**２０２１　年度　　卒業論文**

卒業論文題目

|  |
| --- |
| 観光地検索システムにおけるレスポンス速度を考慮したロードバランサー |

　指導教員　　　　　　鷹野　孝典　教授

神奈川工科大学　情報工学科

　　　　　学籍番号　　　　１８２１０８６

　　　　　学生氏名　　　　　 松尾　祐介

　提出日　２０２１年　月日　指導教員　　印

　受理日　２０２１年　月日　情報工学科長　　印

# **論文要旨**

# 目次

[**1.** **まえがき** 1](#_Toc62641419)

[1.1 はじめに 1](#_Toc62641420)

[1.2 本論文の構成 2](#_Toc62641421)

[**2.** **関連研究** 3](#_Toc62641422)

[2.1 遠隔講義における学習効果と対面講義との学習効果の差異 3](#_Toc62641423)

[2.2 Webサイトの検索機能構築について 3](#_Toc62641424)

[2.3 スライド検索技術との相違点 3](#_Toc62641425)

[**3.** **提案システム** 5](#_Toc62641426)

[3.1 提案システムの概要 5](#_Toc62641427)

[3.2 提案システムの実行手順 5](#_Toc62641428)

[3.2.1 講義動画内容抽出機能と抽出するメタデータ 7](#_Toc62641429)

[3.2.2 文字認識機能と抽出するメタデータ 7](#_Toc62641430)

[3.2.3 検索機能とユーザの検索に対応して提示するメタデータ 7](#_Toc62641431)

[3.2.4 テーブル設計 7](#_Toc62641432)

[**4.** **実装システム** 8](#_Toc62641433)

[4.1 実装システムの概要 8](#_Toc62641434)

[4.2 Webシステムの構築 9](#_Toc62641435)

[4.3 DJangoの利用 10](#_Toc62641436)

[4.3.1 仮想環境の導入 12](#_Toc62641437)

[4.3.2 Webアプリケーションの実行 13](#_Toc62641438)

[4.4 ffmpegを利用したメタデータの抜粋 18](#_Toc62641439)

[4.4.1 メタデータを抜粋するための処理 19](#_Toc62641440)

[4.4.2 ffmpegの実行 19](#_Toc62641441)

[4.5 Tesseract-OCRを利用した文字認識 20](#_Toc62641442)

[4.5.1 Tesseract-OCRの導入 21](#_Toc62641443)

[4.5.2 Tesseract-OCRの利用 21](#_Toc62641444)

[**5.** **実験** 25](#_Toc62641445)

[5.1 実験目的 25](#_Toc62641446)

[5.2 実験環境 25](#_Toc62641447)

[5.2.1 理解度調査の流れ 26](#_Toc62641448)

[5.3 実験 27](#_Toc62641449)

[5.4 実験の考察 28](#_Toc62641450)

[5.4.1 実験の感想一覧 29](#_Toc62641451)

[**6.** **むすび** 30](#_Toc62641452)

謝辞

参考文献

# **まえがき**

## はじめに

WEBサーバは重要なライフラインになりつつある．サービス全体の応答性能や可用性を高めることができる「ロードバランサ」の需要は今後増加傾向になると予想される．ロードバランサは企業に限った話ではない。個人サイトでもサービスが拡大するにつれて負荷分散は重要視される．Googleの発表した情報によると競合サイトと比較し自身のサイトの表示速度が遅いとランキング評価で不利になるとされている．負荷分散をしつつ、自身のサイトを上位にランクインさせるためには応答速度も重要な要素となる．個人サイトで高性能なサーバをいくつも作り，負荷分散するのはコスト面で難しい．性能の低いコンピュータや旧式のサーバ等を利用し，不均一な性能で異種のコンピュータを組み合わせてWEBサイトを構築し，ロードバランスする場合も少なくない．サーバ間の性能や通信装置の性能にバラつきがある場合，応答速度が一定とは限らない．単純に空いているサーバへ割り振ると，遅いサーバにつないでしまいWEBの表示速度が落ちてしまうという課題がある．そういった課題を解決するには，異なる性能のサーバそれぞれの応答速度を計測し，ロードバランサが判断して動的に割り振りを行う必要がある。

本研究では，性能差のあるWEBサーバからなるWEBサイト環境において，応答速度によってサーバの割り振り先を決めるアルゴリズムの提案，WEBサーバの応答速度を考慮したロードバランサの設計と開発を行った。提案システムでは，それぞれのWEBサーバの応答速度を計測し，提案するアルゴリズムに基づいて評価付けを行うが、応答の遅速は個人によって異なる為、Googleが調査した「Webサイトの反応時間の遅延と、それに対するユーザの反応」を参考に評価システムを設計している。また、この情報は評価済み応答速度としてデータベースへ保管される。データベースから現在と過去24時間の状態を抽出し、ロードバランサに与える．こうすることで不均一なサーバからなるWEBサイトにおいて、応答速度を考慮しつつ負荷分散をすることができる．

WEBサーバで負荷分散を行うロードバランサはプロキシ機能を応用している．このプロキシには不特定多数の Client から、特定の Server へ通信を代わりに行うリバースプロキシとClient から不特定の Server にアクセスする際、Client の代わりに通信を行うフォワードプロキシが知られている．WEBサーバで負荷分散するには前者が使われる。NGINXというWEBサーバはコンフィグを書き換えることで、リバースプロキシが行える。また重みづけにも対応している為、応答速度を計測した後、自身の作成した評価アルゴリズムに基づいて重みを変更し、適応させることで自作の動的ロードバランサを作ることができると考えた．通常コンフィグを適用させるにはWEBサーバの再起動が必要だが、これでは稼働率が落ちてしまう。そこで設定ファイルを再起動することなく適応できるreloadと呼ばれる機能を使用した．

実験では教育目的で開発されたシングルボードコンピュータであるRaspberry Pi 4B 4Gを4台使用する。ロードバランサ1台と検索システムを積んだWEBサーバ3台を冗長的に配置する。OS上でネットワークトラフィックをコントロールするTCコマンドやRAMやCPUの使用率を変更できるstressコマンドを使用し異種のサーバを再現する。

本研究では実装した提案方式のプロトタイプと一般的なロードバランサの割り振り方式である「ラウンドロビン」を応答速度に基づいて比較実験することで提案システムの有用性を示す．

## 本論文の構成

　本論文は全６章で構成されている。本章では研究の背景や動機，目的について述べる．第２章ではロードバランスの手法や計測システムの関連研究について述べる．第３章では本研究の提案方式とその特徴である本システムの内容，応答速度をどのように計測・評価し，ロードバランサが判断し割り振りを行うのか述べる。第４章では提案方式を構築した本システムの設計及び実験システムの実装について述べる．第５章では一般的なロードバランサと本システムで比較実験を行う。実験の概要と，比較実験後の結果を述べる．第６章では結論と今後の展望について述べる．

# **関連研究**

## リバースプロキシによるロードバランシング手法

HTTPセッションのハンドオーバによるWEBサーバのロードバランスについて土居幸一郎,後藤滋樹らは既存のwebロードバランシング手法では，実行中のデータ転送を再配分することが出来ない.そこで，リバースプロキシ型のロードバランシング手法を元に，データ転送中であっても，相対的に負荷の高いサーバから負荷の低いサーバへと移し替えられる機構を提案した．4MB以下のファイルでは,ロードバランサを経由しないダウンロードの方が,ロードバランサを経由する場合に比べて,わずかに必要時間が短かった.しかし,4MB以上ではロードバランサを経由する場合の方のダウンロード時間が短くなる場合がある.したがって,音楽や映像のファイルのような比較的サイズの大きなファイルに対しては,はかの要素の寄与度に比べて,提案手法によるオーバヘッドは十分に小さいと言える.ロードバランサを経由しない場合とハンドオーバを行う場合とで,ダウンロードに必要な時間に違いは生じなかった、試行によっては,ハンドオーバを行った場合でロードバランサを経由しない場合のダウンロード時間を下回る場合もあった,よって,提案手法によるハンドオーバによるコストは十分に小さいことがわかった.が増加するに従って,サーバリソースの利用効率が向上す一方で,5接続を超えてからは利用効率が向上していない.その場合は,別の5接続数以下のサーバにデータ転送をハンドオーバすることによって,全体で利用効率を向上させることができる.図5の結果をもとに,2台のバックエンドサーバ同士でクライアント接続数が5台以下となるように調整した場合の平均転送速度と最低転送速度の増加率を表3に示す,平均速度と最低速度を,それぞれ最大で18.7%,31.1%向上させられることが分かった.

## 応答速度評価付けシステムの評価手法

複数のロードバランサによるWEBシステムの応答時間最適化について、河野知行はバランサなし、定率分散、状態分散、応答時問分散の4つのケースを比較した。それらの振り分け技法の特徴を整理すると。パランサがなくとも、処理要求がパランサに偏りなく到着すれば、速い応答時間が確保される。しかし、偏りが生じれば、応答時間は指数的に悪くなる。定率分散は到着する処理要求に偏りが生じても、応答時間が指数的に悪くなる問題は解消できる。但し、偏りがない場合は、バランサがない場合より応答時間は長くなる。状態分散を行うと、応答時間は大幅に短縮される。M/M/n構成の待ち行列のような効果を得ることができる。応答時間分散は、ロジックの複雑さに比例した効果を得ることができない。定率分散の改善型と位置づけられる技法であろう。一般的なロードバランサが提供していると考える処理要求の振り分け技法を検証した。その結果により、状態分散が最も効果的な振り分け技法であるとの結論を得た。

## WEBサーバ計測システムの設計・開発

クラウドに適したWebシステムにおいて堀内晨彦 , 最所圭三らはWEBシステムの概要及び試作システムの評価について述べ、その負荷量の監視機能の改善方法を提案・評価を行った。i. Apacheのserver-statusで負荷量を測定　ii. 外部プログラムで負荷量を測定方法iでの測定結果を図4に,方法iでの測定結果を図5に示す。実験では,前の測定が終了してから1秒後に次の測定を開始する.そのため,プロットの間隔は負荷量の測定にかかった時間+1秒になる.どちらの方法でもスレッド数の増加に伴い,CPU 使用率及び稼働率が上昇しているが,図5のCPU 使用率はCGIを処理するため,すぐに100%に達している,図4では稼働率が1.0に達すると,次の測定までに1分以上かかっている。これはWebサーバデーモンの最大同時処理数以上のアクセスがあり,過負荷になっているためである.また,どちらの図でもアクセスが終わってから負荷量が低下するまでに20秒程度かかっているが,これはWebサーバが各スレッドの行った最後のアクセスを処理しているためである.以上の結果から,アクセス数に応じて増減する稼働率が負荷量に適していることが分かる。CPU使用率は同時アクセス数が少なくても100%に達するため,過負荷の判別ができない.また,稼働率の測定方法としては,Web サーバが過負荷になった場合でも短時間で測定が行える方法iが適しているが,過負荷を脱してから稼働率が低下するまでに時間がかかる.反対に方法iでは過負荷を脱すると直ぐに稼働率が低下するが,過負荷になった場合の測定に時間がかかる.

# **提案システム**

## 提案システムの概要

本研究では性能差のあるWEBサーバからなるWEBサイト環境において応答速度によってサーバの割り振り先を決めるアルゴリズム，応答速度を考慮したロードバランサの設計方法を提案する。

本提案システムでは常に応答速度を計測している為、サーバの故障やロードバランサとWEBサーバ間でボトルネックが生じたとしても，ユーザがシステムの不具合を感じることはない。通常ロードバランサは性能が同一もしくはそれに近い性能で運用されることが前提とされているが，本提案システムの特徴として，異種の不均一な性能のサーバ環境において個人を対象とした負荷分散に着目している。検索システムはWEBサービスの中でも応答速度が重要視されている。また膨大な検索にも耐えられるためには負荷分散が重要であるため効果は高い。冗長的なWEBサービスに搭載するシステム図は〇〇に、説明に関しては〇〇に示す。また評価済み応答速度を保管するデータベースのテーブル設計は〇〇に示す。

## 提案システムの実行手順

本提案システムではサーバの応答速度に変化が生じるとロードバランサが動的に割り振りを変更させるので負荷分散をしつつ最速のWEBサーバへ多くのユーザをつなぐことが出来る。本提案システムの処理は大きく「計測」「評価」「処理」3段階に分けられ、5つのステップを踏んで実行される。STEP-1では想定しているWBEサーバについて説明している。ロードバランサが応答速度に基づいて動的に割り振りを行うまでの流れに関しては、下記のSTEP-2からSTEP-5で説明する。提案システムにおける流れは図〇〇に示している。

STEP-1:冗長的で性能が不均一なWEBサーバを用意。大量アクセスによるCPUやネットワークの負荷を分散させるため，仮想環境で作られるWEBサーバではなく，ここでは物理的な別のサーバを数台用意する。

STEP-2:それぞれのサーバの応答速度を測るため、作成した「応答速度計測プログラム」を使う。ここでは1分に1回各サーバの応答速度を測る.

STEP-3: STEP-2で計測したデータは考案した応答速度評価アルゴリズムを用いてS～D評価を付ける。評価は主観的になりやすい為、Googleが調査したWebサイトの反応時間の遅延と、それに対するユーザの反応を参考に評価する。

STEP-4:評価されたデータは評価済み応答速度としてデータベースへ保管される。

STEP-5:ロードバランサはこのデータベースへアクセスする。サーバの状態に応じて割り振り方法を動的に変化させることが可能になる。応答速度が最も早いサーバへの接続が優先される。

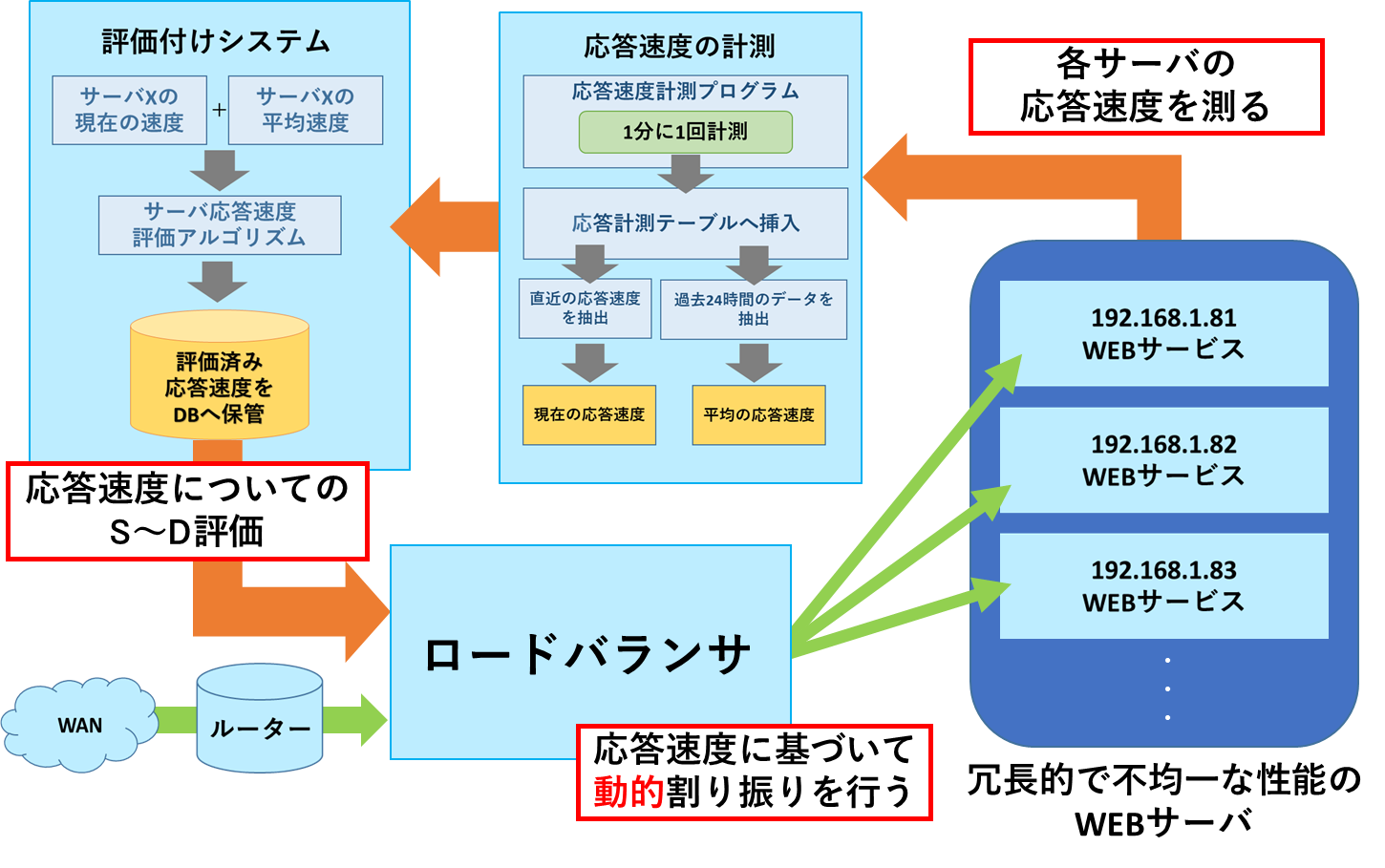


図 3‑1　提案システム図

### 評価付けの方法

実行手順のSTEP-3では評価付けを行う.主観的な評価にならないようGoogleのWebサイト反応

時間の遅延と、それに対するユーザの反応を基に評価している。下記は応答速度と評価の対応表である。

|  |  |
| --- | --- |
| 応答速度 | 評価 |
| 0.000~0.016 | S |
| 0.017~0.099 | A |
| 0.100~0.999 | B |
| 1.000~9.999 | C |
| 10.000以上 | D |

### テーブル設計 monitoring81,82,83

|  |  |
| --- | --- |
| データ名 | データ形式 |
| ID | integer primary key |
| now\_speed | read |
| now\_speed\_score | read |
| Ave\_speed | read |
| Ave\_speed\_score | read |
| datetime | text |

### ロードバランスに使われる技術

WEBにおいて、ロードバランサで負荷分散する場合に使われる技術として「リバースプロキシ」がある。これはプロキシ機能を応用した技術によってユーザからのリクエストを別のWEBサーバへ転送している。動的、静的どちらのロードバランサを作成するときもこの機能は使われている。本提案システムでも，実行手順のSTEP-5でユーザからのアクセスを適宜WEBサーバへ転送する為この技術を使う必要がある．

# **実装システム**

## 実装システムの概要

実装システムでは，物理的なWEBサーバ，ロードバランサ機共に「raspberry pi 4b 4gb」を利用する。また，負荷分散するWEBサービスは「検索システム」を対象とする。この検索システムは観光地のDBと接続され，場所をフォームに入力することで観光地として有名な場所を検索できる。図〇では実装したシステムを説明し，流れは下記に示す．

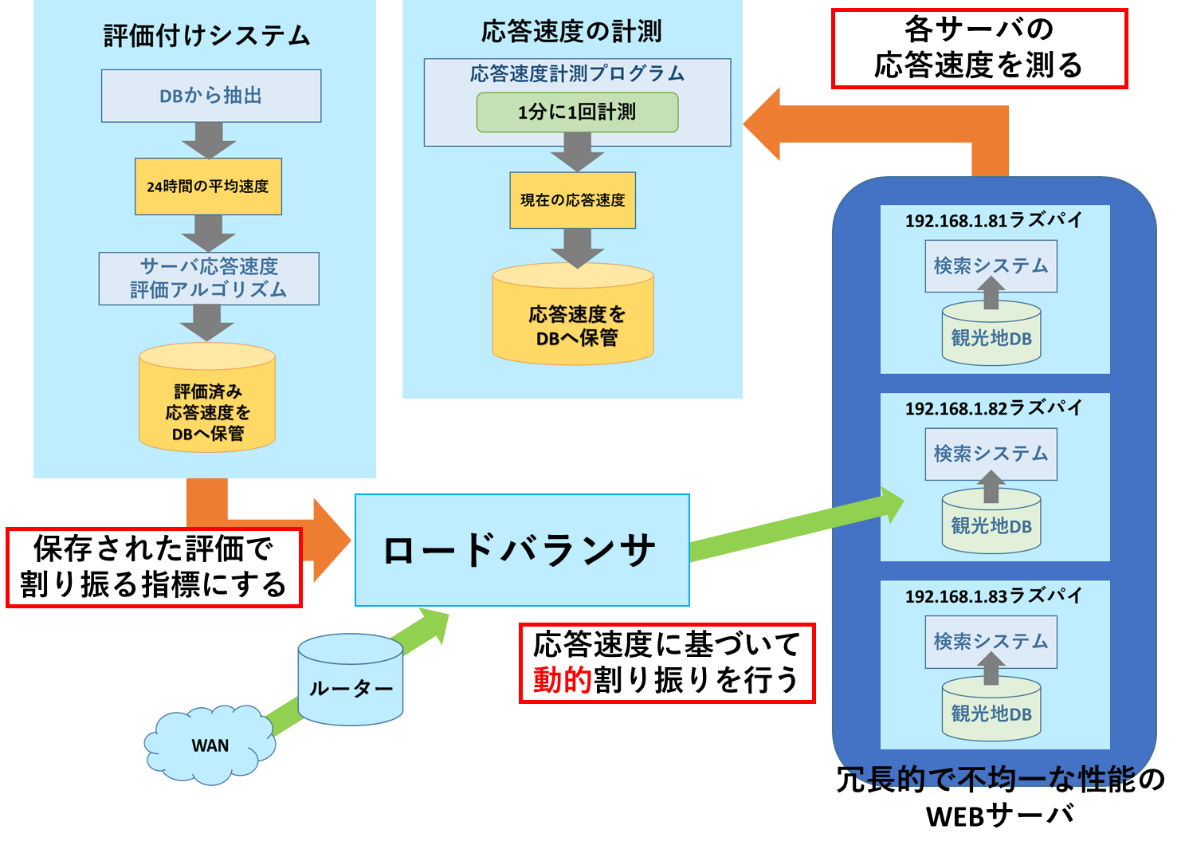


図 4‑1　実装システム図

## ロードバランサの割り振り方法

ロードバランサは過去24時間で評価が最も高いサーバに多く割り振る。評価の良いサーバが変化した場合，ロードバランサの割り振り方法も動的に変化する。

図を載せて解説？

## ロードバランサの構築

動的ロードバランサの構築にはNGINXというWEBサーバを用いる。コンフィグデータを変更して重みづけをプログラムによって書き換え，再読み込みさせることによって実現させる。提案システムのプロトタイプを実装した開発環境は表〇に示す．ロードバランサを作成するために用いたWEBサーバは表〇に示す。

ロードバランサの開発環境

|  |  |  |
| --- | --- | --- |
| 名称 | バージョン | 利用用途 |
| Raspberry Pi | 4B 4G | ロードバランサ |
| Python | 3.8 | 開発言語 |
| Spider(anaconda3) | 4.1.5 | 統合開発環境 |
| Firefox |  | WEBブラウザ |
| Tera Term |  | SSH接続・ファイル転送 |
| SQLITE3 |  | 開発データベース |
| ラズビアン |  | Raspberry PiのOS |
| Windows10 | 20H2 | OS |

提案システム構築に用いたパッケージ

|  |  |  |
| --- | --- | --- |
| パッケージ名称 | バージョン | 備考 |
| SQLITE | 3.0 | 応答速度保管に使用 |
| subprocess |  | プログラムからNGINX適応のシェルを呼び出す時に使用 |
| datetime |  | DBへ日付挿入時に使用 |
| requests |  | 応答速度取得に使用 |

実験システムのうちロードバランサはNGINXとラズベリーパイを用いて構築した．

ラズベリーパイのOSのインストール手順を下記に示す．

１．公式サイトに配布されている「Raspberry Pi imager」をダウンロードする。

https://www.raspberrypi.com/software/

２．「Raspberry Pi imager」を起動後，マイクロSDカードをPCに挿入する。「CHOOS STORAGE」ボタンを選択するとSDカードを選択できる。

３．「CHOOS OS」を選択し「Raspberry Pi Os(32-bit)」を選択する。

４．「WRITE」を選択し，インストールを開始する。

５．書き込み後Raspberry PiにSDカードを挿入し電源を入れて起動する。

Nginxをラズベリーパイにインストール手順を下記に示す．

表　パッケージ一覧を更新

|  |
| --- |
| $ sudo apt update |

リポジトリ追加・削除時には必ず実行しなければならない。

表〇　パッケージを更新

|  |
| --- |
| $ sudo apt upgrade |

パッケージを更新しNGINXのインストール準備をする。

表〇

|  |
| --- |
| $ sudo apt install nginx |

NGINXをインストールする。

ロードバランサの設計・開発ではPythonの統合開発環境である「spider」をWindows上で利用し，テラタームを用いてラズベリーパイに転送した。本プロジェクトファイルの各種機能は下記に示す．

表　プロジェクトファイルの各機能

|  |  |
| --- | --- |
| ファイル名 | 機能説明 |
| measure\_evaluation\_InsertDB.py | 計測、評価、DB挿入プログラム |
| response | データベースの |
| roop\_LoadBalancer.sh | 2つのプログラムを同時に動かしループさせる |
| nginxrestart.sh | NGINXの設定を再読み込みする |
| LoadBalancer.py | 設計したロードバランサ |

### コンフィグ書き換え

|  |
| --- |
| f = open('/etc/nginx/nginx.conf', 'w', encoding='UTF-8')  datalist1 = ['user www-data;\n','worker\_processes auto;\n','pid /run/nginx.pid;\n','include /etc/nginx/modules-enabled/\*.conf;\n','events {\n',' worker\_connections 768;\n','}\n','http {\n',' include /etc/nginx/mime.types;\n','gzip on;\n','gzip\_vary on;\n','upstream backend1{\n']  f.writelines(datalist1)  if minmam == list1[0]:#81が平均最速なら  print('\n過去24時間で「サーバ81」が一番応答速度が速いので優先接続されます。')  datalist2 = ['server 192.168.1.81 weight=10;\n server 192.168.1.82 weight=1;\n server 192.168.1.83 weight=1;\n']  f.writelines(datalist2)      elif minmam == list2[0]:#82が平均最速なら  print('\n過去24時間で「サーバ82」が一番応答速度が速いので優先接続されます。')  datalist2 = ['server 192.168.1.81 weight=1;\n server 192.168.1.82 weight=10;\n server 192.168.1.83 weight=1;\n']  f.writelines(datalist2)    elif minmam == list3[0]:#82が平均最速なら  print('\n過去24時間で「サーバ83」が一番応答速度が速いので優先接続されます。')  datalist2 = ['server 192.168.1.81 weight=1;\n server 192.168.1.82 weight=1;\n server 192.168.1.83 weight=10;\n']  f.writelines(datalist2) |

コンフィグの重みづけを書き換えることでロードバランサの割り振り方法を変更している。ここでは共通の条件をそのまま書き込み24時間のサーバ速度の平均で最も高ければ優先して接続されることになる。優先以外の接続は12回に1回接続されることになる。

### コンフィグを適用させる

表〇　subprocessを利用しnginxrestartのシェルを呼び出す

|  |
| --- |
| subprocess.run(['/home/pi/tools/nginxrestart.sh']) |

表〇　nginxrestar.shの処理

|  |
| --- |
| #! /bin/sh  sudo systemctl reload nginx  exit 0 |

reloadというNGINXを再起動しなくても変化させたコンフィグを適用させる機能を利用する。

### DBの構築と接続

レスポンス速度や評価済みのデータを格納したDBはSQLITEを使用して保管している

下記に使用したデータベースのテーブル設計を示す．

### 表テーブル設計response81,82,83

|  |  |
| --- | --- |
| データ名 | データ形式 |
| ID | integer primary key |
| Speed | read |
| **Datetime** | text |

毎分計測した応答速度をテーブルresponse81,82,83で保管している。

### 表テーブル設計monitoring81,82,83

|  |  |
| --- | --- |
| データ名 | データ形式 |
| ID | integer primary key |
| now\_speed | read |
| now\_speed\_score | read |
| Ave\_speed | Read |
| Ave\_speed\_score | Read |
| datetime | Text |

評価システムで評価された応答速度はテーブルmonitoring81,82,83に保管される。

## 計測・評価システムの構築

応答速度を計測し評価するシステムはロードバランサとは別のプログラムで動いている。

表〇　計測した値を評価する関数

|  |
| --- |
| def eva\_score(list):    #抽出した平均を評価  if 0<=list1[0] and list1[0]<=0.016:  evalution="S"  elif 0.017<=list1[0] and list1[0]<=0.099:  evalution="A"  elif 0.1<=list1[0] and list1[0]<=0.999:  evalution="B"  elif 1.0<=list1[0] and list1[0]<=9.999:  evalution="C"  elif 10.0<=list1[0]:  evalution="D"  return evalution |

表〇　応答速度を計測する処理

|  |
| --- |
| #応答速度計測  url = 'http://192.168.1.81'  res = requests.get(url)  time\_elapsed = res.elapsed.total\_seconds() |

requests.get()で要求を送信してから応答が到着するまでに経過した時間を計測している。ユーザがチェックしやすいよう秒に変換している

表〇　データベースＳＱＬＩＴＥに接続

|  |
| --- |
| #データベースＳＱＬＩＴＥに接続  dbname = 'response'  conn = sqlite3.connect(dbname) |

表〇　応答速度テーブルへ計測結果を挿入

|  |
| --- |
| setdb1 = (None, time\_elapsed,dt\_now)  cur.execute("insert into response81 (id,speed,datetime) values (?,?,?)", setdb1) |

表〇　応答速度テーブルから過去24時間の平均を取り出す

|  |
| --- |
| cur.execute("select avg(speed) from response81 where datetime > datetime(datetime(), '-1 days', '+0 hours');")  list1 = cur.fetchone()  evaluation\_ave=eva\_score(list1) |

## ロードバランサの実行

ロードバランサを実行するには表〇で示したプログラムのうち「measure\_evaluation\_InsertDB.py」と「measure\_evaluation\_InsertDB.py」を同時に実行する必要がある。

そこで「roop\_LoadBalancer.sh」

## 検索システムの構築

提案システムのプロトタイプを実装した開発環境を表4-1に示す．提案システムを作成するために用いたパッケージを表4-2に示す．

表 4‑1　実験システムの開発環境

# **実験**

本章では，提案システムを利用した比較実験について述べる．

## 実験目的

．

## 実験環境

# **むすび**

# **謝辞**

# **参考文献**

# その他

* ソースコード等を記述するときは，表内に挿入し，行間を「固定値：12pt」に設定する．行間を制する者はwordを制する．
* 研究でやったことはすべて，基本的に「本研究では」と称する
* ファイル名は，「1421~-name-thasis.docx」で統一する．
* 目次では、謝辞や参考文献のページ番号は載せない
* 要旨には，背景，課題，提案，実験で確認したことを書く
* 図表番号とラベルの間は半角スペースを入れる
* 節と節の間は改行する．章と章は改ページする
* 本文は日本語：MS明朝　　英語：Cenutty
* ただし数式は「挿入→数式」から入力し，本文中で数式を用いる場合はフォントの形が似ているTimes New Romanを使う．
* 付録にあるプログラムの説明がしたいとき本文の方に，例えば，“（付録３参照）”，“……を付録３に示す．”等と参 照先を記しておく．
* 章の終わりに改ページを行う．
* 改ページはCtrl+Enter
* 今後の課題ではなく展望．修正点とか書かない