

### Intern Requirement

No. of students required	1
Preferred Discipline	<ul style="list-style-type: none"><li>• Computer Science</li><li>• Computer Engineering</li></ul>
Prerequisites/ Skills Required	<ul style="list-style-type: none"><li>• Knowledge Management</li><li>• Experience with Web crawler</li></ul>

### Project Details

Title	Hadoop Web Crawler
Overview/Background	<ul style="list-style-type: none"><li>• As we are living in a information overloaded world, valuable information are pumped into the internet at an exponential rate</li><li>• While the amount of data double every few months, it is very expensive and nearly impossible to double our hard disk capability at the same rate</li><li>• Hadoop, a storage solution pioneered by Google, is the technology employed by them to meet this storage need</li><li>• Web crawler will consume ever more storage and a mean to store these valuable information is mandatory to avoid loss of information</li></ul>
Objectives/Scope/ Deliverables	<ul style="list-style-type: none"><li>• Expose the intern to working knowledge on the cutting edge storage solution – Hadoop</li><li>• Develop a prototype web crawler that scrapes the social web for valuable information</li><li>• Evaluate and propose a feasible implementation plan for storing the huge amount of information crawled</li></ul>
Project Duration	2-4 months