

Elektron Elmin Elmi Verilənlər Problemləri Haqqında

Təhmasib Fətəliyev

AMEA İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan
depart3@it.ab.az

Xülasə – Məqalə elektron elm konsepsiyasının reallaşması çərçivəsində elmi verilənlərin emalı proseslərinin aktual məsələlərinə həsr olunmuşdur. Elmi fəaliyyət zamanı toplanan, saxlanılan və ötürülən elmi verilənlərin həcmnin əhəmiyyətli dərəcədə artması göstərilmişdir. Elm sferasında mühüm Big data təşəbbüsləri tədqiq olunmuş və həmin sahədə tətbiq olunan texnologiyalar haqqında məlumat verilmişdir.

Açar sözlər – e-elm, elmi verilənlər problemləri, AzScienceNet, Data Mərkəz, Big data təşəbbüsləri, grid, cloud computing.

I. GİRİŞ

Son illər ərzində informasiya texnologiyaları sahəsində əldə olunan nailiyyətlər elmin inkişafında əsas tendensiyalardan biri kimi elmi verilənlərin həcmnin əhəmiyyətli dərəcədə artırmış və onların saxlanması və emalı problemini meydana gətirmişdir. Elmi fəaliyyət nəticəsində böyük həcmdə eksperimental verilənlər yaranmış, müxtəlif elmi nəşrlərin, görülən işlər haqqında hesabatların, layihələrin nəticələrinin və s. sayı artmışdır. Hazırda elmi fəaliyyətin təşkili üçün virtual elmi kollektivlərin formalaşdırılması məqsədilə ərazicə paylanmış mühitdə tərəfdaşların axtarışı, həmçinin yaradılmış texnologiyaların və elmi fəaliyyətin nəticələrinin istifadəsi və yayılması ayrıca bir problemə çevrilmişdir. Belə nəticəyə gəlmək olar ki, elmin gələcək inkişafı yalnız böyük həcmli informasiyanın emalı nəticəsində mümkün olacaqdır.

Verilənlərin elmdə rolu həmişə yüksək qiymətləndirilmişdir, lakin elmi verilənlərin qeydiyyatı, emalı və analizi “hesablama elminin” (“*computational science*”) – məsələlərin həlli üçün informatika və hesablama texnikasının ən son nailiyyətlərindən geniş istifadə edilən yanaşmanın inkişafı ilə xüsusi aktuallıq qazanmışdır. Nəticə olaraq, bu sahədə tətbiq edilən elmi metodlar da nəzərə çarpacaq qədər dəyişir. Belə ki, son zamanlar təkcə əldə olunan nəticələr deyil, həmçinin sonrakı emal üçün nəzərdə tutulan və informasiya massivləri şəklində toplanan ilkin verilənlər də təqdim edilir. Nəticədə, bir tərəfdən çox böyük həcmli informasiyanın arxivləşdirilməsi problemi, digər tərəfdən isə informasiyanın axtarışı üzrə xidmətin təşkili, yayılması və istifadəçinin tələb olunan informasiya ilə təmin edilməsi zərurəti yaranır. Bunlarla yanaşı, toplanmış son dərəcədə böyük həcmli elmi verilənlərdən gizli biliklərin aşkarlanması da aktual və əhəmiyyətlidir.

Beləliklə, böyük həcmli verilənlər və onun mühüm tərkib hissələrindən olan elmi verilənlərin yuxarıda qeyd olunan problemlərinin həlli yeni bir elm sahəsinin – verilənlər haqqında elmin (*Data science*), həmçinin qısa bir zaman

ərzində “*Big Data*”, “*Big Data analytics*”, “*Data science*”, “*Datalogy*”, “*Data scientist*”, “*Data-intensiv science*” kimi artıq populyar olan, yeni terminlərin yaranmasına gətirmişdir.

Elmi verilənlər problemlərinin həlli respublikada Azərbaycan Respublikası Rabitə və Yüksək Texnologiyalar Nazirliyinin dəstəyi ilə “Elektron Azərbaycan”-nın tərkib hissəsi kimi həyata keçirilən “e-elm” layihəsi çərçivəsində nəzərdə tutulmuşdur. E-elmın şəbəkə platforması olan AzScienceNet şəbəkəsinin və böyük yaddaş və hesablama resurslarına malik (yaddaş-200 Terabayt, hesablama məhsuldarlığı - 14 Tflops) Data Mərkəzin imkanları buna texniki dəstək verir [1].

Qeyd olunanlar araşdırılan məsələnin aktullağını bir daha təsdiq edir.

II. E-ELM BÖYÜK ELMİ VERİLƏNLƏRİN GENERATORU KİMİ

E-elm mürəkkəb bir sistem olub infrastruktur, verilənlərin generasiyası, toplanması, saxlanması, emalı, axtarışı, analizi, ötürülməsi, təqdim olunması və s. kimi tərkib hissələrə malikdir. E-elmə vahid bir sistem kimi baxıldığı halda isə, onun həll etdiyi məsələlərə uyğun müxtəlif alt sistemlərdən təşkil olunduğunu görmək olar. Elmin informasiya təminatı; elmmetrik (elmmetriya, bibliometriya, vebometriya) təhlillər; intellektual analiz və Big data alt sistemləri böyük həcmli elmi verilənlərin yaranmasına gətirməklə yanaşı, eyni zamanda onun aktual problemlərinin həllini həyata keçirir [2]. E-elmın müxtəlif elm sahələrində tətbiqi böyük həcmli elmi verilənlərin sürətlə artmasında generator rolunu oynayır.

Böyük verilənlər (*big data*) dedikdə elə böyük həcmli və mürəkkəb verilənlər toplusu başa düşülür ki, ənənəvi verilənlər bazasında nəzərdə tutulmuş standart alətlər onların real vaxtda saxlanması, idarə edilməsi və emalının öhdəsindən gələ bilmir. Məsələn olaraq dünyadakı ən böyük eksperimental qurğu olan Böyük Adron Kollayderi – yüklü hissəciklərin sürətləndiricisini göstərmək olar. Kollayder Avropa Nüvə Tədqiqatları Mərkəzində (frans. *Conseil Européen pour la Recherche Nucléaire - CERN*) tikilmişdir. Hissəciklərin toqquşması kollayderin detektorlarında milyonlarla sensorlar vasitəsilə qeyd olunur və nəticədə böyük həcmdə verilənlər yaranır. 2010-cu ildə təcrübələr zamanı toplanmış verilənlər cəmi 13 Petabayt olmuşdur. İndi isə informasiyanın həcmi saniyədə 1 Gbayt aşır. Aydın ki, bu qədər böyük həcmli informasiyanın toplanması və emalını adi hesablama - yaddaş vasitələri ilə həyata keçirmək mümkün deyildir. Beləki, ilkin filtrasiyadan sonra CERN-də ildə 25 Petabayt qədər verilənlər

saxlanılır. İnformasiyanın belə nəhəng həcmnin analizi üçün paylanmış kompüter şəbəkə texnologiyası – *GRID*-dən istifadə olunur [3]. Bundan başqa astrofizikada məşhur olan *SETI* (*Search for Extraterrestrial Intelligence*) layihəsi radioteleskoptan alınan verilənlərin paylanmış emalı vasitəsilə yerdənkənar sivilizasiyaları axtarır. Layihədə böyük həcmdə radiosiqnalların emalında istifadə olunan kompüterlərin ümumi gücü Top 500-ə daxil olan superkompüterlərin gücündən dəfələrlə çoxdur [4]. Eyni zamanda 2000-ci ildə başlanmış *SDSS* (*Sloan Digital Sky Survey*) layihəsi çərçivəsində səmanın rəqəmsal astronomik müşahidələrində (sutkada 200 Gb-a yaxın yığılma sürəti ilə) toplanmış verilənlərin həcmi hazırda 140 Terabayt keçmişdir [4]. Bunlarla yanaşı asrofizika sahəsində tanınan *LSST* (*Large Synoptic Survey Telescope*) - Böyük müşahidə teleskopu göyün əlçatan hissəsinin çəkilməsi üçün nəzərdə tutulmuşdur. Müşahidə zamanı toplanan verilənlərin həcmi (sutkada 15 Terabayt olmaqla) 10 il ərzində 60 Petabayt çatması gözlənilir. Hesablama gücünün isə 150 Teraflops-dan başlayaraq sonda 950 Teraflops-a çatdırılması planlaşdırılır [6]. Təcrübi eksperimentlərlə bol olan tibb və biologiyada aparılan tədqiqatlar müasir informasiya texnologiyaları sahəsində əldə olunan ən yeni nailiyyətlərlə sıx bağlıdır və əsasən böyük həcmli rəqəmsal informasiyanın emalına əsaslanır. Verilənlərin böyük həcmnin analizinin zəruriliyi problemi özünü ən çox bioinformatikada, DNK üzərində işləyərkən göstərir. Fərdiləşdirilmiş tibbdə xəstələr haqqında müxtəlif mənbələrdən toplanmış məlumatları özündə saxlayan elektron tibbi kartların emalı mühüm rol oynayır. Böyük həcmli tibbi verilənlərin analiz edilmə imkanı yeni müalicə üsulları tapmağa və xəstəliklərin təbiətini daha yaxşı başa düşməyə kömək edir. Riyazi modelləşdirməyə və böyük həcmli verilənlər üzərində yüksək sürətli hesablamalara əsaslanan tibbi tədqiqatlar tibbi xidmətin səviyyəsini qaldırır.

III. ELMDƏ BİG DATA TƏŞƏBBÜSLƏRİ

Elm üzrə Beynəlxalq Şuranın (*International Council for Science-ICSU*) məqsədi alimlərin beynəlxalq elmi fəaliyyətinin yayılmasına, beynəlxalq elmi birliklərin fəaliyyətinin əlaqələndirilməsinə, sahələrarası qlobal xarakterli elmi proqramların işlənilməsi və tədqiqatların aparılmasına yardım etməkdir [7]. *ICSU*-nun üzvləri 142 ölkəni təmsil edən 122 elmi qurum və 31 beynəlxalq elmi birlikdir. O, *UNESCO*-nun yanında məsləhətçi statusuna malikdir. *ICSU* fəaliyyəti ilə *Big Data* problemlərinin həllində mühüm rol oynayır. Onun elmi verilənlərə universal və bərabərhüquqlu çıxışın təmin edilməsi, onların daha yaxşı idarə edilməsi üçün yaradılmış aşağıdakı təşəbbüsləri misal göstərmək olar:

- Elmi informasiya və Verilənlərin sahə üzrə qiymətləndirilməsi Paneli;
- İnformasiya və Verilənlər üzrə Strateji Komitə;
- İnformasiya və Verilənlər üzrə Strateji Əlaqələndirmə Komitəsi.

ICSU-nun bu təşəbbüsləri Dünya Verilənlər Sisteminin (*World Data System - WDS*) inkişafında mühüm rol oynamışdır [8]. Bu sistemin məqsədləri aşağıdakılardır:

- Elmi verilənlərə universal və bərabərhüquqlu çıxışı təmin etmək;
- Verilənlərin standartlarına və konvensiyalarına əməl olunmasını həvəsləndirmək;
- Verilənlər və informasiyanın təminatlı keyfiyyətini təmin etmək;
- Verilənlərin daha yaxşı idarə edilməsi üçün şərait yaratmaq.

CODATA (*Committee on Data for Science and Technology*)- Elm və Texnologiya üçün Məlumat Komitəsi *ICSU*-nun sahələrarası komitəsi olub, onun fəaliyyətinin əlaqələndirilməsi, qlobal xarakterli problemlərin işlənilməsi və tədqiqatların aparılmasına kömək göstərir [9]. O, elm və texnikanın bütün sahələri üzrə verilənlərin idarə olunması və əlyətərliyinin yaxşılaşdırılmasına səy göstərir. *CODATA* qarşıya qoyulan məqsədə çatmaq üçün məsələ qrupları, milli üzvlərin fəaliyyəti, konfranslar, simpoziumlar, nəşrlər, ümumi maraq dairəsində digər təşkilatlarla əməkdaşlıq kimi mexanizmlərdən istifadə edir. Onun 2002-ci ildən "*Data Science Journal*"ı nəşr olunur. *CODATA*-nın *ICSU*-nun maliyyələşdirdiyi beynəlxalq əməkdaşlıq elmi proqramlarına dəstək üçün 2014-cü ildə keçirdiyi "Beynəlxalq Elmi Proqramlar üçün *Big Data*: problemlər və perspektivlər" beynəlxalq seminarı *Big Data* problemlərinin və imkanlarının daha yaxşı başa düşülməsini təmin etmək məqsədi daşımışdır. Seminar "Beynəlxalq Elmi Proqramlar üçün *Big Data*: tövsiyələr və tədbirlər haqqında" adlı yekun bəyannamə qəbul etmişdir [10]. Bəyannamədə aşağıdakılar nəzərdə tutulmuşdur:

1. Beynəlxalq elmi proqramların sponsorları üçün tövsiyələr:

- Beynəlxalq elmi proqramların *Big Data* tələblərinə cavab verməsi;
- Cəmiyyət üçün *Big Data*-nın faydalarından istifadə edilməsi;
- Beynəlxalq əməkdaşlıq vasitəsilə *Big Data*-nın anlaşılmasının təkmilləşdirilməsi;
- Qlobal tədqiqat infrastrukturunu vasitəsilə *Big Data*-ya universal çıxışın təbliği;
- *Big Data*-nın idarə olunması problemlərinin tədqiqi;
- *Big Data* sahəsində elmi potensialın və bacarığın artırılmasının təbliği;
- *Big Data*-dan istifadə siyasətinin inkişaf etdirilməsi.

2. *CODATA* işçi qrupuna təklif olunan tədbirlər:

- Beynəlxalq elmi proqramlar üçün *Big Data* sahəsində tematik tədqiqatların artırılması;
- Elm sahələri üzrə *Big Data* həllərinin mübadiləsinin təbliği;

- *Big Data* üçün siyasi, etik və hüquqi məsələlərin tədqiqi;
- *Big Data* üçün idarəçilik və sabitlik problemlərinin tədqiqi.

Maraq doğuran digər təşəbbüs Avropa Birliyinin yeni tədqiqat və innovasiya çərçivə proqramı "Horizon 2020" ilə əlaqədardır. Proqram çərçivəsində İqlim, Enerji, Qida, Səhiyyə, Nəqliyyat, Təhlükəsizlik və İctimai Elmlər sahələrində ən mühüm ictimai problemlərin həlli nəzərdə tutulmuşdur. Bu sahələrin hər birində böyük həcmdə verilənlərin toplanması, inteqrasiyası, emalı, təhlili və vizuallaşdırılması mühüm rol oynayır. Proqramın maliyyələşdirdiyi *Big Data Europe* layihəsinin məqsədi bu məsələlərin həlli üçün açıq *Hadoop*, *Cassandra* və *Storm* kimi *Big Data* texnologiyaları əsasında yeni platformanın işlənməsidir [11].

IV. BÖYÜK VERİLƏNLƏRİN EMALI ÜÇÜN QABAQCIL TEXNOLOGİYALAR

Superkompüter və klasterlər. Superkompüter çox prosessorlu xüsusi, yüksək məhsuldarlıqlı hesablama sistemidir. Klasterlər isə yüksəksürətli əlaqə kanalları ilə əlaqələndirilmiş və vahid hesablama vasitəsi kimi istifadə olunan kompüterlərdən (və ya superkompüterdən) təşkil olunur. Hesablama klasterləri verilənləri bölərək onları qovşaqlarda paralel emal etməklə, tək kompüterlə müqayisədə, hesablama vaxtını əhəmiyyətli dərəcədə azaltmağa imkan verir. Açıq proqram təminatı, sadə şəbəkə texnologiyaları və adi kompüterlər əsasında yüksək məhsuldarlıqlı belə hesablama komplekslərini yaratmaq mümkündür.

Qrid. Bu texnologiya elmi verilənlərlə intensiv əməliyyatlar aparılması üçün elmi təşkilatların öz hesablama güclərini birgə istifadə etməsi konsepsiyasıdır. Mahiyyətə böyük resurslar tələb edən məsələlərin həlli üçün müxtəlif tipli hesablama resurslarının vahid infrastrukturda birləşdiyi paylanmış hesablamaların bir növüdür. Paylanmış hesablamaların üstünlüyü ondan ibarətdir ki, sistemin qovşaqları hətta adi kompüterlərdən təşkil oluna bilər. Beləliklə, nəzəri olaraq superkompüterlərdə olan hesablama güclərini daha aşağı qiymətlə almaq olar. Təəssüf ki, böyük həcmli informasiyanın emal üçün uzaq qovşağa göndərilməsi zərurəti yarandıqda, mövcud kompüter şəbəkələrində verilənlərin ötürülmə sürətinin aşağı olması səbəbindən, bu texnologiya yararlı olmur. *Qrid* layihələrə ABŞ-ın 4 iri superkompüter mərkəzinin resurslarını birləşdirən *TeraGrid*, böyük hesablama gücü tələb edən elmi tədqiqatlar üçün *Open Science Grid*, Avropada 18 kompüter mərkəzini birləşdirən *DEISA* (*Distributed European Infrastructure for Supercomputing Applications*), Avropa İttifaqında yüksək enerjilər, ekologiya və bioinformatika sahəsində verilənlərin emalı üçün qridin qurulması üzrə *DataGrid*, Yaponiyada elmi-tədqiqat fəaliyyəti üçün qrid infrastrukturunun qurulması üzrə *NAREGI* (*National Research Grid Initiative*) və s. göstərmək olar. Hazırda qrid texnologiyasından nüvə fizikasında, ətraf mühitin qorunmasında, hava proqnozunda və iqlim dəyişikliklərinin modelləşdirilməsində, maşınqayırmanın

modelləşdirilməsində, bioloji modelləşdirilmədə və s. geniş istifadə olunur.

Hesablama buludları (Cloud Computing). İstifadəçilərin internet vasitəsilə xidmətlərə, hesablama resurslarına və proqramlara (əməliyyat sistemləri və infrastruktur daxil olmaqla) məsafədən girişini təmin edir. Böyük həcmli verilənlərlə iş zamanı onların əvvəlcədən eyni bir yerdə toplanması və sonra isə analiz edilməsi mümkün deyildir. Doğrudan da, həcmi Petabaytlarla ölçülən verilənlərin analizi zamanı onların şəbəkə vasitəsilə, misal üçün, bir tədqiqat mərkəzindən digərinə ötürülməsi hətta yüksək məhsuldarlıqlı şəbəkələr vasitəsilə də effektiv həyata keçirilə bilməz. Ona görə verilənlər öz yerlərində qalır, bulud texnologiyaları isə onların müxtəlif yerlərdən seçilməsini və emalı üçün zəruri hesablama gücünü təqdim etməklə analizinin aparılmasını təmin edir. Beləliklə, bulud texnologiyaları bir tərəfdən böyük həcmli informasiyanın toplanması və saxlanması mümkün etməli, digər tərəfdən isə bu verilənlərin emalı üçün zəruri prosessor gücünü təqdim etməlidir.

NƏTİCƏ

Big Data texnologiyaları elmdə çoxlu problemlər yaratmaqla yanaşı, eyni zamanda onun inkişafı üçün də yeni imkanlar açır. Böyük həcmli elmi verilənlərin professional idarə edilməsi və analizi nəticəsində yeni biliklərin əldə olunması və elmin sürətli inkişafı üçün yeni imkanlar yaranır. Böyük həcmli elmi verilənlər probleminin həllinin perspektiv istiqaməti hesablama buludları texnologiyasından istifadədir. Bu yolla istifadəçilər böyük həcmli informasiyaya uzaqdan birgə giriş imkanı əldə edir və onların emalı üçün paylanmış hesablama resurslarından daha səmərəli istifadə edir.

TƏŞƏKKÜR NƏMƏ

Bu iş Azərbaycan Respublikasının Prezidenti yanında Elmin İnkişafı Fondunun maliyyə yardımı ilə yerinə yetirilmişdir – **Qrant № EIF-2014-9(24)-KETPL-14/02/1**

ƏDƏBİYYAT

- [1] R.M.Əliquliyev, R.Q.Ələkbərov, T.X.Fətəliyev. Elektron elm: cari vəziyyəti, problemləri və perspektivləri // *İnformasiya Texnologiyaları Problemləri*, 2015, №2, səh.4-15.
- [2] T.X.Fətəliyev. Elektron-elmin informasiya təhlükəsizliyi haqqında // *Beynəlxalq Telekommunikasiya İttifaqının 150 illiyinə həsr olunmuş "İnformasiya təhlükəsizliyinin multidissiplinar problemləri"* üzrə II respublika elmi-praktiki konfransı, Bakı, 14 may 2015, səh.145-147.
- [3] Search for Extraterrestrial Intelligence. <http://setiathome.berkeley.edu>
- [4] Brumfiel G. High-energy physics: Down the petabyte highway // *Nature*. 2011. V. 469. № 7330. pp. 282–283.
- [5] The Sloan Digital Sky Survey. www.sdss.org
- [6] The Large Synoptic Survey Telescope. www.lsst.org/lstt
- [7] International Council for Science. www.icsu.org
- [8] World Data System. www.icsu-wds.org
- [9] CODATA-Committee on Data for Science and Technology. www.codata.org
- [10] Big Data for International Scientific Programmes: Challenges and Opportunities. <http://codata.org/blog/wp-content/uploads/2014/06/CODATA-Big-Data-Workshop-STATEMENT-v07-FINAL.pdf>
- [11] Big Data Europe. www.big-data-europe.eu