NFT

Monday, 15 September 2025 7:07 PM

Choosing DB:

- **Caching Solutions**: Caching is essential for improving performance by reducing database queries. Common caching options include Redis, Memcached, and Hazelcast, with Redis being notably popular.

1. **File Storage**: For storing images and videos, Amazon S3 is recommended for blob storage, accompanied by the use of Content Delivery Networks (CDNs) for efficient content distribution.
    a. CDN : Stores cached versions of content in multiple servers located closer to users
2. **Text Search Functionality**: Tools like Elastic Search and Solr are highlighted for their capabilities in implementing advanced search features, including fuzzy searches for correcting query misspellings.
    a. **Fuzzy Logic** is a form of logic that deals with reasoning that is approximate rather than fixed and exact. Unlike traditional Boolean logic (true/false), fuzzy logic allows values to be partially true.
    b. A **Text Search Engine** is a system designed to find relevant text information from a collection of documents, databases, or the web.
    c. Elastic search and solr are not DB, its search engine, here there can be data lost so we should keep this as secondary source not primary source of data
3. **Time Series Databases**: Applications that track data over time can benefit from databases like InfluxDB and OpenTSDB.
    a. Metric based which are not random read or write, for instance t1 should be > t2, when appending data it should be for all the records within a time range not any random record
4. **Data Warehousing**: For analytics, solutions like Hadoop are suggested for offline reporting purposes.
    a. These are generally used for reporting not for transactional use
5. **Structure vs Un-structure Data**: A flowchart is provided to help determine whether to use relational or non-relational databases based on the data structure and transactional needs.
    a. Structured - mysql, oracle, postgres
    b. Unstructured :
        i. Lot of data and wide verity of queries - document DB like Mongo, couch base, Each record = a “document” (like a JSON object) with key-value pairs, arrays, nested objects.
        ii. **Less queries but more number of rows - Columnar Database** stores data by columns instead of rows. - Casandra, Hbase, Bigquery
6. **Handling Growing Data**: Columnar databases like Cassandra and HBase are recommended for scenarios with high data volume but limited query types.
7. **Combining Database Types**: Emphasis is placed on the necessity to use multiple database types to address diverse functional and non-functional requirements effectively, illustrated through examples from systems like Amazon.