

# NoSQL Road Show, Basel

## NoSQL, NewSQL, Big Data... Total Data The Future of Enterprise Data Management

Matthew Aslett  
Research Manager, Data Management and Analytics

Photo credit: Wilson Loo on Flickr <http://bit.ly/Pf43nE>

# Overview

## NoSQL and NewSQL databases

- Adoption and development drivers

## Big data and Total Data

- Definition and implications

# The 451 Group



# 451 Research

- Matthew Aslett
  - Research manager, data management and analytics
  - With The 451 Group since 2007
  - [www.twitter.com/maslett](http://www.twitter.com/maslett)

## Information Management

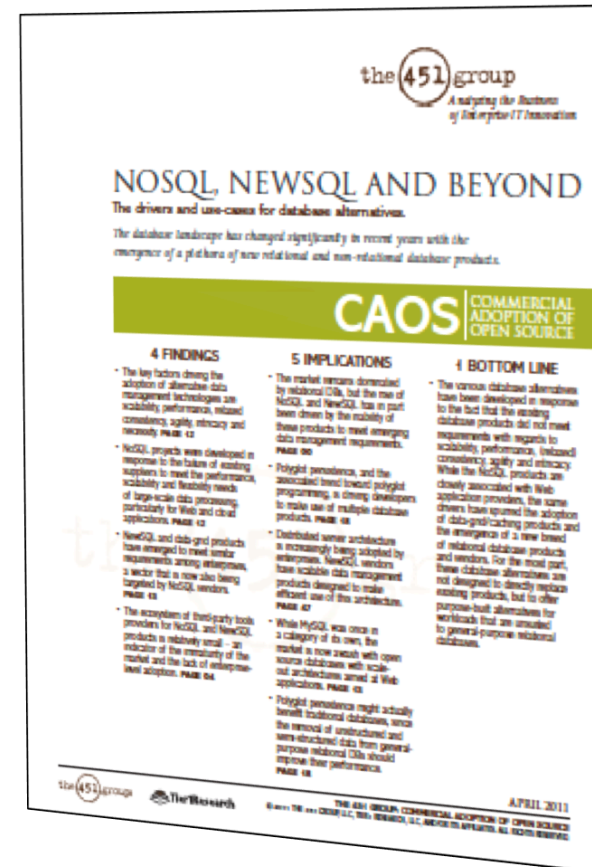
- Operational databases
- Data warehousing
- Data caching
- Event processing

## Commercial Adoption of Open Source (CAOS)

- Open source projects
- Adoption of open source software
- Vendor strategies

# Relevant reports

- NoSQL, NewSQL and Beyond
  - Assessing the drivers behind the development and adoption of NoSQL and NewSQL databases, as well as data grid/caching technologies
- Released April 2011
- Role of open source in driving innovation
- [sales@the451group.com](mailto:sales@the451group.com)



# NoSQL, NewSQL and Beyond

## NoSQL

- New breed of non-relational database products
- Rejection of fixed table schema and join operations
- Designed to meet scalability requirements of distributed architectures
- And/or schema-less data management requirements

# The NoSQL landscape

## Key Value Store

- Aerospike
- HandlerSocket\*
- Redis
- Voldemort
- Membrain
- Oracle NoSQL
- Castle
- RethinkDB
- LevelDB
- Cassandra
- DataStax EE
- Acunu

- HBase
- Hypertable

- Riak
- Couchbase

- DynamoDB
- Redis-to-go

- Accumulo
- ## Big Tables

## Document

- RavenDB
- MongoDB
- CouchDB
- Cloudbant
- Iris Couch
- Mongo Labs
- Mongo HQ

## Graph

- InfiniteGraph
- YarcData
- OrientDB
- NuvolaBase
- DEX
- Neo4j

-as-a-Service

# NoSQL, NewSQL and Beyond

## NoSQL

- New breed of non-relational database products
- Rejection of fixed table schema and join operations
- Designed to meet scalability requirements of distributed architectures
- And/or schema-less data management requirements

## NewSQL

- New breed of relational database products
- Retain SQL and ACID
- Designed to meet scalability requirements of distributed architectures
- Or improve performance so horizontal scalability is no longer a necessity



# The NewSQL ecosystem

## -as-a-Service

- StormDB
- Xeround
- Tokutek

## Storage engines

- Datomic
- Akiban
- GenieDB
- ScaleDB
- MySQL Cluster
- Zimory Scale

- MemSQL
- Drizzle
- VoltDB
- JustOneDB
- ParElastic
- Continuent
- Galera

## New databases

- NuoDB
- SQLFire
- Translattice
- Clustrix
- SchoonerSQL

- ScaleBase
- ScaleArc
- CodeFutures

## Clustering/sharding

# NoSQL, NewSQL and Beyond

## NoSQL

- New breed of non-relational database products
- Rejection of fixed table schema and join operations
- Designed to meet scalability requirements of distributed architectures
- And/or schema-less data management requirements

## NewSQL

- New breed of relational database products
- Retain SQL and ACID
- Designed to meet scalability requirements of distributed architectures
- Or improve performance so horizontal scalability is no longer a necessity

## ... and Beyond

- In-memory data grid/cache products
- Potential primary platform for distributed data management

# SPRAINED RELATIONAL DATABASES



Photo credit: Foxtongue on Flickr  
<http://www.flickr.com/photos/foxtongue/4844016087/>

## Scalability - Hardware economics

- Example project/service/vendor:
  - BigTable, HBase, Riak, MongoDB, Couchbase, Hadoop
  - Xeround, NuoDB
  - Data grid/cache
  
- Associated use case:
  - Large-scale distributed data storage
  - Analysis of continuously updated data
  - Multi-tenant PaaS data layer

## Performance - MySQL limitations

- Example project/service/vendor:
  - Hypertable, Couchbase, Riak, Membrain, MongoDB, Redis
  - Data grid/cache
  - VoltDB, Clustrix
- Associated use case:
  - Real time data processing of mixed read/write workloads
  - Data caching
  - Large-scale data ingestion

## Relaxed consistency - CAP Theorem

- Example project/service/vendor:
  - Dynamo, Voldemort, Cassandra, Riak
  - Amazon DynamoDB
  
- Associated use case:
  - Multi-data center replication
  - Service availability
  - Non-transactional data off-load

## Agility - polyglot persistence, schema-less

- Example project/service/vendor:
  - MongoDB, CouchDB, Cassandra, Riak
  - Google App Engine, SimpleDB,
- Associated use case:
  - Mobile/remote device synchronization
  - Agile development
  - Data caching

## Intricacy - big data, total data

- Example project/service/vendor:
  - Neo4j, GraphDB, InfiniteGraph
  - Apache Cassandra, Hadoop, Riak
  - VoltDB, Clustrix
  
- Associated use case:
  - Social networking applications
  - Geo-locational applications
  - Configuration management database



## Necessity - open source

- Example projects:
  - BigTable: Google
  - Dynamo: Amazon
  - Cassandra: Facebook
  - HBase: Powerset
  - Voldemort: LinkedIn
  - Hypertable: Zvents
  - Neo4j: Windh Technologies
  - Accumulo: NSA

## Necessity

- The failure of existing suppliers to address emerging requirements
- “We couldn’t bet the company on other companies building the answer for us.”

Werner Vogels, Amazon CTO

- The motivation for creating Dynamo was enabling choice and not forcing the relational database to do something it was not designed to do.

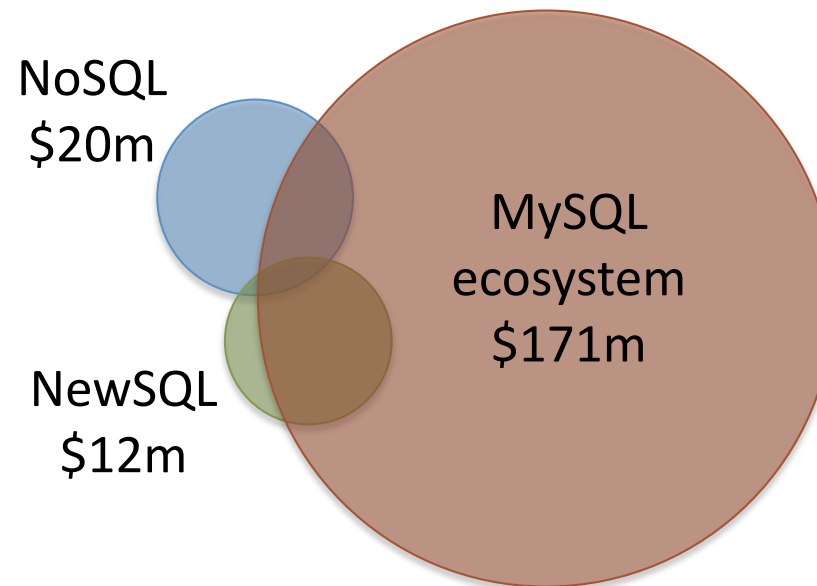
# Relevant reports

## MySQL vs NoSQL and NewSQL: 2011-2015

- Assessing the competitive dynamic
- Released May 2012
- Including market sizing estimates for all three sectors
- Survey of 200+ database users
- [sales@the451group.com](mailto:sales@the451group.com)
- <https://451research.com/report-long?icid=2289>
- [http://blogs.the451group.com/information\\_management/?p=1740](http://blogs.the451group.com/information_management/?p=1740)

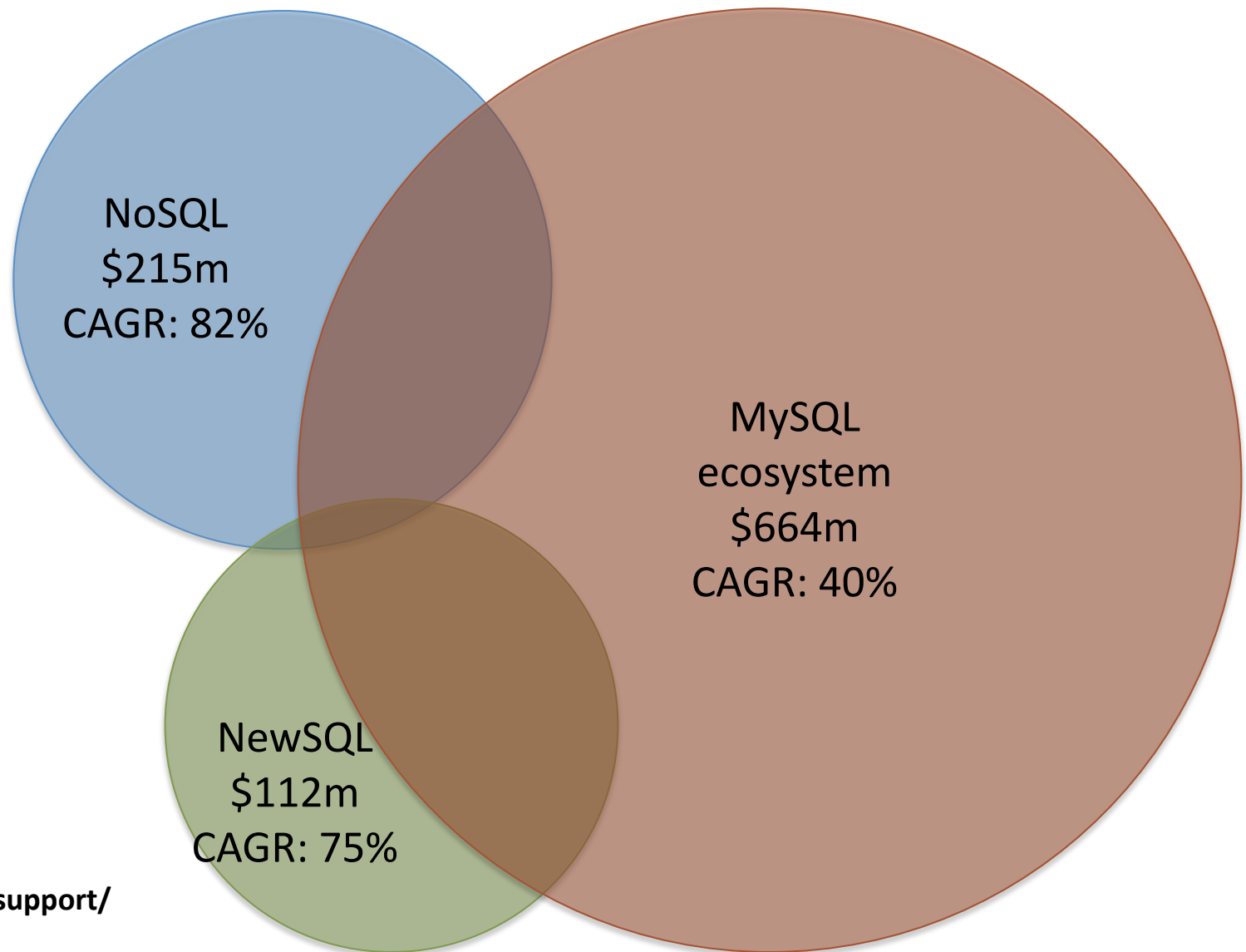


# Revenue estimates: 2011



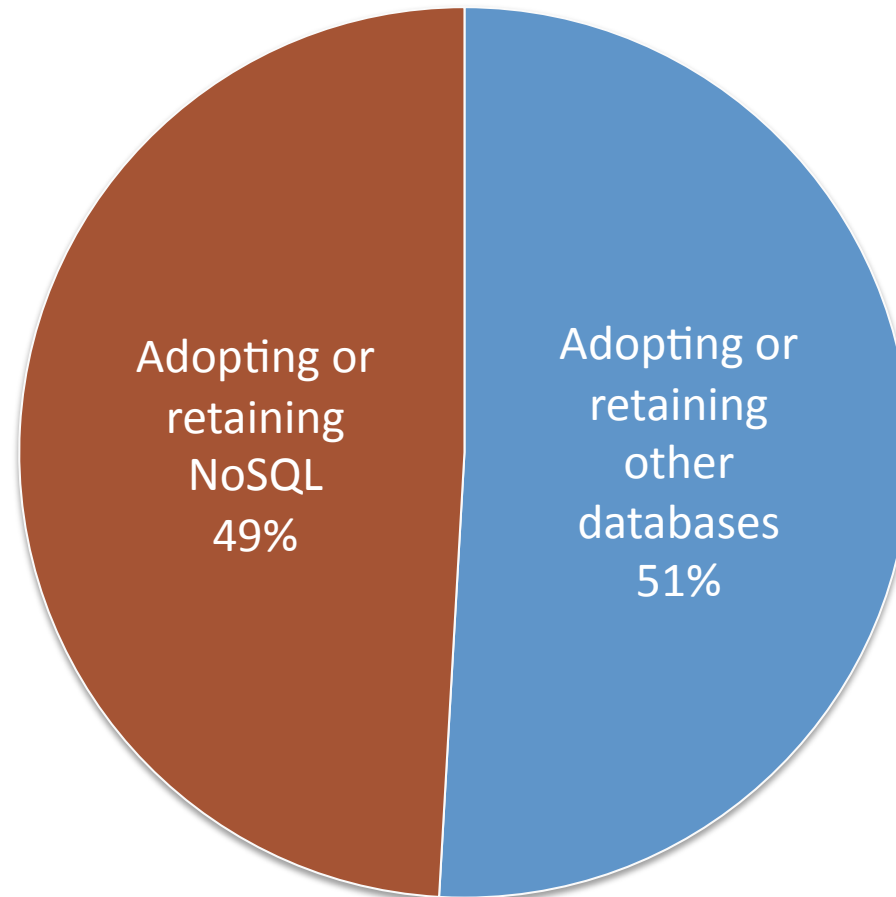
**Includes only software and support/  
subscription revenue.**

# Revenue estimates: 2015



**Includes only software and support/  
subscription revenue.**

# MySQL abandoners - adoption plans

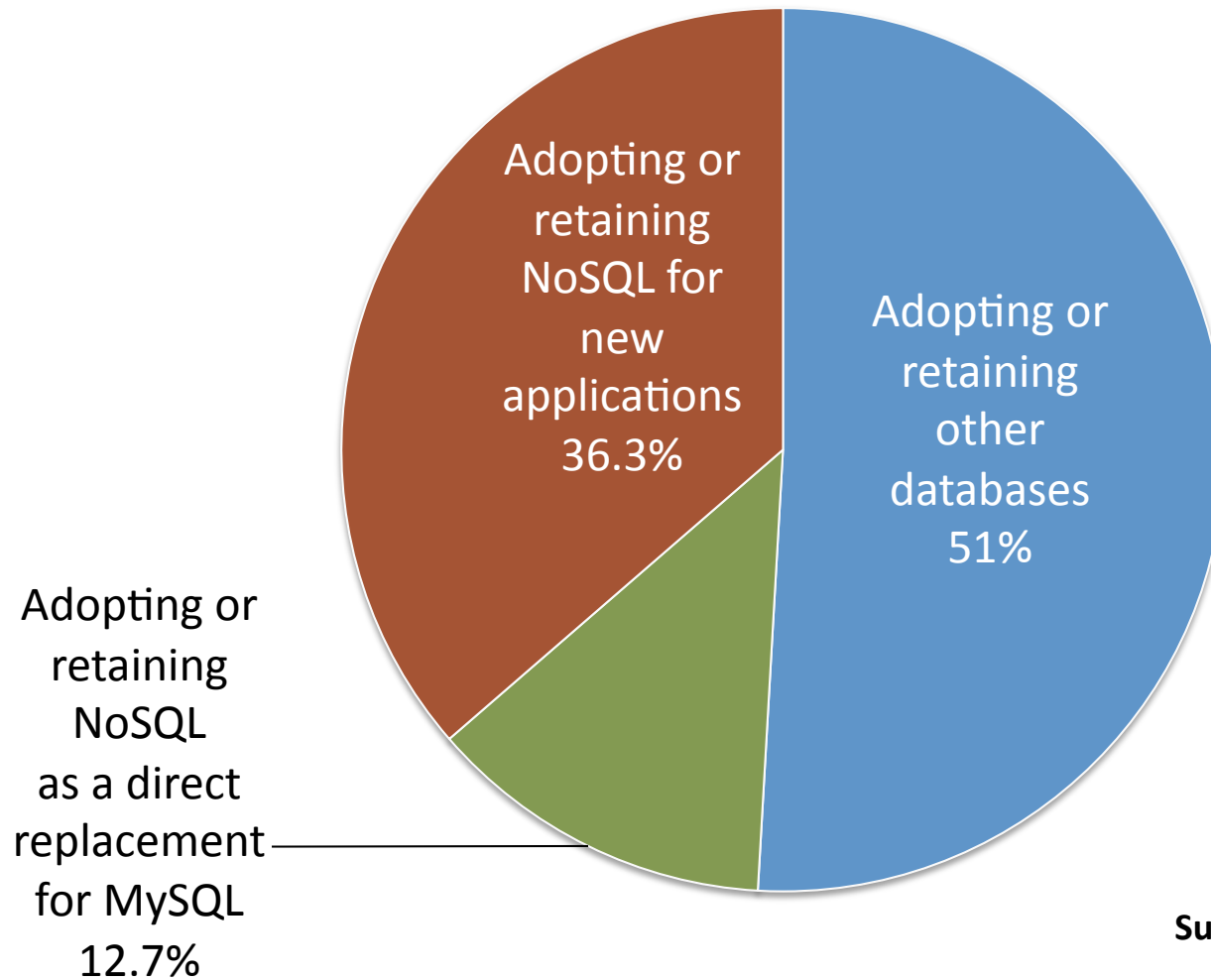


**Survey conducted:** Jan/Feb 2012

**Sample:** 205

**MySQL abandoners:** 55

# MySQL abandoners - adoption plans



Survey conducted: Jan/Feb 2012

Sample: 205

MySQL abandoners: 55

# BIG DATA... TOTAL DATA



Source: Wikimedia. Attribution: Bundesarchiv, Bild 183-N0716-0314 / Mittelstädt, Rainer / CC-BY-SA  
[http://commons.wikimedia.org/wiki/File:Bundesarchiv\\_Bild\\_183-N0716-0314,\\_Fu%C3%9Fball-WM,\\_BRD\\_-\\_Niederlande\\_2-1.jpg](http://commons.wikimedia.org/wiki/File:Bundesarchiv_Bild_183-N0716-0314,_Fu%C3%9Fball-WM,_BRD_-_Niederlande_2-1.jpg)



# 'Big Data'

- “Big data” describes the realization of greater business intelligence by storing, processing and analyzing data that was previously ignored due to the limitations of traditional data management technologies to handle its volume, velocity and/or variety.



## Volume

The volume of data is too large for traditional database software tools to cope with



## Velocity

The data is being produced at a rate that is beyond the performance limits of traditional systems



## Variety

The data lacks the structure to make it suitable for storage and analysis in traditional databases and data warehouses

# 'Big Data'

- “Big data” describes the realization of greater business intelligence by storing, processing and analyzing data that was previously ignored due to the limitations of traditional data management technologies to handle its volume, velocity and/or variety.
- The increased use of interactive applications and websites – as well as sensors, meters and other data-generating machines – has increased the amount and variety of data to store and process.
- The cost of storage, processing and bandwidth has dropped enormously, while network access has increased significantly.
- It is now more economically feasible to store and process many data sets that were previously ignored using clusters of commodity servers and advanced data processing software.

# Beyond 'big data'

- To realize value from data you need to look beyond the data itself.
- Generating value from data is about more than just the volume, variety, and velocity of data.
- “Total Data”
- Not just another name for Big Data
- Inspired by ‘Total Football’ – a new approach to soccer that emerged in the late 1960s
- If your data is big, the way you manage it should be total

# Beyond 'big data'

- The adoption of non-traditional data processing technologies is driven not just by the nature of the data, but also by the user's particular data processing requirements.



## Totality

The desire to process and analyze data in its entirety, rather than analyzing a sample of data and extrapolating the results.



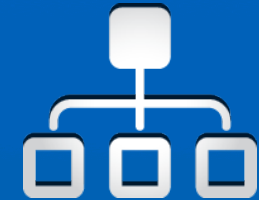
## Exploration

The interest in exploratory analytic approaches, in which schema is defined in response to the nature of the query.



## Frequency

The desire to increase the rate of analysis in order to generate more accurate and timely business intelligence.



## Dependency

The reliance on existing technologies and skills, and the need to balance investment in those existing technologies and skills with the adoption of new techniques.

# Relevant reports

- Total Data – published December 2011
  - Examines the trends behind ‘big data’
  - Explains the new and existing technologies used to store and process and deliver value from data
  - Outlines a Total Data management approach focused on selecting the most appropriate data storage and processing technology to deliver value from big data
  - sales@the451group.com



Thank you. Questions? Comments?  
matt.aslett@451research.com  
@maslett



Photo credit: swisscan on Flickr <http://bit.ly/HCZCRQ>

WWW.451RESEARCH.COM

NEW YORK · BOSTON · WASHINGTON DC · SAN FRANCISCO · SEATTLE · DENVER · LONDON · SAO PAULO · DUBAI · SINGAPORE