|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| khung doi**TỔNG CÔNG TY VIỄN THÔNG VIETTEL**  **------------o0o-----------**  **LogoViettel.jpg**  **SIZING HỆ THỐNG DMP GĐ 2**   |  |  |  | | --- | --- | --- | |  |  |  | | **KHỐI DIGITAL** | **KHỐI CNTT**  **Hà Nội, 07/2019** | **PHÊ DUYỆT** | |

MỤC LỤC

[1. Thông tin hệ thống 3](#_Toc13123356)

[1.1 Thông tin hệ thống 3](#_Toc13123357)

[2. GIẢI PHÁP THIẾT KẾ 4](#_Toc13123358)

[2.1 Mô hình tổng thể 4](#_Toc13123359)

[2.2 Luồng thu thập và cung cấp dữ liệu cho đối tác 5](#_Toc13123360)

[2.3 Mô hình triển khai vật lý 6](#_Toc13123361)

[3. ĐỊNH CỠ MÁY CHỦ ỨNG DỤNG DataNode 7](#_Toc13123362)

[3.1 Cơ sở định cỡ 7](#_Toc13123363)

[3.1.1 Diễn giải giá trị của tương đồng (HT test) 7](#_Toc13123364)

[3.1.2 Diễn giải giá trị định cỡ tới 2020. 8](#_Toc13123365)

[3.2 Định cỡ HDD 15](#_Toc13123366)

[3.3 Định cỡ RAM, CPU 16](#_Toc13123367)

[3.4 Đề xuất thiết bị 21](#_Toc13123368)

[3.5 Định cỡ Card GPU 21](#_Toc13123369)

[3.5.1 Cơ sở định cỡ 21](#_Toc13123370)

[3.5.2 Định cỡ GPU 23](#_Toc13123371)

[3.5.3 Đề xuất thiết bị 24](#_Toc13123372)

[3.6 Định cỡ máy chủ Cloud - DMP WORK, API 24](#_Toc13123373)

[3.6.1 Cơ sở định cỡ 24](#_Toc13123374)

[3.6.2 Định cỡ RAM, CPU, HDD 25](#_Toc13123375)

[3.7 Định cỡ Storage Cloud 26](#_Toc13123376)

[3.7.1 Tính toán SSD 26](#_Toc13123377)

[3.7.2 Tính toán HDD 26](#_Toc13123378)

[4. ĐỊNH CỠ NETWORK 27](#_Toc13123379)

[4.1 SW 10G 27](#_Toc13123380)

[4.2 SW quản trị 1G 27](#_Toc13123381)

[4.3 Rack 27](#_Toc13123382)

[5. TỔNG HỢP NHU CẦU PHẦN CỨNG 28](#_Toc13123383)

[6. Dung lượng xử lý hệ thống DMP test 30](#_Toc13123384)

[7. 10.58.244.166 32](#_Toc13123385)

[8. 10.58.244.167 33](#_Toc13123386)

[9. 10.58.244.169 34](#_Toc13123387)

1. Thông tin hệ thống
   1. Thông tin hệ thống

Hệ thống DMP (Data Management Platform) là hệ thống cung cấp dữ liệu cho quá trình tổng hợp phân tích hành vi của người dùng viễn thông, từ đó đưa ra các dự đoán về thu nhập, điểm tín dụng, đề xuất quảng cáo. Đảm bảo tính chính xác, ổn định và bảo mật của dữ liệu.

Digital đã cung cấp dịch vụ dữ liệu cho 3 đối tác từ năm 2016, cho đến nay đã bộc lộ một số bất cập:

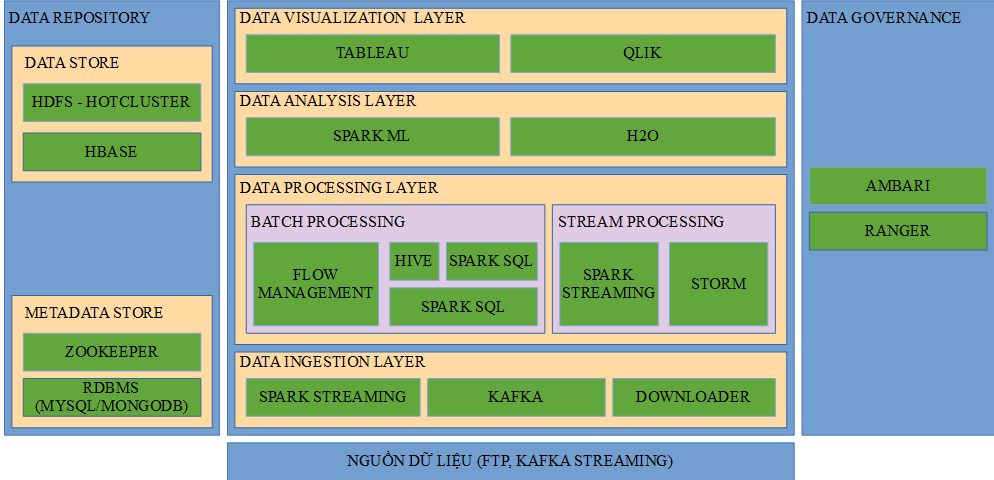
* Chưa có hệ thống phần cứng của riêng mình. Các xử lý làm sạch, mã hóa, loại bỏ dữ liệu nhậy cảm đều được thực hiện nhờ trên phần cứng của đối tác. Dẫn đến khó khăn trong việc cập nhật, bảo trì hệ thống, tính thống nhất trong việc triển khai các giải pháp thuật toán trên các hệ thống khác nhau. Do triển khai trên phần cứng của đối tác không tránh khỏi mất an toàn thông tin.
* Đáp ứng nhu cầu về sự ổn định của dữ liệu. Số đối tác ngày càng tăng lên, nguồn cung dữ liệu không đủ đáp ứng dữ liệu cho nhiều đối tác. Cần có một hệ thống chung để phân phối dữ liệu, đảm bảo tốc độ xử lý.
* Xuất phát từ nhu cầu xây dựng các bài toán phân tích dữ liệu của riêng mình, như dự đoán thu nhập, đòi hỏi phải có một hệ thống phần cứng riêng.

Từ những vấn đề nêu trên đã đặt ra yêu cầu xây dựng một hệ thống DMP theo định hướng sau:

* Thiết kế hệ thống theo định hướng dịch vụ (SOD Service Oriented Design) đảm bảo tích hợp nhanh các hệ thống ngoài và phát triển các dịch vụ mới.
* Xử lý làm sạch, mã hóa và loại bỏ dữ liệu nhậy cảm được thực hiện tập trung, thống nhất.
* Tổng hợp và phân tích dữ liệu đáp ứng nhu cầu các bài toán cụ thể.
* Nguồn dữ liệu, đảm bảo chất lượng cung cấp cho nhiều đối tác khác nhau.

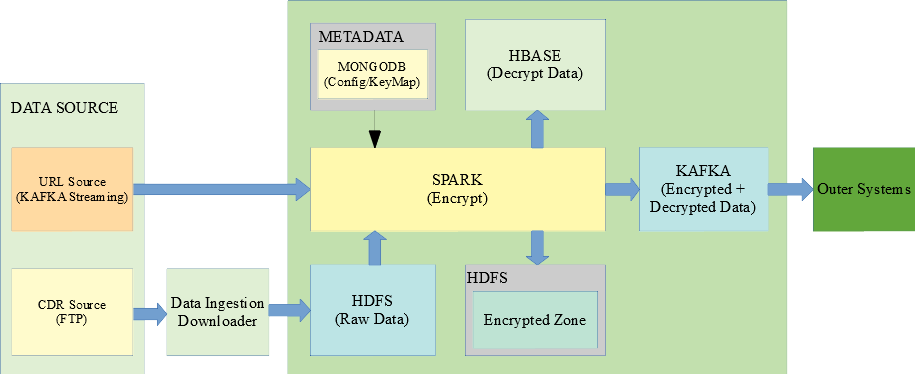
|  |  |  |
| --- | --- | --- |
| **TT** | **Hạng mục** | **Nội dung** |
| 1 | Mô tả hệ thống | * Sizing cụm Hadoop trong 6 tháng. Dựa trên các yêu cầu của hệ thống trong việc cung cấp dữ liệu cho khách hàng, một số yêu cầu đặt ra đối với mô hình thiêt kế về hệ thống * Dự kiến tăng trưởng DMP tới hết năm 2019 như sau:   + Đối tác kết nối đến hệ thống Kafka để lấy dữ liệu: 5 đối tác   + Số dự án tổng hợp và phân tích dữ liệu trực tiếp trên cụm hadoop: 2 dự án. * Dự kiến tăng trưởng giao dịch năm 2019-2020 là 12%. |
| 2 | Đầu mối / đơn vị phát triển | * TTCN – Digital – VTT |
| 3 | Thời gian  phát triển | * Hệ thống được xây dựng và triển khai mới trong năm 2019 |
| 4 | Đầu mối định cỡ | * ToND, DiemDS |
| 5 | Cơ sở định cỡ | * Định cỡ dựa trên hệ thống thử nghiệm đang chạy |
| 6 | Nguyên tắc  định cỡ | * Đảm bảo khả năng dự phòng: Các thiết bị phần cứng phải đảm bảo hoạt động với cơ chế dự phòng active-active hoặc active-standby. * Đảm bảo hiệu suất hoạt động (KPI): Tải CPU không quá 75%, RAM không vượt quá 90% và ổ cứng không vượt quá 80%. * Hệ số dự phòng sai số tính toán: 1.1 |

1. GIẢI PHÁP THIẾT KẾ
   1. Mô hình tổng thể

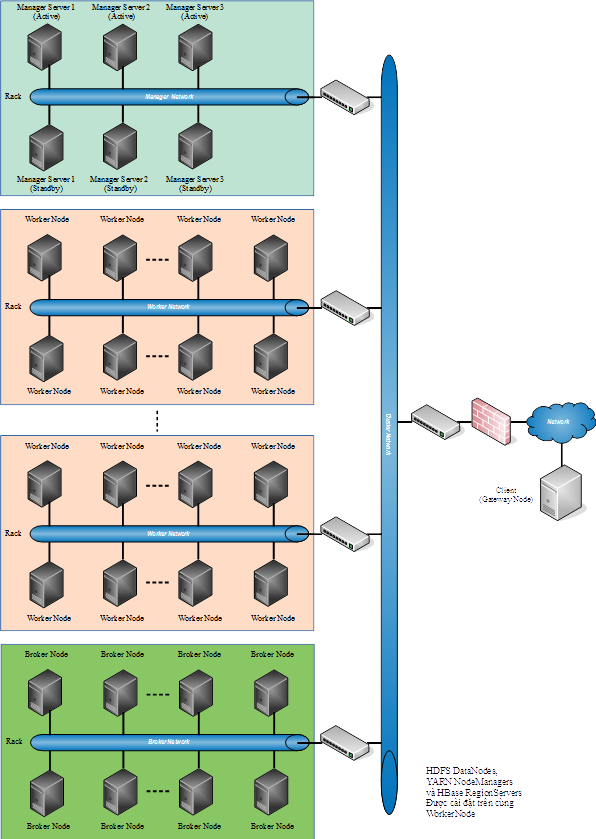


|  |  |  |
| --- | --- | --- |
| **STT** | **Tên thành phần** | **Vai trò** |
| 1. | **DATA INGESTION LAYER (Tầng thu thập dữ liệu)** | |
| 1. | SPARK STREAMING | Thu thập dữ liệu từ các nguồn CDR có dung lượng lớn |
| 2. | KAFKA | - Nguồn dữ liệu cho các Job Spark Streaming. - Chia sẻ dữ liệu với đối tác bên ngoài. |
| 3. | DOWNLOADER | Tiến trình thu thập các dữ liệu CDR được đẩy vào HDFS. |
| 2. | **DATA REPOSITORY** | |
| 1. | HDFS | Lưu dữ liệu khai thác tối đa trong 06 tháng. |
| 2. | HBASE | Lưu các dữ liệu có nhu cầu truy cập nhanh, với tải lớn. |
| 3. | ZOOKEEPER | Cung cấp các dịch vụ trong việc phối hợp các tiến trình làm việc trong môi trường phân tán. |
| 4. | SOLR/ELASTICSEARCH | Cung cấp chức năng tìm kiếm văn bản, sử dụng trong các tác vụ điều hành của Ambari. |
| 5. | RDBMS (MySql/MongoDB) | Cơ sở dữ liệu quan hệ, được sử dụng để lưu thông tin MetaData của hệ thống. |
| 3. | **DATA PROCESSING LAYER** | |
| 1. | YARN | Công cụ quản lý, phân bổ tài nguyên tính toán của hệ thống. |
| 2. | BACTH PROCESSING | Tác vụ xử lý theo lô: - Các job ETL để biến đổi dữ liệu. - Các truy vấn HiveSQL sử dụng trong quá trình phân tích dữ liệu. - Các Spark Jobs, được sử dụng cho các tính toán lớn, đòi hỏi tối ưu cao. |
| 3. | STREAMING PROCESSING | Tác vụ xử lý cần tính real time, thu thập dữ liệu URL từ nguồn KAFKA. |
| 4. | **DATA ANALYSIS LAYER** | |
| 1. | Spark ML | Bộ thư viện học máy, được thiết kế để hực hiện hiệu quả trên môi trường tính toán phân tán. Spark ML hỗ trợ hầu hết các thuật toán học máy phổ biến như: Logictis Regression, Decision Tree, SVM. |
| 2. | H2O | Nền tảng cung cấp các công cụ làm việc với các thuật toán học máy (bao gồm cả các thuật toán Deeplearning) |
| 5. | **DATA VISUALIZATION LAYER** | |
| 1. | TABLEAU | Công cụ phân tích dữ liệu và trực quan hóa. |
| 2. | BIRT REPORT | Cung cụ cung cấp các báo cáo tĩnh. |
| 6. | **DATA GOVERNANCE LAYER** | |
| 1. | RANGER | Công cụ cung cấp tính năng xác thực, phân quyền cho toàn bộ tài nguyên trong hệ thống. |
| 2. | AMBARI | Công cụ quản trị, điều hành cụm dữ liệu:  - Giám sát, cảnh báo trạng thái của tất cả các tiến trình trong hệ thống (HDFS, Yarn, Hive, Spark, Zookeeper,…)  - Thêm/Xóa/Nâng cấp các thành phần.  - Thêm/Xóa các máy chủ. |

* 1. Luồng thu thập và cung cấp dữ liệu cho đối tác



* **Bước 1:** Thu thập dữ liệu
  + Dữ liệu từ CDR source được lưu trên HDFS (Raw Data).
* **Bước 2**: Mã hóa, loại bỏ dữ liệu nhậy cảm
  + Spark Job lấy dữ liệu từ nguồn HDFS (Raw Data) và cấu hình keymap từ META-DATA, thực hiện mã hóa và loại bỏ dữ liệu nhậy cảm. Dữ liệu được mã hóa và dữ liệu giải mã được đẩy vào luồng KAFKA.
  + Dữ liệu mã hóa sẽ được lưu vào vùng ENCRYPTED\_ZONE.
  + Dữ liệu mã hóa sẽ được cung cấp cho các đối tác.
* **Bước 3**: Lưu dữ liệu giải mã.
  + Spark Job đọc dữ liệu giải mã, loại bỏ dữ liệu trùng và lưu vào HBASE.
* **Bước 4:** Tổng hợp và phân tích được thực hiện trên dữ liệu trong vùng ENCRYPTED\_ZONE.
  1. Mô hình triển khai vật lý



1. ĐỊNH CỠ MÁY CHỦ ỨNG DỤNG DataNode
   1. Cơ sở định cỡ

|  |  |  |  |
| --- | --- | --- | --- |
| **TT** | **Đầu vào** | **Giá trị định cỡ đáp ứng đến năm 2019-2020** | **Hệ thống tương đồng (HT DMP test)** |
| 1 | Tổng lưu lượng dữ liệu datamon trong ngày | **11498** GB/ngày. | **207** GB/ngày |
| 2 | Thời gian lưu trữ | 6 tháng   * Do nhu cầu của đối tác, cần lấy dữ liệu lịch sử tối thiểu 6 tháng cho việc tổng hợp phân tích. |  |
| 3 | Yêu cầu backup dữ liệu | 1 lần / ngày   * Dữ liệu được tổng hợp hàng ngày nên cần backup trong trường hợp hệ thống bigdata gặp sự cố. |  |

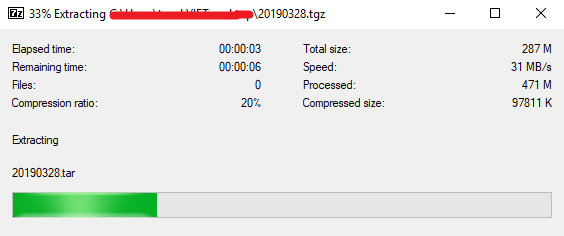
* + 1. Diễn giải giá trị của tương đồng (HT test)
* Các nguồn dữ liệu được xử lý:
  + Dữ liệu từ hệ thống BCCS
* Subscriber

(Dữ liệu subscriber ~100MB/Ngày)

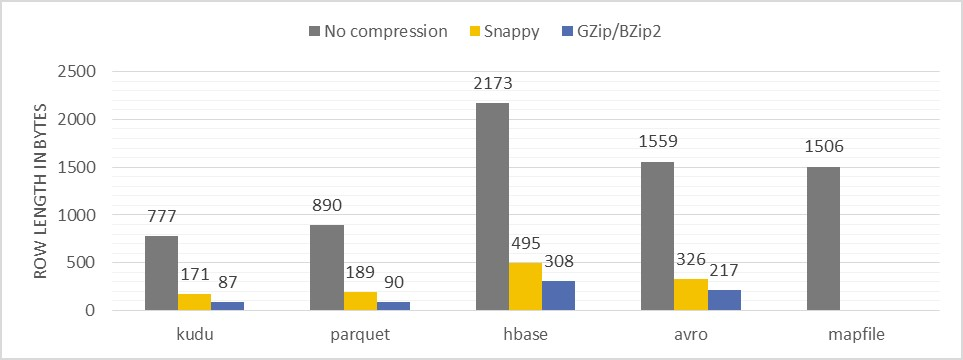
* + Dữ liệu từ hệ thống vOCS
    - VAS\_MPS\_CHARGE
    - CDR\_MSC\_VOICE
    - CDR\_MSC\_SMS
    - REG\_CDR\_21H
    - REG\_CDR\_09H
    - reportIN
    - DailyFeeCdr
    - export\_acct\_book
    - G22
    - G33
    - MOBILE\_mb\_postpaid\_call
    - MOBILE\_mb\_monthly\_fee
    - LoanCdr
    - RechargeList

***Chi tiết kích thước các bảng dữ liệu hàng ngày đang xử lí theo Phụ lục đính kèm.***

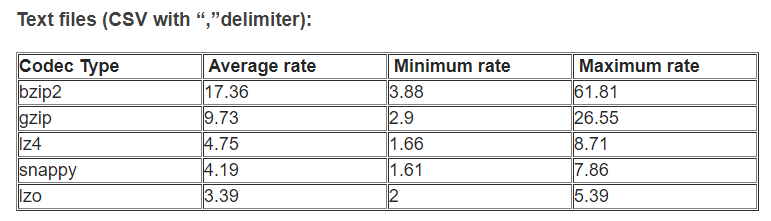
* Tỷ lệ nén lưu trữ: 5 lần.



* Dung lượng dữ liệu xử lý trong ngày: **42395.592** \* 5 = 211977.96MB ~ 207GB.
* Dữ liệu sau khi thu thập từ các nguồn dữ liệu được tổng hợp thành 1 file lớn theo nhóm dữ liệu, để tối ưu lưu trữ và xử lý trên HDFS. Tỉ lệ nến lưu trữ file này là 5 lần (20%). Nên lượng dữ liệu thực tế là **42395.592** \* **5**.
  + 1. Diễn giải giá trị định cỡ tới 2020.
* Dữ liệu nguồn lưu dạng parquet snappy. Tỷ lệ nén 5 lần.



(Tham khảo: <https://db-blog.web.cern.ch/blog/zbigniew-baranowski/2017-01-performance-comparison-different-file-formats-and-storage-engines>)

(Tham khảo <https://blogs.oracle.com/datawarehousing/hadoop-compression-compression-rate-part1>)

* Lưu lượng dữ liệu trung bình theo ngày:

|  |  |  |  |
| --- | --- | --- | --- |
| **Dữ liệu** | **Volume** | **Unit** | **Path** |
| call\_2g\_record | 14.5 | G | /call\_2g\_record/2019/06/01 |
| call\_2g\_record | 14.4 | G | /call\_2g\_record/2019/06/02 |
| call\_2g\_record | 16 | G | /call\_2g\_record/2019/06/03 |
| call\_2g\_record | 18 | G | /call\_2g\_record/2019/06/04 |
| call\_2g\_record | 16.4 | G | /call\_2g\_record/2019/06/05 |
| call\_2g\_record | 13.5 | G | /call\_2g\_record/2019/06/06 |
| call\_2g\_record | 17.5 | G | /call\_2g\_record/2019/06/07 |
| call\_2g\_record | 15.9 | G | /call\_2g\_record/2019/06/08 |
| call\_2g\_record | 16.1 | G | /call\_2g\_record/2019/06/09 |
| call\_2g\_record | 18.4 | G | /call\_2g\_record/2019/06/10 |
| call\_2g\_record | 18.6 | G | /call\_2g\_record/2019/06/11 |
| call\_2g\_record | 18.5 | G | /call\_2g\_record/2019/06/12 |
| call\_2g\_record | 18.4 | G | /call\_2g\_record/2019/06/13 |
| call\_2g\_record | 18.4 | G | /call\_2g\_record/2019/06/14 |
| call\_2g\_record | 17.2 | G | /call\_2g\_record/2019/06/15 |
| call\_2g\_record | 16.8 | G | /call\_2g\_record/2019/06/16 |
| call\_2g\_record | 18.8 | G | /call\_2g\_record/2019/06/17 |
| call\_2g\_record | 18.8 | G | /call\_2g\_record/2019/06/18 |
| call\_2g\_record | 18.4 | G | /call\_2g\_record/2019/06/19 |
| call\_2g\_record | 18.8 | G | /call\_2g\_record/2019/06/20 |
| call\_2g\_record | 18.8 | G | /call\_2g\_record/2019/06/21 |
| call\_2g\_record | 16.9 | G | /call\_2g\_record/2019/06/22 |
| call\_2g\_record | 16.9 | G | /call\_2g\_record/2019/06/23 |
| call\_2g\_record | 18.9 | G | /call\_2g\_record/2019/06/24 |
| cs\_2g\_prod | 27.7 | G | /cs\_2g\_prod/2019/06/01 |
| cs\_2g\_prod | 27.5 | G | /cs\_2g\_prod/2019/06/02 |
| cs\_2g\_prod | 28.6 | G | /cs\_2g\_prod/2019/06/03 |
| cs\_2g\_prod | 31.5 | G | /cs\_2g\_prod/2019/06/04 |
| cs\_2g\_prod | 29.6 | G | /cs\_2g\_prod/2019/06/05 |
| cs\_2g\_prod | 27.5 | G | /cs\_2g\_prod/2019/06/06 |
| cs\_2g\_prod | 32.9 | G | /cs\_2g\_prod/2019/06/07 |
| cs\_2g\_prod | 32.4 | G | /cs\_2g\_prod/2019/06/08 |
| cs\_2g\_prod | 32 | G | /cs\_2g\_prod/2019/06/09 |
| cs\_2g\_prod | 32.8 | G | /cs\_2g\_prod/2019/06/10 |
| cs\_2g\_prod | 32.2 | G | /cs\_2g\_prod/2019/06/11 |
| cs\_2g\_prod | 31.2 | G | /cs\_2g\_prod/2019/06/12 |
| cs\_2g\_prod | 32.1 | G | /cs\_2g\_prod/2019/06/13 |
| cs\_2g\_prod | 32.7 | G | /cs\_2g\_prod/2019/06/14 |
| cs\_2g\_prod | 32.7 | G | /cs\_2g\_prod/2019/06/15 |
| cs\_2g\_prod | 32.1 | G | /cs\_2g\_prod/2019/06/16 |
| cs\_2g\_prod | 32.3 | G | /cs\_2g\_prod/2019/06/17 |
| cs\_2g\_prod | 34.3 | G | /cs\_2g\_prod/2019/06/18 |
| cs\_2g\_prod | 34 | G | /cs\_2g\_prod/2019/06/19 |
| cs\_2g\_prod | 33.1 | G | /cs\_2g\_prod/2019/06/20 |
| cs\_2g\_prod | 33.1 | G | /cs\_2g\_prod/2019/06/21 |
| cs\_2g\_prod | 32.4 | G | /cs\_2g\_prod/2019/06/22 |
| cs\_2g\_prod | 32.8 | G | /cs\_2g\_prod/2019/06/23 |
| cs\_2g\_prod | 32.4 | G | /cs\_2g\_prod/2019/06/24 |
| cs\_3g\_record | 34.3 | G | /cs\_3g\_record/2019/06/01 |
| cs\_3g\_record | 33.9 | G | /cs\_3g\_record/2019/06/02 |
| cs\_3g\_record | 36.9 | G | /cs\_3g\_record/2019/06/03 |
| cs\_3g\_record | 10.7 | G | /cs\_3g\_record/2019/06/04 |
| cs\_3g\_record | 14.5 | G | /cs\_3g\_record/2019/06/05 |
| cs\_3g\_record | 31.9 | G | /cs\_3g\_record/2019/06/06 |
| cs\_3g\_record | 38.9 | G | /cs\_3g\_record/2019/06/07 |
| cs\_3g\_record | 36.4 | G | /cs\_3g\_record/2019/06/08 |
| cs\_3g\_record | 34.2 | G | /cs\_3g\_record/2019/06/09 |
| cs\_3g\_record | 41.3 | G | /cs\_3g\_record/2019/06/10 |
| cs\_3g\_record | 41 | G | /cs\_3g\_record/2019/06/11 |
| cs\_3g\_record | 40.4 | G | /cs\_3g\_record/2019/06/12 |
| cs\_3g\_record | 40.5 | G | /cs\_3g\_record/2019/06/13 |
| cs\_3g\_record | 41.5 | G | /cs\_3g\_record/2019/06/14 |
| cs\_3g\_record | 38.4 | G | /cs\_3g\_record/2019/06/15 |
| cs\_3g\_record | 35.4 | G | /cs\_3g\_record/2019/06/16 |
| cs\_3g\_record | 41 | G | /cs\_3g\_record/2019/06/17 |
| cs\_3g\_record | 40.6 | G | /cs\_3g\_record/2019/06/18 |
| cs\_3g\_record | 40.3 | G | /cs\_3g\_record/2019/06/19 |
| cs\_3g\_record | 40.6 | G | /cs\_3g\_record/2019/06/20 |
| cs\_3g\_record | 40.7 | G | /cs\_3g\_record/2019/06/21 |
| cs\_3g\_record | 37.1 | G | /cs\_3g\_record/2019/06/22 |
| cs\_3g\_record | 35.1 | G | /cs\_3g\_record/2019/06/23 |
| cs\_3g\_record | 40.5 | G | /cs\_3g\_record/2019/06/24 |
| http | 108.5 | G | /http/2019/06/01 |
| http | 110.3 | G | /http/2019/06/02 |
| http | 108.9 | G | /http/2019/06/03 |
| http | 106.8 | G | /http/2019/06/04 |
| http | 105.1 | G | /http/2019/06/05 |
| http | 105.7 | G | /http/2019/06/06 |
| http | 106.7 | G | /http/2019/06/07 |
| http | 108.3 | G | /http/2019/06/08 |
| http | 107.7 | G | /http/2019/06/09 |
| http | 105.1 | G | /http/2019/06/10 |
| http | 106.5 | G | /http/2019/06/11 |
| http | 105.4 | G | /http/2019/06/12 |
| http | 104.9 | G | /http/2019/06/13 |
| http | 105.6 | G | /http/2019/06/14 |
| http | 106.2 | G | /http/2019/06/15 |
| http | 103.2 | G | /http/2019/06/16 |
| http | 105.3 | G | /http/2019/06/17 |
| http | 105.8 | G | /http/2019/06/18 |
| http | 104.2 | G | /http/2019/06/19 |
| http | 104.9 | G | /http/2019/06/20 |
| http | 105.8 | G | /http/2019/06/21 |
| http | 103.1 | G | /http/2019/06/22 |
| http | 103 | G | /http/2019/06/23 |
| http | 102.8 | G | /http/2019/06/24 |
| http\_4g | 145.3 | G | /http\_4g/2019/06/01 |
| http\_4g | 186.8 | G | /http\_4g/2019/06/02 |
| http\_4g | 186.2 | G | /http\_4g/2019/06/03 |
| http\_4g | 184.7 | G | /http\_4g/2019/06/04 |
| http\_4g | 173.2 | G | /http\_4g/2019/06/05 |
| http\_4g | 175.2 | G | /http\_4g/2019/06/06 |
| http\_4g | 178.1 | G | /http\_4g/2019/06/07 |
| http\_4g | 180.2 | G | /http\_4g/2019/06/08 |
| http\_4g | 179.9 | G | /http\_4g/2019/06/09 |
| http\_4g | 180.8 | G | /http\_4g/2019/06/10 |
| http\_4g | 182 | G | /http\_4g/2019/06/11 |
| http\_4g | 181.9 | G | /http\_4g/2019/06/12 |
| http\_4g | 180.7 | G | /http\_4g/2019/06/13 |
| http\_4g | 179.8 | G | /http\_4g/2019/06/14 |
| http\_4g | 180.3 | G | /http\_4g/2019/06/15 |
| http\_4g | 177.6 | G | /http\_4g/2019/06/16 |
| http\_4g | 180.5 | G | /http\_4g/2019/06/17 |
| http\_4g | 181.4 | G | /http\_4g/2019/06/18 |
| http\_4g | 181.4 | G | /http\_4g/2019/06/19 |
| http\_4g | 183.3 | G | /http\_4g/2019/06/20 |
| http\_4g | 183 | G | /http\_4g/2019/06/21 |
| http\_4g | 178.4 | G | /http\_4g/2019/06/22 |
| http\_4g | 165.3 | G | /http\_4g/2019/06/23 |
| http\_4g | 175.6 | G | /http\_4g/2019/06/24 |
| nbap\_channel\_flow | 263.4 | G | /nbap\_channel\_flow/2019/06/01 |
| nbap\_channel\_flow | 257 | G | /nbap\_channel\_flow/2019/06/02 |
| nbap\_channel\_flow | 261 | G | /nbap\_channel\_flow/2019/06/03 |
| nbap\_channel\_flow | 262.8 | G | /nbap\_channel\_flow/2019/06/04 |
| nbap\_channel\_flow | 262.5 | G | /nbap\_channel\_flow/2019/06/05 |
| nbap\_channel\_flow | 264.7 | G | /nbap\_channel\_flow/2019/06/06 |
| nbap\_channel\_flow | 270.3 | G | /nbap\_channel\_flow/2019/06/07 |
| nbap\_channel\_flow | 267.1 | G | /nbap\_channel\_flow/2019/06/08 |
| nbap\_channel\_flow | 256.9 | G | /nbap\_channel\_flow/2019/06/09 |
| nbap\_channel\_flow | 271 | G | /nbap\_channel\_flow/2019/06/10 |
| nbap\_channel\_flow | 270.4 | G | /nbap\_channel\_flow/2019/06/11 |
| nbap\_channel\_flow | 271.6 | G | /nbap\_channel\_flow/2019/06/12 |
| nbap\_channel\_flow | 267.1 | G | /nbap\_channel\_flow/2019/06/13 |
| nbap\_channel\_flow | 274.6 | G | /nbap\_channel\_flow/2019/06/14 |
| nbap\_channel\_flow | 275.4 | G | /nbap\_channel\_flow/2019/06/15 |
| nbap\_channel\_flow | 259.6 | G | /nbap\_channel\_flow/2019/06/16 |
| nbap\_channel\_flow | 274.9 | G | /nbap\_channel\_flow/2019/06/17 |
| nbap\_channel\_flow | 276.5 | G | /nbap\_channel\_flow/2019/06/18 |
| nbap\_channel\_flow | 275 | G | /nbap\_channel\_flow/2019/06/19 |
| nbap\_channel\_flow | 276.7 | G | /nbap\_channel\_flow/2019/06/20 |
| nbap\_channel\_flow | 274.7 | G | /nbap\_channel\_flow/2019/06/21 |
| nbap\_channel\_flow | 269.7 | G | /nbap\_channel\_flow/2019/06/22 |
| nbap\_channel\_flow | 264.6 | G | /nbap\_channel\_flow/2019/06/23 |
| nbap\_channel\_flow | 268.7 | G | /nbap\_channel\_flow/2019/06/24 |
| s6a | 289.2 | G | /s6a/2019/06/01 |
| s6a | 297.9 | G | /s6a/2019/06/02 |
| s6a | 302.4 | G | /s6a/2019/06/03 |
| s6a | 310.1 | G | /s6a/2019/06/04 |
| s6a | 310.1 | G | /s6a/2019/06/05 |
| s6a | 310.1 | G | /s6a/2019/06/06 |
| s6a | 313 | G | /s6a/2019/06/07 |
| s6a | 317.8 | G | /s6a/2019/06/08 |
| s6a | 304.6 | G | /s6a/2019/06/09 |
| s6a | 322.4 | G | /s6a/2019/06/10 |
| s6a | 324.6 | G | /s6a/2019/06/11 |
| s6a | 325.7 | G | /s6a/2019/06/12 |
| s6a | 328.4 | G | /s6a/2019/06/13 |
| s6a | 334 | G | /s6a/2019/06/14 |
| s6a | 344.7 | G | /s6a/2019/06/15 |
| s6a | 330.8 | G | /s6a/2019/06/16 |
| s6a | 349.4 | G | /s6a/2019/06/17 |
| s6a | 351.4 | G | /s6a/2019/06/18 |
| s6a | 278.2 | G | /s6a/2019/06/19 |
| s6a | 353.8 | G | /s6a/2019/06/20 |
| s6a | 335.4 | G | /s6a/2019/06/21 |
| s6a | 321.1 | G | /s6a/2019/06/22 |
| s6a | 259.2 | G | /s6a/2019/06/23 |
| s6a | 329.2 | G | /s6a/2019/06/24 |
| uex\_news | 433 | G | /uex\_news/2019/06/01 |
| uex\_news | 459.9 | G | /uex\_news/2019/06/02 |
| uex\_news | 466.2 | G | /uex\_news/2019/06/03 |
| uex\_news | 427.4 | G | /uex\_news/2019/06/04 |
| uex\_news | 432.5 | G | /uex\_news/2019/06/05 |
| uex\_news | 437.8 | G | /uex\_news/2019/06/06 |
| uex\_news | 409.6 | G | /uex\_news/2019/06/07 |
| uex\_news | 412.9 | G | /uex\_news/2019/06/08 |
| uex\_news | 405.8 | G | /uex\_news/2019/06/09 |
| uex\_news | 431.5 | G | /uex\_news/2019/06/10 |
| uex\_news | 456.1 | G | /uex\_news/2019/06/11 |
| uex\_news | 462.8 | G | /uex\_news/2019/06/12 |
| uex\_news | 459.7 | G | /uex\_news/2019/06/13 |
| uex\_news | 458.6 | G | /uex\_news/2019/06/14 |
| uex\_news | 432.5 | G | /uex\_news/2019/06/15 |
| uex\_news | 427.2 | G | /uex\_news/2019/06/16 |
| uex\_news | 443.1 | G | /uex\_news/2019/06/17 |
| uex\_news | 435 | G | /uex\_news/2019/06/18 |
| uex\_news | 422.9 | G | /uex\_news/2019/06/19 |
| uex\_news | 409.7 | G | /uex\_news/2019/06/20 |
| uex\_news | 418.1 | G | /uex\_news/2019/06/21 |
| uex\_news | 397.8 | G | /uex\_news/2019/06/22 |
| uex\_news | 399.3 | G | /uex\_news/2019/06/23 |
| uex\_news | 404.3 | G | /uex\_news/2019/06/24 |
| uex\_news\_4g | 428.7 | G | /uex\_news\_4g/2019/06/01 |
| uex\_news\_4g | 570.6 | G | /uex\_news\_4g/2019/06/02 |
| uex\_news\_4g | 591.3 | G | /uex\_news\_4g/2019/06/03 |
| uex\_news\_4g | 556.4 | G | /uex\_news\_4g/2019/06/04 |
| uex\_news\_4g | 540.5 | G | /uex\_news\_4g/2019/06/05 |
| uex\_news\_4g | 567 | G | /uex\_news\_4g/2019/06/06 |
| uex\_news\_4g | 542.5 | G | /uex\_news\_4g/2019/06/07 |
| uex\_news\_4g | 585.5 | G | /uex\_news\_4g/2019/06/08 |
| uex\_news\_4g | 577.4 | G | /uex\_news\_4g/2019/06/09 |
| uex\_news\_4g | 645.8 | G | /uex\_news\_4g/2019/06/10 |
| uex\_news\_4g | 603.3 | G | /uex\_news\_4g/2019/06/11 |
| uex\_news\_4g | 647.6 | G | /uex\_news\_4g/2019/06/12 |
| uex\_news\_4g | 634.7 | G | /uex\_news\_4g/2019/06/13 |
| uex\_news\_4g | 627.5 | G | /uex\_news\_4g/2019/06/14 |
| uex\_news\_4g | 568.9 | G | /uex\_news\_4g/2019/06/15 |
| uex\_news\_4g | 550.8 | G | /uex\_news\_4g/2019/06/16 |
| uex\_news\_4g | 621.2 | G | /uex\_news\_4g/2019/06/17 |
| uex\_news\_4g | 578.5 | G | /uex\_news\_4g/2019/06/18 |
| uex\_news\_4g | 562.1 | G | /uex\_news\_4g/2019/06/19 |
| uex\_news\_4g | 539.5 | G | /uex\_news\_4g/2019/06/20 |
| uex\_news\_4g | 550.8 | G | /uex\_news\_4g/2019/06/21 |
| uex\_news\_4g | 521.2 | G | /uex\_news\_4g/2019/06/22 |
| uex\_news\_4g | 501.8 | G | /uex\_news\_4g/2019/06/23 |
| uex\_news\_4g | 522.1 | G | /uex\_news\_4g/2019/06/24 |
| uex\_video | 33.2 | G | /uex\_video/2019/06/01 |
| uex\_video | 35.3 | G | /uex\_video/2019/06/02 |
| uex\_video | 32.1 | G | /uex\_video/2019/06/03 |
| uex\_video | 30.9 | G | /uex\_video/2019/06/04 |
| uex\_video | 32.5 | G | /uex\_video/2019/06/05 |
| uex\_video | 32.5 | G | /uex\_video/2019/06/06 |
| uex\_video | 33 | G | /uex\_video/2019/06/07 |
| uex\_video | 35.3 | G | /uex\_video/2019/06/08 |
| uex\_video | 35.3 | G | /uex\_video/2019/06/09 |
| uex\_video | 31.4 | G | /uex\_video/2019/06/10 |
| uex\_video | 32.5 | G | /uex\_video/2019/06/11 |
| uex\_video | 32.6 | G | /uex\_video/2019/06/12 |
| uex\_video | 32.6 | G | /uex\_video/2019/06/13 |
| uex\_video | 33.3 | G | /uex\_video/2019/06/14 |
| uex\_video | 34 | G | /uex\_video/2019/06/15 |
| uex\_video | 34.3 | G | /uex\_video/2019/06/16 |
| uex\_video | 33.4 | G | /uex\_video/2019/06/17 |
| uex\_video | 33.2 | G | /uex\_video/2019/06/18 |
| uex\_video | 32 | G | /uex\_video/2019/06/19 |
| uex\_video | 33 | G | /uex\_video/2019/06/20 |
| uex\_video | 33.1 | G | /uex\_video/2019/06/21 |
| uex\_video | 33.2 | G | /uex\_video/2019/06/22 |
| uex\_video | 34.3 | G | /uex\_video/2019/06/23 |
| uex\_video | 32.8 | G | /uex\_video/2019/06/24 |
| uex\_video\_4g | 44.9 | G | /uex\_video\_4g/2019/06/01 |
| uex\_video\_4g | 65.5 | G | /uex\_video\_4g/2019/06/02 |
| uex\_video\_4g | 60.8 | G | /uex\_video\_4g/2019/06/03 |
| uex\_video\_4g | 59.9 | G | /uex\_video\_4g/2019/06/04 |
| uex\_video\_4g | 63 | G | /uex\_video\_4g/2019/06/05 |
| uex\_video\_4g | 62 | G | /uex\_video\_4g/2019/06/06 |
| uex\_video\_4g | 63.7 | G | /uex\_video\_4g/2019/06/07 |
| uex\_video\_4g | 68.3 | G | /uex\_video\_4g/2019/06/08 |
| uex\_video\_4g | 69.3 | G | /uex\_video\_4g/2019/06/09 |
| uex\_video\_4g | 62.1 | G | /uex\_video\_4g/2019/06/10 |
| uex\_video\_4g | 63.2 | G | /uex\_video\_4g/2019/06/11 |
| uex\_video\_4g | 62.8 | G | /uex\_video\_4g/2019/06/12 |
| uex\_video\_4g | 63.4 | G | /uex\_video\_4g/2019/06/13 |
| uex\_video\_4g | 66 | G | /uex\_video\_4g/2019/06/14 |
| uex\_video\_4g | 67.5 | G | /uex\_video\_4g/2019/06/15 |
| uex\_video\_4g | 69 | G | /uex\_video\_4g/2019/06/16 |
| uex\_video\_4g | 65.4 | G | /uex\_video\_4g/2019/06/17 |
| uex\_video\_4g | 65.4 | G | /uex\_video\_4g/2019/06/18 |
| uex\_video\_4g | 65.1 | G | /uex\_video\_4g/2019/06/19 |
| uex\_video\_4g | 66.3 | G | /uex\_video\_4g/2019/06/20 |
| uex\_video\_4g | 66.1 | G | /uex\_video\_4g/2019/06/21 |
| uex\_video\_4g | 65.8 | G | /uex\_video\_4g/2019/06/22 |
| uex\_video\_4g | 62.5 | G | /uex\_video\_4g/2019/06/23 |
| uex\_video\_4g | 64.4 | G | /uex\_video\_4g/2019/06/24 |

***Dữ liệu xử lý trung bình ngày:***

|  |  |
| --- | --- |
| **Dữ liệu** | **Average of Volume** |
| call\_2g\_record | 17.2875 |
| cs\_2g\_prod | 31.6625 |
| cs\_3g\_record | 36.0875 |
| http | 105.825 |
| http\_4g | 178.4 |
| nbap\_channel\_flow | 268.175 |
| s6a | 318.4791667 |
| uex\_news | 430.9875 |
| uex\_news\_4g | 568.1541667 |
| uex\_video | 33.15833333 |
| uex\_video\_4g | 63.85 |
| **Total** | **2053** |

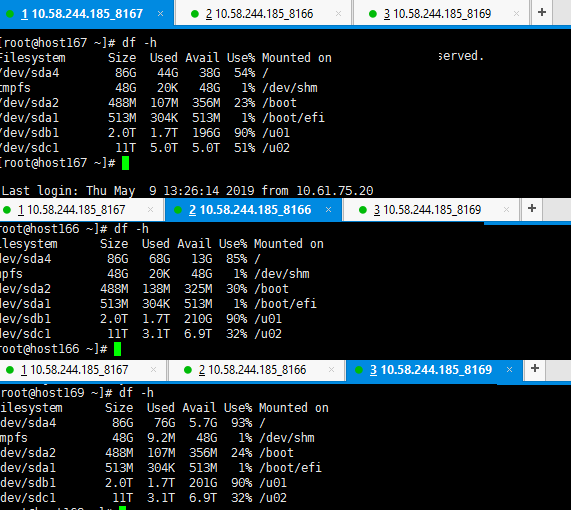
* Tăng trưởng dữ liệu 2019-2020 là 12%/năm.
* Lượng dữ liệu/ngày để định cỡ là: Total \* Hệ số nén \* Hệ số tăng trưởng = **2053 \* 5 \* (1+0.12) = 11496.8 GB/Ngày**.
  1. Định cỡ HDD

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **TT** | **Thông số** | **Thông số** | **Ext** | **Giá trị** | **Ghi chú** |
| 1 |  | **Tổng dữ liệu đến trong ngày** | **T** | **11,496.8** | Trung bình lượng dữ liệu datamon cần xử lý trong ngày (GB). |
| 2 |  | HDFS replication factor | HRF | 3 | Hệ số lưu dự phòng trên hệ thống. Tham khảo: https://www.quora.com/Is-replication-factor-of-3-a-standard-for-HDFS |
| 3 |  | Average compression factor | ACF | 5.00 | Hệ số nén dữ liệu csv dưới dạng parquet snappy trung bình 5 lần. |
| 4 | **Kích thước dữ liệu mỗi ngày (đã nén)** | **Daily volume (đã nén)** | **DVOL** | **6,898** | **DVOL = T \* HRF / ACF.** |
| 5 |  | Calculate Safety Rating | CSR | 1.10 | Hệ số dự phòng sai số tính toán. Theo guideline hướng dẫn của tập đoàn. |
| CSD | 0.80 | Hệ số KPI 80% lưu trữ trong HDD. Theo guideline hướng dẫn của tập đoàn. |
| 6 |  | Retention Month (N): | R | 6 | Số tháng lưu trữ dữ liệu. Theo nhu cầu sản xuất, kinh doanh. |
| 7 |  | Monthly volume: | MVOL | 206,942 | MVOL = (DVOL \* 30) |
| 8 | **Dung lượng lưu trữ sau 6 tháng** |  | **ANM** | **1,241,654** | **ANM = MVOL \* R** |
| 9 |  | Intermediate MapReduce data factor | IMDF | 18% | Dữ liệu MapReduce trung gian. Tham khảo: <https://www.cloudera.com/documentation/other/reference-architecture/PDF/cloudera_ref_arch_metal.pdf>  “Cloudera recommends allocating between 10-25% of the raw disk space for temporary storage as a general guideline.” |
| 10 |  | Non HDFS reserved space factor | NHDFS | 14% | - Log files - Mapreduce shuffle output - Local copies of data files |
| 11 | **Tổng dung lượng trống và xử lý** |  | **TOTAL** | **2,510,698** | **AHS = (ANM \* CSR) / ((1 – NHDFS – IMDF) \* CSD).** |

* 1. Định cỡ RAM, CPU
* Lấy hệ thống hiện tại đang chạy là hệ thống DataNode Test để làm sở cứ tính toán. Hệ thống chạy phân tải đều trên 3 server. Trung bình mỗi ngày chạy 207 GB
* Cấu hình máy chủ **DMP Test** hiện tại:

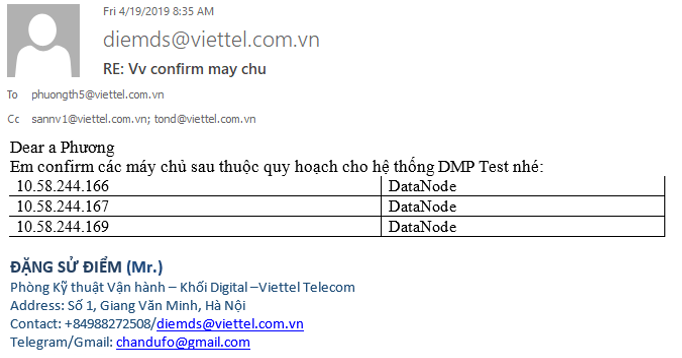
|  |  |
| --- | --- |
| 10.58.244.166 |  |
| 10.58.244.167 |  |
| 10.58.244.169 |  |

Ảnh chụp ổ cứng:



| **TT** | **IP** | **Tên máy** | **CPU** | **RAM** | **HDD** | **Cint rate 2006** | **Link tham chiếu** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 10.58.244.166 | DataNode | Intel E5 2630 v4 | 96 GB | 13.13 TB | 842 | https://www.spec.org/cpu2006/results/res2016q2/cpu2006-20160503-40862.html |
| 2 | 10.58.244.167 | DataNode | Intel E5 2630 v4 | 96 GB | 13.13 TB | 842 | Nt |
| 3 | 10.58.244.169 | DataNode | Intel E5 2630 v4 | 96 GB | 13.13 TB | 842 | Nt |

Sở cứ quy hoạch máy DMP test:



* Tải máy chủ Node hiện tại:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **TT** | **IP** | **Tên ứng dụng** | **SL giao dịch** | **CINT** | **RAM** | **Tải CPU 95** | **Tải RAM 95** | **Cint used** | **RAM used** |
| 1 | 10.58.244.166 | Big Data Cloud | 207GB | 842 | 96 | 18% | 67% | 151.56 | 64.32 |
| 2 | 10.58.244.167 | Big Data Cloud | 207GB | 842 | 96 | 29% | 34% | 244.18 | 32.64 |
| 3 | 10.58.244.169 | Big Data Cloud | 207GB | 842 | 96 | 25% | 85% | 210.5 | 81.6 |
|  |  | **TỔNG** | 207GB |  |  |  |  | 606.24 | 178.56 |

* Tính toán nhu cầu tài nguyên cho CPU/RAM:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **TT** | **Thông số** | **Ex** | **Máy chủ** | **Ghi chú** |
| 1 | CINT hiện tại | A | **686.16** |  |
| 2 | RAM (GB) hiện tại | B | **180.48** |  |
| 3 | Giá trị hiện tại (GB) | C | **207** |  |
| 4 | Cint/1GB giao dịch | D | 3.3148 | =A/C |
| 5 | RAM/1GB giao dịch | E | 0.8719 | =B/C |
| 6 | Giá trị dữ liệu định cỡ (GB) | F | **11,496.8** |  |
| 7 | CINT cần cho giá trị định cỡ | G | 38,109.39 | =DxF |
| 8 | RAM cần cho giá trị định cỡ | H | 10,023.88 | =ExF |
| 9 | CINT nhân hệ số sai số, KPI |  | **55,893.78** | =G\*1,1/0,75 |
| 10 | RAM nhân hệ số sai số, KPI |  | **12,251.40** | =G\*1,1/0,9 |

* Tổng yêu cầu tài nguyên theo tính toán

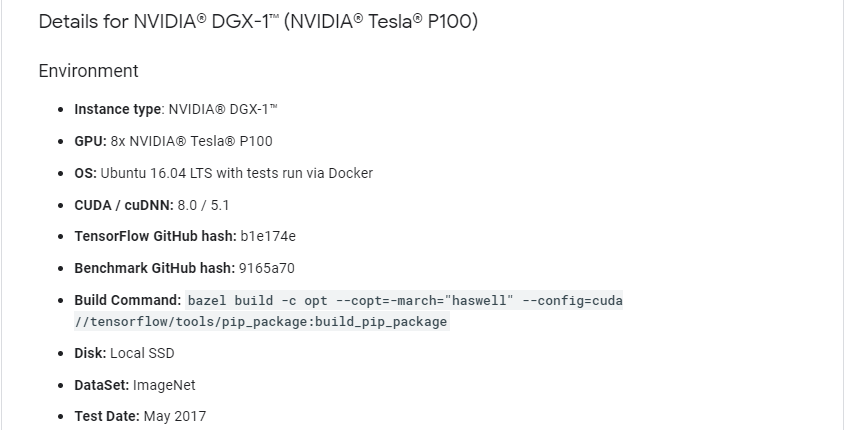
|  |  |  |  |
| --- | --- | --- | --- |
| **Giá trị N** | **Cint CPU Yêu cầu** | **RAM Yêu cầu** | **Yêu cầu HDD (GB)** |
| 1 | **55,893.78** | **12,251.40** | **2,510,698** |
| N=64 | **873.34** | **191.43** | **39,229.66** |

* Đề xuất sử dụng máy chủ có giá trị Cint\_rate = 921, RAM 192 GB (Dự phòng N+1 = 65), ổ cứng 40 TB.
  1. Đề xuất thiết bị

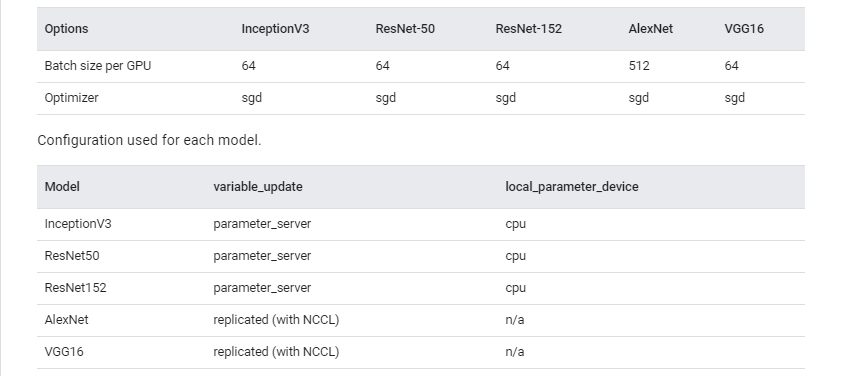
|  |  |  |
| --- | --- | --- |
| **Cấu hình** | **Số lượng** | **Ghi chú** |
| - 2xCPU, CINT ≥ 873  - RAM ≥ 192GB DDR4  - HDD ≥ 40 TB SAS 7.2 Krpm  - HDD OS ≥ 2x600GB 7.2 Krpm  - Card NIC 10G ≥ 2-port 10Gbps  - Card NIC 1G ≥ 2-port 1Gbps  - Nguồn: 2\*500W redundant  - OS Support : Windows, Linux | 65 |  |

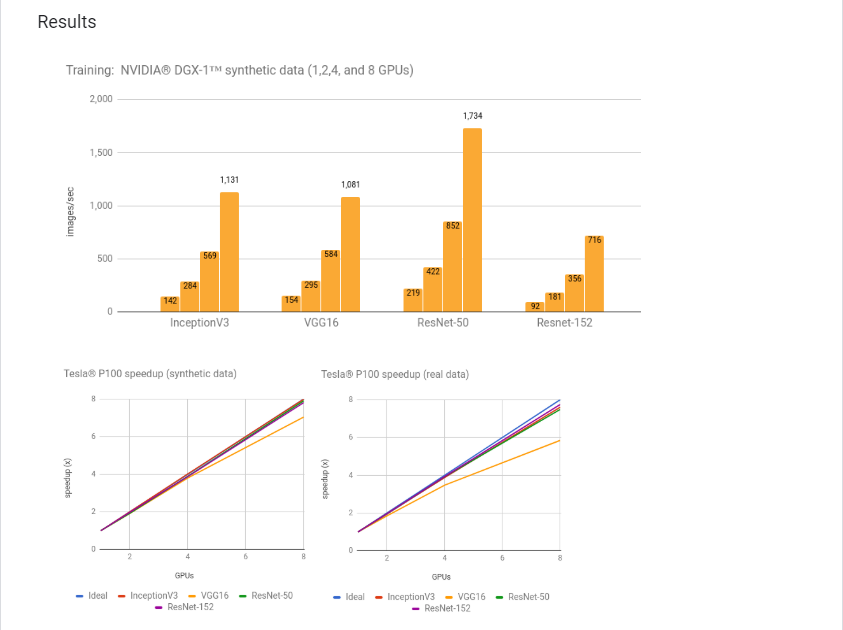
* 1. Định cỡ Card GPU
     1. Cơ sở định cỡ

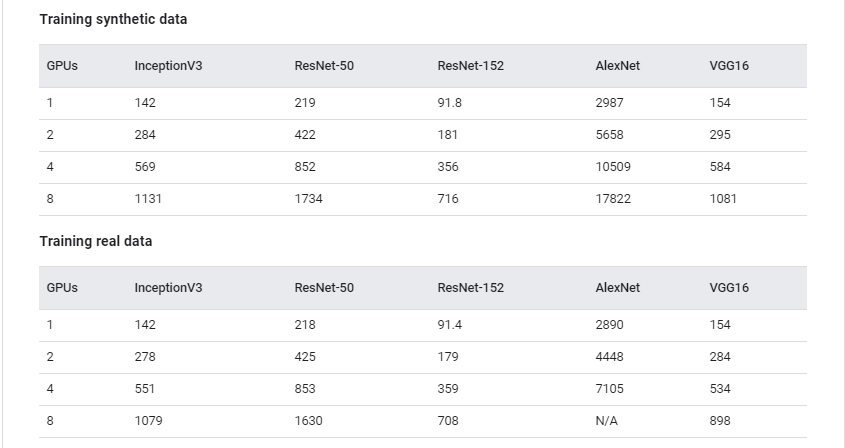
Xuất phát từ nhu cầu nhận dạng ảnh qua chứng minh nhân dân, đáp ứng training tập dữ liệu 5.000.000 ảnh sử dụng mạng nơron Convolutional Neural Network (CNN). Để giảm thời gian training và tăng khả năng thay đổi mô hình trong khoảng thời gian chấp nhận được, chúng ta cần hỗ trợ của GPU. Tham khảo benchmarks tại link <https://www.tensorflow.org/guide/performance/benchmarks>.



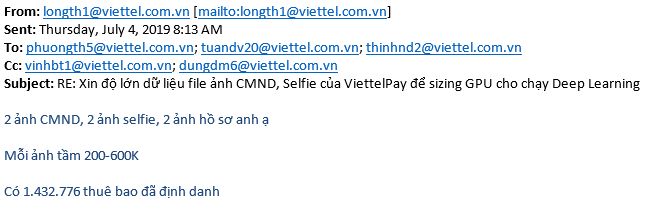
Kích thước 1 ảnh từ nguồn ImageNet: 256 x 256 x 3 = 196608 Bytes ~ 192KBytes.







* + 1. Định cỡ GPU



Dung lương 1 ảnh trung bình: 400Kbytes.

Số lượng ảnh trong kho ảnh: 1432776 \* (2 + 2 + 2) = 8596656 ảnh.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **TT** | **Thông số** | **Ext** | **Giá trị** | **Ghi chú** |
| 1 | Mạng nơron sử dụng |  | **Inception-v3** |  |
| 2 | Dung lượng 1 ảnh trên ImageNet | Stc | **192** | Kbytes |
| 3 | Dung lượng 1 ảnh CMND | D | **8,596,656** | Ảnh |
| 4 | Dung lượng 1 ảnh CMND trung bình | Stb | **400** | Kbytes |
| 5 | Dữ liệu training quy đổi hệ thống tham chiếu | L | 17,909,700 | =D \* Stb / Stc |
| 6 | Số lượng ảnh cần training | Ltt | 14,327,760 | = 80% \* L |
| 7 | Thời gian training | T | 28,800 | Thời gian training dự kiến (8h). |
| 8 | Tốc độ training | V | **497** | Số image được training trong 1 giây. |

* + 1. Đề xuất thiết bị

|  |  |  |
| --- | --- | --- |
| **Cấu hình** | **Số lượng** | **Ghi chú** |
| - **NVIDIA® Tesla® P100** | 4 |  |

* 1. Định cỡ máy chủ Cloud - DMP WORK, API
     1. Cơ sở định cỡ

|  |  |  |  |
| --- | --- | --- | --- |
| **STT** | **Đầu vào** | **Giá trị định cỡ đáp ứng đến năm 2019-2020** | **Hệ thống tương đồng (HT DMP test)** |
| 1 | Tổng dữ liệu đến trong ngày | **11500** GB/ngày. (chi tiết như bên dưới) | **207** GB/ngày |

* + 1. Định cỡ RAM, CPU, HDD
* Lấy hệ thống hiện tại đang chạy là hệ thống DMP API Test để làm sở cứ tính toán. Hệ thống chạy phân tải đều trên gần 5 server.
* Cấu hình
* Cấu hình máy chủ **DMP Test** hiện tại:

| **TT** | **IP** | **Tên máy** | **CPU** | **RAM** | **HDD** | **Cint rate 2006** | **Link tham chiếu** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 10.58.244.168 | DataNode | Intel Silver 4114 | 96 GB | 14 TB | 953 | https://www.spec.org/cpu2006/results/res2016q4/cpu2006-20161101-44994.html |
| 2 | 10.58.244.169 | DataNode | Intel Silver 4114 | 96 GB | 14 TB | 953 | Nt |
| 3 | 10.58.244.170 | DataNode | Intel Silver 4114 | 96 GB | 14 TB | 953 | Nt |

* Tải máy chủ hiện tại:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **TT** | **IP** | **Tên ứng dụng** | **SL giao dịch** | **CINT** | **RAM** | **Tải CPU 95** | **Tải RAM 95** | **Cint used** | **RAM used** |
| 1 | 10.58.244.168 | Big Data | 207G/Ngày | 953 | 96 | 9% | 30% | 85.77 | 28.8 |
| 2 | 10.58.244.169 | Big Data | 207G/Ngày | 953 | 96 | 10% | 32% | 95.3 | 30.72 |
| 3 | 10.58.244.170 | Big Data | 207G/Ngày | 953 | 96 | 9% | 27% | 85.77 | 25.92 |
|  | TỔNG |  |  | 2859 | 288 |  |  | 266.8 | 85.44 |

* Tính toán nhu cầu tài nguyên cho CPU/RAM:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **TT** | **Thông số** | **Ex** | **Máy chủ** | **Ghi chú** |
| 1 | CINT hiện tại | A | **266.84** |  |
| 2 | RAM (GB) hiện tại | B | **85.44** |  |
| 3 | Giá trị hiện tại (Tr) | C | **207** |  |
| 4 | Cint/1Tr giao dịch | D | 1.2891 | =A/C |
| 5 | RAM/1Tr giao dịch | E | 0.4128 | =B/C |
| 6 | Giá trị dữ liệu định cỡ (Tr) | F | **11,500** | =11500/2 |
| 7 | CINT cần cho giá trị định cỡ | G | 14,824.44 | =DxF |
| 8 | RAM cần cho giá trị định cỡ | H | 4,746.67 | =ExF |
| 9 | CINT nhân hệ số sai số, KPI |  | **21,742.52** | =G\*1,1/0,75 |
| 10 | RAM nhân hệ số sai số, KPI |  | **5,801.48** | =G\*1,1/0,9 |

* Tổng yêu cầu tài nguyên theo tính toán

|  |  |  |
| --- | --- | --- |
| **Giá trị N** | **Cint CPU Yêu cầu** | **RAM Yêu cầu** |
| 1 | 21,742.52 | 5,801.48 |
| N=24 | 905.94 | 241.73 |

* Đề xuất sử dụng máy chủ HP DL380 Gen9 (2.00 GHz, Intel Silver 4114) có giá trị Cint\_rate = 953, RAM 192 GB (Dự phòng N+1 = 25)
* Đề xuất thiết bị

|  |  |  |
| --- | --- | --- |
| **Cấu hình** | **Số lượng** | **Ghi chú** |
| - 2xCPU Intel Silver 4114  - RAM 256 GB DDR4  - HDD 4\*2TB SAS 7.2 Krpm  - Card NIC ≥ 2-port 10Gbps  - Card NIC ≥ 2-port 1Gbps  - Nguồn: 2\*500W redundant  - OS Support : Windows, Linux | 25 | https://www.spec.org/cpu2006/results/res2016q4/cpu2006-20161101-44994.html |

* 1. Định cỡ Storage Cloud
     1. Tính toán SSD

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **TT** | **Nội dung** | **Ex** | **Tính toán** | **Đơn vị** | **Ghi chú** |
| 1 | Số lượng file Log | A | 3 | File |  |
| 2 | Dung lượng file Log/ngày | B | 5,000 | MB/Ngày |  |
| 3 | Tổng dung lượng lưu Log file | C | 15,000 | MB/Ngày | =A\*B |
| 4 | Thời gian lưu trữ log theo yêu cầu KD | D | 180 |  |  |
| 5 | **Tổng Nhu cầu dung lượng lưu Log** | **E** | **2,700,000** | **MB** | **=DxC** |
| 6 | Số lượng Request định cỡ (Không tính Peak) | F | 500,000 |  | Trungbình |
| 7 | Kích thước 1 Request khi nén | G | 120 | KB |  |
| 8 | Tổng dung lượng Lưu trữ DB Record | H | 60,000 | MB/Ngày | =GxF |
| 9 | Thời gian lưu trữ nghiệp vụ theo yêu cầu KD | I | 365 | Ngày |  |
| 10 | **Tổng dung lượng lưu DB Record** | **J** | 21,900,000 |  | =IxH |
| 11 | **Tổng dung lượng Log+DB** | **K** | 24,600,000 |  | =J+E |
| 12 | Dung lượng OS | L | 50,000 | MB |  |
| 13 | Tổng dung lượng nhân sai số và dự phòng KPI | M | 33,893,750 | MB | =(H+I)\*1,1/0,8 |
| 14 | Tổng dung lượng sau khi bổ sung Raid 5 | N | 48,419,643 | MB | =M/0.75 |

* + 1. Tính toán HDD

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **TT** | **Nội dung** | **Ex** | **Tính toán** | **Đơn vị** | **Ghi chú** |
| 1 | Số lượng file Log | A | 3 | File |  |
| 2 | Dung lượng file Log/ngày | B | 5,000 | MB/Ngày |  |
| 3 | Tổng dung lượng lưu Log file | C | 15,000 | MB/Ngày | =A\*B |
| 4 | Thời gian lưu trữ log theo yêu cầu KD | D | 1,825 |  |  |
| 5 | **Tổng Nhu cầu dung lượng lưu Log** | **E** | **27,375,000** | **MB** | **=DxC** |
| 6 | Số lượng Request định cỡ (Không tính Peak) | F | 500,000 |  | Trungbình |
| 7 | Kích thước 1 Request khi nén | G | 120 | KB |  |
| 8 | Tổng dung lượng Lưu trữ DB Record | H | 60,000 | MB/Ngày | =GxF |
| 9 | Thời gian lưu trữ nghiệp vụ theo yêu cầu KD | I | 730 | Ngày |  |
| 10 | **Tổng dung lượng lưu DB Record** | **J** | 43,800,000 |  | =IxH |
| 11 | **Tổng dung lượng Log+DB** | **K** | 71,175,000 |  | =J+E |
| 12 | Dung lượng OS | L | 50,000 | MB |  |
| 13 | Tổng dung lượng nhân sai số và dự phòng KPI | M | 97,934,375 | MB | =(H+I)\*1,1/0,8 |
| 14 | Tổng dung lượng sau khi bổ sung Raid 5 | N | 139,906,250 | MB | =M/0.75 |

* Đề xuất thiết bị

|  |  |  |  |
| --- | --- | --- | --- |
| **TT** | **Cấu hình** | **Số lượng** | **Ghi chú** |
| **1** | ***Disk Array*** | - Tương thích EMC Unitity 400 |  |
| **2** | ***HDD Storage*** | - 50 TB SAS SSD - 150 TB SAS 10Krpm |  |

1. ĐỊNH CỠ NETWORK
   1. SW 10G

Sử dụng hạ tầng 10G để đảm bảo tốc độ nội mạng Layer2 cao khi truy xuất dữ liệu big data, cloud.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **TT** | **Tên loại thiết bị** | **SL** | **Số port/1 TB** | **Tổng port** |
| 1 | SAN SW | 2 | 1 | 2 |
| 2 | DMP Big Data | 65 | 1 | 65 |
| 3 | Cloud ViettelPay | 25 | 1 | 25 |
| 4 | Tổng |  |  | 90 |
| 5 | **Tổng** |  | **Kdp = 1.1, KPI = 0.8** | **123.75** |
| 6 | SL SW 48port |  |  | 2.58 |

Số lượng SW 48 port cần là 3 SW chạy đơn, SW sẽ được đấu bond0 để dự phòng N+N. Do vậy cần 6 SW.

* Đề xuất thiết bị

|  |  |  |  |
| --- | --- | --- | --- |
| **TT** | **Cấu hình** | **Số lượng** | **Ghi chú** |
| **1** | ***SW 10G 48port*** | 6 |  |

* 1. SW quản trị 1G

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **TT** | **Tên loại thiết bị** | **SL** | **Số port/1 TB** | **Tổng port** |
| 1 | SAN SW | 2 | 1 | 2 |
| 2 | DMP Big Data | 65 | 1 | 65 |
| 3 | Cloud ViettelPay | 25 | 1 | 25 |
| 4 | Tổng |  |  | 90 |
| 5 | **Tổng** |  | **Kdp = 1.1, KPI = 0.8** | **123.75** |
| 6 | SL SW 48port |  |  | 2.58 |

Số lượng SW 48 port cần là 3 SW chạy đơn để ILO, không cần dự phòng

* Đề xuất thiết bị

|  |  |  |  |
| --- | --- | --- | --- |
| **TT** | **Cấu hình** | **Số lượng** | **Ghi chú** |
| **1** | ***SW 1G 48port*** | 3 |  |

* 1. Rack

Tính toán SL Rack để lắp đặt thiết bị.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **TT** | **Tên loại thiết bị** | **SL** | **Số RackUnit(RU)/1 TB** | **Tổng RU** |
| 1 | SAN SW | 2 | 2 | 4 |
| 1 | SW 1G | 2 | 2 | 4 |
| 1 | SW 10G | 6 | 2 | 12 |
| 2 | DMP Big Data | 65 | 2.5 | 162.5 |
| 3 | Cloud | 25 | 2.5 | 62.5 |
| 4 | DiskArray | 3 | 2.5 | 7.5 |
| 5 | Tổng |  |  | 252.5 |
| 6 | **Tổng** |  | **Kdp = 1.1** | **277.75** |
| 7 | SL Rack 42U |  |  | 6.61 |

* Đề xuất thiết bị

|  |  |  |  |
| --- | --- | --- | --- |
| **TT** | **Cấu hình** | **Số lượng** | **Ghi chú** |
| **1** | ***Rack 42U, chuẩn APC, bao gồm 2 PDU*** | 7 |  |

1. TỔNG HỢP NHU CẦU PHẦN CỨNG

|  |  |  |  |
| --- | --- | --- | --- |
| **TT** | **Dịch vụ thuê máy chủ** | **Thông số** | **SL** |
| 1 | **Vật tư phụ kiện** |  | **1** |
| 2 |  | - Dây nguồn 500m (loại 3x10 mm2) |  |
| 3 |  | - Dây tiếp địa 250m (loại 1x25mm2) |  |
| 4 |  | - 21 Chấu cắm dây nguồn tủ cho Rack |  |
| 5 |  | - Cáp mạng 02 thùng (Cat6 Panduit) |  |
| 6 |  | - 03 hộp RJ45 Cat5; 02 gói lạt thít |  |
| 7 |  | - 100 dây quang 3 mét |  |
| 8 |  | - 100 dây quang 5 mét |  |
| 9 |  | - 100 dây quang 10 mét |  |
| 10 | ***SAN Switch*** | ***Cisco MDS 9148s*** | ***2*** |
| 11 | Port active | 48 port active 16 Gbpcs | 1 |
| 12 | ***Switch 1GE*** | ***HP 5510*** | ***2*** |
| 13 | Port | 48 port 1GE+, 2 port SFP+ | 1 |
| 14 | ***Switch 10GE*** | ***HP 5940*** | ***6*** |
| 15 | Port | 48 port 10GE+, SFP+ | 1 |
| 16 | ***Máy chủ loại 1*** |  | ***65*** |
| 17 | CPU | 2xCPU Siver 4114 hoặc tương đương | 1 |
| 18 | RAM: | 196 GB | 1 |
| 19 | HDD | 10x6TB, 7.2Krpm SAS | 1 |
| 20 | NIC 1GE | 2x1GE | 1 |
| 21 | NIC 2GE | 2x10GE | 1 |
| 22 | Power | 2xPower AC | 1 |
| 23 | ***Máy chủ loại 2*** |  | ***25*** |
| 24 | CPU | 2xCPU Siver 4114 hoặc tương đương | 1 |
| 25 | RAM: | 256 GB | 1 |
| 26 | HDD | 4x2TB, 7.2Krpm SAS | 1 |
| 27 | HBA | 2x10GE | 1 |
| 28 | NIC 1GE | 2x1GE | 1 |
| 29 | NIC 2GE | 2x10GE | 1 |
| 30 | Power | 2xPower AC | 1 |
| 31 | ***Disk Array*** | *Tương thích EMC Unitity 400* | ***2*** |
| 32 | ***HDD Storage*** | *- 50 TB SAS SSD - 150 TB SAS 10Krpm* | ***2*** |
| 33 | **GPU Card** | GPU V100 hoặc tương đương | 8 |
| 34 | **Bản quyền OpenStack** |  | 1 |
| 35 | **Rack 42U** | ***42U, 19Inch, chuẩn APC, kèm 2PDU*** | 8 |

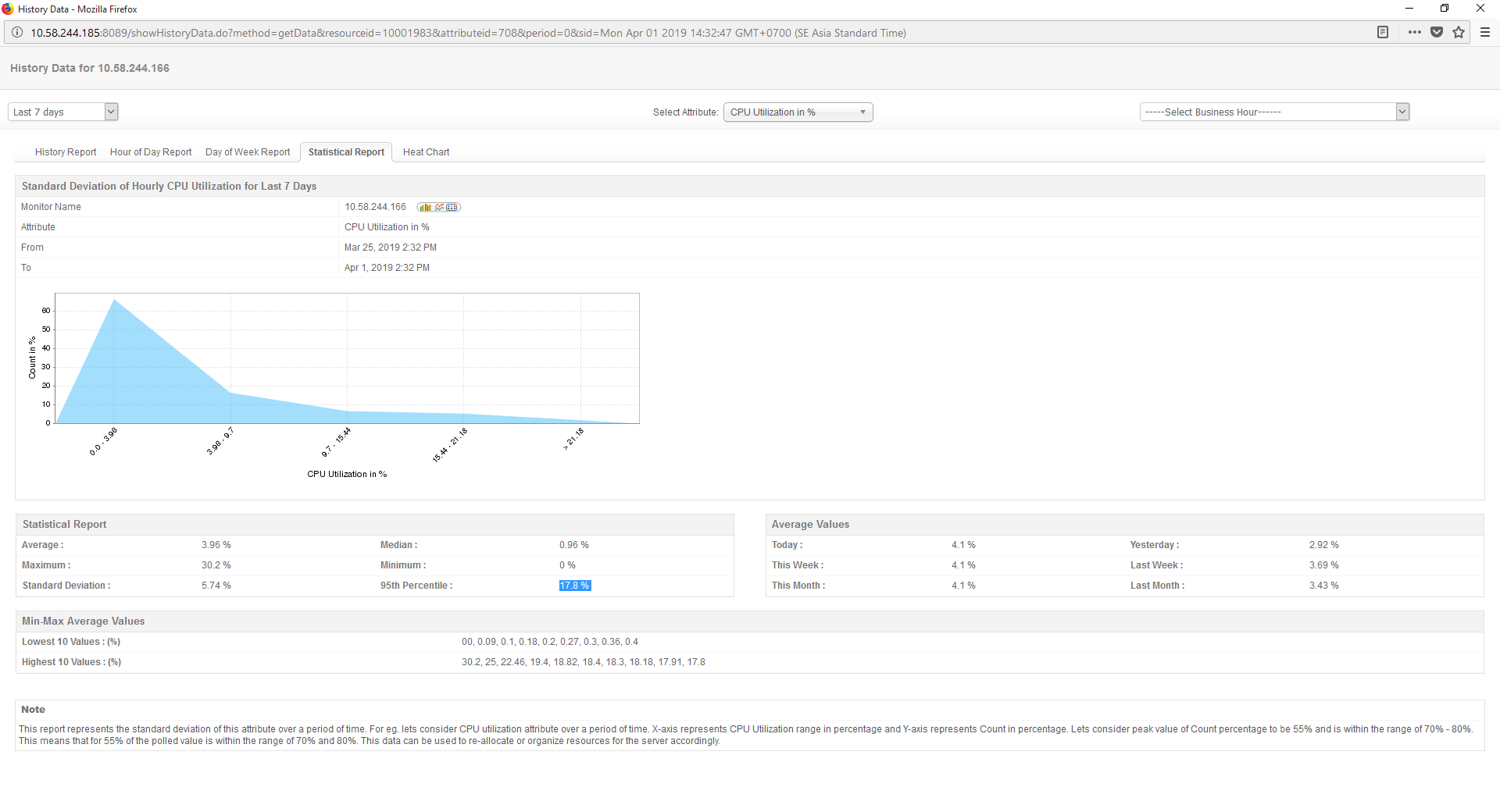
**PHỤ LỤC**

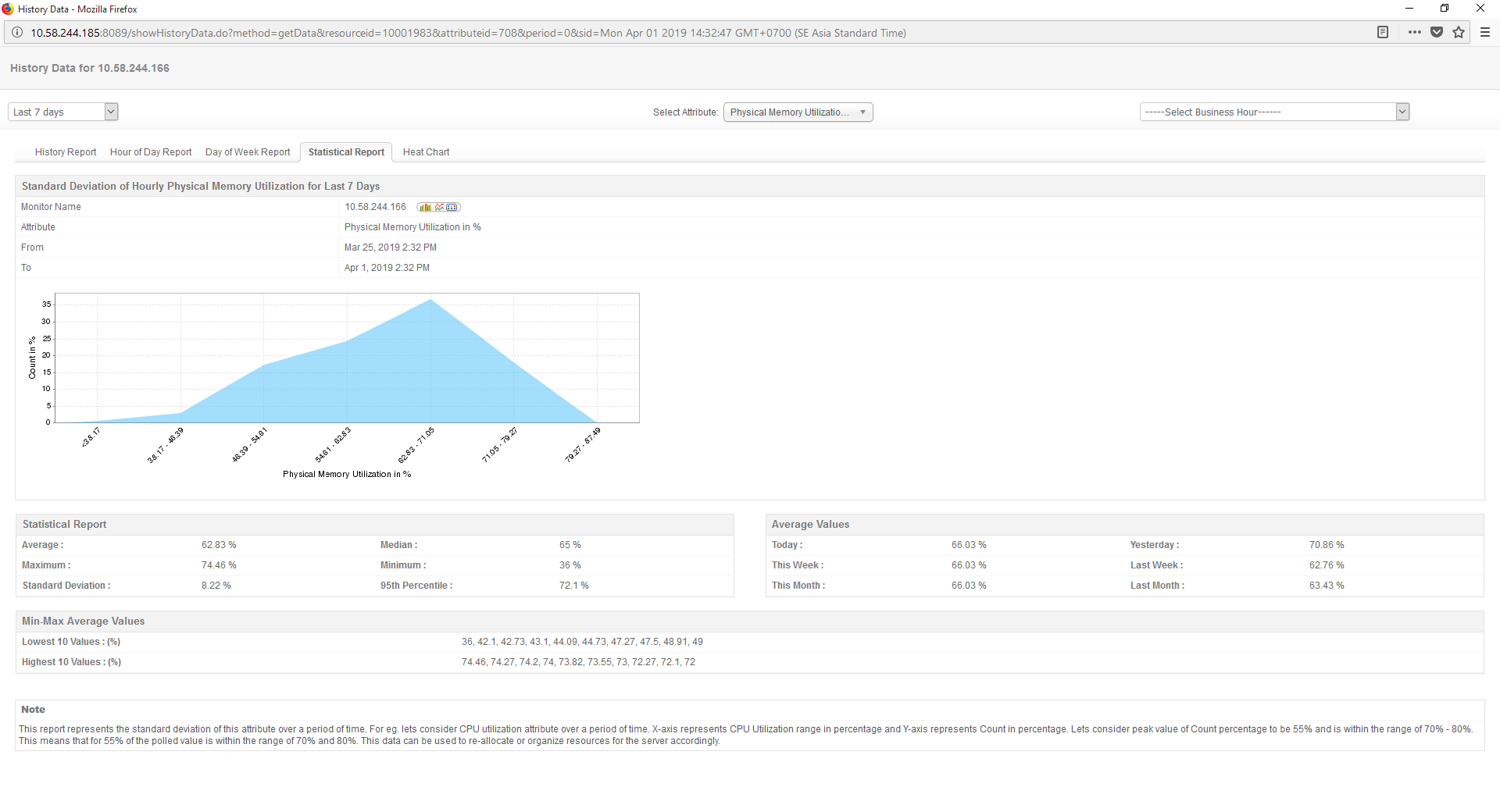
**Tải và dung lượng hệ thống DMP**

1. Dung lượng xử lý hệ thống DMP test

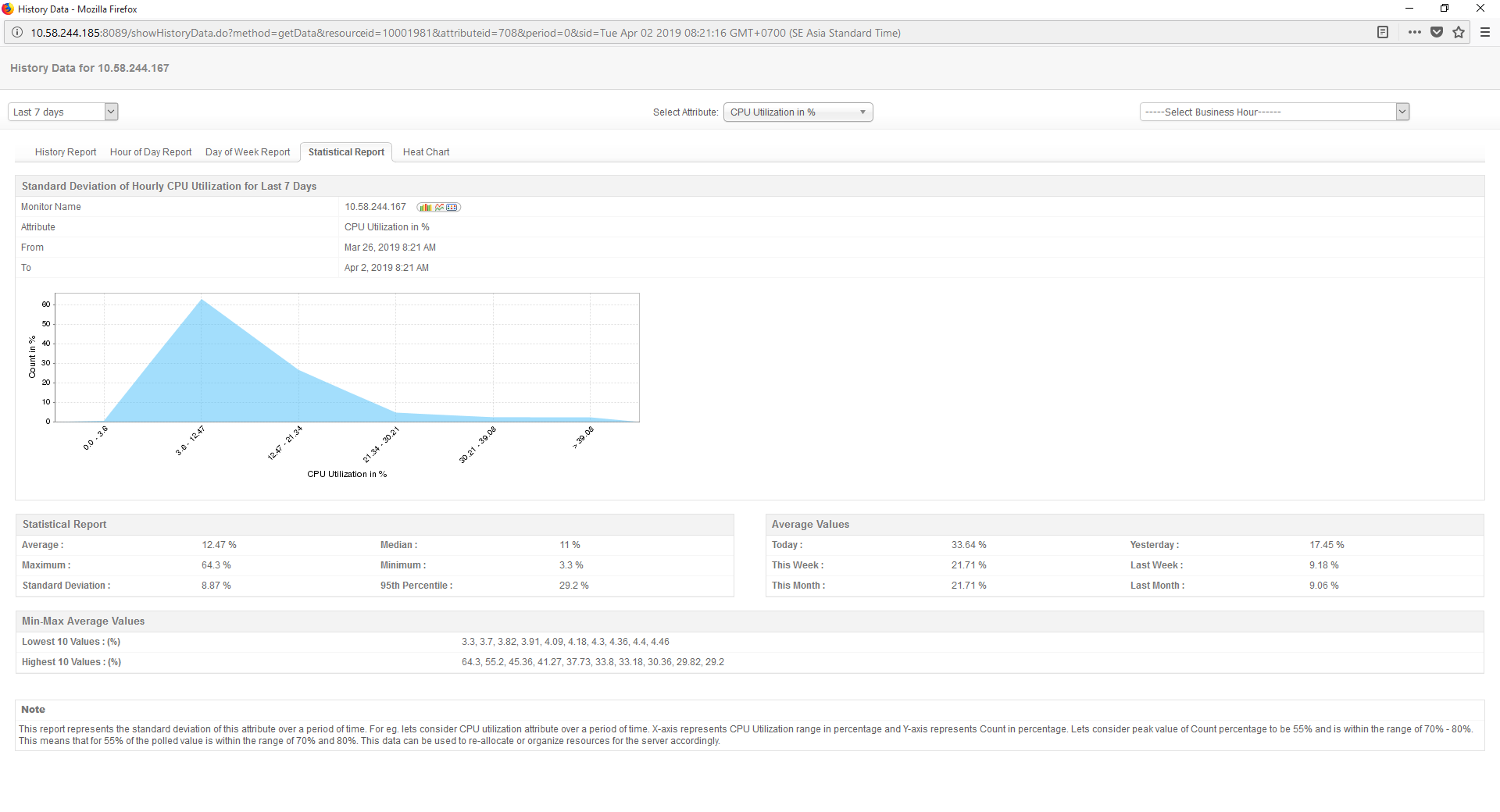
|  |  |  |
| --- | --- | --- |
| **CDR** | **IMAGE** | **AVERAGE**  **(MB)** |
| VAS\_MPS\_CHARGE |  | 1325.0 |
| CDR\_MSC\_VOICE |  | 13737.5 |
| CDR\_MSC\_SMS |  | 6180.0 |
| REG\_CDR\_21H |  | 5314.0 |
| REG\_CDR\_09H |  | 1240.0 |
| reportIN |  | 1924.0 |
| DailyFeeCdr |  | 57.9 |
| export\_acct\_book |  | 432.7 |
| G22 |  | 8375.0 |
| G33 |  | 1645.0 |
| MOBILE\_mb\_postpaid\_call |  | 1796.0 |
| MOBILE\_mb\_monthly\_fee |  | 162.234 |
| LoanCdr |  | 48.828 |
| RechargeList |  | 157.43 |
| Subscriber |  | 25.19 |
| Total |  | **42395.592** |

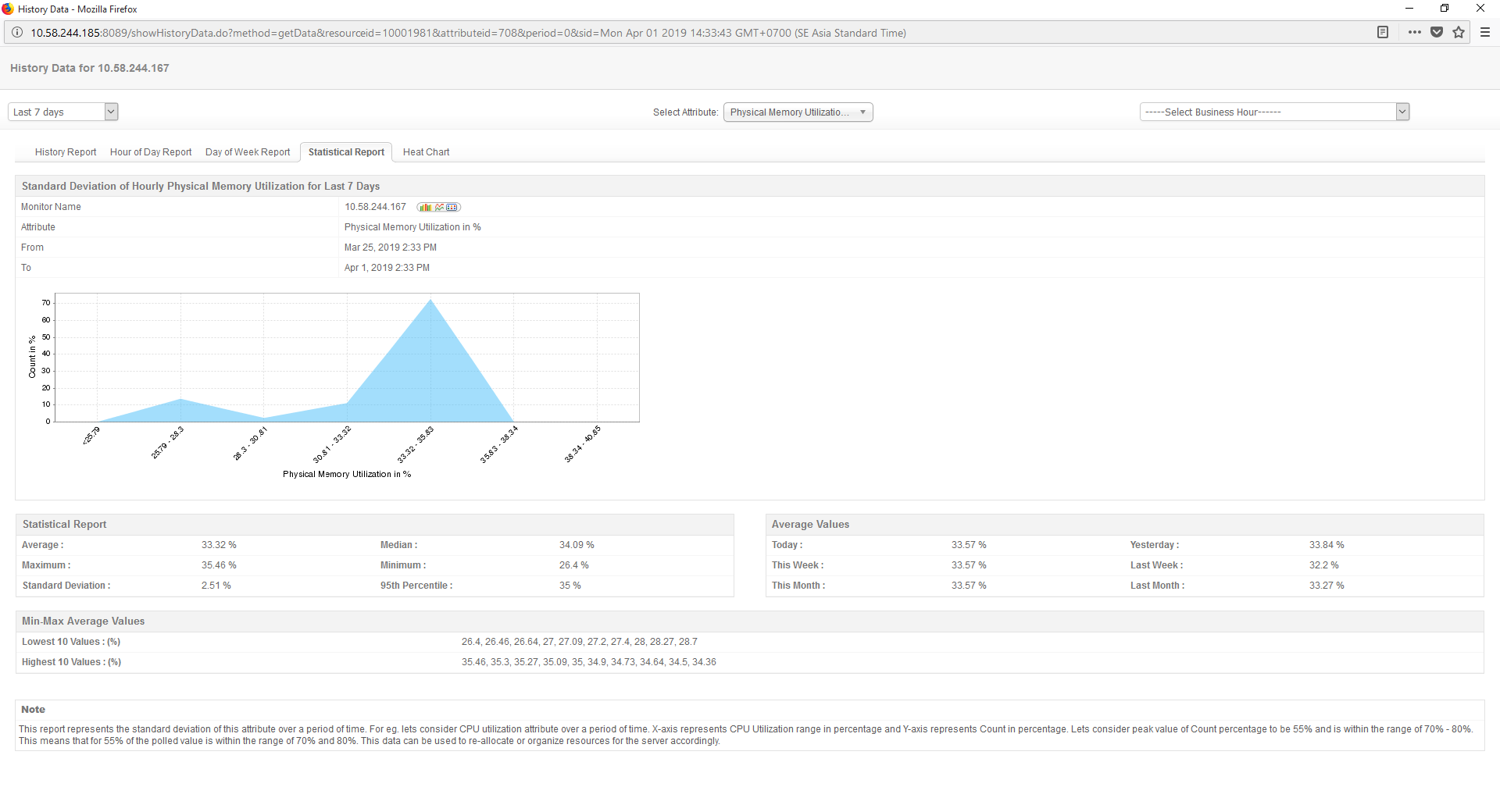
1. 10.58.244.166





1. 10.58.244.167





1. 10.58.244.169

