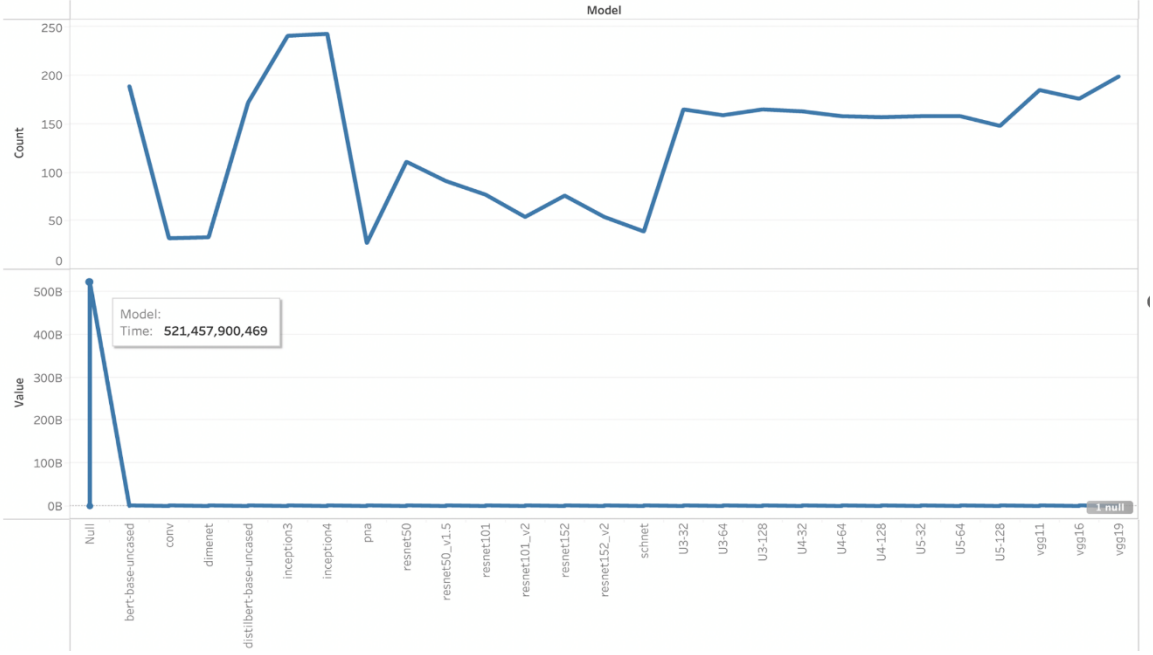
Cloud data Visualization

ABSTRACT

Abstract: Big data will be trans formative in every sphere of life. But just to process and analyze those data is not enough, human brain tends to find pattern more efficiently when data is represented visually. Data Visualization and Analytics plays important role in decision making in various sectors. It also leads to new opportunities in the visualization domain representing the innovative ideation for solving the big-data problem via visual means. It is quite a challenge to visualize such a mammoth amount of data in real time or in static form. In this paper, we discuss why big data visualization is of utmost importance, what are the challenges related to it and review some big data visualization tools.



1. Introduction:

Data visualization is the presentation of data in a pictorial or graphical format, and a data visualization tool is the software that generates this presentation. Data visualization provides users with intuitive means to interactively explore and analyze data, enabling them to effectively identify interesting patterns, infer correlations and causalities, and support sense-making activities. Data visualization can be categorized into two major sub-fields: information visualization and scientific visualization Information visualization is used to visually represent abstract data, such as business data .while scientific visualization represents scientific data, which are usually physical Both information and scientific visualization focus on how to transform data into a visual form, to become understandable information for gaining insight and knowledge. the fundamental process of data visualization, in which data in any form can be transformed into graphical images. When a user reads or looks at a graphical image, the image is interpreted through the human cognitive system for the acquisition of insight or the apprehension of useful information. information is data that has been given meaning by way of relational connection. This "meaning" can be useful but does not have to be. In computer parlance, a relational database makes for information from the data stored within it. Today businesses struggle to just store the massive amount of data whereas analyzing, interpreting and presenting it in meaningful ways is a thought for later . The main challenge of Big Data lies in capturing, storing, analyzing, sharing, searching, and visualizing data. One of the major aspect of Big Data analysis is that we can find interesting pattern in huge data set, but actually the result of the analysis is usually raw numbers and by those numbers it is very difficult to interpret anything. But if those numbers are represented visually then it becomes much easier for our brain to find meaningful patterns and take decision accordingly. shows the benefits of Big data visualization .

Data visualization is certainly not a new thing; it has been around for centuries. Data visualization is easy and quick way to convey messages and represent complex things . We humans are adapted to find patterns in everything we see. Since the data is mounting at such a massive rate the traditional ways of presenting data is obsolete [1]. Compared to traditional data, Big Data is characterized by 5Vs, i.e. huge Volume, high Velocity, high Variety, low Veracity and high Value. Actual challenge is not only to process this huge amount of data but to process data with high diversity. High diversity and uncertainty in data reduces the response time of the application as it has to deal with not only traditional structured data but also with semi and unstructured data.

Challenges:

Traditional visualization tools have reached to their limits when encountered with very large data sets and these data are evolving continuously. Though there are some extensions to traditional visualization approaches but they lag behind by miles. The visualization tool should be able to provide us interactive visualization with as low latency as possible. To reduce the latency, we can do the following things

- Use the pre-computed data
- Parallelize Data Processing and Rendering
- Use a predictive middleware

Big Data visualization tools must be able to deal with semi-structured and unstructured data because big data usually have this type of format. It is realized that to cope with such a huge amount of data there is a need for immense parallelization, which is a challenge in visualization. The challenge in the parallelization algorithm is to break down the problem into a such independent tasks that they can run independently.

The task of big data visualization is to recognize interesting patterns and correlations. We need to carefully choose the dimensions of data to be visualized, if we reduce dimensions to make our visualization low then we may end up losing interesting patterns but if we use all the dimensions we may end up having visualization too dense to be useful to the users. For example: "Given the conventional displays (1.3 million pixels), visualizing every data point can lead to over-plotting, overlapping and may overwhelm user's perceptual and cognitive capacities.

Due to the vast volume and high magnitude of big data, it becomes difficult to visualize. Most of the current visualization tools have low performance in scalability, functionality and response time Methods have been proposed that not only visualize data but processes at the same time. These methods use Hadoop and storage solution and R programming language as compiler environment in the model shows the outline of such a model.

Visualization tools:

Various tools have emerged to help us out from the above-pointed problems. The most important feature that a visualization must have is that it should be interactive, which means that user should be able to interact with the visualization. Visualization must display relevant information when hovered over it, zoom in and out panel should be there, and visualization should adapt itself at runtime if we select subset or superset of data. We reviewed some of the most popular visualization tools.

2. Tableau

u is an interactive data visualization toolthat is focused on Business Intelligence. Tableau provides very wide range of visualization options. It provides option to create custom visualization. It is fast and flexible. It supports mostly all the data format and connection to various servers right from the Amazon Aurora to Cloudera Hadoop and Salesforce. User interface is intuitive, wide variety of charts are available. For simple calculations and statistics one does not require any coding skills but for heavy analytics we can run models in R and then import the results into Tableau. This requires quite a bit of programming skill based upon the task we need to perform.

ORCID(s):

2.1.

2.1.1. Comparison of Tools

We have compared the above tools on the basis of various attributes. These attributes are as follows:

• Open Source: If our tool is open source [16] or not.

• Integration with popular data sources: These include, MapR Hadoop Hive, Salesforce, Google Analytics, Cloudera Hadoop etc.

• Interactive Visualization: Are visualizations created by