

Google uses a large-scale unstructured data processing framework called MapReduce to return search query results. There's an open-source implementation of this framework called Hadoop (there's a book published on this as well called "Pro Hadoop".) There're two options to complete a PhD level research investigation and the corresponding solution:

1. Geospatial Data mining: Geospatial data sets can be huge. Is there an interesting research area to automatically annotate geospatial information from the web and the use Hadoop to process and return analytical results?
2. Augmenting security for Hadoop: Currently, security in Hadoop is very coarse-grained. We can utilize XACML to augment Hadoop's security. We'd need to modify Hadoop open-source to include our security constructs and show that this increases security is useful.

