


디지털 전환을 위한 데이터베이스 선정 시 고려 사항

김 준 MongoDB 기술총괄 상무





For more than a decade,
organizations have been pursuing the promise of

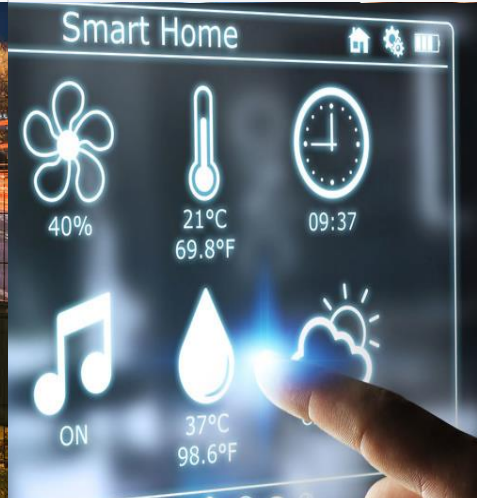
**Digital Transformation,
Digital Innovation**

**New IT
Models**

Mobile

**Smart
Objects**

**Expanded
Use Cases**



Legacy Modernization

88%

CIOs believe they have yet to benefit from their **digital strategy**

Source: Harvey Nash / KPMG CIO Survey 2017

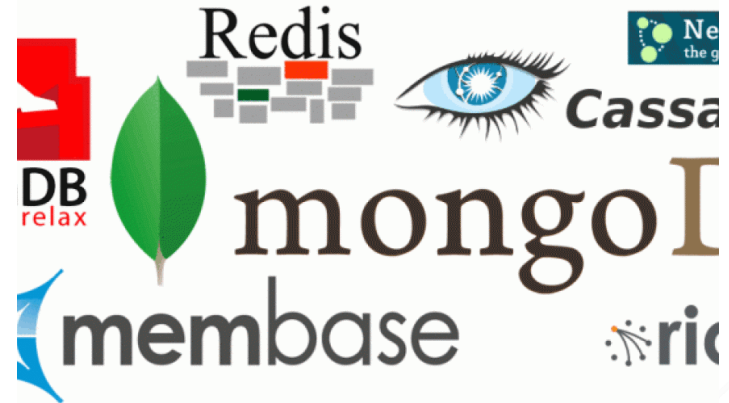
The reason?

DATA.

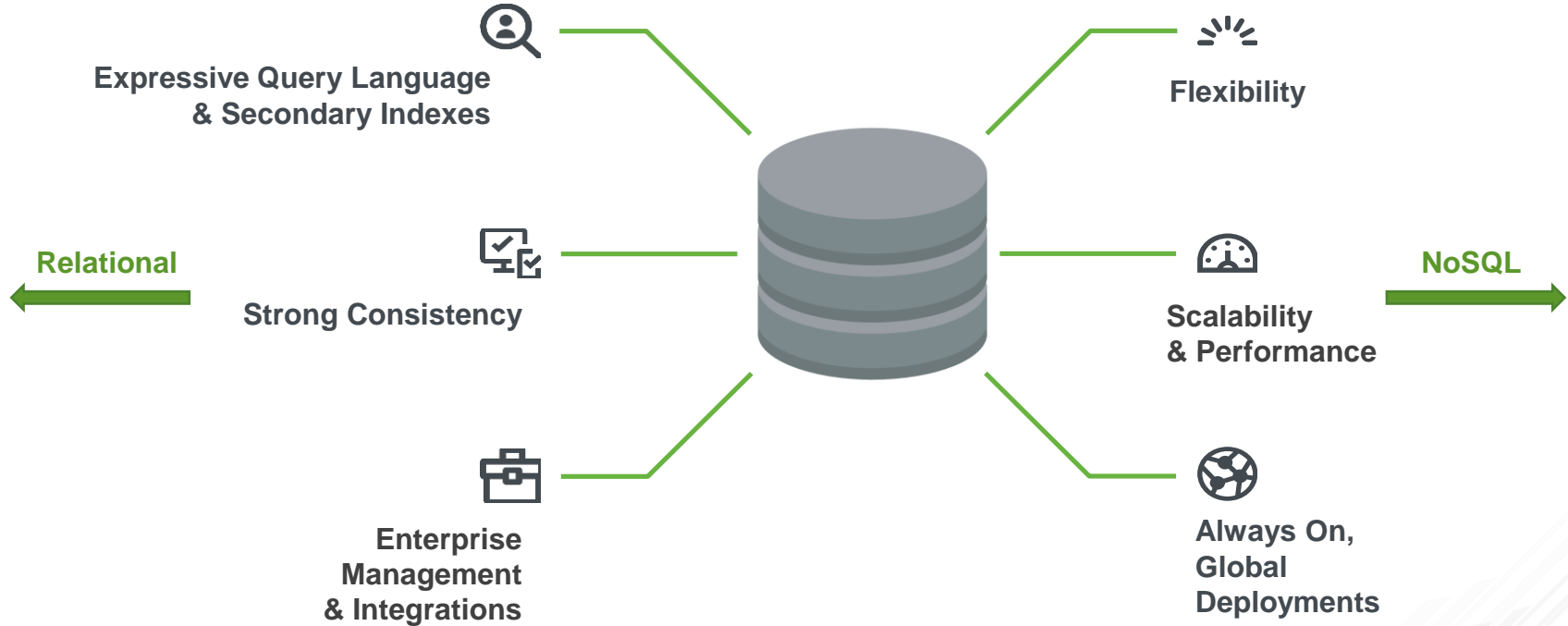
SILOED | COMPLEX | TRAPPED

RDBMSs Vs NoSQLs

How To Use in the Market



Strengths & Weakness



고려사항 1 - Data Model

- 사용하고 있는 데이터 모델이 스키마에 대한 유연성을 제공하고 있는가?
(**schema flexibility**)
- Key-Value 및 Wide-column 모델에서 데이터에 대한 access가 충분한가?
(제약사항이 무엇인가 ? Key를 통한 쿼리만 가능한가 ?)
- 왜 Document data model이 광범위하게 사용되어지고 , 논의 되어 지고 있는가 ? (**broadest applicability all around the world**)
- Document data model이 주는 이점은 무엇인가 ? 기존 object-oriented languages 와의 관계는 ? (**most natural and most productive**)

Document 모델

Relational

Person:

Pers_ID	Surname	First_Name	City
0	Miller	Paul	London
1	Ortega	Alvaro	Valencia
2	Huber	Urs	Zurich
3	Blanc	Gaston	Paris
4	Bertolini	Fabrizio	Rom

Car:

Car_ID	Model	Year	Value	Pers_ID
101	Bentley	1973	100000	0
102	Rolls Royce	1965	330000	0
103	Peugeot	1993	500	3
104	Ferrari	2005	150000	4
105	Renault	1998	2000	3
106	Renault	2001	7000	3
107	Smart	1999	2000	2

no relation



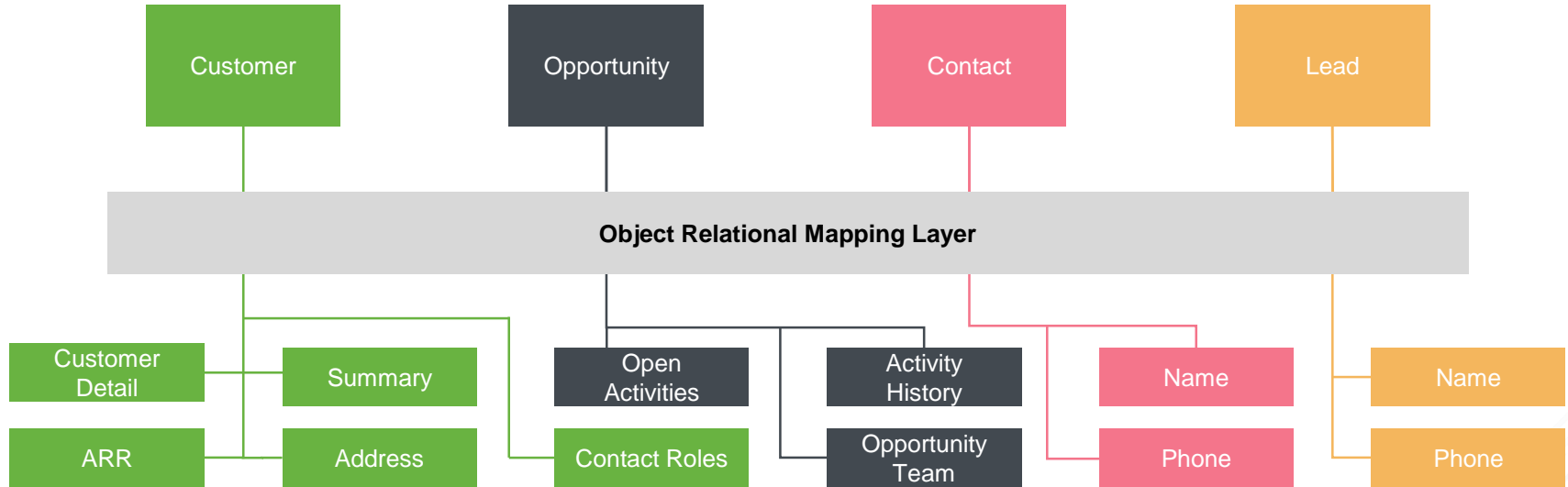
MongoDB Document

```
{
  first_name: 'Paul',
  surname: 'Miller'
  city: 'London',
  location: [45.123,47.232],
  cars: [
    { model: 'Bentley',
      year: 1973,
      value: 100000, ... },
    { model: 'Rolls Royce',
      year: 1965,
      value: 330000, ... }
  ]
}
```

- 다형성
- 배열
- 서브도큐먼트
- JSON/BSON

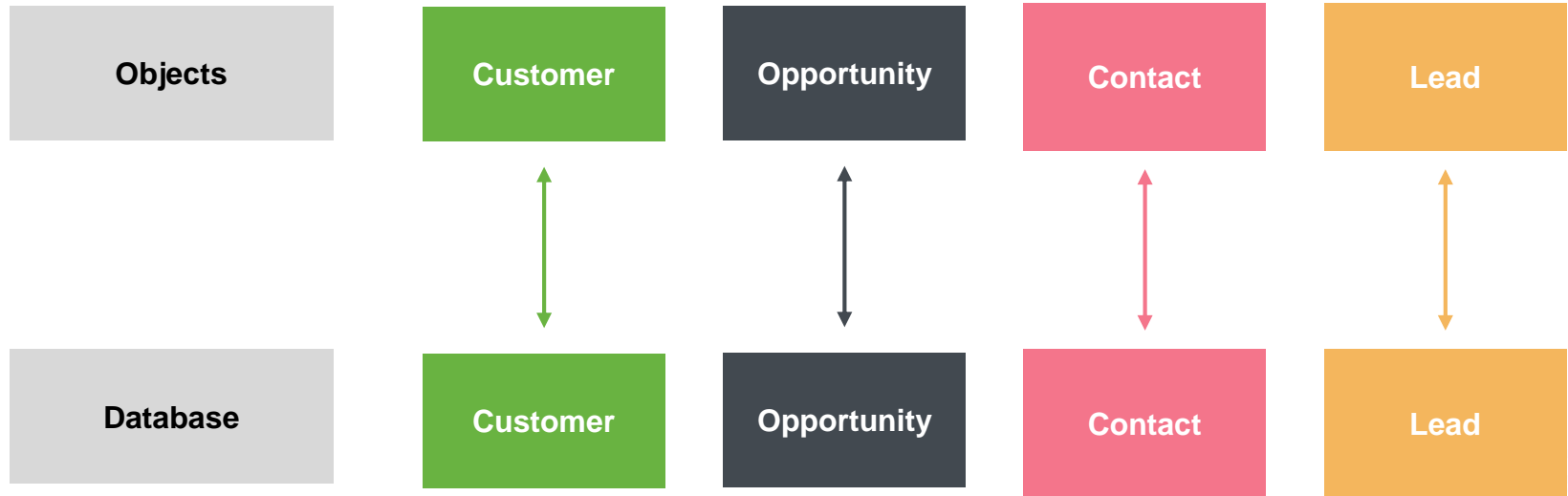
Document 모델 (JSON)

Objects



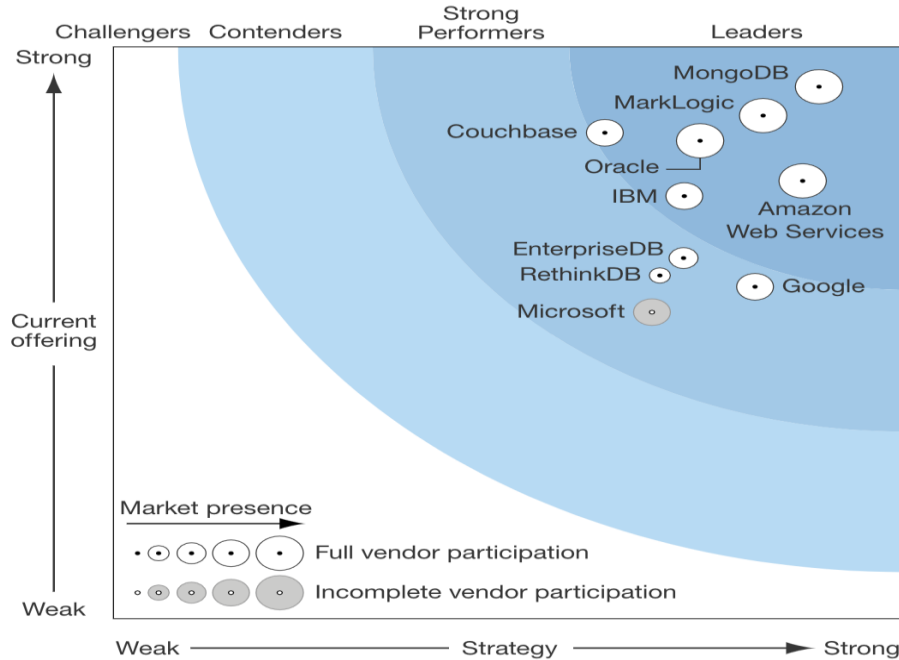
Tables

Document 모델 (JSON)



Document 모델 Leader

FIGURE 3 Forrester Wave™: Document Stores, Q3 '16



The Forrester Wave™
Smart data for smart decisions

Go to Forrester.com to download the Forrester Wave tool for more detailed product evaluations, feature comparisons, and customizable rankings.

고려사항 2. 쿼리 모델

- 데이터 쿼리를 얼마나 효율적으로 할 수 있는가? (**the ability to query data efficiently**)
- 다양한 형태의 워크로드에 대한 풍부한 쿼리 기능을 제공하는가? (**Richest query functionality**)
- Key-Value 및 Wide-column family 는 기본 키로 데이터에 액세스하는 단일 방법을 제공하며 , 매우 제한된 쿼리 기능을 제공하며 기본 쿼리 패턴 이외의 기능을 지원하기 위해 추가 개발 비용과 응용 프로그램 수준 요구 합니다.

풍부한 쿼리 기능

Expressive Queries	<ul style="list-style-type: none">• Find anyone with phone # “1-212...”• Check if the person with number “555...” is on the “do not call” list
Geospatial	<ul style="list-style-type: none">• Find the best offer for the customer at geo coordinates of 42nd St. and 6th Ave
Text Search	<ul style="list-style-type: none">• Find all tweets that mention the firm within the last 2 days
Aggregation	<ul style="list-style-type: none">• Count and sort number of customers by city, compute min, max, and average spend
Native Binary JSON Support	<ul style="list-style-type: none">• Add an additional phone number to Mark Smith’s record without rewriting the document• Update just 2 phone numbers out of 10• Sort on the modified date
JOIN (\$lookup)	<ul style="list-style-type: none">• Query for all San Francisco residences, lookup their transactions, and sum the amount by person
Graph Queries (\$graphLookup)	<ul style="list-style-type: none">• Query for all people within 3 degrees of separation from Mark

효율적인 집계(Aggregation) 쿼리

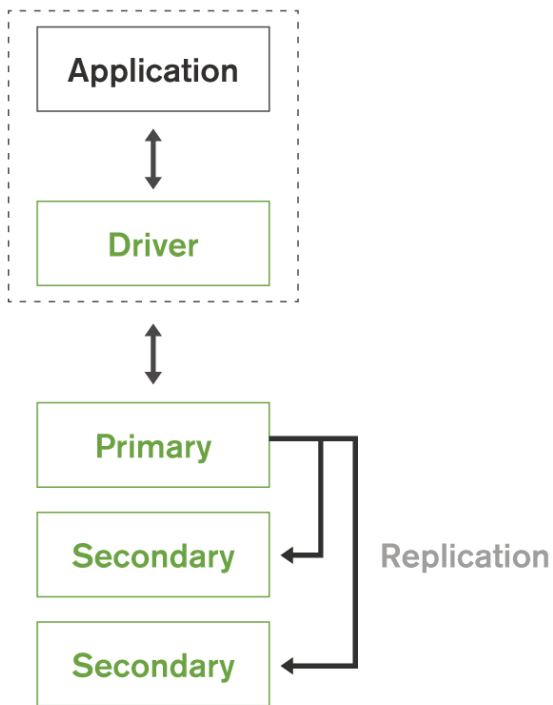
```
SELECT
  city,
  SUM(annual_spend) Total_Spend,
  AVG(annual_spend) Average_Spend,
  MAX(annual_spend) Max_Spend,
  COUNT(annual_spend) customers
FROM (
  SELECT t1.city, customer.annual_spend
  FROM customer
  LEFT JOIN (
    SELECT address.address_id, city.city,
           address.customer_id, address.location
    FROM address LEFT JOIN city
    ON address.city_id = city.city_id
  ) AS t1
  ON
  (customer.customer_id = t1.customer_id AND
   t1.location = "home")
) AS t2
GROUP BY city;
```

```
db.customers.aggregate([
  {
    $unwind: "$address",
  },
  {
    $match: {"address.location": "home"}
  },
  {
    $group: {
      _id: "$address.city",
      totalSpend: {$sum: "$annualSpend"},
      averageSpend: {$avg: "$annualSpend"},
      maximumSpend: {$max: "$annualSpend"},
      customers: {$sum: 1}
    }
  }
])
```

고려사항 3. Architecture & Consistency & Transaction Model

- 무 중단 서비스 – 24 by 7 (365일) 비즈니스 연속성 제공
- 대부분의 응용프로그램, 개발팀은 “**strongly consistent systems** ” 을 기대 하고 요구
- 서로 다른 일관성 모델은 응용프로그램의 일관성 , 가용성 영역에서 서로 다른 트레이드-오프 가 존재
- MongoDB는 쿼리 레벨에서 일관성에 대한 레벨을 정의 (**tunable consistency**)
- MongoDB는 Transaction에 대한 ACID를 보장

고가용성 (High Availability)



Replica Set – 2 to 50 copies

Self-healing shard

Data Center Aware

Addresses availability considerations:

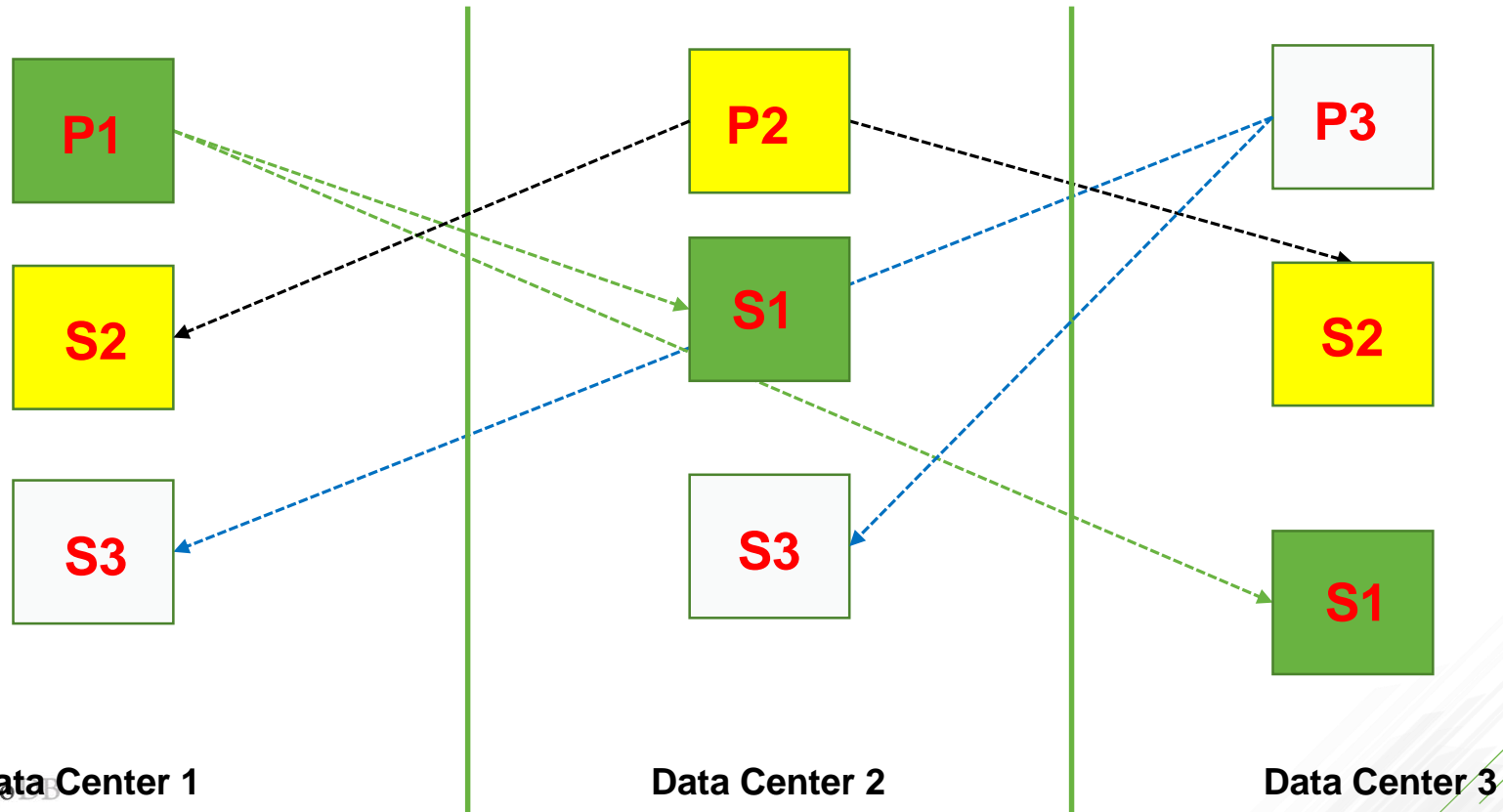
High Availability

Disaster Recovery

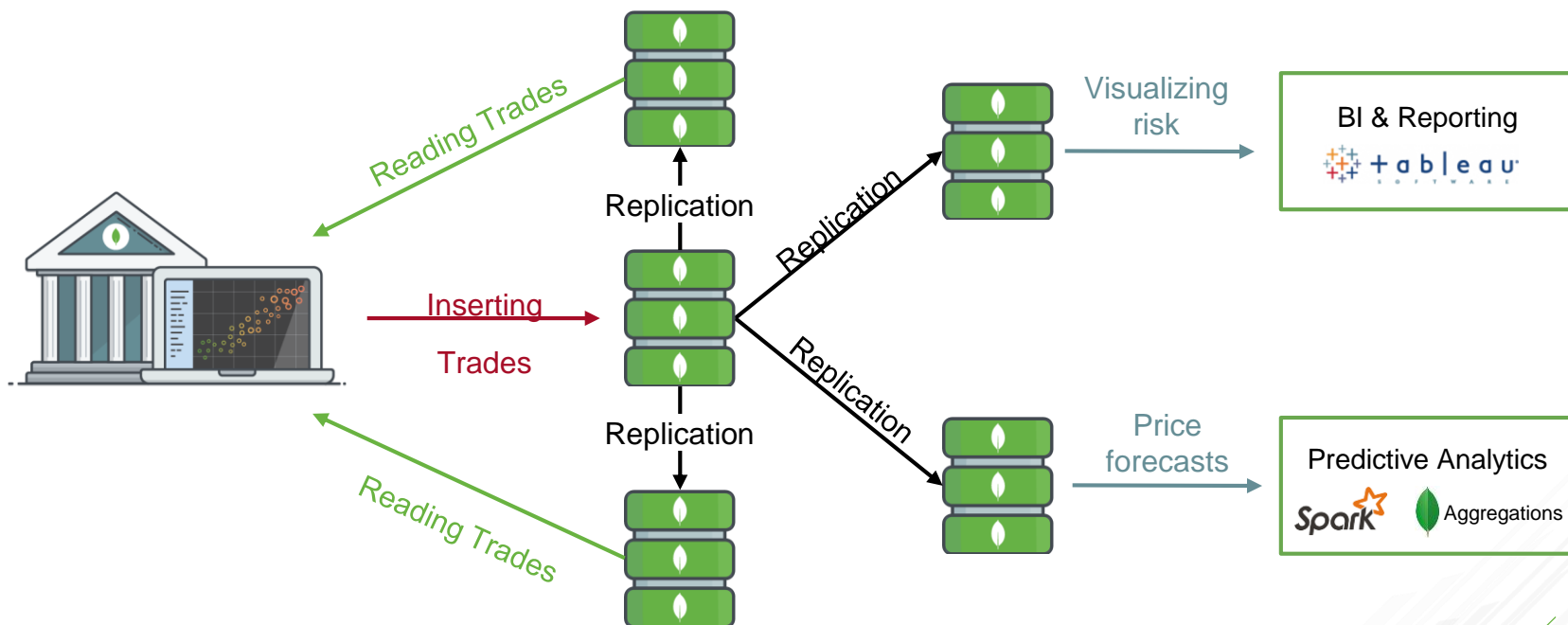
Maintenance

Workload Isolation: operational & analytics

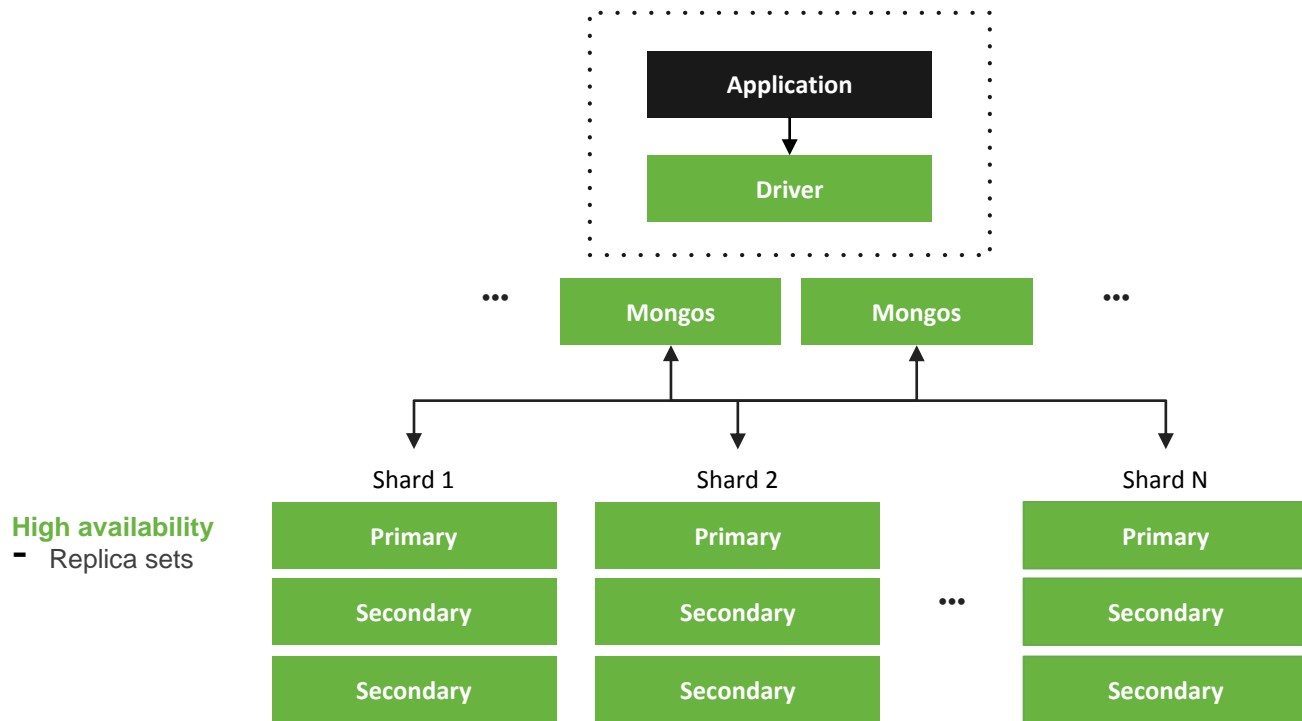
DR 구성의 편이성



다양한 타입의 워크로드에 대한 Isolation



샤딩을 통한 Scale out 지원



ACID Transaction 지원

```
with client.start_session() as s:  
    s.start_transaction()  
    try:  
        collection.insert_one(doc1, session=s)  
        collection.insert_one(doc2, session=s)  
        s.commit_transaction()  
    except Exception:  
        s.abort_transaction()
```

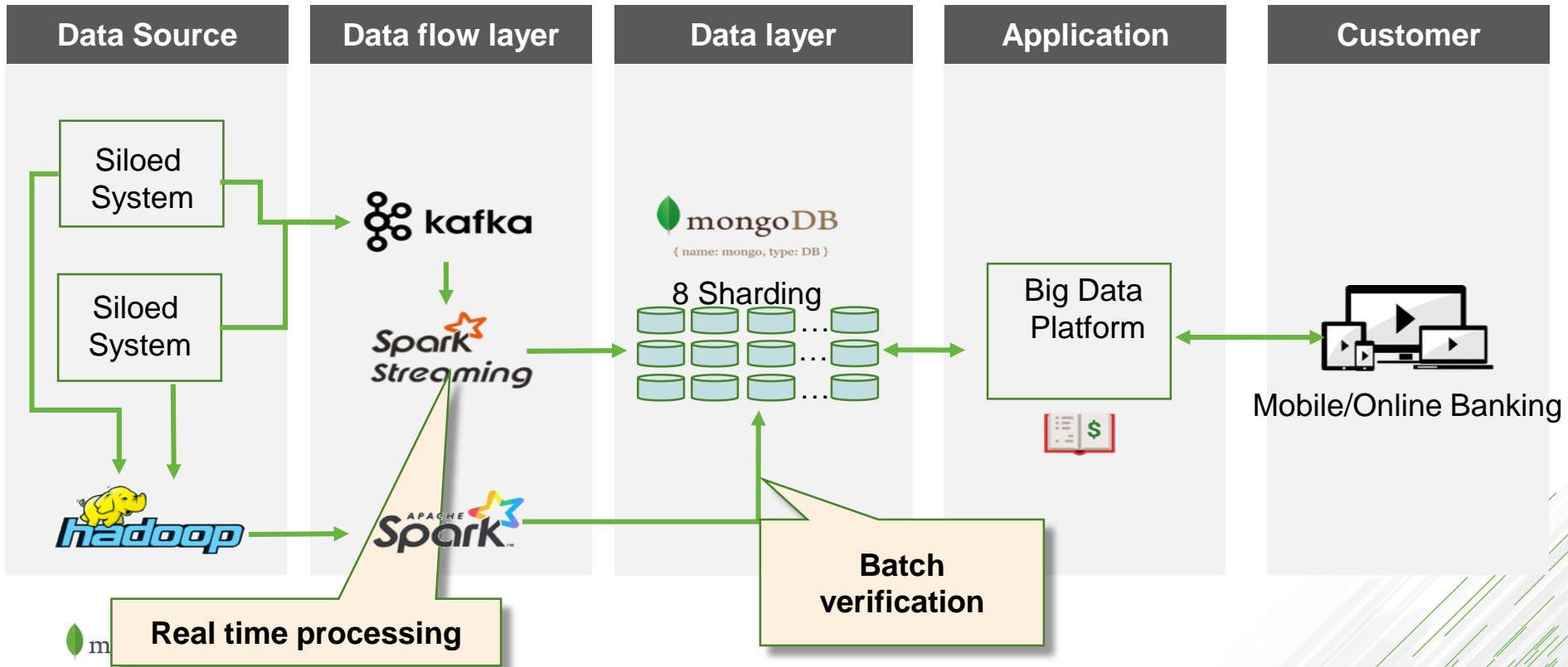
개발자 친화적 API

- 관용적인 프로그래밍 언어 지원
- 관계형 개발자에게 친숙한 api
- 단순한 구현

고려사항 4. API & Tools

- API의 성숙도와 기능은 비 관계형 제품에서는 서로 차이가 있음.
- **MongoDB's idiomatic drivers**는 새로운 개발자의 적용 시간을 최소화하고 응용 프로그램 개발을 단순화 함
- **Not all SQL is created equal.** 비 관계형 데이터베이스가 제공하는 SQL과 유사한 API를 통해 응용 프로그램과 개발자의 요구를 충족시킬 수 있는지 확인 필요
- 다른 시스템, 환경과의 상호 작용에 대한 고려

Data As a Service 사례

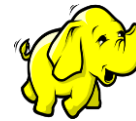


Drivers & Frameworks

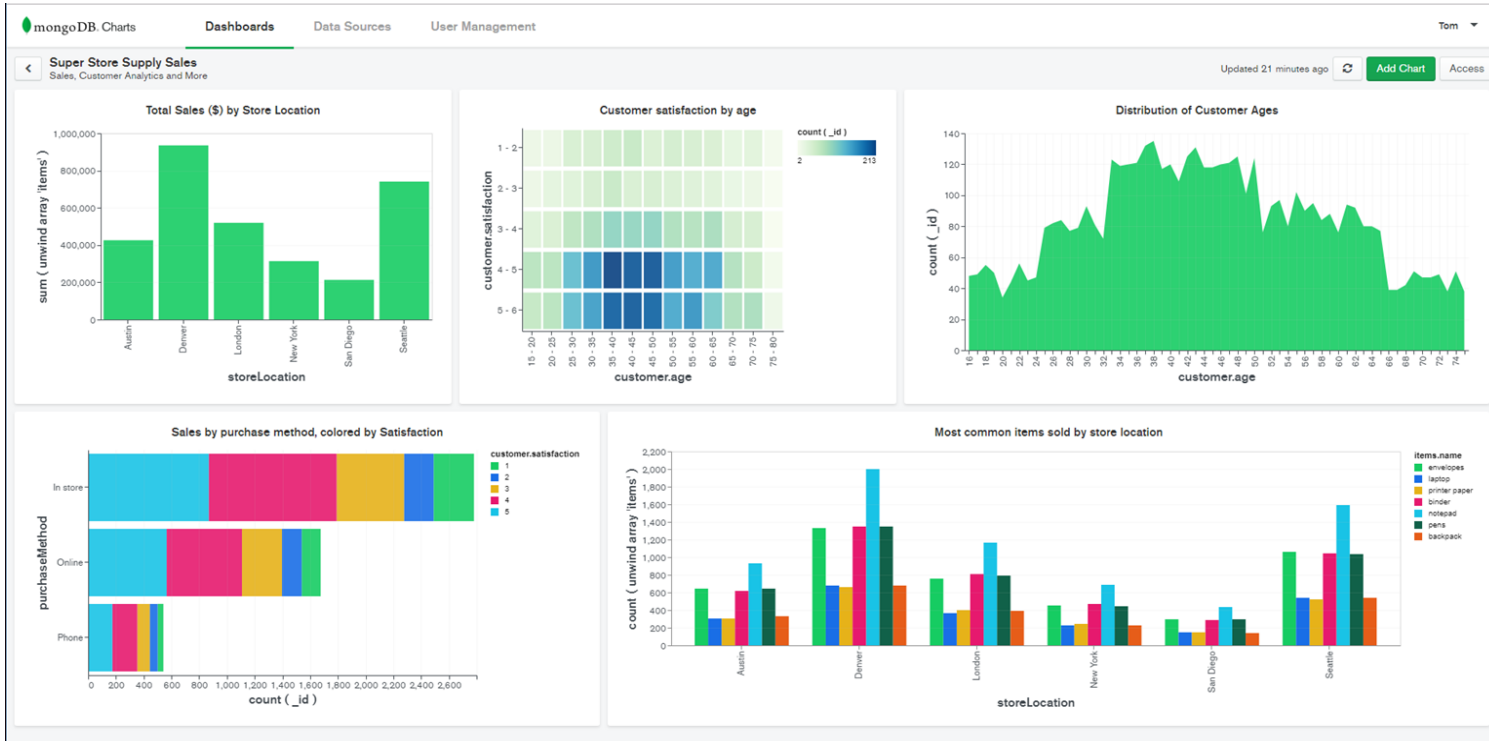
Drivers



Frameworks



Data Visualization - MongoDB Chart



개발자, 운영자를 위한 Tools Kit (Ops-Manager & Compass)

The screenshot displays the MongoDB Ops Manager interface. At the top, it shows 'mongoDB Ops Manager' with version '4.0.2.0187201809051427Z' and 'All Clusters'. The 'CONTEXT' is 'DEMOORG > DEMOPROJECT'. The 'Deployment' section is active, showing 'Processes' for 'myReplicaSet' and 'myReplicaSetQA'. Each replica set card includes a 'DATA METRICS MODIFY' menu, 'TYPE' (Replica Set), 'SSL' (Disabled), 'MONGODB COUNT' (3), 'AUTH' (Disabled), 'DATABASE SIZE', and 'BI CONNECTOR CT'. Performance graphs for 'Operations' (R: 2.9/s, W: 0), 'Disk Usage' (18.7 GB), 'Connections' (21), and 'Disk IOPS' (22.4) are shown for the last 6 hours. The bottom status bar shows the local host ID: 'localhost:8080/v2/5babd423521d5b1507b5ca5fmetrics/replicaSet/5ceb7124521d5b046ce3c849/overview'.

The screenshot shows the MongoDB Compass interface for a query. The query is: `{ "last_login": { "$gte": { "$date": "2013-08-21T11:41:27.000Z"}, "$lte": { "$date": "2013-08-26T06:12:18.000Z"} } }`. The 'Query Performance Summary' shows: Documents Returned: 344184, Index Keys Examined: 374841, Documents Examined: 374841, Actual Query Execution Time (ms): 5071, Sorted in Memory: no, and Query used the following index: last_login. The 'VIEW DETAILS AS' section shows a 'VISUAL TREE' of query stages: SHARD_MERGE (returned 344184, Execution Time 511 ms), SHARD01 (SHARDING_FILTER, returned 113342, Execution Time 1290 ms), and SHARD02 (SHARDING_FILTER, returned 15082, Execution Time 1130 ms). Annotations indicate that the performance summary shows the actual query execution time, details for each stage can be expanded with the 'Details' button, and the clocks provide an estimate of how long the previous stages (gray) and the current stage (blue) took relative to total query time.

고려사항 5. Commercial Support, Community Strength, Freedom from Lock-In


- 데이터베이스(특히 NoSQL) 를 선정하는데 있어 **Community size** , **commercial strength** 는 중요한 요소임
- MongoDB는 Non-relational Database 영역에서는 몇 안되는 상장 회사이며, 데이터베이스 기술로는 과거 20여년 동안 처음 상장한 회사임 .
- 기술 지원은 24*7 (follow the sun support model) 제공
- MongoDB는 자체 인프라에서 실행하거나 모든 주요 공용 클라우드 플랫폼에서 완전히 관리되는 클라우드 서비스를 제공 (MongoDB Atlas)


완전 관리형 데이터베이스 서비스 – MongoDB Atlas


<p>Self-service & elastic</p> <ul style="list-style-type: none">- 필요시 배포, 수정 및 업그레이드는 베스트 Practice기반으로 운영 자동화 제공- 자동화 된 데이터베이스 유지 관리 몇 번의 클릭 또는 API 호출로 확대, 축소 또는 축소 구성의 자동화	<p>Global & cloud-agnostic</p> <ul style="list-style-type: none">- AWS/Azure/GCP 여러 지역에서 사용 가능- 어디에서나 읽기 / 쓰기가 가능한 글로벌 클러스터를 통한 배포 및 다중 리전을 통한고가용성 제공- 클라우드 환경 기반에서 손쉽게 마이그레이션 가능	<p>Enterprise-grade security & SLAs</p> <ul style="list-style-type: none">- 네트워크 격리, VPC 피어링, 종단 간 암호화 및 역할 기반 액세스 제어- 암호화 키 관리, LDAP 통합, 상세 데이터베이스 감사- SOC 2 / 개인 정보 보호/ HIPAA SLA로 안정성 보장
<p>Comprehensive monitoring</p> <ul style="list-style-type: none">- 전 100 개 이상의 메트릭을 통한 KPI에 대한 alerting- 실시간 성능 조연 및 성능 관리자- 모니터링 대시 보드와 통합 할 수 있는 API	<p>Managed backup</p> <ul style="list-style-type: none">- 특정 시점의 데이터 복구- 쿼리 가능한 백업 스냅 샷- 샤딩환경내 일관된 스냅 샷- 클라우드 데이터 이동성	<p>Stitch: Serverless platform services</p> <ul style="list-style-type: none">- 백엔드 로직, 서비스 통합 및 API를위한 단순한 서버리스 기능- 간단한 필드 레벨 액세스 규칙으로 보호되는 프론트 엔드에서의 데이터베이스 액세스- 변경 사항에 실시간으로 반응하는 데이터베이스 및 인증 트리거

다양한 Cloud 환경하에서 사용 - AWS , Azure, GCP

Cloud Provider & Region AWS, N. Virginia (us-east-1) ▾




Google Cloud Platform


Azure

Create a **free tier cluster** by selecting a region with **FREE TIER AVAILABLE** and choosing the **M0** cluster tier below.

★ recommended region ⓘ



<p>NORTH AMERICA</p> <ul style="list-style-type: none">🇺🇸 N. Virginia (us-east-1) ★ FREE TIER AVAILABLE🇺🇸 Ohio (us-east-2) ★🇺🇸 N. California (us-west-1)🇺🇸 Oregon (us-west-2) ★🇨🇦 Montreal (ca-central-1)	<p>EUROPE</p> <ul style="list-style-type: none">🇮🇪 Ireland (eu-west-1) ★🇬🇧 London (eu-west-2) ★🇫🇷 Paris (eu-west-3) ★🇩🇪 Frankfurt (eu-central-1) ★ FREE TIER AVAILABLE <p>SOUTH AMERICA</p> <ul style="list-style-type: none">🇧🇷 Sao Paulo (sa-east-1)	<p>AUSTRALIA</p> <ul style="list-style-type: none">🇦🇺 Sydney (ap-southeast-2) ★ <p>ASIA</p> <ul style="list-style-type: none">🇯🇵 Tokyo (ap-northeast-1) ★🇰🇷 Seoul (ap-northeast-2)🇸🇬 Singapore (ap-southeast-1) ★ FREE TIER AVAILABLE🇮🇳 Mumbai (ap-south-1) FREE TIER AVAILABLE
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

다중 Region을 통한 구성

- Electable Node에 대한 multi region 구성으로 single region에 대한 outage를 대비 할 수 있다.

Electable nodes for high availability







Configure 3, 5, or 7 nodes across multiple regions to better withstand data center outages

Region	Nodes
 Seoul (ap-northeast-2)  HIGHEST PRIORITY	<input type="text" value="3"/>
+ Add a region	

- Read Only Node를 통해 (Local Node) 구성을 통해 latency 문제를 해결 할 수 있다.

Read-only nodes for optimal local reads




Add replicas in additional regions to optimize for local reads in any of your service areas

Region	Nodes
 Seoul (ap-northeast-2) 	<input type="text" value="1"/> 
 Tokyo (ap-northeast-1) 	<input type="text" value="1"/> 
+ Add a region	

- Analytic Node를 통해 (Local Node) 구성을 통해 latency 문제를 해결 할 수 있다.

Analytics nodes for workload isolation

Isolate queries on read-only nodes that will not contend with your operational workload

Region	Nodes
 Seoul (ap-northeast-2) 	<input type="text" value="1"/> 

고객지원 포탈 - MongoDB Support Portal

MongoDB Help Center

MongoDB Paid Support

MongoDB offers help with training, upgrading, and more



COMMERCIAL SUPPORT

Named Technical Services Engineers

Commercial Support

Our Support Policy



TRAINING & CONSULTING

Major Version Upgrade Assistance

Dedicated Consulting Engineer

Developer & Operations Focused

Consulting

Customer Support Portal

Commercial Support subscribers can sign in to read the Knowledge Base, escalate an issue, or create and manage cases.

Sign in

MongoDB Documentation & Community Support



Getting Started



MongoDB Manual 4.0



Cloud Products



Guides

Why MongoDB



Why MongoDB?

Intelligent Operational Data Platform



**Best way to work
with data**



**Intelligently put data
where you need it**



**Freedom
to run anywhere**

What do you mean ? Stack Overflow Developer Survey 2018 , 2019



Most Wanted Databases

MongoDB

18.6%

Elasticsearch

12.2%

PostgreSQL

11.4%

Redis

9.7%

MySQL

7.5%

Microsoft Azure (Tables, CosmosDB, SQL, etc)

7.3%

Google Cloud Storage

7.3%

Cassandra

6.1%

Amazon DynamoDB

5.7%

Google BigQuery

5.6%

SQL Server

4.2%

Neo4j

3.9%

Amazon RDS/Aurora

3.5%

MariaDB

3.4%

Amazon Redshift

3.3%

SQLite

3.3%

Memcached

2.7%

Apache Hive

2.6%

Apache HBase

2.4%

Oracle

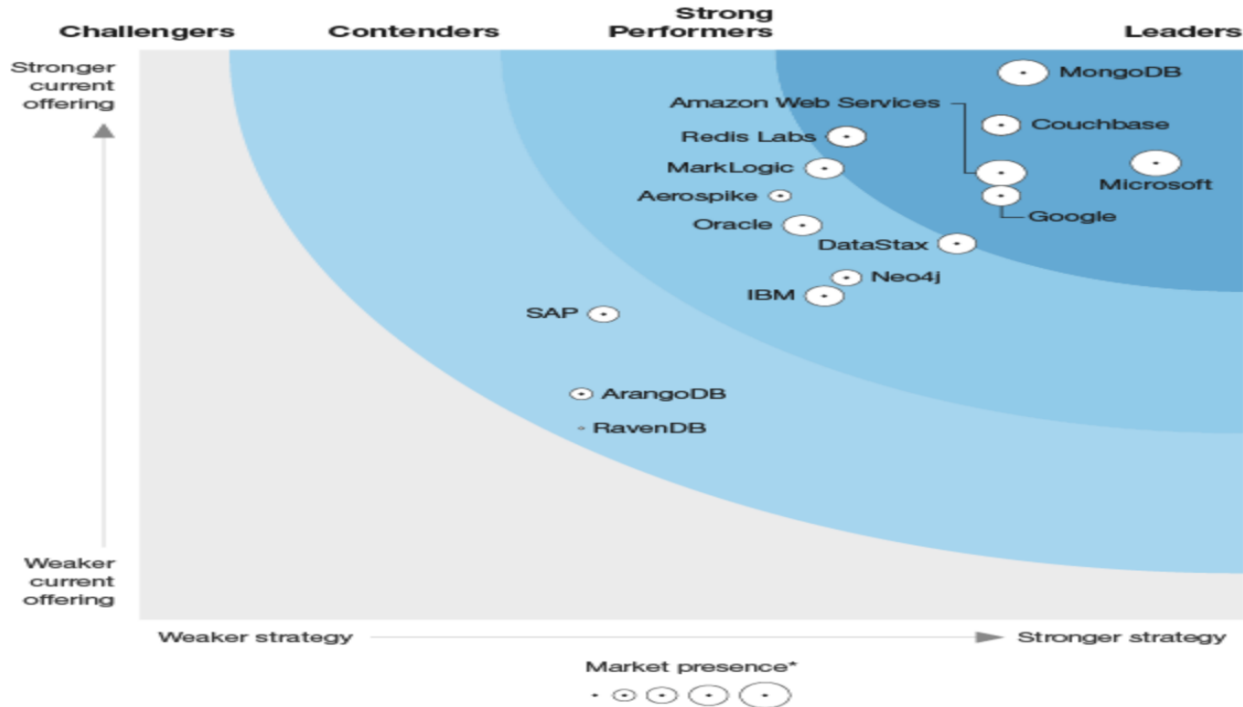
2.3%

IBM Db2

0.7%

The Forrester Wave™: Big Data NoSQL, Q1 2019

THE FORRESTER WAVE™
Big Data NoSQL
Q1 2019



The Forrester Wave™: Database-As-A-Service, Q2 2019

THE FORRESTER WAVE™
Database-As-A-Service
Q2 2019



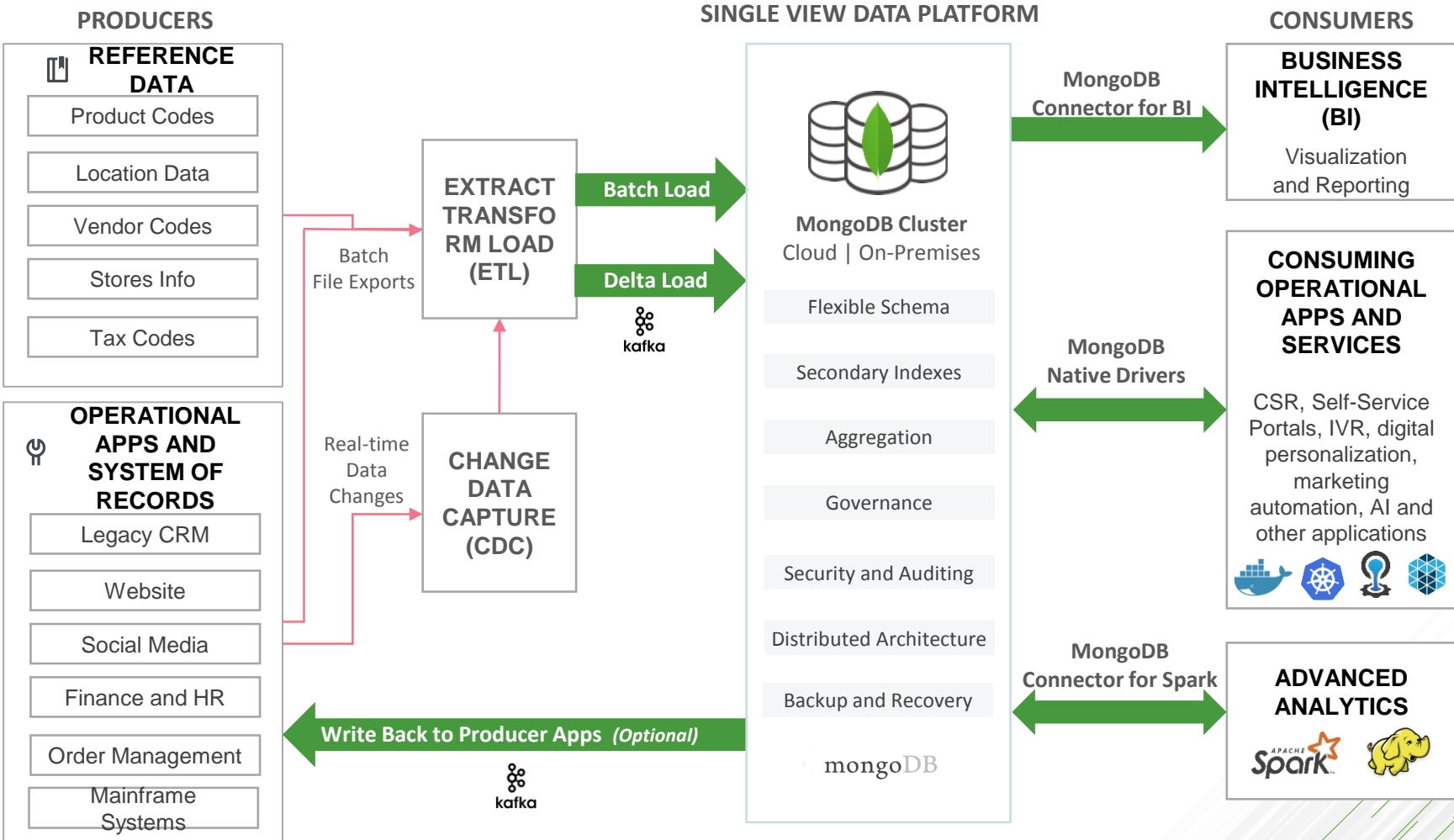
DB-Engines.com

350 systems in ranking, February 2020

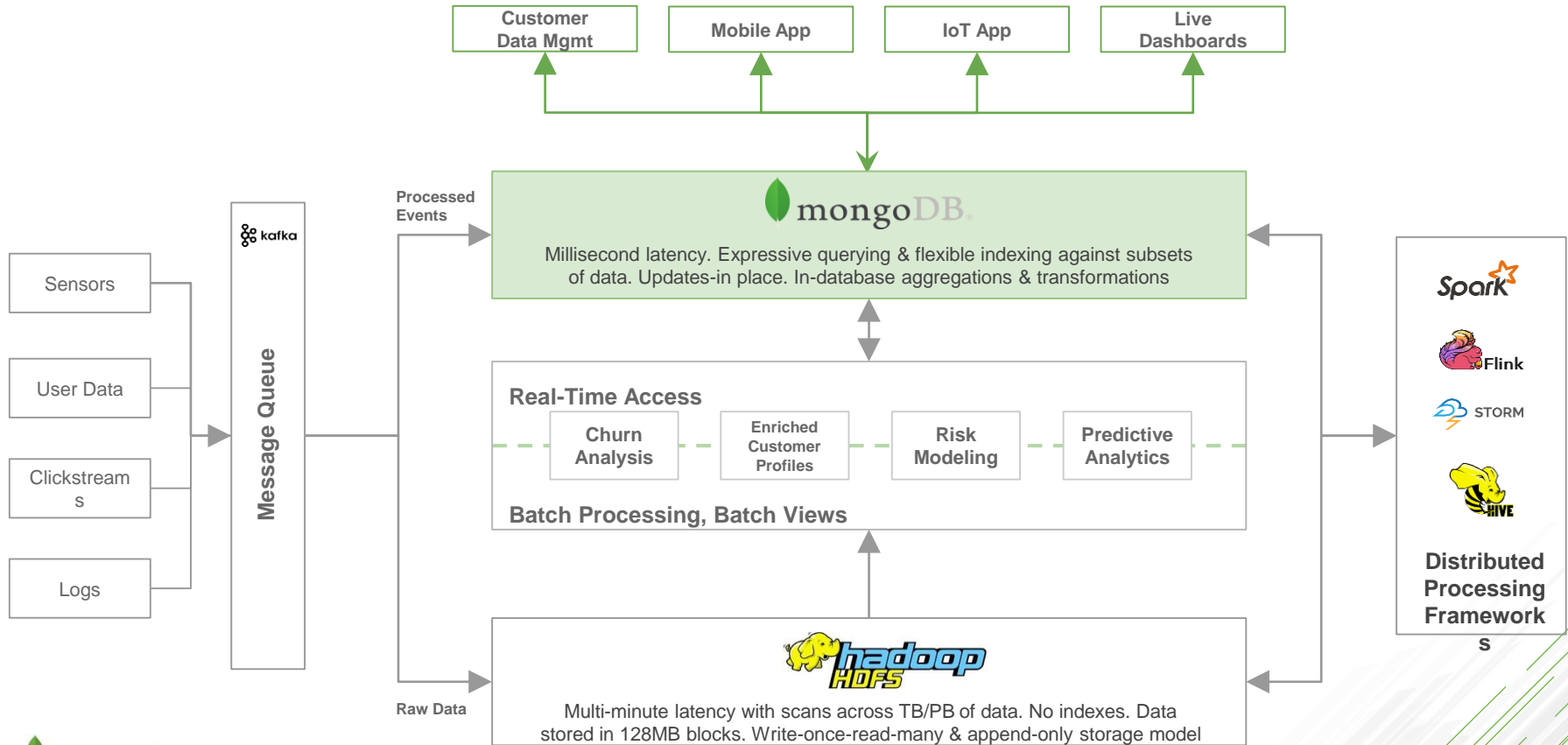
Rank			DBMS	Database Model	Score		
Feb 2020	Jan 2020	Feb 2019			Feb 2020	Jan 2020	Feb 2019
1.	1.	1.	Oracle	Relational, Multi-model	1344.75	-1.93	+80.73
2.	2.	2.	MySQL	Relational, Multi-model	1267.65	-7.00	+100.36
3.	3.	3.	Microsoft SQL Server	Relational, Multi-model	1093.75	-4.80	+53.69
4.	4.	4.	PostgreSQL	Relational, Multi-model	506.94	-0.25	+33.38
5.	5.	5.	MongoDB	Document, Multi-model	433.33	+6.37	+38.24
6.	6.	6.	IBM Db2	Relational, Multi-model	165.55	-3.15	-13.87
7.	7.	8.	Elasticsearch	Search engine, Multi-model	152.16	+0.72	+6.91
8.	8.	7.	Redis	Key-value, Multi-model	151.42	+2.67	+1.97
9.	9.	9.	Microsoft Access	Relational	128.06	-0.52	-15.96
10.	10.	10.	SQLite	Relational	123.36	+1.22	-2.81

Reference Architecture

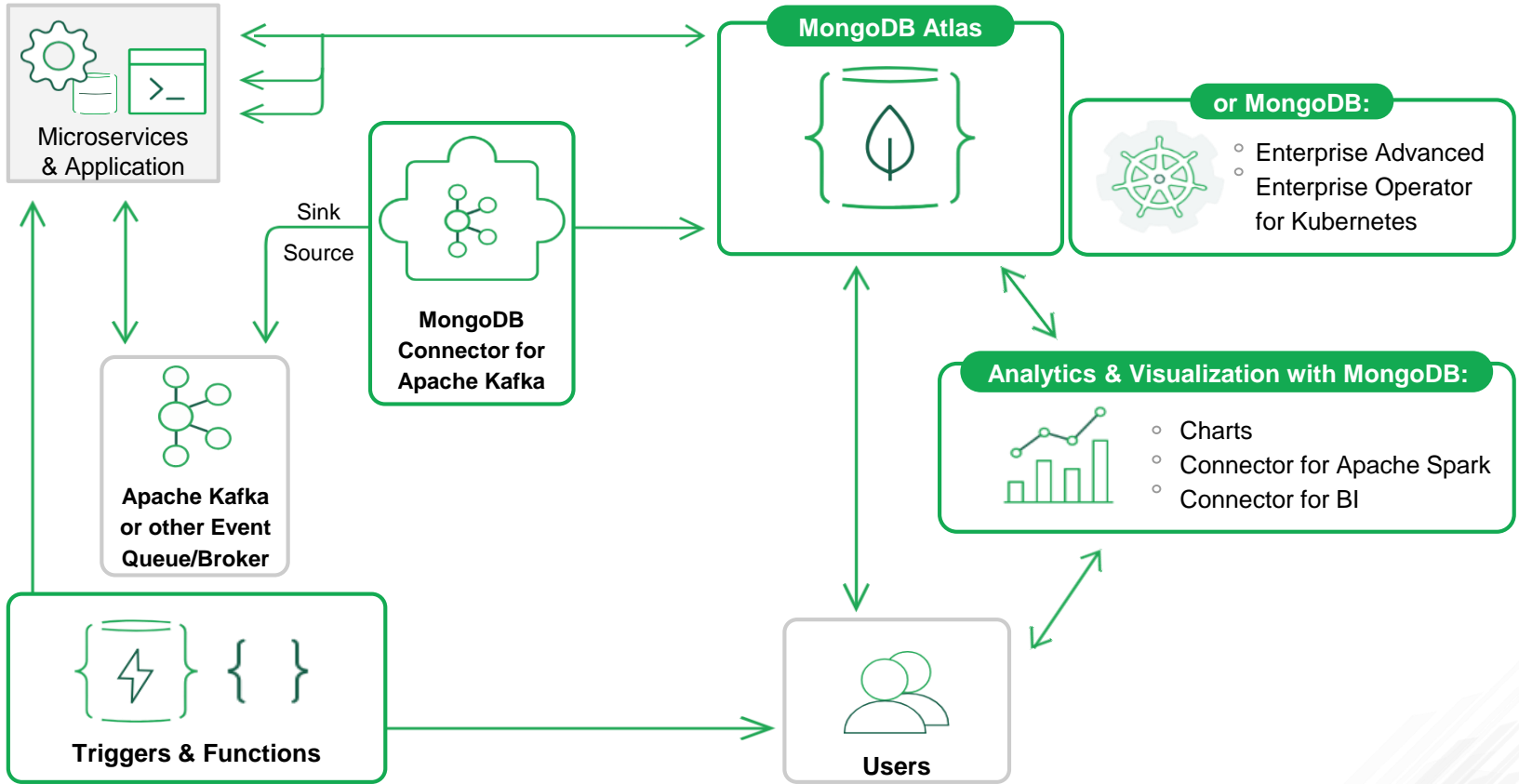




기술 및 데이터 통합의 해결



MongoDB Event-Driven Architecture



Register MongoDB Atlas

Estimate the price of your deployment

1 Choose your cloud provider



2 Choose region

3 Choose Cluster Size

Please select a region



4 Customize your storage

No additional options for this cluster



Please select region for price estimate

M0 Free Tier

Shared RAM ■ 512 MB Storage

Start with a free 3-node replica set, a sandbox environment for getting started with MongoDB.

Get Started Free