NoSQL (examples)

Sources:
Pramod J. Sadalage and Martin Fowler
Types of NoSQL

- Method for storing/retrieving data in a non-tabular relation format.
  - Graphs
  - Key/Value
  - Documents
Graph NoSQL DBs

- Good for representing networks (social/industrial)
- Query graphs DBs with traversals.
- Repeated traversals indexed for optimization.
Graph NoSQL DBs

Examples:

- Neo4j
- DEX
- Infinite Graph
- FlockDB
Key/Value NoSQL Store

- Way for storing data in a schema-less way.
- Key created for each record.
- Bin created for each piece of data.
- Each bin has key/value.
- No data - no bin
Key/Value NoSQL Store

Examples:

- Aerospike
- LevelDB
- Tarantool
Document style NoSQL DBs

• Notion of storing information in a document.
• Data is encapsulated.
• Can be stored in XML/JSON/BSON
• Easy to map objects in code to documents in a DB.
Document style NoSQL DBs
MongoDB (Cross platform)

- Different data types stored in different collections (like tables)
- Stores data in JSON-like format.
- Extensible queries (and, or, in, not, regex, etc…)
- Data very object-like
- Can have nested data
- No joins
- No transactions
Document style NoSQL DBs
RavenDB (.NET)

- Much like Mongo, documents stored in collections
- Collections are indexed to increase performance
- Query language is lucene
- Client API or RESTful requests
- Comes with built in RavenStudio
- Is said to fully supports ACID
Advantages to using NoSQL

- Queries are very fast
- Less overhead (easier to deploy)
- No need to develop schemas (flexible)
- Cost effective and mainly open source
- Data can be easier to visualize
- More object-oriented
- Scale out instead of up
Disadvantages of NoSQL

- ACID not ensured
- Query languages vary
- Usually doesn’t support joins
- Very narrow focus (mainly data storage)
- Performance > Consistency
- Lack of maturity compared to relational databases
Data predictions
Data predictions

Data Growth
Driven by Unstructured Data

80 Exabytes
Growing 50% Per Year

300 Exabytes
88% Unstructured Data

12% Structured Data

2013
2015

Storage Options
Backup & Archive
Application
Database
Backup

Source: ESG Digital Archive Market Forecast
Data (side issue)

Economics of Tiered Storage
Tape is the Foundation: Most of the Data Stored at the Lowest Cost

- 80% Archival (>90 days): Tape
  - Very infrequent / no changes
  - Offsite / offline / nearline protected

- 15% Recent (7-90 days): SAS/SATA
  - Infrequent changes
  - Modifications change classification to “Current”
  - Slight access delay acceptable

- 5% Current (<7 days): Flash/Performance Disk
  - Frequent changes
  - Immediate access
  - Instantaneous protection

ORACLE
Enterprise dilemma

- As enterprises grow, their data needs also grow.
- Continued dealings with extremely sensitive data (financial, commerce, etc.).
- Need the ability to scale and maintain ACIDity.
- SQL can be expensive to scale.
“NewSQL” (?)

- An answer to the enterprise dilemma.
- Emerging middle ground between SQL and NoSQL.
- Main types
  - New architecture
    - General-purpose
    - In-memory
  - SQL Engines
  - Transparent Sharding