**STORAGE AND ORGANIZATION OF LARGE GEOSPATIAL DATASETS ON A DISTRIBUTED BLOCKCHAIN**

***A Report submitted***

***in partial fulfillment for the Degree of***

**B. Tech**

**in**

**Computer Science and Engineering**

***by***

**SHIVANSHU SINGH**

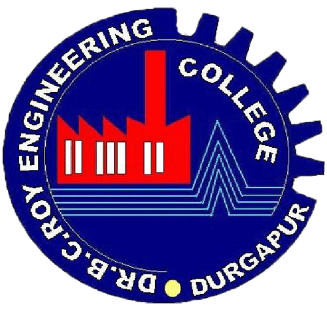
**AYUSH TAH**

pursued in

Department of Computer Science and Engineering

**DR BC ROY ENGINEERING COLLEGE, DURGAPUR**

To



**DR BC ROY ENGINEERING COLLEGE**

**DURGAPUR**

**NOVEMBER, 2022**

**CERTIFICATE**

This is to certify that the project report entitled “Storage and Organization of large geospatial datasets on a distributed Blockchain” submitted by Ayush Tah and Shivanshu Singh to the Dr BC Roy Engineering College, Durgapur, in partial fulfillment for the award of the degree of B. Tech in (Computer Science and Engineering) is a bona fide record of project work carried out by him/her under my/our supervision. The contents of this report, in full or in parts, have not been submitted to any other Institution or University for the award of any degree or diploma.

……………………………

Prof. Debkanta Chakraborty

Supervisor,

Department of Computer Science and Engineering

Durgapur ……………………………..

November, 2022

**DECLARATION**

I declare that this project report titled “Storage and Organization of large geospatial datasets on a distributed Blockchain” submitted in partial fulfillment of the degree of B. Tech in (Computer Science and Engineering) is a record of original work carried out by me under the supervision of Prof. Debkanta Chakraborty, and has not formed the basis for the award of any other degree or diploma, in this or any other Institution or University. In keeping with the ethical practice in reporting scientific information, due acknowledgements have been made wherever the findings of others have been cited.

………………………………..

Ayush Tah

12000119129

………………………………..

Shivanshu Singh

12000120131

Durgapur

Dd/mm/yyyy

**ACKNOWLEDGMENTS**

We take this opportunity to thank Dr. Sanjay S. Pawar, Principal - BCREC, Prof.

Debkanta Chakraborty – Mentor, and other faculty members who helped in preparing the guidelines.

I extend my sincere thanks to one and all of BCREC family for the completion of

this document on the project report format guidelines.

Shivanshu Singh

Ayush Tah

**ABSTRACT**

With the current rapid expansion and development of global navigation systems, information and communication technology, and sensor technology and sheer number of geo satellites being launched an unprecedented amount of geospatial data is being generated every smallest of intervals. The geospatial data we’re talking about consists of but not limited to individual trajectories, location-based services, remote sensing images and so on, thus there has been a persisting need for the ability of sharing and use of spatial information reliably. As any data of such magnitude and importance is concerned data privacy, integrity and security play a crucial role both in sharing and storage of such data, however, it is quite a challenge to guarantee data privacy and security in geospatial data when is so extensively used in myriad of technologies. In addition, geospatial data is stored in many different formats and calibrated by different standards. Any effort to compare, combine or map data first requires a significant amount of data scrubbing and reformatting thus making a claim to the integrity of data becomes difficult. Data integrity depicts the data consistency and accuracy while sharing or storing the data, which quantifies the validity and fidelity of data. To preserve privacy and integrity of geospatial data, geospatial data should be validated to satisfy the security requirements, and be protected from unauthorized modifications.

Recently Blockchain has been realized as a possible solution to the problems we face with geospatial data because of the tamper-proof, traceable, trust-free, transparent, and decentralized characteristics is possess. Blockchain achieves this de-centralized and security characteristics by integrating consensus mechanism, asymmetric cryptographic algorithms, distributed data storage to name a few. Hence, Blockchain presents itself as a major player in finding a solution for the problem we’re trying to tackle.

**TABLE OF CONTENTS**

**DESCRIPTION PAGE NUMBER**

CERTIFICATE

DECLARATION

ACKNOWLEDGEMENTS

ABSTRACT

LIST OF FIGURES

LIST OF TABLES

ABBREVIATION/NOTATION/NOMENCLATURES

# CHAPTER 1: INTRODUCTION

## Geospatial Data

### Vector Data

### Raster Data

## Blockchain

## Motivation

**LIST OF FIGURES**

**FIGURE TITLE PAGE NUMBER**

**1.1 Vector Data x**

**1.2 Raster Data xi**

**1.3 A Block in Blockchain xi**

**1.4 Genesis Block xii**

**LIST OF TABLES**

**ABBREVIATIONS/ NOTATIONS/ NOMENCLATURE**

**CHAPTER 1**

**INTRODUCTION**

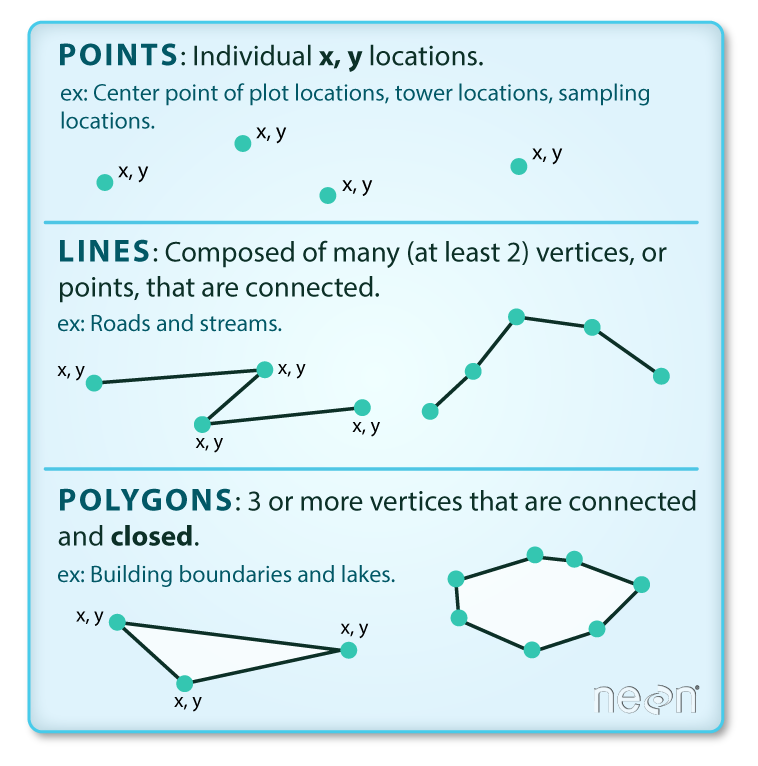
Before we dive in to this, let’s just see a brief introduction to all the technological terms and technologies involved in this project.

**1.1 GEOSPATIAL DATA**

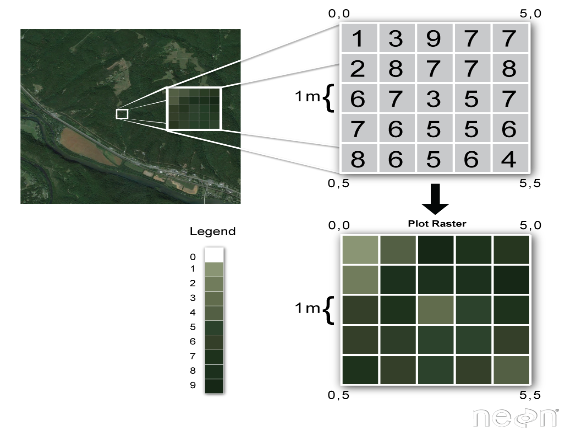
Geospatial data contains the data associated with a place on earth that contains but isn’t limited to latitude and longitude(location), characteristics of objects, event or phenomenon(attributes) with temporal information(time) all of which we can use to draw a map of the said place. The location info may/may not be static or dynamic in nature.

There are two main types of primary geospatial data models:

* **Vector Data**: Co-ordinate based model where geographical locations are represented as points, lines, or polygon features. This works best when used to represent geographical features with defined boundaries. Each geographical feature has associated tabular/spreadsheet data(attributes).

 (Fig 1.1)

* **Raster Data:** Mostly used to represent dynamic data (e.g., temperatures) but can also be used to represent discrete data (e.g., land cover). It uses equal sized pixels like blocks arranged in a grid to represent geographic data.

(Fig 1.2)

* Minor data types that are part of the geospatial dataset include point clouds, census data, cellphone data, social media data etc.

Some common file formats that are associated with this dataset are:

• shapefiles (.shp)

• file geodatabases (.gdb)

• Keyhole Markup Language (.kml or .kmz)

• GeoTIFF (.tiff)

• comma-separated values (.csv)

• GeoJSON (.json)

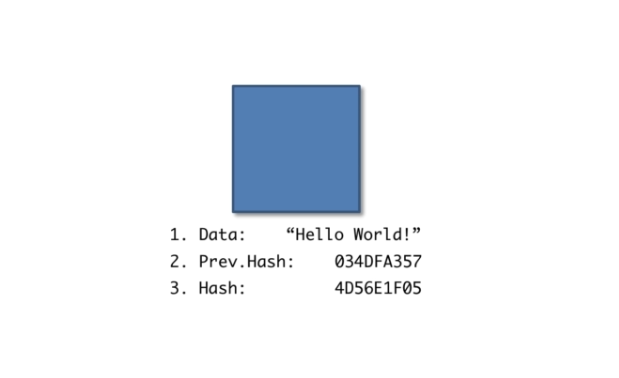
• GPS Exchange Format (.gpx)

• LiDAR point cloud data (.las)

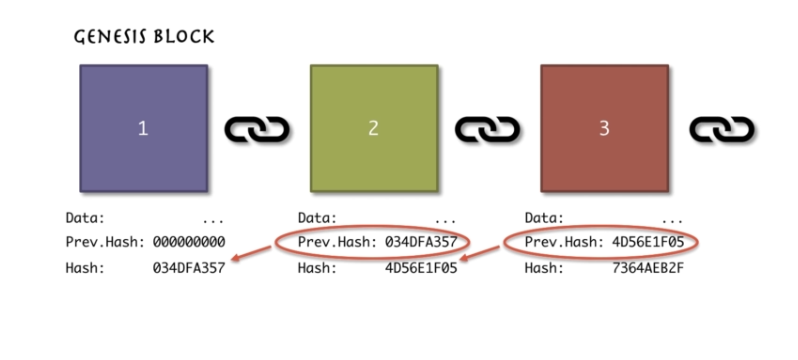
We will work with the (.tiff) format as this was the file type present in the dataset provided in the project on SIH where we took it form.

**1.2 BLOCKCHAIN**

A blockchain is type of distributed ledger technology (DLT) that has a continuously growing list of records, called blocks, which are linked and secured using cryptography. A block contains a data and contains a previous hash and its own hash. It was created by a person or group using the pseudonym Satoshi Nakamoto in 2008 to serve as the public distributed ledger for bitcoin cryptocurrency transactions, based on previous work by Stuart Haber, W. Scott Stoenetta and Dave Bayer.

 (Fig 1.3)

The first block in a chain is called the "Genesis Block". It doesn’t have any previous hash just its own. The blocks that come after the genesis block has previous hash.

 (Fig 1.4)

This is the reason blocks are called to be cryptographically linked together.

**1.3 MOTIVATION**

After a lot of time doing research on what should be our final year project should be. We stumbled upon this problem on the SIH platform and as we wanted to invest our times in something that had substance and was beneficial to our world. Even though we may not be able to realize the entire scope of what we set out to achieve we will definitely try our best to put a dent in this problem.

**Geospatial applications by industry**

* Governments can take insights about health, disease and weather and use them to better advise the public when a natural disaster strikes, or an emergency health event occurs.
* Electric utilities providers can use data to help predict possible service disruptions and optimize maintenance and crew schedules.
* Insurers can do a more accurate job of projecting risks and warning policy holders about potential issues they may soon be facing.
* Farm and agricultural lenders can improve the methodology they use to assess credit risk scores and reduce bad loan placements.
* COVID-19 mapping - COVID-19 mapping is performed using geospatial analytic models, based on population data, livestreaming video, maps and weather.
* Vegetation management - Through user defined functions (UDFs), geospatial analytics enables those involved in vegetation management to assess water and moisture levels.
* Tornado tracking - User defined functions are also useful at helping meteorologists work with incoming data to chart the path of tornadoes that could be moving through an area.
* Wildfire mapping - Having relevant data, such as satellite imagery, census data and wind forecasts, in one platform lets incident commanders chart wildfire growth and movement.