International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-6, August, 2019

# Groundwater Quality Mapping using Spark-GIS

**Devidas, Balasubramani.R**

***Abstract*: Contamination of underground water has become a common problem. Various chemicals get mixed in the water and make it contaminated. Certain effective measures have to be taken in order to clean these contaminants. For this to happen, the regions where the chemical contents are more have to be found out. This is implemented using Apache Spark framework and GIS tools.***

***Keywords* : Apache Spark, PySpark, GIS.**

## I. INTRODUCTION

This is an International reputed journal that published With the increasing number of population, it has led to various side effects. One of them includes the contamination of the ground water. The ground water may contain few of the dangerous chemicals such as Fluoride, Nitrate and Iron. In order to perform necessary measures, the quantity of these contaminants has to be determined. This is achieved in the proposed work, using a framework called Apache Spark. All the necessary data across the state Karnataka is collected and processed using Apache Spark framework. The processed data is then pictured on a map over the GIS tool (Google Maps) is created based on the processing of surveyed data of Groundwater Quality of Shallow Aquifers of India, CGWB, 2010.

the final paper/camera ready submission. It is be sure that contents of the paper are fine and satisfactory. Author (s) can make rectification in the final paper but after the final submission to the journal, rectification is not possible. In the formatted paper, volume no/ issue no will be in the right top corner of the paper. In the case of failure, the papers will be declined from the database of journal and publishing house. It is noted that: 1. Each author profile along with photo (min 100 word) has been included in the final paper. 2. Final paper is prepared as per journal the template. 3. Contents of the paper are fine and satisfactory. Author (s) can make rectification in the final paper but after the final submission to the journal, rectification is not possible.

## II. APACHE SPARK

Apache spark is a quick, open source dispensed popular cause cluster computing framework.

Revised Manuscript Received on October 30, 2019.

* Correspondence Author

**Devidas ,** B.E. and M.Tech degrees from Visvesvaraya Technological University, Belagavi, India,

**Dr. Balasubramani R,** B.E. and M.Tech degrees from Visvesvaraya Technological University, Belagavi, India,



The fundamental leap forward from Hadoop is that, the spark uses in-memory processing of statistics pipelines and allows sharing between the processing steps. Spark permits 4 distinctive kinds of processing and statistical analysis. Spark finds it areas of usage in:

• Streaming: Streaming is relevant in processing incoming data in close to actual time.

• Batch: The jobs such as large map-reduce jobs that are performed by manipulating huge datasets are done in this mode.

• Interactive: Interactive is specially used for surveying the facts as huge chunks of information are in reminiscence and because of the very brief reaction time of spark.

• Iterative: Machine learning algorithms such as a gradient descent work in this mode where convergence is achieved by repetitively accessing the data.

## III. PYSPARK

Pyspark isn't always a transliterated model of spark on a java enabled dialect of python consisting of Jython. Pyspark gives incorporated api bindings round spark and it also encourages the full practice of using python surroundings among all the nodes in the cluster with the help of pickle python serialization and, prominently, elements get right of entry to the rich ecosystem of python's libraries which includes: information processing inclusive of Pandas or the Scikit-Learn. When a spark program is initialized, it first has to generate the SparkContext item. It states spark the way to get right of entry to the cluster. PysparkContext is created using the python. The spark JVM SparkContext is bound to the python software by a gateway called Py4j. The JVM SparkContext serializes the application codes and closures and then forwards them to cluster for the further execution. The cluster supervisor assigns assets & schedules, and then ships the closures to the spark employees within the cluster who spark off python virtual machines as required. The spark-worker is controlled similar to an executor controlling storage, cache and the computation.

## IV. METHODOLOGY

The data is read in Apache Spark (pyspark) into a DataFrame (rows and columns) and cleaned the data which is not required for the analysis. Different tasks are performed for identifying the region of Karnataka affected by the contaminants like Fluoride, Iron and Nitrate. The data after the tasks performed are visualized for a better understanding of regions affected by the contaminants based on lower, middle and upper levels of contamination. Finally, a mapped layer of these data is created onto the Google Maps for picturing the affected regions for a better understanding of the problem.

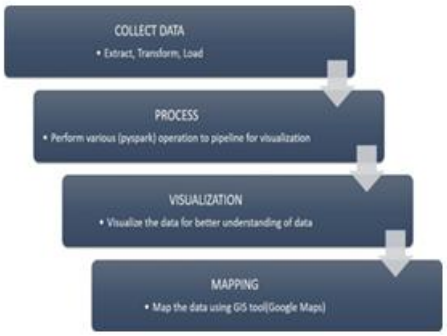


**Fig1: Stepwise Methodology**

## V. RESULTS

### A. Fluoride


**Fig2: Mapped areas of Fluoride based on different districts of Karnataka**

The map in *Fig2* shows the area of Fluoride contamination in the known districts of Karnataka. The following *Fig3* shows the bar graph of total contaminated by Fluoride in Karnataka.

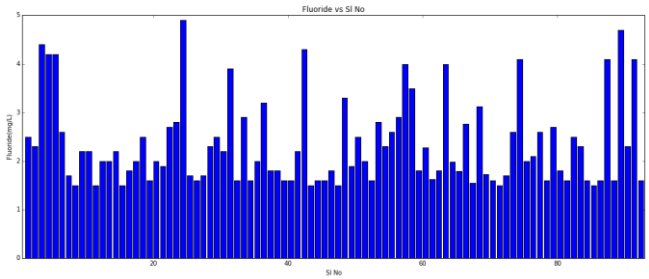


*Fig3: Bar Graph representation of areas contaminated by Fluoride*

### B. Iron

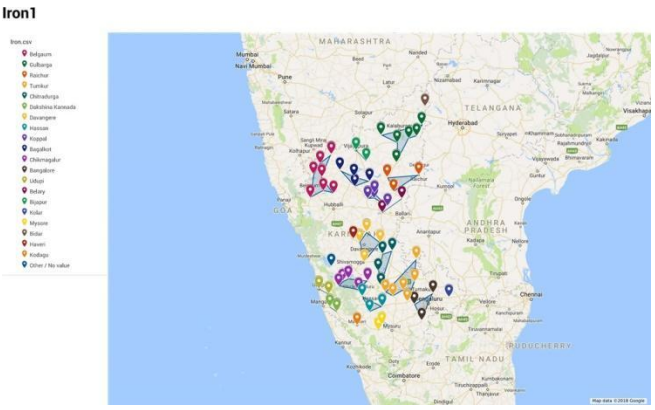


**Fig4: Mapped areas of Iron based on different districts of Karnataka**

The map in *Fig4* shows the area of Iron contamination in the known districts of Karnataka. The following *Fig5* shows the bar graph of total contaminated by Iron in Karnataka.

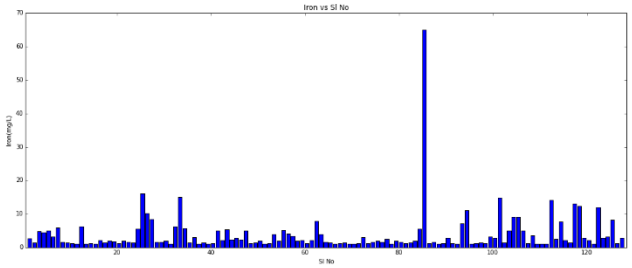


*Fig5: Bar Graph representation of areas contaminated by Iron*

### C. Nitrate

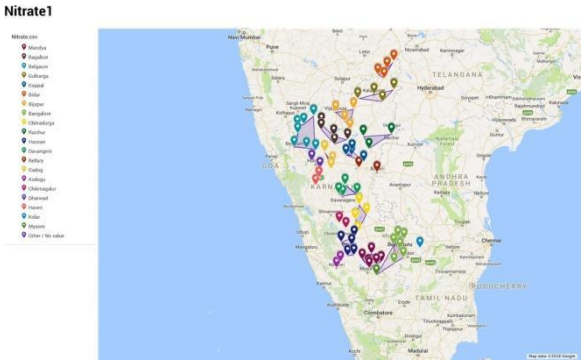


**Fig6: Mapped areas of Nitrate based on different districts of Karnataka**

The areas in the known districts of Karnataka where the Nitrate contamination in the soil is detected are marked in the map above in *Fig6.* Total contamination of Nitrate is represented in a bar graph as shown below in *Fig7.*

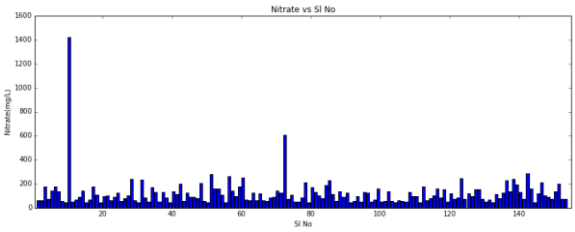


**Fig7: Bar Graph representation of areas contaminated by Nitrate**

## VI. RESULTS

The proposed work is intended to find the regions around Karnataka affected by contaminants like Fluoride, Iron and Nitrate. These regions are pictured on a map using GIS tool (Google Maps). From these data the conclusions are drawn as below:

- The level of Fluoride contents in groundwater is majorly concentrated in Central and Northern regions of Karnataka while also considerable regions affected are towards the Eastern boundary of Karnataka.
- The level of Nitrate contents in groundwater lay upon a very huge area of Karnataka regions except the coastal regions.

The level of Iron contents in groundwater is almost entirely attributed along all the regions of Karnataka.

Note that the equation is centered using a center tab stop. Be sure that the symbols in your equation have been defined before or immediately following the equation. Use "(1)", not "Eq. (1)" or "equation (1)", except at the beginning of a sentence: "Equation (1) is . . ."

### A. Figures and Tables

*Positioning Figures and Tables:* Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation "Fig. 1", even at the beginning of a sentence.

**Table- II: Name of the Table that justify the values**

| Table Head | Table Column Head | | |
|---|---|---|---|
| | *Table column subhead* | *Subhead* | *Subhead* |
| copy | More table copy[a] | | |

a. Sample of a Table footnote. (*Table footnote*)

The figure, graph, chart can be written as per given below schedule.

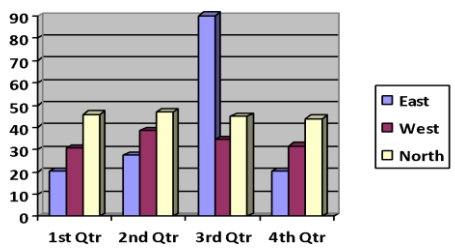


**Fig. 1.Example of a figure caption. (*figure caption*)**

## VII. FUTURE WORK

In the proposed work, Fluoride, Iron and Nitrate contaminated across Karnataka are determined. This process can be replicated with the other states of India and based on their contaminant levels an area can be specified based on different permissible levels of contaminant and visualized and mapped.

## REFERENCES

1. A. M, ""Spark SQL: Relational Data Processing in Spark"," New York, W.-K. Chen, Linear Networks and Systems (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.
2. B. J. Y. B. S. E. V. S. F. J. T. A. Meng X., "Machine Learning in Apache Spark," Journal of Machine Learning Research, vol. 34, no. 17, pp. 1-7, 2016.
3. M. Zaharia, "Apache Spark," Communications of the ACM, New York, 2016.
4. S. S. T. Bokhari M. U., "Metrics for Requirement Engineering and Automated Requirement Tools," in *INDIACom*, New Delhi, 2011.
5. G. M. F., "Geographical Information Science," *International Journal of Geographical Information Systems*, vol. I, no. 6, pp. 31-45, 1992.
6. N. R. B. J. R. P. J. R. V. Miguel R. Luaces, "A Generic Framework for GIS Applications," 2004.
7. G. J. K. M. A. J. R. A. A. K. P. Bhattacharya, "Epidemiological Study of Arcenicism in the City of Antofagasta," *Trace Substances in Environ Health*, vol. 5, pp. 13-24, 1971.
8. Z. W. Z. Z. Shen Zhaoli, "Basis of Hydrochemical," Beijing: Geologicall Publishing House, pp. 83-84, 1993.
9. C. H. L. L. Ma Zhenmin, "Controlling Effect of Quaternary Hydrogeological Structure on Contaminating Sensitivity of Shallow Groundwater," Earth Science- Journal of China University of Geosciences, vol. 25, pp. 472-475, 2000.
10. D. R. H. S. E. R. Doug Cain, "Priliminary Evaluations of Regional Groundwater Quality in Relation to Land Use," *Ground Water*, vol. 2, pp. 230-244, 1989.
11. B. B. Sikharani Hazarika, "Assessment of Groundwater Quality with Reference to Fluoride, Arsenic and Iron in Tea Garden Belt of Lakhimpur District, Assam, India," *Pelagia Research Library*, vol. 4, pp. 93-97, 2013.
12. E. A. I. R. S. A. Z. W. Y. W. N. F. A. I. A. F. Sabahi, "Americal Journal of Environmental Sciences," vol. 5(3), pp. 256-266, 2009.
13. B. B. S. H. P. Borah K. K., "International Journal of Chemical Sciences," vol. 7(3), pp. 1563-1574, 2009.
14. M. J. F. M. Meloun M., "Chemometrics for Analutical Chemistry," PC-aided Statistical Data Analysis; Ellis Horwood Ltd: Chinchester, England, vol. 1, 1992.
15. A. W. APHA, "Standard Methods for the Examination of Water and Wastewater," APHA, AWWA, WPCF, New York, Washington DC, 1992.
16. B. B. S. H. P. Borah K. K., "Archives of Applied Science Research," vol. 1(2), pp. 159-164, 2009.

## AUTHORS PROFILE

**Devidas ,** received the B.E. and M.Tech degrees from Visvesvaraya Technological University, Belagavi, India, in 2006 and 2010 in Information Science and Computer Science and Engineering respectively.He is currently pursuing his Ph.D in the area of Data Mining. His area of Interest are DAta mining,Machine Learning and Data Science.

**Dr. Balasubramani R,** obtained PhD in Image Processing and had published more than 50 research papers in reputed conferences and journals. His research interests include Artificial Intelligence, Machine Learning, Deep Learning and, Data Science. He is the member of various professional bodies like ACM, IAENG and CMI. Currently he is guiding eight research scholars for PhD.