Key-Value Stores

Chandreyi Chakraborty, Emmanuel Etukudoh, Rebecca Evans, Navin Senthil, Sebastian Stephens

CS 4440

March 3, 2025

Image Sources: www.slidegeeks.com

Agenda

- Introduction
- Product Overview
- Technical Details
- Sample Applications
- Market Analysis
- Future Trends
- Research References
- Questions

Introduction

- What are Key-Value Stores?
- Key Terms
  - Key
  - Value
  - NoSQL
- Why use Key-Value Stores?
  - High-Speed Data Access
  - Scalability
  - Simplicity
- Example Business Use Cases

Image Sources: www.chatgpt.com

Features/Functionalities

- Schema-less design
- Simple data model
- CRUD using keys
- Supports complex objects
  - arrays, nested dictionaries, images, videos, etc.
- Sorted keys
  - Used to iterate over keys efficiently
- Secondary Key Support
  - Supports multiple keys
- Replication and Partitioning
- ACID Support

	Type	Performance	Scalability	Persistence	Use Cases	Consistence	Cost
Dynamo	NoSQL database service	Highly scalable, low latency reads and writes	Automatically scales	Supports w/ configurable options	Web/Mobile applications, gaming	Strong or eventual	Pay-as-you-go
Redis	In-memory data store and cache	Fast due to in-memory storage	Limited by available memory	Supports but primarily in memory	Caching, real-time analytics, session store	Eventual	Expensive for large datasets
Couchbase	NoSQL database service	Low latency, in-memory caching	Horizontally scalable	Durable writes, in-memory + disk storage	Real-time apps, caching, session management	Tunable	Priced per node

Amazon DynamoDB

- Intro
- Architecture
- Use Cases
- Technical Differentiators

Image Sources: www.medium.com/@syedshamail3/understanding-the-internal-architecture-of-amazon-dynamodb-e9dccb8be8de

Amazon DynamoDB Architecture

- Data Model
  - Tables
  - Items
  - Attributes
  - Primary Keys
- Storage and Partitioning
  - Automatically Partitioned
  - Multiple AZs
- Indexing
  - GSI: Global Secondary Indexes
  - ISI: Local Secondary Indexes

Amazon DynamoDB Use Cases

- Web and Mobile Applications
- E-commerce and Retail
- IoT and Real Time Analytics
- Serverless Applications
- Financial Banking Services

DynamoDB Technical Differentiation

- Fully Managed and Serverless
- Predictable Performance
- High Availability and Durability
- Security and Compliance
- AWS Integration

Redis

- Intro
- Architecture
- Use Cases
- Technical Differentiators

Image Sources: https://iconduck.com/icons/13186/redis-original-wordmark

Redis Architecture

- Data Model
  - Strings/Lists
  - Sets/Ordered Sets
  - Hashes
  - Other data structures
- Memory Storage and Persistence
  - RDB(Redis Database Backup)
  - AOF(Append-Only File)
- Clustering/ Performance Optimization
  - Redis Sentinel/Cluster and Replication
  - Pipelining/Memory Eviction

Redis Use Cases

- Caching
- Real-Time Analytics
- Message Queues
- Session Management
- ML/AI
- Gaming

DynamoDB Technical Differentiation

- High-Performance In-Memory Storage
- Flexible Data Structures
- High Availability and Scalability
- Pub/Sub Messaging
- Persistence and Durability Options

Sample Applications: Snapchat

- *Existing systems struggling with scalability and performance during peak traffic.*
- *Explored high throughput and low latency AWS DynamoDB.*
- *Offload operational burden of managing the database.*
- *Cost benefits with pay as you go pricing and on demand scaling.*
- *Use cases: messaging and friend graph.*

Image Sources: www.snapchat.com, www.scalefactory.com/blog/2022/01/18/a-decade-of-dynamodb/

“As the number of Snapchatters grew to hundreds of millions, we did some technical diligence and found that Amazon DynamoDB was the right technology choice for us.”

- Saral Jain, Snapchat Director of Engineering and Head of Infrastructure.

Sample Applications: Prime Video

- *Run user watchlist and bookmarking services.*
- *Global databases provides users who don't live together with consistent application experience.*
- *Table below represents a possible data model that supports watchlists.*

Primary key		Attributes			
Partition key: userId	Sort key: listId#titleId				
user123	list1234#metadata	listId	listMetadata		
		list123	Data		
	list123#title0564	titleMetadata1	titleMetadata2	titleId	listParent
		Data	Data	Melodramatic Drama	list123
	list123#title1523	titleMetadata1	titleMetadata2	titleId	listParent
		Data	Data	Generic Action Movie	list123
	list123#title8431	titleMetadata1	titleMetadata2	titleId	listParent
		Data	Data	Predictable RomCom	list123
	list234#metadata	listId	listMetadata		
		list234	Data		
	list234#title 5781	titleMetadata1	titleMetadata2	titleId	listParent
		Data	Data	Campy Horror	list234

Image Sources: https://aws.amazon.com/blogs/database/amazon-dynamodb-use-cases-for-media-and-entertainment-customers/, www.scalefactory.com/blog/2022/01/18/a-decade-of-dynamodb/

Market Analysis

- *Table from IDC's "The Business Value of Amazon DynamoDB" shows US based organizations using DynamoDB*

TABLE 1

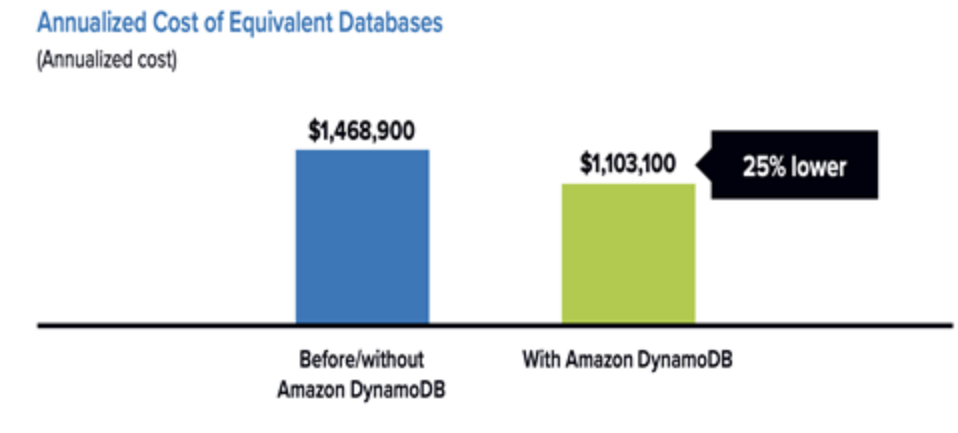
Demographics of Interviewed Organizations

	Average	Median
Number of employees	16,450	5,250
Number of IT staff	2,416	250
Number of business applications	140	100
Annual revenue	$9.92B	$904.00M
Countries	United States	
Industries	Financial services/banking (3), biotechnology, healthcare, IT services, marketing	

Image Sources: Business Value White Paper, sponsored by Amazon Web Services, Inc. May 2024 | IDC #US51795724

Market Analysis

- *Figures from IDC's "The Business Value of Amazon DynamoDB" demonstrates the benefits and savings from using Amazon DynamoDB.*



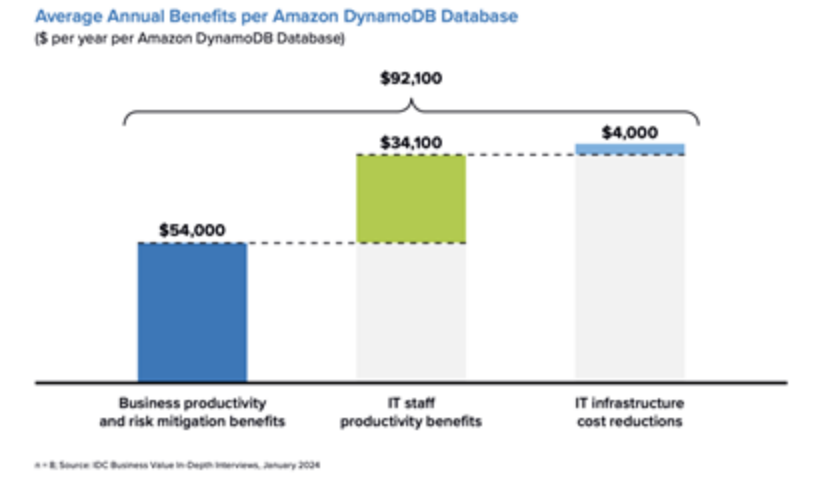


Image Sources: Business Value White Paper, sponsored by Amazon Web Services, Inc. May 2024 | IDC #US51795724

Market Analysis

- *Demonstrates rapid growth in Snapchat revenue and need for scalability.*

	Year Ended December 31,		
	2017	2016	2015
	(in thousands, except per share amounts)		
Consolidated Statements of Operations Data:			
Revenue	$ 824,949	$ 404,482	$ 58,663
Costs and expenses:			
Cost of revenue	717,462	451,660	182,341
Research and development	1,534,863	183,676	82,235
Sales and marketing	522,605	124,371	27,216
General and administrative	1,535,595	165,160	148,600
Total costs and expenses	4,310,525	924,867	440,392
Loss from operations	(3,485,576)	(520,385)	(381,729)
Interest income	21,096	4,654	1,399

Image Sources: https://www.sec.gov/Archives/edgar/data/1564408/000156459018002721/snap-10k_20171231.htm

Future Trends - Cloud Computing

- Performance and persistence breakthrough
  - SSD-based storage instead of block storage
- Edge Computing
  - RedisEdge and Badger
- Serverless and Cloud-Native (Since 2010s)
  - Fully managed and elastic for commercial use

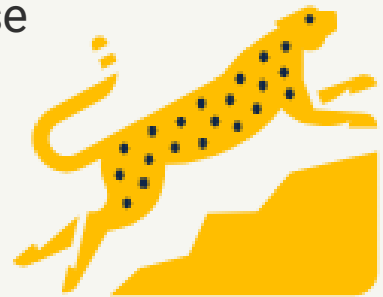


Image Sources: www.rocksdb.org

Future Trends - AI Integration Direction

- Multi-model support
  - Graph, vector, time-series data
- Semantic Search at Scale
  - NeuralDB
- Recommendation Systems
  - E-Commerce, Social Media

Image Sources: www.linkurious.com/blog/visualize-cosmos-db-graph-linkurious/

Future Trends - Challenges

- Data Security
  - Homomorphic Encryption
    - Allows for computations on encrypted data
  - Confidential Computing
    - Isolates and encrypts data in use
- Consistency vs. Performance Trade-offs
- Energy Efficiency and Sustainability

Research Reference 1

Dynamo: Amazon’s Highly Available Key-Value Store

- Overview
  - Introduction to DynamoDB as Amazon’s distributed key-value storage system
  - Designed for high-reliability services with control over availability, consistency, cost, and performance
  - Originated to support Amazon’s e-commerce platform
- Key Contributions
  - Highly Available Storage System
  - Evaluation of tradeoffs
  - Real-World Applications
- Relevance
  - Critical to Amazon’s infrastructure and e-commerce operations
  - Adopted by other major platforms, including Disney and Snap Inc.

Image Sources: https://abiabi0707.medium.com/overview-of-aws-dynamodb-844bf943f209

Research Reference 2

Supply of a key value database redis in-memory by data from a relational database

- Overview
  - Key-value databases provide a flexible, schema-less alternative to relational databases.
  - Optimized for caching and fast data access with minimal overhead.
- Key Contributions
  - Developing an automated approach for migrating relational databases to Redis.
  - Redis enhances performance with in-memory storage and advanced data structures.
- Relevance
  - Redis is widely used for caching, real-time analytics, and session storage in large-scale systems.

Image Sources: https://iconduck.com/icons/13186/redis-original-wordmark

Thank You

Cited Sources

- *Giuseppe DeCandia, Deniz Hastorun, Madan Jampani, Gunavardhan Kakulapati, Avinash Lakshman, Alex Pilchin, Swaminathan Sivasubramanian, Peter Vosshall, and Werner Vogels. “Dynamo: Amazon’s Highly Available Key-Value Store,” in Proceedings of the 21st ACM Symposium on Operating Systems Principles (SOSP), 2007, pp. 205–220.*
- *Michael Stonebraker and Andrew Pavlo. “What Goes Around Comes Around... And Around...” in SIGMOD Record, Vol. 53, No. 2, June 2024, pp. 21–37.*
- *MongoDB - “What is a Key Value Database?” https://www.mongodb.com/resources/basics/databases/key-value-database*
- *A. E. Alami, M. Bahaj and Y. Khourdifi, "Supply of a key value database redis in-memory by data from a relational database," 2018 19th IEEE Mediterranean Electrotechnical Conference (MELECON), Marrakech, Morocco, 2018, pp. 46-51, doi: 10.1109/MELCON.2018.8379066.*
- *Huang, Feixiong, et al. The Journal of Supercomputing, vol. 81, no. 1, 22 Oct. 2024, doi:10.1007/s11227-024-06526-7.*
- *Thorne, James, et al. “Database Reasoning Over Text.” arXiv.Org, 2 June 2021, arxiv.org/abs/2106.01074. Accessed 03 Mar. 2025.*