

Capstone Project Write Up - Cosmic Microwave Background Data

Abel Flores Prieto

May 6, 2021

Scope

This project focuses on the data used for computing the Cosmic Microwave Background or CMB (see [1](#)). In this project I focus to easily enable access to data used for computing the CMB. The data used for this project is simulated by [Sehgal et al. \(2010\)](#) and can be found [here](#). This simulated data works as a tool to cross-check multiple calculations from real surveys, including power spectrum calculations.

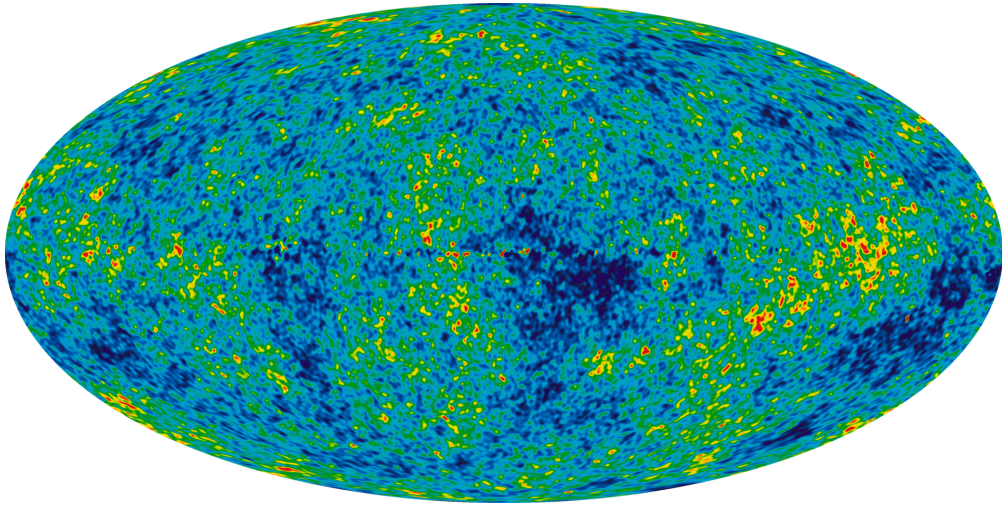


Figure 1: Cosmic Microwave Background temperature, remnant of the beginning of the universe.

The data can be downloaded from the [Legacy Archive for Microwave Background Data Analysis](#). The data used in this project is simulated by a team of researchers and includes realistic simulations of the microwave sky. The data is composed of Halo Catalogs and Galaxy Catalogs.

The end solution of the project is to have a data lake in S3 so that I, or anyone (with the right permissions) can access the data for further processing. The data lake will sit in one of my S3 buckets, `s3://abelfp-physics-data/` and will have two main tables, with the possibility to expand: `cmb_simulated` for simulated galaxy catalogs, and `halo_simulated` for the halo

catalogs. The galaxy catalogs come in DAT files with over 80 GB of data while the halo catalogs come in as ASCII.GZ files with over 800 MB of data. All of this data was uploaded to `s3://abelfp-physics-data-raw/`. The catalogs come with a description of the data and what each column represents. These files can be found under the `/textttdata/` directory in this project, all descriptions are TXT files.

Galaxy Catalogs

Halo Catalogs