
AWS
re:Invent

DAT308

Migrating databases to Amazon DocumentDB (with MongoDB compatibility)

Meet Bhagdev
Senior Product Manager
AWS

Joseph Blair
Senior Software Engineer
The Washington Post

AWS re:Invent

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

Agenda

- What is Amazon DocumentDB?
- Why customers are migrating to Amazon DocumentDB
- Common sources of migration
- Migration methods
- Best practices
- Demo: Minimal downtime migration to Amazon DocumentDB
- Customer spotlight – The Washington Post



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

What is Amazon DocumentDB?



AWS re:Invent

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

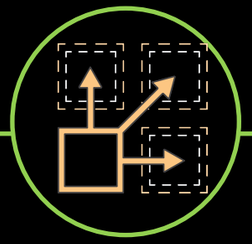
Amazon DocumentDB (with MongoDB compatibility)

Fully managed



Managed by AWS: no hardware provisioning; auto patching, quick setup, secure and automatic backups

Scalable



Separation of compute and storage enables both layers to scale independently; scale out to 15 read replicas in minutes

MongoDB compatible



Compatible with MongoDB 3.6 and 4.0; use the same SDKs, tools, and applications with Amazon DocumentDB

Fully managed and scalable document database service that supports MongoDB workloads

Why customers are migrating to Amazon DocumentDB



Purpose-built, operational database for JSON data



Flexible schema



Cloud-native architecture with autoscaling storage and I/Os



Low total cost of ownership and consumption-based billing

Common sources of migration



Relational



Key-value



Document



Search

Migration tools

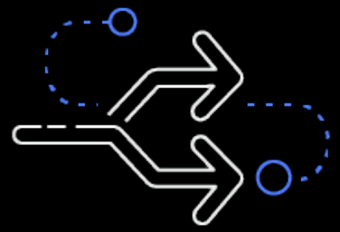


AWS Database Migration Service (AWS DMS) – easily and securely migrates and / or replicates your databases and warehouses to AWS with minimal downtime



Native Mongo tools and utilities:

- **mongodump** – creates a binary (BSON) backup of a MongoDB database
- **mongorestore** – enables you to restore a binary (BSON) backup of a database that was created with the mongodump utility



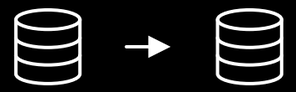
Custom solutions:

- AWS Glue to extract and transform data from a source and load it into Amazon DocumentDB
- Amazon Managed Streaming for Apache Kafka to move streaming data from a given source to Amazon DocumentDB as the sync

Migration methods



Offline migration using mongodump and mongorestore
Get started quickly, great for POCs



Online migration using AWS DMS
Near-zero-downtime migration



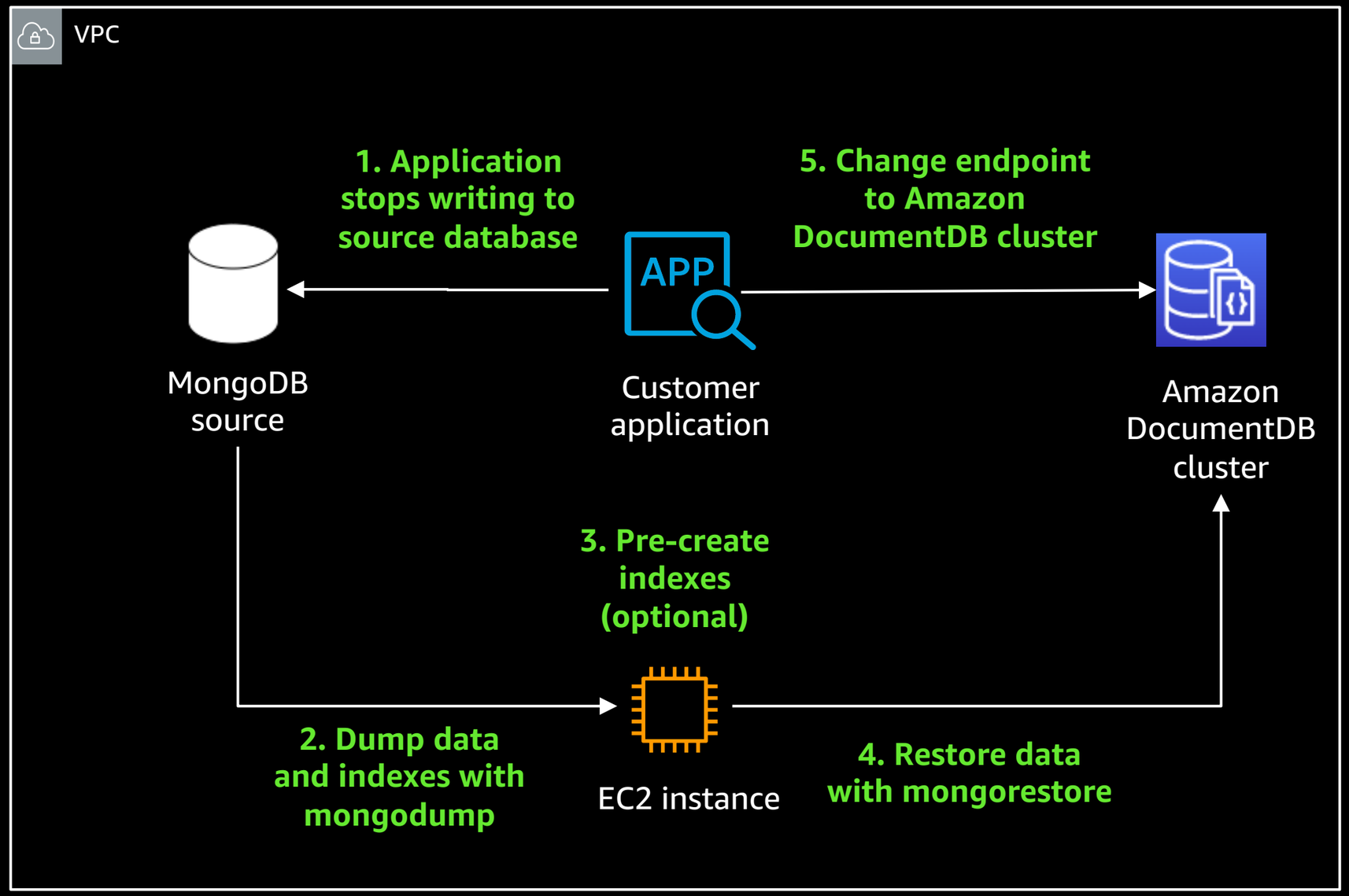
Sources for both relational and document databases

You can use AWS Database Migration Service (AWS DMS) **free** (for 6 months) to easily migrate to Amazon DocumentDB

All options support migrations from on premises and Amazon EC2, for both replica sets and sharded clusters

https://docs.aws.amazon.com/documentdb/latest/developerguide/docdb-migration.html

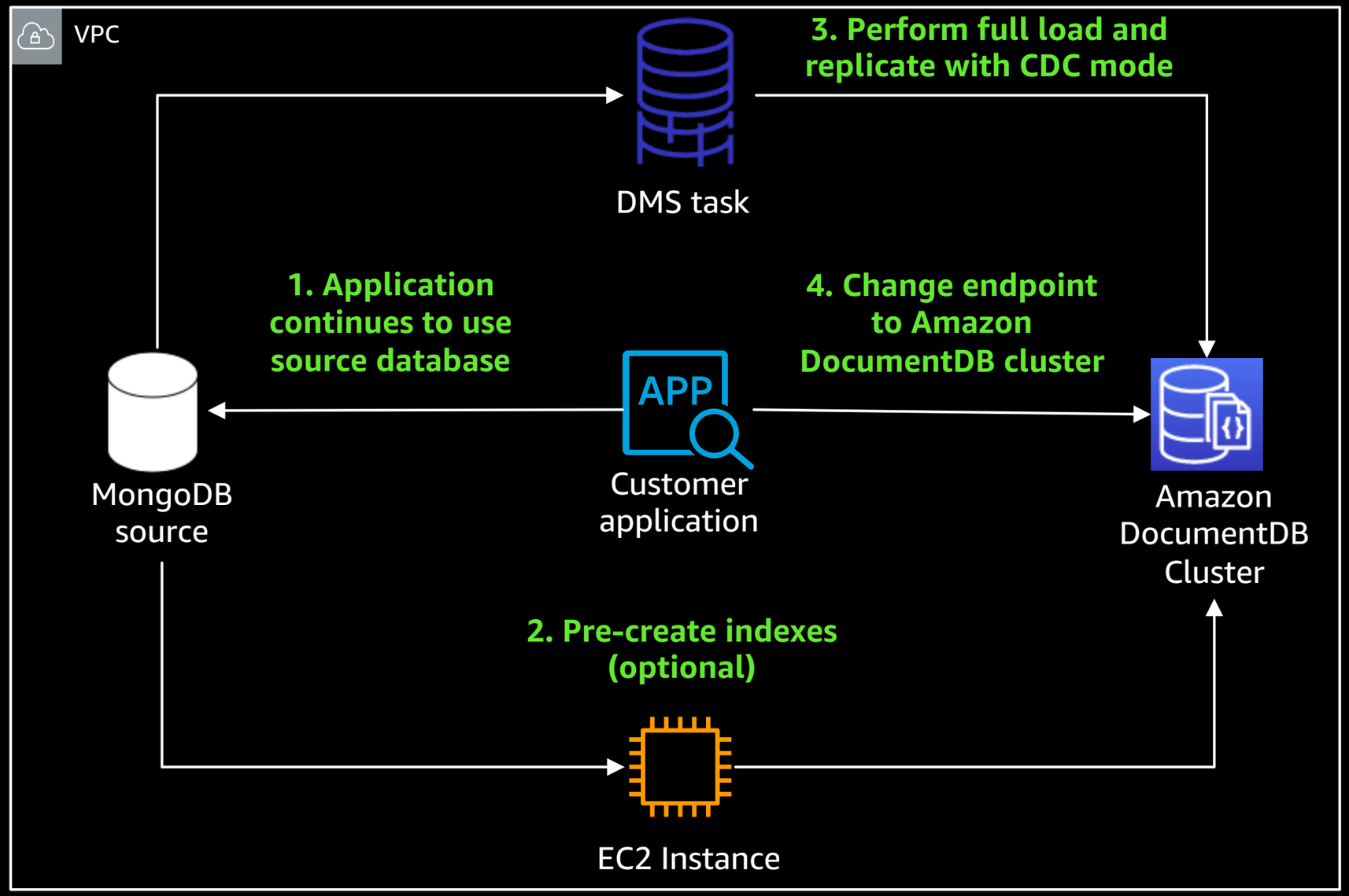
Migration approach: Offline



Highest downtime

Simplest approach

Migration approach: Online



Minimal downtime

Medium complexity

Migrations best practices

1. Verify API compatibility

2. Size your cluster based on your needs

3. Estimate costs prior to the migration

4. Create indexes prior to the migration
   - Use the Amazon DocumentDB index tool (on GitHub) to migrate indexes

Migrations best practices

5. Size your oplog to capture all change events during the migration

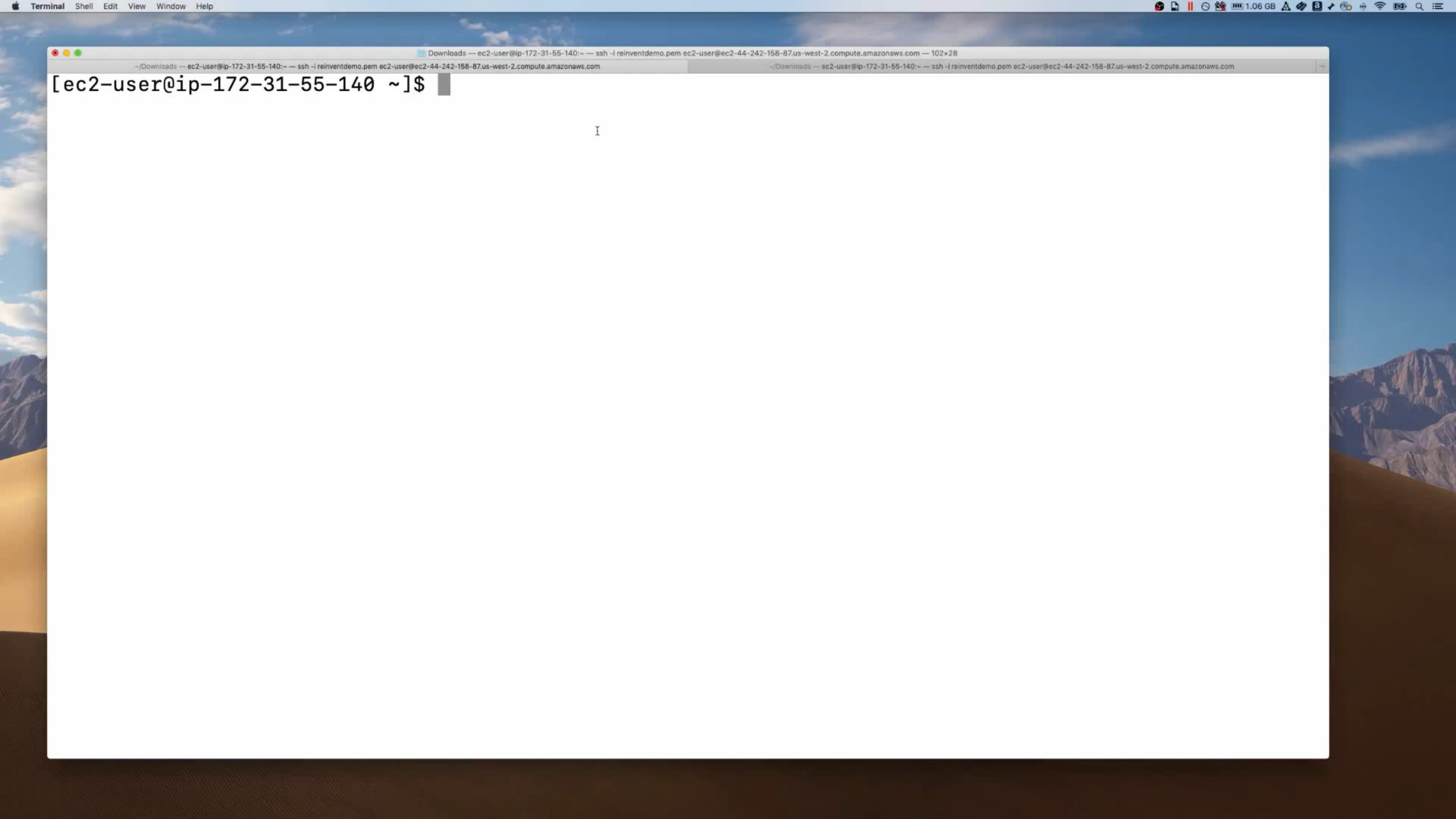
6. Scale up your primary instance at the time of migration to speed up the full load

7. For larger workloads tune the parameters to speed up the migration
   - AWS DMS – specify segments to parallelize the full load for the migration
   - mongorestore – set numInsertionWorkersPerCollection to match the number of vCPUs on your Amazon DocumentDB cluster’s primary instance

8. No substitute for testing

- What Is Amazon DocumentDB?
- Get Started Guide
- Quick Start using AWS CloudFormation
- Best Practices
- **Functional Differences with MongoDB**
- Querying
- Supported MongoDB APIs, Operations, and Data Types
- Migrating to Amazon DocumentDB
- Security
- Backing Up and Restoring
- Managing Amazon DocumentDB
- Monitoring Amazon DocumentDB
- Developing with Amazon DocumentDB
- Quotas and Limits
- Troubleshooting
- Resource Management API Reference
- Document History

https://docs.aws.amazon.com/documentdb/latest/developerguide/what-is.html


Terminal Shell Edit View Window Help
1.06 GB
Downloads — ec2-user@ip-172-31-55-140:~ — ssh -i reinventdemo.pem ec2-user@ec2-44-242-158-87.us-west-2.compute.amazonaws.com — 102×28
~/Downloads — ec2-user@ip-172-31-55-140:~ — ssh -i reinventdemo.pem ec2-user@ec2-44-242-158-87.us-west-2.compute.amazonaws.com
~/Downloads — ec2-user@ip-172-31-55-140:~ — ssh -i reinventdemo.pem ec2-user@ec2-44-242-158-87.us-west-2.compute.amazonaws.com
[ec2-user@ip-172-31-55-140 ~]$

Customer spotlight: Arc Publishing by The Washington Post



AWS re:Invent

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

Who are we?

arc publishing

- Arc Publishing is a digital experience platform built by The Washington Post
- Provides an integrated ecosystem of cloud-based tools that helps organizations create and distribute content, drive digital commerce, and deliver powerful experiences to both internal and external audiences
- Today, Arc supports publishers, broadcasters, and brands around the world
- Arc is now available in AWS Marketplace

1.5 billion
Unique monthly visitors

1,200
Sites

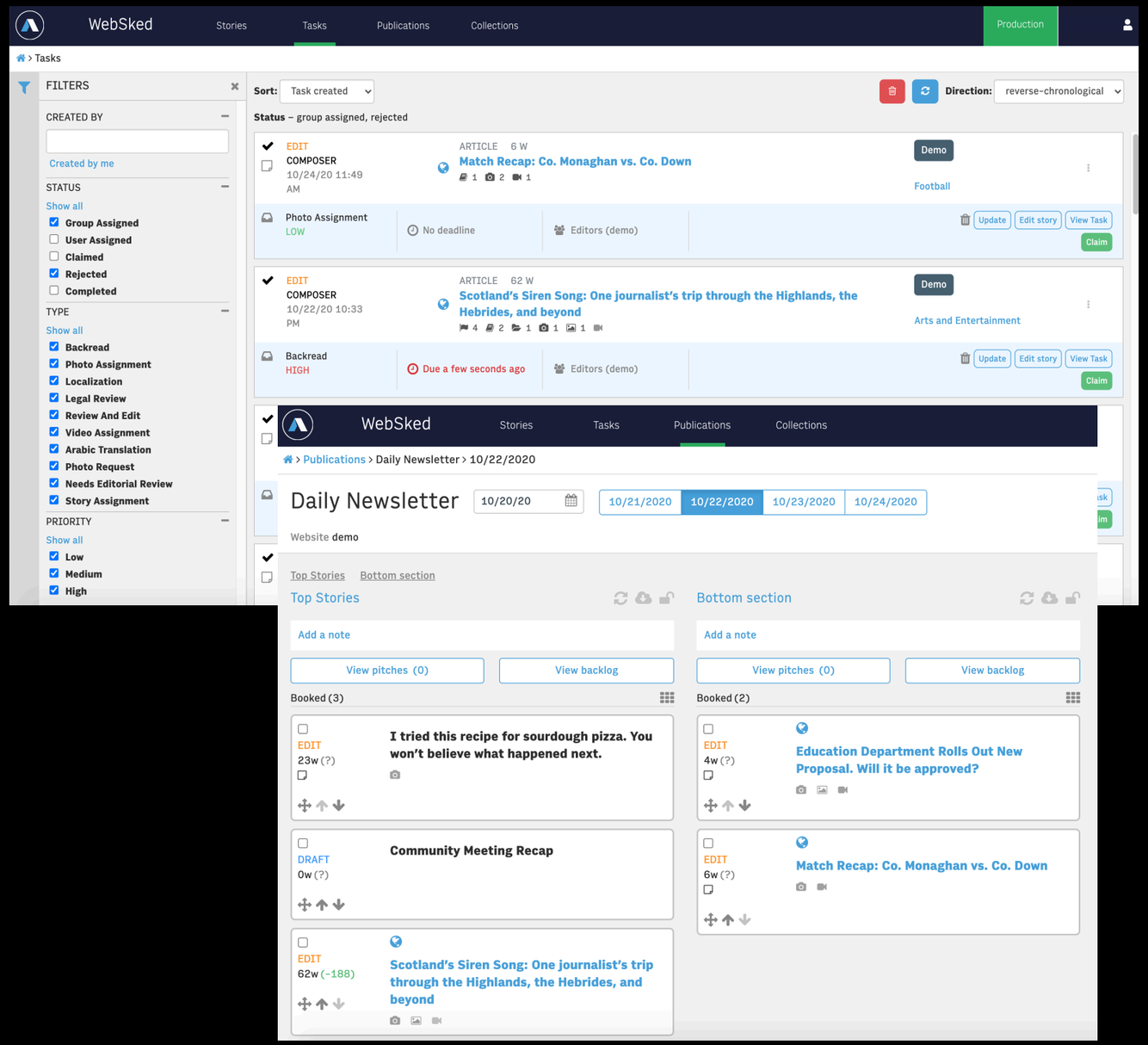
23
Countries

50 million
Commerce users

What do we use Amazon DocumentDB for?

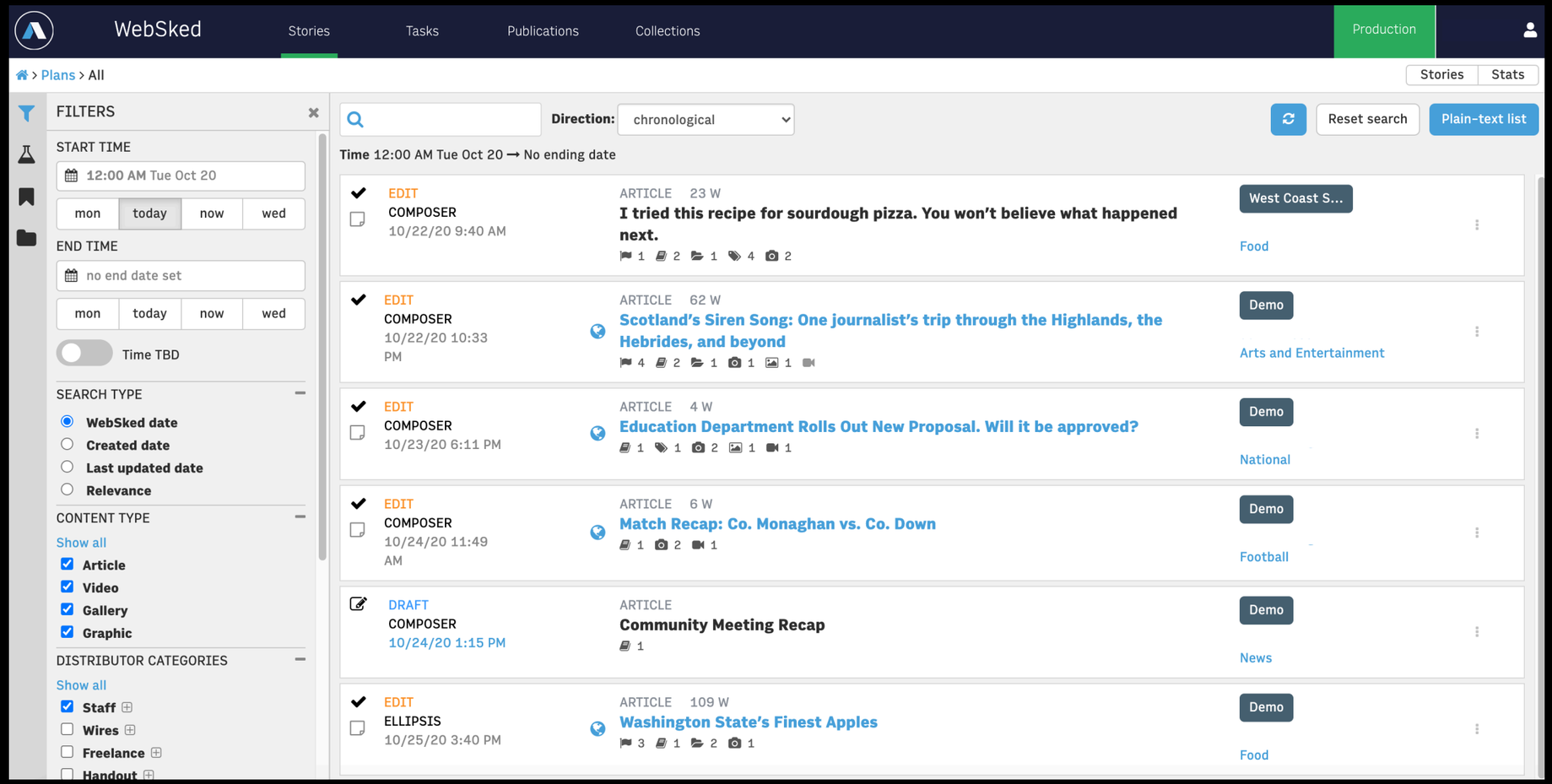
- WebSked is an editorial-facing planning, workflow and curation tool

- Amazon DocumentDB serves as the primary data store for many of WebSked's feature areas, including:
  - Workflow and task coordination for Arc content
  - Curation and budgeting of content for use in consumer-facing platforms, including both 24/7 and time-bucketed platforms
  - Configuration data such as notification subscriptions, user profile data, and user groups

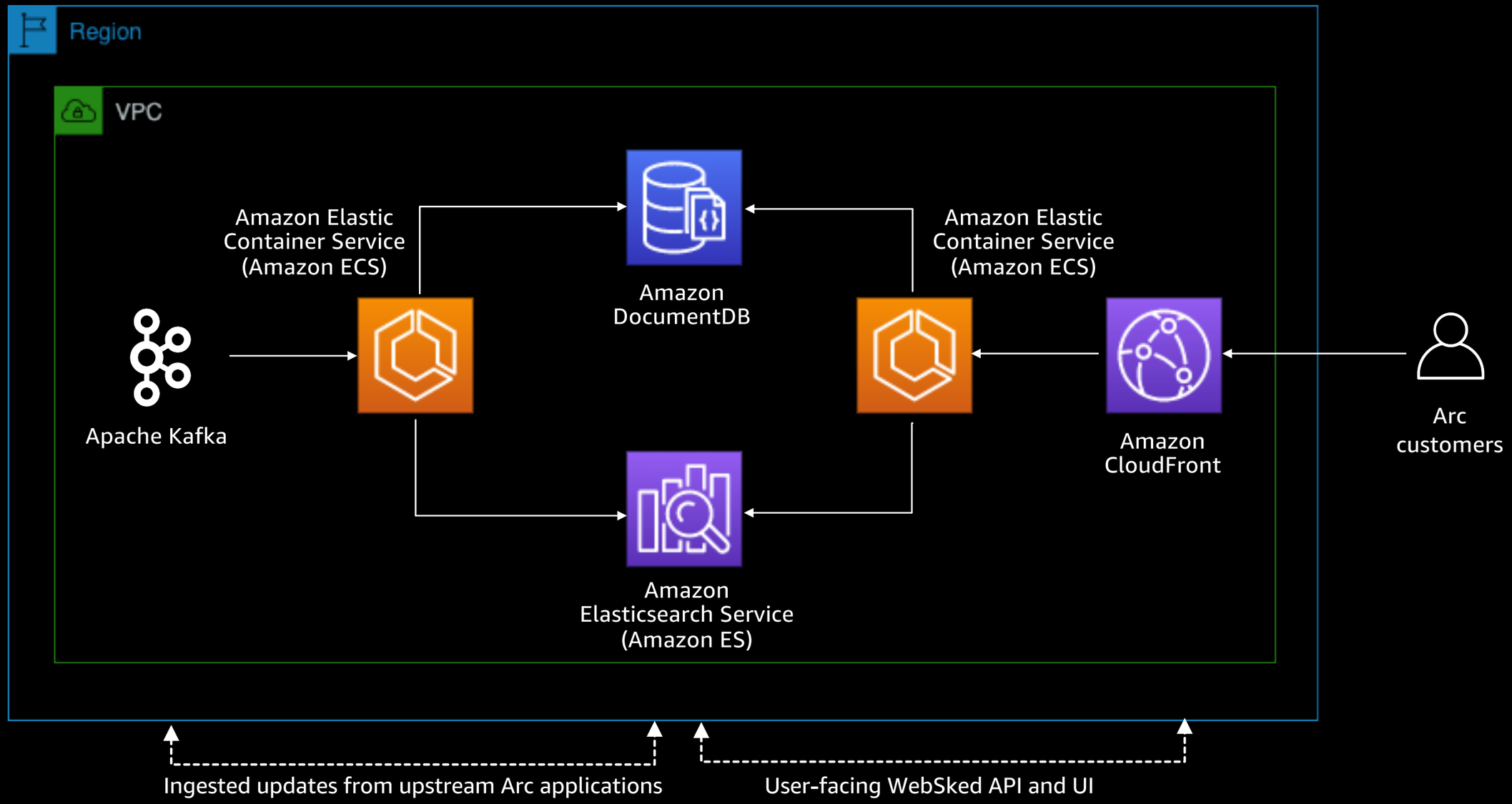


What do we use Amazon DocumentDB for?

- Amazon DocumentDB also serves as the secondary data store for editorial-facing content search, along with Amazon Elasticsearch Service



Architecture



Scale on Amazon DocumentDB

45,000

IOs per second

1,300

Connections

240

Migrations

240

Databases

How we migrated to Amazon DocumentDB

- Used AWS Database Migration Service (AWS DMS) with change data capture (CDC) for minimal downtime
- Opted to pause ingestion process to avoid conflicts between change data capture from the old cluster and writes on the new cluster
- Decided to do this in a maintenance window, with deployments staggered over a few weeks
- Actual cutover took only as long as the two ECS deployments (under 10 minutes each)

Benefits of migrating to Amazon DocumentDB

- Offload database management tasks such as patching, security, backup, HA, monitoring, and scaling of our MongoDB databases to AWS
- Enables us to focus on delivering customer value
- Well integrated with our infrastructure-as-code best practices
  - Amazon DocumentDB clusters are a native AWS CloudFormation resource
- Access to AWS support and other benefits of an existing relationship with AWS

Thank you!



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.