

Bulud Test Mühitlərinin Böyük Verilənlərin Emalında Reallaşma İmkanlarının Tədqiqi və Analizi

Fərqanə Abdullayeva

AMEA İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan
farqana@iit.ab.az

Xülasə — Məqalə bulud test mühitlərinin böyük verilənlərin emalında reallaşma imkanlarının tədqiqi məsələsinə həsr olunmuşdur. Bu məqsədlə, bulud simulyasiya proqramlarının, ticari buludların, bulud test mühitlərinin məqsədləri araşdırılmış, onların böyük verilənlərin emalında tətbiqi imkanlarının dərinədən analizi aparılmışdır.

Açar sözlər — *bulud texnologiyaları; test mühiti; böyük verilənlər; CloudSim; CloudAnalyst; GreenCloud; OpenCirrus; Google Apps*

I. GİRİŞ

Son zamanlar bulud texnologiyaları təşkilat və korporasiyaların infrastrukturunda yeni texnologiya kimi meydana gəlmişdir və elmin bir sıra müxtəlif tədqiqat sahələrinə ciddi təsir göstərmişdir. Bulud xidmətlərinin uğurla həyata keçəcəyinə zəmanəti təmin etmək məqsədi ilə bu sahədə yaradılmış modellər istifadəçiyə servis xidməti şəklində təklif edilməmişdən əvvəl test edilməlidir.

Bulud texnologiyalarında tədqiqatların aparılmasına maneə törədən başlıca amil eksperimentlərin yoxlanmasını təmin edən tədqiqat platformasının yaradılmasındadır.

Bulud texnologiyaları əsasən kommersiya məqsədi daşıyır və kommersiya məqsədli buludlar təbiətə tədqiqat platforması kimi olduqca realdır. Lakin onlar tədqiqatçıların etibarlı eksperimentlərinin aparılması üçün onlara kifayət qədər idarəetmə hüququ vermir. Bundan əlavə, simulyasiya, riyazi modelləşdirmə və ya kiçik prototiplərdən istifadə etməklə aparılan tədqiqatlar nəticəsində yaradılmış bulud məhsullarının irimiqyaslı real buludlarda tətbiqi hər zaman uğurlu alınır.

Sadalanan problemləri aradan qaldırmaq üçün bulud sistemlərini dəstəkləyən test mühitlərinin (testbeds) yaradılması istiqamətində aktiv şəkildə işlər aparılır [1]. Bu növ platformaların yaradılması həmçinin ABŞ-ın Milli Elm Fondunun (National Science Foundation, NSF) da diqqət mərkəzindədir. Belə ki, NSF-in 2012-ci ildə elan etdiyi çağırışların başlıca mövzularını məhz elə bulud texnologiyaları üçün test mühitlərinin yaradılması təşkil edirdi [2], 2014-cü ildə elan etdiyi çağırış bulud texnologiyaları üçün “Chameleon” və “CloudLab” adlı test mühitlərinin yaradılması ilə bağlı idi [3].

Təqdim olunan məqalədə böyük verilənlərin emalında bulud test mühitlərinin mövcud problemlərinin müəyyən edilməsi istiqamətində araşdırmalar aparılmışdır. Bulud simulyatorlarının, test mühitlərinin böyük verilənlərin

emalında yararlı olmasını təmin etmək üçün təklif və tövsiyələr verilir.

II. BULUD SİMULYASIYA PROQRAMLARI

Hazırkı dövrdə bulud sistemlərinin təbiətinin öyrənilməsinə həsr olunmuş çoxsaylı tədqiqatlar aparılır və bu tədqiqatların eksperimental yoxlanmasını həyata keçirmək üçün möhtəşəm simulyasiya proqramları yaradılır. Bu tip proqramlara aşağıdakılar aiddir [4]:

CloudSim. Bulud texnologiyaları infrastrukturlarını və tətbiqi proqram xidmətlərini modelləşdirməyə, simulyasiya və sınaq etməyə imkan verən ümumi simulyasiya vasitəsidir. CloudSim Melburn Universitetinin (ing. University of Melbourne) “Kompüter elmləri və proqram mühəndisliyi” şöbəsinin CLOUDS laboratoriyasında yaradılmışdır və GridSim proqram təminatının genişlənmiş formasıdır. CloudSim – verilənlər anbarını, veb-servisləri, virtual maşınlar arasında resursların paylanmasını, federallaşmış buludları və s. modelləşdirməyə imkan verir. CloudSim toolkit 3.0 proqram təminatı 2012-ci ilin yanvar ayında istifadəyə verilmişdir. CloudSim sistemindən HP və başqa təşkilatlar, o cümlədən dünyanın bir sıra universitetləri bulud resurslarının paylanması, verilənlər mərkəzinin resurslarında enerji qənaətinin idarə edilməsi, bulud texnologiyalarının optimallaşdırılması və elmi fəaliyyətlərdə istifadə edirlər. Bu sistemin çatışmazlığı onun qrafiki istifadəçi interfeysinə olmamasıdır.

CloudAnalyst. Bulud texnologiyaları mühitini və tətbiqi proqramları analiz etmək üçün CloudSim əsasında qurulmuş vizual modelləşdirmə vasitəsidir.

GreenCloud. Verilənlər mərkəzinin kompüter serverləri, şəbəkə kommutatorları, rabitə əlaqələri kimi IT qurğularının enerji sərfiyyatını modelləşdirməyə xidmət edir. GreenCloud simulyatoru NS2 (Network Simulator) şəbəkə simulyatorunun genişlənmiş formasıdır, GreenIT və ECOCLOUD layihələri çərçivəsində işlənib hazırlanmışdır.

MDCSim. 2009-cu ildə yaradılmışdır, çoxsəviyyəli verilənlər mərkəzlərini dərininə analiz etməyə xidmət edən çevik, miqyaslı bilən, kompleks simulyasiya platformasıdır. Təyinatı real verilənlər mərkəzlərini təqlid etməkdir, bunun üçün verilənlər mərkəzlərinin serverlər, müxtəlif vendorların təqdim etdiyi kommutatorlar, kommunikasiya əlaqələri kimi xüsusi aparat təminatı xarakteristikaları ilə təchiz edilmişdir. Simulyator verilənlər mərkəzlərinin ötürücülük qabiliyyətini, cavabvermə vaxtını, enerji sərfiyyatını qiymətləndirə bilər.

iCanCloud. Böyük həcmdə saxlanları olan bulud infrastrukturalarını (minlərlə qovşağa malik) simulyasiya etmək üçün 2010-cu ildə yaradılmış platformadır. Platforma açıq kodludur və istifadəyə yararlılıq, çeviklik, məhsuldarlıq və miqyaslılıq kimi xüsusiyyətlərə malikdir. Təyinatı xüsusi aparat qurğusundakı müəyyən tətbiqi proqramın qiyməti ilə məhsuldarlığı arasında nisbəti proqnozlaşdırıb çəkiləcək xərclər haqqında istifadəçiləri məlumatlandırmaqdır.

NetworkCloudSim. CloudSim simulyatorunun genişlənmiş formasıdır, verilənlər mərkəzlərindən fərqli olaraq verilənlər axınına dəstəkləyir.

VirtualCloud. Bulud texnologiyaları mühiti yaratmağa və yeni yaradılmış həlləri test etməyə imkan verən simulyatordur.

Miqyaslılıq və eksperimentlərin həyata keçirilməsində təkrarlamaları mümkünlüyü baxımından üstünlük təşkil etdiyinə baxmayaraq simulyasiya platformaları həqiqi bulud texnologiyalarını tam əks etdirə bilmir. Bu səbəbdən tədqiqatçılar real bulud test mühitlərindən istifadə olunmasına üstünlük verirlər. *Ticari bulud test mühitləri* və *bulud test mühitləri* real bulud test mühitləridir. Bu test mühitləri haqqında ətraflı məlumat aşağıda verilir.

III. TİCARİ BULUD TEST MÜHİTLƏRİ

Bulud texnologiyaları mühitini eksperimental məqsədlərə sazlamak üçün zəruri şərt burada aparat təminatı infrastrukturalarına giriş imkanının və onu idarə etmək üçün proqram təminatının olmasıdır. Tədqiqatçılar eksperimental tədqiqatların aparılması üçün həm ümumi məqsədlərə xidmət edən ticari bulud servislərindən və həm də elmi məqsədlər üçün xüsusişədirilmiş bulud test mühitlərindən istifadə edir. Ticari bulud servisləri böyük miqdarda investisiya xərcləri tələb etmədən sürətli və kiçikölçülü eksperimentlərin aparılması üçün əlverişlidir, xüsusişədirilmiş bulud test mühitləri isə uzun sürən eksperimentlərin aparılması üçün əlverişlidir.

Elmi tədqiqatlarda istifadə olunan *ticari bulud test mühitləri* aşağıdakılardır:

Amazon Elastic Compute Cloud (EC2). Buludda miqyaslanan hesablaşma resursları təqdim edən veb-servisdir, əksər elmi tədqiqatlarda istifadə olunmuşdur. İstifadəçilərin virtual maşınların sayında dəyişiklik etməsinə imkan yaradır. EC2 platformasının tətbiqini və ya onun məhsuldarlığının qiymətləndirilməsini hədəfə alan məqalələrə [5, 6, 7] ədəbiyyatlarını göstərmək olar.

Amazon S3 saxlanc xidməti (S3 storage services). İstənilən həcmdə verilənləri saxlamağa imkan yaradan, Amazon təşkilatı tərəfindən təqdim edilən bulud xidmətidir, 2006-cı ildə yaradılmışdır. Bu platformanın üzərində eksperimentlərin aparıldığı məqaləyə [8] aid etmək olar. Burada S3 və EC2 platformalarından istifadə edərək verilənlərin ötürülmə anında məhsuldarlıqda baş verən dəyişikliklər tədqiq edilmişdir.

Google App Engine. İstifadəçilərin proqram tətbiqlərini Google-nin infrastrukturunda yaratmasına və icra etməsinə şərait yaradan PaaS xidmətidir. Burada verilənlər saxlancının tutumu istifadəçinin tələbatına və trafik həcminə görə artırıla

bilir. Bu platformanın məhsuldarlığının qiymətləndirilməsinə ədəbiyyat [9]-də baxılmışdır.

Google Apps. Təbiətinə görə SaaS tipli proqram tətbiqidir. İstifadəçilərə ənənəvi ofis proqram təminatlarının (məsələn, *Word* redaktoru) funksional imkanlarına bənzər veb-tətbiqlər təqdim edir, onların bir-biri ilə asanlıqla qarşılıqlı əlaqə qurmasına şərait yaradır. Burada bütün tətbiqlər onlayn saxlandığından və onlara giriş veb-brauzer vasitəsi ilə mümkün olduğundan, istifadəçilər öz hesablarına İnternet şəbəkəsinə qoşulmuş istənilən kompüterdən daxil ola bilər və lokal olaraq əlavə proqram təminatı quraşdırmağa zərurət qalmır. [10]-də Google Apps əsasında e-təlim mühitinin yaradılması üçün yanaşma təklif edilmişdir.

Windows Azure. Platforma xidmətidir, təyinatı Windows proqram tətbiqlərini icra etməkdir. Burada proqram tətbiqləri və virtual maşınlar yaratmaq və onları idarə etmək üçün platforma təqdim edir. [11]-də Windows Azure platformasının məhsuldarlığının yüksəldilməsi üçün yanaşma təklif edilir.

IV. BULUD TEST MÜHİTLƏRİ

OpenCirrur. Bulud texnologiyaları sahəsində tədqiqatların aparılması üçün 2008-ci ildə yaradılmış ən böyük test mühitidir. Mövcud alternativlərdən fərqli olaraq paylanmış verilənlər mərkəzlərinin federallaşmasını təmin edə bilər. Layihə HP, Intel, Yahoo! və bir çox təşkilatların birgə təşəbbüsü nəticəsində yaradılmışdır. OpenCirrur (<http://opencirrur.org>) tədqiqatçılara fiziki və virtual maşınlardan, o cümlədən vahid giriş, monitorinq, saxlanc kimi global serverlərdən ibarət bulud steki təqdim edir. OpenCirrur dörd məqsədə xidmət edir: bulud texnologiyalarında sistem səviyyəsində tədqiqatların aparılmasına köməklik, proqram tətbiqləri səviyyəsində tədqiqatların aparılmasına köməklik, eksperimental verilənlər çoxluğunun təqdim edilməsi, buludlar üçün açıq kodlu sistemlərin yaradılması. Test mühitinin tərkibinə şimali Amerika, Avropa və Asiya ölkələrindən olan on bölmə daxildir. [12]-də xüsusi buludun OpenCirrur test mühitində məhsuldarlığının qiymətləndirilməsi məsələsinə baxılmışdır.

Açıq Bulud Test mühiti (Open Cloud Testbed, OCT). Açıq Bulud Konsorsiumu (Open Cloud Consortium, OCC) tərəfindən 2009-cu ildə yaradılmışdır və bulud texnologiyalarını dəstəkləyən geniş sahəli test mühitidir [13]. OCT bulud texnologiyaları tədqiqatlarını inkişaf etdirmək, müxtəlif bulud arxitekturlarının cəlb etdiyi eksperimental tədqiqatların aparılmasını mümkün etmək, bulud interoperabeliyinin öyrənilməsi üçün platforma təqdim etmək, bulud texnologiyaları bencmarklarını dəstəkləmək məqsədi ilə yaradılmışdır. Hazırda o ABŞ-ın Baltimor (John Hopkins University), Çikaqo (University of Illinois at Chicago və StarLight) və San Dieqo (California Institute for Telecommunications and Information Technology) şəhərlərindəki dörd verilənlər mərkəzini özündə birləşdirir. OCT təşəbbüsünün ikinci fazasında buraya həmçinin Kembric (MIT Lincoln Lab) və Pitsburq (Pittsburgh Supercomputer Center / Carnegie Mellon University) şəhərlərindəki daha iki verilənlər mərkəzi də əlavə edilmişdir.

Bu test mühiti digər paylanmış infrastrukturular üçün problem sayılan olduqca böyük verilənlər axınının tələblərini ödəyə bilər. OCT platformasının tərkibində Eucalyptus, Hadoop, CloudStore (KosmosFS), Sector/Sphere və Thrift kimi müxtəlif bulud sistemləri və xidmətləri quraşdırılmışdır və onlar sistemdə tədqiqatların aparılması üçün əlçatır formada mövcuddur. OCT-nin tərkibinə daxil olan bu sistemlər benchmark tədqiqatların aparılması prosesini asanlaşdırır. Burada həmçinin bulud texnologiyaları stekində eksperimental tədqiqatların həyata keçirilməsini mümkün etmək üçün şəbəkə kitabxanaları, monitoring sistemləri də yaradılmışdır.

Elmi tədqiqatların aparılması üçün əlçatır olan digər bulud test mühitlərinə Science Clouds, Virtual Computing Lab, Future Grid, Grid'5000, Okeanos, Illinois Cloud Computing Testbed (CCT), IEEE Intercloud Testbed infrastrukturalarını göstərmək olar.

V. BÖYÜK VERİLƏNLƏR VƏ BULUD TEST MÜHİTLƏRİ

Hazırda mövcud olan bulud test mühitləri böyük verilənlər axınının emalında öz təsirini itirir. Bu səbəbdən yuxarıda sadalanan bulud simulyatorlarının, test mühitlərinin böyük verilənlərin emalında yararlı olmasını təmin etmək üçün bu bulud vasitələrinin tərkibinə paralelləşmə yarıda bilinən MapReduce və s. modellərini daxil edirlər. Bu tip yanaşmalara [14, 15] ədəbiyyat mənbələrində baxılmışdır.

NƏTİCƏ

Test mühitləri tədqiqatçıların paralel və bulud texnologiyaları sahəsində innovativ yanaşmalarını reallaşdırmaq və test etmək məqsədi ilə yaradılır. Mövcud bulud test mühitləri tədqiqatçılara paylanmış sistemlərdən istifadə edərək öz eksperimentlərini həyata keçirmək imkanı vermək üçün onlara infrastruktura təqdim edir. Lakin bulud test mühitləri böyük verilənlərlə bağlı tədqiqatların aparılmasını dəstəkləyə bilmir. Bu test mühitlərinin böyük verilənlərin tələblərini ödəmək məqsədi ilə onların modifikasiyası istiqamətində aktiv şəkildə araşdırmalar aparılmalıdır. Bu tip araşdırmalar həmçinin nəhəng elmi maliyələşdirmə qurumlarının da prioritet məsələləri siyahısına daxil edilməlidir.

MİNNƏTDARLIQ

Bu iş Azərbaycan Respublikasının Prezidenti yanında Elmin İnkişafı Fondunun maliyyə yardımı ilə yerinə yetirilmişdir – **Qrant № EIF-2013-9(15)-46/16/1**

ƏDƏBİYYAT

- [1] G. Sakellari, G. Loukas, “A survey of mathematical models, simulation approaches and testbeds used for research in cloud computing,” *Simulation Modelling Practice and Theory*, 2013, Vol. 39, pp. 92–103.
- [2] NSF Report on Support for Cloud Computing, 2012, 21 p.
- [3] Enabling a new future for cloud computing, 2014, http://nsf.gov/news/news_summ.jsp?cntn_id=132377
- [4] A.K. Pathan, M.M. Monowar, S. Khan, “Simulation technologies in networking and communications: Selecting the best tool for the test,” 2014, 648 p.
- [5] S. Ostermann, A. Iosup, N. Yigitbasi, R. Prodan, T. Fahringer, D. Epema, “A performance analysis of EC2 cloud computing services for scientific computing, Proc. of the first international conference cloud computing (CloudComp),” 2009, pp. 115-131.
- [6] A. Iosup, S. Ostermann, M. Yigitbasi, R. Prodan, T. Fahringer, D. Epema, “Performance analysis of cloud computing services for many-tasks scientific computing, IEEE transactions on parallel and distributed systems,” 2011, Vol. 22, No 6, pp. 931-945.
- [7] S. Islam, K. Lee, A. Fekete, A. Liu, “How a consumer can measure elasticity for cloud platforms, Proc. of the 3rd ACM/SPEC international conference on performance engineering,” 2012, pp. 85–96.
- [8] J. Wang, P. Varman, C. Xie, “Avoiding performance fluctuation in cloud storage, Proc. of the international conference on high performance computing,” 2010, pp. 1–9.
- [9] A. Iosup, N. Yigitbasi, D. Epema, “On the performance variability of production cloud services, Proc. of the 11th IEEE/ACM international symposium on cluster, cloud and grid computing (CCGrid),” 2011, pp. 104-113.
- [10] J. Alemany, X. Perramon, L. Panadès, “Improving the quality of the practicum. The use of moodle and google docs in monitoring the practicum process after the EHEA,” 2012.
- [11] W. Lu, J. Jackson, J. Ekanayake, R. Barga, N. Araujo, “AzureBlast: A case study of developing science applications on the cloud,” Proc. of the 19th ACM international symposium on high performance distributed computing, 2010, pp. 413-420.
- [12] C. Baun, M. Kunze, “Performance measurement of a private cloud in the OpenCirrus testbed,” Proc. of the parallel processing workshops, 2010, pp. 434-443.
- [13] R. Grossman, Y. Gu, M. Sabala, C. Bennet, J. Seidman, J. Mambratti, “The open cloud testbed: A wide area testbed for cloud computing utilizing high performance network services,” Proc. of the GridNets, 2009, pp. 1-6.
- [14] Jung J., Kim H., “MR-CloudSim: Designing and implementing MapReduce computing model on CloudSim,” Proc. of the international conference on ICT convergence (ICTC), 2012, pp. 504-509.
- [15] Laszewski G., Fox G. “The FutureGrid testbed for big data, Cloud computing for data-intensive applications,” 2014, pp. 27-59.