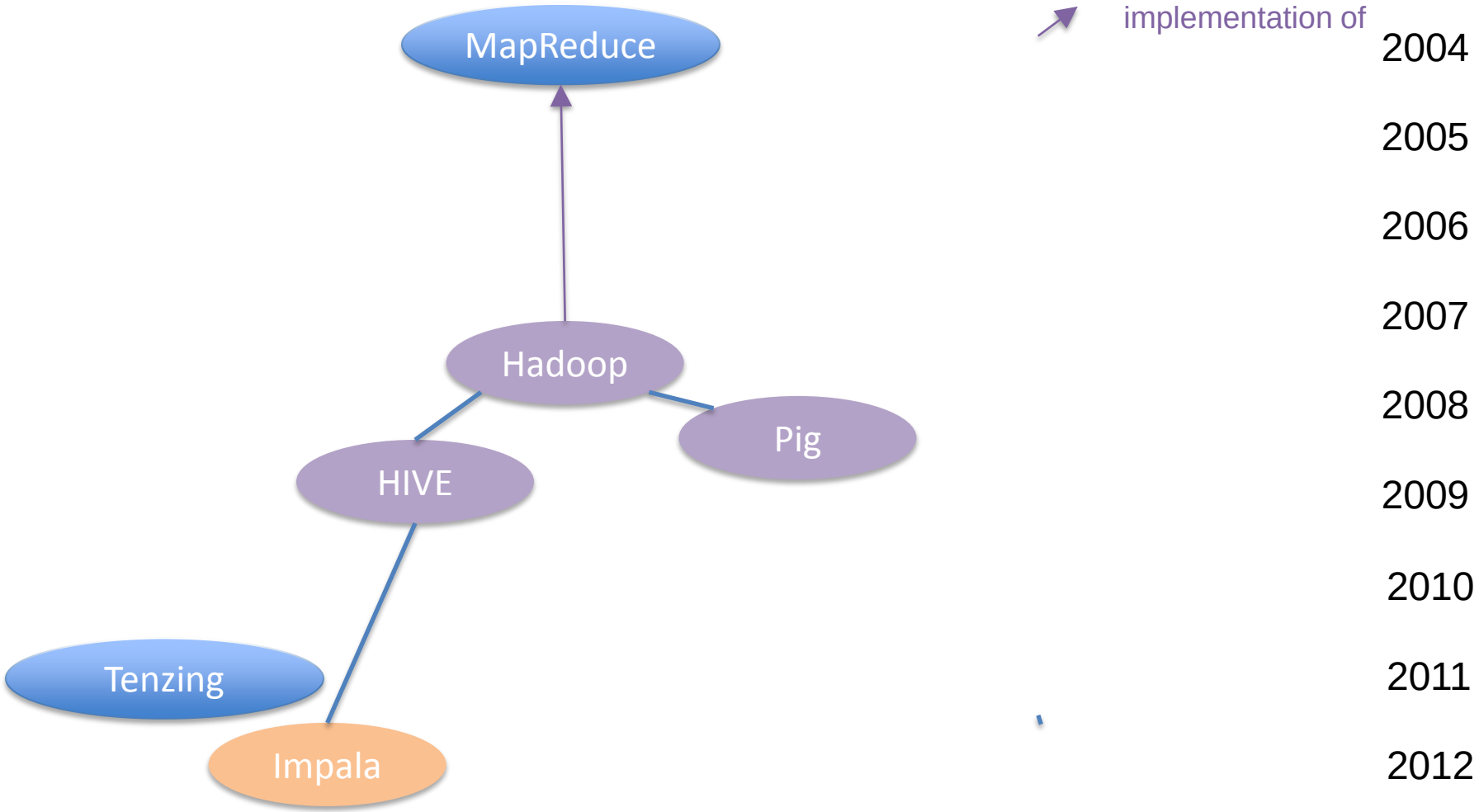
Year	System/ Paper	Scale to 1000s	Primary Index	Secondary Indexes	Transactions	Joins/ Analytics	Integrity Constraints	Views	Language/ Algebra	Data model	my label
1971	RDBMS	O	✔	✔	✔	✔	✔	✔	✔	tables	sql-like
2003	memcached	✔	✔	O	O	O	O	O	O	key-val	nosql
2004	MapReduce	✔	O	O	O	✔	O	O	O	key-val	batch
2005	CouchDB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

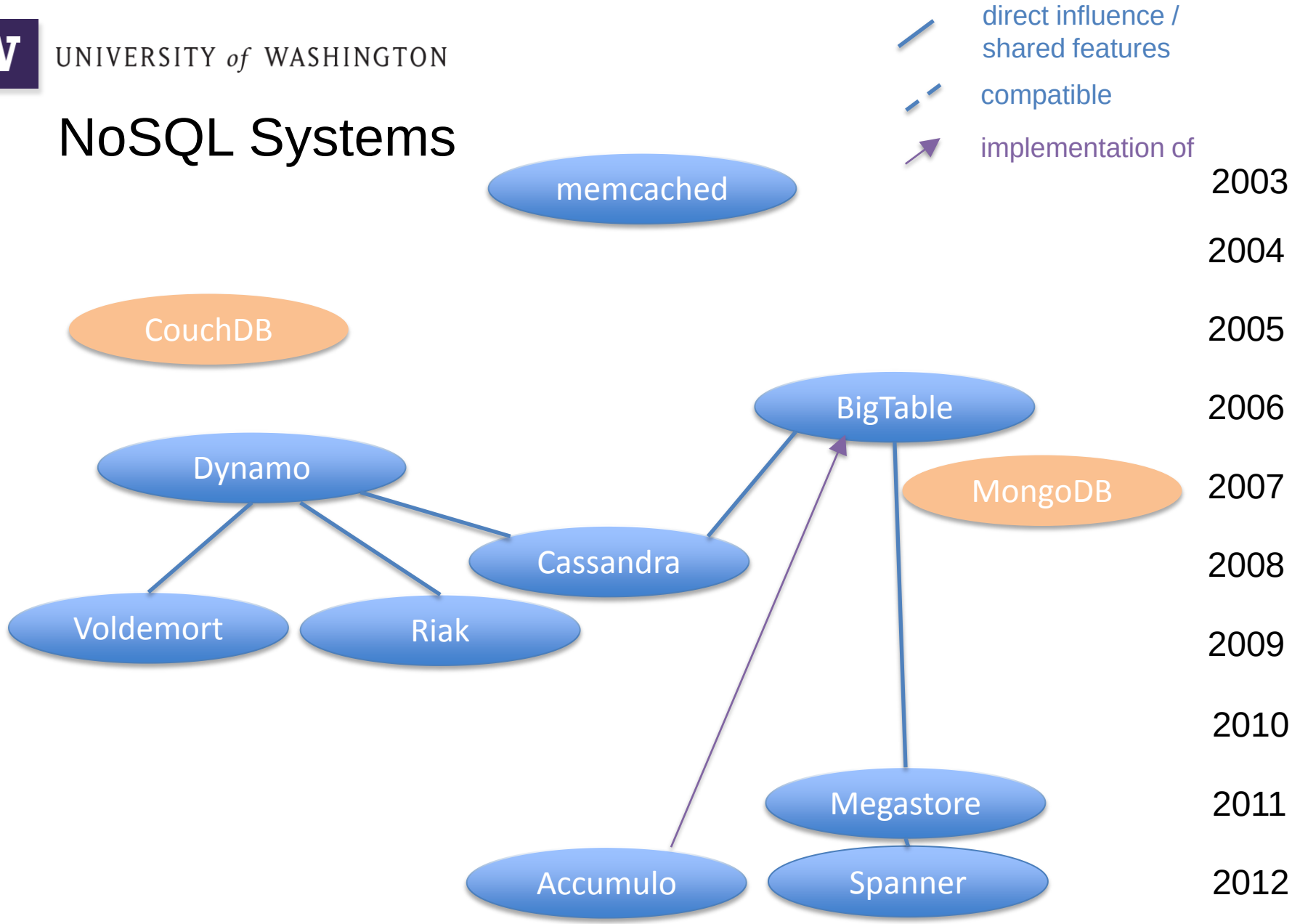
MapReduce-based Systems

MapReduce-based Systems



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

NoSQL Systems



Year	source	System/ Paper	Scale to 1000s	Primary Index	Secondary Indexes	Transactions	Joins/ Analytics	Integrity Constraints	Views	Language/ Algebra	Data model	my label
1971	many	RDBMS	O	✔	✔	✔	✔	✔	✔	✔	tables	SQL-like
2003	other	memcached	✔	✔	O	O	O	O	O	O	key-val	lookup
2004	Google	MapReduce	✔	O	O	O	✔	O	O	O	key-val	MR
2005	couchbase	CouchDB	✔	✔	✔	record	MR	O	✔	O	document	filter/MR
2006	Google	BigTable (Hbase)	✔	✔	✔	record	compat. w/MR	/	O	O	ext. record	filter/MR
2007	10gen	MongoDB	✔	✔	✔	EC, record	O	O	O	O	document	filter
2007	Amazon	Dynamo	✔	✔	O	O	O	O	O	O	key-val	lookup
2007	Amazon	SimpleDB	✔	✔	✔	O	O	O	O	O	ext. record	filter
2008	Yahoo	Pig	✔	O	O	O	✔	/	O	✔	tables	RA-like
2008	Facebook	HIVE	✔	O	O	O	✔	✔	O	✔	tables	SQL-like
2008	Facebook	Cassandra	✔	✔	✔	EC, record	O	✔	✔	O	key-val	filter
2009	other	Voldemort	✔	✔	O	EC, record	O	O	O	O	key-val	lookup
2009	basho	Riak	✔	✔	✔	EC, record	MR	O			key-val	filter
2010	Google	Dremel	✔	O	O	O	/	✔	O	✔	tables	SQL-like
2011	Google	Megastore	✔	✔	✔	entity groups	O	/	O	/	tables	filter
2011	Google	Tenzing	✔	O	O	O	✔	✔	✔	✔	tables	SQL-like
2011	Berkeley	Spark/Shark	✔	O	O	O	✔	✔	O	✔	tables	SQL-like
2012	Google	Spanner	✔	✔	✔	✔	?	✔	✔	✔	tables	SQL-like
2012	Accumulo	Accumulo	✔	✔	✔	record	compat. w/MR	/	O	O	ext. record	filter
2013	Cloudera	Impala	✔	O	O	O	✔	✔	O	✔	tables	SQL-like

A lot of these systems give up joins!

Joins

- Ex: Show all comments by “Sue” on any blog post by “Jim”

- Method 1:
  - Lookup all blog posts by Jim
  - For each post, lookup all comments and filter for “Sue”
- Method 2:
  - Lookup all comments by Sue
  - For each comment, lookup all posts and filter for “Jim”
- Method 3:
  - Filter comments by Sue, filter posts by Jim,
  - Sort all comments by blog id, sort all blogs by blog id
  - Pull one from each list to find matches

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

NoSQL Criticism

- Two value propositions
  - Performance: “I started with MySQL, but had a hard time scaling it out in a distributed environment”
  - Flexibility: “My data doesn’t conform to a rigid schema”

NoSQL Criticism: flexibility argument

- Who are the customers of NoSQL?
  - Lots of startups
- Very few enterprises. Why? most applications are traditional OLTP on structured data; a few other applications around the “edges”, but considered less important

NoSQL Criticism: Flexibility Argument

- No ACID Equals No Interest
  - Screwing up mission-critical data is no-no-no
- Low-level Query Language is Death
  - Remember CODASYL?
- NoSQL means NoStandards
  - One (typical) large enterprise has 10,000 databases. These need accepted standards