SAN DIEGO STATE UNIVERSITY
Leadership Starts Here

SDSU
presents
MS Computer Science
THESIS DEFENSE

Friday,
October 21, 2016
12:00pm GMCS 418

Akhil Gokhale

Educational Tool for Hadoop

Abstract

You do not see a lot of good informative tutorials/information on Hadoop EcoSystem which helps you in mastering Hadoop from beginning to the end. Most of the Hadoop tutorials available on the web describe everything as if there will be no errors given by the system during the learning process and all information about Hadoop is always conveyed in an expert language and not in a simple easy to understand language which at times is difficult to grasp.

The above problem gave rise to have a tool that will provide you a useful learning experience. Dr,Eckberg (Chair of Thesis) expressed his desire to have a tool that will help the Computer Science students of our university to learn Hadoop by just using this tool and with that marked the birth of this tool, called an Educational Tool for Hadoop. The beauty of this tool is that all the tutorials are in the form of pdf which can be opened through the tool and saved for further use at any point of time. The second advantage the tool provides is the simplicity of the language used in these tutorials because these tutorials have been practically designed and performed by the author and then described as simply as possible. The advantage of that it has been made with a typical student mindset in view, and has been created by a master's student very familiar with that mindset, so that the experience that it gives to the end user is hopefully exactly what he needs.

This tool basically covers three essential parts of the Hadoop Ecosystem namely Hive, HBase and Sqoop.Each section of the tool takes you through a complete learning process for every language covered in this ecosystem. All the practical examples and the outputs attached in the pdf documents that are accessed through this tool make learning even easier. This tool has been developed using JAVA along with the eclipse IDE. As all tutorials have been designed and performed by the author on a standard laptop and as all the outputs have been attached in the pdf's we can say that the tool fully illustrates the use of Hadoop along with Hive, HBase and Sqoop. Hadoop, plus components like Hive, HBase and Sqoop, is called a Hadoop Ecosystem.

Thesis Committee

Carl Eckberg, Thesis Chair, Department of Computer Science
Roger Whitney, Department of Computer Science
Fredrick Harris, Department of Electrical & Computer Engineering