

# NoSQL

# Where we are

- Data science
  1. Data Preparation (at scale)
  2. Analytics
  3. Communication
- Databases
  - Key ideas: Relational algebra, physical/logical data independence
- MapReduce
  - Key ideas: Fault tolerance, no loading, direct programming on “in situ” data

## NoSQL and related systems, by feature

<i>Year</i>	<i>System/ Paper</i>	<i>Scale to 1000s</i>	<i>Primary Index</i>	<i>Secondary Indexes</i>	<i>Transactions</i>	<i>Joins/ Analytics</i>	<i>Integrity Constraints</i>	<i>Views</i>	<i>Language/ Algebra</i>	<i>Data model</i>	<i>my label</i>
1971	RDBMS	0	✓	✓	✓	✓	✓	✓	✓	tables	sql-like
2003	memcached	✓	✓	0	0	0	0	0	0	key-val	nosql
2004	MapReduce	✓	0	0	0	✓	0	0	0	key-val	batch
2005	CouchDB	✓	✓	✓	record	MR	0	✓	0	document	nosql
2006	BigTable/Hbase	✓	✓	✓	record	compat. w/MR	/	0	0	ext. record	nosql
2007	MongoDB	✓	✓	✓	EC, record	0	0	0	0	document	nosql
2007	Dynamo	✓	✓	0	0	0	0	0	0	ext. record	nosql
2008	Pig	✓	0	0	0	✓	/	0	✓	tables	sql-like
2008	HIVE	✓	0	0	0	✓	✓	0	✓	tables	sql-like
2008	Cassandra	✓	✓	✓	EC, record	0	✓	✓	0	key-val	nosql
2009	Voldemort	✓	✓	0	EC, record	0	0	0	0	key-val	nosql
2009	Riak	✓	✓	✓	EC, record	MR	0			key-val	nosql
2010	Dremel	✓	0	0	0	/	✓	0	✓	tables	sql-like
2011	Megastore	✓	✓	✓	entity groups	0	/	0	/	tables	nosql
2011	Tenzing	✓	0	0	0	0	✓	✓	✓	tables	sql-like
2011	Spark/Shark	✓	0	0	0	✓	✓	0	✓	tables	sql-like
2012	Spanner	✓	✓	✓	✓	?	✓	✓	✓	tables	sql-like
2013	Impala	✓	0	0	0	✓	✓	0	✓	tables	sql-like

Year	System/ Paper	<i>Scale to 1000s</i>	Primary Index	Secondary Indexes	Transactions	Joins/ Analytics	Integrity Constraints	Views	Language/ Algebra	Data model	my label
1971	RDBMS	✓	✓	✓	✓	✓	✓	✓	✓	tables	sql-like
2003	memcached	✓	✓	○	○	○	○	○	○	key-val	nosql
2004	MapReduce	✓	○	○	○	✓	○	○	○	key-val	batch
2005	CouchDB	✓	✓	✓	record	MR	○	✓	○	document	nosql
2006	BigTable (Hbase)	✓	✓	✓	record	compat. w/MR	/	○	○	ext. record	nosql
2007	MongoDB	✓	✓	✓	EC, record	○	○	○	○	document	nosql
2007	Dynamo	✓	✓	○	○	○	○	○	○	ext. record	nosql
2008	Pig	✓	○	○	○	✓	/	○	✓	tables	sql-like
2008	HIVE	✓	○	○	○	✓	✓	○	✓	tables	sql-like
2008	Cassandra	✓	✓	✓	EC, record	○	✓	✓	○	key-val	nosql
2009	Voldemort	✓	✓	○	EC, record	○	○	○	○	key-val	nosql
2009	Riak	✓	✓	✓	EC, record	MR	○			key-val	nosql
2010	Dremel	✓	○	○	○	/	✓	○	✓	tables	sql-like
2011	Megastore	✓	✓	✓	entity groups	○	/	○	/	tables	nosql
2011	Tenzing	✓	○	○	○	○	✓	✓	✓	tables	sql-like
2011	Spark/Shark	✓	○	○	○	✓	✓	○	✓	tables	sql-like
2012	Spanner	✓	✓	✓	✓	?	✓	✓	✓	tables	sql-like
2012	Accumulo	✓	✓	✓	record	compat. w/MR	/	○	○	ext. record	nosql
2013	Impala	✓	○	○	○	✓	✓	○	✓	tables	sql-like

Scale was the primary motivation!