AWS re:Invent

CMP320

How Dropbox leverages hybrid cloud for scale and innovation

Anuj Dewangan
Senior Solutions Architect
Amazon Web Services

Andrew Fong
Head of Infrastructure
Dropbox

AWS re:Invent

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

aws





Change 2

re:Invent Presentations due!

October

Change 1

It’s a brand new presentation!

December 1

December 3

December 4

Agenda

Why do customers need Hybrid Cloud?

What have customers been asking for?

How does AWS help?

The Dropbox-AWS Hybrid Cloud

Hybrid cloud use cases

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Support applications need to run on-premises



Traditional enterprise applications

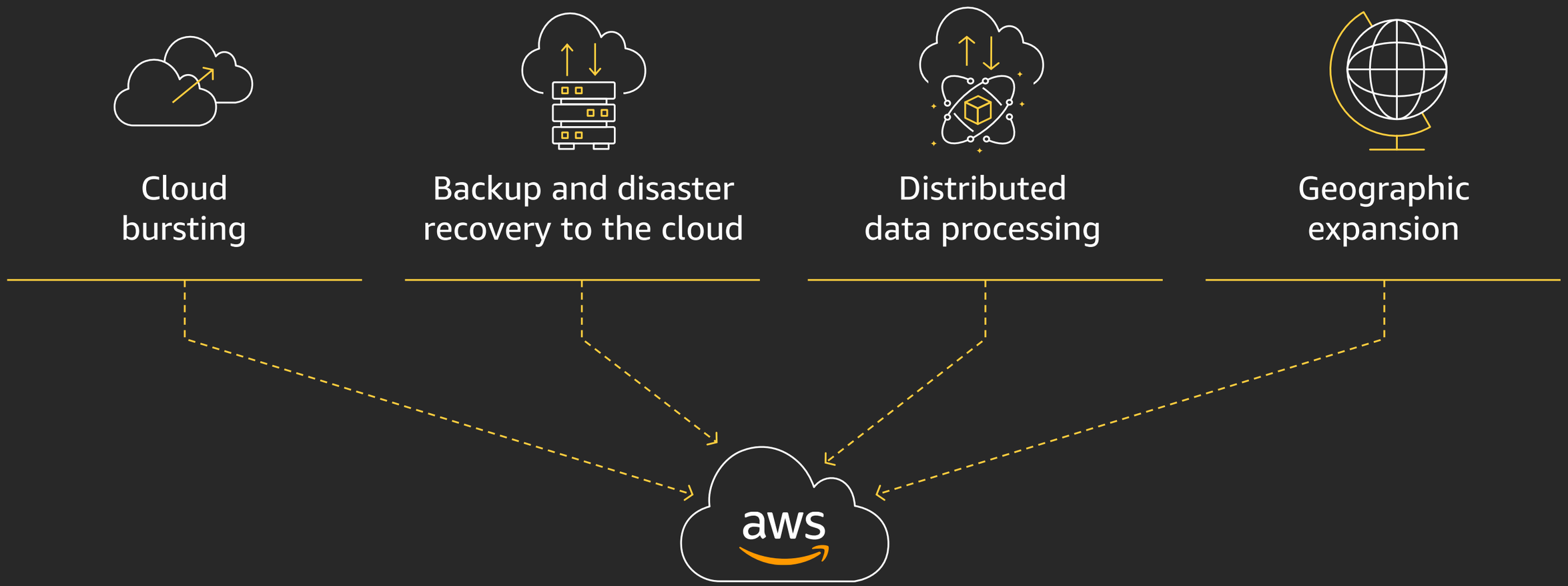


Low-latency applications



Local data processing applications

Data center extension

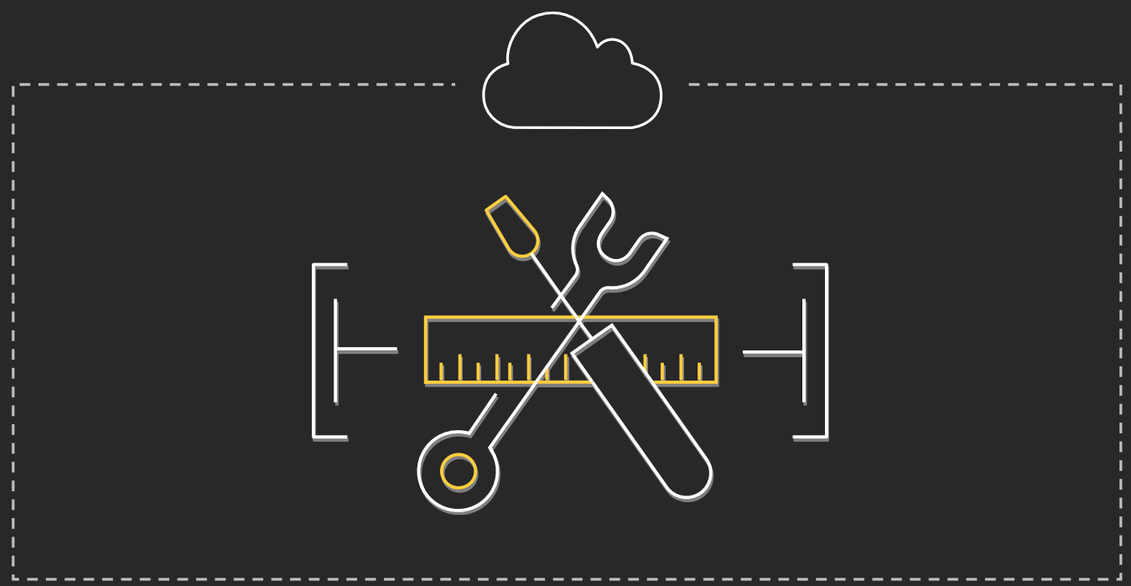


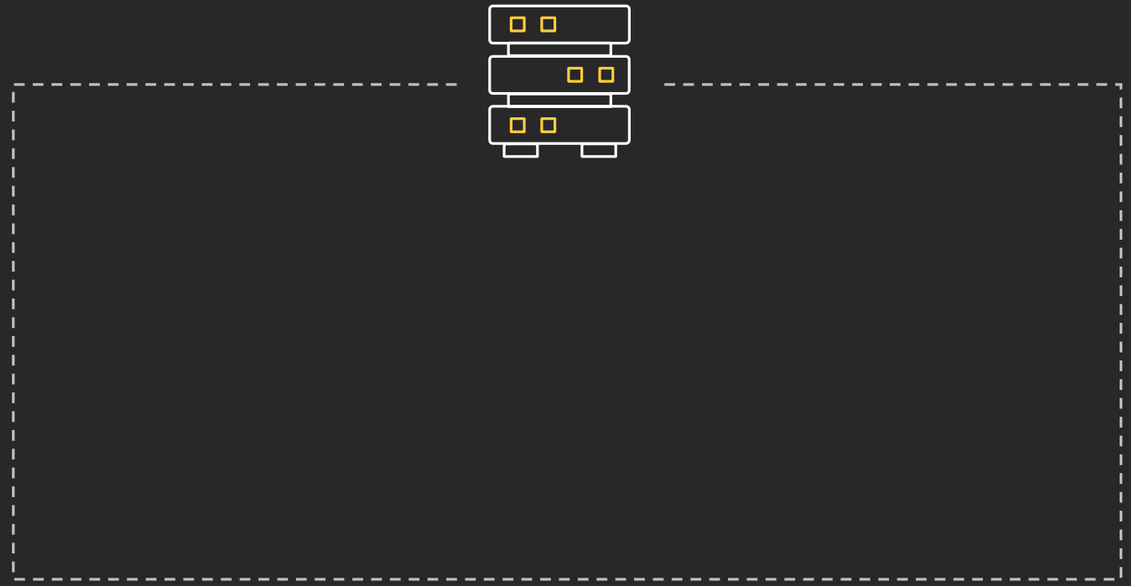
Applications migrating to cloud



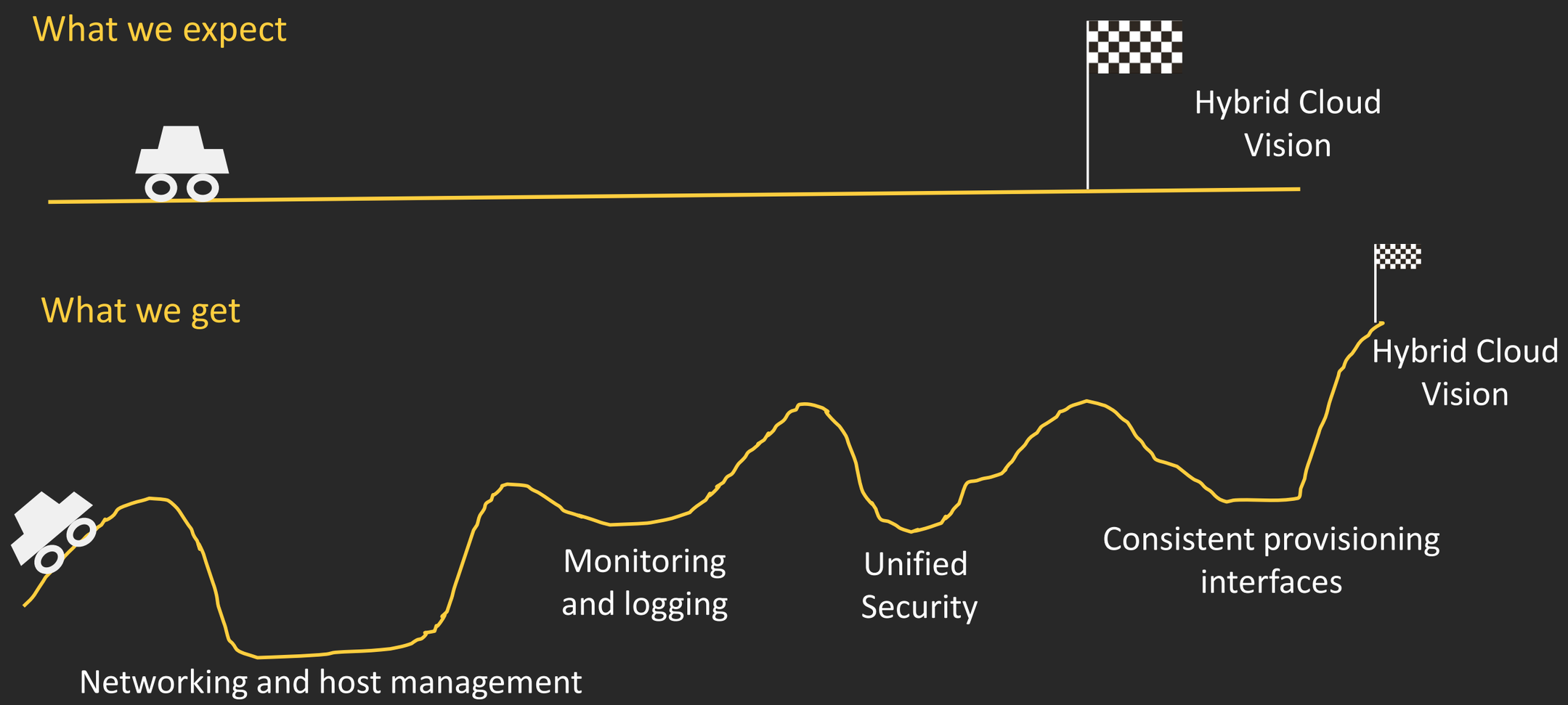
- Migration to the cloud can be a multiyear process for large enterprises
- Want consistent management and operations across the cloud and on-premises environments during the transition

Cloud services on-premises





Journey to the hybrid cloud…

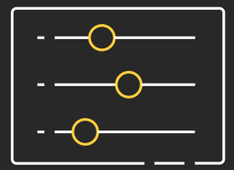


Building a hybrid cloud is hard!

Hybrid architecture goals



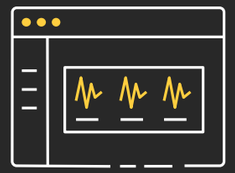
Operational consistency



Simple to control, manage, secure



Build once, deploy anywhere

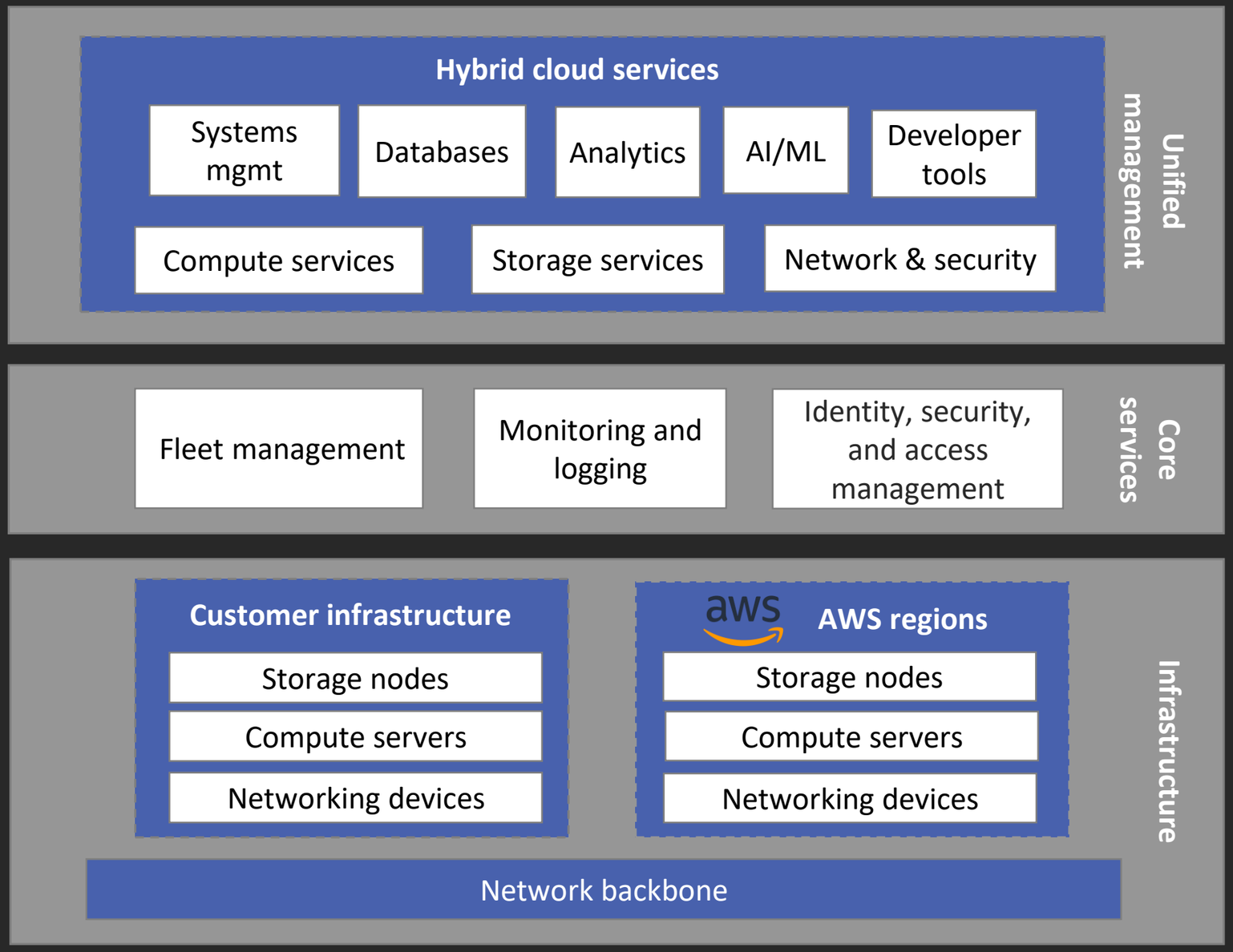


Enterprise-class app SLA



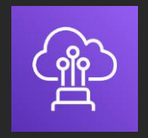
Existing skill sets and tools

Operations and management framework

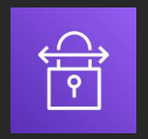


Family of hybrid services

Network & Content Delivery



AWS Direct Connect



AWS Site-to-Site VPN



Elastic Load Balancing



Amazon CloudFront



Amazon Route 53

Device Management

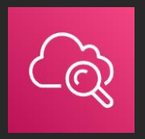


AWS Systems Manager



AWS OpsWorks

Monitoring and logging



Amazon CloudWatch

Security, Identity and Compliance



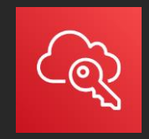
AWS Identity and Access Management



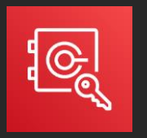
AWS Directory Service



AWS Key Management Service



AWS Single Sign-On



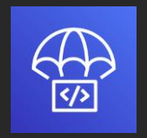
AWS CloudHSM

Application Integration



AWS Step Functions

Developer Tools



AWS CodeDeploy

Family of hybrid services

Compute



Amazon Elastic Kubernetes Service

Storage



AWS Storage Gateway



AWS Snowmobile

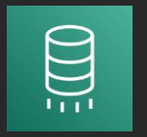


AWS Backup

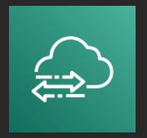


AWS Snowball Edge

Migration & Transfer



AWS Database Migration Service



AWS DataSync



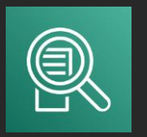
AWS Server Migration Service



AWS Transfer for SFTP



AWS Snowball



AWS Application Discovery Service



AWS Migration Hub



AWS Snowball Edge



AWS Snowmobile

Same AWS infrastructure & services on-premises

AWS Outposts



AWS-designed, Nitro infrastructure for the *same* security and performance

Automatically managed and updated as part of AWS Regions

Single point of service and support

- Same infrastructure
- Same programming interface
- Same operations & management
- Use existing skill set

Available in two options:

- Native AWS
- VMware Cloud on AWS

VMware solution for on-premises & cloud

VMware Cloud on AWS



Jointly engineered solution delivers the best of VMware and AWS for customers

VMware SDDC technologies you know and trust delivered as a service on AWS

- Simple and consistent operations
- Flexible consumption with elastic resources
- Direct access to AWS native services

Dropbox–AWS hybrid cloud

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Dropbox today

We're the world's first smart workspace that helps people and teams

We've created a single organized place for all your content to help you stay focused on your most important work and in sync with your teammates

We have a long runway for growth

We have a rare combination of virality, scale, and cash generation

Dropbox-AWS hybrid infrastructure

Why Hybrid:



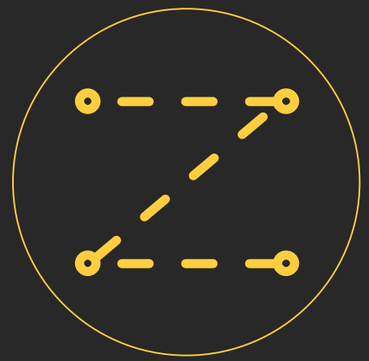
Scale

Powers Dropbox’s massive scale and global expansion



Innovation

Tightly coupled hardware and software innovations



Agility

Faster product development and flexibility in making product changes

Dropbox-AWS hybrid infrastructure provides scale

The hybrid infrastructure:

- Provides the technology resources to power the massive scale of Dropbox application
- Powers global expansion
- Enables economies of scale

3EB+

Exabytes of data
(3,000,000,000,000 GB)

5B+

Connections to
shared content

600M+

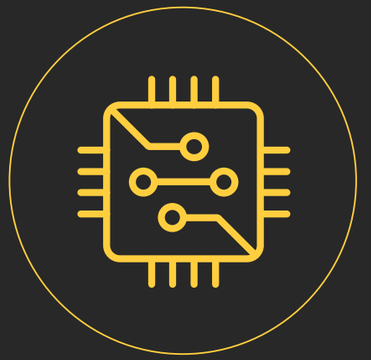
Registered users

Dropbox-AWS hybrid infrastructure enables innovation



Software

- Innovative solutions leveraging user behavior unique to Dropbox
- Strong focus on continuously optimizing the infrastructure



Hardware

- Close partnership with AWS and supply base
- Pushing the envelope through technology adoption

Dropbox-AWS hybrid infrastructure enables agility



Faster time to market

- Product launches
- International expansion



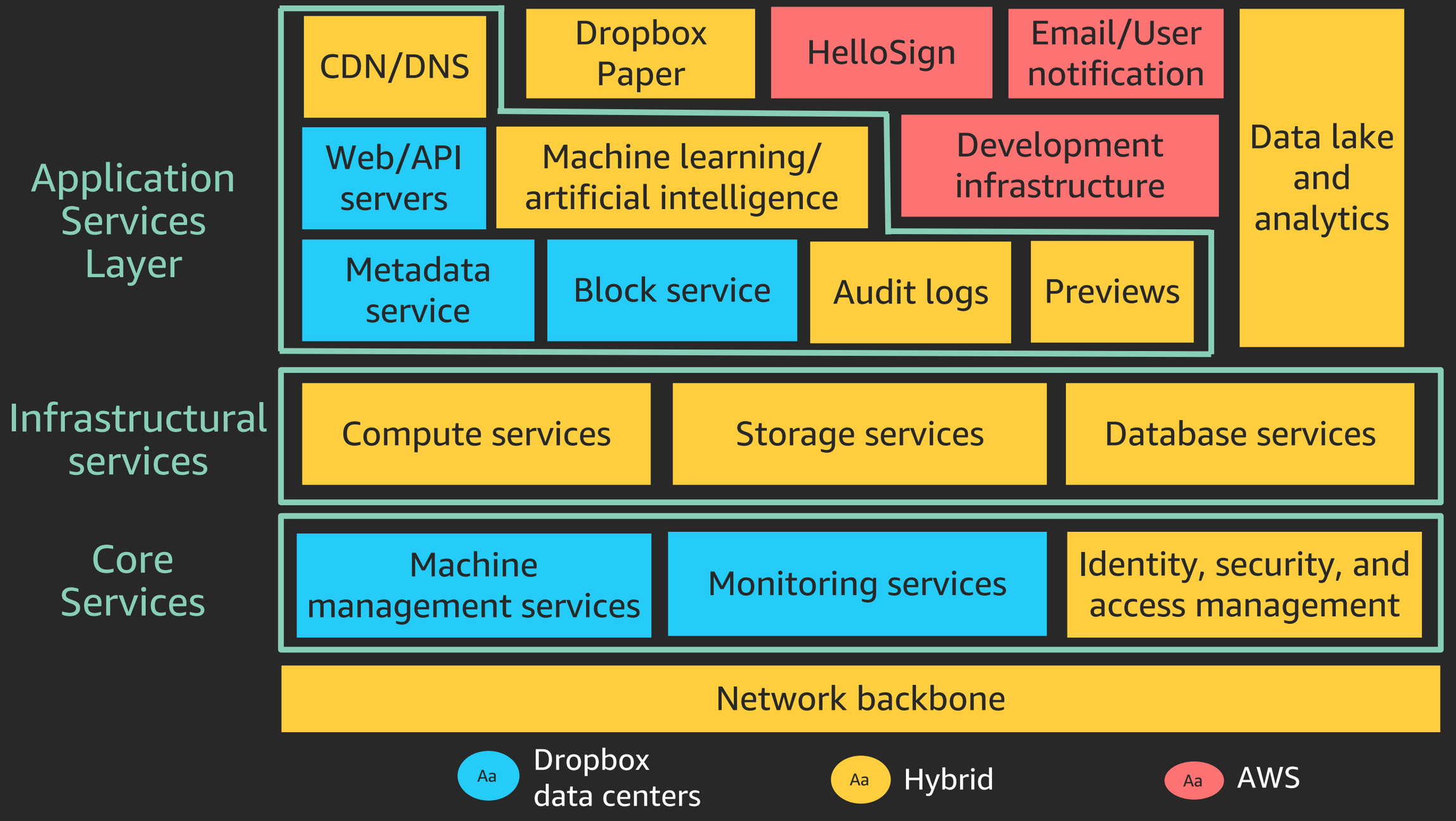
Flexibility

- Engineering partnership between Dropbox and AWS to support innovation

Hybrid workloads

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Dropbox software stack



Hybrid cloud strategy



Leverage scale, agility,
& innovation from AWS

Low latency to our
application services layer

Previews: Distributed data processing

What is Previews?

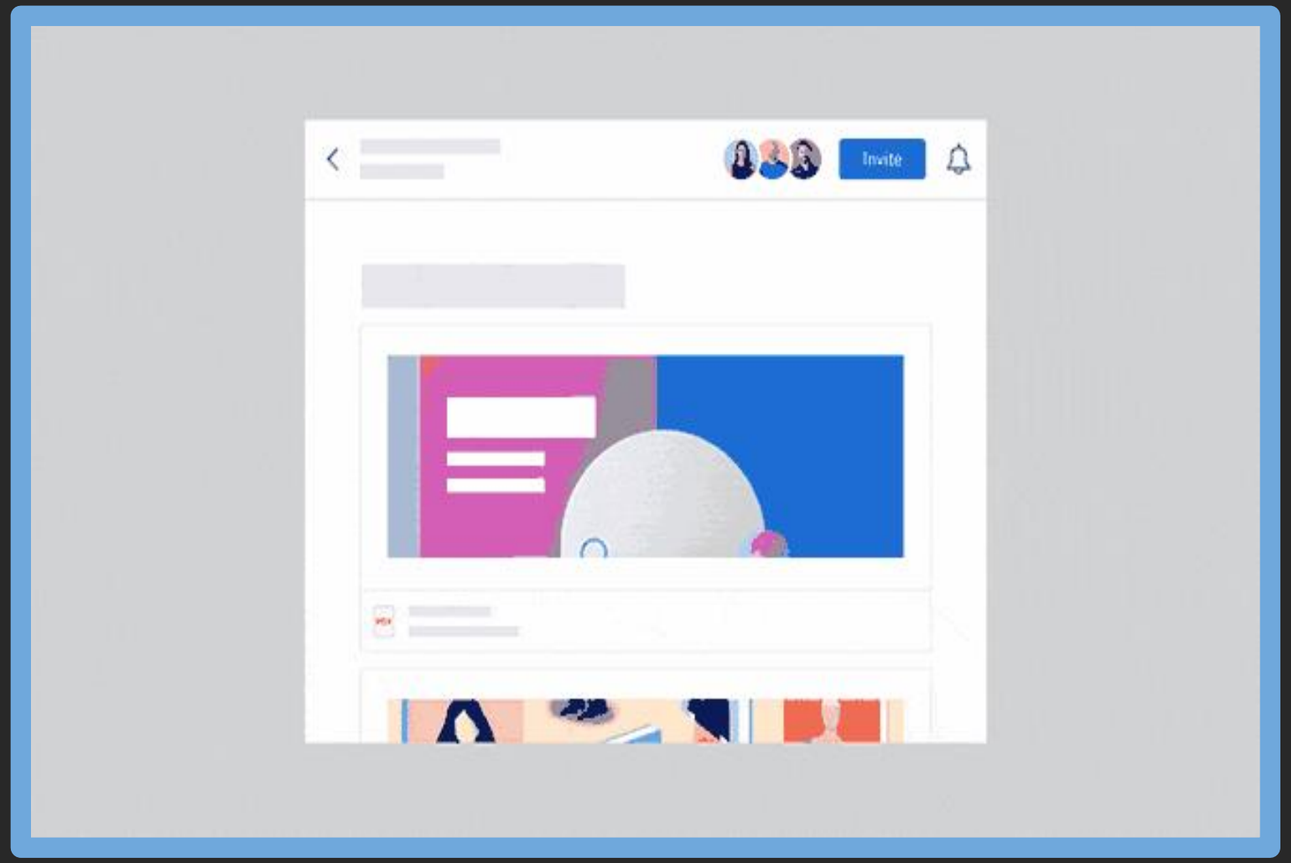
- Rich Microsoft document previews
- Contained VM environment to host Microsoft Windows OS and Office built on i3.metal

How does it work?

- Microsoft user files sent to AWS for conversion to PDF
- PDFs are sent back to serving path for previews
- 14440 vCPUs at steady state

Why is Previews hybrid?

- User files sourced from on-premises MagicPocket
- Innovation of Amazon EC2 Nitro system provides high performance and nested virtualization for high VM density



Audit Logs: Distributed data processing

What is Audit Log?

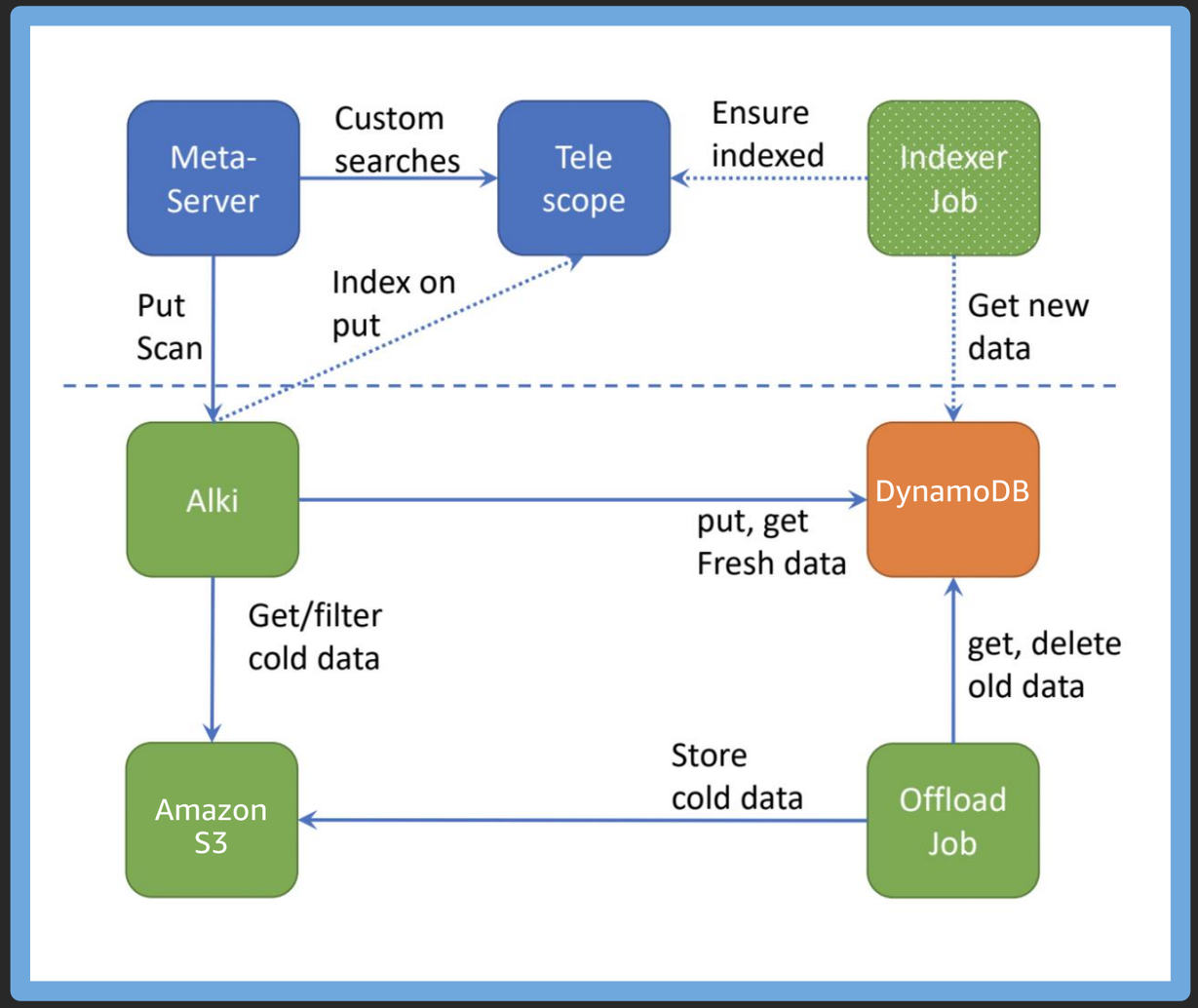
- A cost-effective key-value store for serving audit records to customers

How does it work?

- Audit events read/write from AWS
- Amazon DynamoDB for hot storage and Amazon Simple Storage Service (Amazon S3) for cold storage
- 100TB+ migrated from on-premises; ~20K writes/sec

Why is Audit Logs hybrid?

- Audit records ingested from on-premises servers
- Leverage DynamoDB's low-latency, high throughput for reads and writes of audit records



Analytics: Distributed data processing

Why is Analytics important?

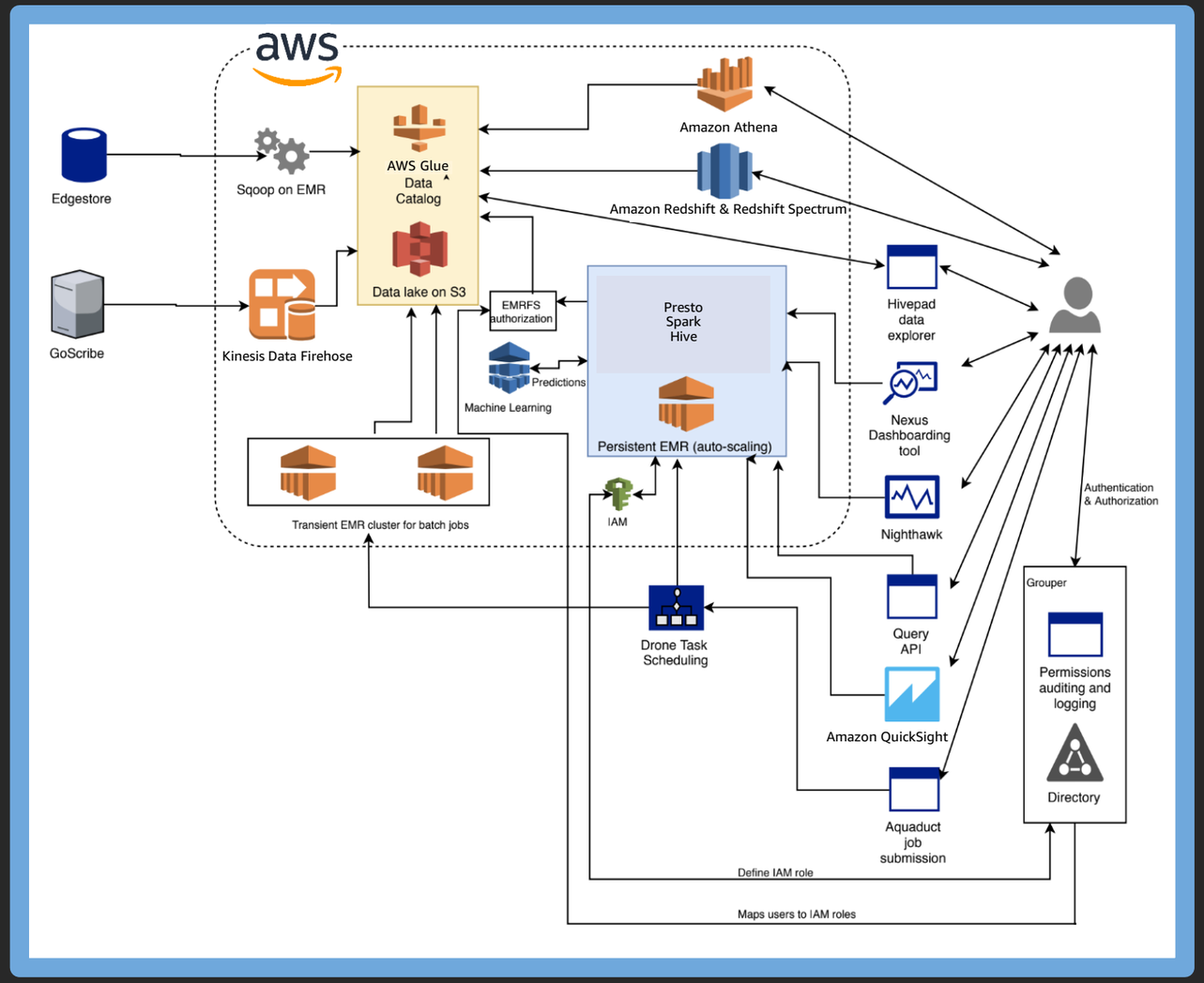
- Metrics for business intelligence and application usage

How does it work?

- Data ingested from on-premises sources to Amazon S3-based data lake
- Hadoop/Hive for batch and interactive queries on AWS
- 22+ PB migrated; 100K+ jobs/day

Why distributed data processing for analytics?

- Data pipeline is hybrid with on-premises sources
- Leverage large portfolio and pace of innovation of AWS in the analytics space



Machine learning—distributed data processing

Why is machine learning important for DBX?

- Dropbox smart workspaces (machine intelligence)—be organized, stay focused, and get in sync

How is AWS leveraged for machine intelligence?

- DBXLearn—integration of Dropbox model training infrastructure with Amazon SageMaker
  Built a hybrid pipeline for training and inference
  Training on AWS and inference in the serving path on-premises
  Amazon SageMaker provides:
  - Scalable and elastic compute
  - An interactive environment
  - Ease of integration w/ a variety of ML frameworks

  MI/ML team investigating use of GPU instances for training models

- Amazon Textract and Amazon Rekognition image being evaluated for document analysis and image labeling

Cloud bursting & disaster recovery

Cloud bursting

- Trampoline—*Extending MagicPocket functionality to include Amazon S3*
- Storage bursting to scale on-premises storage capacity

Disaster recovery

- Backup and restore of databases & configurations

How are we leveraging AWS?

- Cloud bursting for user file storage
- Backup of Edgestore and configurations to Amazon S3



Dropbox Paper: Agility and elasticity

What is Dropbox Paper?

- Dropbox Paper is a collaboration tool for teams to create, coordinate, and communicate around content

How are we leveraging AWS?

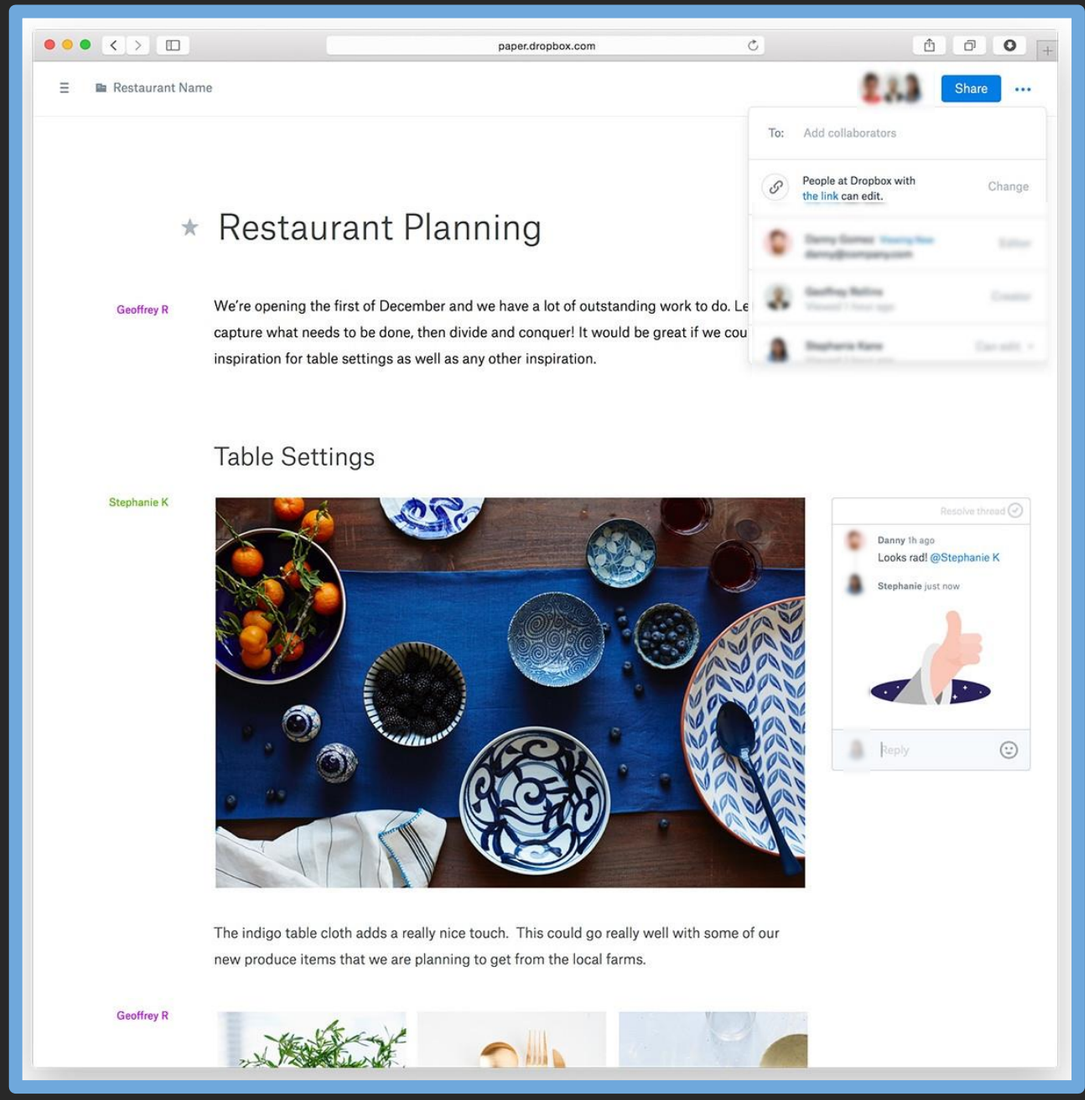
- Amazon ElastiCache clustered Redis
- Services: Amazon RDS for MySQL, EC2
- Amazon Simple Queue Service (Amazon SQS)





Why hybrid?

- Leveraging on-premises machine management services
- Leveraging the agility of AWS and managed services



Global expansion

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Dropbox-AWS global hybrid infrastructure



Global expansion

AWS serving markets

- Singapore
- Hong Kong
- Frankfurt
- Tokyo
- Sydney

APJ—Global Expansion:

- What is APJ @ Dropbox?
- Why was it important for DBX?
- How was AWS critical for success?

Hybrid Cloud with AWS has unleashed agility, scale and innovation for Dropbox, and it can for you too

Thank you!

Anuj Dewangan
anujddew@amazon.com

Andrew Fong
fong@dropbox.com

AWS re:Invent

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

aws



Please complete the session survey in the mobile app.

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.