

Database Systems Summary and Overview

Professor Xiannong Meng
Bucknell University
Spring 2018

Database Landscape

- Relational database systems (RDBMS)
 - E.g., sqlite
- NoSQL database systems
 - E.g., mongodb
- Large scale distributed database
 - E.g., Hadoop

<https://datajobs.com/what-is-hadoop-and-nosql>

RDBMS

- Databases are based on relations or tables;
- Relations are represented by entity relations in logic;
- Columns in a table represent the attributes of a relation;
- Rows in a table are entries in the relation;
- Relations have keys, super-keys, primary keys, and secondary keys;
- Functional dependencies guide the design theories;
 - Normal forms

RDBMS

- Relations are bound by constraints and dependencies;
- Database index helps search and retrieve database contents;
- Database views gives a snapshot of the database;
- Relational Algebra is a logical language that describe the basic operations of a database while SQL is an implementation;
- We discussed one aspect of the database implementation, index with B+ tree;

What is a NoSQL database?

- A **NoSQL** (originally referring to “non SQL” or “non relational”) database provides a mechanism for storage and retrieval of data that is modeled in means other than the tabular relations used in relational databases (SQL databases).

<https://en.wikipedia.org/wiki/NoSQL>

Four types of NoSQL databases

- **Column** : A column of a distributed data store is a NoSQL object of the lowest level in a keypace.
- **Document database** : A document-oriented database, or document store, is a computer program designed for storing, retrieving and managing document-oriented information.
- **Key-value database** : A data storage paradigm designed for storing, retrieving, and managing associative arrays, a data structure more commonly known today as a dictionary or hash.
- **Graph database** : A database that uses graph structures for semantic queries with nodes, edges and properties to represent and store data.

Large scale distributed databases

- **Hadoop** is not a type of database, but rather a software ecosystem that allows for massively parallel computing. It is an enabler of certain types NoSQL distributed databases (such as HBase), which can allow for data to be spread across thousands of servers with little reduction in performance.
- A staple of the Hadoop ecosystem is **MapReduce**, a computational model that basically takes intensive data processes and spreads the computation across a potentially endless number of servers (generally referred to as a Hadoop cluster).

<https://datajobs.com/what-is-hadoop-and-nosql>

A quick and simple example of Map-Reduce

- We learned the basics of Map-Reduce in CSCI 203!
- Map-Reduce is a two-step processing
 - **map** (f, t) -- Applies *any arbitrary function* f to every element in list t
 - **reduce** (f, t) -- Applies f (a function of two arguments) *cumulatively* to the items of t
- See `MapReduce_example.py` for an example

```
myList = [-100, 30, -23, 32, 13, 100, 22, 101]

#-----
# EXAMPLE OF MAP
#-----
def ftoC(degreesF):
    return (degreesF - 32) * (5/9)

def list_ftoC(tempList):
    # Map applies ftoC to each element in the list
    # It returns a MAP object that is ITERABLE.
    tempMap = map(ftoC, tempList)
    # We transform it into a list before returning
    return list(tempMap)

#-----
# EXAMPLE OF REDUCE
#-----
def maxOfTwo(x, y):
    if x > y:
        return x
    else:
        return y

def maxOfList(numList):
    return reduce(maxOfTwo, numList)

print('Temperature list in Fahrenheit : ', myList)
tempList = list_ftoC(myList)
print('Max temperature in Celsius : ' + '{:1f}'.format(maxOfList(tempList),2))
print('Min temperature in Celsius : ' + '{:1f}'.format(minOfList(tempList),2))

Python 3.6.0 (tags/v3.6.0:41df7923611, Dec 23 2016, 07:18:10) [MSC v.1900 32 bit (Intel)] on win32
Type "copyright", "credits" or "license()" for more information.
>>>
= RESTART: C:/Users/home/Desktop/bu-work/cs305/lectures/mapreduce_example.py =
Temperature list in Fahrenheit : [-100, 30, -23, 32, 13, 100, 22, 101]
Max temperature in Celsius : 38.33
Min temperature in Celsius : -73.33
```