# CS 848 – Modern Database Systems
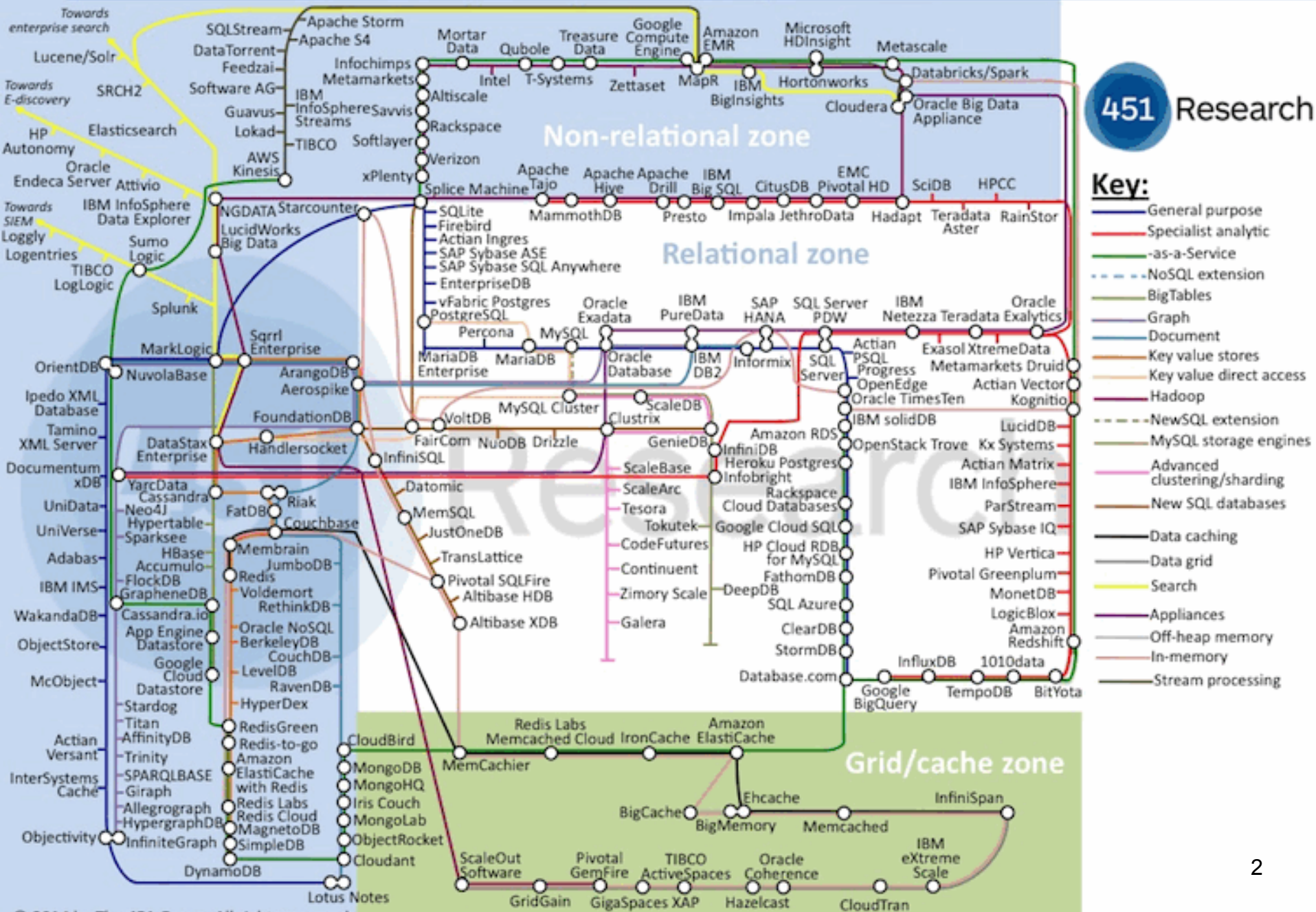
## M. Tamer Özsu

tamer.ozsu@uwaterloo.ca
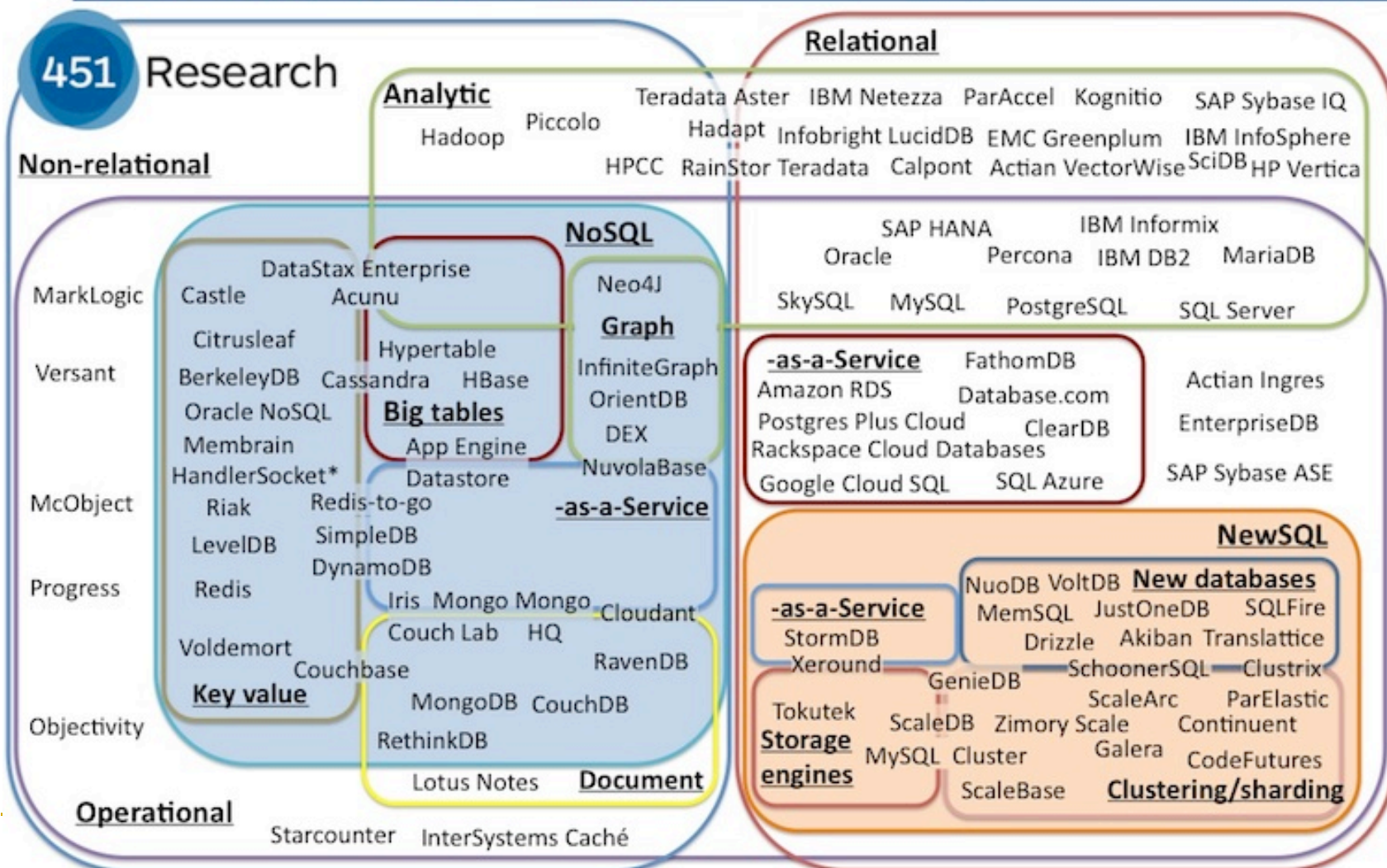
http://cs.uwaterloo.ca/~tozsu

# Data Platforms Landscape Map – February 2014



**451 Research**

**Non-relational zone**

**Relational zone**

**Grid/cache zone**

## Key:

- General purpose
- Specialist analytic
- -as-a-Service
- NoSQL extension
- BigTables
- Graph
- Document
- Key value stores
- Key value direct access
- Hadoop
- NewSQL extension
- MySQL storage engines
- Advanced clustering/sharding
- New SQL databases
- Data caching
- Data grid
- Search
- Appliances
- Off-heap memory
- In-memory
- Stream processing

Towards enterprise search
Lucene/Solr
Towards E-discovery
SRCH2
HP Autonomy
Elasticsearch
Oracle Endeca Server
Attivio
Towards SIEM
IBM InfoSphere Data Explorer
Loggly
Logentries
TIBCO LogLogic
Sumo Logic
Splunk

SQLStream
DataTorrent
Feedzai
Software AG
Guavus
Lokad
AWS Kinesis
Apache Storm
Apache S4
Infochimps
Metamarkets
IBM InfoSphere Streams
TIBCO
Altiscale
Savvis
Rackspace
Softlayer
Verizon
xPlenty

Mortar Data
Qubole
Intel
T-Systems
Zettaset
MapR
Treasure Data
Google Compute Engine
Amazon EMR
Microsoft HDInsight
IBM BigInsights
Hortonworks
Cloudera
Metascale
Databricks/Spark
Oracle Big Data Appliance

NGDATA
Starcounter
LucidWorks Big Data

SQLite
Firebird
Actian Ingres
SAP Sybase ASE
SAP Sybase SQL Anywhere
EnterpriseDB
vFabric Postgres
PostgreSQL
Percona
MySQL
MariaDB Enterprise
MariaDB

Splice Machine
Apache Tajo
MammothDB
Apache Hive
Apache Drill
Presto
IBM Big SQL
Impala
CitusDB
JethroData
EMC Pivotal HD
Hadapt
SciDB
Teradata Aster
HPCC
RainStor

Oracle Exadata
IBM PureData
SAP HANA
SQL Server PDW
IBM Netezza
Teradata
Oracle Exalytics

MarkLogic
OrientDB
NuvolaBase
Ipedo XML Database
Tamino XML Server
Documentum xDB
UniData
UniVerse
Adabas
IBM IMS
WakandaDB

Sqrrl Enterprise
ArangoDB
Aerospike
FoundationDB
DataStax Enterprise
Handlersocket

Oracle Database
IBM DB2
Informix
SQL Server
Actian PSQL Progress OpenEdge
Oracle TimesTen
IBM solidDB
OpenStack Trove
Amazon RDS
Heroku Postgres
Infobright
Rackspace Cloud Databases
Google Cloud SQL
HP Cloud RDB for MySQL
FathomDB
DeepDB
SQL Azure
ClearDB
StormDB
Database.com

Exasol XtremeData
Metamarkets Druid
Actian Vector
Kognitio
LucidDB
Kx Systems
Actian Matrix
IBM InfoSphere
ParStream
SAP Sybase IQ
HP Vertica
Pivotal Greenplum
MonetDB
LogicBlox
Amazon Redshift
InfluxDB
1010data
TempoDB
BitYota

VoltDB
MySQL Cluster
Clustrix
ScaleDB
GenieDB
FairCom
NuoDB
Drizzle
InfiniSQL
Datomic
MemSQL
JustOneDB
TransLattice
Pivotal SQLFire
Altibase HDB
Altibase XDB
ScaleBase
ScaleArc
Tesora
Tokutek
CodeFutures
Continuent
Zimory Scale
Galera

Cassandra
Neo4J
Hypertable
Sparksee
HBase
Accumulo
FlockDB
GrapheneDB
Cassandra.io
App Engine Datastore
Google Cloud Datastore
Stardog
Titan
AffinityDB
Trinity
SPARQLBASE
Giraph
Allegrograph
HypergraphDB
InfiniteGraph
Objectivity
InterSystems Cache
Actian Versant
McObject
ObjectStore

YarcData
FatDB
Riak
Couchbase
Membrain
JumboDB
Redis
Voldemort
RethinkDB
Oracle NoSQL
BerkeleyDB
CouchDB
LevelDB
RavenDB
HyperDex
RedisGreen
Redis-to-go
Amazon ElastiCache with Redis
Redis Labs
Redis Cloud
MagnetoDB
SimpleDB
DynamoDB
Lotus Notes

CloudBird
MongoDB
MongoHQ
Iris Couch
MongoLab
ObjectRocket
Cloudant
MemCachier

Redis Labs
Memcached Cloud
IronCache
Amazon ElastiCache

ScaleOut Software
GridGain
Pivotal GemFire
GigaSpaces XAP
TIBCO ActiveSpaces
Hazelcast
Oracle Coherence
CloudTran
BigCache
BigMemory
Ehcache
Memcached
InfiniSpan
IBM eXtreme Scale
Google BigQuery

2

# Older but a bit more helpful

## The evolving database landscape

**451** Research

**Relational**

**Non-relational**

**Analytic**

Hadoop   Piccolo

Teradata Aster   IBM Netezza   ParAccel   Kognitio   SAP Sybase IQ

Hadapt   Infobright LucidDB   EMC Greenplum   IBM InfoSphere

HPCC   RainStor Teradata   Calpont   Actian VectorWise SciDB HP Vertica

MarkLogic

Versant

McObject

Progress

Objectivity

**NoSQL**

DataStax Enterprise

Castle          Acunu

Citrusleaf          Hypertable

BerkeleyDB   Cassandra   HBase

Oracle NoSQL       **Big tables**

Membrain          App Engine

HandlerSocket*   Datastore

Riak   Redis-to-go

LevelDB   SimpleDB

DynamoDB

Redis

Iris  Mongo Mongo

Voldemort   Couch Lab   HQ

Couchbase

**Key value**      MongoDB   CouchDB

RethinkDB

Lotus Notes   **Document**

**Operational**

Starcounter   InterSystems Caché

Neo4J

**Graph**

InfiniteGraph

OrientDB

DEX

NuvolaBase

-Cloudant

RavenDB

SAP HANA          IBM Informix

Oracle          Percona   IBM DB2   MariaDB

SkySQL   MySQL   PostgreSQL   SQL Server

**-as-a-Service**   FathomDB

Amazon RDS          Database.com          Actian Ingres

Postgres Plus Cloud          ClearDB          EnterpriseDB

Rackspace Cloud Databases

Google Cloud SQL   SQL Azure          SAP Sybase ASE

**-as-a-Service**

**NewSQL**

NuoDB VoltDB **New databases**

**-as-a-Service**   MemSQL   JustOneDB   SQLFire

StormDB          Drizzle   Akiban Translattice

Xeround          SchoonerSQL — Clustrix

GenieDB

Tokutek          ScaleArc          ParElastic

**Storage**   ScaleDB   Zimory Scale          Continuent

**engines**   MySQL  Cluster   Galera   CodeFutures

ScaleBase   **Clustering/sharding**

# Two parts

- A brief overview of database technology
  - Warning: this is **not** a tutorial; just an overview of basic functionality
- … and technology changes that impact database systems
- New systems
  - NewSQL
  - NoSQL
  - Column stores
  - MapReduce-based systems
  - Stream processing systems
  - Graph processing (both querying and analysis) systems

# Course Information

- Prerequisites: need to know database technology (internals) fairly well.
  - DO NOT take the course if you don't and you don't feel comfortable reading database research papers
- Will discuss 4papers each week
- There is no textbook for the course; I will indicate some that have parts of what we will cover
- The web page for the course:

  https://cs.uwaterloo.ca/~tozsu/courses/cs848/W15/

# Workload & Evaluation

- Research Project: 40%
  - Usually in groups of 2
  - Can be experimental or conceptual
  - Pick a problem (in consultation with me) and
    - Understand it fully and the literature about it
    - Show me how you would solve it and what is novel about your solution
  - Each group presents at the end of the course (public presentation)
- Class presentation(s): 20%
- 2 critical reviews of research papers: 20%
- Class participation: 20%
  - Participation in class discussions
  - Short reviews of papers presented each week (1/2 page)
  - Evaluation of other students' presentations