Dokumentacija

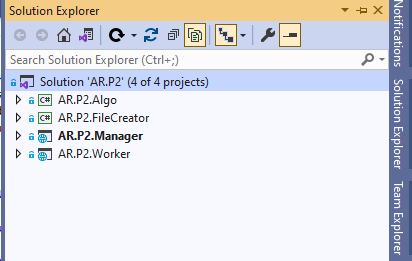
# Uvod

Rešenje ima za cilj implementaciju Fast Fourier Transform (FFT) algoritma. Konkretno, Cooley Tukey varijantu FFT[[1]](#endnote-1).

Osnovni skup zahteva za implementaciju je omogućiti SIMD optimizacije u programskom jeziku višeg nivoa i višeprocesorsku paralelizaciju algoritma, te kombinovanje ta dva pristupa i njihovo poređenje po brzini i propusnom opsegu pri obradi zahteva.

Rešenje se sastoji od sledećih projekata, gde ime svakog sadrži prefiks „AR.P2“

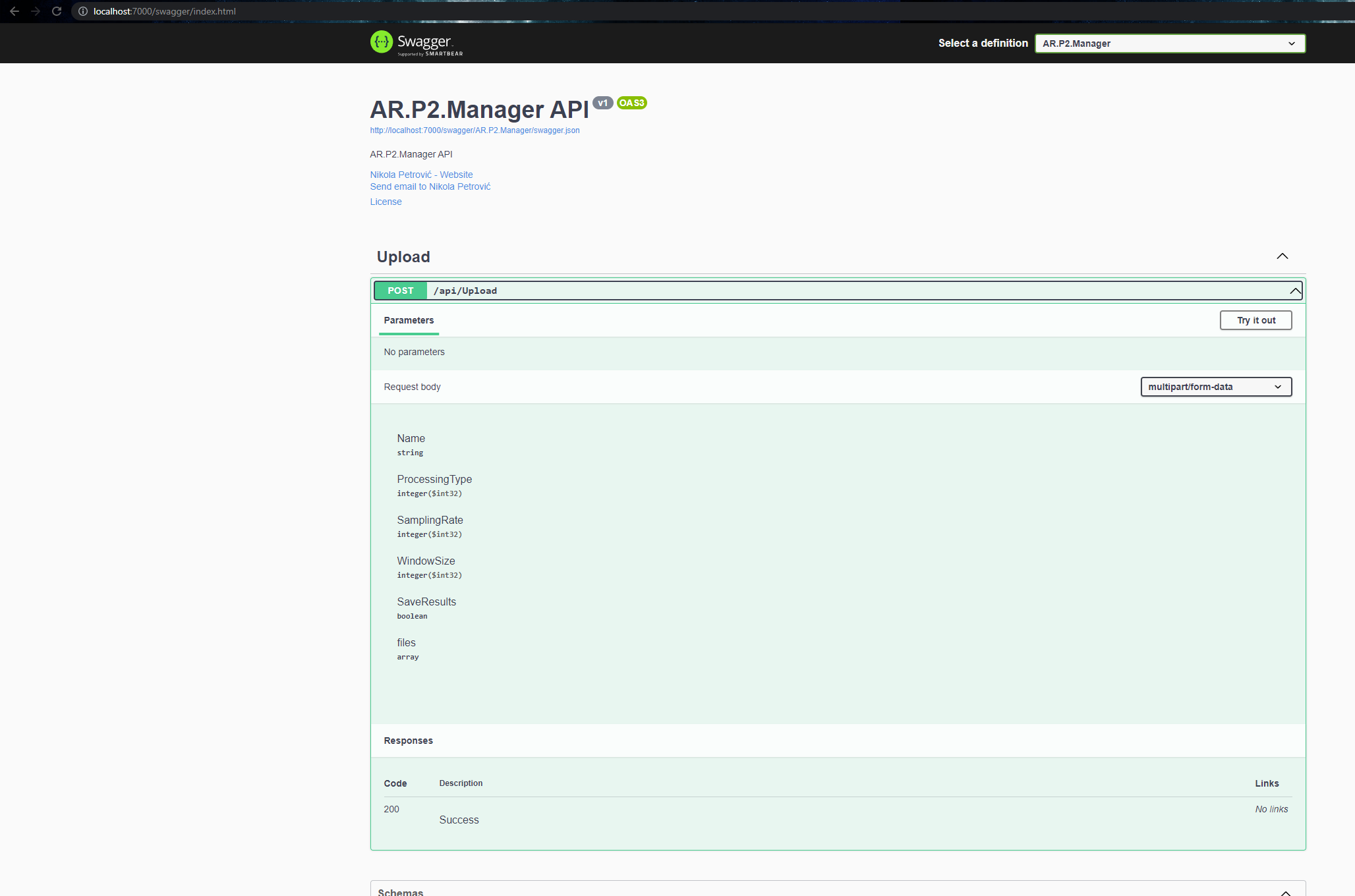
1. Algo – C# .NET 5.0 biblioteka; implementira varijante FFT algoritma
2. Docs – Dokumentacija rešenja
3. FileCreator – C# .NET CLI alat  
   Kreira fajlove floating point PCM signala koji se mogu proslediti Manager servisu za obradu
4. Manager – C# .NET 5.0 REST Web.API servis koji koristi Algo da bi obrađivao fajlove FFT algoritmom



Prikaz 1 Pregled rešenja

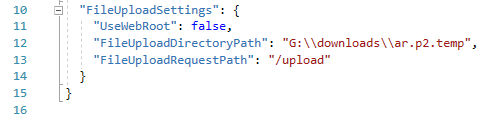
# Specifičnosti Implementacije

Manager REST API implementira Swagger[[2]](#endnote-2) varijantu OpenAPI standarda za dokumentovanje API endpointa. Swagger dokument je dostupan GET pozivom */swagger/index.html* relativno u odnosu na host adresu Manager servisa, te se može koristiti za ručno testiranje endpoint-a.



Prikaz 2 Swagger API dokumentacija

Moguće je konfigurisati direktorijum u koji se pohranjuju fajlovi iz zahteva promenom *apssettings.Development.json* fajla pre pokretanja Manager servisa, i takođe i relativnu putanju sa koje će Manager servis da servira fajlove koje poseduje.



Prikaz 3 Konfigurisanje Manager servisa

U scenarijima sa manje raspoložive radne memorije, moguće je dati instrukciju .NET Garbage Collector-u da bude konzervativniji sa alokacijom memorije i je češće oslobađa. To je moguće postići dodavanjem linije

*<ServerGarbageCollection>false</ServerGarbageCollection>*

u PropertyGroup od Project tag-a za Manager.csproj fajl.

Prilikom skupljanja metrika za dokumentaciju, ovo nije korišteno, te je memorijska iskorištenost visoka.

# Strategija benchmark-a

Za benchmark je korišten PowerShell. Verzija PowerShell-a je priložena u dodatku.

Strategija benchmark-a je napraviti 125 sekvencijalnih ili paralelnih zahteva, pri čemu svaki zahtev sadrži 4 fajla za obradu, pri čemu se modifikuje način obrade fajlova i njihova veličina između različitih sesija merenja benchmark-a.

Dva opsega veličine fajlova (svaki od 4 fajla iz zahteva pripada jednom opsegu) su korištena

1. 3-5MB
2. 340-345MB

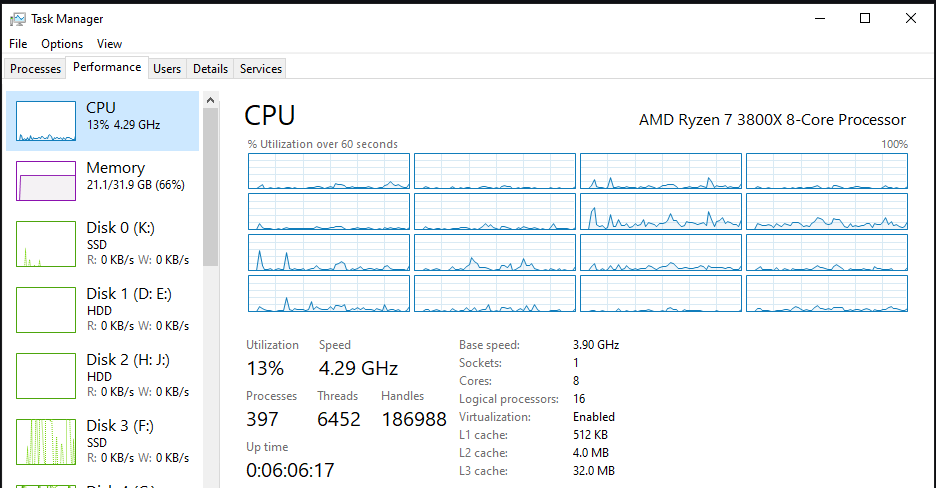
# Paralelizam zahteva

Osnovni slučaj paralelizacije je paraleizacija na nivou zahteva. Upotrebom PoweShell alata, moguće je slati zahteve Manager servisu, pri čemu slanje svakog narednog zahteva čeka završetak prethodnog zahteva.

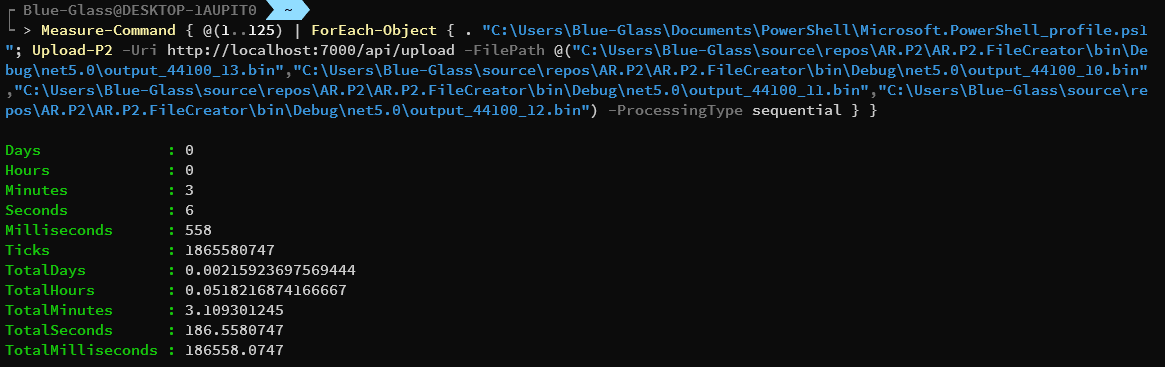
Za posmatranje paralelizma zahteva, korišteni su fajlovi u prvom opsegu.

## Osnovni slučaj

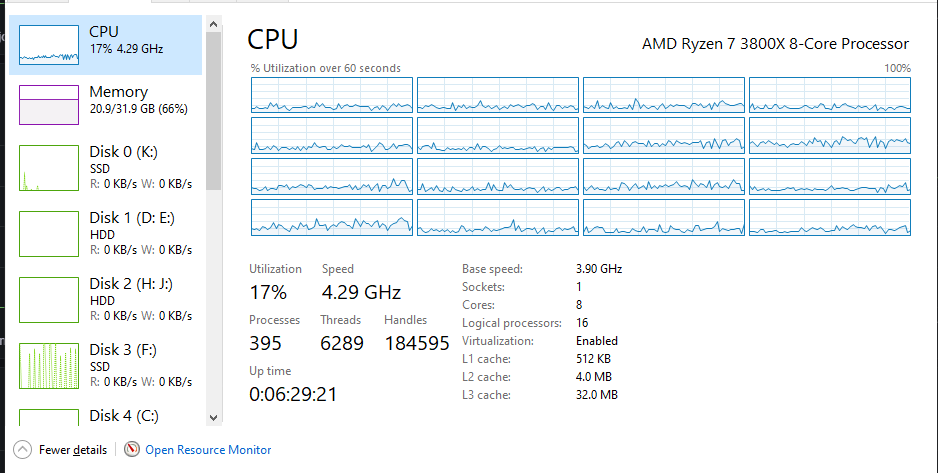
U osnovnoj varijanti, fajlovi prisutni u zahtevu se sekvencijalno upload-uju. Kada je završen upload svih fajlova, fajlovi se takođe sekvencijalno obrađuju.



Prikaz 4 Opterećenje “Idle” Baseline

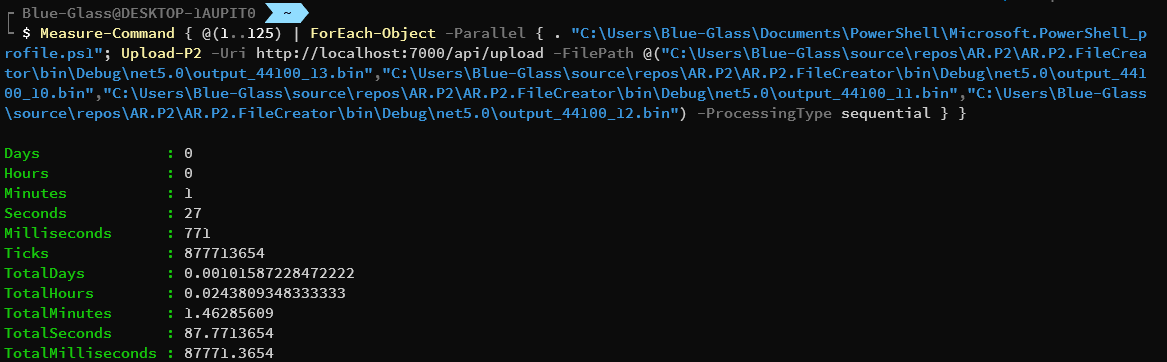


Prikaz 5 Osnovni slučaj – potpuno sekvencijalna obrada

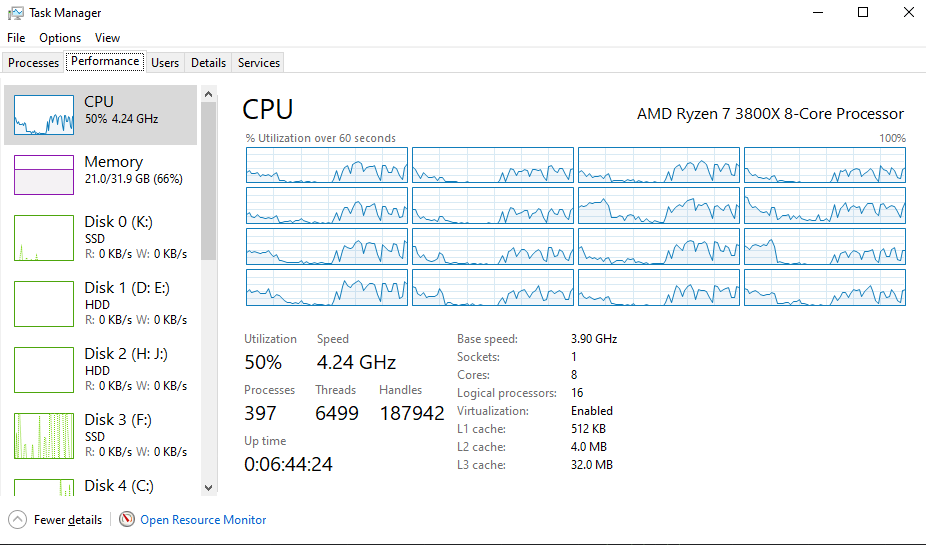


Prikaz 6 Oteprećenje - osnovni slučaj - potpuno sekvencijalna obrada

Vidimo da potpuno sekvencijalna obrada traje 3min 6s, pri čemu samo jedno od CPU jezgara uzima većinu opterećenja.



Prikaz 7 Osnovni slučaj - sekvencijalna obrada uz paralelizaciju zahteva



Prikaz 8 Opterećenje - Osnovna varijanta uz paralelizaciju zahteva

Paralelizacija zahteva je omogućena na nivou PowerShell alata i na nivou ASP.NET Framework-a, koji za svaki dolazeći zahtev instancira novi izvršni kontekst u okviru Task-a. To dvoje zajedno ne garantuje potpunu iskorištenost CPU-a, jer oni nužno ne impliciraju paralelizam, već konkurentnost.

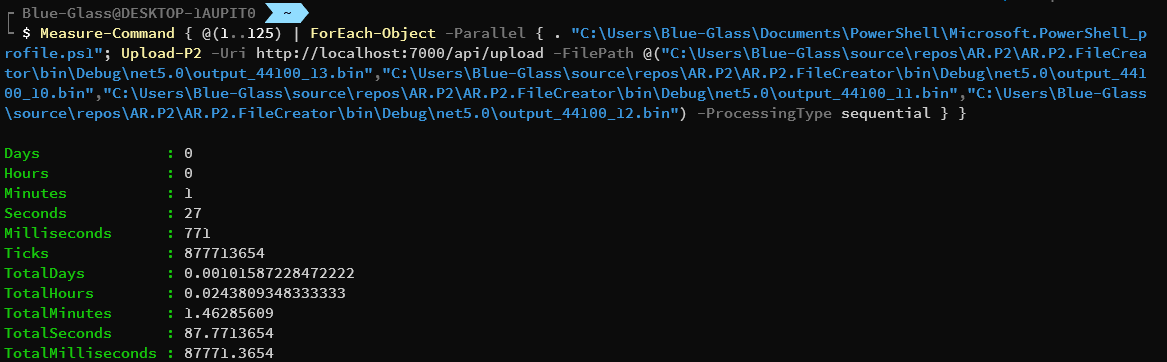
Pritom, vidimo da je obrada istih podataka trajala skoro tri puta kraće, te je moguće ostvariti je bez ikakve modifikacije koda u odnosu na osnovnu implementaciju.

Za ostala poređenja po brzini i propunsom opsegu, nadalje će se koristiti samo paralelizacija zahteva radi uniformnosti, te neće biti posebno naznačena u analizi. Za pretpostaviti je da će za fajlove iz istog ili većeg opsega overhead koji postoji pri ostvarivanju paralelizma zahteva biti zanemariv u odnosu na vreme procesuiranja fajlova.

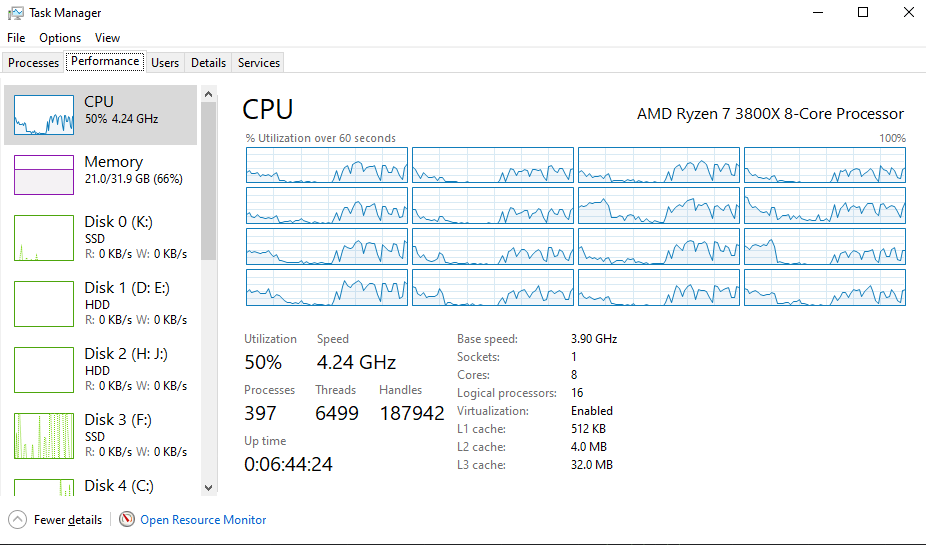
# Poređenje po brzini i propusnom opsegu – fajlovi srednje veličine

Podaci na graficima metrika su usrednjeni na intervalu od 3s.

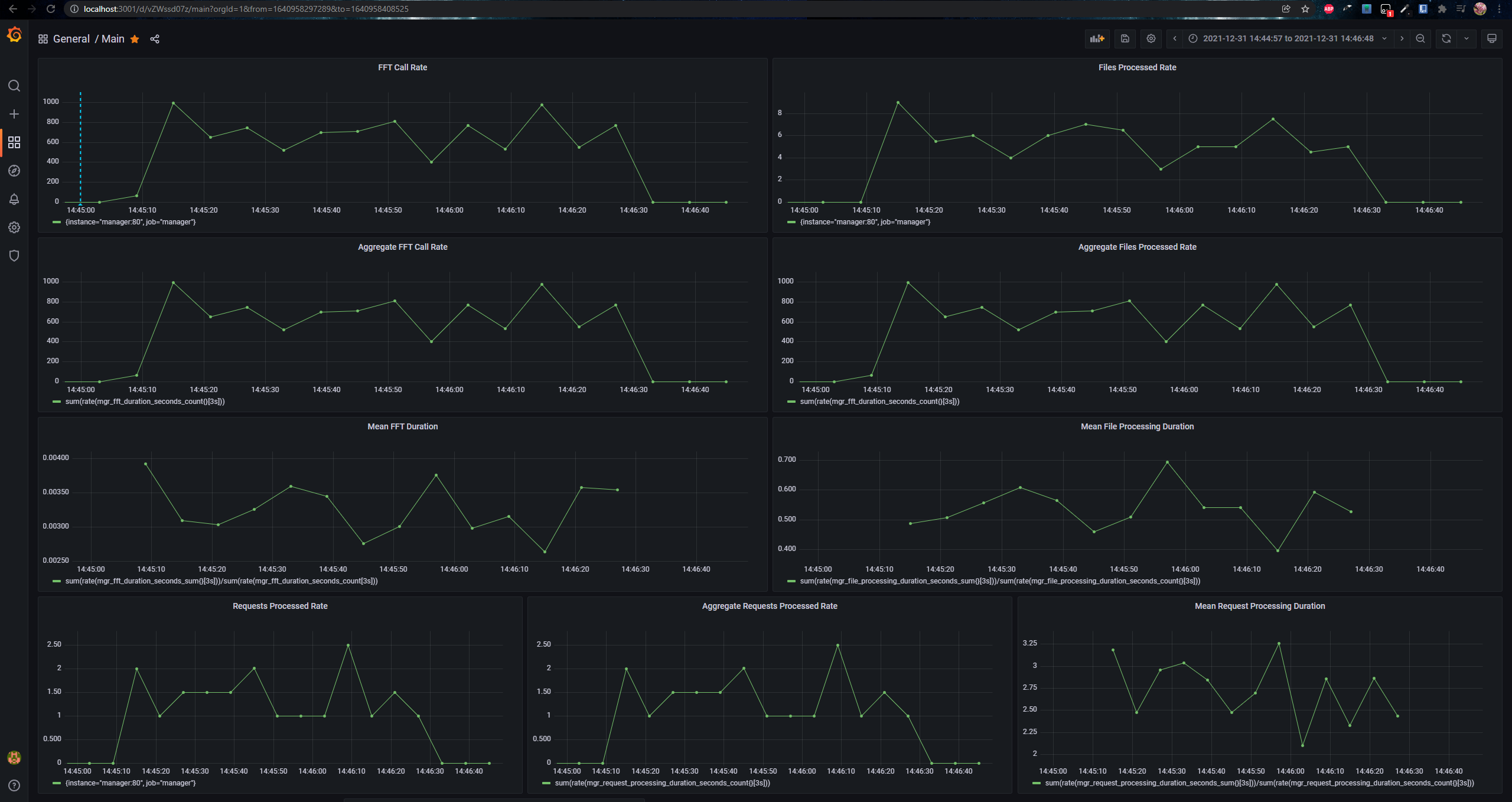
## Osnovna varijanta



Prikaz 9 Osnovna varijanta



Prikaz 10 Opterećenje - Osnovna varijanta uz paralelizaciju zahteva



Prikaz 11 Metrike - Osnovni slučaj

Na grafu broja obrađenih zahteva u sekundi i na grafu opterećenja CPU-a primećujemo udubljenja i ispupčenja koja su jednim delom rezultat IO rada. IO rad koji se dešava je pohranjivanje fajlova pri upload-u, učitavanje pohranjenih fajlova sa diska u memoriju – te pri IO-bound operacijama imamo slabiju iskorištenost CPU-a. To možemo obrazložiti time da pri 125 zahteva od po 4 fajla, broj izvršnih jedinica procesora i .NET Thread Pool niti (iako su neke privremeno oslobođene zbog async i await poziva) je još uvek značajno manji od broja dolazećih fajlova koji zahtevaju IO, tako da se brzo kreiraju niti koje čekaju IO.

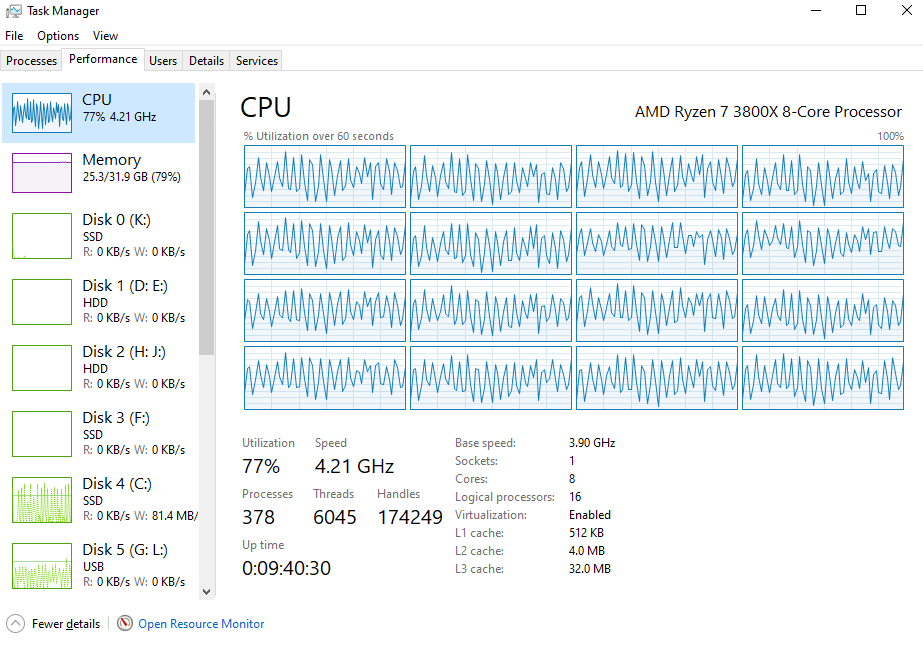
|  |  |
| --- | --- |
| Metrika | Opseg, srednja vrednost |
| Broj poziva FFTRecurse metode [1/s] | [404, 995], ~ avg. 700 |
| Srednje trajanje FFTRecurse poziva [ms] | [2.64, 3.92], ~avg. 3.25 |
| Broj obrađenih fajlova u sekundi | [3, 9], ~avg. 6 |
| Srednje trajanje obrade fajlova [ms] | [396, 694], ~avg. 550 |
| Broj obrađenih zahteva u sekundi | [1, 2.5], ~avg 1.4 |
| Srednje trajanje obrade zahteva [s] | [2.10, 3.26], ~avg. 2.60 |

## Paralelno obrađivanje

Pri paralelnom obrađivanju, fajlovi se konkurentno upload-uju i konkurentno započinje njihova pojedinačna obrada, te u okviru obrade pojedinačnih fajlova opet imamo konkurentnu obradu na nivou pojedinačnih „prozora“ FFT algoritma.



Prikaz 12 Paralelno obrađivanje



Prikaz 13 Opterećenje – paralelno obrađivanje

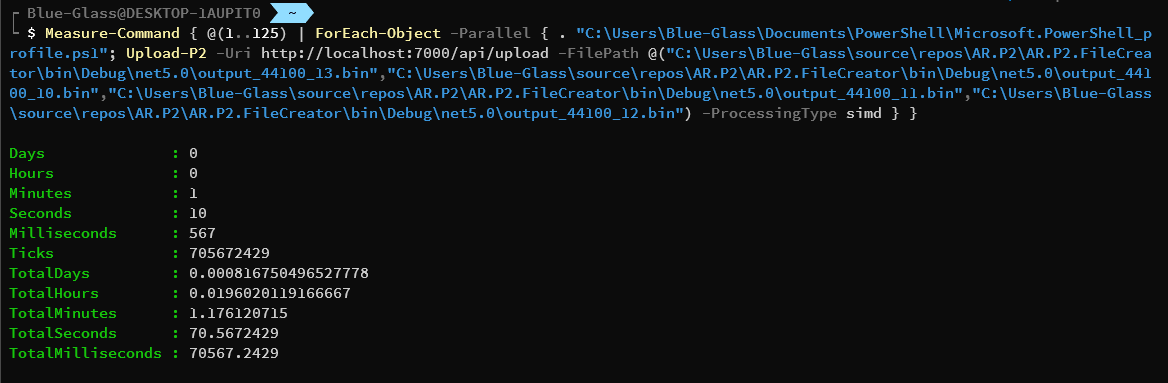


Prikaz 14 Metrike – paralelno obrađivanje

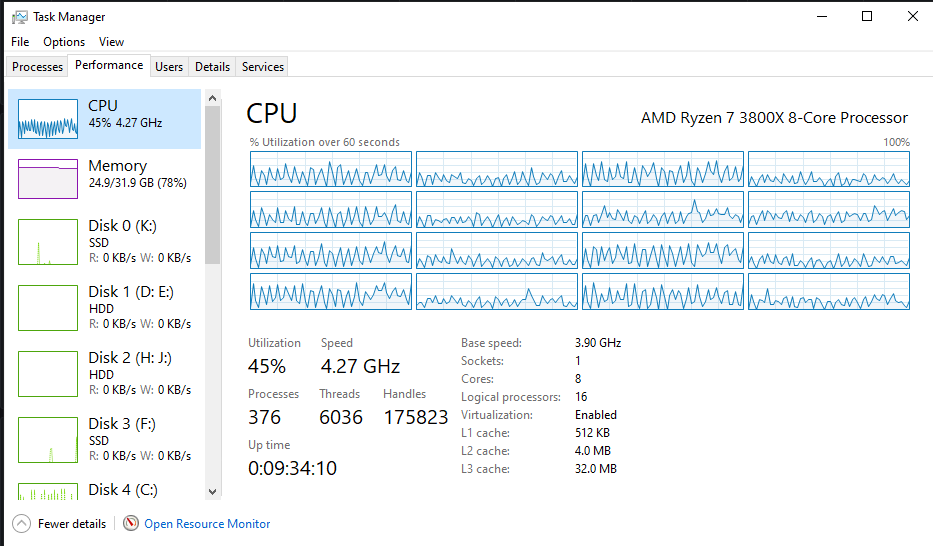
|  |  |
| --- | --- |
| Metrika | Opseg, srednja vrednost |
| Broj poziva FFTRecurse metode [1/s] | [227, 1232], ~ avg. 750 |
| Srednje trajanje FFTRecurse poziva [ms] | [3.06, 10.3], ~avg. 7.42 |
| Broj obrađenih fajlova u sekundi | [2, 11.5], ~avg. 7.5 |
| Srednje trajanje obrade fajlova [ms] | [349, 869], ~avg. 760 |
| Broj obrađenih zahteva u sekundi | [0.5, 3], ~avg 1.7 |
| Srednje trajanje obrade zahteva [s] | [0.825, 2.93], ~avg. 2 |

Na grafu opterećenja procesora vidimo izražene cikluse većeg intenziteta nego u osnovnoj varijanti. Ovo je rezultat paralelizacije na nivou ulaznih fajlova i na nivou pojedinačnih fajlova. Za jedan fajl od 10s sa sampling rate-om od 44100 imamo 441000 uzoraka, pa primenom FFT prozora od 4096 uzoraka imamo 107 prozora za izračunati, odnosno 107 poziva FFTRecurse metodi u zasebnom Task-u. Dakle, velik broj context-switch-eva između različitih niti, te shodno tome i graf iznad.

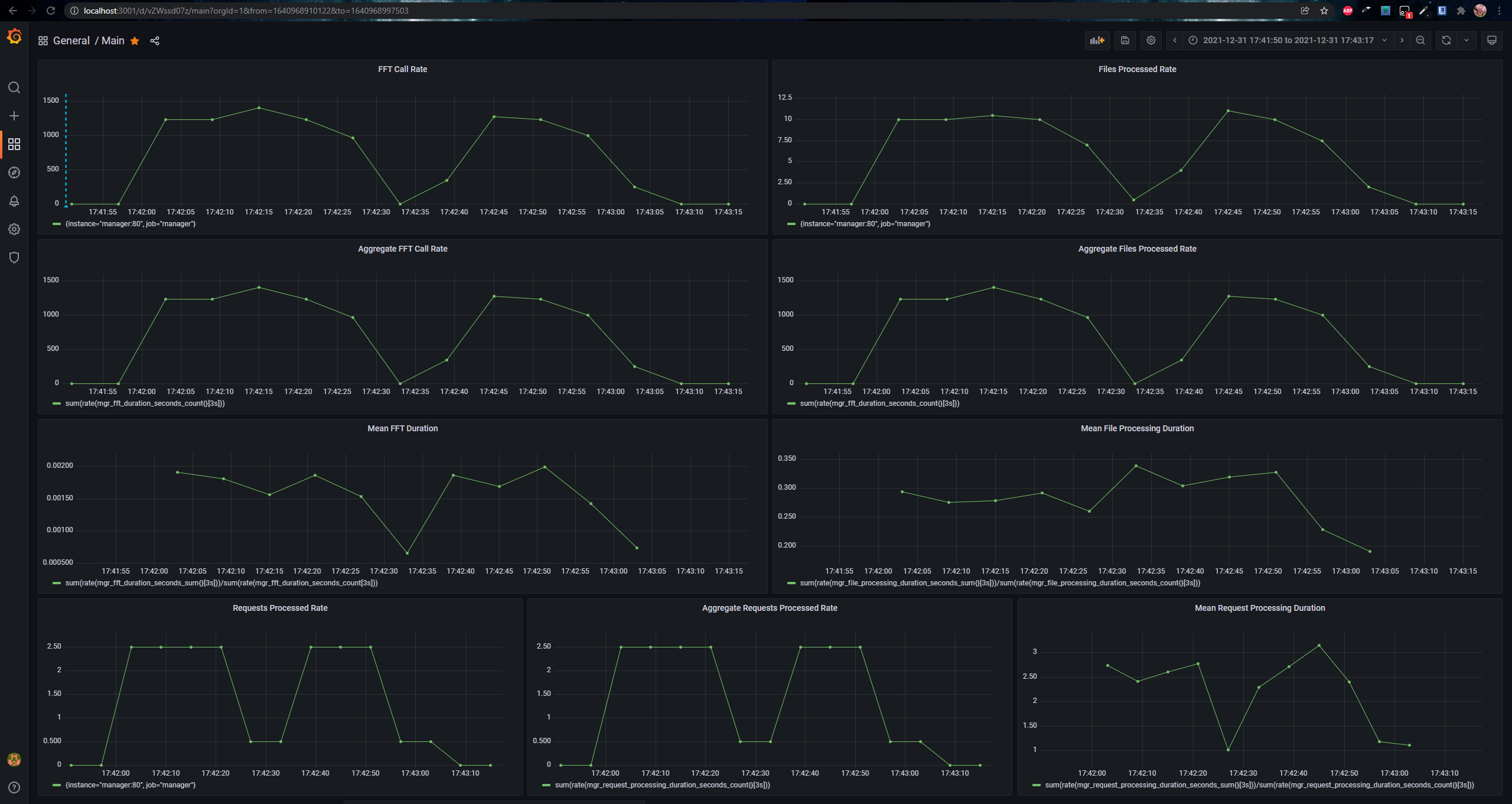
## SIMD obrađivanje



Prikaz 15 SIMD obrađivanje



Prikaz 16 Opterećenje - SIMD obrađivanje



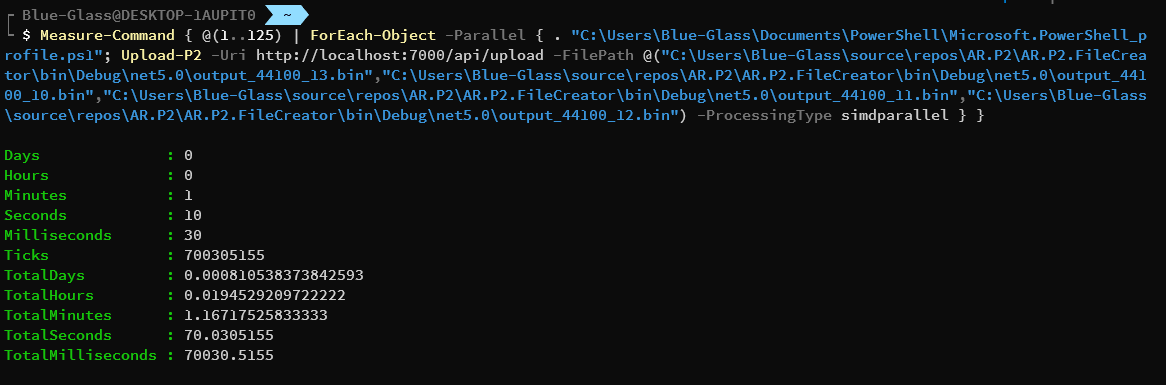
Prikaz 17 Metrike - SIMD obrađivanje

|  |  |
| --- | --- |
| Metrika | Opseg, srednja vrednost |
| Broj poziva FFTRecurse metode [1/s] | [0.501, 1403], ~ avg. 960 |
| Srednje trajanje FFTRecurse poziva [ms] | [0.655, 1.99], ~avg. 1.45 |
| Broj obrađenih fajlova u sekundi [1/s] | [0.501, 11], ~avg. 7.3 |
| Srednje trajanje obrade fajlova [ms] | [191, 339], ~avg. 260 |
| Broj obrađenih zahteva u sekundi [1/s] | [0.5, 2.5], ~avg 1.7 |
| Srednje trajanje obrade zahteva [s] | [1.01, 3.15], ~avg. 2.30 |

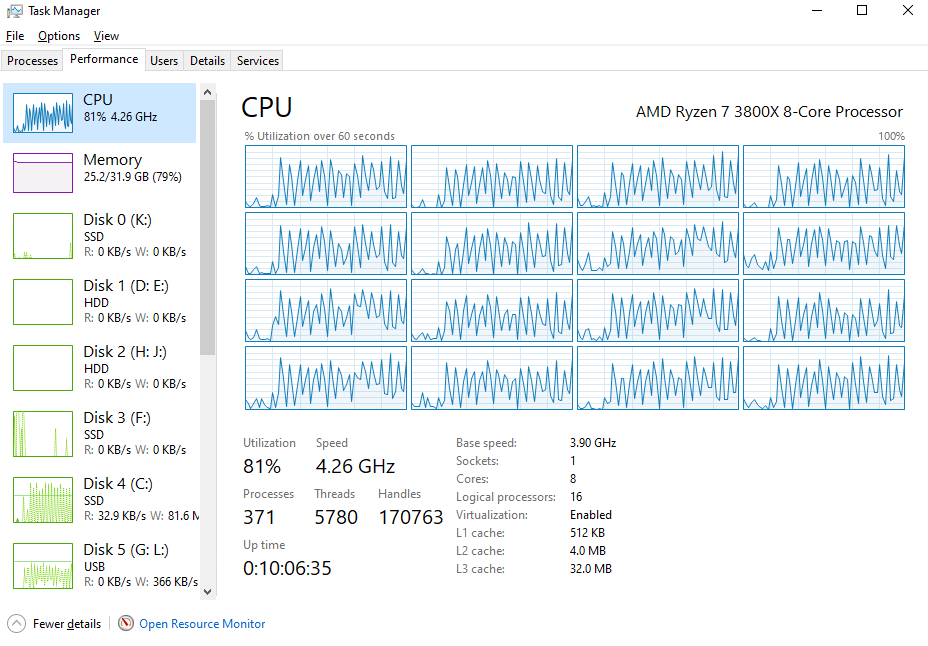
Primetimo asimetrično opterećenje među procesorima pri SIMD obrađivanju.

Pojedinačni FFTRecurse pozivi traje kraće od prethodnih varijanti - skoro upola, što omogućava veći broj FFTRecurse poziva u sekundi. Interesantno je da je SIMD varijanta kraće trajala od paralelne varijante, ali to može biti rezultat standarne devijacije trajanja IO operacija.

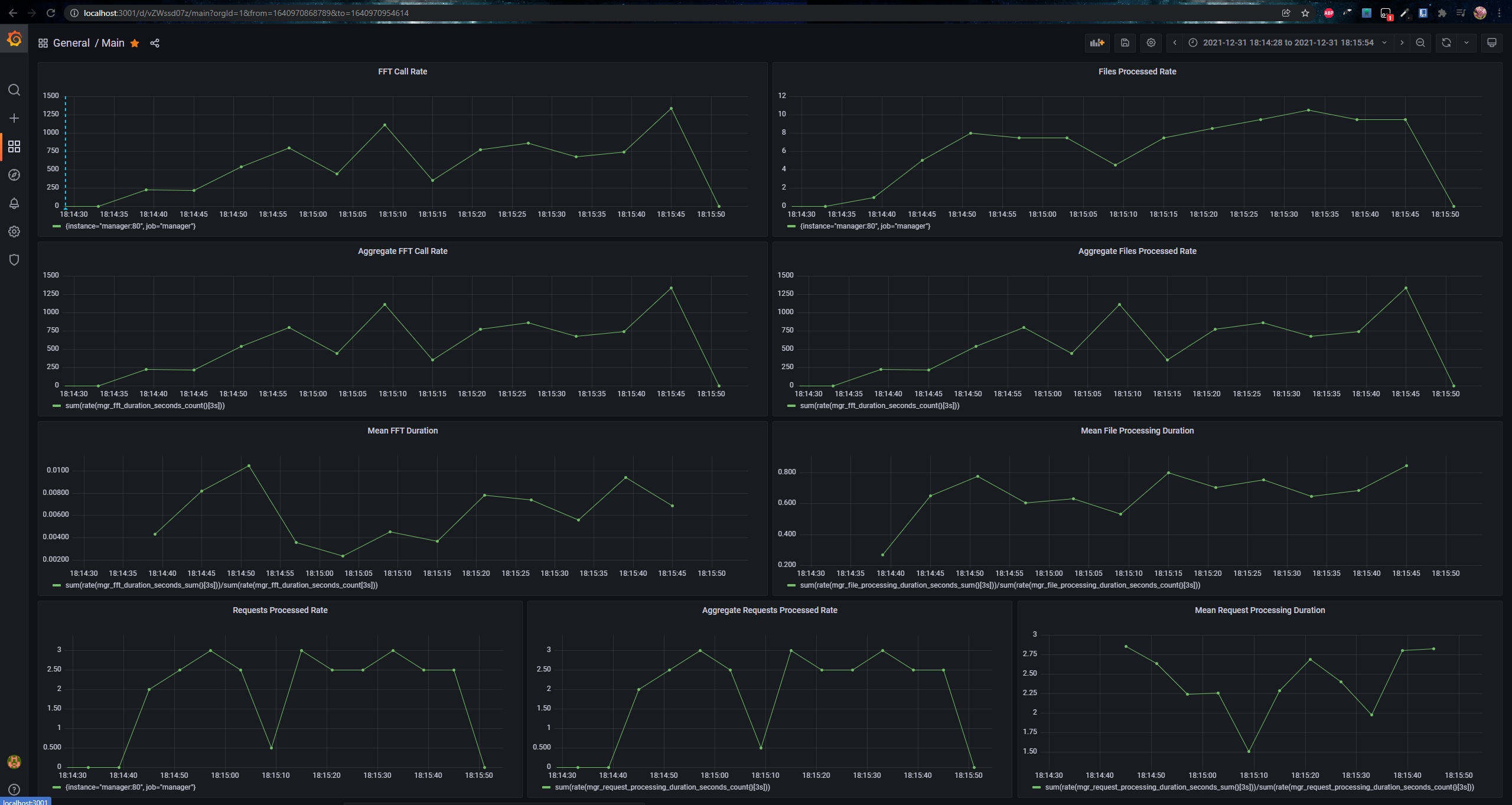
## SIMD paralelno obrađivanje



Prikaz 18 SIMD paralelno obrađivanje



Prikaz 19 Opterećenje - SIMD paralelno obrađivanje



Prikaz 20 Metrike - SIMD paralelno obrađivanje

|  |  |
| --- | --- |
| Metrika | Opseg, srednja vrednost |
| Broj poziva FFTRecurse metode [1/s] | [219, 1340], ~ avg. 770 |
| Srednje trajanje FFTRecurse poziva [ms] | [2.36, 10.5], ~avg. 7.5 |
| Broj obrađenih fajlova u sekundi [1/s] | [1, 10.5], ~avg. 7.5 |
| Srednje trajanje obrade fajlova [ms] | [267, 845], ~avg. 700 |
| Broj obrađenih zahteva u sekundi [1/s] | [0.5, 3], ~avg 2.25 |
| Srednje trajanje obrade zahteva [s] | [1.51, 2.85], ~avg. 2.26 |

Od prethodnih varijanti, ova se pokazala najbržom, što se može ogledati i u povećanoj iskorištenosti CPU-a, ali još uvek je značajan broj context-switch-ova i IO operacija, što je onemogućilo 100% iskorištenost CPU-a.

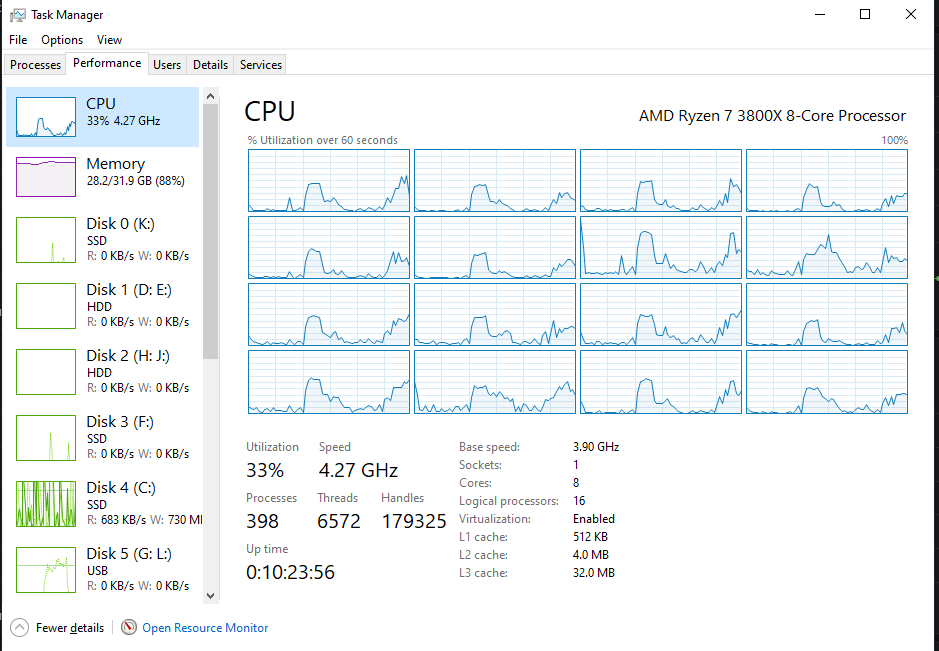
# Poređenje po brzini i propusnom opsegu – veliki fajlovi

Napravljena su dva poziva sa po dva fajla iz drugog opsega, te u tabelarnom prikazu su izostavljene metrike koje se odnose na broj obrađenih fajlova i zahteva u sekundi.

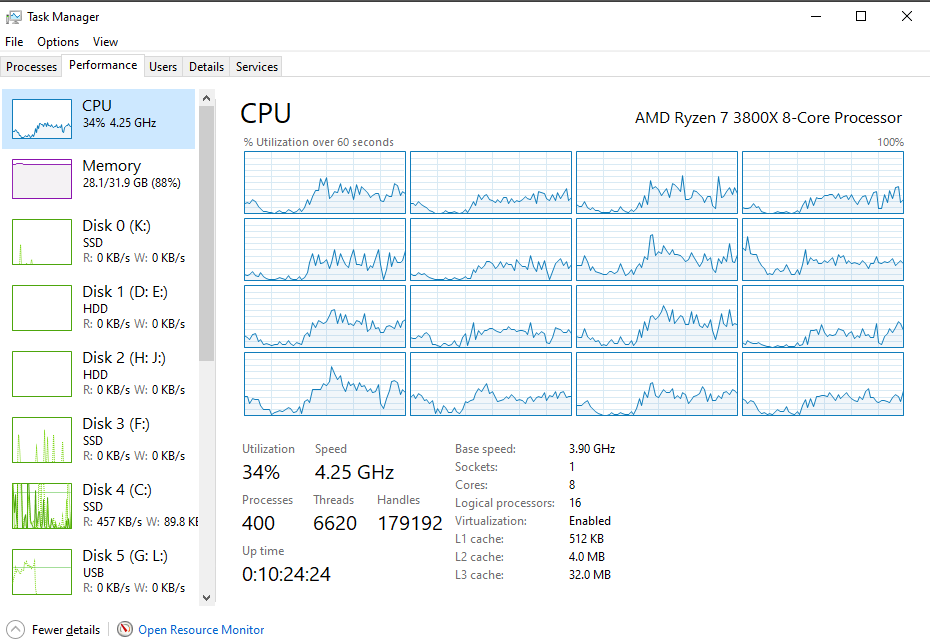
## Osnovna varijanta



Prikaz 21 Osnovna varijanta, veliki fajlovi



Prikaz 22 Opterećenje - osnovna varijanta, veliki fajlovi - upload



Prikaz 23 Osnovna varijanta, veliki fajlovi - obrađivanje

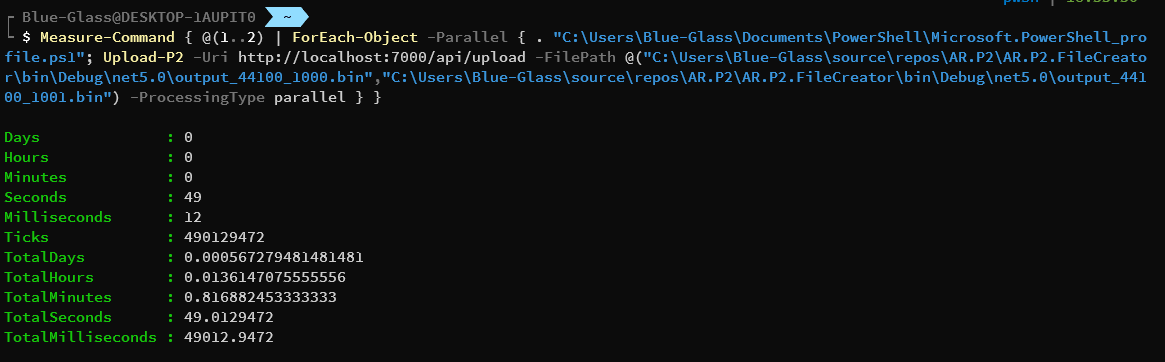
## 

Prikaz 24 Metrike - osnovna varijanta, veliki fajlovi

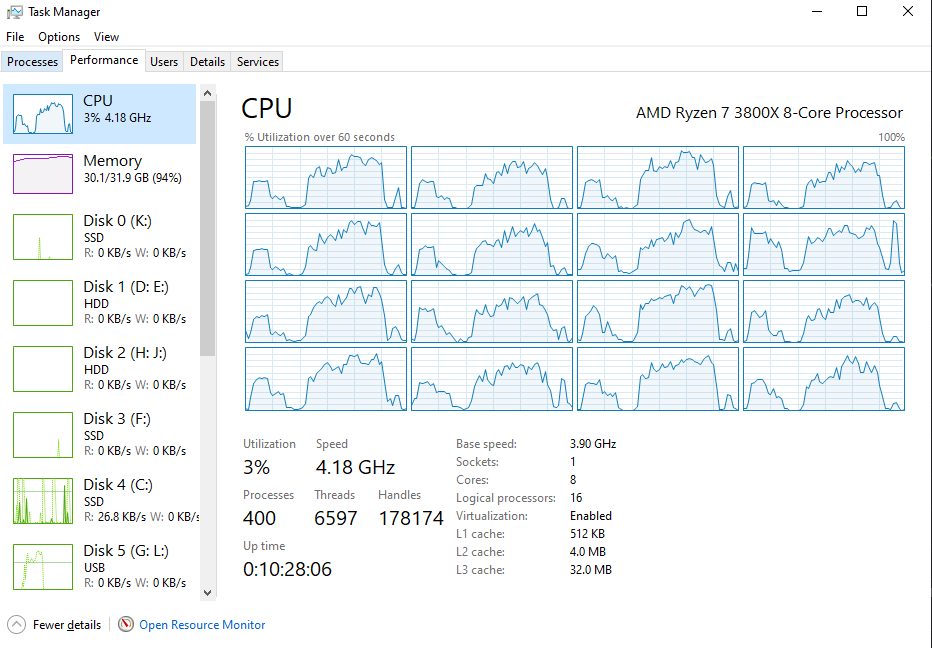
|  |  |
| --- | --- |
| Metrika | Opseg, srednja vrednost |
| Broj poziva FFTRecurse metode [1/s] | [123, 1006], ~ avg. 871 |
| Srednje trajanje FFTRecurse poziva [ms] | [1.84, 2.3], ~avg. 2.05 |

Pri upotrebi sekvencijalne varijante sa velikim fajlovima, performanse su najlošije u pogledu obrađenih zahteva i fajlova, jer je vreme potrebno za upload i obradu svakog fajla izraženo u odnosu na vreme obrade.

## Paralelno obrađivanje



Prikaz 25 Paralelno obrađivanje, veliki fajlovi



Prikaz 26 Opterećenje - paralelno obrađivanje, veliki fajlovi



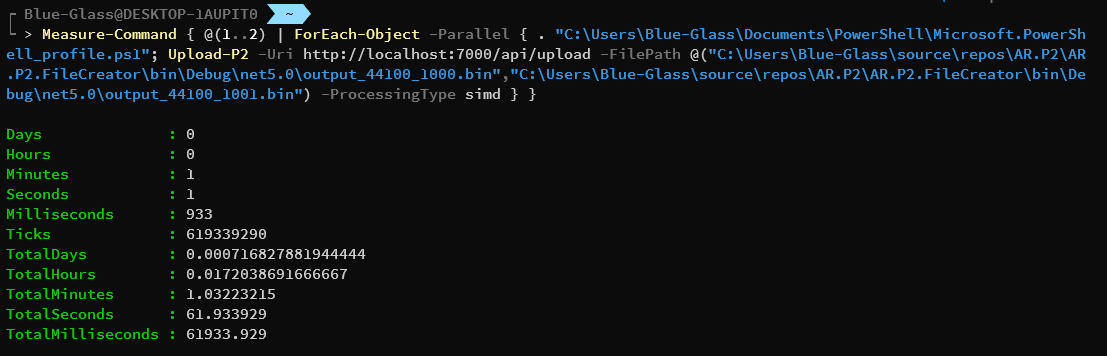
Prikaz 27 Metrike - Paralelno obrađivanje, veliki fajlovi

|  |  |
| --- | --- |
| Metrika | Opseg, srednja vrednost |
| Broj poziva FFTRecurse metode [1/s] | [959, 1636], ~ avg. 1300 |
| Srednje trajanje FFTRecurse poziva [ms] | [2.53, 7.18], ~avg. 5.50 |

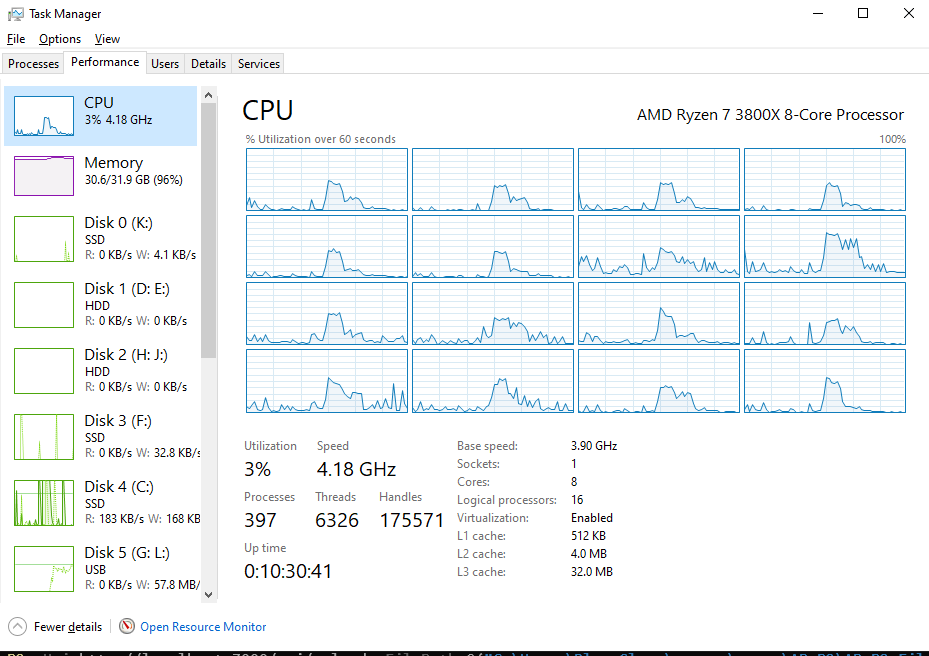
Paralelizacija na nivou fajla i na nivou oba fajla ima značajan uticaj pri obradi velikih fajlova. Pošto su u pitanju ukupno 4 IO stream-a, IO problemi nisu toliko izraženi, dok CPU-bound operacije FFTRecurse metode imaju priliku povećati iskorištenost CPU-a.

Vidimo značajno veću iskorištenost CPU-a.

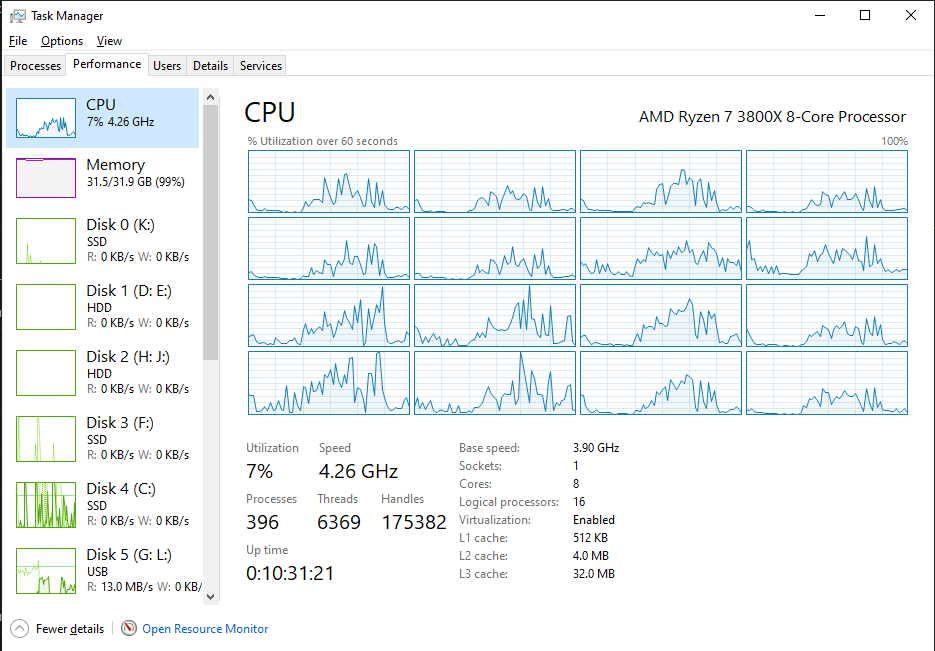
## SIMD obrađivanje



Prikaz 28 SIMD obrađivanje, veliki fajlovi



Prikaz 29 Opterećenje - SIMD obrađivanje, veliki fajlovi - upload



Prikaz 30 Opterećenje - SIMD obrađivanje, veliki fajlovi - obrađivanje



Prikaz 31 Metrike - SIMD obrađivanje, veliki fajlovi

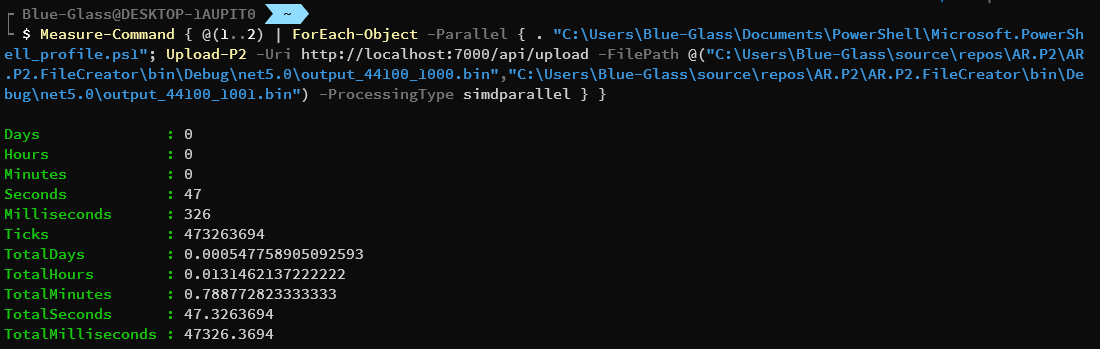
|  |  |
| --- | --- |
| Metrika | Opseg, srednja vrednost |
| Broj poziva FFTRecurse metode [1/s] | [973, 1502], ~ avg. 1000 |
| Srednje trajanje FFTRecurse poziva [ms] | [1.2, 1.41], ~avg. 1.2 |

Interesantno je da je SIMD varijanta nešto sporija od paralelne. Ovo možemo obrazložiti time što veći stepen paralelizma i iskorištenosti CPU-a rezultuje u većem broju izvršenih FFTRecurse poziva u sekundi u paralelnoj varijanti, iako svaki poziv traje kraće u SIMD varijanti. To vidimo i na osnovu manje iskorištenosti CPU-a tokom rada.

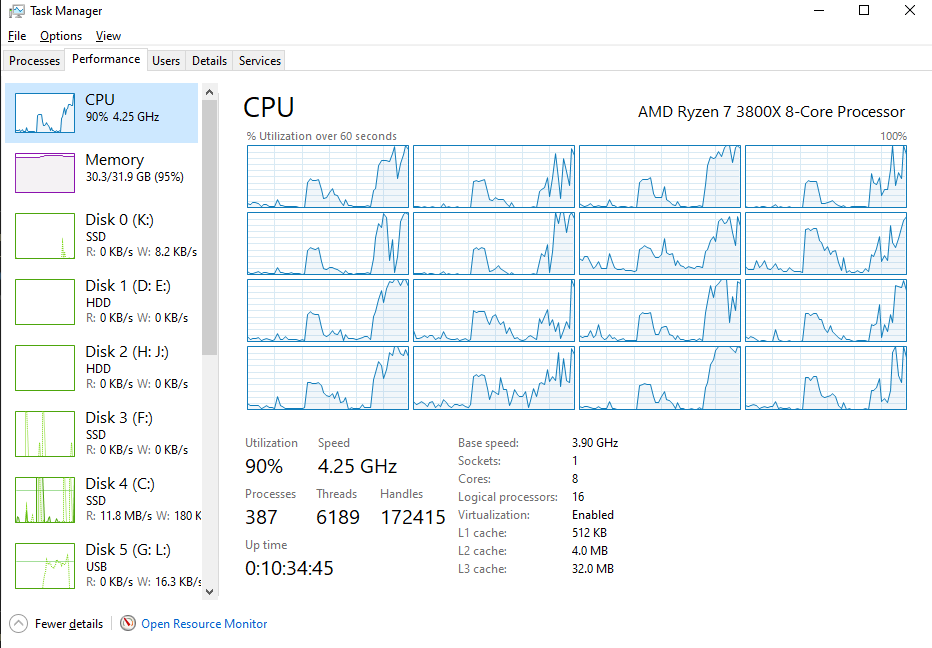
Pritom, značajan pad između perioda većeg opterećenja je uzrokovan IO operacijama između obrade dva fajla, pošto se fajlovi sekvencijalno obrađuju SIMD-om.

Takođe je izraženo kraće trajanje svakog FFTRecurse poziva, što pokazuje da su SIMD optimizacije uspešne.

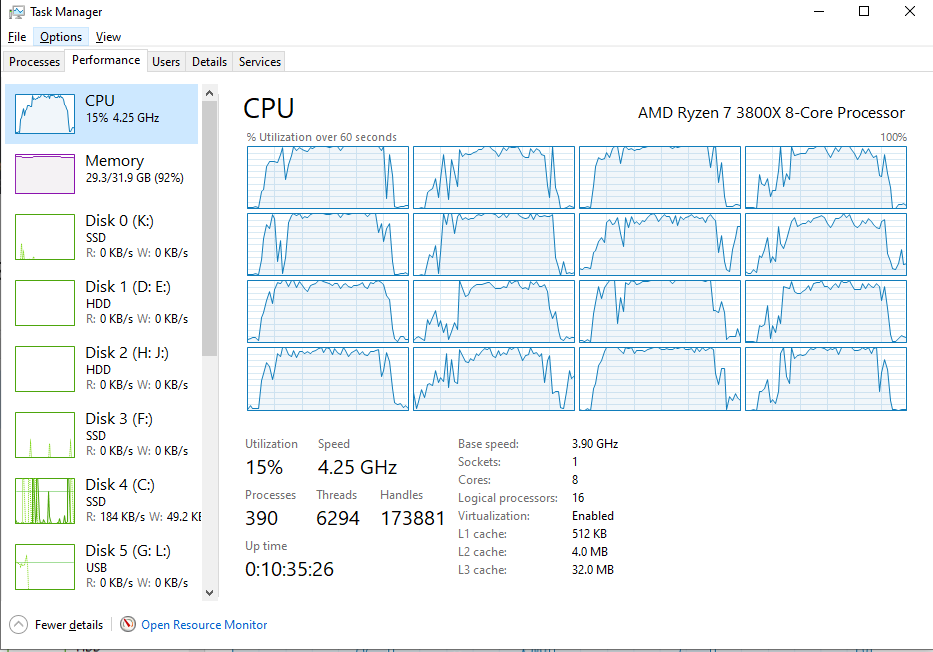
## SIMD paralelno obrađivanje



Prikaz 32 SIMD paralelno obrađivanje, veliki fajlovi



Prikaz 33 Opterećenje - SIMD paralelno obrađivanje, veliki fajlov - upload



Prikaz 34 Opterećenje - SIMD paralelno obrađivanje, veliki fajlovi - obrađivanje

# 

Prikaz 35 Metrike - SIMD paralelno obrađivanje, veliki fajlovi

|  |  |
| --- | --- |
| Metrika | Opseg, srednja vrednost |
| Broj poziva FFTRecurse metode [1/s] | [1689, 1754], ~ avg. 1700 |
| Srednje trajanje FFTRecurse poziva [ms] | [3.52, 8,31], ~avg. 6.5 |

Interesantno je da, iako imamo 100% iskorištenost CPU-a većinu vremena rada i iako je ovo najbrža varijanta, nije mnogo brža u poređenju paralelnom.

Interesantno je videti najveći broj FFTRecurse poziva u metrikama.

# Napomena

Nakdnadno su izvršene sledeće modifikcije:

1. Dodati su XUnit testovi za proveru rada FFT metoda
2. Dodata je preciznija AVX Cos funkcija, koja je precizna na intervalu [-pi,pi], ali se ne koristi
3. Metoda Fma.MultiplyAdd je zamenjena sa Fma.MultiplyAddSubtract
4. Dodat je poziv metode Avx2.Permute4x64 radi korekcije rada algoritma
5. Korigovana je metoda FftService.ParallelProcessing, radi uklanjanja suvišnih dinamičkih alokacija

S tim u obziru, zbog smanjenja dinamičkih alokacija, iako je dodata jedna AVX instrukcija, moguće je da obe paralelne varijante budu još brže od njihovih odgovarajućih sekvencijalnih varijanti.

# Dodatak

Napomena: u trenutnoj verziji dokumentacije, sve Prometheus/Grafana metrike pod nazivom „Mean FFT Duration“ pri paralelizovanim varijantama (obična paralelna i SIMD-paralelna), greškom, obuhvataju pored vremena izvršavanja FFTRecurse metode i vreme izvršavanje Operations.GetFftResults metode. Pošto ta metoda ima vremensku kompleksnost Olog(n), njen uticaj nije zanemarljiv na ukupno vreme pomenute metrike. S tim u obziru, stvarne vrednosti metrike „Mean FFT Duration“ očekujemo da budu potencijalno dosta manje za paralelne varijante.



Prikaz 36 PoweShell i OS verzija

1. Cooley, J. and Tukey, J. (1965) "An algorithm for the machine calculation of complex Fourier series", Mathematics of Computation, 19(90), pp. 297-301. doi: 10.1090/s0025-5718-1965-0178586-1. [↑](#endnote-ref-1)
2. Swagger. 2021. Swagger. [ONLINE] Available at: <https://swagger.io> [Accessed 31 December 2021]. [↑](#endnote-ref-2)