# “Big Data” Course Project Proposal

Fill out the table in the next page. Follow the instructions below:

1. **Title**
2. **Group members**: 1-3 students. Project word load should be appropriately determined. Each student should have as little as 2-3 assignments’ workload.
3. **Short Description**: Briefly explain the problem you want to solve. Two possible tracks: (1) an application of mapreduce on a large data set to provide a scalable solution for a real-world problem and possibly developing a nice mapreduce pattern, (2) developing and/or improving a distributed data storage and processing framework for a better scalable solution, e.g. a fast linked data storage system for fast access, query, update, delete operations for the Web of Data (linked data such DBpedia).
4. **Big Data source(s)**: List the data sourcesyou will use in the project. Real-world data is preferred. List data sources by name, brief description, source, size, Web link. Two possible venues: (1) **IoT data** (sensor data, mobile data, etc.), (2) **Social Media data** (twitter data, etc.).
5. **Related work**: Find related papers and list their shortcomings. Explain what is done, what is missing. At least 5 recent (2012-) papers should be listed. List the papers in the following format (google scholar citation, APA format, cited numbers, and link to the source):

* Hogan, A., Zimmermann, A., Umbrich, J., Polleres, A., & Decker, S. (2012). [Scalable and distributed methods for entity matching, consolidation and disambiguation over linked data corpora](http://mail.websemanticsjournal.org/index.php/ps/article/download/224/221). *Web Semantics: Science, Services and Agents on the World Wide Web*, *10*, 76-110.  
  **Cited**: 40

1. **Methods**: Briefly explain the method(s) you want to try/develop to solve the problem, such as data processing methods, machine learning, data mining methods, or statistical approaches.
2. **Target conferences and journals**:
   1. Conference: name, organization, location, deadline, web site
   2. Journal: name, publisher, index, impact, issue/year, web site

|  |  |  |
| --- | --- | --- |
| **Project Proposal – Big Data Course** | | |
| Project title |  | |
| Group members |  | |
| Data sources | Name |  |
| Description |  |
| Size |  |
| Web link |  |
| Description of the project |  | |
| Related work |  | |
| Methods |  | |
| Target conferences and journals with links |  | |