TECHX

Asia's Largest

AI & Cloud

Conference 2024

15 – 16, November 2024

Chennai Trade Center, Chennai

It's me

Santosh Hegde

Senior Director – Couchbase Inc

About Me

17+ years of Database and Distributed Systems experience at IBM Software Labs and Visa Inc. Currently leading R&D at Couchbase.

Analytics beyond relational databases

Unlocking the power of schemaless data

Ever since there have been software applications that produced data...

the next logical step...

is to ask questions of that data, get answers, and improve the business

Application Evolution

- Modern applications are powered by Web and Mobile

- They are complex and integrate multiple systems

- The data format battle has been won by JSON

- Data is more heterogeneous than ever

This application evolution brings about new challenges for Analytics

Analytics Challenges | JSON Applications with RDBMS Analytics

1. Relational Rigidity

JSON's developer flexibility does not align with traditional relational analytic structures

2. Multi-source Data

Silos of data must be consolidated from unstructured and structured origin sources

3. Slow Analytic Turnaround

Days or weeks to get changes for new metrics, queries and reports from BI / Data Warehouse team
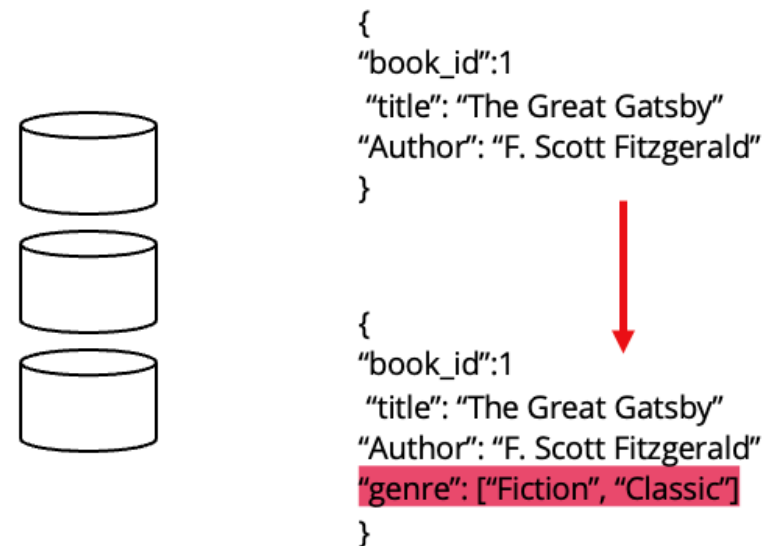
4. No Actionability

Difficult to incorporate newly learned information based on calculated metrics to improve application functionality and AI conversations

Capella Columnar is designed to overcome these challenges

Analytics Challenges | Relational Rigidity

JSON Transactional Data

```
{
"book_id":1
 "title": "The Great Gatsby"
"Author": "F. Scott Fitzgerald"
}
```

```
{
"book_id":1
 "title": "The Great Gatsby"
"Author": "F. Scott Fitzgerald"
"genre": ["Fiction", "Classic"]
}
```

Rigid Analytic Warehouse

Books		
Book_ID	Title	Author
1	The Great Gatsby	F. Scott Fitzgerald

Genres	
ID	Type
1	Fiction
2	Classic

BookToGenre	
Book ID	GenreID
1	1
1	2

JSON not intended for Analytics
- Unsupported structures:
  - Subdocuments
  - Arrays, Geo coordinates

Built for OLTP applications

Imports must match fixed schema
- Central fact table with star and snowflake-style dimensions
- Data integrity-driven

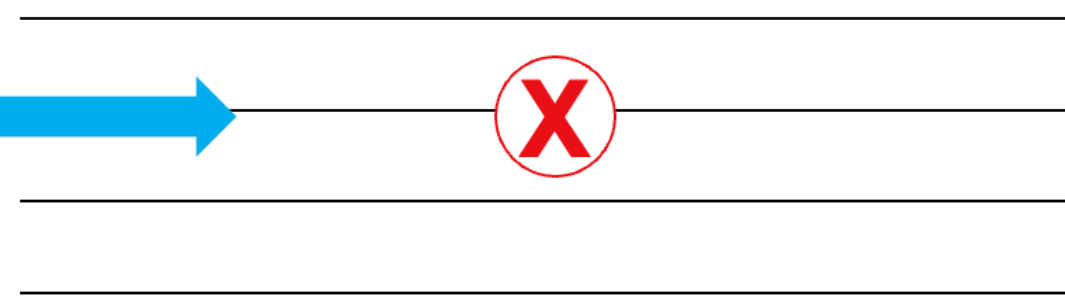**Built for OLAP analytics**

Analytics Challenges | Multi Source Data

JSON & other data sources


```
{
  "book_id":1
  "title": "The Great Gatsby"
  "Author": "F. Scott Fitzgerald"
  "genre": ["Fiction", "Classic"]
}
```

Real-time data stream

Multi format files

Multiple Ingestion Pipelines & File Imports

Rigid Data Warehouse

Books		
Book_ID	Title	Author
1	The Great Gatsby	F. Scott Fitzgerald

Genres	
ID	Type
1	Fiction
2	Classic

BookToGenre	
Book ID	GenreID
1	1
1	2

Analytics have multiple sources
- JSON (nested and heterogeneous)
- Flat files are common
- RDBMS data are common
- Application data is real-time data

ETL & data pipelines hate change
- New fields in JSON documents
- Null and missing values
- Fields of different data types
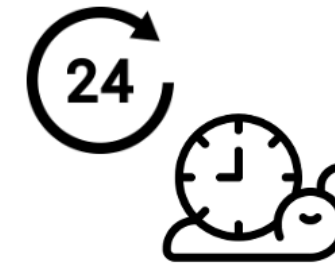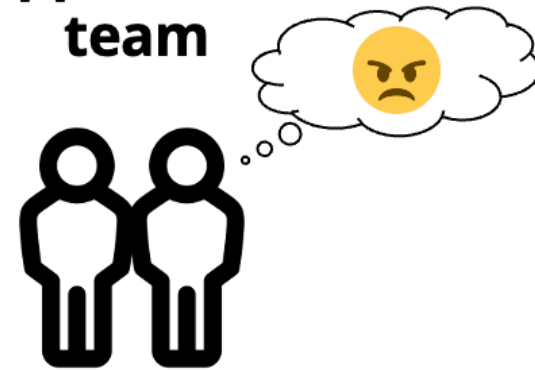- All break the pipeline

This leads to extra work
- This can require additional fields and tables to be created in the warehouse
- Breakage is common

Analytics Challenges | Slow Analytic Turnaround

Application team

- Requests for report and analytics changes must go through another team bureaucracy
(e.g., Data Engineers, BI Analysts, BI Consumers)

- Analytics team handles requests from many teams across the organization

BI/DW team

- Need to balance requests leads to time delays
- Often, they don't know the data as well as the application team

Analytics Challenges | No Actionability

NoSQL JSON
Transactional DB

Calculated Metrics & Insights

Relational
Analytical Store

- Analytic stores are not designed to writeback to transactional databases

- **Reverse ETL programs or custom code** needed to extract metrics data to the transactional database

- Often costly and time consuming to set up and maintain

Capella Columnar | Next Generation JSON Analytics

Capella Columnar

NoSQL

Schema Flexibility

High Availability
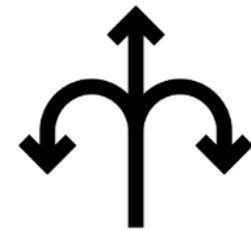
Scalability

Developer Friendliness

Analytical

Columnar Storage

Cost Based Optimizer

Massive Parallel Processing

Storage/Compute Separation

Capella Columnar | Next Generation JSON Analytics

Zero ETL for JSON

- JSON automatically processed
- Schema on Read
- Changes handled
- Transformations happen automatically

Multi-source Ingestion

- Capella & Couchbase Server
- MongoDB
- MySQL & PostgreSQL
- Files: CSV, TSV, Parquet, Avro, etc.
- Batch and streaming
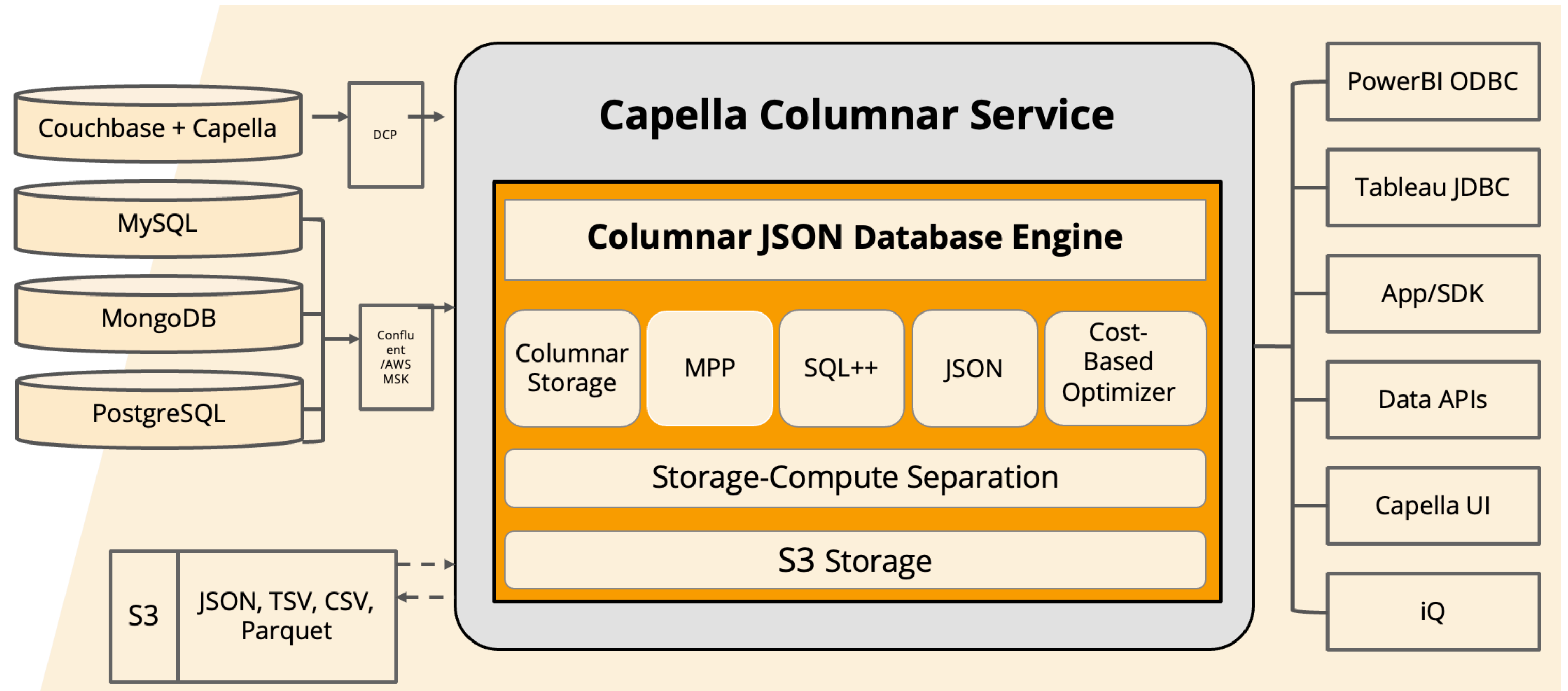
Conversational Analytics

- Developer-controlled analytics using iQ
- Capella with Columnar for operational and analytical applications

Operational Writeback

- Drive action from calculated metrics
- Feed AI prompts
- Build trustworthy AI-enabled features

Capella Columnar | Service Architecture

Capella Columnar Service

Columnar JSON Database Engine

- Columnar Storage
- MPP
- SQL++
- JSON
- Cost-Based Optimizer

Storage-Compute Separation

S3 Storage

Data sources (left):
- Couchbase + Capella → DCP
- MySQL
- MongoDB
- PostgreSQL → Confluent /AWS MSK
- S3 — JSON, TSV, CSV, Parquet

Consumers (right):
- PowerBI ODBC
- Tableau JDBC
- App/SDK
- Data APIs
- Capella UI
- iQ

Capella Columnar | What does it enable

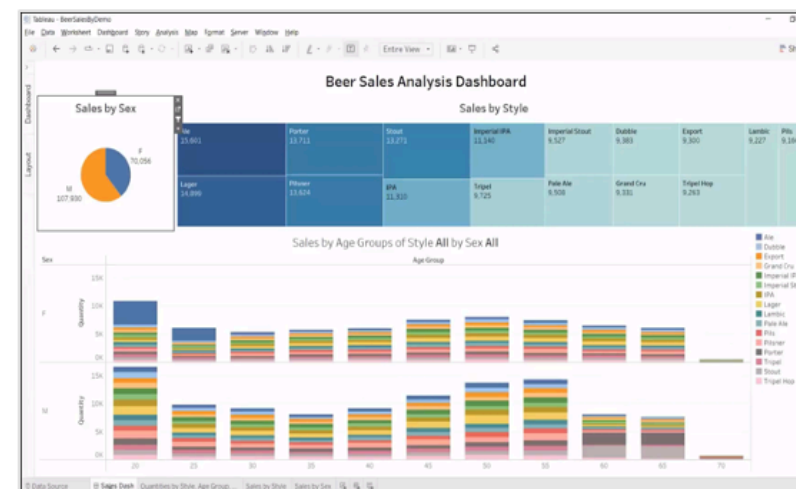## More Personalized User Experiences

- Contextualized user experiences incorporating AI
- Driving custom offers

"You've hit your saving goal, book the trip!"

Operational Analytics

- Operational visibility
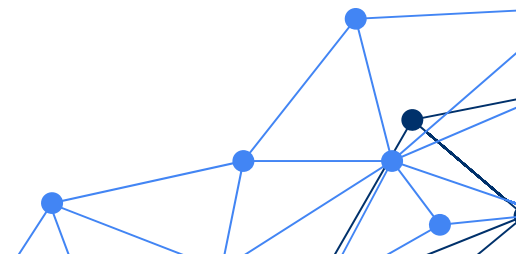- Smoother DevOps
- Faster anomaly detection

Customer-Facing Metrics

- Embedded analytic information
- For B2B or user-facing apps

Questions ??

THANK YOU!

Santosh Hegde

Email: santosh.hegde@couchbase.com
LinkedIn: santosh-hegde-cb