University of Victoria
Computer Science

Presented By:
Harshit Jain
Kirti Agarwal
Roshni Jain

Analysis and Evaluation of Translation Services

Using *Map Reduce*

Outline

- Problem
- Motivation
- Related Work
- Proposed Solution
- Analysis and Evaluation
- Results and Conclusion
- Challenges and Future Work
- References

Problem

- Language barrier restricts people to get connected
- Various research approaches in the field of translation, uses the bulk load APIs and are not lightweight in terms of network usage
- We want to compare and learn the issues such as latency, geographical location, and accuracy of the different translation APIs while translating the large size of data (big data).

Motivation

Important for : Students, Elderly people, Private and Public sector

- Translation services assist the users of different language origin to interact with each other without being bounded by language constraints
- MapReduce, a lightweight programming model, which can be used for translating massive data [1]
- Amazon Elastic Compute Cloud (EC2) service, provided by Amazon Web Services
  - Provides access to geographically located server instances on demand as a service.

Related Work

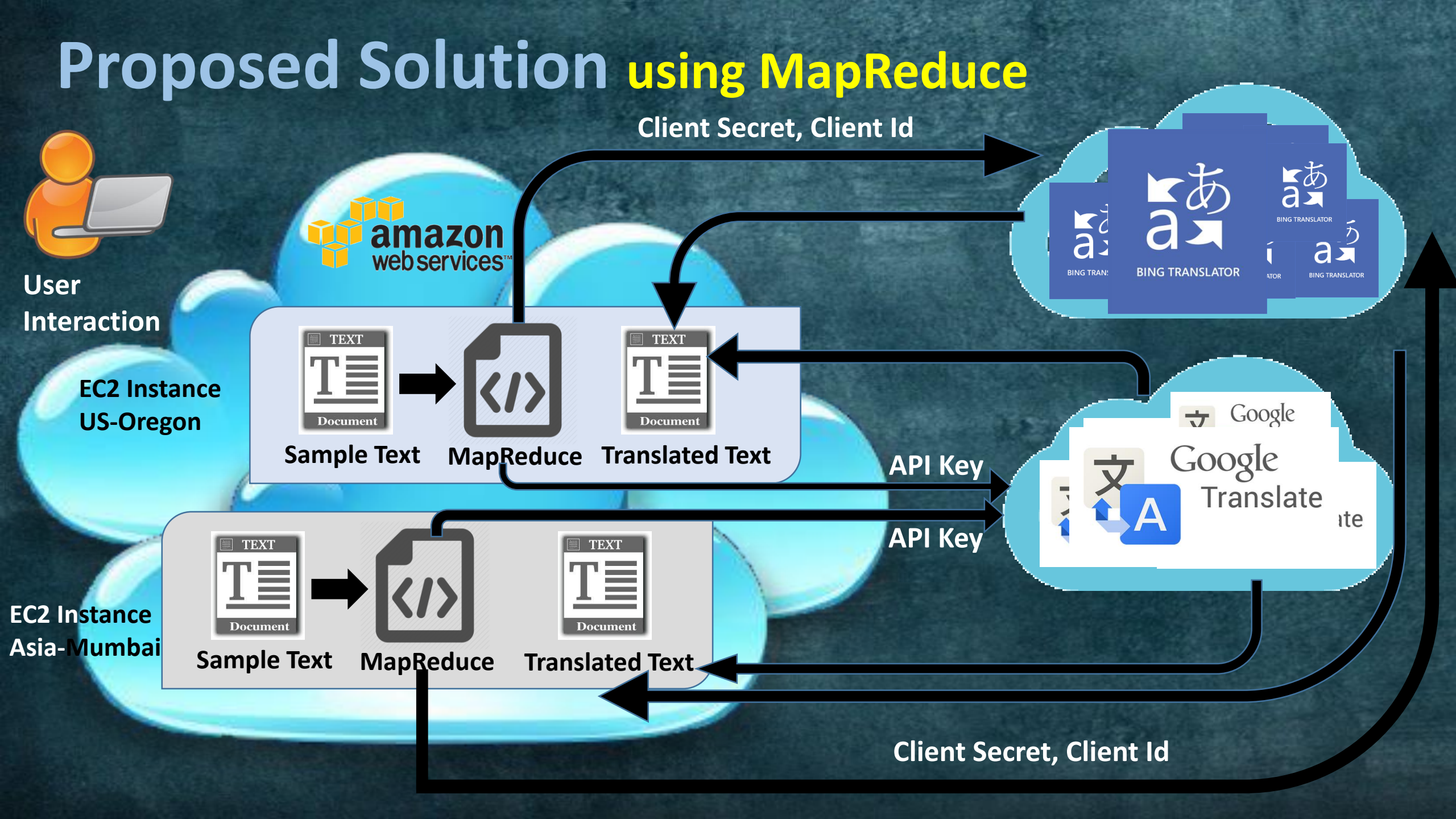
- Bhojraj Singh Dhakar et al., did a survey, by evaluating translation quality of online translation system in various fields like technology, medicine, news [2]

- Chunyu Kit et al., Evaluated a large-scale corpus of legal texts by means of BLEU/NIST scoring [3]

  What's out there: Skype translator, Google chrome page translator, Google Hangout and Whatsapp (soon)

Note: So far the researchers have mainly focused on the translation quality evaluation in the field of MT

Proposed Solution using MapReduce

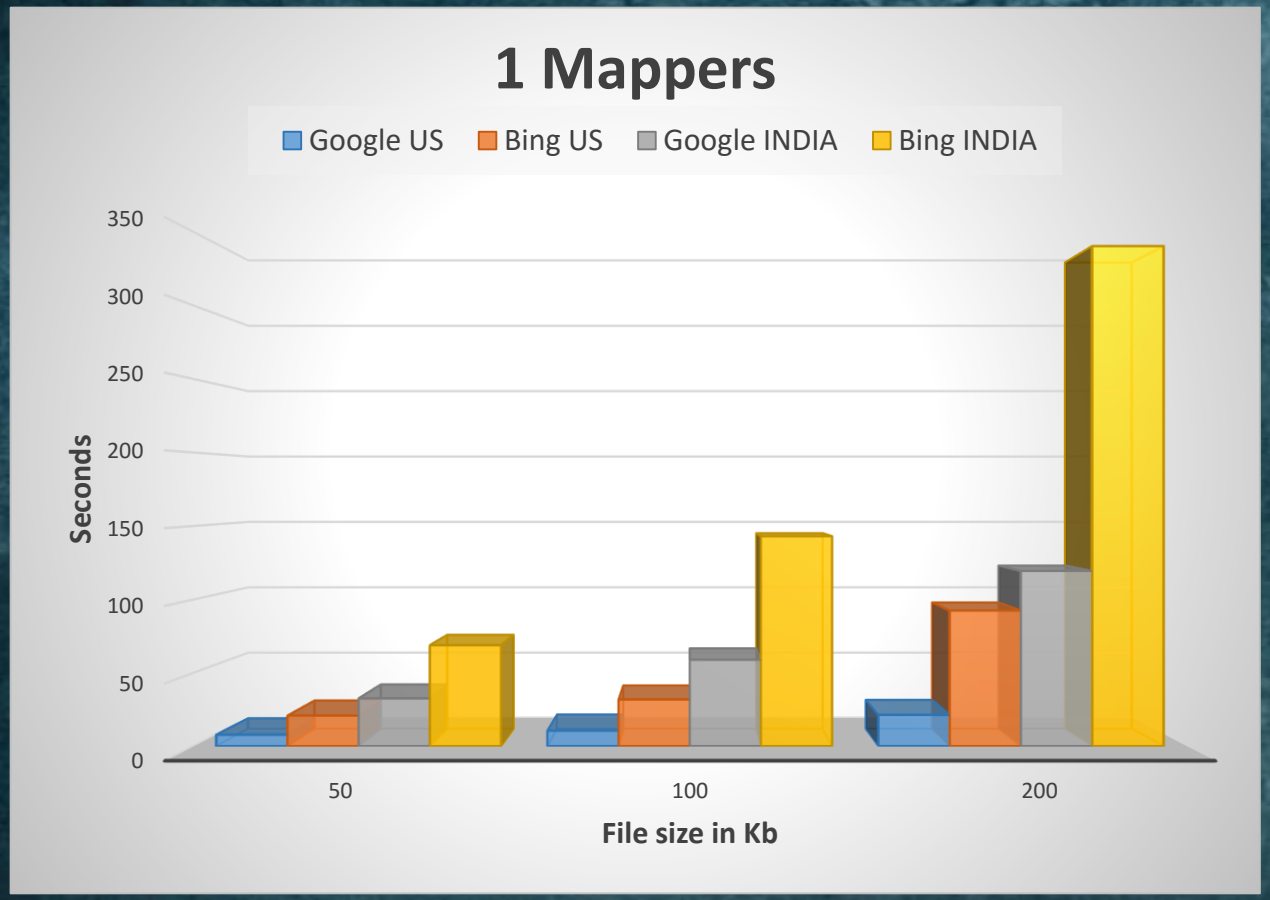


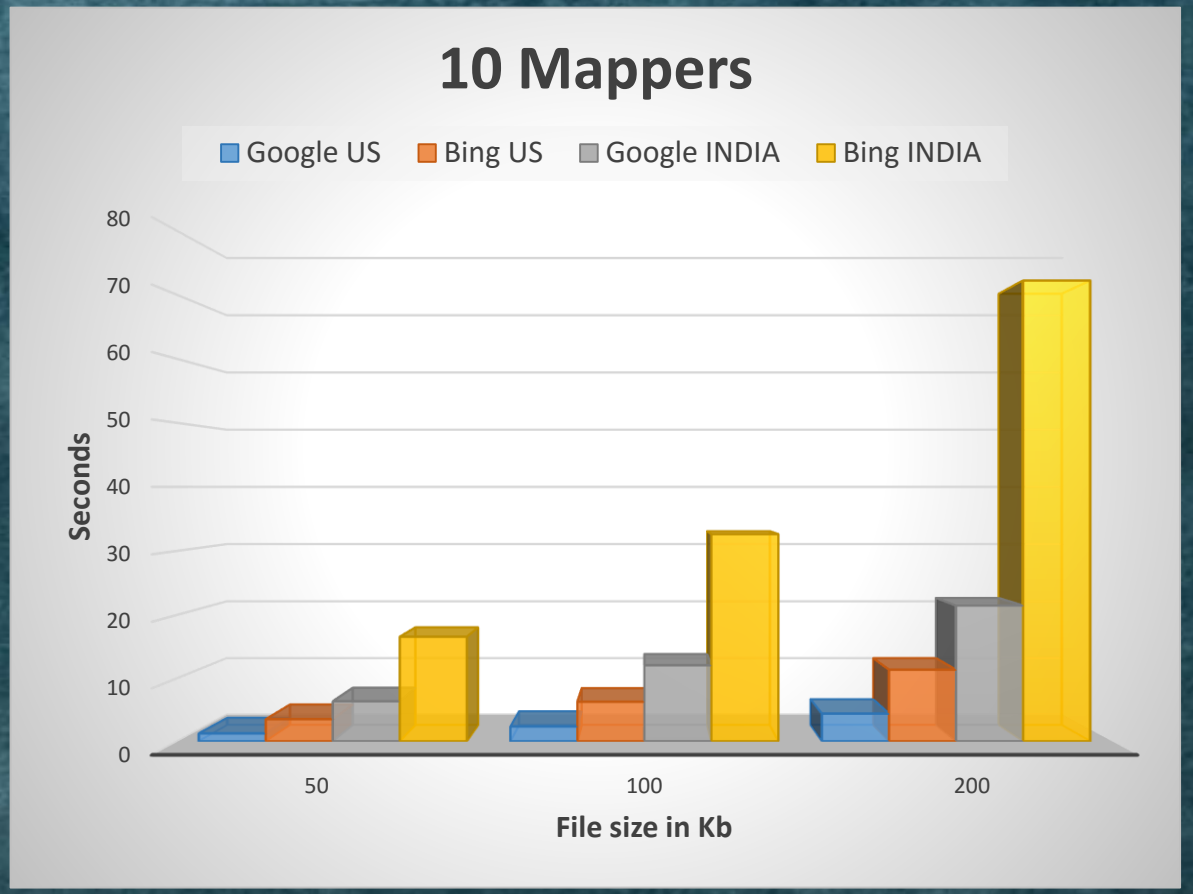
Translators Analysis and Evaluation

- Between two Translators
- Location Based

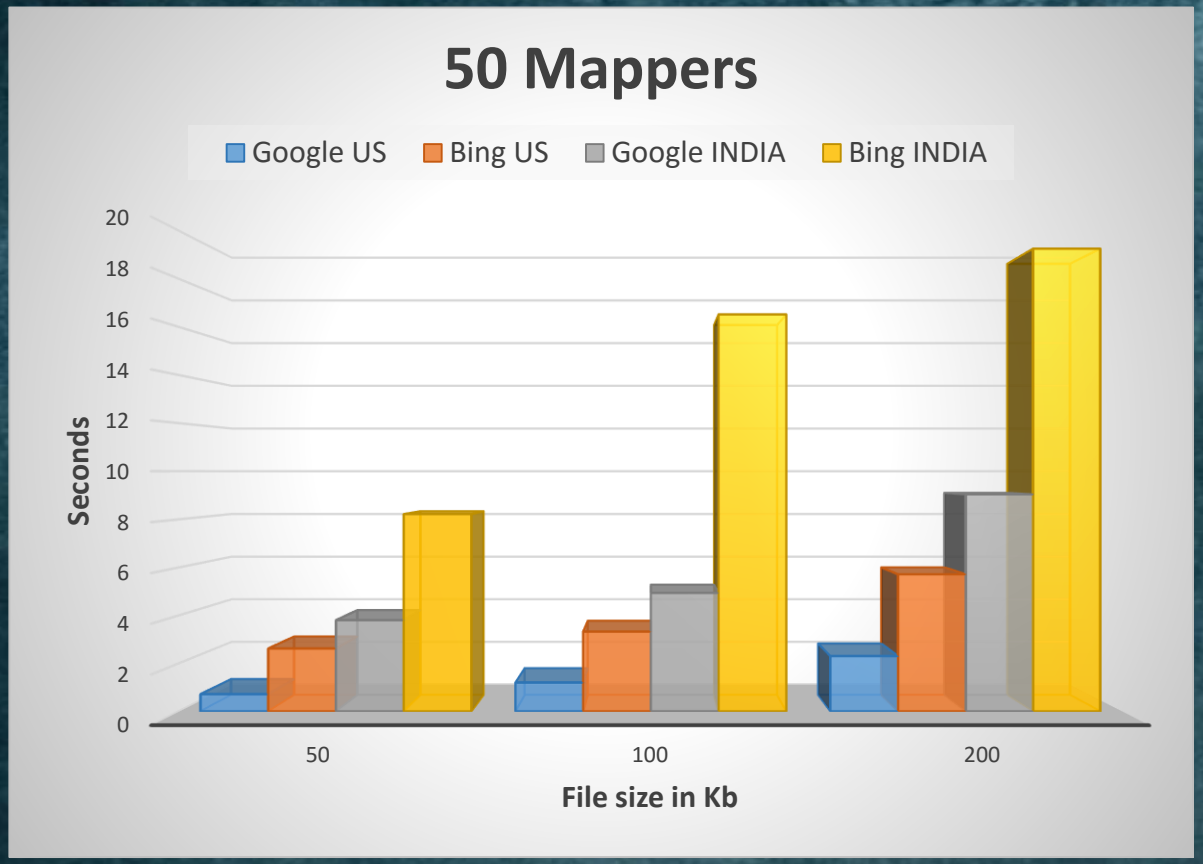
- Mappers Based
- Language Based

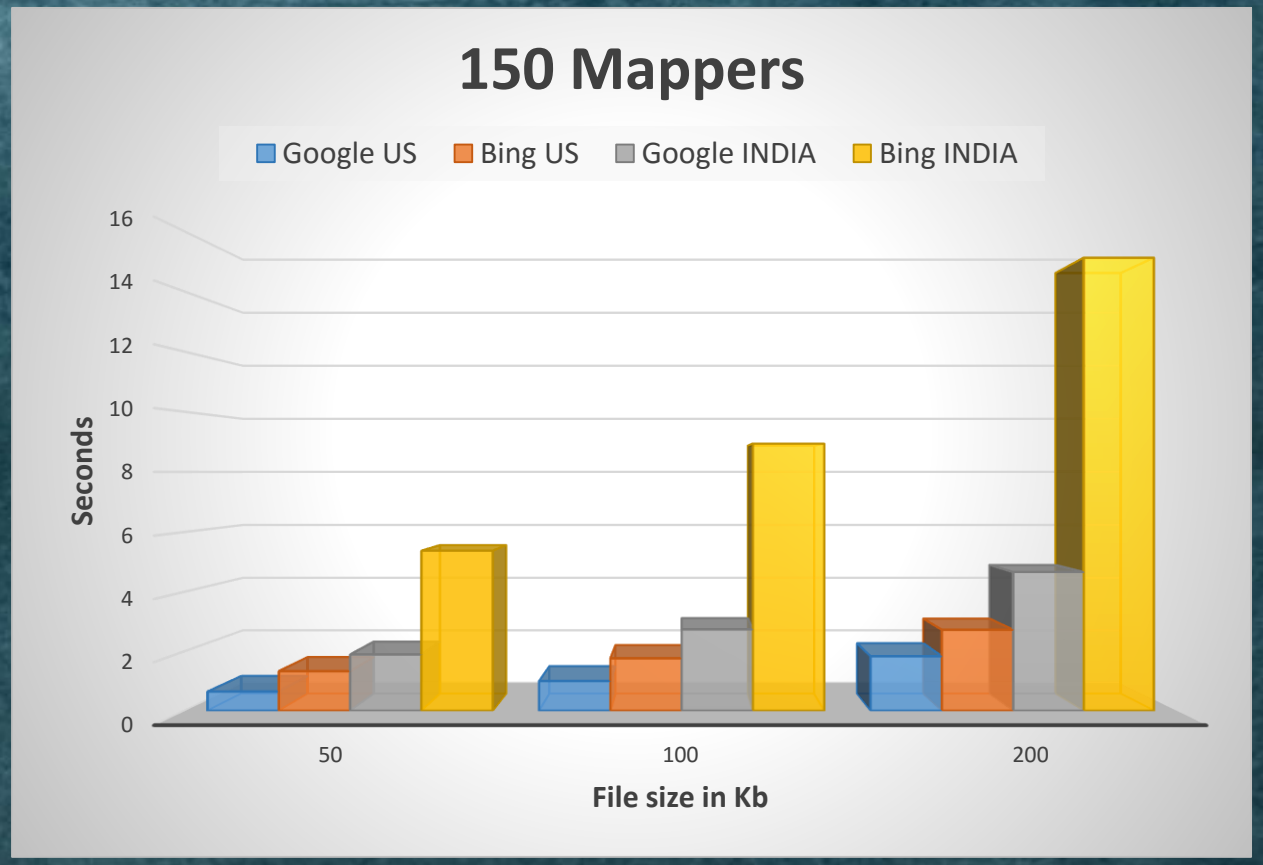
Translators and Location Based Evaluation(1)



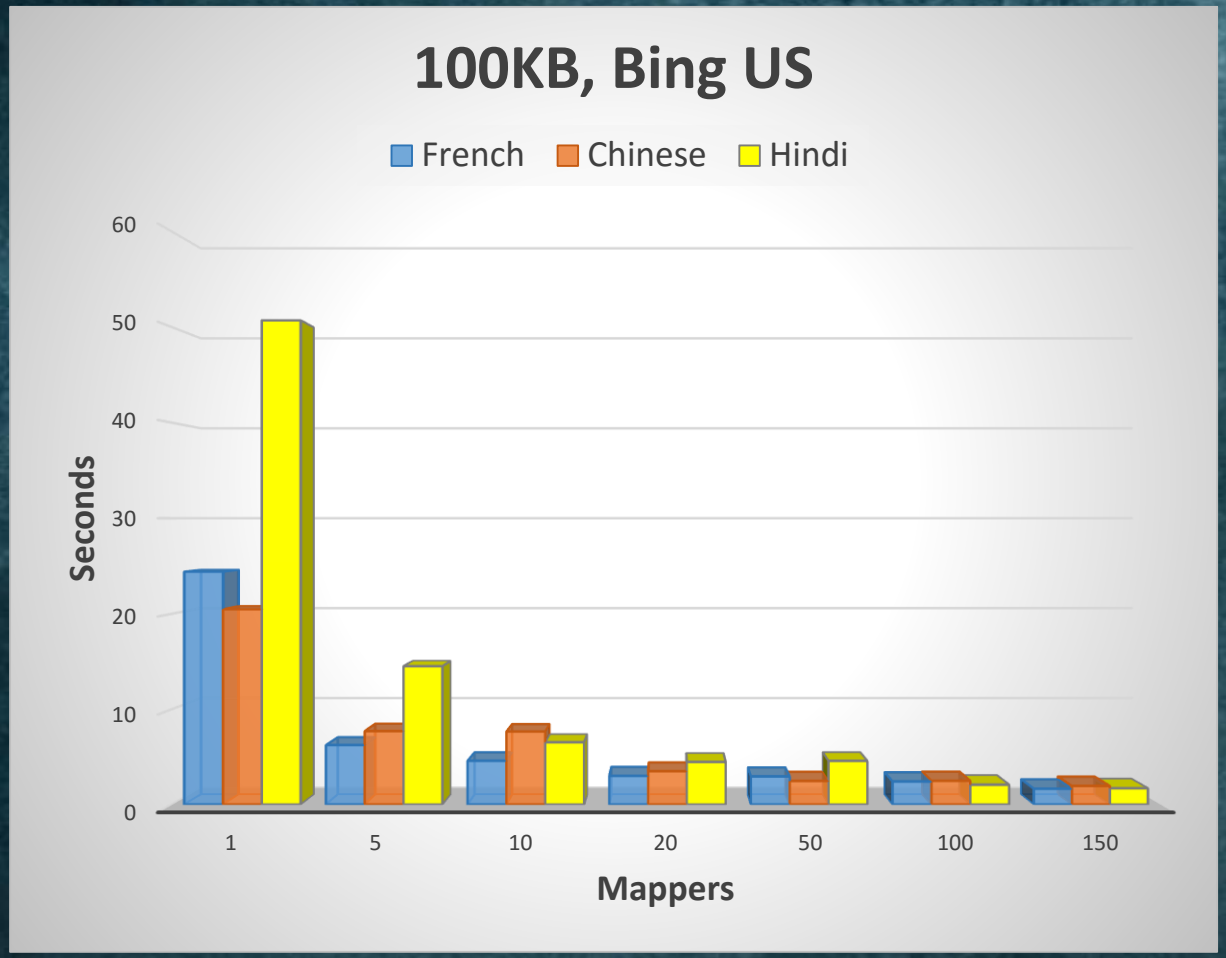


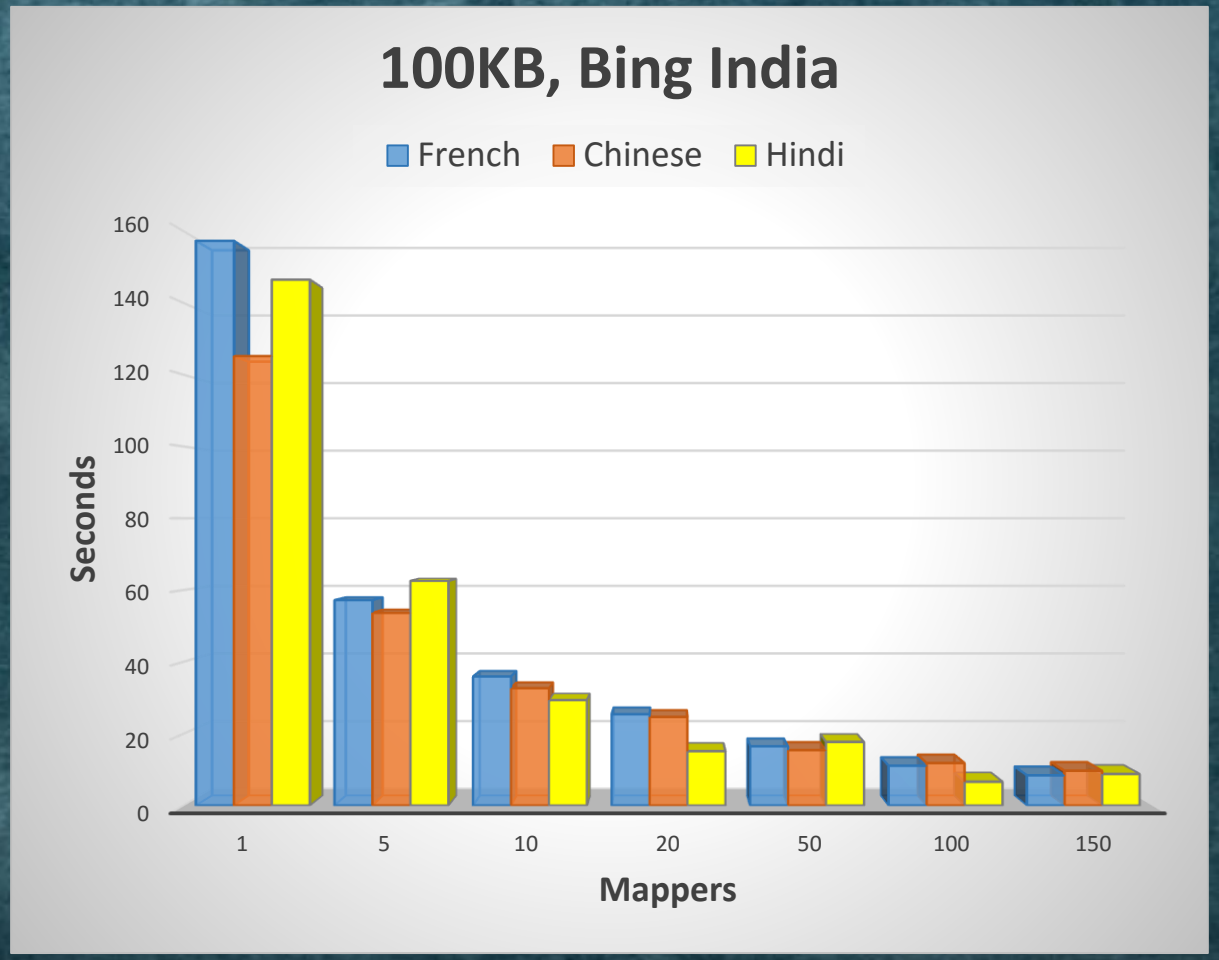
Translators and Location Based Evaluation(2)



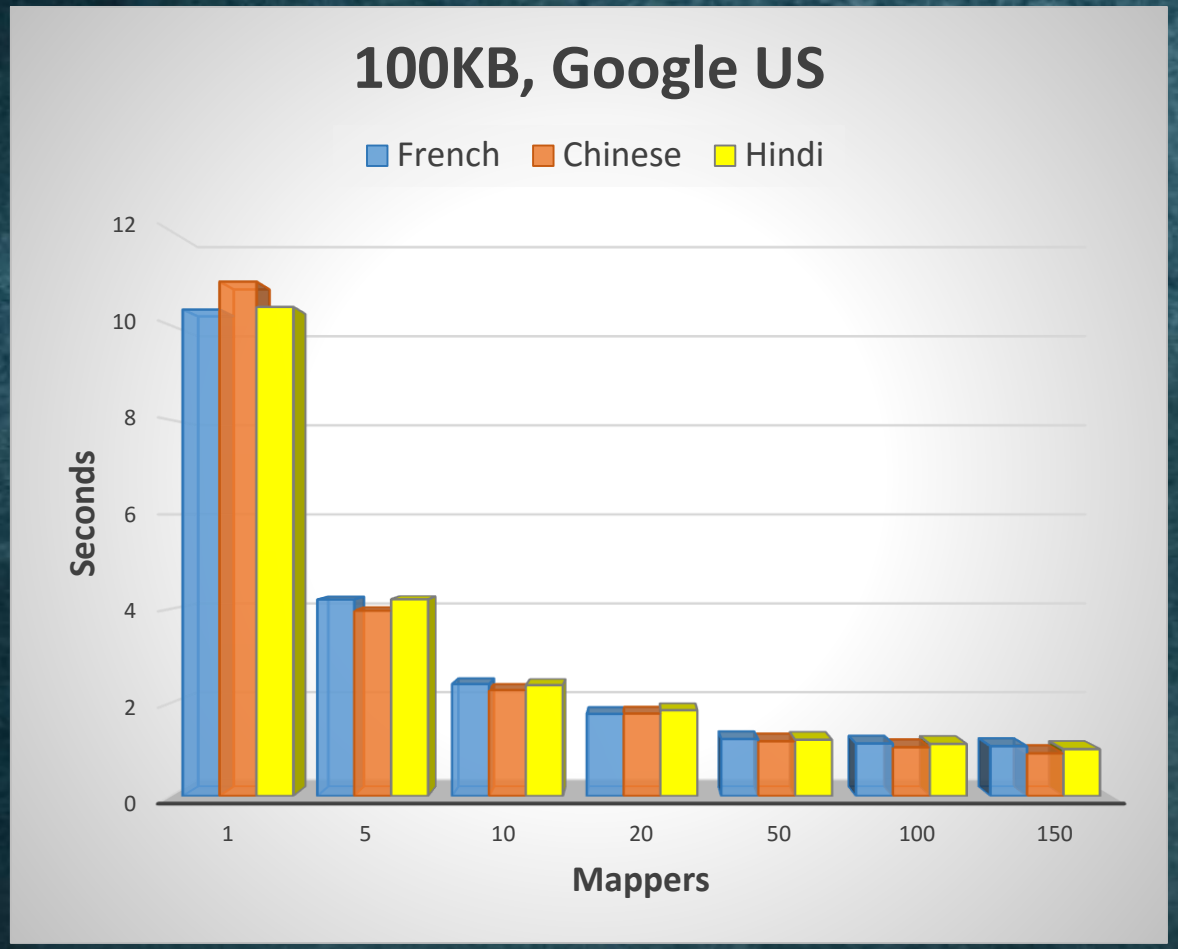


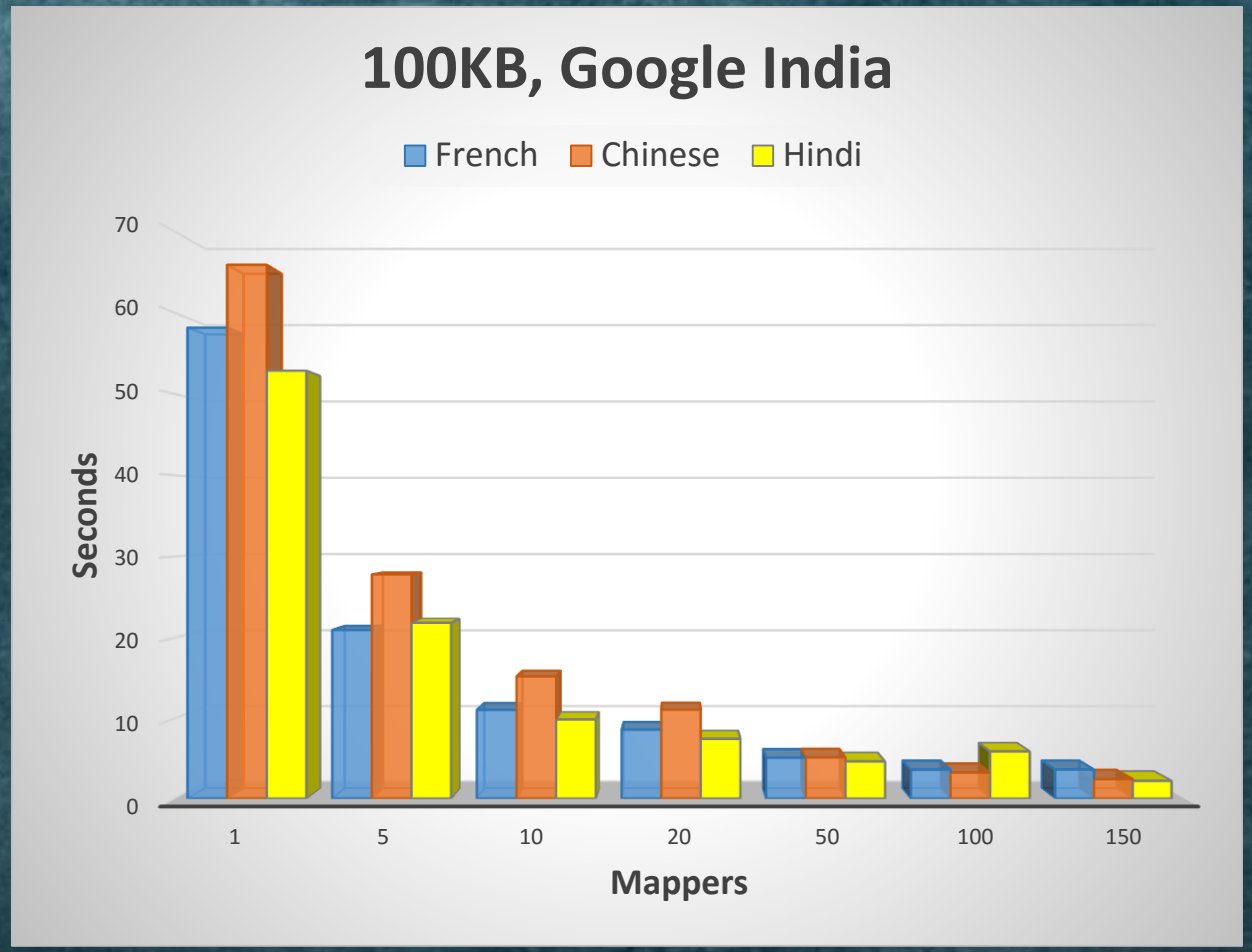
Mappers and Language Based Evaluation(1)





Mappers and Language Based Evaluation(2)





Results and Conclusion(1)

Translator Evaluation

Suppose we have 10 mappers, 100KB of text, we found that :

- **For EC2 instance US-Oregon, the latency of the Google translator is 62.4% less than the Bing translator**
- **Similarly, for EC2 instance Asia-Mumbai, the latency of the Google translator is 63% less than the Bing translator**

Location Based Evaluation

For 10 mappers and 100 KB of text, we found that :

- **The latency of Google translate API in US-Oregon is 80.5% less than in Asia-Mumbai**
- **The latency of Bing translate API in US-Oregon is 80.8% less than in Asia-Mumbai**

Results and Conclusion(2)

Mappers Evaluation

As we are increasing the number of mappers, the latency is almost decreasing for both the translator and for both the locations.

Language Based Evaluation

For all the three languages we evaluated, there is no prominent difference in latency for both the translators.

Results and Conclusion(3)

From the above results, after comparing two translation services with:

- Various file sizes
- Varying the number of Mappers
- Three different languages, and
- Two different locations (US-Oregon, Asia-Mumbai)

We deduce that the Google translate API outperforms the Bing Translate APIs

Challenges

Due to the limited free access to API keys, we have a restriction on input file size of up to 200 kb

Future Work

- Translated text quality evaluation
- A chatting application can be build over this to transfer translated files between different users
- Embedded text translation

References

1. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6612229
2. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.437.7864&rep=rep1&type=pdf
3. http://heinonline.org/HOL/LandingPage?handle=hein.journals/llj100&div=28&id=&page=
4. https://lintool.github.io/MapReduceAlgorithms/ed1.html

Here is the link of code and data evaluation:
https://github.com/jharshit/Translate-Service-Evaluation

Thank You!!!