11$^{th}$ Annual Conference on Foundations of Nano-Science: Self-Assembled Architectures and Devices (FNANO14)

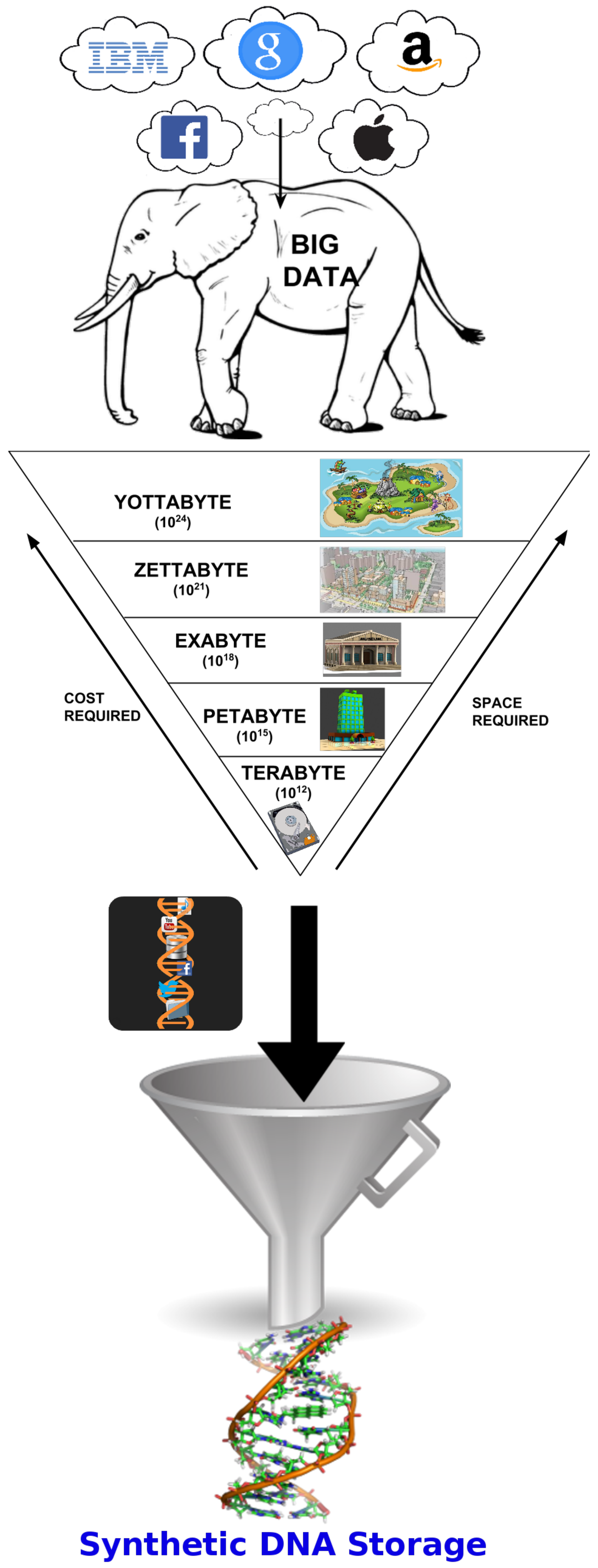
# DNACloud: A Potential Tool to Store Big Data on DNA

**Shalin Shah, Dixita Limbachiya and Manish K. Gupta**

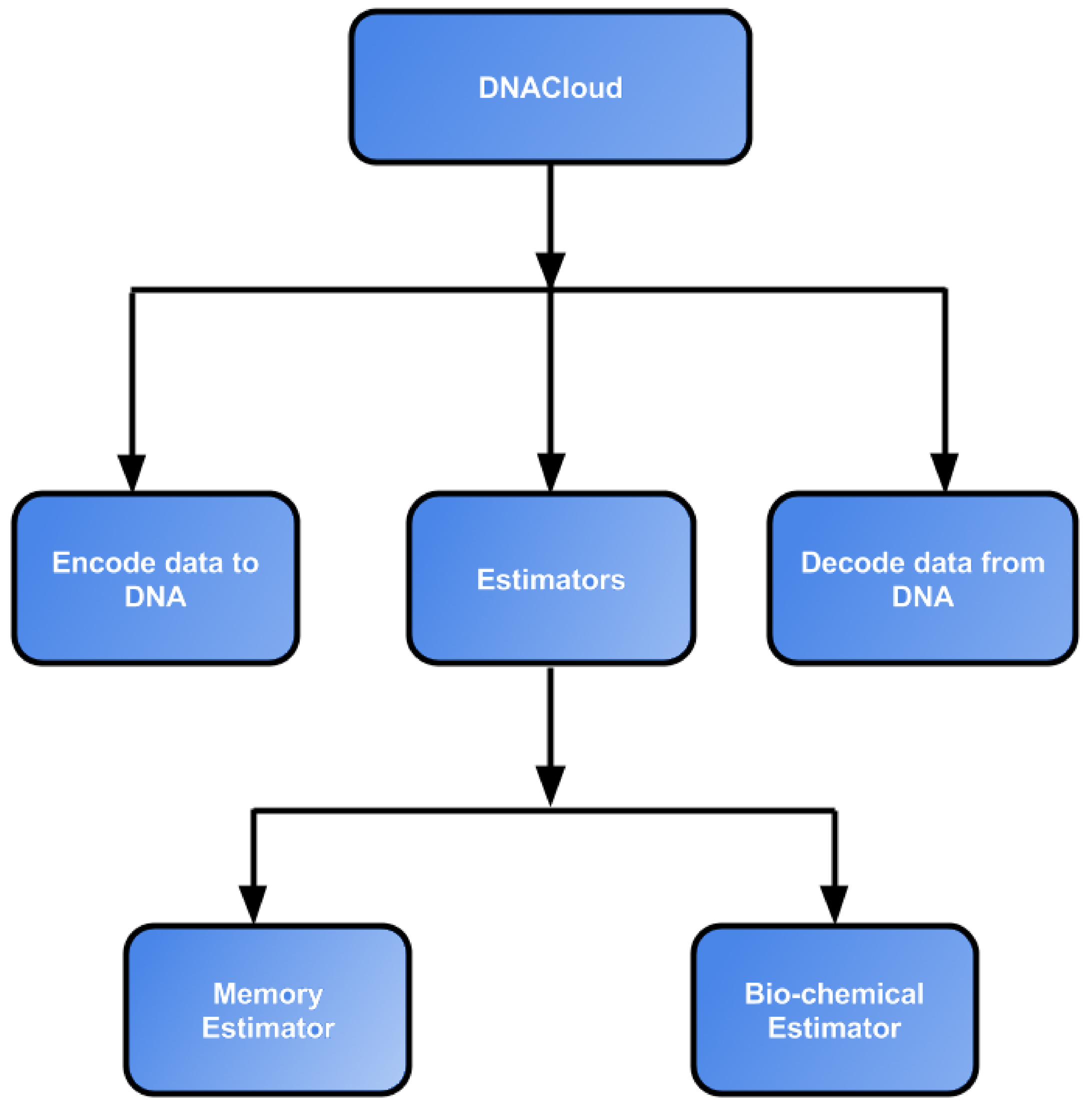
*Dhirubhai Ambani Institute of Information and Communication Technology, Gandhinagar, India.*

## Abstract

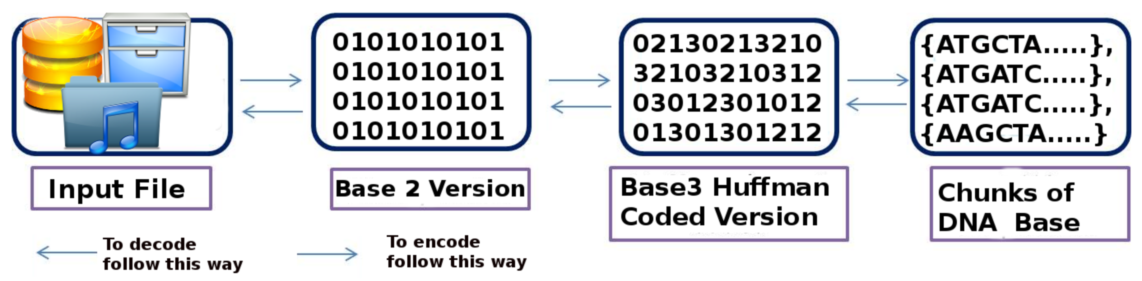
The term Big Data is usually used to describe huge amount of data that is generated by humans from digital media such as cameras, internet, phones, sensors etc. However before one can use the data, one has to address many issues for big data storage. Motivated by Goldman and his team [1], we have developed a software called DNACloud which makes it easy to store the data on the DNA.



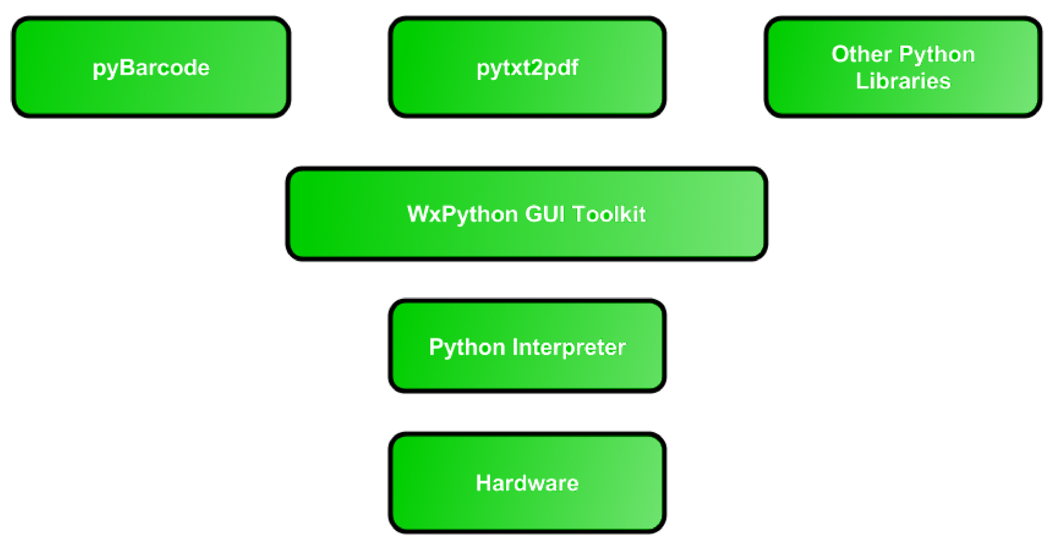

## Workflow




## Main Modules




## Software Architecture




## Software Output

DNACloud produces three different types of output files apart from .dnac file one of them is the figure . Other files which are generated by the software include output text file of Memory and pdf file generated by Bio-chemical Estimator.

**Output of LATEX file generated using DNACloud**

## Testing

We have performed some encoding and decoding operation on some sample data files, results for which are shown in the table below.

| File_Type | File size(Bytes) | DNA required |
|---|---|---|
| Text | 48680 | $1.06 \times 10^{-16}$gms |
| Audio | 85799 | $1.88 \times 10^{-16}$ gms |
| Video | 33745571 | $7.41 \times 10^{-14}$ gms |
| Image (HD) | 473206 | $1.04 \times 10^{-15}$ gms |

Comparison of the file formats encoded by DNACloud. Different file types were encoded and decoded using this tool.

## Advantages

DNA data storage is a stellar technology when it comes to data store, some of them are as follows:

1. Dense storage medium
2. No maintenance, electricity requirement
3. Long term data storage
4. Portable

## Future Challenges

DNA data storage technology has many forthcoming challenges, some of them are as follows:

1. Re-writable
2. Cost-effective
3. Cheap synthesizing and sequencing techniques
4. Length of DNA string
5. Data security
6. Efficient encoding and decoding algorithm

## Estimator

Our tool, DNACloud, consists of two storage estimators which are described as below:

1. Memory Estimator: This estimator provides approximate values for the amount of physical memory required, amount of secondary memory required, amount of DNA required etc.
2. Biochemical Properties Estimator: This module provides the minimum and maximum boiling point among-st all the oligonucleotides residing in the .dnac files.

## References

[1] Nick Goldman, Paul Bertone, Siyuan Chen, Christophe Dessimoz, Emily M LeProust, Botond Sipos, and Ewan Birney. Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature*, 2013.

**Web**: www.guptalab.org/dnacloud
**Email**: dnacloud@guptalab.org