# Eco-Friendly Hadoop

## Shiv Kumar[1], Shrawan kr.shrama[2],Hitesh kumar swarnkar[3]

[1,2]Computer Science & Engineering, Mewar University, Gangrar, Chittorgargh-312901, India.
[3]Faculty of Computer Application, Pacific University, Pacific HillUdaipur-313024, India

*Abstract : Hadoop is a framework to handle large amount of data. Most important part of Hadoop architecture is Map Reduce method to process the distributed data and second most important part is Hadoop distributed file system to store distributed data. The way Hadoop framework handles or manipulates or extract data we can say that it is eco friendly because it provides same CPU time to all job using JOB FAIR SCHEDULING algorithm to save power or work energy. Similarly Hadoop frame work also shares Hadoop cloud using Capacitive scheduling algorithm. So, once again electric power is saved using resource sharing. We can save energy, resources, cost and materials also by using Green building techniques, solar energy and time sharing.*

*In this way, we can save our environment or we will reduce the effect on environment by using Hadoop framework, Green building methods, renewable energy like wind and solar, resource sharing methods as well as by time sharing. So, Our Eco system is not affected or if affected then in minimized amount which can be ignore. So, we can say that Hadoop is a Eco friendly frame work.*

*Keywords: hadoop; energy techniques; green building techniques; resource sharing techniques; time sharing techniques, hadoop as service techniques*

## I. INTRODUCTION

There are two major words in the title of this paper. First one is Eco friendly Techniques while second one is Hadoop

**Eco Friendly Techniques.**

Eco system is directly related to green color of natures. Natures color will be green if and only if environment of the earth will be clean. Due to industrialization and urbanization Earth's environment is degrading day by day and second by second.

- Energy Techniques
- Green Building Techniques
- Chemical Techniques
- Biological Techniques and others

**Green energy** is energy that is produced in such a way as to minimize its negative impact on the environment. Traditional energy sources, most notably fossil fuels, produce greenhouse gases that are believed to be the primary cause of an effect known as global warming or climate change. Sources of green energy, such as solar, wind, geothermal, and hydro energy, are developed and promoted as alternative sources that make little or no contribution to climate change. Even nuclear energy is sometimes considered a green energy source, because some types of nuclear technology produce much less waste than oil and coal, The industrial revolution that began in the second half of the 18th century changed the world, with new methods of machine-based manufacturing leading to a profound increase in economic growth, population growth, and quality of life. At the time, the long-term consequences could not even be conceived of, much less felt; however, the burning of vast quantities of fossil fuels, such as coal and oil, has caused a great deal of harm to the environment. Most climatologists agree that the use of fossil fuels has contributed significantly to global warming. This term refers to the measured increase in the Earth's surface temperature since the late 19th century and the environmental effects of this change[1].

**Green building,** or sustainable design, is the practice of increasing the efficiency with which buildings and their sites use energy, water, and materials, and reducing building impacts on human health and the environment over the entire life cycle of the building. Green building concepts extend beyond the walls of buildings and can include site planning, community and land use planning issues as well[2].

**Green chemistry** is to reduce chemical related impact on human health and virtually eliminate contamination of the environment through dedicated, sustainable prevention programs. Green chemistry searches for alternative,

environmentally friendly reaction media and at the same time strives to increase reaction rates and lower reaction temperatures. Green chemistry reduces pollution at its source by minimizing or eliminating the hazards of chemical feedstock's, reagents, solvents, and products[3].

**Green computing** is the environmentally responsible and eco-friendly use of computers and their resources. Green computing is the environmentally responsible and eco-friendly use of computers and their resources. In broader terms, it is also defined as the study of designing, manufacturing / engineering, using and disposing of computing devices in a way that reduces their environmental impact[4].

## II. HADOOP

The Apache™ Hadoop® project develops open-source software for reliable, scalable, distributed computing. The Apache Hadoop software library is a framework that allows for the distributed processing of large data sets across clusters of computers using simple programming models. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage. Rather than rely on hardware to deliver high-availability, the library itself is designed to detect and handle failures at the application layer, so delivering a highly-available service on top of a cluster of computers, each of which may be prone to failures [5].

The project includes these modules:

❖ **Hadoop Common**: The common utilities that support the other Hadoop modules.
❖ **Hadoop Distributed File System (HDFS™)**: A distributed file system that provides high-throughput access to application data.
❖ **Hadoop YARN**: A framework for job scheduling and cluster resource management.
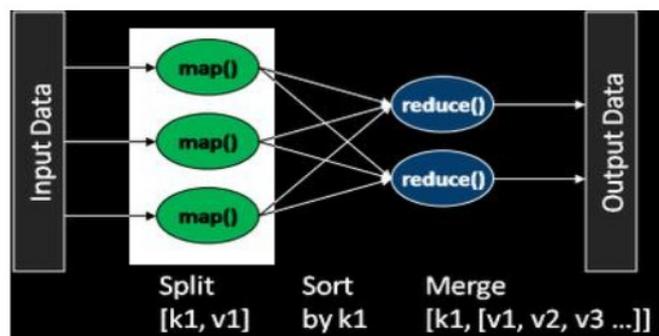❖ **Hadoop MapReduce**: A YARN-based system for parallel processing of large data sets (Fig. 1).



Figure 1: Working of Map Reduce

**2.1 Implementing Eco Friendly techniques in hadoop**
  **A. Map Reduce Technique as Eco Friendly *Techniques.***

Map Reduce is a framework used by Hadoop to handle large data. Map Reduce is not using general job scheduling algorithm like FCFS, SJF, PRIORITY SHEDULING, ROUND ROBIN SCHEDULING etc. to handle user's job. Fair scheduling is a method of assigning resources to jobs such that all jobs get, on average, an equal share of resources over time. When there is a single job running, that job uses the entire cluster. When other jobs are submitted, tasks slots that free up are assigned to the new jobs, so that each job gets roughly the same amount of CPU time.

While default Hadoop scheduler forms a queue of jobs, this lets short jobs finish in reasonable time while not starving long jobs. It is also an easy way to share a cluster between multiple of users. Fair sharing can also work with job priorities - the priorities are used as weights to determine the fraction of total compute time that each job gets. So, we can also customize FAIR SCHEDULING [6] algorithm on the basis of our requirements (Fig. 2).
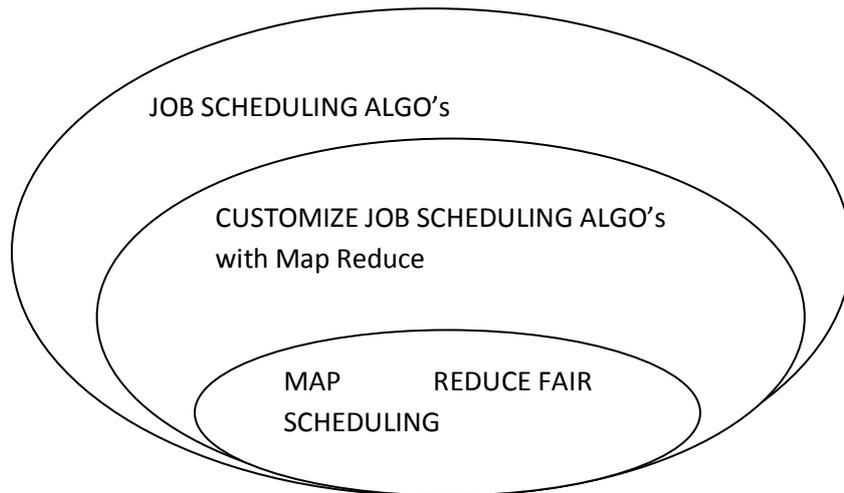
Figure 2: Map Reduce algorithm

In this way we can say that every user's jobs get same amount of time whether user is special one or general which saves users time and energy as well as electric power used. That is why it is Eco friendly in nature.

There is several green energy techniques like. Solar energy, Wind energy, Bio gas energy etc. Others Example are Geothermal Energy, Ocean (Tidal) Energy, Hydropower, Biomass, Land Fill Gas (LFG), Municipal Solid Waste (MSW)

But Solar energy(fig. 3) and Wind energy are mostly used in industries as green energy techniques because they do not disturbs environment and don't produces waste material like thermal power plant. As we know that Solar/Wind techniques produces DIRECT CURRENT. So we have to use DC/AC inverter to convert DC to ALTERNATIVE CURRENT.



Figure 3: Solar Energy

Our data center can generate its own SOLAR ENERGY using solar power computer system (Fig. 4). We want to save energy by generating because of following reasons [7]:
1)      Because energy loses due to power generation and transmission is about 40%.
2)      The ability to survive grid outage in developing countries.
3)      Due to lower cost of establishment.
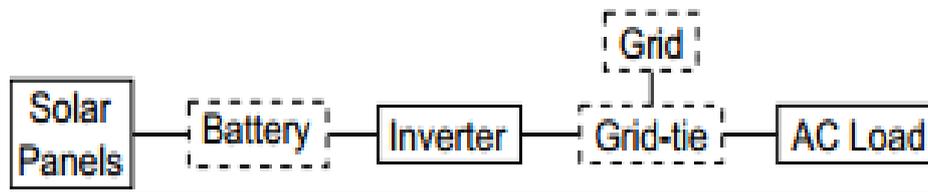


Figure 4: Solar Power Computer System

**B.      Green Building Techniques as Eco Friendly Techniques.**
        The growth and development of our communities has a large impact on our natural environment. The manufacturing, design, construction, and operation of the buildings in which we live and work are responsible for the consumption of many of our natural resources [8].

| DESIGN | ADVANTAGES |
|---|---|
| Seal Cable Cutouts | Lower energy usage |
| Hot & Cold Aisles | Lower storage temperature |
| Min. Gaps in Rows | System reliability increases |
| Use of Blanking Panels | Increases life of data centers |
| Use Longer Rows | Uptime is better |
| Use Area Density That Matches Air Flow Requirements | Increase number of server or storage density. |
| Use Power Management Features Efficiently | To save power |
| Use Power Efficient Platform | |

TABLE 1:

### C. Resource Sharing Techniques as Eco Friendly Techniques.

`       The central idea is that the available resources in the Hadoop cluster are shared among multiple organizations that collectively fund the cluster based on their computing needs [8]. There is an added benefit that an organization can access any excess capacity not being used by others. This provides elasticity for the organizations in a cost-effective manner.

Sharing clusters across organizations necessitates strong support for multi-tenancy since each organization must be guaranteed capacity and safe-guards to ensure the shared cluster is impervious to single rouge application or user or sets thereof. The Capacity Scheduler provides a stringent set of limits to ensure that a single application or user or queue cannot consume disproportionate amount of resources in the cluster. Also, the Capacity Scheduler provides limits on initialized/pending applications from a single user and queue to ensure fairness and stability of the cluster.[9]

The primary abstraction provided by the Capacity Scheduler is the concept of queues. These queues are typically setup by administrators to reflect the economics of the shared cluster.

To provide further control and predictability on sharing of resources, the Capacity Scheduler supports hierarchical queues to ensure resources are shared among the sub-queues of an organization before other queues are allowed to use free resources, there-by providing affinity for sharing free resources among applications of a given organization.

Here sharing can be done up to the grass root label from sharing of cluster to the sharing of vehicles, offices, hardware as well as software sharing

### D. Time Sharing Techniques as Eco Friendly Techniques.

Hadoop nodes should use the concept of time sharing. For example, when email, database, web, or ERP applications are idle, the compute power available should be transferred to Hadoop nodes that are analyzing improvements in business performance [10].

### E. Hadoop as a service Techniques as Eco Friendly Techniques.

Hadoop should be available to the company as a shared service. This is one of the most cost-effective ways to provide Hadoop as a service. In this model, it is available to all departments based on chargeback accounting. Even with shared services, virtualization still allows for enough isolation to meet independent business and security needs.

## III.      CONCLUSIONS

On applying all these techniques we are not only saving power i.e. electricity but also getting data quickly with less resource and cost with better performance  with respect to time without effecting the natures process or eco system. So, Hadoop is Eco friendly frame work.

## REFERENCES

[1] http://www.ifpaenergyconference.com/Green-Energy.html
[2] http://en.wikipedia.org/wiki/Green_building
[3] http://www.sigmaaldrich.com/chemistry/greener-alternatives/green-chemistry.html
[4] http://www.techopedia.com/definition/14753/green-computing
[5] http://hadoop.apache.org/
[6] http://hadoop.apache.org/docs/r1.0.4/fair_scheduler.html
[7] http://static.usenix.org/event/hotpower/tech/full_papers/Kaushik.pdf
[8] http://www.findthatpdf.com/search-84313247-hPDF/download-documents-Bauer_Building_the_Green_Data_Center_FINAL.pdf.htm
[9] http://hadoop.apache.org/docs/r2.4.0/hadoop-yarn/hadoop-yarn-site/CapacityScheduler.html
[10] http://blogs.vmware.com/vfabric/2013/04/why-virtualize-hadoop.html

**Shiv Kumar** received the M. Tech. degree in Computer Science and Engineering from Mewar University Chittorgargh in 2012. During 2007-2013, he stayed in Canon India Private limited Center of Excellence center and India Software Center Noida and Gurgaon of India. He knows with Mewar University, Chittorgargh, India.

**Shrawan Kumar Sharma** is currently pursuing masters degree program {M.Tech} in Computer science and engineering in Mewar University, India. His research interest includes soft computing, Mobile computing, DBMS, Network Security, Graphic programming, Fuzzy and Data Mining and etc.

**HITESH KUMAR SWARNKAR** is currently pursuing masters degree program {MCA} in Computer science in Pacific University, India. His research interest includes DBMS, Computer Network, programming, Artificial intelligence and data processing and etc.