3. RESEARCH METHODOLOGY & WORK PLAN

In our research we will follow qualitative method of research methodology, as our research is experimental based using real data. One of the most important steps in the qualitative research process is analysis of data. There are seven qualitative analysis tools: methods of constant comparison, keywords-in-context, word count, classical content analysis, domain analysis, taxonomic analysis, and componential analysis. Constant comparison analysis likely is the most commonly used type of analysis for qualitative data in computer science. To perform a constant comparison analysis, the researcher first reads through the entire set of data (this also could be a subset of the data). After doing so, the researcher chunks the data into smaller meaningful parts.

Then, the researcher labels each chunk with a descriptive title or a “code.” The researcher takes pains to compare each new chunk of data with previous codes, so similar chunks will be labeled with the same code. After all the data have been coded, the codes are grouped by similarity, and a theme is identified and documented based on each grouping. In our research data analyzed need to use more than one data analysis method. For example, if the researcher had utilized word count alone, and focused on “different,” the issue of advocating would not have emerged. Likewise, if the researcher stops with only one domain, as shown in the text, the concept of taking years to understand would not emerge.

Thus, we contend that in our research need to utilize at least two, if not more, types of data analysis tools in order to triangulate results.

As our research is experimental based we will use some sample case studies and implement our idea in that. To implement our research we will use following tools and technology:

**IoT Beacons:**
Beacons are small wireless sensor devices that continuously transmit a simple radio signal”. In most cases the signal is picked up by nearby smart phones using Bluetooth Low Energy (BLE) technology or WIFI signals. iBeacon is a brand name created by Apple Inc. It is a new technology that Apple has implemented in its location framework in iOS 7 and newer operation systems. It lets Apple phones that are using iOS7 (or newer versions) constantly scan the environment for Bluetooth Low Energy devices such as beacons.
Apache Spark for real time data analytics:

Apache spark is a powerful open source processing engine provides real time stream as well as batch processing with speed. It is 100 times faster than Map Reduce. Provides API’s in java, python and Scala. Generality combine SQL using Spark SQL and can be used for streaming and complex analytics. In our research we will use spark eco system for real time data analytics as well as batch processing.

Flume:

This tool is used for the collection, aggregation of streaming event data. This tool is more reliable, scalable, manageable, and customizable, also gives high performance. In our research we will use this tool to retrieve data from beacons in real time.

MLib:

MLib is a spark subproject providing machine learning primitives. It is built on Apache spark, a fast and general engine for large scale data processing. MLib is Spark’s scalable machine learning library consisting of common learning algorithms and utilities, including classification, regression, clustering, collaborative filtering, and dimensionality reduction. In MLib v1.0, we support both dense and sparse input in a unified way, which introduces a few breaking changes. If the data is sparse, we store it in a sparse format instead of dense to take advantage of sparsity in both storage and computation.

Spark SQL:
Spark SQL is a Spark module for structured data processing. Unlike the basic Spark RDD API, the interfaces provided by Spark SQL provide Spark with more information about the structure of both the data and the computation being performed. Internally, Spark SQL uses this extra information to perform extra optimizations. There are several ways to interact with Spark SQL including SQL, the Data Frames API and the Datasets API. In our research we will use Spark SQL for analyzing JSON format data by creating data frames and data sets.

6.1 Work Plan
In this research we will use IoT devices (Beacons) to detect signals in the form of WIFI or Bluetooth from any nearby smart phones. After this we will retrieve sensor data from IoT devices and transform it into JSON data format by using flume/Kafka. Now this JSON data we will analyze in real time using apache spark eco system. Our objective is to find valuable information from the JSON data, visualize this information on dashboard and then connect it to internet. Use this information for sale promotion in a real time. We will analyze this data in following ways:

So by following above work plan and methodology purposed system will be able to do following task:
1. Automatically detect the mobile phone and engage the customer.
2. Provide real value with the right content at the right time and location.
3. Help users find info so they can easily come for shopping.
4. Deliver only relevant experiences by acknowledging customer references.