DATASTAX

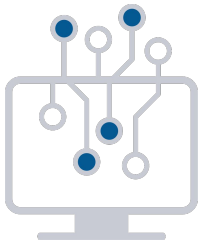
Relational (RDBMS) to NoSQL Migration

Ankit Patel | DataStax | Principal Strategy Architect

“We cannot solve our problems with the same thinking we used when we created them.”

- Albert Einstein

The Digital Era - The Need to Modernize



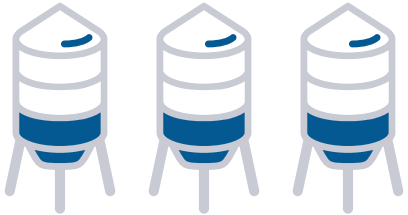
Digital

Data-Driven

AI Enabled

DATASTAX

The Modern Era
SAD (Silos Affects Delivery) Speed of Data Matters!



Data access

Source: https://www.pinterest.com/pin/573716440029920090/

Legacy processes

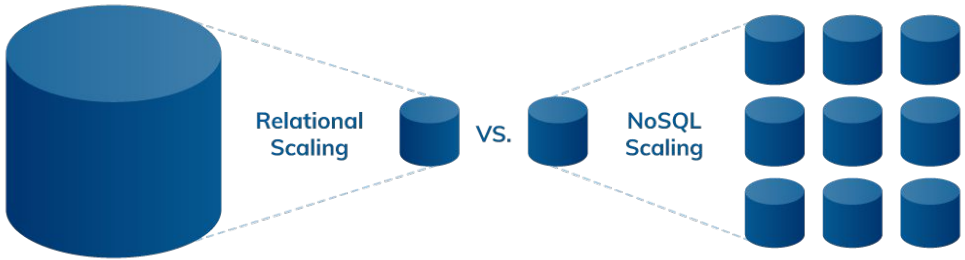
Resistance to change

Lack of data analytical skills

DATASTAX

NoSQL - The Future
What is a NoSQL (Not-only-SQL) Database?

- Non Relational Database - supports ability to access data using other forms besides Structured Query Language (SQL)
- Designed to be used by Cloud Applications' need to handle massive amounts of Data in real-time
- Provides ability to overcome scale, performance, data storage, data model, and data distribution limitations



DATASTAX

NoSQL vs RDBMS....

	When to use NoSQL?	When to use RDBMS?
Applications	Decentralized (scalable) microservice applications	Centralized monolithic applications
Availability	100% availability, zero-downtime	Moderate to high
Data	Low latency structured/semi/unstructured data @ high velocity	Structured data @ moderate velocity & latency
Transactions	Simple transactions & queries	Complex nested transactions & joins
Scalability (Reads/Writes)	Horizontal (Linear) scaling	Vertical scaling

Cassandra: The Best NoSQL Database of Choice

Zero Downtime

Active-everywhere,
masterless, scales linearly

Zero Lock-in

Best NoSQL database for
cloud-native and microservices

Global Scale

#1 choice of world's largest
consumer internet applications

If you use a website or a smartphone today,
you're touching a Cassandra backend system.

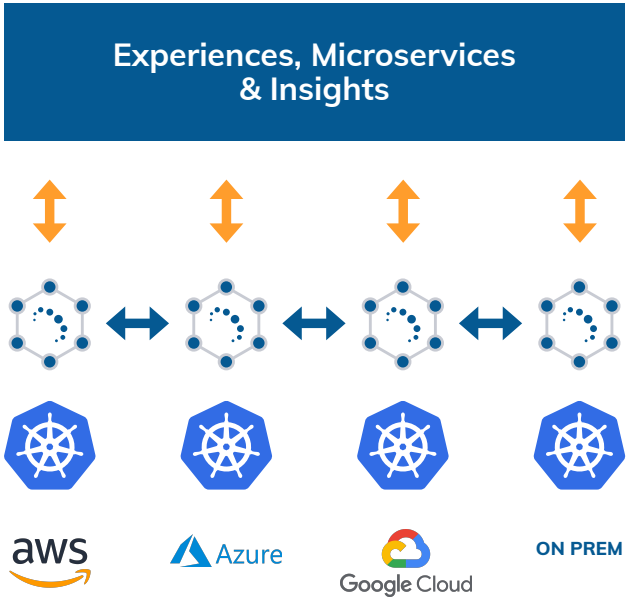
Source: https://sdtimes.com/data/apache-cassandra-4-0-beta-now-available/

DATASTAX

Cassandra: Cloud Native NoSQL Database

Why?

With Cassandra masterless architecture, easily achieving 100% uptime across on-prem, single cloud, hybrid, and/or multi-cloud deployments is engraved in the technology.



Cassandra: What is CQL?

- CQL – Cassandra Query Language
- Similar to syntax compared to SQL
- Standard way to communicate to DSE C* cluster for reading/writing data.
- Feature rich language that allow you to manage the cluster (managing schema/permissions, managing roles, JSON support, UDF/UDA support…)
- Example Read: select * from keyspace.table where partition_key=<value>;
- Example Writing Data: insert into keyspace.table (partition_key,clustering_key,value1) values (‘A’,’B’,’C’);

Cassandra: What is a Keyspace?

- Similar to schema in RDBMS
- Container for multiple tables
- Replication Strategy is set at the keyspace level (Example: SimpleStrategy, NetworkTopologyStrategy)
- Replication Factor defined at the keyspace level
- DURABLE_WRITES is set at the keyspace level. Setting to false will bypass the commit log.
- Example to create keyspace:
  CREATE KEYSPACE test WITH replication = {'class': NetworkTologyStrategy', 'DC1': '1'} AND durable_writes = true;

Cassandra: What is a Table?

- Same as RDMBS table
- Contains a primary key
- Always has partition key as part of primary key
- Optionally can define a clustering key (ordering can be defined)
- Both partition and clustering key can be composed of multi-column
- A of parameters can be adjusted at the table level (compaction, compression, gc_grace_seconds, time to live, etc..)

Cassandra: Example Create Table

```
CREATE TABLE test.sample_table (
    par_key1 uuid,
    par_key2 uuid,
    clust_key1 timestamp,
    clust_key2 int,
    value1 text,
    value2 double,
    PRIMARY KEY ((par_key1, par_key2), clust_key1, clust_key2)
) WITH CLUSTERING ORDER BY (clust_key1 DESC, clust_key2 ASC)
```

Cassandra: What is Replication Factor

- Replication factor determines how many copies of your data are stored in the Cassandra Cluster.
- Each copy is stored in a different node.
- Replication Factor can be defined by datacenters that you've setup
- This is a parameter set at the keyspace level within the cluster.

Cassandra: What is Consistency Level

- This parameter is set by the client on individual queries
- This parameter combined with replication factor can help you achieve the consistency requirement the specific use case is looking for.
- Some of the different values are

  ONE

  LOCAL_ONE

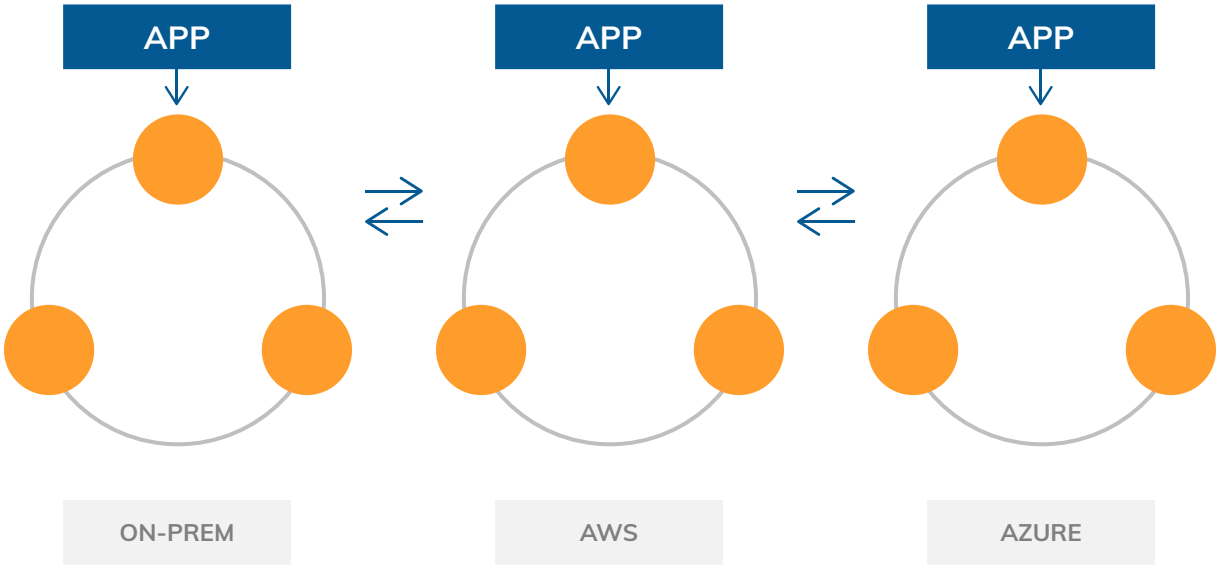
  QUORUM

  EACH_QUORUM

  LOCAL_QUORUM

  ALL

Cassandra - Read/Write in Action



Replication - 3 per DC

Consistency - Per Read/Write Request from Client

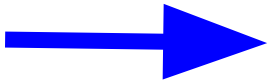
Application - Active/Active Deployment across DC for Read/Write

How can My Enterprise get from an RDBMS Based Design to Cassandra Based Architecture?

- Structured Data is the norm for both
- Re-evaluate the need for ACID transactions with Lightweight-transactions (LWT) in Cassandra
- Take advantage of Cassandra Performance
  - Move Joins to Application Stack
  - Denormalization & Data Duplication is efficient
  - Choose type of Index wisely based on Latency/TPS requirements
- Thoroughly plan the Data Model in Cassandra

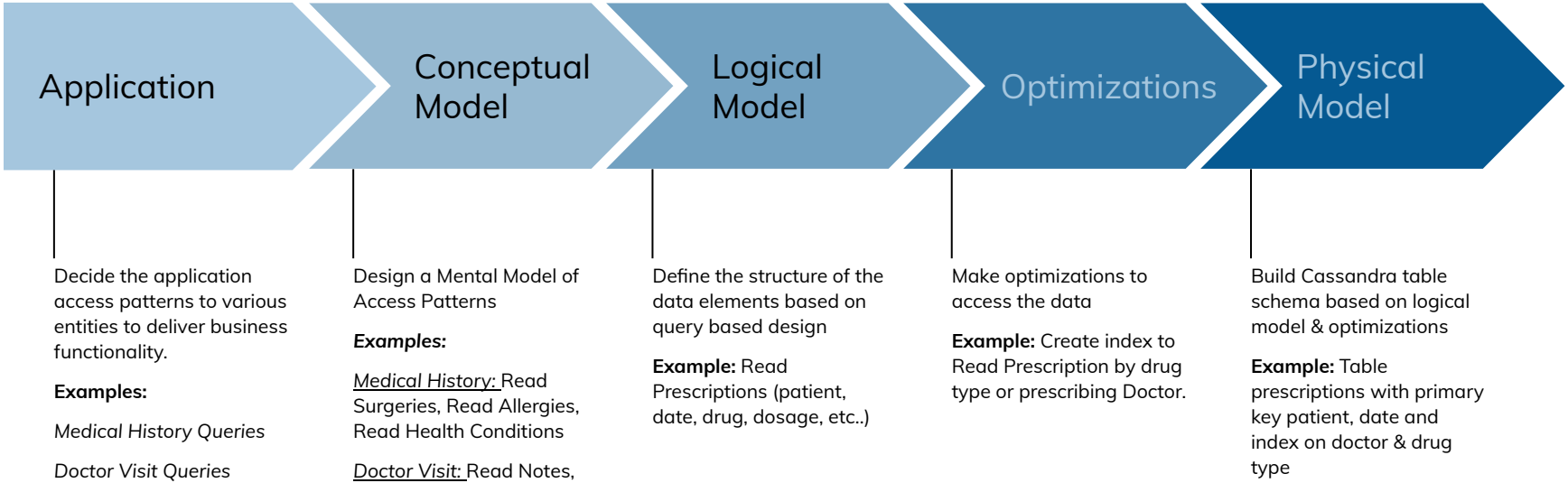
ERD to Query Based

ERD Based Design



Query Based Design

5 Steps to Query Based Design



DataStax Enterprise: Cassandra Data Platform

OUTCOMES

AI-Scale Experiences, Microservices and Insights

STRATEGIC

Kubernetes Operator (Cloud-Native Automation + Elasticity)

Developer and DevOps APIs (K8S, CQL, REST, GraphQL, gRPC)

ACCELERATED

Multi-Model Data

(All Data Styles)

Operational Analytics

(Spark, Pipelines, Streaming)

Enhanced Search

(Enhance Any Query)

Graph Engine

(Relate Data Across Partitions)

Extensible Integration

(Kafka, Elastic, Bulk Loading)

TRUSTED

Operational Reliability (Advanced Performance, Enterprise Security, Monitoring)

FOUNDATIONAL

Apache Cassandra NoSQL Database (100% Uptime, Zero-Lock-In, Global Scale)

Tools

Thought Leadership

Enterprise Support

Partnerships

OSS Commitment

DATASTAX

DataStax Astra: Cassandra Made Easy in the Cloud

Cassandra-as-a-Service

Cloud-native
Database-as-a-Service built
on Apache Cassandra

No Operations

Eliminate the overhead
to install, operate, and
scale Cassandra

Powerful APIs

Out-of-the-box REST
and GraphQL endpoints
and browser CQL shell

Cloud Native

Powered by our
open-source Kubernetes
Operator for Cassandra

Zero Lock-in

Deploy on AWS or GCP and
keep compatibility with
open-source Cassandra

10 Gig Free Tier

Launch a database in the
cloud with a few clicks,
no credit card required

DATASTAX

Use Case #1 - C&S Wholesale Grocers - Supply Chain

- Delivers over 140,000 food and non-food items to from over 50 warehouse locations
- Operates over 18 million square feet of storage
- Some of C&S's customers are Safeway, Target, Stop & Shop
- Traditional solutions slowing down distribution efficiency & impeding innovation
- Business growth leading to Technology Innovation

Use Case #1 - C&S - The Challenge

- Supply Chain Process in local RDBMS to warehouse
- Business need to consolidate warehouse data for ease of management via mobile app
- The transaction volumes were in the thousands per several seconds
- Needed real-time view of all the working parts of the manufacturing operations. Warehouse ➜ locations ➜ pallet
- Data Platform capable of operational analytics

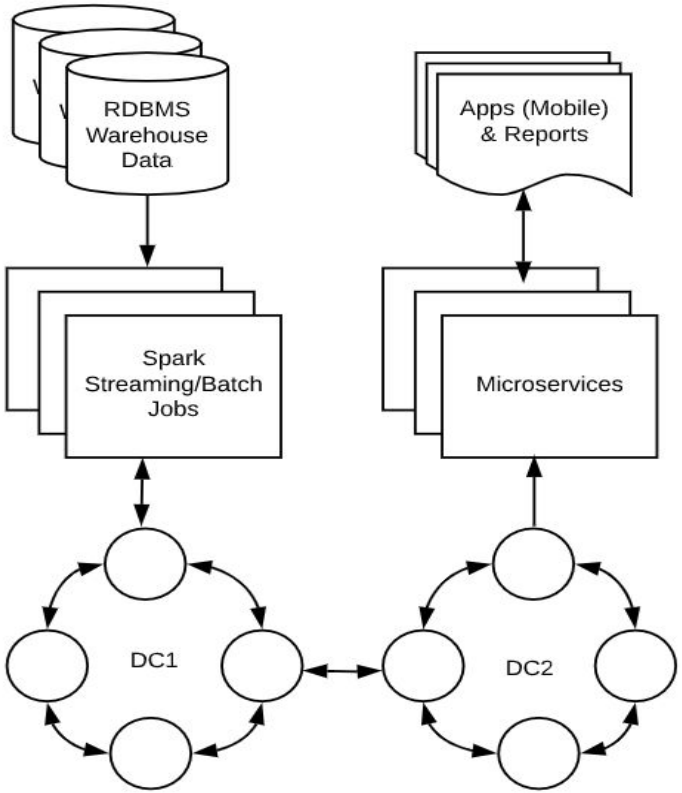
Use Case #1 - C&S - Why Cassandra?

- Scalable
- High Transaction Volume
- Low Latency
- High Availability - Warehouse operations 24/7
- Ease of Development for Microservices & Mobile App
- Multi-DC Deployment Capability
- Ease of Operational Analytics

Use Case #1 - C&S - Business Benefits

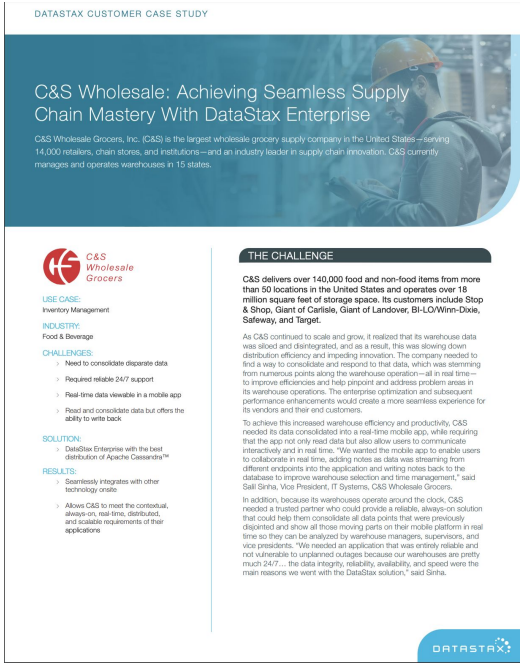
- 5 year ROI projection to save multi-millions
- Able to optimize management capabilities of consolidated warehouse operations
- Achieve remarkable efficiency in data pipeline
- Transactions - Read/Write Thousands in seconds
- Supports 300+ Users processing ~ 300k records in 5 mins

Use Case #1 - C&S - The Architecture



C&S - Case Study

> We needed an application that was entirely reliable and not vulnerable to unplanned outages because our warehouses are pretty much 24/7...



DATASTAX CUSTOMER CASE STUDY

C&S Wholesale: Achieving Seamless Supply Chain Mastery With DataStax Enterprise

C&S Wholesale Grocers, Inc. (C&S) is the largest wholesale grocery supply company in the United States—serving 14,000 retailers, chain stores, and institutions—and an industry leader in supply chain innovation. C&S currently manages and operates warehouses in 15 states.

C&S Wholesale Grocers

USE CASE:
Inventory Management

INDUSTRY:
Food & Beverage

CHALLENGES:
- Need to consolidate disparate data
- Required reliable 24/7 support
- Real-time data viewable in a mobile app
- Read and consolidate data but offers the ability to write back

SOLUTION:
- DataStax Enterprise with the best distribution of Apache Cassandra™

RESULTS:
- Seamlessly integrates with other technology onsite
- Allows C&S to meet the contextual, always-on, real-time, distributed, and scalable requirements of their applications

THE CHALLENGE

C&S delivers over 140,000 food and non-food items from more than 50 locations in the United States and operates over 18 million square feet of storage space. Its customers include Stop & Shop, Giant of Carlisle, Giant of Landover, BI-LO/Winn-Dixie, Safeway, and Target.

As C&S continued to scale and grow, it realized that its warehouse data was siloed and disintegrated, and as a result, this was slowing down distribution efficiency and impeding innovation. The company needed to find a way to consolidate and respond to that data, which was stemming from numerous points along the warehouse operation—all in real time—to improve efficiencies and help pinpoint and address problem areas in its warehouse operations. The enterprise optimization and subsequent performance enhancements would create a more seamless experience for its vendors and their end customers.

To achieve this increased warehouse efficiency and productivity, C&S needed its data consolidated into a real-time mobile app, while requiring that the app not only read data but also allow users to communicate interactively and in real time. "We wanted the mobile app to enable users to collaborate in real time, adding notes as data was streaming from different endpoints into the application and writing notes back to the database to improve warehouse selection and time management," said Salil Sinha, Vice President, IT Systems, C&S Wholesale Grocers.

In addition, because its warehouses operate around the clock, C&S needed a trusted partner who could provide a reliable, always-on solution that could help them consolidate all data points that were previously disjointed and show all those moving parts on their mobile platform in real time so they can be analyzed by warehouse managers, supervisors, and vice presidents. "We needed an application that was entirely reliable and not vulnerable to unplanned outages because our warehouses are pretty much 24/7... the data integrity, reliability, availability, and speed were the main reasons we went with the DataStax solution," said Sinha.

DATASTAX

https://www.datastax.com/resources/case-study/cs-wholesale-achieving-seamless-supply-chain-mastery-datastax-enterprise

DATASTAX

Use Case #2 - Financial Services - Mobile Banking

- Very competitive retail banking market
- Need to keep up with demand growth in digital banking
- Have high customer satisfaction rates
- Achieve efficient DR & Business Continuity Plans

Use Case #2 - Financial Services - The Challenge

- # of Transactions in RDBMS was not easily scalable
- DR was not easy
- Achieving Latency metrics was harder as volumes increased
- Downtime or poor experience would translate to customer churn

Use Case #2 - Financial Services - Why Cassandra?

- Deploy 3 DC Cluster
- Microservices Architecture
- Scale Application Stack w/ Database
- Achieve low latency SLA (<20ms on avg)
- DR Strategy was solid w/ High Availability
- Capable of processing billions of transactions per month

Some Other Common Use Cases

- Customer 360/SVOC
- Omnichannel & Global Payments
- IoT/Time Series/eCommerce Data (sensors, tick data, user interactions, shopping cart)
- Fraud Detection
- Online/Mobile Banking
- Inventory Management
- Recommendations (products & services)
- Regulatory Compliance
- Alerts & Monitoring (Credit card transactions)
- Global Payments
- Portfolio Management
- Loan Authorization
- Authentication (Mobile Logins)

Thank You!

DATASTAX®

Ankit Patel
Principal Strategy Architect @ DataStax
https://www.linkedin.com/in/ankit-p-patel