CLOUD-BASED DATA ANALYTICS FRAMEWORK FOR MOBILE APP EVENT ANALYSIS

PRANAV VILAS VAIDYA, JANAKI MEENA M, SYED IBRAHIM SP

School of Computing Science Engineering, VIT Chennai, Chennai, Tamil Nadu, India. Email: vaidyapranav.vilas2015@vit.ac.in

ABSTRACT

Mobile analytics studies the behavior of end users of mobile applications and the mobile application itself. These mobile applications, being an important part of the various business products, need to be monitored and the usage patterns are to be analyzed. The data collected from these apps can help to drive important business strategies by identifying the usage patterns. Enriching the data with information available from other sources, such as sales/ service information, provides holistic view about the solution. Thus, here, we aim at exploring some set of tools that give capabilities as event trailing with higher extraction of its linguistics. If the application is used worldwide, the data generated out of it is Big Data, which traditional systems cannot handle. We therefore propose a special framework for efficient data collection, storage, and processing at Big Data scale on cloud platform.

Keywords: Mobile analytics, Mobile metrics, Data collection, Segmentation analysis, Retention analysis, Acquisition, Engagement, Retention.

INTRODUCTION

The existing quality of cell phones and alternative mobile devices capable of connecting to the internet, including the dramatic increase in quality of powerful sensible phones such as Apple's iPhone, has resulted into a surge in interest among businesses throughout the planet. Unfortunately, the number of accessible experience for deploying mobile sites across various networks, devices, and application stacks is presently lacking. In things like this, net analytics tools are powerful champions that are suitable for use of activity to assist and verify the standard and efficaciousness of deployed sites and applications.

Mobile analytics shares several terms and ideas, which are identical to online analytics, and recognized by analytics experts. A smartphone and a personal computer (PC) will access the online and realize constant content employing a browser designed by a specific company. Constant metric report for a desktop user does not mean constant because he/she would be a mobile user also.

The key to the current distinction is in understanding the context and capabilities of mobile move with shopper behavior. The foremost and vital distinction is that the nature of shopper engagement is heavily influenced by the intimacy and private nature of mobile. At this point, mobile analytics is an understanding of this engagement.

Need of mobile analytics
Mobile analytics can have the following impact on the organization:
- Raising the end user experience: Mobile analytics should bear high resemblance to traditional web analytics, in terms of concepts, for the sake of existing marketers who are used to working with traditional methods. Each one is helpful for understanding how the customers consume content and what changes are created to optimize their browsing behavior. Analytics reports are generated with cause and effect eventualities to show the segments that need attention so as to boost the client expertise [1].
- Real-time knowledge collection: Knowledge is the cornerstone of partaking business campaigns. Mobile analytics harnesses the ability of knowledge to make additional relevant, partaking messages which will increase. Constant observation will deliver up-to-date information to refine the user expertise through options such as interactive visualization, associate interface to use filters, and varied views of knowledge. A whole lifecycle perspective on the way people use reports/apps to boost user engagement is essential to succeed with mobile analytics [1].
- Delivering custom content: Once businesses have established the programs and options that job for his or her customers, they will higher the future investments. With the addition of knowledge on the connecting device, mobile sites are progressively optimized to suit the precise interests of every shopper [1].
- Industry challenges in mobile analytics: Agencies, vendors, developers, and merchants are the various stakeholders in mobile-promoting industry. Those people face a variety of challenges that are required to be addressed brazenly by everyone operating in that area.
- Implementing analytics across mobile is hampered by knowledge Silo. A knowledge Silo could be a repository of mounted knowledge that a company does not often use in its daily operations.
  Knowledge from a mobile could be obtained through various ways. High level of complexity is involved in mobile analytics because of multiple sources of data. Knowledge could be obtained through SMS, mobile or web app behavior knowledge, app store knowledge, location-based knowledge, and mobile ad and QR code knowledge, all tend to reside in separate knowledge Silos. Compiling this information together is difficult; therefore the breaking down of these silos should be adapted as a common practice by the industry. [2,3].
- Tagging and chase users remain a technical hurdle: When new insights are got about the business, the findings shall be shared with others. Assembling is one of the ways to share the insights that have been found. Assembling analytics provides visualization of the data which will in turn help to explore and predict the data. Tagging and chase mobile user behavior are the crucial ways for assembling analytics; however, it is still in development stage for determining mobile capability. JavaScript could be a normal tagging mechanism, however this can be solely effective for smartphones capable of handling JavaScript and fully omits other feature phones. This can be turning into more difficult as users switch to smartphones with – modern mobile OS browsers – iOS, Android, and Windows Phone browsers – that all support JavaScript [2]. Pixel-based chase has been introduced as an answer for non-JavaScript-enabled devices; however, it is not normal to observe. Third party cookies, the bread and butter of online chase, are not supported by several platforms or are turned off by default, as on iOS, creating comprehensive net chase harder [2].
- Privacy remains a preponderating issue: Shoppers expect that some information are collected concerning their activities; however, this must be balanced with the customer's privacy – providing shoppers'

transparency, notice, and choice. The mobile-promoting trade works to make a self-regulative approach around these customers; whereas brand marketers and their media and analytics partners should be conscious of the foremost difficult gospel and also the management that customers should be ready to exert with any mobile device [2].

- Regional and national variations impact technical capabilities and rules: There are unit regional issues to be considered while implementing mobile analytics programs. The analytics made should also consider the lawful constraints of a network. Due to various network and native infrastructure, what is also technically or lawfully attainable in one country might not be attainable in another. Every country has its own privacy and knowledge assortment rules that confirm what will and cannot be done to contextually alter relevant programs [2,4].

LITERATURE SURVEY

- Mobile analytics is a growing field. This section of the paper reviews the research work carried out in mobile analytics and web analytics. Kumar *et al*. [5] discuss about multiple types of web analytics metrics and the techniques of data collection required for deriving them. Different sources and methods such as cookies, log files, page tagging, web beacons, and packet sniffing for data collection are suggested. They additionally mentioned completely different metrics and mechanism of evaluating websites and grouping information regarding behavior of the user. Identifying key dimensions is of utmost necessity to perform successful analytics. Batrinca and Treleaven [6] present detailed information about the software tools and their usages for social media analytics. Challenges such as scraping, data cleansing, holistic data sources, data protection, data analytics, and analytics dashboards are discussed. For efficiency, three types of database use are suggested: Flat file, relational database, and NoSQL database. They recommend MapReduce paradigm for social media analytics. Cassandra/Hive database, along with Hadoop and Mahout Machine learning implementations, can be used [6]. Easy availability of application programming interface (API)

and in turn access to huge amounts of data calls for special platforms and techniques. Verbert *et al*. [7] shed light on the importance of dashboards. Data completeness is directly dependent on the types of trackers and how the tools are used. An in-depth analysis of the data completeness is necessary to build dashboards that are useful for learners. Chen *et al*. [8] investigate the chance of privacy leakage through mobile analytics services and show the easiness with which an external body will extract an individual's profile and mobile applications usage information through two of the major mobile analytics services, i.e., Google Mobile App analytics and Flurry that have the highest market share. Access to information such as device ID can reveal all the information from Flurry. According to Chen *et al*. [8], it is possible not only to spoof the identity of an Android device but also to extract each user's profile. To tackle this risk, they suggest deprecation of permanent unique identifiers (Android ID for Android and UUID for iOS). Peterson *et al*. [9] emphasize on identifying the most important matrices from accuracy point of view. The technology that can be used needs to be thoroughly accessed. Understanding the specifics of solutions from the vendors is necessary. Integration with fixed web analytics is critical [9]. A framework for mobile analytics to grasp user engagement with mobile application is mentioned in a study by Fowler *et al*. [2]. It additionally focuses on business challenges to mobile analytics.

- Stodder [10] examines the impact of mobile computing on the development, management, and implementation of applications and services for business intelligence (BI) and analytics. Most organizations will have both native and web browser-based mobile BI applications. In some firms, ease of deployment on mobile is accelerating BI adoption at a faster pace than on PCs and workstations. Mobile BI and analytics can give road warrior sales, service, and support personnel with an information resource that they have never had before. Cloud computing and its services have a great potential to address the scalability issues [10]. A number of mobile analytics tools exist, and Table 1 shows the details of the tools available and the details about a firm's app that can be extracted from the tools.

Table 1: Head-to-head comparison of mobile analytics tools with respect to matrices

Metric	Google analytics + Firebase analytics	Flurry	Amazon mobile analytics	Answers on fabric	Countly	MixPanel	Adobe mobile analytics
Users - New and returning users	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sessions - Number of times our app was used on a particular day	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Session lengths - The period of time between app open and close	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Frequency - How many repeat visits there have been	Yes	Yes	Yes	No	Yes	Yes	Yes
Cohort analysis - Group of users who share a common characteristic	Yes	Yes	No	No	No	Yes	Yes
Error and crash reporting	Yes	Yes	No	Yes. specialized Crashlytics kit	Yes	Yes	Yes
Segmentation - Divide users based on demographics, technology, behavior, date of first visit, etc.	Yes	Yes	No	No	Yes	Yes	Yes
Funnels - Lets you see dropoff and conversion in any multi-step process	Yes	Yes	Through custom events	Through custom events	Yes. only through custom events	Yes	Yes
Retention - Percentage of users that use our app on a particular day and then again in the next coming days	Yes	Yes	Yes	Yes	Yes	Yes	Yes

PROPOSED FRAMEWORK

The proposed framework takes into consideration the scenario where the data extracted out by the mobile analytics tools need to be enriched with information available from other sources, such as sales/service information from an organization. The visualizations need not be real time. Data of this nature classify into Big Data since the devices keep on generating large amount of data for multiple applications that are real time and unstructured. The main motivation behind designing the framework is to ensure the data privacy.

Fig. 1 shows the framework proposed for mobile analytics. The framework takes all the advantages provided by the cloud platform and ensures that the cost of infrastructure stays low. IaaS is used here, and using a cloud platform also increases the reliability and scalability of the framework.

The framework also takes care and ensures that the data flow through the pipeline and does not suffer any bottleneck and is immune to system failures and data loss.

- Vendor servers: These are the servers that host the data collected by various mobile apps. The mobile application developers integrate the Software development kits provided by the third party tools into their applications. The vendors keep on recording the statistics and events.
- Data-fetching API: The vendor servers are incapable to make a push to the data. Therefore, there is a need to make a pull to the data. This pull request can be made through the APIs provided by the vendors. The APIs are widely of RESTful nature. The vendors may have a limit on the number of calls that can be made and predetermined nature of data that are sent as response. The need for this arises because the vendors do not store the data on their location forever. There is a predetermined time after which they execute a flush out and retain only the recent data. There is a need to get these data out before they are flushed out. This block sends out data to the next block in small chunks so as to ease out the processing.
- API for pushing and acknowledgment: This block serves as a dedicated process for indexing and file handling. This block is separated out from data-fetching API to avoid a single point of failure. This block continuously keeps a track of which files are fetched from the vendors, which have been uploaded to permanent storage on cloud. This block can take care of sending data to Hadoop cluster forward. Sometimes, the data might need to be broken down or restructured. These actions can be done in this block. After every successful reception of data from the fetching API, it sends out an acknowledgment. The raw data received from the vendors need to be uploaded onto a permanent storage as well for future references.

This block keeps on pushing data as they come from the fetching API to the cloud storage.
- Hadoop cluster: This block serves as the central processing unit for the framework. It may serve various purposes as follows:
  - Preprocessing: Transformation of the data, such as converting forms, aggregations, derivations, splits, merges, and sorts.
  - Enrichment: Mixing up of data from others sources within the organization, adding necessary attributes, etc.
  - Predictions/machine learning: Numerous algorithms can be executed to get the most out of the data as per the requirements. Predictions can be made or neural networks can be used for learning patterns. For huge data as such, running them on a cluster works out as the best option.
- Intermediate process: Hadoop cluster might generate intermediate results that may need future references. Such results need to be pushed to permanent storage since clusters are prone to crashes. This block serves the purpose of storing the intermediate results or the final findings to the permanent storage. It keeps on performing the pull operation periodically on the Hadoop cluster. A separate block for this task ensures that the data are not lost in case of cluster failure and the damage stays minimum. On the other hand, this block also keeps on pushing the data to the cloud storage.
- Cloud storage: This block serves as a permanent storage for the whole framework. The data in every other block of the framework are prone to loss. Introducing this block in the framework ensures that the problem of data loss and backup is taken care of. Replication is a strategy used for ensuring that the data are not lost. A good replication factor has to be chosen such that the data are never lost and also the storage is not very much higher.
- Database store: This is a structured database that stores the results and findings of the jobs executed on the Hadoop cluster. This database needs to be structured because the visualization application demands so. After numerous transformations and applying various algorithms, the data may be taken care of by semi-structured databases. In addition, if the demands of visualization are in real time, these databases are the most appropriate to serve the purpose.
- Visualization logic: Various visualization tools such as Tableau and R can work in this block. They can query for the required data from the database store and produce visualizations.

ADVANTAGES OF THE PROPOSED FRAMEWORK

- The proposed framework is free from single point of failures. The framework is designed in such a way that, even if the components fail, there will be no or, in extreme cases, minimum loss of data which can be easily regained.
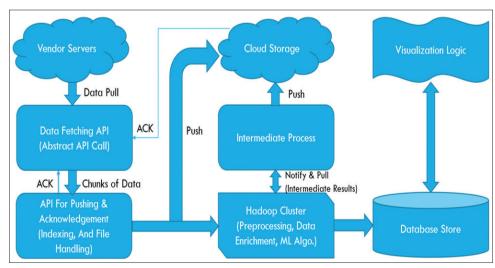- The framework can handle Big Data very easily since it is cloud

Fig. 1: Proposed framework

based and uses Hadoop for distributed processing. The framework can be scaled up and down according to the requirements. Since we are proposing the cloud platform, the infrastructure is effectively managed; henceforth, the framework is very cost effective.

- The framework is flexible since it can be fully cloud based or partially cloud based. If it is implemented as fully cloud based, it inherits all the advantages of the cloud computing paradigm. The blocks are designed and placed such that the complexity of the system becomes least. It therefore eases up the maintenance efforts. Since the data reside with cloud service providers, data privacy and security is well taken care. Though the framework is designed specifically for mobile analytics, it can be molded and used for other data processing as well.

DRAWBACKS OF THE PROPOSED FRAMEWORK

Due to redundancy and distributed processing, the framework does create latency with respect to real time processing and visualization.

CONCLUSION

Mobile analytics has established its existence in business by providing completely different analytic services to vendors and business individuals. It is noticed that almost all users access internet from laptop to mobile. Companies track mobile websites over applications and giant enterprises extend their existing analytics investments. Mobile analytics vendors are moving quicker which signifies that there are great days ahead for mobile analytics and BI. Thus, through this paper, we tried to overview the requirement of mobile analytics with its advantages in addition to industrial challenges. Dashboards' availableness is extremely vital for many mobile BI and analytic users. Dashboards will instantly increase speed by delivering information during a manner that renders it usable and essentially qualifies for many users of mobile

devices. Along with dashboards, performance metrics and KPIs are also key options that organizations wish to examine on mobile devices. A framework is proposed which takes into consideration the issues of privacy and scalability of Big Data processing with respect to mobile analytics. The framework needs to be extensively tested to discover and eliminate drawbacks and point of failures.

REFERENCES

1. Astonkar MR, Buchade A. Exploring mobile analytics for business intelligence. Int J Comput Sci Mobile Comput 2014;3(12):603-11.
2. Fowler S, Becker M, Scullin L, Ricci M. The mobile marketing association primer on mobile analytics. Mobile Marketing Association. 2012. p. 1-17.
3. Muthusamy G, Ramalingam R. Mobile Enterprise Analytics in 60 Minutes. Whitepaper by Cognizant 20-20 Insights. December; 2013. p. 1-5.
4. Sleeper M, Consolvo S, Staddon J. Exploring the benefits and uses of web analytics tools for non-transactional websites. Proceedings of the 2014 Conference on Designing Interactive Systems, 2014, June. p. 681-4.
5. Kumar L, Singh H, Kaur R. Web analytics and metrics: A survey. Proceedings of the International Conference on Advances in Computing, Communications and Informatics, 2012, August. p. 966-71.
6. Batrinca B, Treleaven PC. Social media analytics: A survey of techniques, tools and platforms. AI Soc 2015;30(1):89-116.
7. Verbert K, Govaerts S, Duval E, Santos JL, van Assche F, Parra G, *et al*. Learning dashboards: An overview and future research opportunities. Pers Ubiquitous Comput 2014;18(6):1499-514.
8. Chen T, Ullah I, Kaafar MA, Boreli R. Information leakage through mobile analytics services. Proceedings of the 15th Workshop on Mobile Computing Systems and Applications, 2014, February. p. 15.
9. Peterson E, Berger M, Pottjegort T. The Truth about Mobile Analytics. Camas, WA: Web Analytics Demystified and Nedstat; 2009. p. 1-24.
10. Stodder D. Mobile business intelligence and analytics. TDWI Best Practices Report. 2012. p. 681-4.