February 21, 2018

Net Neutrality Value Pack using Network Data Analytics

Vidya V. Bharadwaj
Hewlett Packard Enterprise

Prashanth G.A
Hewlett Packard Enterprise

M Mahesh Babu
Hewlett Packard Enterprise

Suhasan Chirravuri
Hewlett Packard Enterprise

Recommended Citation

Bharadwaj, Vidya V.; G.A, Prashanth; Babu, M Mahesh; and Chirravuri, Suhasan, "Net Neutrality Value Pack using Network Data Analytics", Technical Disclosure Commons, (February 21, 2018)
https://www.tdcommons.org/dpubs_series/1059

Net Neutrality Value Pack using Network Data Analytics

Abstract

The advent of mobile internet and the phenomenal growth of the use of smart phones has brought data onto the forefront, creating newer revenue streams for the operators. The data/Internet connection now needs to cater to diverse traffic, just as a city must manage the flow of various vehicles and pedestrians on its streets. In the data world, usage of data ranges across various applications like streaming-video, real time gaming, B2B & M2M applications. Such diverse customers often blame their operators for throttling data flows to the phones or computers. This causes significant delays and losses in data transmission. Any lapses of providing connectivity and continuity to network will create a large number of dissatisfied customers and unwarranted reduction of customer base. Network neutrality is an idea, that all operators should treat all data that travel over their networks fairly, without improper discrimination in favor of particular apps, sites or services. However it is a complex, controversial topic and is an important part of a free and open Internet. It aims at enabling access, choice, and transparency of Internet offerings, there by empowering users to benefit from full access to services, applications, and content available on the Internet. Implementing network neutrality legitimately without discrimination in favor of particular applications, sites or services have been a challenge faced by operators globally. This paper describes a Net Neutrality value pack using the Smart Profile Server (SPS). SPS is an enterprise application which forms the middleware to collect & analyze the network data to build and expose a data model having network traffic info w.r.t. session throughput, speed classification, page reloads etc. for a given customer/subscriber at a given time & location using the analytic database (DB). This data model can be either exposed as a REST [1] based interface as a smart profile view with fine grain access control or tied to 3rd party dashboard tools to act as a window to subscribers & regulation agencies to determine if the operator is truly net neutral.

Problem statement

The telecom industry worldwide has experienced a paradigm shift from voice to data. The rapid adoption of smart devices, plethora of useful applications/online services, and increased mobile & broadband penetration are leading to huge growth of data traffic. This also means increase in the consumption of bandwidth. Operators have also started facing competition from Over-the-Top (OTT) players in their traditional voice space. With an objective of enhancing revenue streams and to compete from OTT players, operators have been exploring new opportunities for generating revenues from subscribers and the content providers. Some of the models attempted by operators, such as charging higher data tariffs for VoIP services, charging content/application providers and providing the content free to subscribers (called “zero rating” plans), have raised concerns about Network Neutrality. In May 2015, Airtel, a tier-1 operator in India faced severe backlash when they introduced Airtel Zero. Under the scheme, application firms signed a contract, in which Airtel provides the apps for free to its customers. The reports of Flipkart, an e-commerce firm, joining this scheme drew negative response. This lead to Flipkart walking out of the scheme and Dept. of Telecom releasing a Net Neutrality principles [2] document.

This in principle must be upheld to protect the future of our open Internet. However, there have been several incidents such as operators forging packets to tamper with certain kinds of traffic or slowing down or even outright blocking protocols or applications which can be a serious threat to network neutrality principles. Practicing net neutrality comes up against a variety of constraints like secure network from attacks, legal compatibility, maintain acceptable level of Quality of Service (QoS) for real time services etc. For example, a data packet carrying emergency service information versus a data packet carrying video/email have to be treated differently. In principle if an operator implements an application-agnostic congestion control for a legitimate reasons, then they cannot be considered to be against Net Neutrality. However application-specific control within the “Internet traffic” may be against the principles of Net Neutrality. Today in the absence of standard definition of Net Neutrality, reliability on the operator generated QoS reports, an ecosystem with emerging business models and technologies can lead to the discrimination by fixed or mobile operators with market power in favor of their own applications, content and services, thus harming both competition and consumers.

With prior understanding of the mobile network elements, operators data sources, we realized that there is a huge opportunity to capture additional value by building an integration layer of analyzed data that bridges the gap of ideal and practical world to provide holistic view of the network usage w.r.t the total speed (upload +download), latency, retries etc. at a given time & location for a subscribers accessing the same content/application.

Our solution

The logical architecture of SPS is shown as three main layers in Fig.1.

Data Collection Layer collects subscriber’s data using real time and/or in batches interfaces and converts them to standard file format such as Comma Separate Values (CSV) files. The subscriber data can be

- Call Detail Record (CDR) having voice usage details of subscribers such as calling number, called number, call duration, results of the call indicating whether or not the call was connected, any fault condition encountered etc.
- Internet Protocol Detail Record (IPDR) having the data usage details of subscribers such as IP address, service usage details like session details, data volume details etc.
- Metadata having subscriber details from device related data, Customer Relationship Management (CRM) data etc.

Data Analysis Layer is based on algorithms developed by us for various statistical computations and analysis to support the values packs created and deployed on SPS. The analysis is run and result set is stored in an analytic DB called SPS Analytics Repository. A Value Pack is a unit for collection of many analysis and a data model supporting them. Each analysis processes the input facts and produces result facts. Several Value Packs can be deployed into an instance of SPS.

Data Exposure Layer acts as a single-source (analytics DB), single-protocol (REST based), single transaction (synchronous call) interface for internal/external applications service providers and subscribers with fine grained access control and security to the view analysis results.

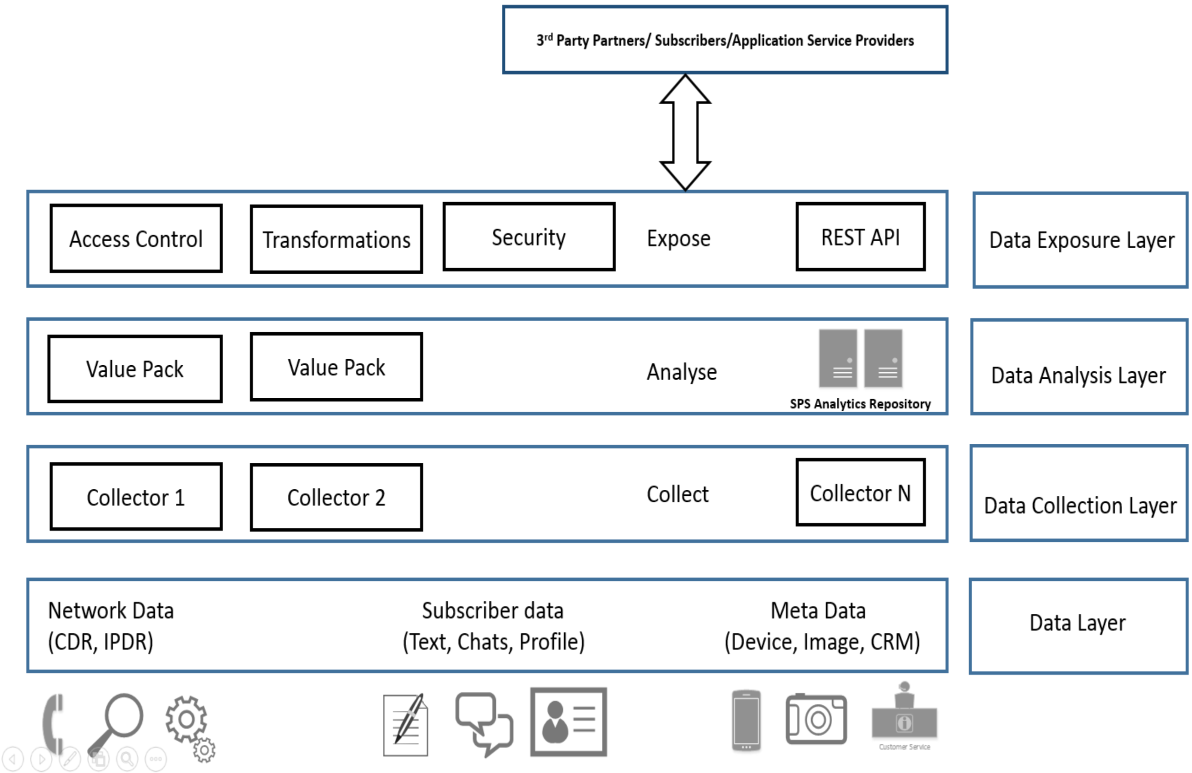


Figure 1 Logical Architecture of SPS

A proposal of a Net Neutrality Value Pack is as follows:

- Design of data model in a star schema [3], which is independent of the original data sources in an analytic DB. A snapshot of the Net Neutral Value Pack data model is shown in Fig.2. This data model has
  - Dimension tables contains several identifiers for a fact such as subscriber information, date/time, and website and access information. The dimension tables helps in filtering, grouping and labelling of the data. These dimensions are referenced in a Fact Table using a primary key column that has foreign key toward the dimension table.
  - Fact tables contains measures related to a set of dimensions. A Fact table may be the result of an analysis from one or several input fact table. For example :
    - Subscriber_Session_Throughput_Fact_Daily_TB is related to a daily throughput data for a given IP session and a subscriber dimension, date and time dimension possibly with an access dimension (which is mobile network cell etc.); it can hold measures such as total packets send upstream/downstream, total packets retransmitted upstream/downstream etc. These tables are indexed by dimension keys/dimension columns etc.
    - Subscriber_Content_Download_Slow_KPI_Rollup_TB is an output fact table which stores the result of an analysis which is a KPI computed from an expression involving measures of a Total_No_Packets_Retransmitted_Downstream,Volume_Downlink etc. for a given subscriber and a given website.

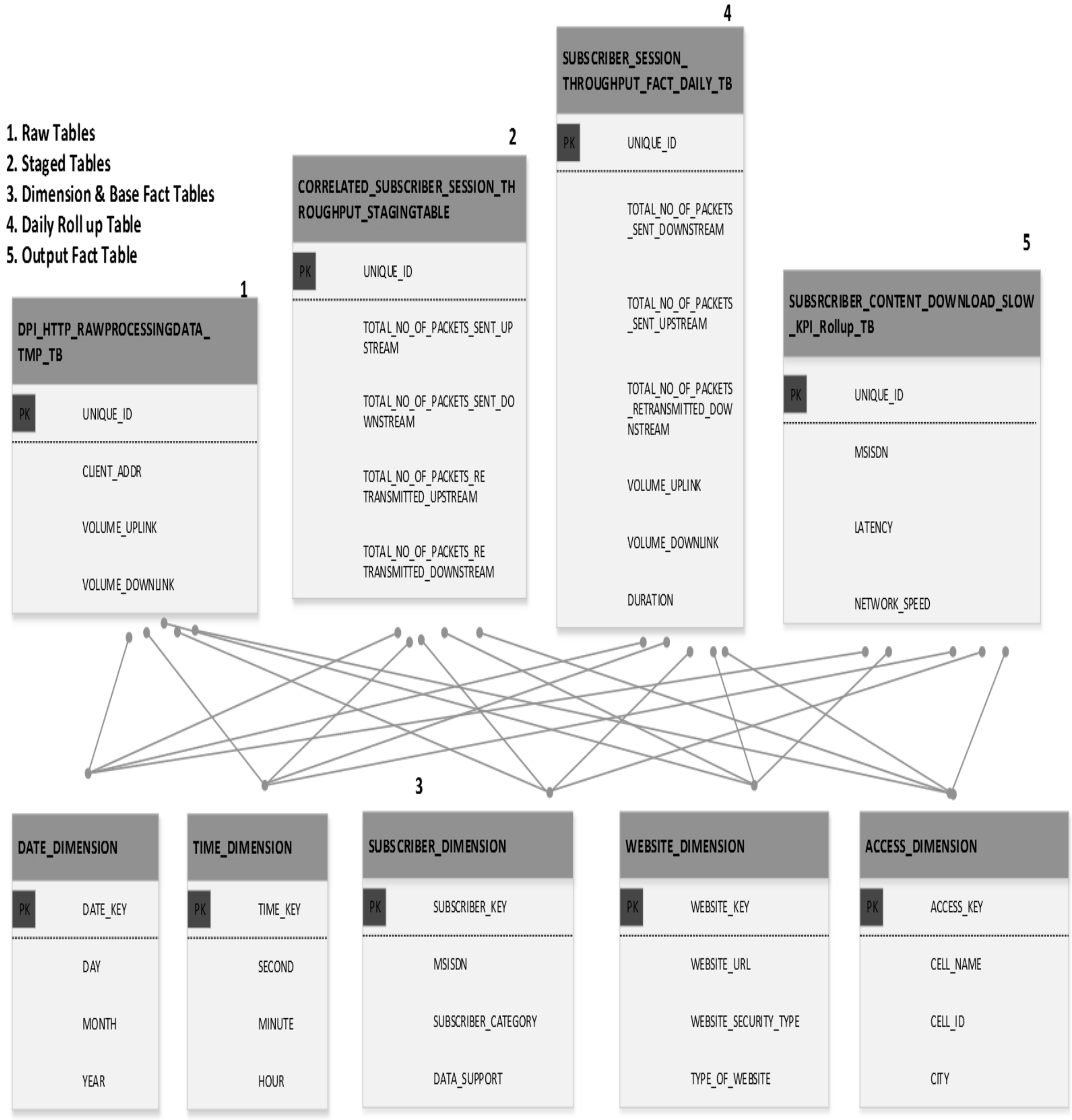


Figure 2 Snapshot of Net Neutral Value Pack data model

- Develop statistical algorithms and analytical functions to compute the mobile network's Key Performance Indicators (KPI) which would trigger a violation of Net Neutrality, if guaranteed to a user/customer/subscriber or content provider. Thus the most commonly cited issues became the basis of our initial definition of the Net Neutral KPI such as Unstable Data service connections, Slow speeds for upload/downloads, High Latency & Jitter, Incomplete page/content downloads.

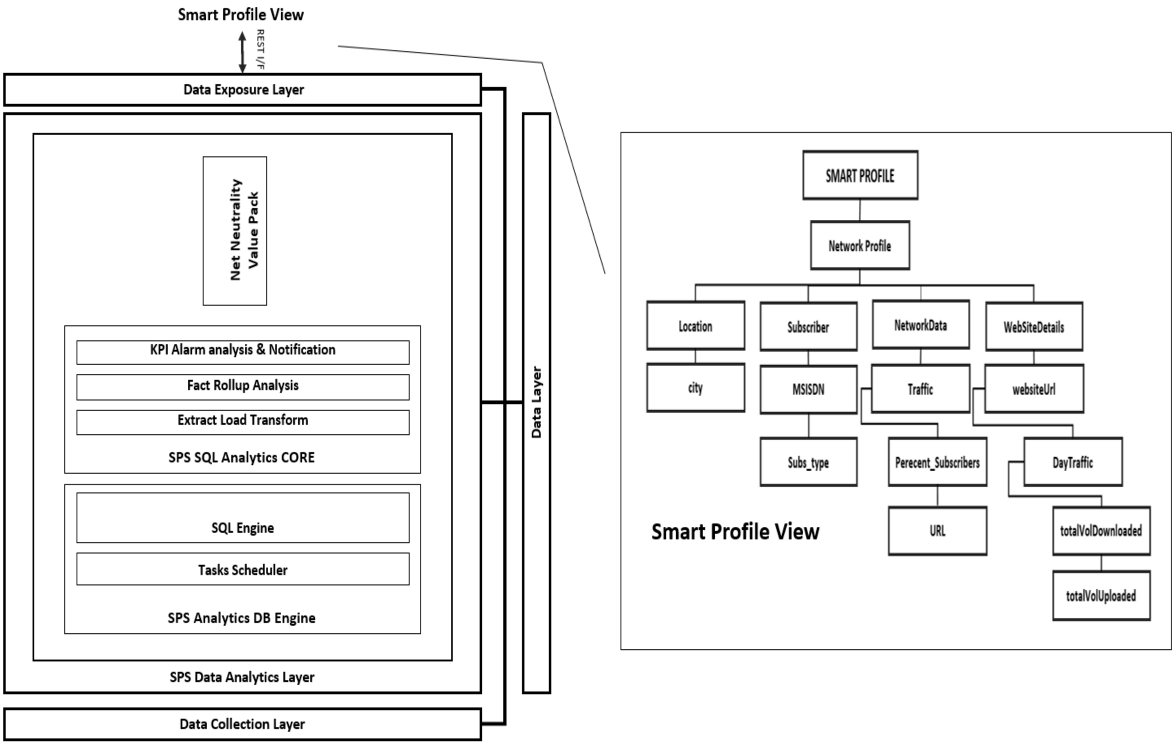


Figure 3 High Level Architecture of SPS Data Analytics Layer with Net Neutral Value pack and a sample Smart Profile View as exposed by REST interface

The high level architecture of SPS Data Analytics Layer with Net Neutrality Value Pack is as shown in Fig.3. This layer can be further divided as:

- SPS Data Analytics DB Engine consisting of an
  - SQL Engine which generates the SQL for Net Neutrality data model to fetch the analyzed result set
  - Tasks Scheduler which schedules and monitors the analysis using standard CRON[4] expression
- SPS SQL Analytics CORE uses the data from the collection layer to perform the following sequence of actions
  - Extract Load Transform – extracts the raw data from the CSV files, loads the bulk of raw collected facts into the staging tables on analytic DB and finally transforms the loaded staging tables into fact tables with appropriate dimensions.
  - Fact Rollup Analysis – either outputs the facts aggregating measures or creating measures from dimensions, from an input fact table (E.g. Subscriber_Session_Throughput_Fact_TB can be rolled up into a daily subscriber session throughput fact table where the data volume of each usage is aggregated per day/24 hours for each subscriber). Furthermore the input facts can be filtered (defined as fact views). Roll up fact table measures are called KPI (Key Performance Indicator).
  - KPI Alarm Analysis raises or clear alarms based on one fact table measure as defined in the Net Neutrality KPI algorithms.

Data Exposure Layer builds a single and secure client profile view based on the results of the Data Analysis Layer. The JDBC connector directly retrieves the analyzed results from the analytic DB. The result set is transformed using XQUERY [5] for compliance with consumer defined data model and present it as a Smart Profile View. The Smart Profile View is a hierarchical XML document which is built dynamically from the analyzed results set. Hence parts of the Smart Profile can be viewed as Resources using XPATH [6] like notation. These XPATHs form part of the REST interface definition.

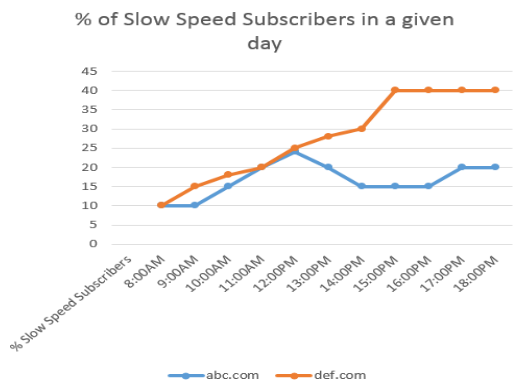
Now consider a scenario when a regulation agency wants to determine the “percentage of subscribers experiencing slow speeds when trying to download content from http://www.abc.com when compared to http://www.def.com at a given location”. Now let us define a KPI to quantify slow speed as a throughput of less than 256kbps for 5% or more of their downloads whose size is greater than 1 MB. KPI measures for “download speed”, “no.of downloads” & “download size” is calculated using statistical computation on the VOLUME_DOWNLINK and DURATION column of daily roll-up table marked as 4 in Fig.2 for the given two WEBSITE_URL at the same ACCESS_ID. The inputs to calculate these measures can be passed as parameters to the SQL. The result set of this analyzes is stored in the output fact table marked as 5 in Fig.2.

Finally the output fact table holds the % of subscribers with comparative speeds experienced at abc.com and def.com at a given time and location. The Data Exposure Layer provides a profile view of this analyzed data via a REST interface. A sample REST URL to query such KPI would be as follows:

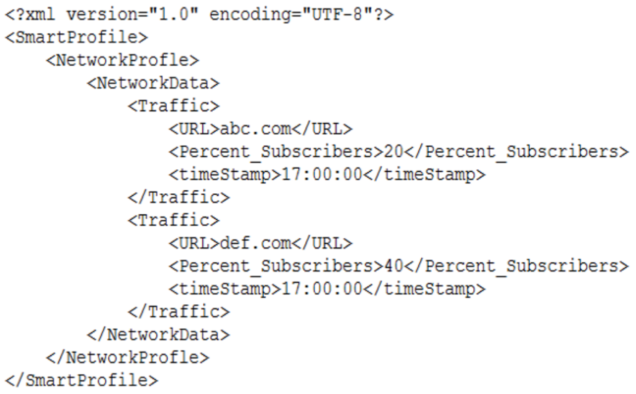
*http://localhost:8080/SPS_DEL_Rest/DSQUERY?inputDimension=/SmartProfile/NetworkProfile/Location&param=city*bangalore&inputDimension=/SmartProfile/NetworkProfile/DateDetails*&param=dateVal*(BETWEEN::2017-09-17,2017-10-18)&inputDimension=/SmartProfile/NetworkProfile/WebSiteDetails&param=websiteUrl*abc.com,websiteUrl*def.com&inputDimension=/SmartProfile/NetworkProfile/WebSiteDetails/DayTraffic&param=totalVolDownloaded*(gt 1000000)&rule=QUERY_PCL&param=OUTPUT*/SmartProfile/NetworkProfile/NetworkData/Traffic/URL&param=OUTPUT*/SmartProfile/NetworkProfile/NetworkData/Traffic/Percent_Subscribers&xpath=/SmartProfile/NetworkProfile*

Additionally placeholders such as username and password are defined to provide a fine grained access to the smart profile.

The smart profile view can be accessed based on the roles such as subscriber or a regulation agency. The output fact tables can also be integrated with any 3rd party tools for dashboard view of the data as shown in Fig.4



Dashboard View



```
<?xml version="1.0" encoding="UTF-8"?>
<SmartProfile>
    <NetworkProfle>
        <NetworkData>
            <Traffic>
                <URL>abc.com</URL>
                <Percent_Subscribers>20</Percent_Subscribers>
                <timeStamp>17:00:00</timeStamp>
            </Traffic>
            <Traffic>
                <URL>def.com</URL>
                <Percent_Subscribers>40</Percent_Subscribers>
                <timeStamp>17:00:00</timeStamp>
            </Traffic>
        </NetworkData>
    </NetworkProfle>
</SmartProfile>
```

Smart Profile View

Figure 4 Sample Results of percentage of Slow Speed Subscribers

Competitive approaches

IBM Tivoli, Nagios, CA, Cacti, MRTG, Sandvine have solutions to address either the traffic or network monitoring space but not both. Most of these solutions use MySQL, Oracle etc. which are designed for row based transactions, than for analyzing data.

There are several critical differentiators of this solution.

- Implementing Net Neutrality KPI with analytical algorithms is highly query intensive. Hence usage of columnar based analytic DB, which works on projections rather on indexes, provides significantly faster execution of SQL queries.
- The application service providers are abstracted from the know-how of underlying data models, development of complicated statistical algorithms, SQL etc.
- A near real time exposure of the analyzed results with a dashboard will enable operators to closely monitor both the customer engagement index and network management operations. The REST URL will provide the end customers/users a transparency to validate the QoS of the operator.

References

[1] Representation State Transfer (REST) - http://www.w3.org/TR/ws-arch/#relwwwrest

[2] Net Neutrality DoT Committee Report - http://www.dot.gov.in/sites/default/files/Net_Neutrality_Committee_report%20%281%29_0.pdf

[3] Star data model - http://en.wikipedia.org/wiki/Star_schema

[4] CRON - https://en.wikipedia.org/wiki/Cron

[4] XQuery - http://www.w3.org/TR/xquery/

[5] XPath - http://www.w3.org/TR/xpath/

Authors

Vidya V Bharadwaj, Prashanth G.A, M Mahesh Babu, Suhasan Chirravuri
vbharadwaj@hpe.com, prashanth.ga@hpe.com, mahesh.m5@hpe.com, suhasan.chirravuri@hpe.com