# 1 H20

H20 is a open source platform that can execute highly advanced and complex machine learning algorithms in faster and scalable way, regardless of the size, format and location of the data[5]. It achieves this by serializing the data faster between nodes and clusters that stores huge amount of data. Data processing is done in memory thus providing faster response[5]. It uses fine grain parallelism technique for processing of distributed data archiving 100x faster speed as compared to traditional mapreduce without compromising on accuracy [5].

# 2 Google Dremel

With vast amount of publicly available data over the internet / cloud , there was a need of technological system/framework that is deployed on cloud which can execute on demand queries in faster and scalable way for read only multi level nested data. Along with that a system that uses structured query language, which is widely adapted and extensively used by the developers for writing queries to avoid the learning curve of new language. To fill this gap Google came up with Dremel. It is a interactive ad hoc query system that lets the user query the large dataset providing them results with much faster speed compared to traditional technologies[4]. By combining multi-level execution trees and columnar data layout, it is capable of running aggregation queries over trillion-row tables in seconds[4]. Dremel is capable of scaling up to thousands of CPUs and petabytes of data[4].

# 3 Google Genomics

With the size of medical data getting increased exponentially from petabytes to exabytes rapidly, Google came up with Google Genomics as extension to Google cloud platform.It helps the life science community organize the worlds genomic information and make it accessible and useful[1]. Researchers are able to apply Google powerful technologies such as Google Search and Maps to securely store, process, explore, and share large, complex genomics datasets [1]. Multiple genome repositories data can be processed using Google Genomics within seconds as it is backed by Google bigtable and Spanner technologies[1]. It is based on open standard from Global Alliance of Genomics and Health achieving higher level of interoperability for genomics data[1]. It is fully integrated with Google cloud virtual machine, storage and SQL/NoSQL databases[1].It helps analysing Genomic data in real-time with BigQuery, in literate programming style with Cloud Datalab, in batch with GATK on Google Genomics, with Apache Spark or Cloud Dataflow, or with a Grid Engine cluster[1].

# 4 Google Vision

Google Cloud Vision API has made the herculean task of correct labeling/classification of images simple. With exponential increase in different types of data including images, voice , video are transformed into digital form, stored and transmitted over network. There was a dire need of automated technology solution that can correctly classify / label images with high level of confidence; Google Vision API provides such platform to researchers and developers. It quickly classify images into thousands of predefined meaning ful categories[3]. It does this by encapsulating powerful machine learning models ( KNN , Regression ) etc for classification of images. [3] It help detects objects and faces within images, finds and reads printed words contained within images through OCR[3]. Developers can build meta data on their image catalog, moderate offensive content, or launch new marketing campaign/scenarios through image sentiment analysis. It can be accessed through REST API. [3]

# 5 Google Cloud Spanner

Cloud Spanner is the enterprise-grade, globally-distributed database service built for the cloud[2]. This technology combines the benefits of relational database structure with non-relational horizontal scale. This is unique combination that delivers high-performance transactions and strong consistency across rows, regions, and continents with high availability and enterprise-grade security[2]. Cloud Spanner revolutionizes database administration and management and makes application development more efficient. [2].It is fully managed , can be easily deployed and has built in synchronous replication and maintenance feature[2]. It takes advantages of all critical features of relational databasesuch as schemas, ACID transactions, and SQL queries (ANSI 2011) thus reducing the need of high learning curve for developers who are well proficient in SQL. Client libraries that can connect to spanner is language independent. These libraries can be developed in C sharp, Go, Java, Node.js, PHP, Python, and Ruby. Already existing JDBC driver with popular third-party tools can be used to connect with spanner[2]. It is purposely built for global transactional consistency[2].

# References


[1] Google. Google genomics. Web.

[2] Google. Google spanner. Web.

[3] Google. Google vision api. Web.

[4] jlong gromer shiva mtolton theov melnik, andrey. Dremel: Interactive analysis of webscale datasets. Web.

[5] web. H2o. Web.