

### Intern Requirement

No. of students required	1
Preferred Discipline	<ul style="list-style-type: none"><li>• Computer Science</li><li>• Computer Engineering</li><li>• Electrical Engineering</li><li>• Information System Management</li></ul>
Prerequisites/ Skills Required	<ul style="list-style-type: none"><li>• Programming (Java preferred)</li><li>• Interest in information system infrastructure</li><li>• Knowledge in data management technologies preferred</li></ul>

### Project Details

Title	Tackling the Big Data Problem in Clickstream Analysis
Overview/Background	<p>Major companies around the world are currently faced with the "big data" problem of having to process huge volume of information collected.</p> <p>A particularly challenging arena of "big data" problem concerns the analysis of clickstream. Companies such as Facebook and Google are analyzing tens of terabytes of clickstreams on a daily basis. In order to facilitate such a large-scale analysis, both companies have developed new data processing paradigms and implemented state-of-the-art data processing infrastructures.</p>
Objectives/Scope/ Deliverables	<p>In this project, the intern would experiment with the latest distributed data storage and processing technologies such as Hadoop, Hbase and Hives, to understand the pros and cons of each technology, for use in a clickstream analysis platform.</p> <p>The intern will be expected to deliver a technical report outlining the pros and cons of each technology, supported by experiments conducted on prototype systems.</p>
Project Duration	<input checked="" type="checkbox"/> 2-4 months, or <input checked="" type="checkbox"/> 6 months