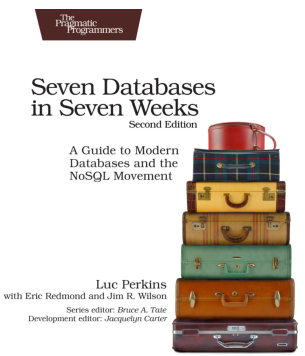
UNIVERSITY OF MARYLAND

UMD DATA605 - Big Data Systems

6.4: CouchDB

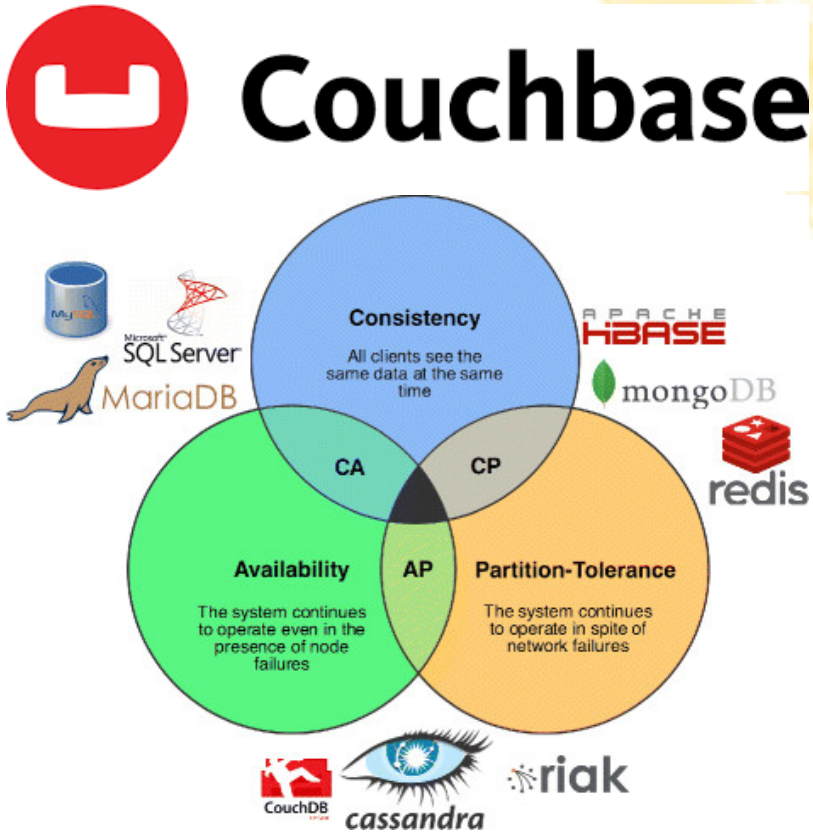


Instructor: Dr. GP Saggese - gsaggese@umd.edu

- **References**:
  - All concepts in slides
  - Seven Databases in Seven Weeks, 2e

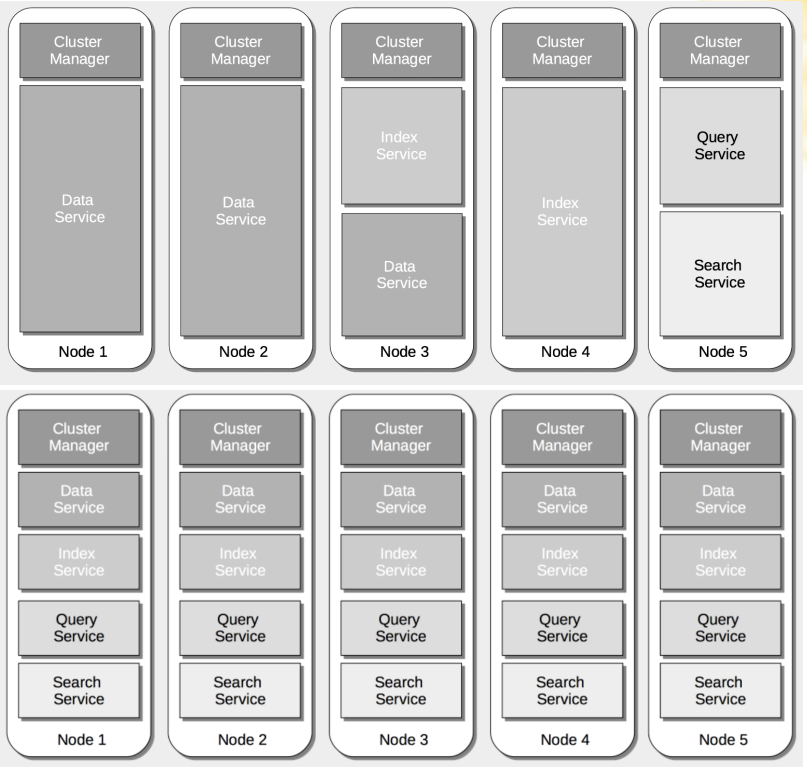
Couchbase

- NoSQL document-oriented DB (like MongoDB)
- Couchbase = merge of CouchDB and membase
  - *CouchDB*
    - Open source document store
    - HTTP RESTful API to add, update, delete documents
    - Supports all 4 ACID properties
  - *membase*
    - Distributed key-value store (like Redis)
    - Scales up and down
    - Highly available and partition tolerant
  - Uses HTTP protocol to query and interact with objects
    - No query language
  - Objects stored in *buckets*
    - Collection of JSON docs, no special relation
- From CAP perspective:



Architecture

- Every Couchbase node consists of **different services**:
  - Data service
  - Index service
  - Query service
  - Cluster manager component
- Services can run on separate nodes
- **Data replication**
  - Across nodes
  - Across data centers
- **Data service**
  - Writes data asynchronously to disk after acknowledging to client
  - Optionally synchronous: ensure data is written to more than one server before acknowledging a write



Queries

- **Can create multiple views over documents**
  - Views optimized/indexed by Couchbase for fast queries
  - Re-indexed when documents change
  - Perform full-text searches using indexes
- **Perform well when:**
  - Infrequent changes to document structure
  - Know query types in advance
- **Query**
  - Uses custom query language N1QL (“nickel”)
  - Extends SQL to JSON documents
  - Queries over multiple documents using server-side joins
- **Map-reduce support**
  - (Map) Define a view with document columns of interest
  - (Reduce) Optionally define aggregate functions over data