1. **NoSQL Project Information**

* The main objective of this project is for you to learn NoSQL concepts and programming by doing a hands-on project of your own interest.
* You can implement any NoSQL application using any NoSQL database and any language of your choice.
* It doesn’t have to be a big data project. For example, you can cut down a public data set to make it fit in your laptop. By the same token, you don’t need a cluster of servers.
* Any NoSQL application that is not well supported by relational databases will be accepted. Basically, many applications dealing with semi-structured or unstructured data, e.g., twitter data, Wikipedia data, or sensor data, will be relevant. However, trivial applications, e.g., Word Count in Hadoop or simple store/load of data into/from a NoSQL database, won’t be accepted.
* For the interest of time, a project extending the underlying the NoSQL database system is not recommended. Although it can be a good project, it is too challenging to finish in one semester.
* If you use an existing application, you have to reference it, significantly extend it, and justify why the extension is non-trivial. (Just changing variable names, indentation, or minor changes of the program/data structure is not significant. It can be considered as plagiarism.)
* Alternatively, you can implement your application yourself on top of an existing NoSQL database. Even if your application is not new, if you design and implement it by yourself and it works correctly with reasonable performance, it will be acceptable.

1. **Example NoSQL Databases**

* Batch systems: Apache Casandra, Apache Spark, Hadoop
* Stream processing engine: Apache Storm, Apache Heron, Apache Edgent (runs on edge devices), Spark Streaming (part of Spark)
* Document database: MongoDB, CouchDB

1. **Example Datasets**

* [**Tweet datasets**](https://github.com/shaypal5/awesome-twitter-data#id5)

## [RIoTBench: A Real-time IoT Benchmark for Distributed Stream Processing Platforms](https://github.com/dream-lab/riot-bench)

* **Wikipedia Data:** <https://towardsdatascience.com/wikipedia-data-science-working-with-the-worlds-largest-encyclopedia-c08efbac5f5c>
* **Google Dataset Search Engine:** <https://datasetsearch.research.google.com/>

1. For surveys, you can read “[A Survey on NoSQL Databases](https://pdfs.semanticscholar.org/60e4/7cebd354b958d8b9b1576d1645b9fe04440a.pdf)”, “[A Survey of NoSQL Stores](https://pdfs.semanticscholar.org/60e4/7cebd354b958d8b9b1576d1645b9fe04440a.pdf)”
2. **Database Conferences**

* **ACM SIGMOD Conference**
* **VLDB Conference**
* **IEEE ICDE**