

IT17

Aqila Aqsa - 5027211032

Mutiara Nurhaliza - 5027221010

Markas pusat meminta laporan hasil benchmark dengan menggunakan apache benchmark dari load balancer dengan 2 web server yang berbeda tersebut dan meminta secara detail dengan ketentuan:

- Nama Algoritma Load Balancer
- Report hasil testing apache benchmark
- Grafik request per second untuk masing masing algoritma.
- Analisis

Nginx Load Balancer

```
Server Software:      nginx/1.10.3
Server Hostname:      10.72.2.4
Server Port:          80

Document Path:        /
Document Length:      171 bytes

Concurrency Level:     10
Time taken for tests:  0.184 seconds
Complete requests:     1000
Failed requests:       500
  (Connect: 0, Receive: 0, Length: 500, Exceptions: 0)
Total transferred:    5941500 bytes
HTML transferred:     5746000 bytes
Requests per second:   5437.29 [#/sec] (mean)
Time per request:      1.839 [ms] (mean)
Time per request:      0.184 [ms] (mean, across all concurrent requests)
Transfer rate:         31548.52 [Kbytes/sec] received

Connection Times (ms)
  min  mean[+/-sd] median  max
Connect:    0    0  0.1    0    1
Processing:  1    2  0.5    2    5
Waiting:    1    2  0.5    2    4
Total:      1    2  0.6    2    6

Percentage of the requests served within a certain time (ms)
 50%    2
 66%    2
 75%    2
 80%    2
 90%    2
 95%    3
 98%    4
 99%    4
100%    6 (longest request)
```

Algoritma: Round Robin

Hasil Testing Apache Benchmark:

Analisis Kinerja Server:

- Server menggunakan Nginx versi 1.10.3.
- Dilakukan pengujian pada alamat IP server dengan port 80.
- Pada pengujian ini, dilakukan 1000 permintaan (request) dengan tingkat konkurensi (concurrency level) sebesar 10.

- Waktu yang dibutuhkan untuk melakukan pengujian adalah 0.184 detik.
- Dari 1000 permintaan, terdapat 500 permintaan yang gagal, semuanya karena panjang respons yang tidak sesuai (Length: 500).
- Total data yang ditransfer adalah 5941500 bytes, dengan 5746000 bytes berupa HTML.
- Rata-rata permintaan per detik adalah 5437.29 [#sec].
- Waktu rata-rata yang dibutuhkan untuk menangani setiap permintaan adalah 1.839 ms.
- Waktu rata-rata untuk menangani setiap permintaan, diukur secara keseluruhan, adalah 0.184 ms.
- Tingkat transfer rata-rata adalah 31548.52 Kbytes/detik.

Waktu Koneksi:

Waktu minimum yang dibutuhkan untuk koneksi adalah 0 ms, dengan waktu maksimum 1 ms.

Waktu pemrosesan (processing) terjadi dalam rentang 1 hingga 5 ms.

Waktu menunggu (waiting) sebesar 1 hingga 4 ms.

Waktu total yang dibutuhkan untuk koneksi dan pemrosesan berkisar dari 1 hingga 6 ms.

Persentase Waktu Respon:

50% permintaan dilayani dalam waktu 2 ms.

66% permintaan dilayani dalam waktu 2 ms.

75% permintaan dilayani dalam waktu 2 ms.

80% permintaan dilayani dalam waktu 2 ms.

90% permintaan dilayani dalam waktu 2 ms.

95% permintaan dilayani dalam waktu 3 ms.

98% permintaan dilayani dalam waktu 4 ms.

99% permintaan dilayani dalam waktu 4 ms.

Permintaan terlama membutuhkan waktu 6 ms.

Analisis:

Requests per Second (RPS): Dengan rata-rata 5437.29 permintaan per detik, server mampu menangani lalu lintas dengan tingkat yang sangat tinggi. Ini menunjukkan kemampuan server untuk menangani beban lalu lintas yang besar dalam periode waktu yang singkat.

Failed Requests: Dari 1000 permintaan yang dilakukan, 500 permintaan gagal. Ini menunjukkan tingkat kegagalan yang tinggi, dengan setengah dari permintaan mengalami kegagalan. Kegagalan ini terutama disebabkan oleh panjang respons yang tidak sesuai, yang mengindikasikan adanya masalah pada konfigurasi server atau aplikasi.

Time per Request: Waktu rata-rata yang dibutuhkan untuk menangani setiap permintaan adalah 1.839 ms. Ini menunjukkan respons yang cepat dari server terhadap permintaan yang diterima.

Connection Times: Waktu koneksi dan pemrosesan yang relatif rendah (dari 1 hingga 6 ms) menunjukkan bahwa server dapat menangani permintaan dengan efisien dan cepat.

Percentage of Requests Served within a Certain Time: Sebagian besar permintaan dilayani dalam waktu yang sangat singkat, dengan 90% permintaan dilayani dalam 2 ms atau kurang. Ini menunjukkan respons yang konsisten dan cepat dari server.

Cara Kerja Algoritma Round Robin:

Algoritma Round Robin digunakan untuk mendistribusikan lalu lintas secara merata di antara beberapa server backend. Cara kerjanya adalah sebagai berikut:

- Setiap permintaan yang diterima oleh server load balancer akan dialihkan secara bergantian (secara berputar) ke setiap server backend yang tersedia.
- Saat satu server backend menerima permintaan, server load balancer akan mengalihkan permintaan berikutnya ke server backend berikutnya dalam daftar.
- Proses ini akan terus berlanjut, sehingga setiap server backend memiliki kesempatan yang sama untuk melayani permintaan.
- Dengan demikian, algoritma Round Robin memastikan bahwa beban lalu lintas terdistribusi secara merata di antara semua server backend yang terhubung.

Apache Load Balancer

```
Server Software:      Apache/2.4.18
Server Hostname:      10.72.2.4
Server Port:          80

Document Path:        /
Document Length:      166 bytes

Concurrency Level:     10
Time taken for tests:  2.436 seconds
Complete requests:    1000
Failed requests:       334
  (Connect: 0, Receive: 0, Length: 334, Exceptions: 0)
Total transferred:    357334 bytes
HTML transferred:     166334 bytes
Requests per second:  410.52 [#/sec] (mean)
Time per request:      24.359 [ms] (mean)
Time per request:      2.436 [ms] (mean, across all concurrent requests)
Transfer rate:         143.26 [Kbytes/sec] received

Connection Times (ms)
      min      mean[+/-sd] median    max
Connect:    0        0  0.1      0      2
Processing:  0       12 144.2      1    2384
Waiting:    0       12 144.2      1    2384
Total:      0       12 144.2      1    2384

Percentage of the requests served within a certain time (ms)
 50%      1
 66%      1
 75%      1
 80%      1
 90%      1
 95%      1
 98%      7
 99%     21
100%    2384 (longest request)
```

Report Hasil Testing Apache Benchmark:

Method ByRequests

Analisis Kinerja Server:

- Server menggunakan Apache versi 2.4.18.
- Dilakukan pengujian pada alamat IP server dengan port 80.
- Pada pengujian ini, dilakukan 1000 permintaan (request) dengan tingkat konkurensi (concurrency level) sebesar 10.
- Waktu yang dibutuhkan untuk melakukan pengujian adalah 2.436 detik.
- Dari 1000 permintaan, terdapat 334 permintaan yang gagal, semua karena panjang respons yang tidak sesuai (Length: 334).
- Total data yang ditransfer adalah 357334 bytes, dengan 166334 bytes berupa HTML.
- Rata-rata permintaan per detik adalah 410.52 [#sec].
- Waktu rata-rata yang dibutuhkan untuk menangani setiap permintaan adalah 24.359 ms.
- Waktu rata-rata untuk menangani setiap permintaan, diukur secara keseluruhan, adalah 2.436 ms.
- Tingkat transfer rata-rata adalah 143.26 Kbytes/detik.

Waktu Koneksi:

- Waktu minimum yang dibutuhkan untuk koneksi adalah 0 ms, dengan waktu maksimum 2 ms.
- Waktu pemrosesan (processing) terjadi dalam rentang 0 hingga 2384 ms.
- Waktu menunggu (waiting) sebesar 0 hingga 2384 ms.
- Waktu total yang dibutuhkan untuk koneksi dan pemrosesan berkisar dari 0 hingga 2384 ms.

Persentase Waktu Respon:

50% permintaan dilayani dalam waktu 1 ms.

66% permintaan dilayani dalam waktu 1 ms.

75% permintaan dilayani dalam waktu 1 ms.

80% permintaan dilayani dalam waktu 1 ms.

90% permintaan dilayani dalam waktu 1 ms.

95% permintaan dilayani dalam waktu 1 ms.

98% permintaan dilayani dalam waktu 7 ms.

99% permintaan dilayani dalam waktu 21 ms.

Permintaan terlama membutuhkan waktu 2384 ms.

Analisis:

Complete Requests dan Failed Requests: Jumlah permintaan yang berhasil (1000) dan yang gagal (334). Jadi kemungkinan ada masalah pengelolaan sumber daya.

Requests per Second: Rata-rata 410.52 permintaan per detik merupakan angka yang cukup tinggi.

Time per Request: Waktu rata-rata untuk menangani setiap permintaan adalah 24.359 ms, yang sebagian besar dihabiskan dalam proses pemrosesan. Waktu ini dapat diterima, tergantung pada jenis aplikasi dan tingkat kebutuhan respons instan.

Connection Times: Waktu koneksi dan pemrosesan yang berkisar dari 0 hingga 2384 ms menunjukkan variasi yang cukup besar dalam penanganan permintaan. Masih kurang stabil.

Percentage of Requests Served within a Certain Time: Sebagian besar permintaan dilayani dalam waktu sangat singkat (1 ms). Namun, adanya beberapa permintaan yang membutuhkan waktu lebih lama (hingga 2384 ms) menunjukkan kurangnya konsistensi.

Algoritma/Metode ByRequest:

Cara kerja algoritma "byrequests":

- Pemilihan Server: Setiap kali ada permintaan masuk, server load balancer akan memilih server backend yang akan melayani permintaan tersebut. Pemilihan ini didasarkan pada jumlah permintaan yang telah dilayani oleh masing-masing server.
- Perhitungan Permintaan: Server load balancer melacak jumlah permintaan yang telah dilayani oleh setiap server backend yang terhubung. Ini bisa dilakukan dengan mencatat

jumlah permintaan yang telah dilayani sejak server backend tersebut mulai diaktifkan, atau dalam periode waktu tertentu.

- Pemilihan Server Berdasarkan Jumlah Permintaan: Saat ada permintaan masuk, server load balancer akan memilih server backend yang telah melayani jumlah permintaan paling sedikit. Dengan kata lain, server dengan jumlah permintaan yang lebih rendah akan memiliki prioritas untuk melayani permintaan berikutnya.
- Pengalihan Permintaan: Permintaan kemudian akan dialihkan ke server backend yang dipilih. Ini dilakukan dengan meneruskan permintaan dari server load balancer ke server backend yang dipilih, sehingga server backend dapat memproses permintaan tersebut dan memberikan respons kembali ke klien.
- Penyimpanan dan Pembaruan Data: Setelah pemrosesan permintaan selesai, server load balancer akan memperbarui data jumlah permintaan yang telah dilayani oleh setiap server backend. Data ini digunakan untuk keputusan pemilihan server pada permintaan selanjutnya.