

Bulud İnfrastrukturunda Yükün Proqnozlaşdırılması Üsulu

Fərqanə Abdullayeva

AMEA İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan

farqana@iit.ab.az

Xülasə— Məqalədə bulud infrastrukturunda hostlarda yüklənməni utillizasiya parametrlərinə görə proqnozlaşdırma üsul təklif edilmişdir, müxtəlif klassifikasiya alqoritmlərinin müqayisəli təhlili aparılmışdır. Hostun izafi yüklənmiş və az yüklənmiş olduğu vəziyyətləri yüksək dəqiqliklə proqnoz edə bilən klassifikator əsas aşkarlama vasitəsi kimi təyin edilmişdir. Yanaşmanın test edilməsi “Google cluster trace” verilənlər bazasından götürülmüş “task events” loq yazıları üzərində aparılmışdır. Eksperimental yoxlanma nəticəsində NaiveBayes alqoritmı hostun yüklənmiş və ya az yüklənmiş olduğunu daha dəqiq proqnoz etmişdir və aşkarlama dəqiqliyi 0.95 faiz təşkil etmişdir.

Açar sözlər— bulud texnologiyaları, bulud proqramı, yükün proqnozlaşdırılması, proqnozlaşdırma modelləri, maşın təlimi, Naive Bayes, Nearest neighbor, neyron şəbəkə

I. GİRİŞ

Son zamanlar texnologiyalar cəmiyyəti çox sürətlə mövcud proqramlarını təzələməklə bulud xidmətlərindən istifadə edə bilən proqram təminatları yaratmağa cəhd edirlər. Bu proqramlara əsasən elastiklik, virtuallaşma, xərclərinin azaldılması və s. kimi vacib xüsusiyyətlər daxil etməyə çalışırlar.

Bulud proqramında etibarlılıq, miqyaslılıq, əlçatırlıq, dayanıqlıq, yüksək performans parametrlərinin olması vacib şərtlərdir.

Bulud proqramının RAM, CPU, Bw, disk yaddaşı və s. kimi resursları böyük həcmə malik olduğuna baxmayaraq, bulud fəlsəfəsinin yük balanslaşdırılması, imtinalara dayanıqlığın təmin edilməsi, dinamik planlaşdırma, enerji səmərəliliyinin təmin edilməsi kimi tədqiqat problemləri vardır.

Bulud sistemlərində yükün proqram təminatının yerinə yetirildiyi anda idarə edilməsi üçün yükün maşın təlimi üsulları vasitəsi ilə proqnoz edilməsi və virtual maşın miqyası yanaşması əsasında tənzimlənməsi mühüm məsələdir.

Yükü balanslaşdırmaq üçün ilk addım hostlarda yüklənmənin izafi və ya az olduğunu proqnoz etməkdir [1].

Hostlarda yükün proqnozlaşdırılması üçün parametrlərə sərhəd qiymətləri təyin edirlər [2].

Təqdim olunan məqalədə bulud texnologiyaları mühitində hostlarda yüklənmənin izafi və ya az olduğunu proqnozlaşdırma üsul təklif olunur. Yanaşmada bulud proqramının RAM sərfiyyatı, ötürmə zolağının sərfiyyatı, yaddaş tutumu, CPU sərfiyyatı kimi resursların izafi və ya az yükləndiyinin

proqnozlaşdırılması üçün bu parametrlərə sərhəd qiyməti təyin edilir. Hostların izafi və ya az yükləndiyini müəyyən etmək üçün k sayda yaxın qonşu (K -Nearest Neighbors, KNN), süni neyron şəbəkəsi (*Artificial Neural Network, ANN*), Sadəlovh Bayes (*Naive Bayes*) alqoritmləri “Google cluster trace” verilənləri üzərində test edilmişdir. Təklif edilmiş yanaşmanın effektivliyi müxtəlif qiymətləndirmə kriteriyalarından istifadə etməklə yoxlanmışdır.

II. BULUDUN İŞ YÜKÜ

Bulud xidmətlərinin generasiya etdiyi iş yükünə aşağıdakılar aiddir [3]:

- **Verilənlərə xidmət** (*Data serving*). İrimiyyəli və ya proqramlar üçün yaddaş qurğusu qismində çıxış edən verilənlər bazarıdır;

- **Veb-axtarış** (*web search*). Veb indekslərə görə axtarışı həyata keçirir;

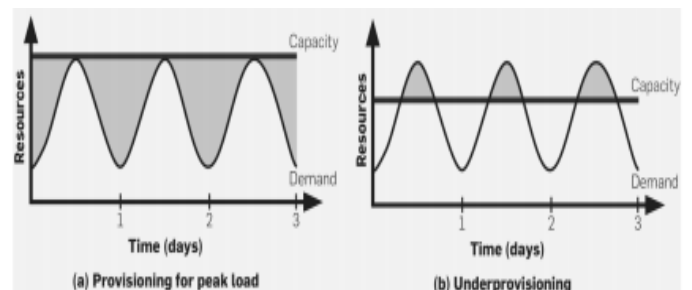
- **MapReduce**. Paralel emal sistemidir, paketlərlə emal olunan bulud-yönümlü layihələrin yerinə yetirilməsi üçün istifadə edilir;

- **Media axını** (*media streaming*). Media fayllarının müxtəlif bitlərdə və formatlarda paketlənməsini və ötürülməsini həyata keçirir;

- **SAT solver həlli**. Bul məsələlərinin həllini həyata keçirir.

III. BULUD RESURLARININ AYRILMASI

Bulud texnologiyalarının sorğu əsasında xidmət, resursların elastik təqdim edilməsi imkanları istifadəçilərin kompüter resurslarından lazım olan miqdarda istifadə etməsinə imkan yaradır. Bulud texnologiyaları resursları elastik miqyaslına bildiyindən bu texnologiyaların tətbiqi müştərilərin xidmətdən kənar qalması, yüklənmə riskinin olması faktlarını azaldır [4].



Şəkil 1. Dəyişkən yük üçün resursların ayrılması

Şəkil 1.-də resurs sərfiyyatının statik halı təsvir edilmişdir. Sol tərəfdəki təsvirdə resursun tutumu pik yüklənməni təmin etmək üçün əvvəlcədən ayrılmışdır. Boz rənglənmiş hissədə ayrılmış resursun tutumu istifadə olunmamış qalır. Sağ tərəfdəki şəkildə isə resursun tutumunun pik yüklənmədən aşağı olduğu vəziyyət təsvir edilmişdir. Burada ayrılmış resurslar tam istifadə edilmişdir, lakin sərhəd xəttinin yuxarısındakı boz rənglənmiş sahə müştərinin xidmət olunmamış sorğularını göstərir.

Statik yanaşmanın sadalanan problemlərini aradan qaldırmaq üçün bulud texnologiyalarının tətbiqi zamanı istifadəçilər resursları miqyaslaşdırma bilər, yükü resurslar arasında paylaşdırmaq imkanı əldə edir.

Yükün effektiv paylaşdırılmasının təmin edilməsi üçün vacib məsələ dəqiq qərarların qəbul edilməsidir. Burada qərarların qəbulunda hər hansı səhvə yol verilməsi bulud infrastrukturunu yüklənməyə məruz qoya bilər, bu isə effektivliyin azalması ilə nəticələnər.

IV. YÜKÜN PROQNOZLAŞDIRILMASI METODU

Yükün effektiv proqnozlaşdırılmasına sadə yanaşma bulud infrastrukturunun server kriteriyalarının (prosessor sərfiyyatı, hostda emal olunan tapşırıqların sayı, yaddaş sərfiyyatı və s.) sərhəd qiymətinə nəzərən izlənməsini təmin etməkdir. Bu prosesi həyata keçirmək üçün aşağıdakı yanaşma təklif edilir:

Addım 1. “Google cluster trace” verilənlər bazasının “task_events” loq yazılarından yüklənmə parametrlərinin əldə olunması;

Addım2. Hostda emal olunan tapşırıqların sayına görə sərhəd qiymətinin təyin edilməsi (bizim halda tapşırıqların sayı normalda 4-dür). Hosta emal olunan tapşırıqların sayı üçün sərhəd qiyməti provayder tərəfindən təyin edilir.

Addım 2. Bu verilənlərin 4 sərhəd qiymətinə nəzərən iki sinfə ayrılması (0 – az yüklənmiş host, 1 – izafi yüklənmiş host);

Addım 3. Bu verilənlər üzərində müxtəlif maşın təlimi üsullarının test edilməsi;

Addım 4. Əgər tapşırıqların sayı təyin edilmiş sərhəd qiymətindən böyükdürsə, host izafi yüklənmiş sinfinə daxil edilsin;

Addım 5. Əgər tapşırıqların sayı təyin edilmiş sərhəd qiymətindən kiçikdirsə host az yüklənmiş sinfinə daxil edilsin.

V. İSTİFADƏ OLUNAN PROQNOZ MODELƏRİ

Yükün intellektual proqnozlaşdırılması üçün aşağıdakı maşın təlimi üsulları istifadə edilmişdir:

D. Naive Bayes alqoritmi

Naive Bayes üsulunda dəyişənlərin bir-birindən asılı olmaması əsas şərt kimi qəbul olunur. Obyektin bu və ya digər sinifə aid olması ehtimalını Bayes düsturu vasitəsi ilə tapır. Metodun adında sadələvh (naive) sözü dəyişənlərin bir-birindən asılı olmadığına görə istifadə edilmişdir.

Fərz edək ki, $P(y = c_r)$ hər hansı bir y obyektinin $c_r (y = c_r)$ sinfinə aid olması ehtimalıdır. Fərz edək ki, E asılı olmayan dəyişənlərin müəyyən qiymət almasını göstərən hadisə $P(E)$ isə həmin hadisənin baş vermə ehtimalıdır.

Metodun ideyası obyektin c_r sinfinə aid olmasını müəyyən edən şərti ehtimalı hesablamaqdır.

Ehtimal nəzəriyyəsiindən məlumdur ki, hər hansı E hadisəsinin baş vermə ehtimalı aşağıdakı düsturla hesablanır:

$$P(y = c_r | E) = \frac{P(E | y = c_r) * P(y = c_r)}{P(E)} \quad (1)$$

Burada qaydalar generasiya olunur, bu qaydaların şərt hissində bütün asılı olmayan dəyişənlər uyğun mümkün qiymətlərlə müqayisə edilir. Yekun hissədə asılı dəyişənin bütün mümkün qiymətləri göstərilir. Yəni $x_1 = c_1^k, \dots, x_n = c_n^k$, $y = c_r$ və s. Beləliklə bu qaydalar bütün elementlər üçün qurulur. Burada y – öyrədilən seçim daxilindəki bütün obyektlər, c_r – hər hansı r -ci sinif, x_1, \dots, x_n – asılı olmayan dəyişənlər, c_i^k – dəyişənlərin aldığı bütün mümkün qiymətlərdir.

Bütün bu qaydaların hər biri üçün Bayes düsturuna əsasən onların ehtimal aşağıdakı düsturla hesablanır:

$$P(E | y = c_r) = P(x_1 = c_1^k | y = c_r) P(x_n = c_n^k | y = c_r) \quad (2)$$

E. K-nearest neighbors alqoritmi

K sayda ən yaxın qonşu alqoritmi metrik klassifikasiya alqoritmidir, obyektlərin uyğunluğunu qiymətləndirir. Burada klassifikasiya olunan obyekt təlim verilənlərində bu obyektə yaxın olan obyektlərin təşkil olunduğu sinifə aid edilir.

Qonşu obyektlər arasında məsafə aşağıdakı düstur vasitəsi ilə hesablanır:

$$D(\hat{x}, x_i) = \sum_{j=1}^d (\hat{x}_j - x_{ij})^2 \quad (3)$$

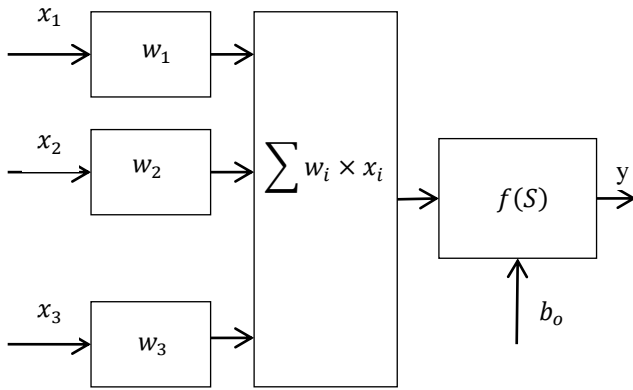
F. Neyron şəbəkələri

Süni neyron aşağıdakı elementlərdən təşkil edilir (şəkil 2):

- hər birinin w_i çəki əmsalı olan x_i giriş verilənləri çoxluğu;
- x_i giriş verilənlərini w_i çəkili ilə cəmləyən cəmləyici element;
- cəmləyicinin çıxış (qeyri-xətti) qiymətlərinin çevrilməsini həyata keçirən $f(S)$ aktivləşdirmə funksiyası.

i -ci əlaqə vasitəsi ilə daxil olan hər bir x_i qiyməti əlaqənin w_i çəkisinə vurulur. Belə olduqda neyronun çəkili cəmi aşağıdakı kimi hesablanır:

$$S = w_1 \times x_1 + w_2 \times x_2 + \dots + w_n \times x_n + b_0 = \sum_{i=1}^n w_i \times x_i + b_0 \quad (4)$$



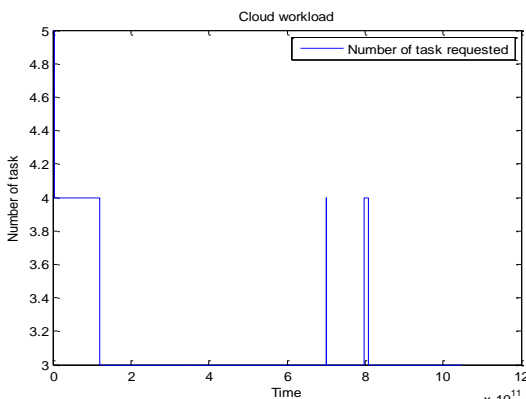
Şəkil 2. Sünəy neyron

Təqdim olunan məqalədə yükün proqnozlaşdırılması məsələsi üçün giriş neyronlarının sayının 5 çıxış neyronlarının sayının isə 2 olması nəzərdə tutulmuşdur. Tədqiqatın aparılması prosesində istifadə olunan verilənlər bazasında yüklənmənin proqnozlaşdırılması 5 əlamət üzrə identifikasiya olduğundan giriş neyronlarının sayı 5 götürülür. İdentifikasiyanın nəticəsi yüklənmiş və az yüklənmiş siniflərdən ibarət olduğu üçün modelin çıxış neyronlarının sayı 2 götürülür.

VI. EKSPERİMENT

Bulud infrastrukturunda yükün göstərilən modellər əsasında proqnozlaşdırılması üçün eksperimentlərin aparılmasında “Google cluster trace” verilənlər bazasından “task_events” loq yazıları istifadə edilmişdir [5]. Bu loq yazı cpu_requested, disk_space_requested, number_of_task_requested və time kimi xarakteristikalar (əlamətlər) ibarətdir.

Şəkil 3-də “Google cluster trace” verilənlər bazasından götürülmüş təsnif edilməmiş verilənlər əsasında CPU və yaddaşa edilən sorğuların zaman sırası təsvir edilmişdir. Təklif edilmiş yanaşmaya əsasən bulud resursunun yüklənmə dərəcəsini müəyyən etmək üçün əvvəlcə bu təsnif edilməmiş verilənlərin təyin edilmiş sərhəd qiymətinə əsasən klasterizasiya aparılır. Burada sərhəd qiyməti 4 götürülərək verilənlər bazasının verilənləri az yüklənmiş və yüklənmiş kimi iki sinfə bölünür. Burada bir hostda emal olunan tapşırıqların sayının 4-dən kiçik bərabər olduğu vəziyyətdə host az yüklənmiş, böyük olduğu halda isə yüklənmiş kimi qiymətləndirilir.



Şəkil 3. “Google cluster traces” verilənlər bazasında tapşırıqların sayının zaman sıraları

DOI: 10.25045/NCSofEng.2017.24

Resursun yükləndiyini müəyyən etmək üçün hostda emal olunacaq maksimal tapşırıqların sayını göstərən sərhəd qiyməti provayder tərəfindən təyin olunur.

G. Dəqiqliyin yoxlanılması

Təklif edilmiş modelin yüklənməni proqnozlaşdırma dəqiqliyini qiymətləndirmək üçün aşağıdakı kriteriyalar istifadə edilmişdir:

1) Dəqiqlik (precision):

$$Precision = TP / (TP + FP) \quad (5)$$

2) Tamlıq (recall):

$$Recall = TP / (TP + FN) \quad (6)$$

3) F-ölçü (F-measure):

$$F - measure = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (7)$$

4) Doğru müsbət hallar (True Positives, TP):

$$TPR = TP / positive \quad (8)$$

5) Doğru mənfi hallar (True Negatives, TN):

$$TNR = TN / negative \quad (9)$$

6) Doğruluq (accuracy):

$$Accuracy = (TP + TN) / (P + N) \quad (10)$$

Məqalədə resursun yüklənmiş və ya az yüklənmiş olduğunu müəyyən etmək üçün NaiveBayes, K-nearest neighbors, Neural network klassifikatorlarının müqayisəli təhlili aparılmışdır.

Cədvəl 1-dən görüldüyü kimi Naive Bayes klassifikatoru hostun yüklənmə səviyyəsini daha yüksək dəqiqliklə identifikasiya edə bilir.

CƏDVƏL 1. “GOOGLE CLUSTER TRACE” VERİLƏNLƏR BAZASININ TEST NƏTİCƏLƏRİ

	Naive Bayes	K-nearest neighbors	Neural network
TP	0.95	0.60	0.03
TN	0.79	1	1.00
FP	0.21	0.00	0.00
F-measure	0.11	0.75	0.06
Recall	0.95	0.60	0.03
Acuracy	0.79	0.99	0.98
Precision	0.95	0.60	0.03

NƏTİCƏ

Məqalədə bulud mühitində resursların utilizasiya parametrlərinə görə hostlarda yükün proqnozlaşdırılması üsulu təklif edilir. Yükün proqnozlaşdırılması bulud mühitində yük balanslaşdırılmasını təmin etmək üçün zəruridir. Yükün proqnozlaşdırılmasında yüksək dəqiqlik verə bilən alqoritmi seçmək məqsədi ilə NaiveBayes, K-nearest neighbors, Neural network maşın təlimi üsullarının eksperimental nəticələri

müqayisə edilmişdir. Sadalanan alqoritmlərdən *Naive Bayes* klassifikatoru hostun yüklənmə səviyyəsini daha yüksək dəqiqliklə proqnozlaşdırma bildiyi göstərilmişdir.

ƏDƏBİYYAT

- [1] A. Bala, I. Chana, “Prediction-based proactive load balancing approach through VM migration,” *Engineering with Computers*, 2016, vol. 32, no 4, pp. 581-592.
- [2] E. Caron, F. Desprez, A. Muresan, “Forecasting for Cloud computing on-demand resources based on pattern matching,” Technical report, INRA, 2011, 24 p.
- [3] C. Vazquez, R. Krishnan, E. John, “Time Series Forecasting of Cloud Data Center Workloads for Dynamic Resource Provisioning,” *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, vol. 6, no. 3, pp. 87-110.
- [4] M. Armbrust, A. Fox, R. Griffith, A.D. Joseph, R.Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica et al., “A view of cloud computing,” *Communications of the ACM*, vol. 53, no. 4, pp. 50-58, 2010.
- [5] Google Cluster Data, “ClusterData2011_2 traces,” <https://github.com/google/cluster-data>