

Performance Comparison between Hive, Spark-Sql & Flink-Sql through IVR Data Analysis

Ramanjot Kaur*, Dr. Raman Chadha

*Research Scholar, M.Tech(CSE), CGC Technical Campus, Jhanjeri.

Professor, HOD(CSE), CGC Technical Campus, Jhanjeri,

Abstract: Companies that utilize automated IVR systems have a veritable treasure trove of data that can be analyzed to improve the quality of the customer experience. After all, many customers who are greeted by linear thinking IVR systems instead of human voices already assume that their self-service experience is going to be less than favorable. Analyse the call Centre Performance includes various parameters like Cross-Team Visibility, Monitor Interactions in Real Time, Simplify reporting, Evaluate and streamline journeys etc. This paper focus on an approach in which IVR data is analysed and comparison is done based on HIVE, SPARK and FLINK frameworks.

I. Introduction

Ad-hoc querying is very important for all the business domains (retail, telecom, healthcare, finance, etc.) for generating valuable insights. There are tons of SQL engines available in the industry out of which we have chosen three diverse engines whose underlying layers are different:

- Apache Hive – It runs on the top of Apache Hadoop’s MapReduce, which is a batch engine.
- Spark-SQL – It runs on the top of Apache Spark, which provides diverse capabilities like stream, batch, interactive, graph processing, etc. But at the core Spark is a batch engine.
- Flink-SQL – It runs on the top of Apache Flink, which is a unified platform, can handle real-time, stream, batch, interactive, native-iterative, graph processing, etc. At the core Flink is a true streaming engine.

IVR & Call Center Services

First Data call center services and flexible IVR system help your customers access their accounts so you can maintain a high-level of customer satisfaction.

Benefits for Financial Institutions

Whether you want to outsource your entire call center functions or implement a self-service Interactive Voice Response (IVR) solution, First Data has the IVR systems and call centers you need to be successful. We handle an average of 30 million calls per month with customer satisfaction as the top priority.

- Highly trained call center agents take your customer’s call
- Save money with a fully hosted phone banking solution
- Provide customers 24x7x365 access to account information
- Resolve simple issues with IVR

Key Components

- Available ports precisely calculated to ensure little to no hold time for customers
- First Data managed and owned phone lines with resource upgrades as necessary result in less expenses for you
- Highly scalable call centre technology that allows you to start small and expand on demand
- Full system integration delivers up-to-date, real-time account information
- Voice and touch-tone responses available on IVR, while agents handle the more complex and sensitive issues

IVR Outsourcing Services

Interactive Voice Response (IVR) is a system which allows customers to interact with a computer through speech recognition, or alternatively, through the telephone keypad. IVR outsourcing is very useful in situations when a contact center executive is not necessary, yet customers still require some degree of interaction. IVR services from Invensis Technologies, a leading Call Center Outsourcing company, are a cost-effective option for gaining an advantage over competitors and building customer satisfaction.

IVR can assist you to:

- Be available to your customers whenever required and from any location
- Provide information about your business' offerings
- Allow customers to enter passwords and access confidential information
- Direct customer's calls to the specific company department or office
- Distribute automated outbound calls for surveys, up-selling, post sales support etc.
- Customize menus, on-hold functions and transfer options
- Conduct voice blasting services
- Create automated account payment facility
- Make customer satisfaction (C-Sat) calls
- Take orders
- Conduct market surveys

II. Simulation Setup And Query Generation

IVR Data Analysis - KPIs & Solution

IVR data analyzation is done based on HIVE, Spark and Flink framework comparison by taking various examples in different scenarios and deeply analysing each of them by using different set of queries for each framework. We have proposed six different queries depending upon the nature and type of the IVR data of the call centers and all these queries are discussed and explained in the below section .

A. IVR Data Analysis using HIV

1. Create ivr_data table in Hive

```
create EXTERNAL table ivr_data
```

```
(  
mob_no string,  
date_time string,  
customer_category string,  
menu_path string,  
call_transferred string,  
talk_timeint,  
transection String  
)
```

```
ROW FORMAT DELIMITED
```

```
FIELDS TERMINATED BY ',';
```

2. Load data in ivr_data table

```
load data local inpath '/home/dataflair/Data/ivrData003' into table ivr_data ;
```

3. Hive Query

1) Number of calls where total handling time is less than 30 sec but still it was transferred to customer care

```
SELECT count(talk_time)  
FROM ivr_data  
WHERE talk_time<30 and call_transferred LIKE 'RAT';
```

2) Avg talk time of calls transferred to customer care

```
SELECT AVG(talk_time)  
FROM ivr_data  
WHERE call_transferred LIKE 'RAT';
```

3) Avgtalk time of calls that were not transferred to customer care

```
SELECT AVG(talk_time)  
FROM ivr_data  
WHERE call_transferred LIKE 'CD';
```

4) Avg Talk time of "D" Customer category where menu path is "PRE_HOST10_JINGLE_PP-PREF_WELCOME_PP-PREF_PROMO_PP|M001"

```
SELECT AVG(talk_time)  
FROM ivr_data  
WHERE customer_category LIKE 'D' AND menu_path LIKE  
'PRE_HOST10_JINGLE_PP-PREF_WELCOME_PP-PREF_PROMO_PP|M001%';
```

5) Which IVR path has maximum number of customers.

```
SELECT ivr_path, max_cust_count
```

```

FROM
  (SELECT ivr_path , count(talk_time) as customer_count
  FROM
    (SELECT split(menu_path,'\\')[1] as ivr_path,talk_time
    FROM ivr_data) as a
  WHERE ivr_path LIKE 'M001:2' or ivr_path LIKE 'M001:9'
  GROUP BY ivr_path ) as b

JOIN
  (SELECT max(cust_count) AS max_cust_count
  FROM
    (SELECT ivr_path , count(talk_time) as cust_count
    FROM
      (SELECT split(menu_path,'\\')[1] as ivr_path,talk_time
      FROM ivr_data) as c
    WHERE ivr_path LIKE 'M001:2' or ivr_path LIKE 'M001:9'
    GROUP BY ivr_path) as d ) as e

```

ON max_cust_count=customer_count;

6) Total talk time where IVR Path is M001:2

```

SELECT SUM(talk_time)
  FROM ivr_data
  WHERE menu_path LIKE '%M001:2%';

```

B. IVR Data Analysis using Spark

- 1) Number of calls where total handling time is less than 30 sec but still it was transferred to customer care
- 2) Avg talk time of calls transferred to customer care
- 3) Avg talk time of calls that were not transferred to customer care
- 4) Avg Talk time of "D" Customer category where menu path is "PRE_HOST10_JINGLE_PP-PREF_WELCOME_PP-PREF_PROMO_PP|M001"
- 5) Which IVR path (first level path only like 9 after 1 or 2 after 1) has maximum number of customers
- 6) Total talk time where IVR Path is M001:2

C. IVR Data Analysis using Flink

- 1) Number of calls where total handling time is less than 30 sec but still it was transferred to customer care
- 2) Avg talk time of calls transferred to customer care
- 3) Avg talk time of calls that were not transferred to customer care
- 4) Avg Talk time of "D" Customer category where menu path is "PRE_HOST10_JINGLE_PP-PREF_WELCOME_PP-PREF_PROMO_PP|M001"
- 5) Which IVR path has maximum number of customers.
- 6) Total talk time where IVR Path is M001:2

III. Results

IVR Data Analysis - Performance comparison between Hive, Spark &Flink-

Query	Hive(Time in seconds)	Spark(Time in seconds)	Flink(Time in seconds)
Number of calls where total handling time is less than 30 sec but still it was transferred to customer care	299.819	65.628	103.521
Avg talk time of calls transferred to customer care	526.75	68.097	108.171
Avg talk time of calls that were not transferred to customer care	459.74	84.090	101.856
Avg Talk time of "D" Customer category where menu path is "PRE_HOST10_JINGLE_PP-PREF_WELCOME_PP-PREF_PROMO_PP M001"	580.33	75.986	86.853
Which IVR path (first level path only like 9 after 1 or 2 after 1) has maximum number of customers	548.059	169.209	119.379
Total talk time where IVR Path is M001:2	233.468	73.835	96.788

Table 1: IVR Data analysis Comparison between Hive, Spark &Flink

[11]. Number of calls where total handling time is less than 30 sec but still it was transferred to customer care

From the results we concluded that Hive has almost thrice the value observed in flink and almost five times from that of spark when Number of calls where total handling time is less than 30 sec but still it was transferred to customer care parameter is concerned. The figure below describes the clear picture of all the three

values observed in seconds as per number of calls are transferred to the customer care where total handling time is less than 30 seconds.

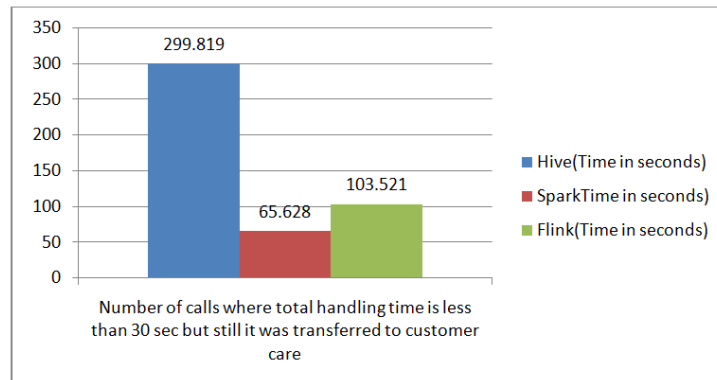


Fig. 1. Number of Calls transferred to customer care in Hive, Spark and flink

[12]. Average talk time of calls transferred to customer care

Figure 2 below describes the average talk time of the calls transferred to the customer care. Again it is clear from the figure that the Hive has the maximum value as compared to both the other frameworks. Here again the spark has the least value and its maximum value is only 68.097 seconds which is far less that of the Flink at 108.171 seconds and Hive at 526.75 seconds respectively.

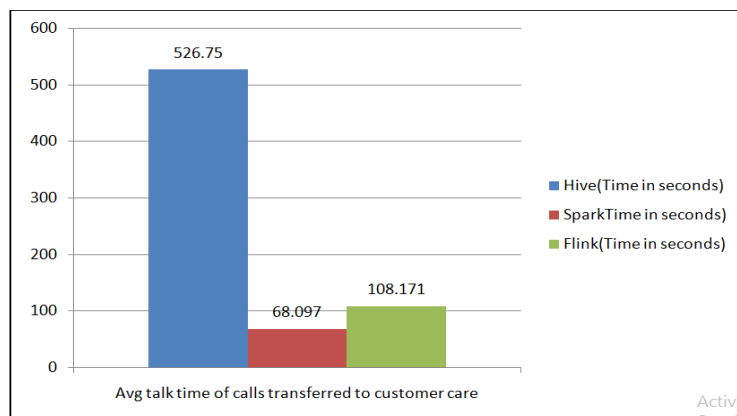


Fig.2. Average talk time of calls transferred to customer care

[13]. Average talk time of calls that were not transferred to customer care

Another parameter is considered and is analysed by three different frameworks in the call center. It is the average talk time of calls that were not transferred to customer care. As per the observations this parameter also shows almost the same results as it were observed in the previous parameter that focuses on average talk time of calls transferred to customer care. Themain thing noticeable here is the performance of Spark improves a bit in this by almost 14 seconds, whereas the values of other two is decreased by a marginal difference as compared to the previous values as shown in fig 3.

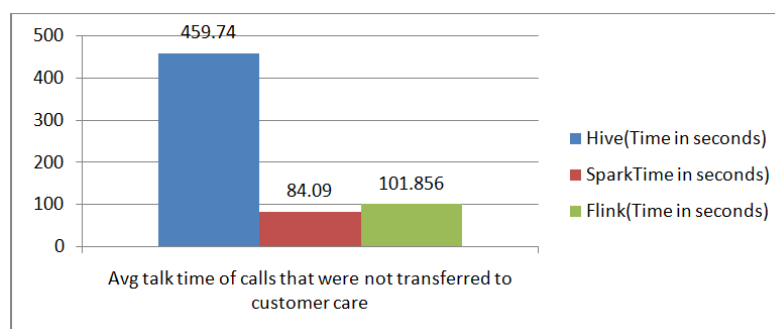
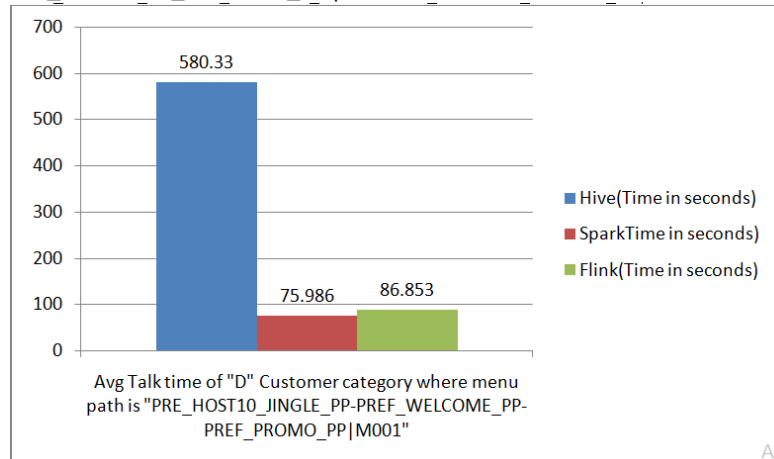


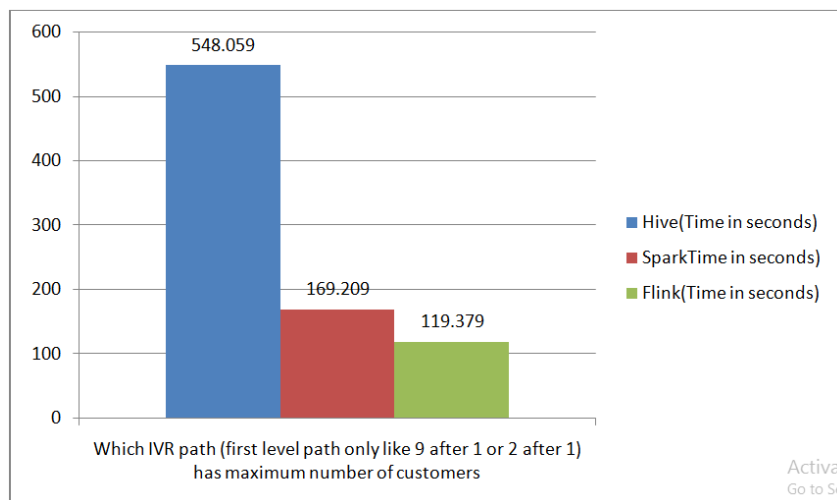
Fig.3. Averagetalk time of calls that were not transferred to customer care

[14]. Avg Talk time of "D" Customer category where menu path is "PRE_HOST10_JINGLE_PP-PREF_WELCOME_PP-PREF_PROMO_PP|M001"

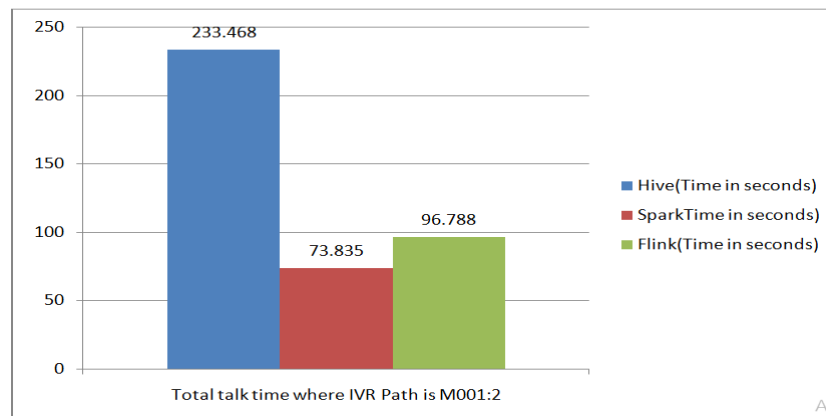


[15]. Which IVR path (first level path only like 9 after 1 or 2 after 1) has maximum number of customers

This factor focuses on the IVR path that has maximum number of customers, and as per the observations Hive again leads in this scenario and the main noticeable thing in this is that here the value of Flink degraded by a huge value and it almost goes down and stabilises at just the value of 119.38 seconds. It shows that it has the least number of customers.



[16]. Total talk time where IVR Path is M001:2



IV. Conclusion

From the detailed discussion and analyzation if IVR data we have concluded that in all the cases related to IVR data, HIVE outperforms in all the scenarios and is much ahead of Spark and Flink in almost every query that is analysed. The comparison table is made which explains how each framework differs in results of various queries implemented on the IVR data and the value of results are evaluated in terms to the response time measured in seconds for all the frameworks.

References

- [1] Tom White, "Hadoop Definitive Guide". O'Reilly.
- [2] XiaopengLi(2015) et. al" Performance Comparison of HiveImpala and Spark SQL". 2015 7th International Conference on Intelligent Human-Machine Systems and Cybernetics.
- [3] Dillon Chrimes(2016) et. al . , "Interactive Healthcare Big Data Analytics Platform under Simulated Performance", 2016 IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress.
- [4] Kuo M.H., Sahama T., Kushniruk A.W., Borycki E.M., Grunwell D. Health Big Data Analytics: Current Perspectives, Challenges and Potential Solutions. *Int J Big Data Intelligence* 2014; 1(12): 114–126.
- [5] Sun J., Reddy C.K. Big Data Analytics for Healthcare. Tutorial presentation at the SIAM International Conference on Data Mining, Austin, TX, 2013.
- [6] Nelson, R., Staggers, N. *Health Informatics: an interprofessional approach*. Mosby, an imprint of Elsevier Inc.; 2014. Saint Louis, MO.
- [7] Moselle, K. Data Management in the Island Health Secure Research Environment. Enterprise Architecture at Vancouver Island Health Authority. Working Draft 5; 2015. Victoria, BC.
- [8] <https://www.quora.com/What-are-the-differences-between-batch-processing-and-stream-processing-systems>
- [9] Hadoopproject. <http://hadoop.apache.org/>.