Introduction to Amazon DocumentDB
(with MongoDB compatibility)

Fast, scalable, and fully managed MongoDB-compatible database service

Joseph Idziorek, AWS Principal Product Manager
Purpose built
The right tool for
the right job

## Data categories and common use cases

<table>
<thead>
<tr>
<th>Relational</th>
<th>Key-value</th>
<th>Document</th>
<th>In-memory</th>
<th>Graph</th>
<th>Search</th>
<th>Time-series</th>
<th>Ledger</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referential integrity, ACID transactions, schema-on-write</td>
<td>Low-latency, key lookups with high throughput and fast ingestion of data</td>
<td>Indexing and storing documents with support for query on any attribute</td>
<td>Microseconds latency, key-based queries, and specialized data structures</td>
<td>Creating and navigating data relations easily and quickly</td>
<td>Indexing and searching semistructured logs and data</td>
<td>Collect, store, and process data sequenced by time</td>
<td>Complete, immutable, and verifiable history of all changes to application data</td>
</tr>
<tr>
<td>Lift and shift, EMR, CRM, finance</td>
<td>Real-time bidding, shopping cart, social</td>
<td>Content management, personalization, mobile</td>
<td>Leaderboards, real-time analytics, caching</td>
<td>Fraud detection, social networking, recommendation engine</td>
<td>Product catalog, help and FAQs, full text</td>
<td>IoT applications, event tracking</td>
<td>Systems of record, supply chain, health care, registrations, financial</td>
</tr>
</tbody>
</table>
AWS: Purpose-built databases

Relational
- Amazon RDS
  - Aurora
- MySQL
- PostgreSQL
- SQL Server
- MariaDB

Key-value
- Amazon DynamoDB
- Memcached
- Redis

Document
- Amazon DocumentDB

In-memory
- Amazon ElastiCache
- Redis
- Memcached

Graph
- Amazon Neptune

Search
- Amazon Elasticsearch Service

Time-series
- Amazon Timestream

Ledger
- Amazon Quantum Ledger Database
Agenda
What's the plan?

What is a document database?
Introduce Amazon DocumentDB
Challenges and capabilities
Demos

{ "Hello": "Amazon DocumentDB",
  "Getting Started": "https://aws.amazon.com/documentdb/getting-started/"
}
What is a document database?
Evolution of document databases

JSON became the de facto data interchange format

Friction when converting JSON to the relational model

Object-relational mappings (ORMs) were created to help with this friction

Document databases solved the problem
Document databases

- Data is stored in JSON-like documents
- Documents map naturally to how humans model data
- Flexible schema and indexing
- Expressive query language built for documents (ad hoc queries and aggregations)

JSON documents are first-class objects of the database

```json
{
  id: 1,
  name: "sue",
  age: 26,
  email: "sue@example.com",
  promotions: ["new user", "5\%", "dog lover"],
  memberDate: 2018-2-22,
  shoppingCart: [
    {product:"abc", quantity:2, cost:19.99},
    {product:"edf", quantity:3, cost: 2.99}
  ]
}
```
Document databases help developers build applications faster and iterate quickly.
Use cases for document data

- Content management
- Mobile
- Personalization
- Catalog
- Retail and marketing
- User profiles
Use cases for document data

User profiles

```json
{
  id: 181276,
  username: "sue1942",
  name: {first: "Susan",
          last: "Benoit"},
  ExploidingSnails: {
    hi_score: 3185400,
    global_rank: 5139,
    bonus_levels: true
  },
  promotions: ["new user","5%","snail lover"]
}
```
Challenges of existing document databases

- Hard to set up
- Hard to manage
- Hard to scale
- Hard to secure
- Hard to back up
What is Amazon DocumentDB?

Fast, scalable, and fully managed MongoDB-compatible database service
Amazon DocumentDB
Fast, scalable, and fully managed MongoDB-compatible database service

**Fast**
- Millions of requests per second with millisecond latency; twice the throughput of MongoDB

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

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

**MongoDB compatible**
- Compatible with MongoDB 3.6; use the same SDKs, tools, and applications with Amazon DocumentDB
Challenges with traditional database architectures

- Not designed for the cloud
- Single monolithic architectures
  - Scale monolithically
  - Fail monolithically

- Application
  - API
  - Query processor
  - Caching
  - Logging
  - Storage
Challenges with traditional databases: Scaling

**Scenario:** Spike in traffic and you want to add additional read capacity quickly

Node 1 | Node 2 | Node 3 | Node 4
Challenges with traditional databases: Scaling

**Scenario:** Scale up to run large analytical workloads on a replica
Challenges with traditional databases: Recovery

**Scenario**: An instance experiences a failure and you want to recover quickly.

![Diagram showing replication process](image)
Amazon DocumentDB: Modern cloud-native architecture

What would you do to improve scalability and availability?

1. Decouple compute and storage
2. Distribute data in smaller partitions
3. Increase the replication of data (6x)
Amazon DocumentDB: Modern cloud-native architecture

1. Decouple compute and storage

- Compute layer
- Storage layer

API
Query processor
Caching
Logging
Storage

Scale compute
Scale storage
Amazon DocumentDB: Modern cloud-native architecture

2 Distribute data in smaller partitions

Distributed storage volume

AZ1 | AZ2 | AZ3
Amazon DocumentDB: Modern cloud-native architecture

3. Increase the replication of data (6x)
Amazon DocumentDB: Modern cloud-native architecture
Amazon DocumentDB: Scaling

**Scenario:** A spike in traffic and you want to add additional read capacity quickly

AZ1  | AZ2  | AZ3  
---|---|---
Distributed storage volume

- Adding a new availability zone (AZ3) to distribute the load.
Amazon DocumentDB: Failure recovery

**Scenario:** An instance experienced a failure and you want to recover quickly
Amazon DocumentDB: Failure recovery

**Scenario:** Six-way replication across three Availability Zones provides the ability to handle AZ + 1 failures

![Diagram showing six-way replication](image)
Demo: Getting started with Amazon DocumentDB
Fast, scalable, and fully managed MongoDB-compatible database service

Fast
Millions of requests per second with millisecond latency

More throughput
Separation of storage and compute layers offloads replication to the storage volume so that your instances can do more work; twice the throughput of MongoDB

Optimizations
Database engine optimizations to reduce the number of IOs and minimize network packets in order to offload the database engine

Flexible
Scale up an instance in minutes for analytical queries and scale down at the end of the day
Flexible

Durability and replication are handled by the distributed storage volume

**Scenario 1:** Dev/test with a single instance

**Scenario 2:** Read scaling in minutes

**Scenario 3:** Scale-up and scale-out for analytics
Scalable
Fast, scalable, and fully managed MongoDB-compatible database service

Scale out in minutes
Scale out read capacity by adding additional replicas (up to 15 replicas); adding replicas takes minutes regardless of data size

Scale up in minutes
Scale up and down instances in minutes (15.25 GiB memory to 244 GiB memory)

Storage scales automatically
Storage volumes automatically grow from 10 GB to 64 TB without any user action

Load balancing
Load balancing across instances with replica sets
Fully managed
Fast, scalable, and **fully managed** MongoDB-compatible database service

**Pay-as-you-go pricing; enterprise grade**
On-demand, pay-as-you-go pricing enables you to pay only for the resources that you need and only when you use them.

**Automatic failure recover and failover**
Replicas are automatically promoted to primary; failing processes are automatically detected and recovered; no cache warmup needed.

**Point-in-time recovery**
Automated backups are stored in Amazon S3, which is designed for 99.999999999% durability.

**Durable, fault-tolerant and self-healing storage**
Data at rest is replicated six ways across three AZs; handle AZ + 1 failures.
Fully managed
Fast, scalable, and fully managed MongoDB-compatible database service

Automatic patching
Up to date with the latest patches

AWS Support
AWS Support provides people, technology, and programs to help you achieve success

Monitoring
More than 20 key operational metrics for your clusters at no extra charge

Integrated
Deeply integrated with AWS services
MongoDB compatible
Fast, scalable, and fully managed MongoDB-compatible database

- **MongoDB 3.6**
  - Compatible with MongoDB Community Edition 3.6

- **Same drivers, tools**
  - Use the same MongoDB drivers and tools with Amazon DocumentDB; as simple as changing an application connection string

- **Replica sets**
  - Read scaling is easy with automatic replica set configurations

- **Migration with AWS DMS**
  - Live migrations with AWS DMS; free for 6 months
Migration
Migrate to Amazon DocumentDB with Amazon Database Migration Service (AWS DMS)

AWS DMS is free to use for 6 months if you are moving to Amazon DocumentDB

- Migrate between on-premises MongoDB and Amazon DocumentDB
- Migrate self-hosted MongoDB databases to Amazon DocumentDB
- Data replication for virtually no downtime
- Migrate from replica sets and sharded clusters

DMS: 100,000+ Databases migrated
Demo: Live migration to DocumentDB with AWS DMS
AWS DMS demo architecture
Migrate from MongoDB to Amazon DocumentDB with AWS DMS

Demo app: https://www.meteor.com/tutorials
Security and compliance

Amazon VPC
- Strict network isolation with Amazon Virtual Private Cloud (VPC)

Encryption by default
- Encryption at rest with AWS KMS and customer-managed AWS keys; encryption in transit with TLS

Safe defaults
- Best practices are the defaults

Compliance
- Amazon DocumentDB has been assessed to comply with PCI DSS, ISO 9001, 27001, 27017, and 27018, SOC 2, in addition to being HIPAA eligible
Backup

Automatic backups
- Automatic, incremental, and continuous backups

No performance impact
- Backups do not affect database performance

35 days of PITR
- Point-in-time recovery (PITR) for up to 35 days

Archive snapshots
- Keep snapshots for as long as you need
Amazon DocumentDB: Backups streamed to S3

- Availability Zone 1
  - Instance (primary)

- Availability Zone 2
  - Instance (replica)

- Availability Zone 3
  - Instance (replica)

Distributed storage volume

Amazon S3
Demo: Backup
Getting started
Getting started with the AWS Management Console, AWS CLI, and AWS CloudFormation

Learn more: https://aws.amazon.com/documentdb/
Customers

"Our developers love the document model as it enables them to move fast and iterate quickly when building applications. Amazon DocumentDB integrates deeply with AWS services and has the potential to provide us with a robust, highly scalable, and cost effective database service. With Amazon DocumentDB, our developers will be able to move faster and focus more on innovating on behalf of our customers versus managing a database."

Sunjay Pandey, Vice President - Capital One.

"At Hudl, we utilize a significant amount of AWS services, as we're always looking for opportunities to get out of the business of managing our own infrastructure. Our developers love the MongoDB API and document model. We're very excited about the launch of Amazon DocumentDB, as it fits perfectly into our short-term and long-term architectural plans. Amazon DocumentDB has the capabilities we're looking for, and it's great to see AWS Database Migration Service (DMS) support from day one."

Brian Kaiser, CTO - Hudl.

"To provide our readers with the best possible experience, the Washington Post engineering team leverages AWS database services because they offer self-service provisioning without compromising operational excellence. From automated backup to multi-AZ failover, Amazon DocumentDB provides all the key features we need to build the world class systems that power both our Arc Publishing business and our machine learning platform. Document databases support the unstructured data that is prevalent in media, and with Amazon DocumentDB, we can now leverage AWS for all our critical database needs."

Patrick Cullen, Director of Data Science and Artificial Intelligence - The Washington Post.

"Dow Jones uses a number of specialized databases including MongoDB to provide a variety of services for our customers. We are excited about collaborating with AWS around Amazon DocumentDB, which meets key needs we expressed to AWS in order to simplify our operations and free up our developers to invest in innovative experiences for our customers rather than undifferentiated operations."

Ramin Beheshti, Chief Product & Technology Officer - Dow Jones
Amazon DocumentDB
with MongoDB compatibility

- Scalability
  - CPU & MEMORY
  - Compute & Storage Decoupled
  - Up to 15 Read-Replicas

- Availability
  - Self-Healing
  - 6-Way Replication
  - Fully Managed
  - Auto Failover

- Security
  - Encrypt In-Flight
  - Encrypt at Rest
  - Amazon Key Management Service

- Database Migration Service
- On-Premises

- AZ1, AZ2, AZ3

- Storage
  - 64 TB
  - Automatic

- DocumentDB Instance
- Replica
- Encrypted Backups
  - Manual Snapshots
  - Point-In-Time Recovery

MongoDB 3.6 Compatible

Twitter: @awsgeek
Purpose built
The right tool for the right job

Thank you