**Contents**

1. Introduction

2. How it works.

3. Things to look after.

4. Useful Stuff

5. References

# 1 Introduction:

This passion project was an idea conceived from past experiences in big companies. Here in we monitored the coding style of each developer, logged their mistakes during the real-time coding and uploaded those on cloud. Using PIG scripts/Map Reduce programs and visualize the result using tableau/Chart.js. Idea is to get into Developers shoe and understand their issues.

# 2 How it works:

Implementation has been carried out in 3 phases.

1. Processing Log files and collecting useful information.

2. Applying analytics on the information collected in the previous step to get useful insights.

3. Display insights to the end user

## 2.1 Processing Log files and collecting useful information:

In our programming life we use different types of development tools such as Visual Studio, Net Beans, and Android Studio etc. Choice of the tool is largely driven by language we wish to use. While coding with such tools whenever there is any exception, a log file is generated at a predefined location by underlying tool. Developer tools have their own format defined for log files.Eg. Visual Studio has log file in xml format, Net Beans and Android studio have log files in text format. This log file contains detailed description about the error.

Our goal is to collect error data and apply process on it to generate useful information.

We decided to go for Big Data technology as we are dealing with large number of log files here.

Big Data would provide us with Parallel Distribution, Processing, and scalability.

A script is written in Pig Scripting language which does following tasks:

* Read every log file
* Extract data such as error type, user name, error description, date time etc.
* Dump this information in a container for further use.

At the end of above process, we have information such as error type, user name, error description, date time available to us to apply analytics.

To create a demo, we selected few log files of Visual studio tool and uploaded them on Azure Cloud. Log files have XML formatted data so we decided to go for Pig scripting as it has a very good support for data in XML format. We created a hadoop cluster in Azure cloud to Process log files and collect useful information. This information is stored at a temporary location on azure storage which will be used for further processing.

## 2.2 Applying Analytics:

Why Analytics?

* Analytics makes processes more strategic to the business
* Analytics can combine disparate data streams to improve business results
* predictive *and* prescriptive modeling

And it goes on and on and on…

Here we applied simple analytics on the information collected above using Pig script. We are going to calculate total number of unique exceptions.

This would give us some useful insight like:

* Exception having highest count should be given quick attention as most of the developers are facing it on the larger scale.
* If Exception is occurring repeatedly in a close time interval, we might say that developer is not understood / not aware of the underlying concept.
* We can visualize “Exception vs. Exception Count” at Technology level, Developer level, and Project level etc.
* Cases when developer is not able to finish task within given time frame, we can refer to above report to get insights.
* Based on the Exception Count we will be able to define areas where a developer needs to be given exposure/training.

A pig script would apply analytics and result of which would be stored in the Mongo DB for further display.

## 2.3 Display:

To display analytics to the end user we made a simple MEAN Stack application which will fetch data from Mongo DB database and display it in the form of Bar chart. We used Chart.js JavaScript library for the same.

# 3 Things to look after:

So far we have considered and implemented a simple analytics and realized the importance of it. In the similar way we can implement complex business analytics to get more insights.

We can also automate this process to make it near real-time.

# 4 Useful Resources:



Contains Input, Output files used, Pig script, Jar files needed to run pig script, images of azure etc.

# 5 References:

* [RUNNING APACHE PIG (PIG LATIN) AT APACHE HADOOP ON WINDOWS AZURE](https://cloudcelebrity.wordpress.com/2012/01/11/running-apache-pig-pig-latin-at-apache-hadoop-on-windows-azure/)
* [Learn to Build Modern Web Apps with MEAN](https://thinkster.io/mean-stack-tutorial)
* [Use Pig with Hadoop on HDInsight](https://azure.microsoft.com/en-in/documentation/articles/hdinsight-use-pig/)
* [What is Hadoop in the cloud? An introduction to the Hadoop ecosystem in HDInsight](https://azure.microsoft.com/en-in/documentation/articles/hdinsight-hadoop-introduction/)