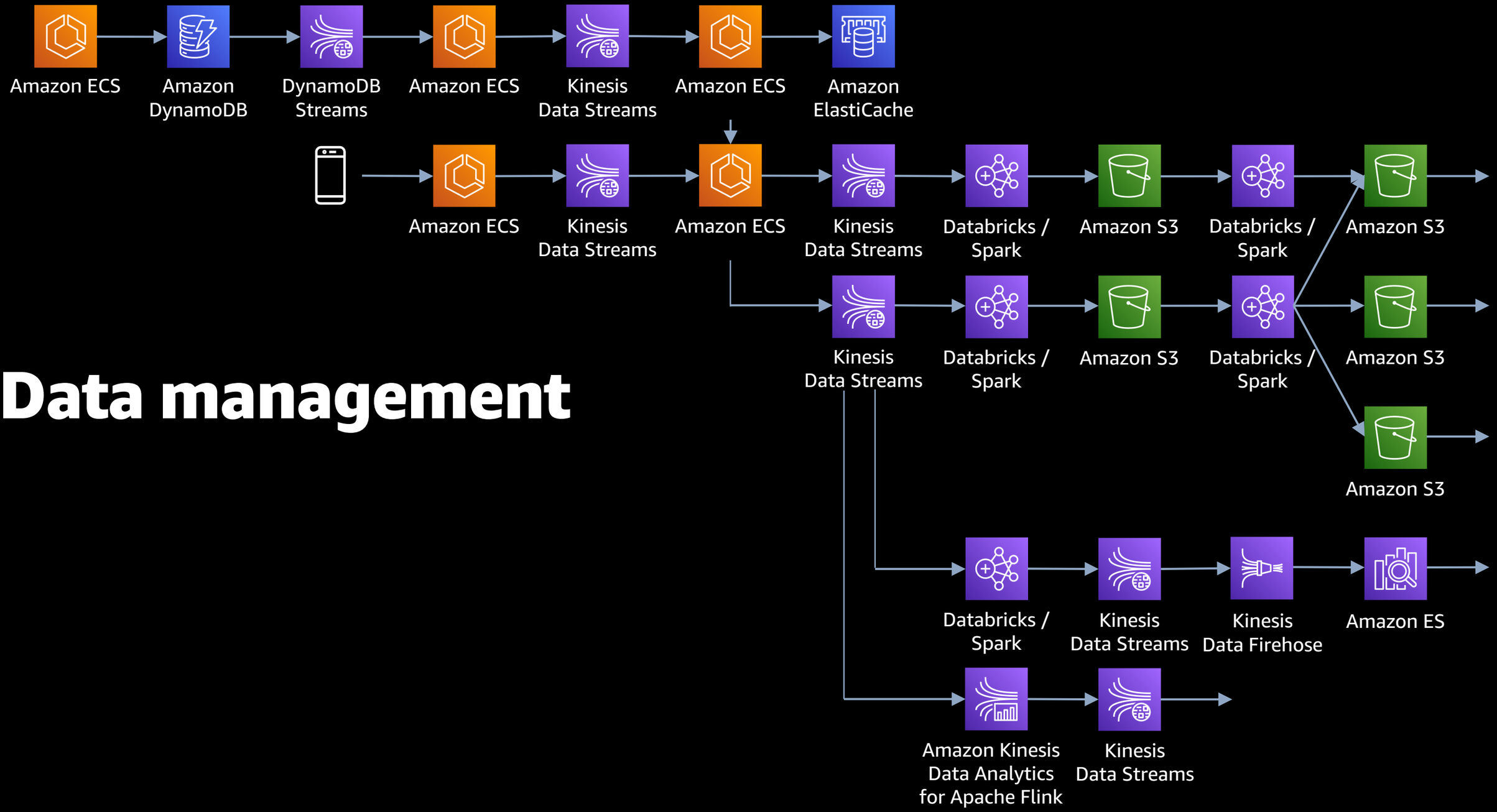
Ubiquity



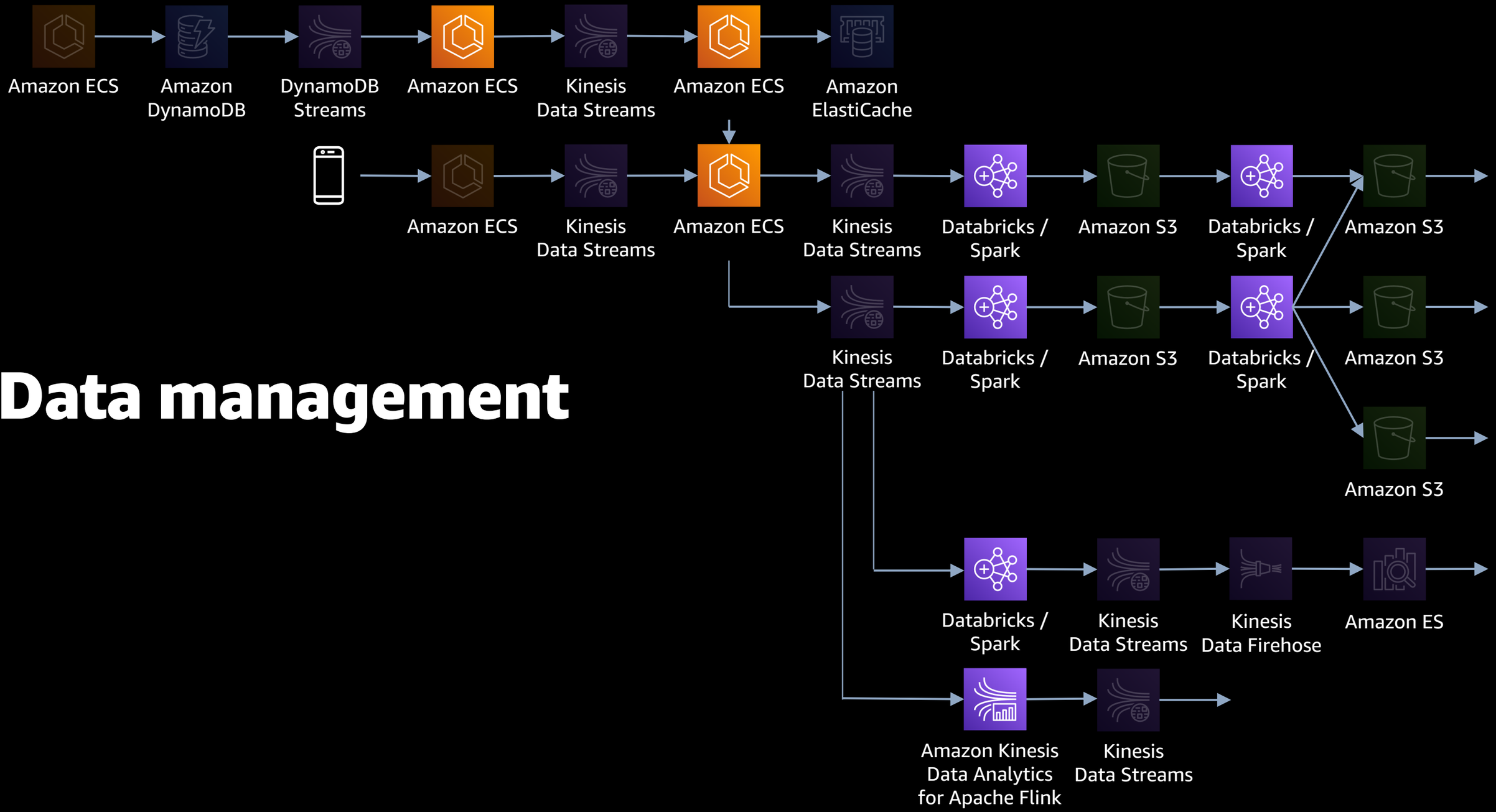
AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

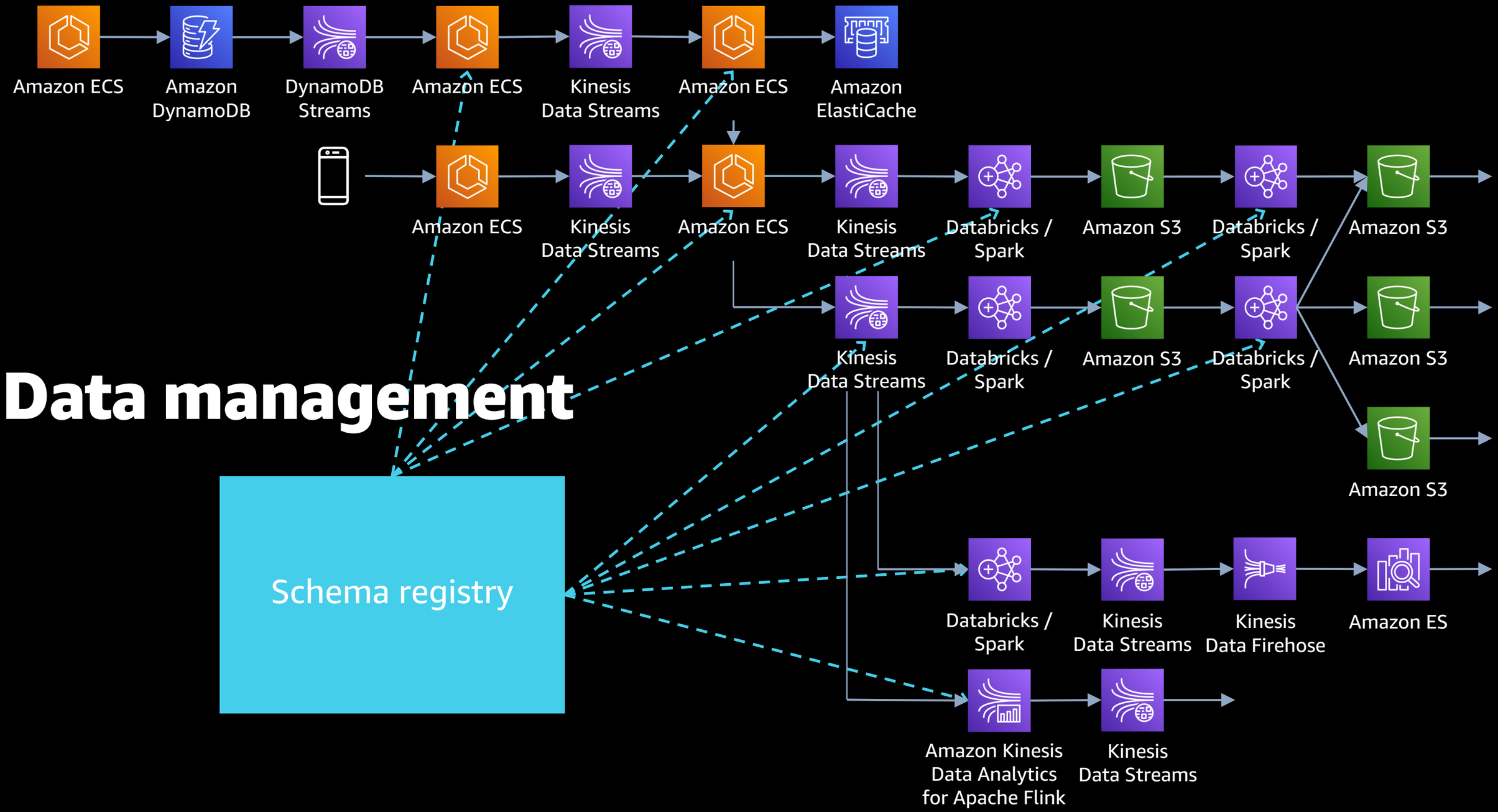
Data management



Data management



Data management



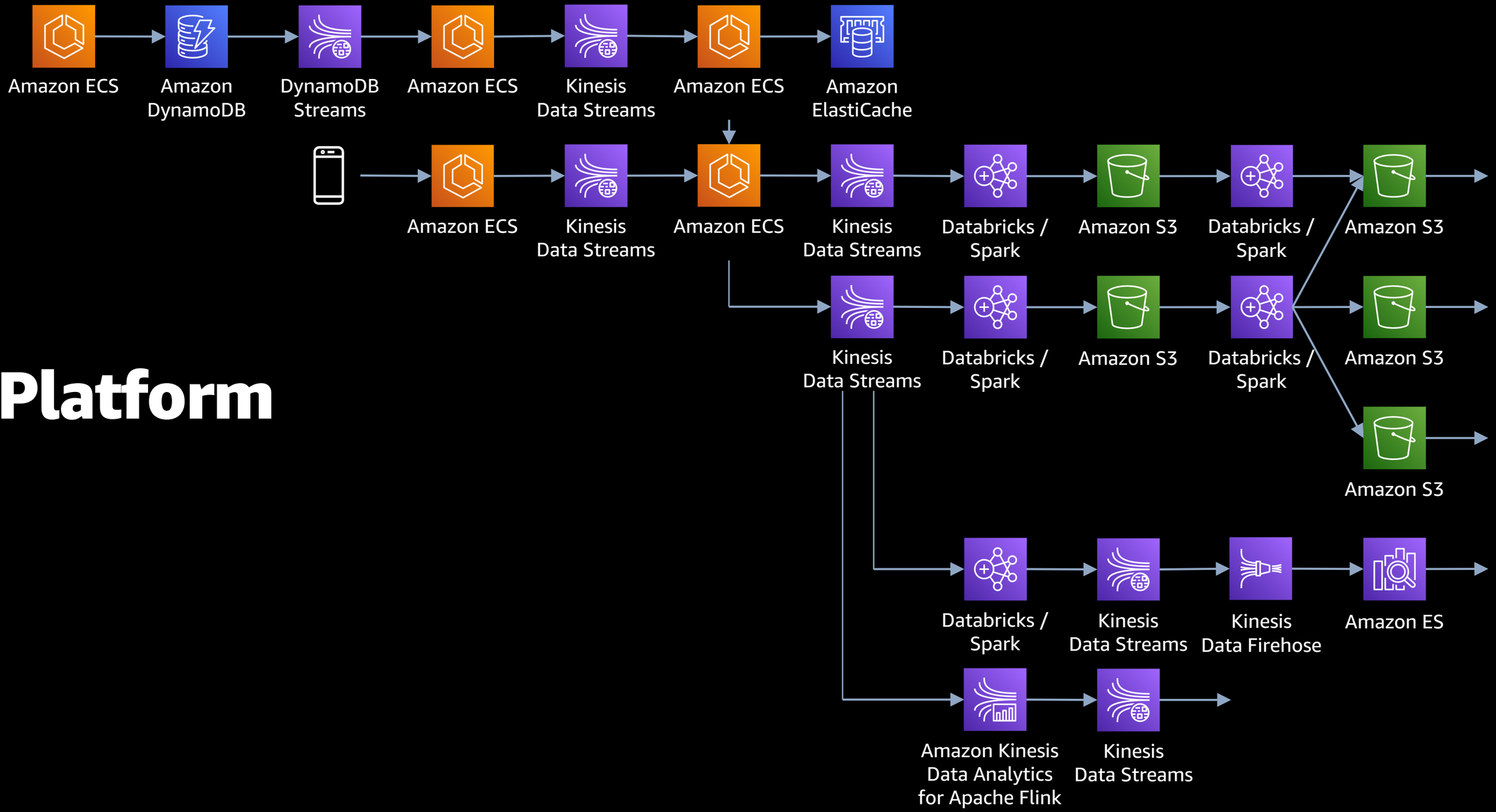
Platform

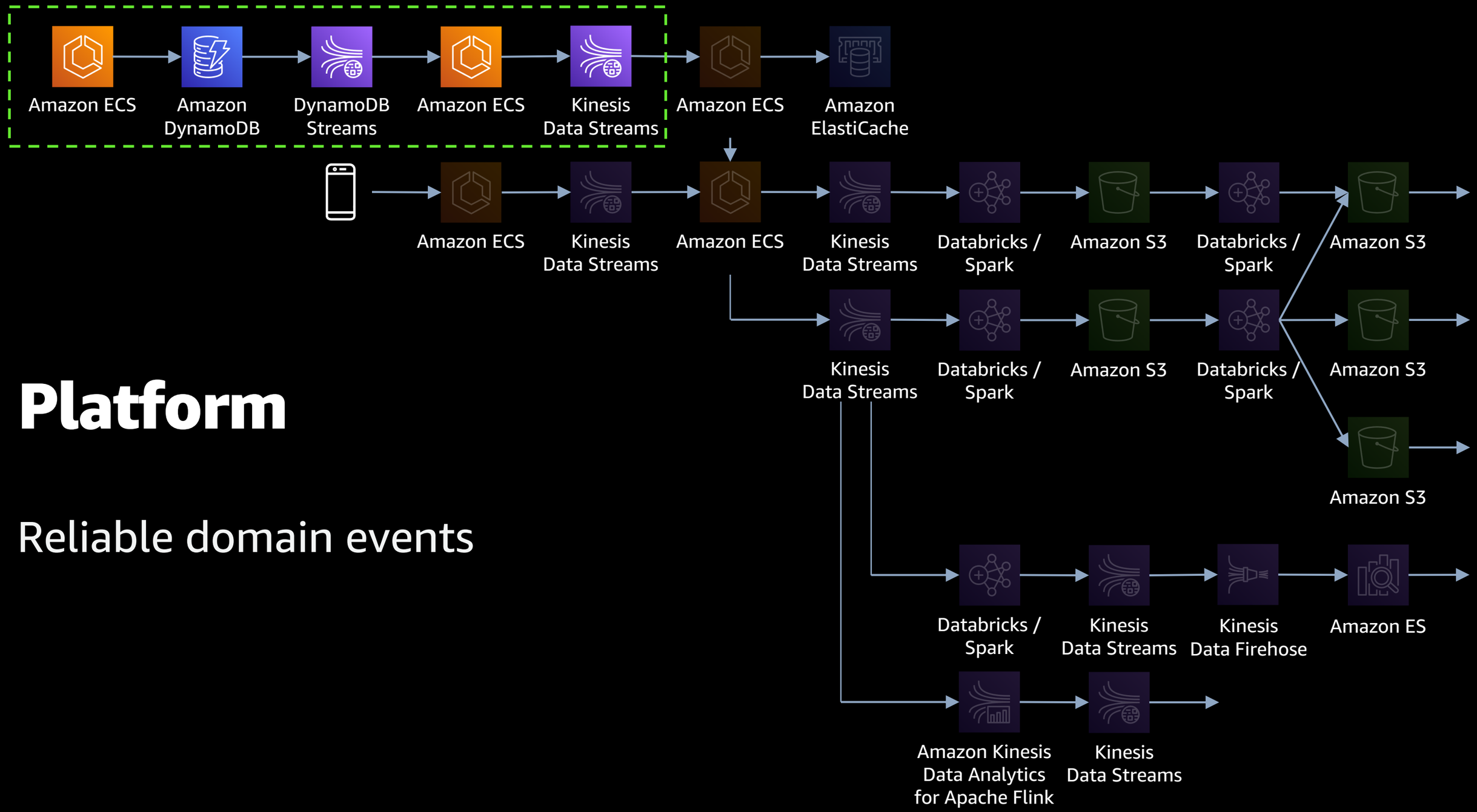


AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

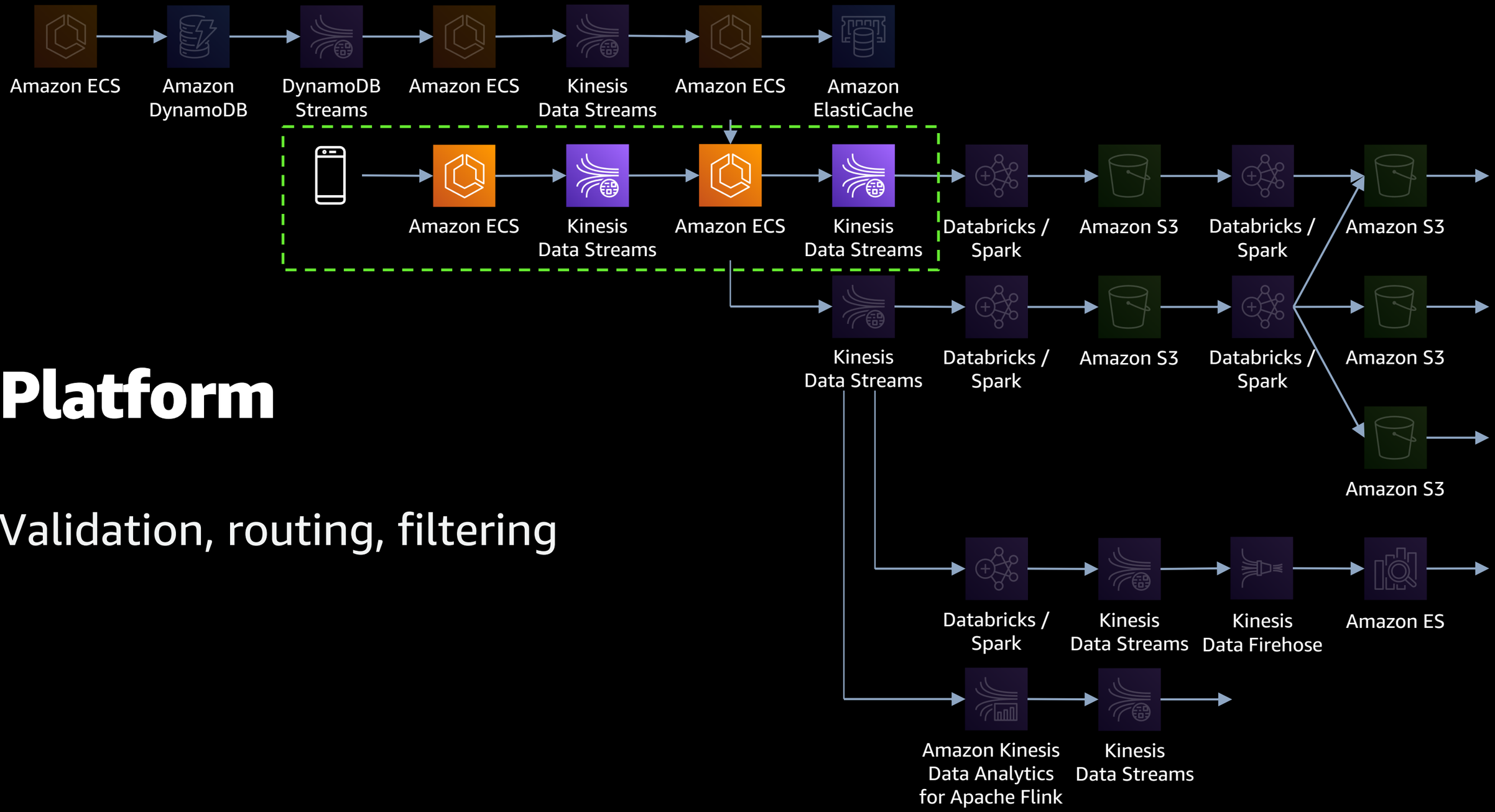
Platform


Amazon ECS
Amazon DynamoDB
DynamoDB Streams
Amazon ECS
Kinesis Data Streams
Amazon ECS
Amazon ElastiCache
Amazon ECS
Kinesis Data Streams
Amazon ECS
Kinesis Data Streams
Databricks / Spark
Amazon S3
Databricks / Spark
Amazon S3
Kinesis Data Streams
Databricks / Spark
Amazon S3
Databricks / Spark
Amazon S3
Amazon S3
Databricks / Spark
Kinesis Data Streams
Kinesis Data Firehose
Amazon ES
Amazon Kinesis Data Analytics for Apache Flink
Kinesis Data Streams



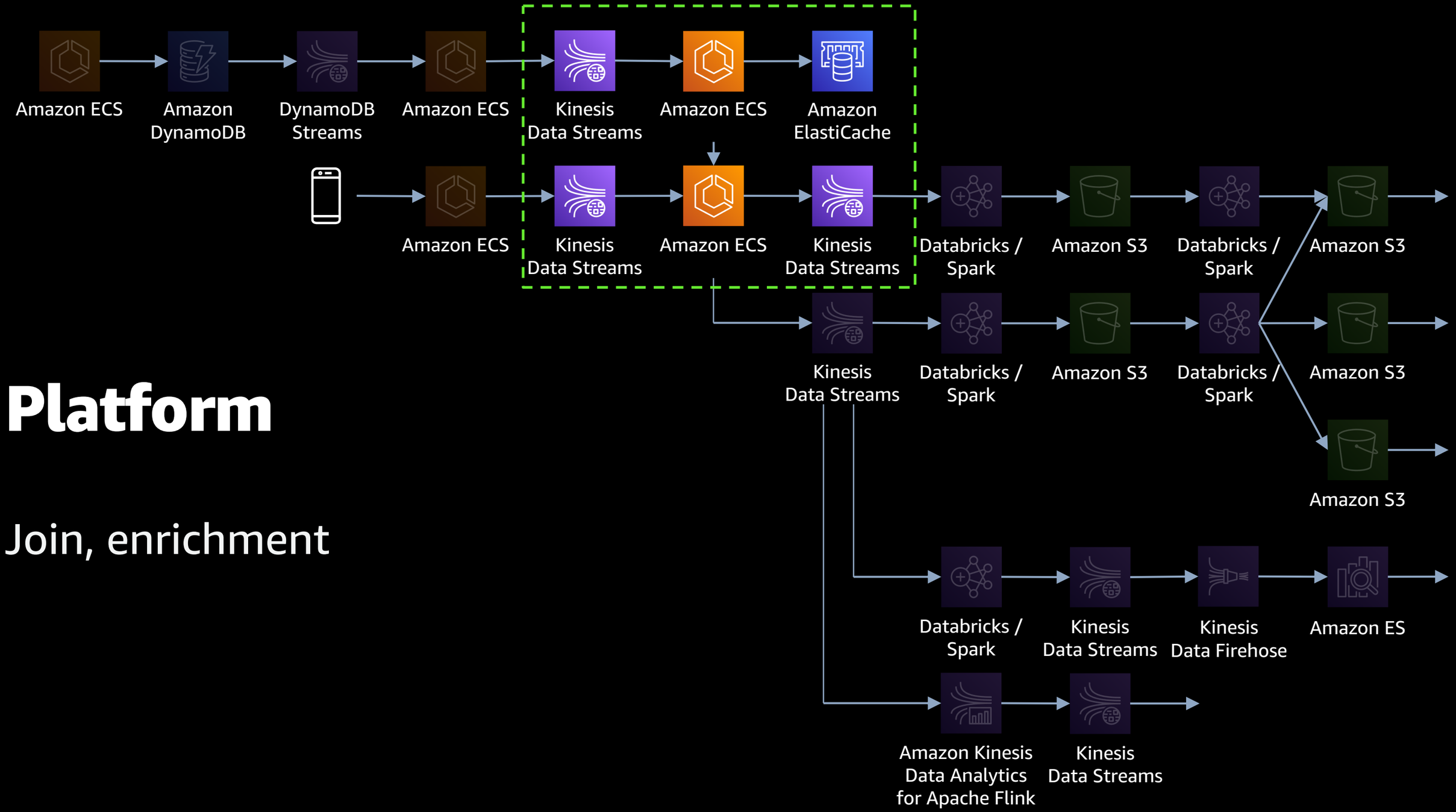
Platform

Reliable domain events



Platform

Validation, routing, filtering

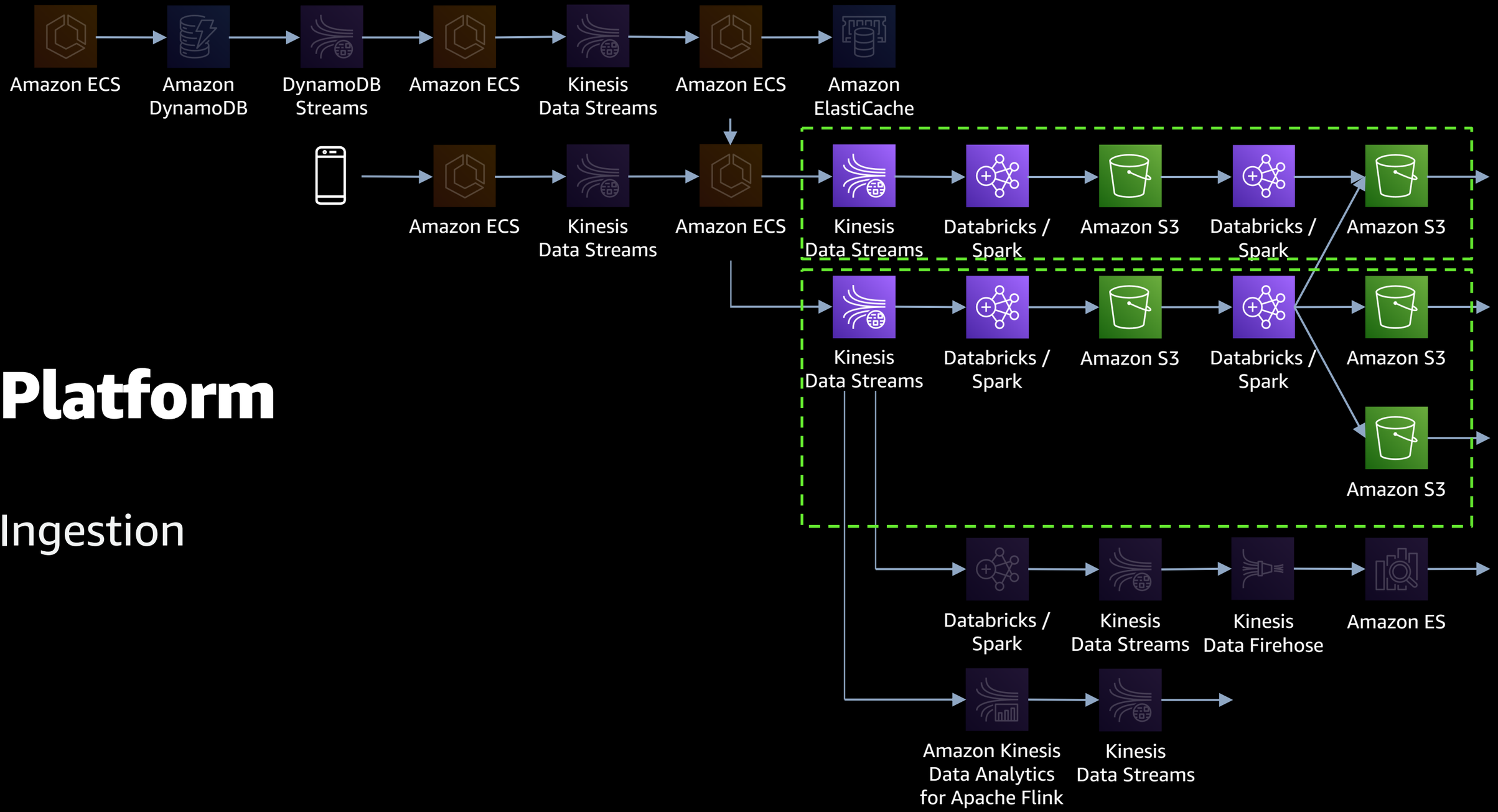


Platform

Join, enrichment

Platform

Ingestion



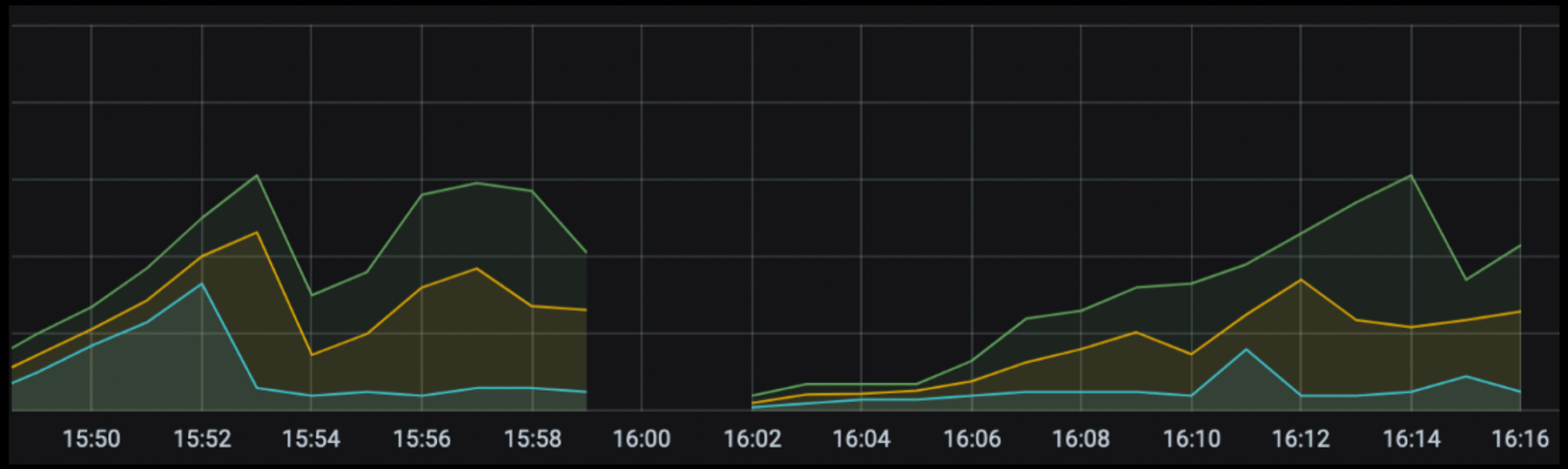
Platform

Streaming application maturity

- Architecture patterns
- Automated testing
- Performance testing and management
- Elasticity and auto scaling
- Deployment
- Observability, alerting
- Reliability and resilience
- Operations simplicity
- Multi-region replication and failover

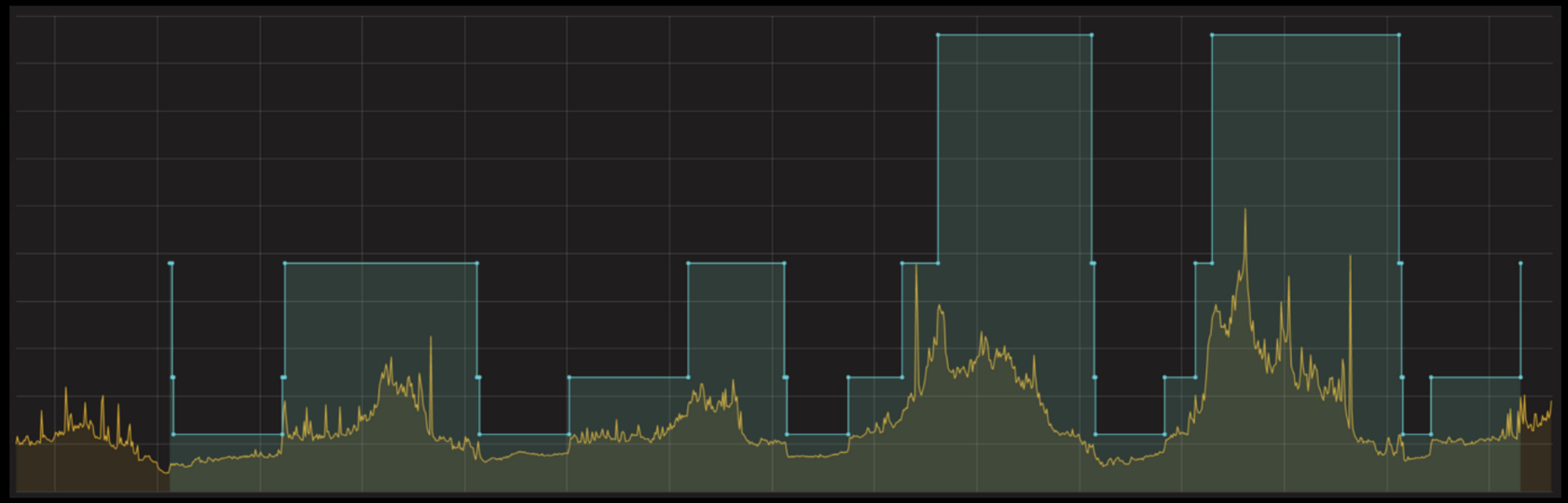
- Data lineage
- Self-healing
- Distributed tracing
- Cost efficiency
- Discoverability
- Traffic routing
- Guarantees
- Streaming as a service platform
- Etc.

Platform



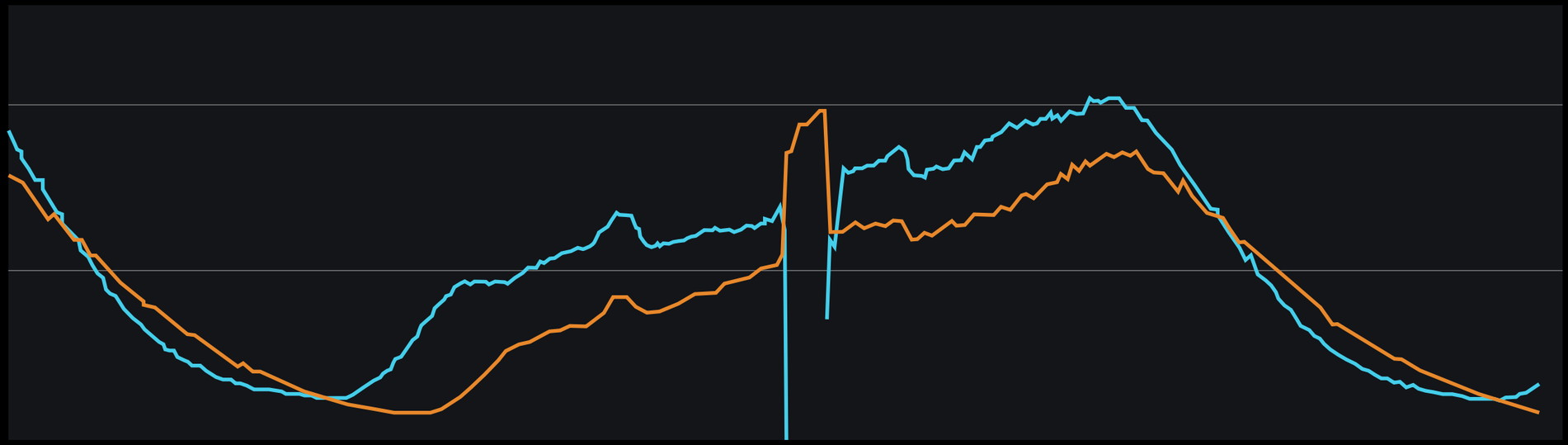
- Configurable trade-offs
- Latency management
- Deployment patterns

Platform



- Stream elasticity
- Application elasticity
- Elasticity trade-offs

Platform



- Reliability
- Delivery semantics
- End-to-end management
- Failure scenarios

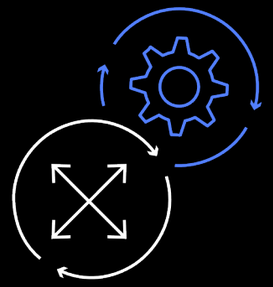
Culture



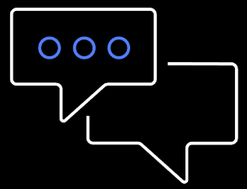
AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Culture



Training



Collaboration



Tooling



Ease of use



Integrations

Conclusion

- Data-driven organization and data democracy
- Ubiquitous data
- Streaming data platform
- Culture and tools
- Build on top of Amazon Kinesis

Credits and resources

Resources

- Disney Technology Blog – https://medium.com/disney-streaming
- Delivering data in real-time via auto scaling Kinesis streams – https://medium.com/disney-streaming/delivering-data-in-real-time-via-auto-scaling-kinesis-streams-72a0236b2cd9
- Testing asynchronous pipelines with fs2 and weaver-test – https://medium.com/disney-streaming/testing-asynchronous-pipelines-with-fs2-and-weaver-test-f0ffd37676d
- Open source project weaver-test – https://github.com/disneystreaming/weaver-test/

Credits

- Tom LeRoux
- Christian Villoslada
- Petr Zapletal
- Nick Burkard
- Matt Jankowski
- Ben Morris
- Jess Geddes
- Daniel Spiewak
- Diego Pineda
- Eric Meisel
- Anthony Garo
- Benoit Louy
- Mark Harrison
- Evan Kaplan
- Olivier Melois
- Rekha Bachwani
- User Services team
- Subscription team
- Streaming Data Platform team
- Data Engineering team
- API Services team
- Data Governance & Instrumentation team
- Experimentation team
- And the whole Disney+ team!

Thank you!

Martin Zapletal

Twitter @zapletal_martin
LinkedIn martinzapletal

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Please complete the session survey

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.