

Title: YouTube Data Ecosystem.

Scope & Objective: We'll design a comprehensive RDMS to model YouTube's data intricacies and manage retrieval efficiently, followed by translating it to a Cassandra database with adjustments for large-scale handling and a query-first design. A performance analysis will then highlight the differences and advantages between the two databases in terms of efficiency, scalability, and maintenance.

Translation process: The translation process from RDBMS to Cassandra involves defining and normalizing the RDMS schema, adapting it to Cassandra's denormalized structure, migrating the data with consistency, and converting SQL queries to Cassandra's query language, ensuring data integrity and optimized access patterns.

Potential Real-World Implications: This process will demonstrate how platforms like YouTube can efficiently manage vast data via Cassandra for scalability and fault tolerance, offering insights into complex data system management and serving as a valuable educational resource on database design and migration.

Data Reference:

We are using a combination of some dummy data and some original Youtube data from below source:

- <https://data.world/popculture/donald-glovers-this-is-america-youtube-comments>