International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-6, March 2020

Monitoring High Throughput Distributed System using Statistical Data Analysis

Divya Jain, Swarnalatha P.

Abstract: ***Monitoring high throughput distributed system by using a statistical analysis of the historical time series of an Instrumentation Data . The Pipeline has been made to process the information which can be otherwise called data pipeline, is a lot of information handling components associated in arrangement, where yield of one component is the contribution of the next one . Several codes are giving different visualization for statistical analysis of data. Network and Cloud Data Centers generate a lot of data every second; this data can be gathered as period arrangement information. A time-series is a grouping taken at progressive similarly dispersed focuses in time that implies at a particular time interval to a particular time, the estimations of explicit information that was taken is known as information of a time-series. This time-series information can be gathered utilizing framework measurements like CPU, Memory, and Disk utilization . The TICK and ELK Stack is abbreviation for a foundation of open source instruments worked to make collection, storage, graphing, and alerting on time arrangement data incredibly easy. As an information collector, using Telegraf, for storing and analyzing information and the time-series database InfluxDB and Elasticsearch. For plotting and visualizing used Grafana and Kibana. Watchman is utilized for alert refinement and once system metrics usage exceeds the specified threshold, the alert is generated and sends it to the Telegram.***

Keywords: ELK, TICK, Watchman, Monitoring, Grafana

I. INTRODUCTION

In DTH_LAPU DTH stands for Direct-to-Home and LAPU stands for Local Area Payment Unit . Direct-to-Home (DTH) television is a strategy for accepting satellite TV by methods for signals transmitted from direct-communicate satellites. LAPU (Local Area Payment Unit) is the customary energize framework that is consolidated in everyone's life. LAPU recharge is done either by mobile or LAPU SIM, which is a SIM that is utilized to revive portable and DTH SIM cards. DTH_LAPU is an application which is serving DTH Recharge or checking the Balance of DTH by Mobile i.e., smart phones or by keypad phones (old time phones) or by mobile application.

Manuscript received on March 15, 2020.
Revised Manuscript received on March 24, 2020.
Manuscript published on March 30, 2020.
* Correspondence Author
Divya Jain, School of Computer Science and Engineering Vellore Institute of Technology Vellore, India
Prof. Swarnalatha P., School of Computer Science and Engineering Vellore Institute of Technology Vellore, India

The request will be send to the Load Balancer which is a gadget that goes about as a reverse proxy and appropriates system or application traffic over various servers. It is utilized to expand limit (concurrent users) and reliability of applications. They improve the general execution of utilization by diminishing the weight on servers related with managing and keeping up application and network sessions, just as by performing application-explicit errands. Load Balancer will send request to the Interface server which is sent on weblogic server through them get the condition of HTTP request. Different status code have different meaning (100-199) Informational responses, (200-299) Successful responses, (300-399) Redirects, (400-499) Client errors, and (500-599) Server errors. Then if the response is Informational or Successful the request will be sent to the application server on which java code is deployed which will assign different thread ID to different processes. Filebeat has been introduced on each Interface servers and Application servers . Filebeat is a lightweight shipper for sending and bringing together log information screens the log documents or areas that you determine, gathers log occasions, furthermore, advances them which will send all the server information to redis. Redis is in-memory data structures, for example, strings, hashes, lists, sets, arranged sets with range queries, bitmaps, hyperlogs, geospatial files with radius queries and streams . All the redis will send the information to Logstash which can pull from practically any information source utilizing input modules, apply a wide assortment of information changes and improvements utilizing channel modules, and ship the information to countless goals utilizing output plugins. Then in output plugin we have given two paths to dump the data i.e., InfluxDB which is a stage for storing, collecting, visualizing and managing time-series data . It is quicker than MySQL . Presently InfluxDB is the mostly used time-series database . And elasticsearch provides real-time search and analytics for all types of data . Whether or not you have organized or unstructured content, numerical information, or geospatial information, elasticsearch can effectively store and file it such that supports quick searches. You can go a long ways past basic information recovery and total data to find patterns and patterns in your data. What's more, as your information and query volume develops, the dispersed idea of elasticsearch empowers your sending to become flawlessly directly alongside it . Then the data which is sent into the InfluxDB will be visualized in grafana which is utilized for metric analytics and visualization suite . It is most ordinarily utilized for imagining time arrangement information for foundation and application investigation. And the elasticsearch data will be visualized in kibana which is representation apparatus predominantly used to break down enormous volume of logs in the structure of line graphs, bar graph, pie chart, heat maps, region maps, coordinate maps, gauge, goals, timeline, etc .

The representation makes it simple to anticipate or to see the adjustments in patterns of mistakes or other noteworthy occasions of the input source.

II. METHODOLOGY

A. Load Balancer

With the load balancer, you'll split the workload and balance it between two or more servers. As a result, we can configure our infrastructure to maximize activity, optimize resource allocation, and provide a littlest measure of response time [1].

Utilizing a load balancer is recommended in all cases, regardless of whether we require in any event one of the following:

By utilizing load balancer in all circumstances [2], we can avail the following one or more requirements

- Continuity of the guaranteed service
- Manages a lot of traffic
- Prepare sudden peaks in the application

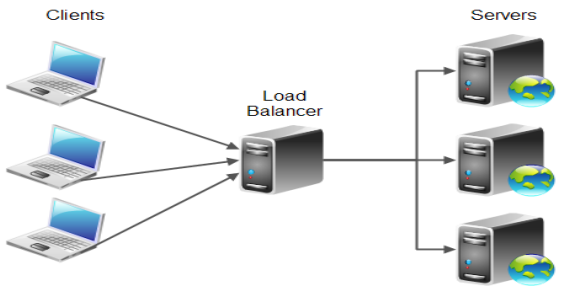

Figure 1 Load Balancer [2]

B. Elasticsearch

RDBMS support services that have a schema. Unstructured or semi- structured information need to be indexed. It has gotten imperative to make another stage to satisfy the interest of association because of the difficulties looked by conventional information [3]. Elasticsearch is a big data innovation which is schema less. To search big Data this tool is used. Elasticsearch makes use of the idea of denormalization for search. Elasticsearch makes use of the indexing idea. It is a record arranged instrument. Once the report is delivered, it could be searched within a next second as Elasticsearch is real time.

Elasticsearch is utilized for plenty use cases like analytics shop, automobile completer, spell checker, alerting engine, and as a trendy cause file keep; full text seek is considered one of it. It is a robust search engine that offers a short complete text search over various files. It searches within full textual content fields to discover the document and return the maximum applicable result first. The relevancy of files is ideal as Elasticsearch makes use of Boolean model to locate file. When a record coordinates a query, Lucene ascertains its score for that inquiry, joining the scores of each coordinating timeframe. The significance of the record can be determined the use of practical scoring capacity [3].

C. Logstash

Logstash is an open source information collection engine. It additionally does indexing on massive quantity of data or logs which might be gathered from exceptional servers. This information is passed to the elasticsearch for further query processing. It can dynamically unify data from variety of assets and normalize the records into framework of consumer's choice. It is a tool to accumulate, system, and forward occasions and log messages. Assortment is finished by means of configurable information modules together with crude socket/packet correspondence, document tailing, and numerous message bus clients. It does the four main duties along with parsing the facts and logs, extracting the records and logs, managing the logs and structuring it. We get server logs event viewer logs and application custom logs [4].

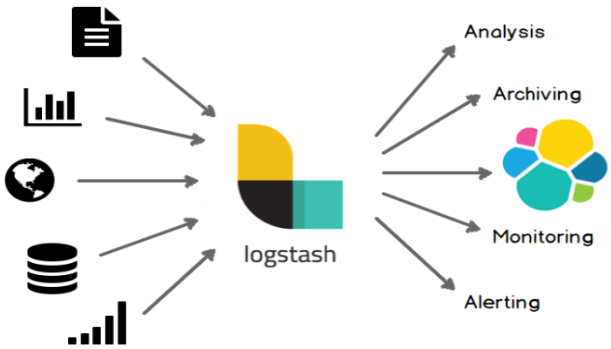

Figure 2 Logstash [4]

D. Kibana

Kibana turned into designed as a perception stage for Elasticsearch. It provides web-based totally interface for seek, view and examine statistics stored in Elasticsearch cluster. The main view of Kibana is divided into four main components - Discover, Visualize, Dashboards and Management [10].

Kibana is an open source information representation stage that permits you to connect with your data through dazzling, incredible designs that might be mixed into custom dashboards that help you rate insites out of your information far and broad. This tool does the responsibilities inclusive of exploring, visualization and discovering data. Depending upon the query and the JSON response the outcome is generated. Kibana creates tables, graphs, pie charts etc. Thus, kibana does clean representation of big volumes of records and provides analytics. Our system use these equipment enables different agencies producing massive quantity of facts in managing, storing and gives information, illustration of records in smooth graphical forms. It converts complicated and unstructured information to structured information and indexing for its smooth representation [4].

E. Telegraf

Telegraf has a consolidation of plugins and rules for retrieving the measurements from the diverse system metrics. It takes distinctive metrics from Application Program Interface (API) or it will pay attention for measurements by means of StatsD and Kafka client services (TICK stack). Also, Telegraf has output plugins for sending the gathered metrics records to different data stores like Influxdb, graphite, OpenTSDB, and Kafka [5].

Telegraf is a daemon which can run on any server and accumulate a wide sort of measurements from the framework (CPU, memory, change, and so forth.), commonplace offerings (MySQL, redis, postgres, and many others.), or third-party APIs. It is module driven for both collection and output of information so it is effectively extendable. It is also written in Go, which implies that it's miles a compiled and standalone binary that may be accomplished on any device without a want for external dependencies.

F. InfluxDB

Influxdb is a stage for storing, collecting, visualizing and handling time-series information . It is quicker than MySQL.

InfluxDB is utilized as an information store for many use cases which include enormous amounts of time stamped records and also used in lots of application metrics monitoring and real time analytics [5].

Here are a portion of the features that InfluxDB presently supports that make it an exquisite desire for operating with time-series information.

- Custom excessive overall execution information store written particularly for time collection facts. The TSM engine lets in for high ingest speed and information compression.
- Written absolutely in Go. It incorporates into a solitary paired without outside dependencies.
- Simple, excessive performing writes and queries HTTP APIs.
- Modules support for other information ingestion conventions which include Graphite, collected, and OpenTSDB.
- Expressive SQL like query language tailored to easily query aggregated data.
- Tags permit series to be listed for immediate and efficient queries.
- Retention guidelines efficiently auto-expire stale data.
- Non-stop queries automatically compute mixture records to make frequent queries extra efficient.

G. Filebeat

Filebeat is a lightweight shipper for sending and concentrating log information. Mounted as a specialist on servers, Filebeat screens the log documents or areas that you determine, gathers log exercises, and advances them both to Elasticsearch or Logstash for ordering.

Here's how Filebeat works: When Filebeat begins, it starts off-evolved at least one inputs that appearance inside the locations you've distinctive for log data. For each log that Filebeat finds, Filebeat begins a collector. Every collector peruses a single log for new substance and sends the new log information to libbeat, which totals the events and sends the accumulated information to the output that you've configured for Filebeat [8].

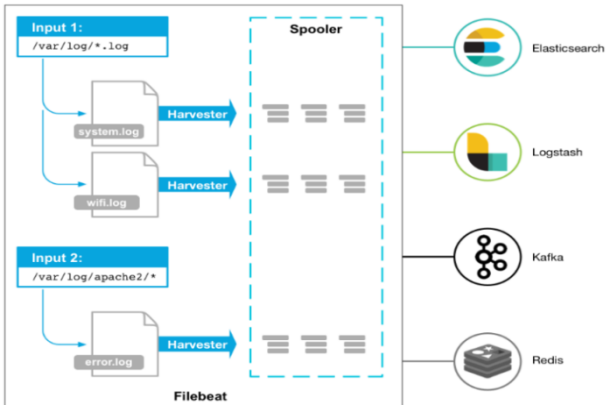

Figure 3 Filebeat [8]

H. Redis

Redis is an open source, in-memory information structure store, utilized as a database, reserve and message dealer. It helps data structures including strings, hashes, lists, sets, sorted sets with range queries bitmaps, hyper logs, geospatial indexes with radius queries and streams . Redis has coordinated replication, Lua scripting, LRU expulsion, exchanges and various degrees of on-disk determination, and gives over the top accessibility by means of Redis Sentinel and programmed dividing with Redis cluster [6].

Redis does now not assist complicated queries or indexing, but has help for data structures as values. Being quite simple it is moreover fastest KVS implementations for plenty fundamental operations. Speed and data structures are intriguing blend that can be for simple thing tolerance. Data structures in Redis are extra usually known as Redis Data types. These include strings, lists, sets, sorted sets and hashes . Redis gives commands that can be utilized to alter these sorts. For example list underpins typical rundown activities, similar to push and pop [7].

I. Grafana

Grafana is an open source metric evaluation and visualization suite . It is most ordinarily utilized for visualizing time arrangement information for foundation and application investigation.

Grafana is one of the most generally utilized visualization tools with regards to time-series data analytics. Grafana licenses querying, visualizing and understanding the measurements from any time-series database by method for various powerful and reusable dashboards containing a huge variety of graphs, charts and other plugins. It additionally lets in to seamlessly defining alert guidelines and thresholds for the most critical metrics. Grafana will persistently assess and send notifications to frameworks like slack, email, Telegram, and so on [9]. Create, discover, also, share dashboards and foster information driven subculture:

- Visualize: Rapid and bendy client side charts with a large number of choices. Panel plugins for plenty exclusive methods to visualize metrics and logs .
- Dynamic Dashboards: Make dynamic and reusable dashboards with format factors that show up as dropdowns at the highest point of the dashboard.

- Explore Metrics: Discover your information through ad-hoc queries and dynamic drilldown. Split view and assess distinctive time extents, queries and information sources next to each other.
- Explore Logs: Experience the enchantment of changing from measurements to logs with protected name channels. Quick inquiry through the entirety of your logs or spilling those remains.
- Alerting: Visually define alert rules for your most important metrics. Grafana will continuously evaluate and send notifications to systems like slack, Pager Duty, VictorOps, OpsGenie.
- Mixed Data Sources: Mix different data sources in the same graph. You can specify a data source on a per-query basis. This works for even custom data sources.

J. Watchman

Watchman exists to watch documents and record when they change. It can likewise trigger activities when coordinating documents change.

Watchman Monitoring is a Software as a Service (SaaS) presenting which monitors the health of Mac, Linux, and home windows computer systems. Our monitoring system offers hourly reviews on health problems, for example, disk I/O errors, reinforcement capability, and RAID status. Watchman monitoring consists of two primary components:

1. Monitoring Agent
   The monitoring agent is introduced utilizing one among numerous quick and handy strategies. Once introduced, the operator runs hourly and surveys its discoveries to the Watchman monitoring Server.
2. Monitoring Server
   The watchman monitoring server gathers audits from monitored machines and informs endorsers as issues are distinguished. The server dashboard gives a subscriber at-a-look repute of all monitored computers, access to powerful search, inventory and demographic facts.

IV. PROPOSED ARCHITECTURE

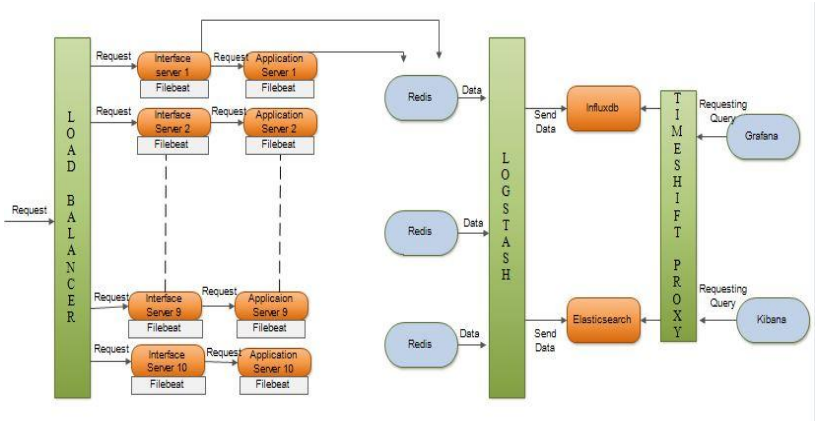

Figure 4 Proposed Architecture

V. RESULT AND DISCUSSIONS

STK is SIM toolkit. In STK we get to recognize that how much request has come for recharge or stability from logs provided.

We monitor STK utility which serves the STK requests through shell script, which frequently check whether the STK application code is running or not.

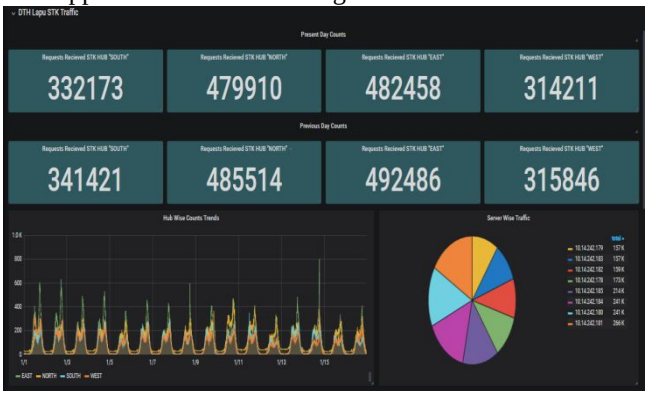

Figure 5 DTH Lapu STK Traffic

- The application generates a real time logs which are stored in a file.
- The document is perused by Filebeat and sent to the Redis which is in-memory data structure store, utilized as a database.
- Information from the redis is devoured by Logstash configuration file.

Figure 6 STK Process Status on Application servers

- The request gets landed on LB when it enters the Airtel network

- Load Balancer will send request to interface servers .
- On Interface server weblogic is deployed.
- We regularly read weblogic access logs to monitor response time and response code for particular request, where 200 means everything is fine and "500" response code implies there is a server error, which is a critical application problem.

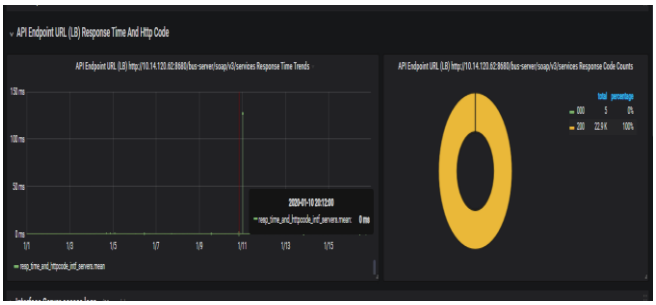

Figure 7 API Endpoint URL (Load Balancer) Response Time and HTTP Code

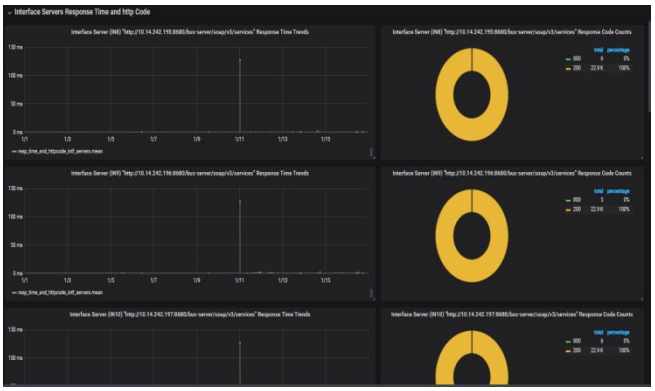

Figure 8 Interface Servers Response Time and HTTP Code

- Interface server will create there access logs.

Figure 9 Interface Server access logs

- Interface server will send request to Application server on which java code is deployed from which a thread ID is assigned to each processes.
- In application server we get to know the API performance from the application logs.

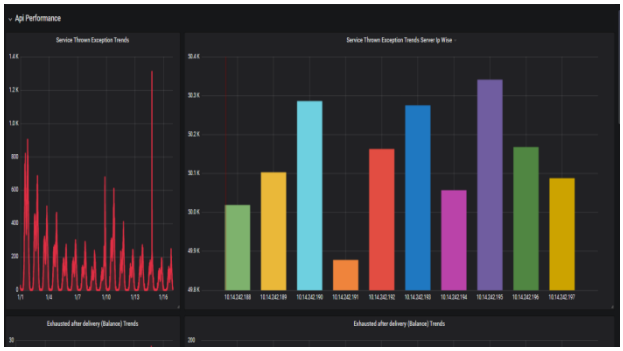

Figure 10 API Performance

Figure 11 Counts from Application Logs that how much request came

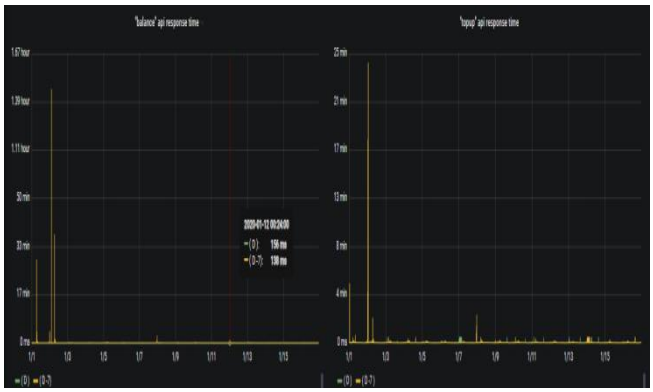

Figure 12 Graphical Representation of counts

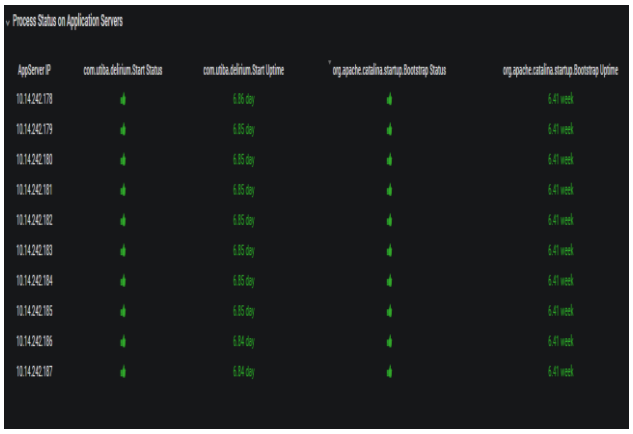

Figure 13 Process Status on Application Servers

Retrieval Number: F9810038620/2020©BEIESP
DOI:10.35940/ijrte.F9810.038620
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Figure 14 Health Metrics App Servers

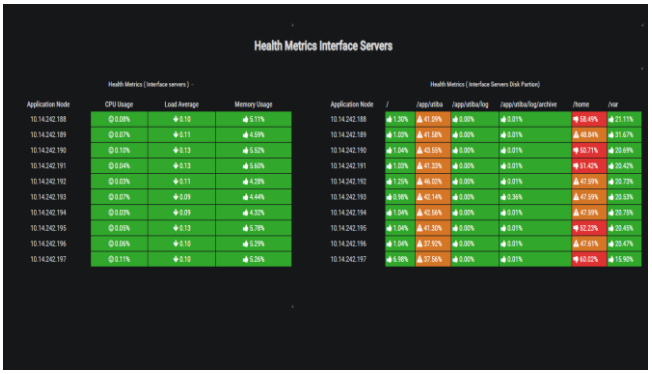

Figure 15 Health Metrics Interface Servers

Figure 16 Graphical Representation of Health Metrics

- If threshold bleach its range which has set then alert will be send alert to the telegram.
- The alert configuration (code) written in watchman tool in which we have set threshold limits which is written in a query and in the configuration we have written the input file from which we need to take the input and where we need to send the alerts i.e., telegram.

Figure 17 Telegram Alert

VI. CONCLUSION

In this paper, we are monitoring end to end health of DTH_LAPU . Develop over the previous decade to keep up this monstrous framework in a shrewd, effective, and adaptable way. Elasticsearch, Logstash, Kibana, Telegraf, Influxdb, Grafana, Filebeat, Redis, Watchman was used to capture, transform, enrich, store, index, alerts, select relevant time slots and generate graphs that were integrated in a dashboard for combined visualization and analysis. By checking, we can improve the general execution of utilization by diminishing the weight on servers related with overseeing and keeping up application and system meetings and sparing the time by getting alarms on wire through which we find a workable pace where the issue has happened, by performing application-specific tasks.

REFERENCES

1. Rimal, Prasad B, Choi E, Lumb V (2009) A taxonomy and survey of cloud computing systems. Proceedings of 5th International Joint Conference on INC, IMS and IDC, IEEE .
2. Sinha PK (1997) Distributed operating Systems Concepts and Design. IEEE Computer Society Press .
3. Kalyani, D., and D. D. Mehta. "Paper on searching and indexing using elasticsearch." Int. J. Eng. Comput. Sci 6 (2017): 21824-21829 .
4. Sunny Advani, Meghna Mridul (2016). "Log analytics using ELK stack on Cloud platform." IJARCCE.2016.5413 .
5. Girish, L. (2018). Efficient Monitoring of Time Series Data Using Dynamic Alerting .
6. https://redis.io/topics/introduction .
7. Paksula, M. (2010). Persisting objects in redis key-value database. University of Helsinki, Department of Computer Science .
8. https://www.elastic.co/guide/en/beats/filebeat/current/filebeat-overview.html .
9. Paul, D., & Sala, P. (2019). Real-Time Server Monitoring and CNN Inference on FPGA .
10. Bajer, Marcin."Building an IoT data hub with Elasticsearch, Logstash and Kibana." 2017 5th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW). IEEE, 2017 .

AUTHORS PROFILE

Divya Jain pursuing M.Tech in Computer Science and Engineering from Vellore Institute of Technology, Vellore . And received a B.Tech degree in Computer Science and engineering degree from the Govt. Women Engineering College Ajmer .

Swarnalatha P is an Associate Professor, in the School of Computer Science and Engineering, VIT University, at Vellore, India. She pursued her Ph.D degree in Image Processing and Intelligent Systems. She has published more than 120 papers in International Journals/International Conference Proceedings/ National Conferences. She is having 18+ years of teaching experiences. She is a member of IACSIT, CSI, ACM, IEEE (WIE), and ACEEE.