

# NoSQL Overview + Elasticsearch Quick Dive

---

October 10, 2017  
Christopher Thielen

Credit: Wikipedia, [egghead.io](http://egghead.io) (Will Button)  
No copyright infringement intended.

# NoSQL is a Misleading Name

---

- NoSQL != anti-SQL
- “No SQL” as in “databases that don’t implement SQL”, not “databases that are better/cooler/newer than SQL” (some NoSQL databases are quite old)
- “Not Only SQL” is a good way to think of it, e.g. complements SQL engines
  - SQL + Elasticsearch is a good option at UCD

# History

---

- Modern NoSQL movement (and “NoSQL” terminology) began around 1998 with full steam ~2010.
- Necessitated by architectural requirements of “Web 2.0”
  - Replication (geography) in SQL is hard
  - SQL performance is hit-and-miss, flexibility very difficult (e.g. migration frameworks)
  - What if we think outside the box?

# What's Inside the Box?

---

- Tabular Data
- ACID

# What's Inside the Box? Tabular Data

---

- Is your data really tabular?
  - Social networks
  - Pattern detection (fraud, product recommendations)
  - Lists (sessions, logs)
  - Schema-less data

# What's Inside the Box? ACID

---

- ACID
  - Atomicity: Transactions are all or nothing, no partial writes
  - Consistency: All transactions start/end at valid states (with regard to triggers, constraints, etc.)
  - Isolation: Concurrent transactions behave like serial transactions
  - Durability: Power loss, crash does not affect completed transactions
- Do we always need ACID? e.g. what about “good enough” consistency?

# Types of NoSQL Databases

---

- Key-value store (e.g. Redis)
- Document store (values are documents)
- Graph databases (e.g. Neo4j)
- Object databases (think Hibernate without the SQL translation)
- Multi-model (why choose one?)

# Pitfalls to NoSQL Adoption

---

- SQL is massively popular and “good enough”
- NoSQL databases are all different with no standard query language
- Fundamental shift in modeling:
  - List of students, list of grades but no joins. How do I find out a student’s grades? Or all students with a 4.0 GPA?



# Advantages to NoSQL Adoption

---

- Super fast queries, often less code
- Very flexible (works well with agile)
- Easier to learn/use than SQL (see certain Banner views)
- Possibly cheaper, less disk space, etc.

Elasticsearch

“The greatest thing since sliced bread.”

**-Scott Kirkland on Elasticsearch**

# Elasticsearch

---

- Full-text search engine usable as a NoSQL data store
- HTTP verb interface (GET, POST, PUT) often using JSON
- Easily scalable, self-balancing (if you need it)
- Free, open source, based on Apache Lucene
- Self-host, AWS, Azure, etc.

# Elasticsearch Schema

---

- Index - a container for types (e.g. one index per app)
- Type - a container for similar documents (like class types, object types, etc.)
- Document - the content to search/store (instances of a type)

Demo

# Demo Server Is Open

---

- Elasticsearch is great! I want folks to try it.
- The meeting notes posted online will contain CURL examples ready for copy-and-paste.
- Feel free to play around. I won't delete the ES instance until Friday (easy to set up your own via packages, Docker, etc.)
- Ideas to try:
  - How many times does someone say “Eat my shorts”?
  - How many times do characters other than Bart Simpson say it?

Thanks.