



## Purpose-built databases: Choose the right tool for each job

**Thiago Nogueira Dolabella**  
Principal Technical Account Manager, AWS

# Agenda

- Data requirements for modern applications
- Why purpose-built databases?
- Picking the right tool for the job
- Demo
- Conclusion

# Data requirements for modern applications

# Rapid expansion of data requirements

## Explosion of data



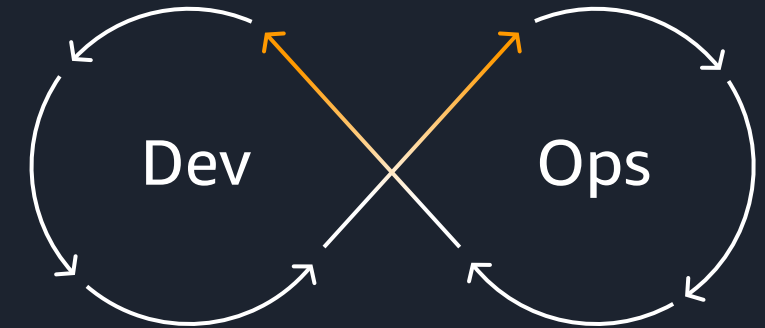
Data grows 10x every 5 years  
driven by network-  
connected smart devices

## Microservices change data and analytics requirements



Microservices architecture  
decreases need for "one size fits all"  
databases and increases need for  
real-time monitoring and analytics

## Accelerated rate of change driven by DevOps



Transition from IT to  
DevOps increases rate of  
change

# The best tool for a job usually differs by use case



## Build new applications with purpose-built databases

# Why consider purpose-built databases?



Scale



Performance



Availability

# AWS purpose-built databases



# Relational data

Divide data among tables

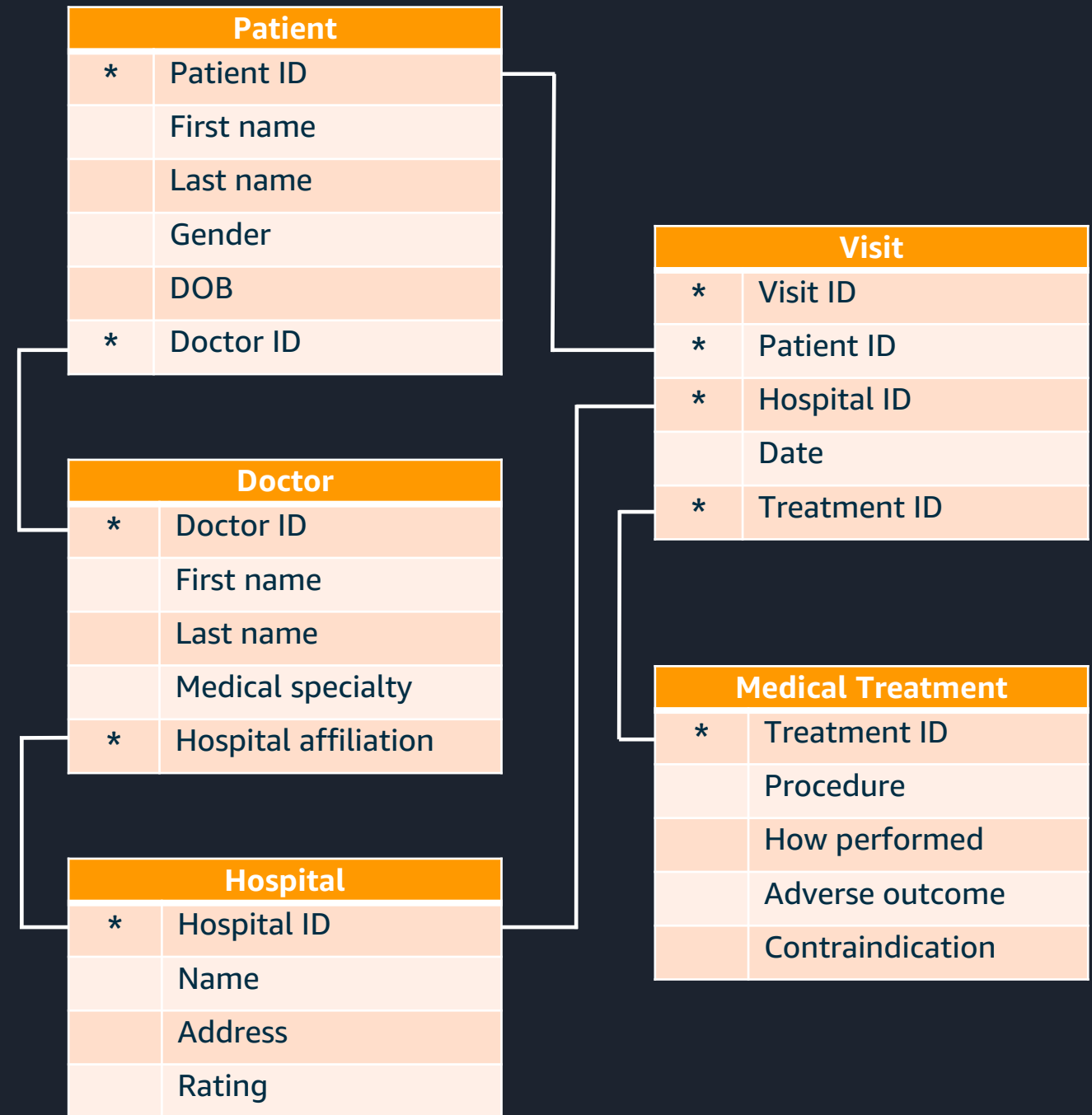
Highly structured

Relationships established via keys enforced by the system

Data accuracy and consistency

 Amazon Aurora

 Amazon RDS



# Key–value database

Simple key–value pairs  
Partitioned by keys  
Resilient to failure  
High-throughput, low-latency reads and writes  
Consistent performance at scale

```
PUT {  
  TableName: "Gamers",  
  Item: {  
    "GamerTag": "Hammer57",  
    "Level": 21,  
    "Points": 4050,  
    "Score": 483610,  
    "Plays": 1722  
  }  
}
```

```
GET {  
  TableName: "Gamers",  
  Key: {  
    "GamerTag": "Hammer57",  
    "ProjectionExpression": "Points"  
  }  
}
```

Gamers				
Primary key	Attributes			
Gamer tag	Level	Points	High score	Plays



## Amazon DynamoDB



Duolingo uses AWS databases to serve up **over 31 billion items** for 80 language courses with **high performance and scalability**

Primary database: **Amazon DynamoDB**

- 24,000 reads and 3,000 writes per second
- Personalize lessons for users taking 6B exercises per month

Transactional data: **Amazon Aurora**

- Maintain user data

# Why document databases?



Documents map naturally to how humans model data



Documents (objects/JSON) are common application data models



Document databases store JSON-like documents



Document databases provide flexible schema and indexing



Ad hoc querying and aggregations

```
#Ad hoc queries
db.col.find( { "name.last": "Smith" } )

#Count of documents
db.col.find({ product: "xyz", region: "msp" }).count()

#Return distinct fields
db.col.distinct( "region" )

#Power aggregations over documents
db.col.aggregate([
  { $match: { product: "xyz" } },
  { $group: { _id: "$c_id", rev: { $sum: "$price" } } },
  { $sort: { rev: -1 } }
])
```



## Amazon DocumentDB



FINRA is a not-for-profit organization authorized by the U.S. Congress to protect investors and ensure market integrity

### Challenge:

FINRA's existing solution to store millions of documents as XML in a relational database was consuming too much storage, required custom tooling, and was difficult to manage

### Solution:

**Amazon DocumentDB (with MongoDB compatibility)** as a managed JSON document store, making it simpler to query and index regulatory documents, reduce development cycles, and extend usability of data



# In-memory databases: usage patterns



Caching



Real-time  
analytics store



Gaming  
leaderboards



Geospatial



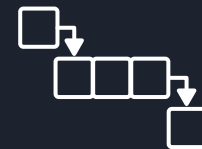
Media  
streaming store



Session  
store



Chat apps  
pub/sub



Job  
queue

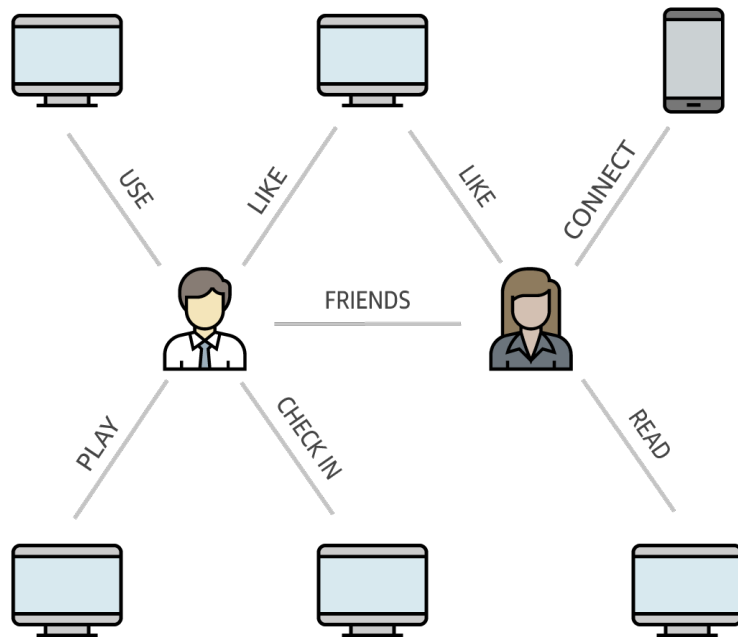


Machine learning  
real-time model scoring



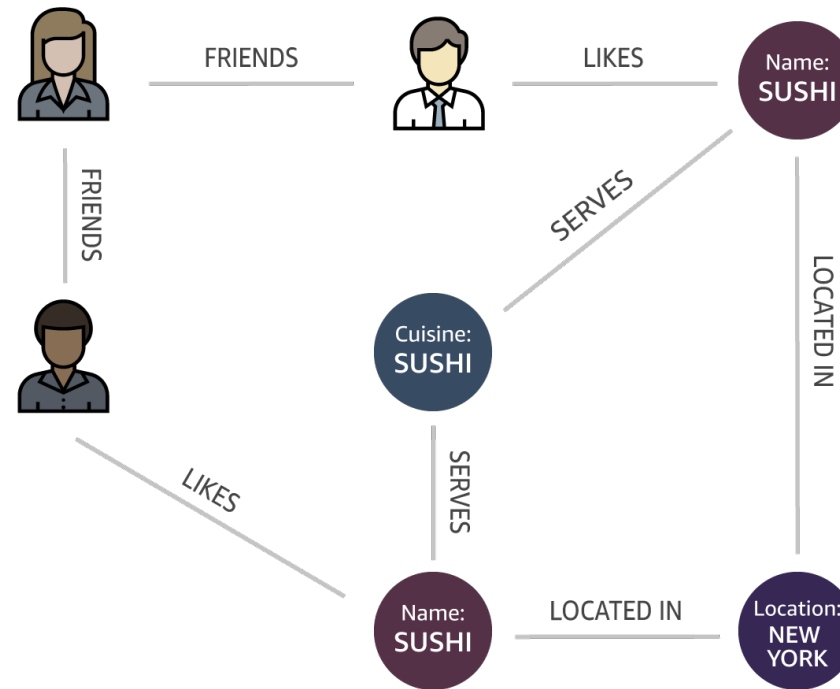
## Amazon ElastiCache

# Graph database: usage patterns



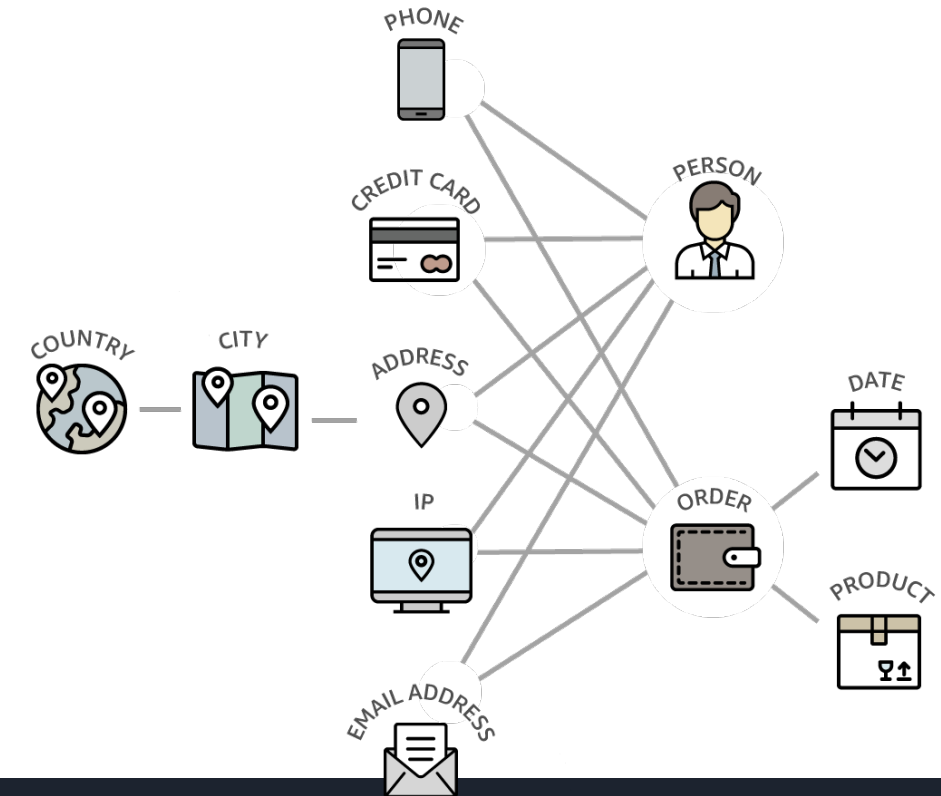
## Device and Social Graphs

Manage complex relationships between large datasets with billions of relationships



## Personalization & Recommendations

Easily navigate entities connected with many-to-many relationships



## Pattern Detection

Calculate strength, weight or quality of relationships

# Amazon Neptune - Graph database for highly connected data



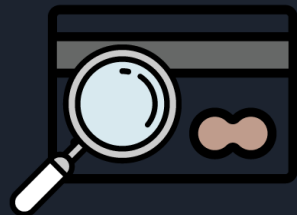
Social networking



Recommendations



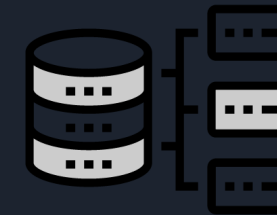
Knowledge graphs



Fraud detection



Life sciences

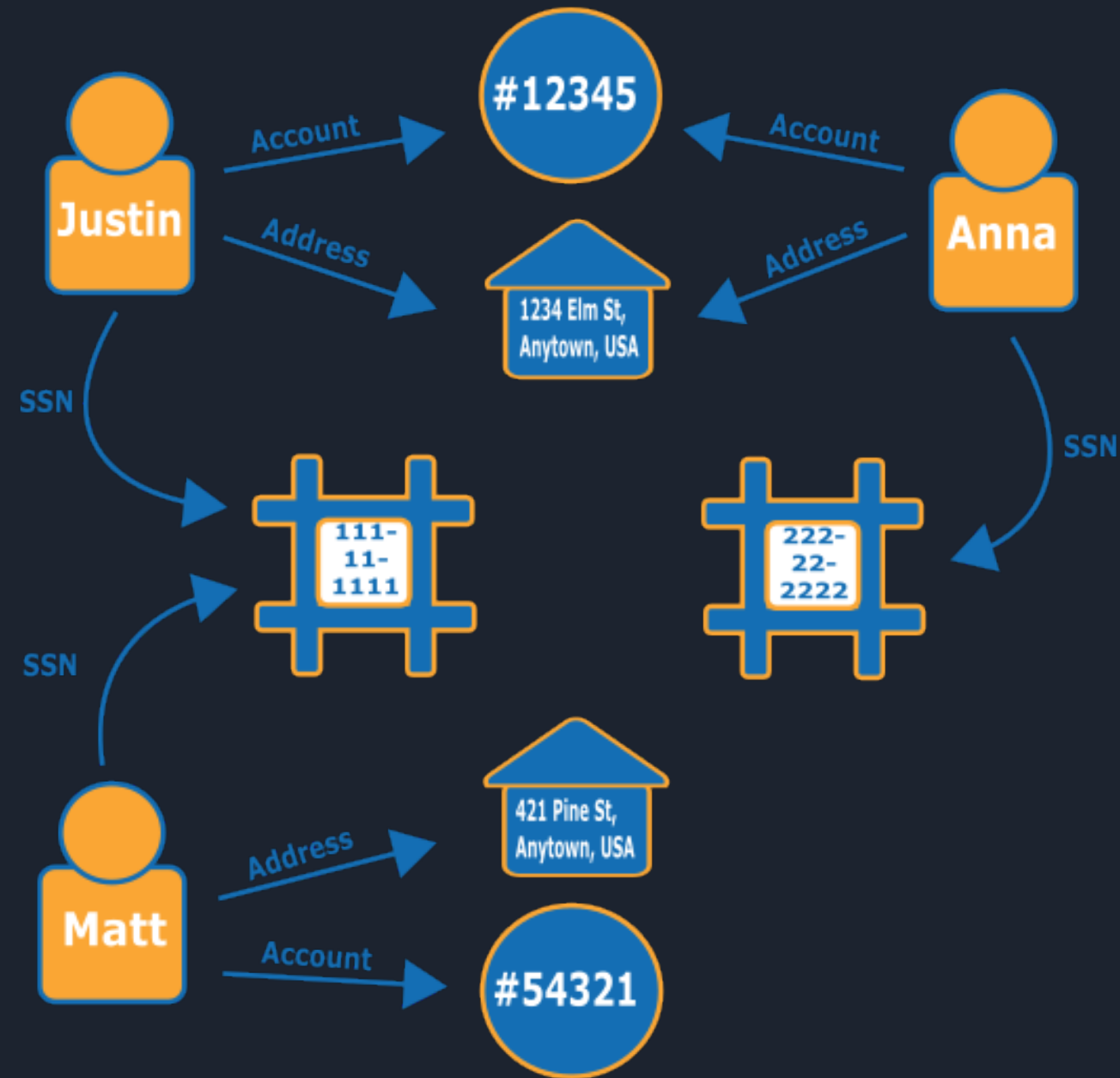


Network and IT operations

 **Amazon Neptune**

# Graph database use case example

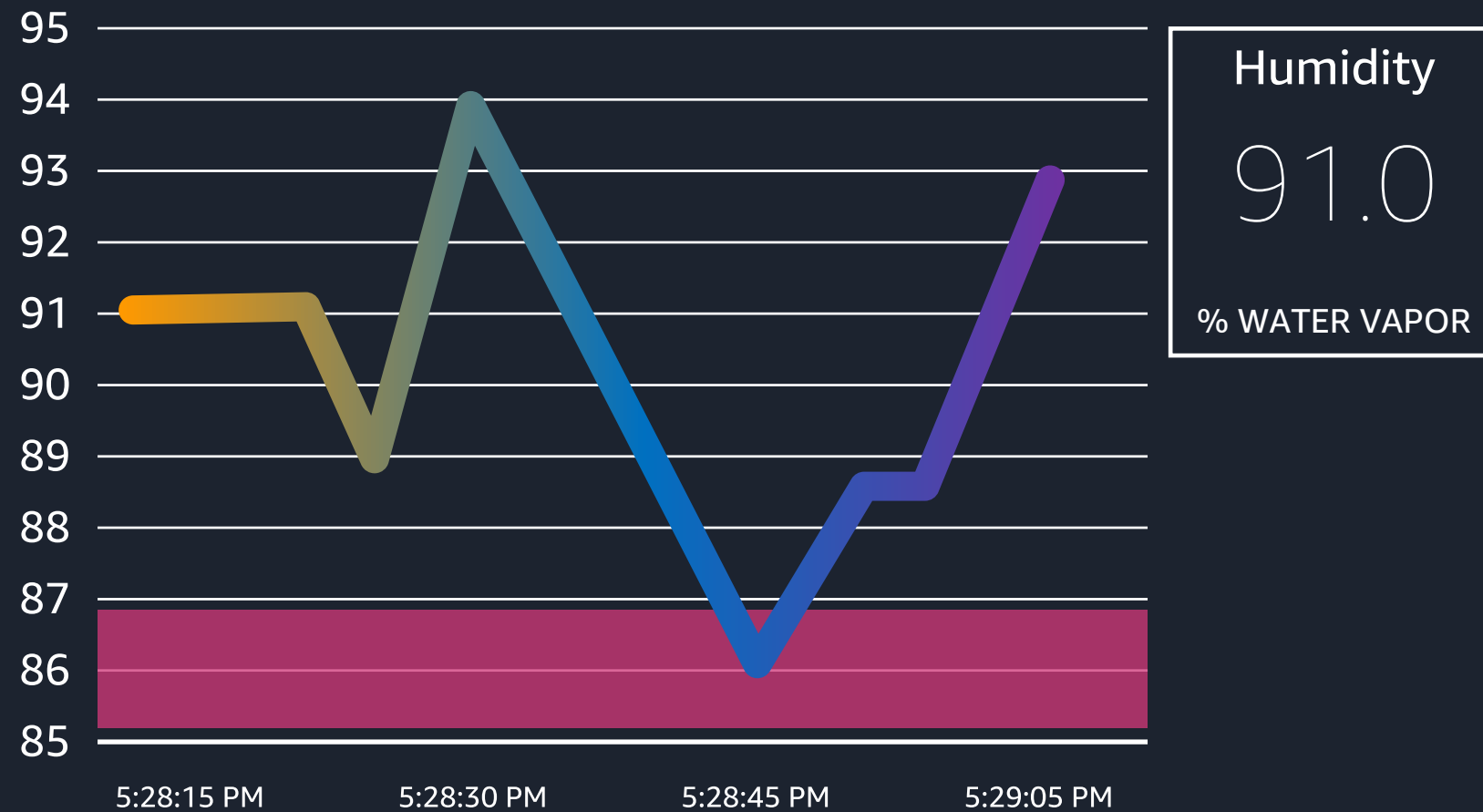
 Amazon Neptune



# Demo

 Amazon Neptune

# Time series use cases



① Application events

② IoT sensor readings

③ DevOps data

 Amazon  
Timestream

# Ledger databases



## Banking and finance

Keeping track of transactions, trades, and accounts



## Manufacturing

Recording components used in manufacturing



## Ownership

Maintaining records of asset ownership

Amazon  
 **Quantum Ledger Database**

# Wide column: Apache Cassandra

- Open-source, wide-column data store
- Large scale applications that require fast read and write performance
- Use cases:
  - User profiles
  - Device metadata
  - Time-series data
  - Transaction logging
- Cassandra Query Language (CQL)



Amazon Keyspaces (for Apache Cassandra)

# Benefits of purpose-built databases



Better  
performance



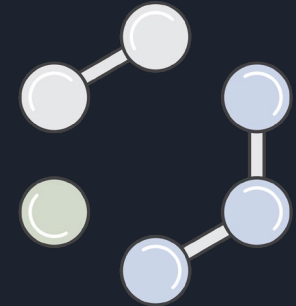
Better scale



More  
functionality



Easier to  
debug



Independence  
between teams

# AWS Training and Certification



## Training for the Whole Team

Explore tailored Data or Database learning paths for customers and partners



## Flexibility to Learn Your Way

Build cloud skills with free digital Data training courses such as "The elements of Data Science", or dive deep with classroom training



## Validate Skills with AWS Certification

Demonstrate expertise with a Data industry-recognized credential (Data analytics and Database Specialty AWS Certifications)

[aws.amazon.com/training/](https://aws.amazon.com/training/)

# Visit the Data, Databases, and Analytics Resource Hub for more resources

Dive deeper with these newly created whitepapers and e-books to help you uncover new insights and value from your data

- An introduction to cloud databases
- Enter the purpose-built database era
- Harness the power of data
- Creating a modern analytics architecture
- The data-driven enterprise
- ... and more!









<https://tinyurl.com/aws-data-databases-analytics>

**Visit resource hub »**

# Thank you for attending AWS Data, Databases, and Analytics Online Series

We hope you found it interesting! A kind reminder to **complete the survey**.  
Let us know what you thought of today's event and how we can improve the event experience for you in the future.

-  [aws-apac-marketing@amazon.com](mailto:aws-apac-marketing@amazon.com)
-  [twitter.com/AWSCloud](https://twitter.com/AWSCloud)
-  [facebook.com/AmazonWebServices](https://facebook.com/AmazonWebServices)
-  [youtube.com/user/AmazonWebServices](https://youtube.com/user/AmazonWebServices)
-  [slideshare.net/AmazonWebServices](https://slideshare.net/AmazonWebServices)
-  [twitch.tv/aws](https://twitch.tv/aws)

# Thank you!