

Drift Analysis based Complexity: REST & BULK API based Data Accessing System Model

¹Naeem Hassan ²Shahzad Latif ³Muhammad Usama ⁴Muhammad Armaghan
⁵Shazia Saqib ⁶Tahir Alyas ⁷Muhammad Adnan Khan

^{1,3,4,5,6,7} Computer Science Department, Lahore Garrison University, Lahore, 54000, Pakistan.

²Department of Computer Science, Shaheed Zulfikar Ali Bhutto Institute of Science and Technology (SZABIST), Islamabad campus.

Abstract

As with every passing day, digital data storage is increasing due to the cheap and safe facilities. The time to access data to perform operations on that data is going more complex. As the billions of users are communicating with the internet that's why data transfer rate and security factors are the main challenge of technologists. There are two types of techniques which are dealing with the data processing: batching big data and streaming processing. In this article, two data accessing methods named as Rest Application Program Interface and Bulk Application Program Interface are considered. We implement these two techniques in the real time scenario by taking an example of a Pvt. Limited organization. Which has its main server that is accessed by their clients. The organization implemented these two data accessing algorithms in different time. Simulation results shown that Bulk API based data accessing approach gives lesser time as compared to Rest API based approach.

1. Introduction

Now a day many organizations are using different software's to enhance their business. In today's world software is a key need of medium to large organizations [1]. Some organizations are using software's to only operate simple tasks like resource management but there are also many organizations which are using software's to automate their business domains [2]. It is understood if any organization use correct software according to their business domain, then surely organization business procedures execute efficiently and accurately [3] [4].

If an organization have small data set than execution time does not matter a lot in data processing. But if some organization have large data set and single function process an action which have to be applied on thousands of records then there is very high chance of operation/procedure crash, if function is processing data one by one. May be one reason for crash is time out error.

To overcome this problem, information technology industry introduced a term e.g; "bulk data processing". In this technique algorithm identifies same kind of tasks in data set and apply action on same kind of task in one call. On the other hand, if algorithm is not equipped with bulk data

processing capability then algorithm will process data one by one [5] [6].

For example, a function gets 10,000 records to process, and there are 5000 records for type 'A' processing and other 5000 records are for type 'B' processing. Then Bulk data processing algorithm identifies same tasks and make 2 calls for operation. But if algorithm has not capability of bulk data processing then it makes 10,000 calls to operate data.

1.1 Rest Application Programming Interface

Every web-based application has its Application Programming Interface (API) [9]. The concept of REST based architecture was introduced by Thomas Fielding, in year 2000 [10]. The REST API is working on the same principle as SOAP or other API's, it was used to increase the scalability and performance. It takes one Request from User and provides results.

1.2 Bulk Application Programming Interface

Bulk API is another method to implemented web based or server-based systems. Rather than REST API its performance is much higher.

2. Proposed System Model

In this article, two real life scenarios will be used to decide the best algorithm for data accessing from bulk API & Rest API. Representational State transfer [7] when data grow in thousands.

WHITEHATS organization made a system years ago to process user's data and send this data to "Salesforce".

Salesforce is another application which save user's data and process it for marketing purpose.

Eventually when WHITEHATS organization grow their users increase from hundreds to millions. When user grows in millions at this time system does not able to process even thousands of users in single request and system became down.

This is due to use of rest API calling for each user data e.g. if system get request to process data of 10 thousand users

then system will call Salesforce 10 thousand times using its rest API to update data on Salesforce.

And according to standards max execution time for each request will be at max 120 seconds but unfortunately due to bad use of rest API it is consuming more than 5 minutes to execute, and most of the times it fails to execute.

After this scenario WHITEHATS organization realize the need of bulk API to overcome this issue. As world is moving towards big data processing which will be highly used in machine learning and artificial intelligence, WHITEHATS organization also keep this thing in mind and developed a system using bulk data processing to overcome existing problem of user data migration to Salesforce. In this article, one of the components of system will be used as example to elaborate system model.

2.1 Proposed System model with REST API

In this section Proposed REST API is used for data accessing on given below system model as shown in figure 1.

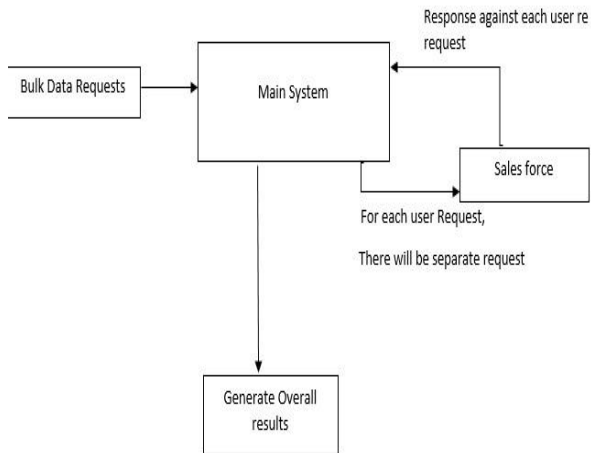


Fig. 1 Proposed REST API based System model

In the figure 1, there is a main system API which will first take requests from user/ users in the Bulk form. Then, API will perform search operation on the Sales force data force and finally response is generated through the main system API again. In the REST API approach, the main system will perform separate cycle against each request.

2.2 Proposed System model with BULK API

Figure 2 illustrate the Bulk API based system model used to process user's data requests.

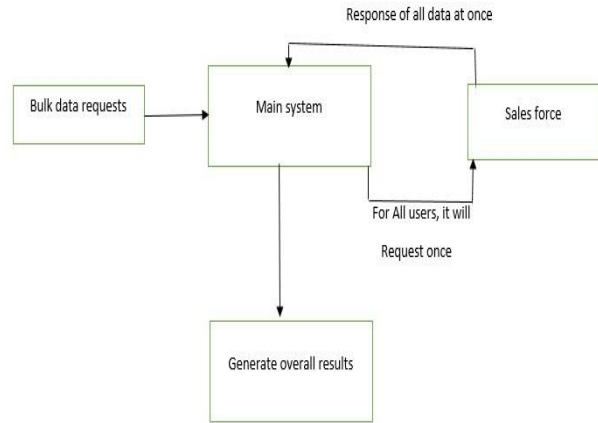


Fig. 2 Proposed BULK API based System model

In the Bulk API, the main system will perform search operation once for all requests.

3. Complexity Analysis

In this section find out the complexity analysis of both proposed solutions. Section 3.1 contain the complexity analysis of REST API based solution & section 3.2 contains the detailed complexity analysis of BULK API based solution. Complexity analysis play a very import rule in the selection of candidate solution. Complexity analysis is performed in various domains like MUD, Wireless communication etc. [11] [12] [13].

3.1 Time Complexity for REST API based Solution

In this section Complexity is calculated using drift analysis Salesforce function will be called for each iteration against user data as shown given below:

```

for: i-> 'n' number of users
{
    SalesforceRestAPI (each user);
}
  
```

In the above function it is understood that the time complexity of Salesforce function is “n” because it takes same time for per user as it is independent of type of data to process. So, sales force function is inside the loop it will be called “n” times because loop is iterating over n users [8].

Time complexity for “outer for loop”

$$T(n) = \left(\sum_{i=1}^n i \right)$$

$$T(n) = \frac{n(n+1)}{2}$$

Time complexity for inner salesforce function = n

Time complexity for complete algorithm:

$$T(n) = \frac{n(n+1)}{2} * n$$

$$T(n) = \frac{n^3 + n^2}{2}$$

$$T(n) = O(n^3) \text{ ----- (i)}$$

Total time complexity of proposed system model with REST API is $O(n^3)$.

3.2 Time Complexity for Bulk API based Solution

In Bulk API based solution Salesforce function will be called only one time for all users. This solution use array to hold user data for Salesforce processing. In each iteration array will be filled and this array will be used as input to Salesforce function after completion of loop.

Declare: array of users to process in bulk = bulk_users
for: i-> 'n' number of users
{
 Push "each user" in "bulk_users" array;
}
SalesforceBulkAPI(bulk_users);

In the above algorithm it is understood that the time complexity of Salesforce function is "n" because it takes same time for per user as it is independent of type of data to process. So, sales force function is outside the loop it will be called only one time because loop is iterating over n users and loop is filling user's data in array.

Time complexity for outer for loop

$$T(n) = (\sum_{i=1}^n i)$$

Time complexity for array = 1

Time complexity for complete for loop

$$T(n) = \frac{n(n+1)}{2}$$

Time complexity for salesforce function = n

Time complexity for complete algorithm:

$$T(n) = \frac{n(n+1)}{2} + n$$

$$T(n) = \frac{n^2 + 3n}{2}$$

$$T(n) = O(n^2) \text{ -----(ii)}$$

Time complexity of Bulk API based solution is $O(n^2)$.

4. Simulation and Results

In the simulation, the Time function for both REST and Bulk API's are checked on different values of data requests like when the data requests are 1000, 10,000 and 50,0000.

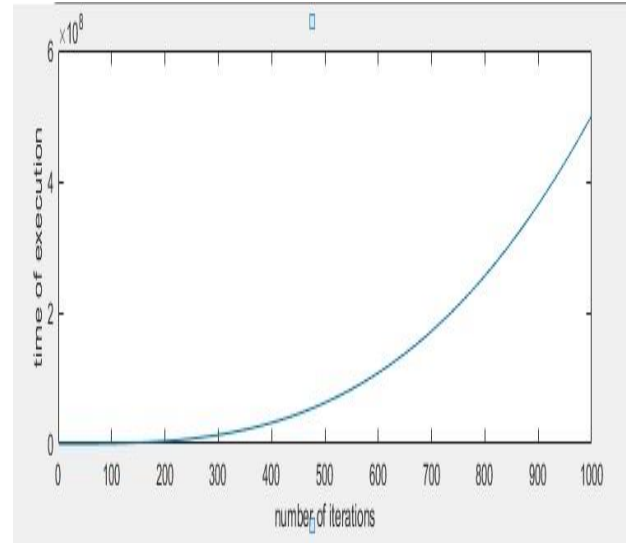


Fig. 3 Time Complexity of Proposed REST API based system model with 1000 request.

Figure 3 shown the time complexity of the proposed REST API based system model with 1000 request in terms of number iterations.

It clearly observed that if the number of iterations is increased then proposed REST API based solution complexity is also increased.

Figure 4 shown the time complexity of the proposed Bulk API based system model with 1000 request in terms of number iterations respectively. It clearly observed that proposed REST API based solution time complexity is directly proportional to the number of iterations.

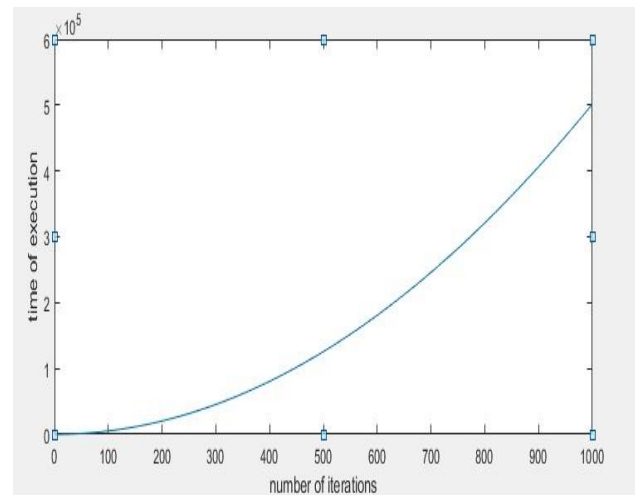


Fig. 4 Time Complexity of Proposed BULK API based system model with 1000 request.

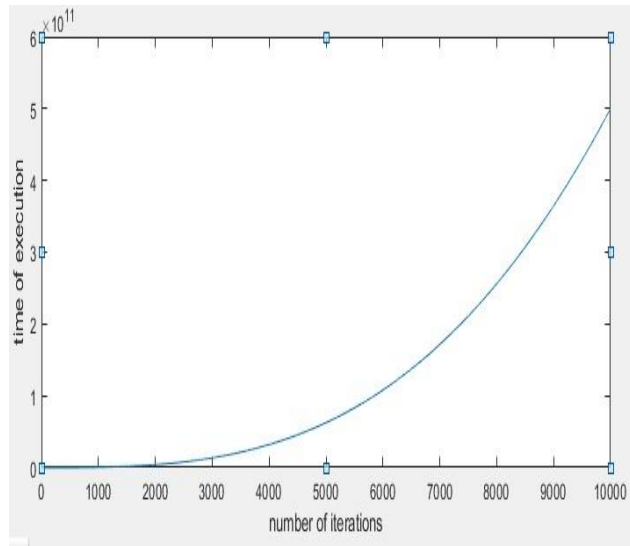


Figure 5: Time Complexity of Proposed REST API based system model with 10000 request.

Figure 5 shown the time complexity of the proposed RSET API based system model with 10000 requests in terms of number iterations respectively. It clearly observed that proposed REST API based solution time complexity is directly proportional to the number of iterations.

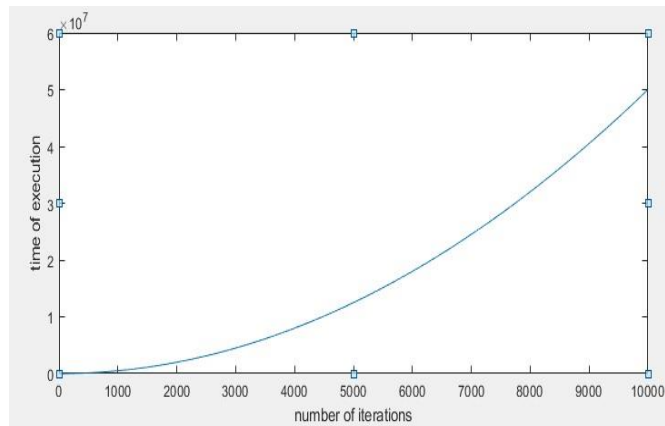


Figure 6: Time Complexity of Proposed BULK API based system model with 10000 request.

Figure 6 shown the time complexity of the proposed Bulk API based system model with 10000 requests in terms of number iterations respectively. It clearly observed that proposed REST API based solution time complexity is 5×10^7 at 10000 iteration.

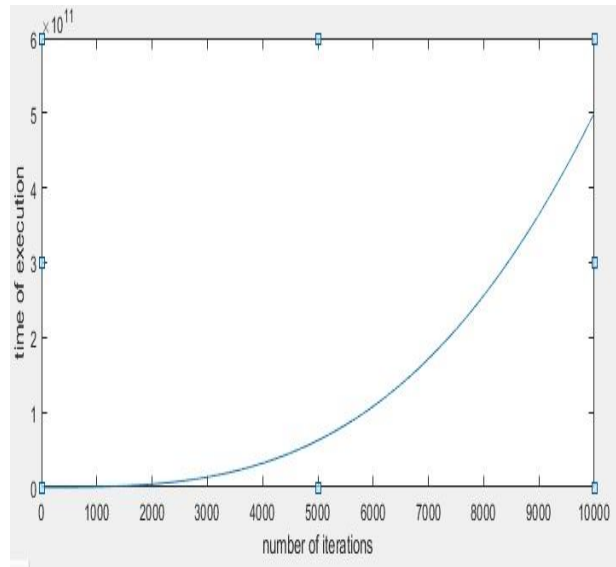


Figure 7: Time Complexity of Proposed REST API based system model with 50000 request.

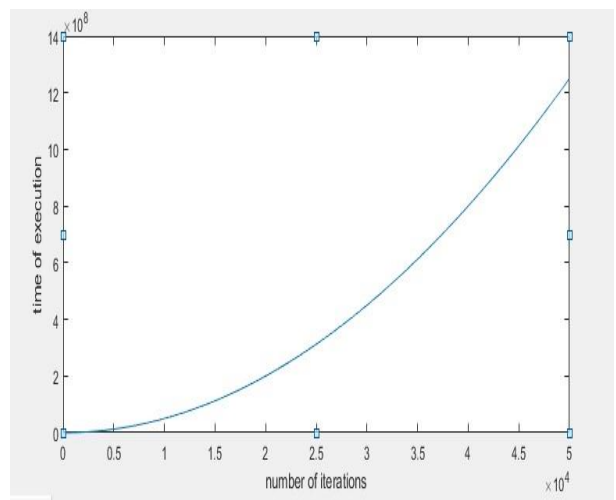


Figure 8: Time Complexity of Proposed BULK API based system model with 50000 request.

Figure 7 shown the time complexity of the proposed RSET API based system model with 50000 requests in terms of number iterations respectively. It clearly observed that proposed REST API based solution time complexity is 5×10^{11} at 10000 iteration.

Figure 8 shown the time complexity of the proposed BULK API based system model with 50000 requests in terms of number iterations respectively. It clearly observed that proposed REST API based solution time complexity is 12.2×10^8 at 10000 iteration.

5. Conclusion

In this article, our main focus was on the processing of requests through salesforce database. Drift analysis is used for time complexities of both REST and BULK API based solution. Simulation results shown that when user requests in API is in 10's, the time factor might be ignoring. But when the data requests are in increased like 1000, 10,000 or 50,000 then both solution performance effects in terms of time. There is major difference in the time complexities of both API's. i.e. when the data requests are 1000 the time complexity graph for REST API shows that it is of order of 108, while for same data requests the time complexity is of order 105, which is much lower.

Same is the case when data requests are 10,000 and 50,000. hence, it is clear that using bulk data processing approach is better than rest API approach when number of processing's are very large.

Proposed BULK API based solution gives less complexity as compared to REST API based Solution.

References

- [1] Madhavi Vaidya, Shrinivas Desphande, Vilas Thakare, "Design and analysis of Large Data processing techniques". International Journal of Computer Application's (0975-8887), Volume 100-No.8, August 2014.
- [2] Tongya Zheng, Gang Chen, Xinyu Wang, Chun Chen, Xingen Wang, Sihui Luo, "Real time intelligent big data processing: technology, platform, and applications, X. et al. Sci. China Inf. Sci. (2019) 62: 82101.
- [3] Salvador Garcia, Sergio Ramirez, Julian Luengo, Jose Manuel Benitez & Francisco Herrera, "Big data processing: methods and prospects", Big Data Anal 1, 9 (2016) doi:10.1186/s41044-016-0014-0
- [4] Tarkowska A, Carvalho-Silva D, Cook CE, Turner E, Finn RD, Yates AD (2018) eleven quick tips to build a usable rest api for life sciences. Plos Comput Biol 14(12): E1006542.
- [5] Todd W. Sprenkle, Srinivasa D. Murthy, Anil Khatri. Direct Bulk data transfers. United States Patent number US5805920A. retrieved from: <https://patents.google.com/patent/US5805920A/en>
- [6] Lior Aronovich, Kien K.Huynh. Concurrent Bulk processing of tree-based data structures. United States Patent number US10223409B2. Retrived from: <https://patents.google.com/patent/US10223409B2/en>
- [7] Shyue Ping Ong, Shreyas Cholia, Anubhav Jain, Mirin Brafman, Dan Gunter, Gerbrand Ceder, Kristin A. Persson. The material Application Programming Interface (API): A simple, flexible and efficient API for materials data based on Representational State Transfer (REST) principles.
- [8] Gashkov, S.B. The arithmetic computational complexity of linear transforms. Moscow Univ. Math. Bull. 69, 251-257 (2014) doi: 10.3103/S0027132214060047
- [9] Neumann, Andy & Laranjeiro, Nuno & Bernardino, Jorge. (2018). An Analysis of Public REST Web Service APIs. IEEE Transactions on Services Computing. PP. 1-1. 10.1109/TSC.2018.2847344.

- [10] [10] R. T. Fielding, "Architectural Styles and the Design of Network-based Software Architectures," Univ. of California, Irvine, 2000.
- [11] Iqbal, K., Khan, M. A., Abbas, S., & Hasan, Z. (2019). Time complexity analysis of GA-based variants uplink MC-CDMA system. SN Applied Sciences, 1(9), 953.
- [12] Asif, M., Khan, M. A., Abbas, S., & Saleem, M. (2019, January). Analysis of Space & Time Complexity with PSO Based Synchronous MC-CDMA System. In 2019 2nd international conference on computing, mathematics and engineering technologies (iCoMET) (pp. 1-5). IEEE.
- [13] Tariq, Z. B., Khan, M. A., Abbas, S., & Fatima, A. (2019, January). Complexity analysis of DE based CE-UD MIMO system. In 2019 2nd international conference on computing, mathematics and engineering technologies (iCoMET) (pp. 1-5). IEEE.



Cloud Computing

Mr. Naeem Hassan is doing M.Phil. from the Department of Computer Science, Lahore Garrison University, Lahore, Pakistan. He completed his Bachelor of Computer Sciences from the Department of Computer Science, Government College University, Lahore, Pakistan. Naeem's research interests primarily include Computer Architecture, Algorithms and



Computational intelligence and Fuzzy Systems

Shahzad Latif received the M.Sc and M.Phil degrees in Electronics from Quaid-i-Azam, University-Islamabad Pakistan in 2005 and 2008, respectively, and Ph.D. degree in electronic engineering from School of engineering and Applied Sciences (SEAS), ISRA University, Islamabad, Pakistan in 2018. His research interests include Cognitive radio networks,



technologies PVT. LTD. Usama's research interests primarily include Big data processing, parallel processing, Artificial Intelligence, Cloud Computing and Emotional Intelligence.

Mr. Muhammad Usama is doing M.Phil. from the Department of Computer Science, Lahore Garrison University, Lahore, Pakistan. He completed his Bachelor of Computer Sciences from the Department of Computer Science, The University of Lahore, Lahore with distinction and certificate of excellence. Currently, he is working as a software developer, Devnym



research interests primarily include Big data processing, parallel processing, Artificial Intelligence, Cloud Computing and Emotional Intelligence.

Mr. Muhammad Armaghan is doing M.Phil. from the Department of Computer Science, Lahore Garrison University, Lahore, Pakistan. He completed his Bachelor of Computer Sciences from the Department of Computer Science, Lahore Garrison University, Lahore, Pakistan. Currently, he is working as a Sr software developer at alpha square. Armaghan's

International Conferences. Khan's research interests primarily include MUD, Image Processing & Medical Diagnosis, Channel Estimation in Multi-Carrier Communication Systems Using Soft computing with various publications in journals and conferences of international repute.



University, Islamabad, Pakistan. Shazia's research interests primarily include Image Processing, Deep Learning, Machine Learning etc.

Dr. Shazia Saqib is currently working as Dean at the Faculty of Computer Science, Lahore Garrison University, Lahore, Pakistan. She completed her Ph.D. & M.Phil. from the Department of Computer Science, Government College University, Lahore, Pakistan. She completed her Master of Computer Sciences from the Department of Computer Science, Quaid-i-Azam



Department of Computer Science NCBA&E, Lahore, Pakistan. Tahir's research interests primarily include Cloud Computing, IoT & Intelligent Age

Dr. Tahir Alyas is currently working as an Assistant Professor at the Department of Computer Science, Lahore Garrison University, Lahore, Pakistan. He completed his Ph.D. from the School of Computer Science, NCBA&E, Lahore, Pakistan. He completed his Master of Computer Sciences from the Department of from the School of Computer Science, NCBA&E, Lahore, Pakistan. He completed his Master of Computer Sciences from the



BS degrees from the International Islamic University, Islamabad, Pakistan by obtaining scholarship award from the Punjab Information & Technology Board, Govt of Punjab, Pakistan. Prior to joining the Lahore Garrison University, Khan has worked in various academic and industrial roles in Pakistan. He has been teaching graduate and undergraduate students in computer science and engineering for the past 11 years. Presently, he is guiding 04 Ph.D. scholars and 04 M.Phil. Scholars. He has published about 130 research articles in International Journals as well as reputed

Dr. Muhammad Adnan Khan is currently working as an Assistant Professor at the Department of Computer Science, Lahore Garrison University, Lahore, Pakistan. He completed his Ph.D. from ISRA University, Pakistan by obtaining scholarship award from the Higher Education Commission, Education Commission, Islamabad, Pakistan. He also completed his M.Phil. &