

## UOT 004.9

*Aliquliyev R.M.<sup>1</sup>, İmamverdiyev Y.N.<sup>2</sup>, Abdullayeva F.C.<sup>3</sup>*

AMEA İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan

<sup>1</sup>[r.aliguliyev@gmail.com](mailto:r.aliguliyev@gmail.com), <sup>2</sup>[yadigar@lan.ab.az](mailto:yadigar@lan.ab.az), <sup>3</sup>[farqana@iit.ab.az](mailto:farqana@iit.ab.az)

### NEFT-QAZ SƏNAYESİ ÜÇÜN BIG DATA ANALİTİKANIN CLOUD COMPUTING PLATFORMASINDA ANALYTICS-AS-A-SERVICE KİMİ REALLAŞDIRILMA İMKANLARININ TƏDQIQI

*Neft-qaz sənayesində toplanan verilənlərin həcmnin artması bu sektorda ciddi problemlərin meydana çıxmasına səbəb olmuşdur. Məqalədə böyük həcmli verilənlərin neft-qaz sənayesində yaratdığı problemlər, bu sahədə big data analitikanın tətbiqinin müasir vəziyyəti təhlil edilir. Nəhəng təşkilatların neft-qaz sənayesi üçün istehsal etdiyi big data analitika platformaları və dünyanın iri neft-qaz şirkətlərinin big data analitika sahəsində təcrübəsi araşdırılır. Big data analitika üçün böyük həcmli verilənlərin analizi üsulları və cloud computing platformasından istifadənin məqsədləri tədqiq olunur. Neft-qaz sənayesində big data analitikanın Analytics-as-a-Service xidməti kimi reallaşması üçün təklif və tövsiyələr verilir.*

**Açar sözlər.** *Big data analitika, OLAP, cloud computing, Analytics-as-a-Service, neft-qaz kəşfiyyatı və hasilatı, Hadoop, MapReduce, verilənlər elmi.*

#### Giriş

Rəqəmsal dövrdə sensor qurğularının geniş istifadəsi bioinformatika, sosial şəbəkələr, neft-qaz sənayesi kimi sahələrdə böyük verilənlər çoxluğunun toplanmasına səbəb olmuşdur. Bu tip kütləvi verilənlərin az vaxt sərf etməklə sürətli analiz edilməsi hazırkı biznes mühitinin əsas rəqabət qabiliyyətlilik cəhətidir. Lakin hazırda toplanan verilənlərin məlumat “sel”i halı alması analitikanın qarşısında böyük problemlər qoymuşdur [1].

Müasir cəmiyyətdə böyük verilənlər probleminə məruz qalan sahələrdən biri də dövlət sektorudur [2]. İnkişaf etmiş ölkələrdə dövlət sektorunda böyük verilənlərin meydana gətirdiyi çətin problemlərin tədqiqinin aparılması istiqamətində ciddi addımlar atılır. Bu məqsədlə 2012-ci ildə Obama (ABŞ prezidenti) administrasiyası böyük verilənlərin öyrənilməsi və inkişaf etdirilməsi üzrə təşəbbüs irəli sürmüşdür [3].

İnformasiya və kommunikasiya texnologiyalarının neft və qaz sənayesi sahəsinə tətbiqi ilə əldə olunan verilənlər 21-ci əsrin iqtisadiyyatının (rəqəmsal iqtisadiyyatın) *nefti* kimi qiymətləndirilir. [4]-də qeyd edilir ki, informasiya texnologiyaları əvvəlki dövrlərə nisbətən hər bir qurğudan (aktivdən) daha çox barel neft əldə etməyə imkan verir. Neft-qaz şirkətlərində istifadə olunan qurğuların təkmilləşdirilməsi, proseslərin avtomatlaşdırılması və bir sıra əməkdaşlığın təşkili bu sektorda verilənlərin həcmnin artmasını daha da sürətləndirmişdir. Bəzi ekspertlər hesab edir ki, verilənlərin həcmi hər il beş dəfə artır [5]. Beynəlxalq verilənlər korporasiyasının 2013-cü il hesabatında neft və qaz sənayesi üçün verilənlərin həcmi 2,7 zetabayt həcmində qiymətləndirilmişdir. Neft sənayesində verilənlər sensorlardan, fəza və qlobal mövqetəyinetmə sistemi koordinatlarından, metereoloji xidmətlərdən, seysmik verilənlərdən və başqa ölçmə qurğularından toplanır [6]. Verilənlərin toplandığı digər mənbələr sosial media, elektron poçt, mətn, şəkillər və multimedia hesab olunur. Burada verilənlər strukturlaşdırılmamış və ya yarım-strukturlaşdırılmış şəkildə mövcuddur. Bu səbəbdən verilənləri ənənəvi verilənlər anbarında saxlamaq bu verilənlərə müntəzəm müraciət və analiz etmək olduqca mürəkkəbdir və ya bahalıdır. Əsasını innovativ analitika təşkil edən big data analitika bu tip verilənlərin emalı üçün yeni həllərin yaradılmasında başlıca mövqeyə malikdir.

Big data analitika rəqabət üstünlüyü yarada bilən, yeni ideyalar çıxarmağa, böyük mənfəət əldə etməyə imkan verən dəyərli biliklərin əldə olunmasını təmin edən texnologiya kimi qiymətləndirilir [7].

Big data analitikanın neft sənayesində tətbiqi neft kəşfiyyatı və hasilatında qabaqlayıcı qərarlar qəbul etməyə, neftin aşkarlanması sürətini artırmağa, bu prosesə sərf olunan xərcləri azaltmağa və s. imkan verir.

Analitik sistemlərin heç birinin tələb olunan əlçatırlığı və miqyaslılığı təmin edə bilməməsi; analitiklər, sahə ekspertləri, iş prosesinin icraçıları, sorğu icraçıları kimi müxtəlif istifadəçi rollarını özündə cəmləşdirməməsi; analitik proqram tətbiqlərinin təşkilatın infrastrukturunda quraşdırılması; onların satın alınmasının, konfigurasiya və idarə edilməsinin istifadəçinin özü tərəfindən həyata keçirilməsi; bu sistemlərin istifadəçinin dinamik dəyişə bilən tələbləri ilə qarşılaşa bilməməsi big data analitikanın cloud computing texnologiyası ilə birgə istifadəsini zəruri edən şərtlərdəndir. Sadalanan xüsusiyyətlərə malik analitik sistemlər yaratmaq, istifadəçini dinamik analitika vasitələri ilə təmin etmək üçün buludların təqdim etdiyi analitika xidmətindən (*Analytics-as-a-Service, AaaS*) istifadə olunması əvəzolunmaz vasitə hesab olunur. AaaS vahid analitik sistem və ya proqram deyil, servislər təqdim edən platformadır. AaaS platformasının tərkibinə çox sayda analitik proqramlar daxil olur və adətən, onu servis səviyyəsi müqaviləsinə əsaslanan analitik idarəetmə sistemi kimi şərh edirlər. Burada əsas məqsəd buludların girişinin istənilən məkandan təşkili, verilənləri paylaşma və miqyaslılıq kimi xüsusiyyətlərindən istifadə etməklə, müxtəlif rollarda çıxış edən istifadəçilər çoxluğuna xidmət etmək üçün analitik sistemlərin buludda yerləşdirilməsidir.

Big data analitikanın müştərilərə AaaS xidməti şəklində təqdim olunmasına həsr olunmuş bir sıra yanaşmalar irəli sürülmüşdür [8–10]. Lakin təklif edilən yanaşmalarda Big Data analitikanın neft-qaz sənayesi üçün AaaS xidməti kimi reallaşması imkanlarının tədqiqinə baxılmamışdır.

Təqdim olunan məqalədə verilənlər analitikası, təkmilləşdirilmiş analitika, big data analitika, verilənlər elmi, verilənlər üzrə alim anlayışlarının mahiyyəti tədqiq olunur. Big data analitikanın formalaşması istiqamətində təklif edilmiş yanaşmaların geniş analizi aparılır. Big data analitikanın neft-qaz sənayesində tətbiqinin mövcud vəziyyəti araşdırılır. Neft-qaz sənayesində big data analitikanın bulud xidməti kimi reallaşması üçün bir sıra təklif və tövsiyələr verilir.

### **Big data analitika sahəsində bəzi anlayışlar**

Big data analitikaya tərif verməzdən öncə aşağıdakı anlayışlara aydınlıq gətirək.

*Analitika* – qərar qəbulunu dəstəkləmək üçün analiz və çıxarış üsullarını əhatə edir. Bu üsullar verilənlərdən müəyyən məna kəsb edən obrazı, baş verən dəyişiklikləri aşkarlamağa, insanların bu dəyişikliklərə reaksiyasını müəyyən etməyə imkan verir [11].

*Təkmilləşdirilmiş analitika* – gələcək nəticələri proqnozlaşdırmaq üçün istifadə olunan analitik üsullar qrupudur [12]. Bu baxımdan təkmilləşdirilmiş analitika yeni müştəri segmentlərinin aşkarlanması, ən yaxşı provayderin müəyyən edilməsi, satış mövsümünün müəyyən edilməsi və s. kimi məsələlərin həllində əvəzolunmaz vasitə hesab olunur. Adətən proqnoz analitikası, verilənlərin əldə olunması (*data mining*), statistik analiz, mürəkkəb SQL, verilənlərin vizuallaşdırılması, süni intellekt, mətn analitikası və s. kimi vasitələr təkmilləşdirilmiş analitika qrupuna aid edirlər.

Ədəbiyyat mənbələrində təkmilləşdirilmiş analitika əvəzinə bəzi hallarda aşkarlama analitikası terminindən də istifadə olunur. Bütün bu üsullar uzun müddətdir mövcuddur, onların bəziləri 1990-cı illərdə meydana gəlmişdir. Bu üsullar çox böyük, multiterabaytlarla ölçülən verilənlər çoxluğunu analiz etməyə yaxşı adaptasiya olunduqları üçün hazırda böyük verilənlərin analizində geniş tətbiq olunurlar.

*Big data analitika* – təkmilləşdirilmiş analitika üsullarının böyük verilənlər çoxluğuna tətbiqidir [13]. Tərifdən görünür ki, böyük verilənlər analitikası iki obyektə əhatə edir: böyük verilənlər və analitika. Digər tərifdə *big data analitika* – müxtəlif tipli verilənlərdən (strukturlaşmış və strukturlaşmamış) keyfiyyətli bilik əldə etmək üçün yaradılmış, verilənlərin əldə edilməsini, aşkarlanmasını və analizini yüksək sürətlə həyata keçirməyə

imkan verən yeni nəsil texnologiyadır [14].

Big data analitikanın məqsədi şəxslərin əhəmiyyətli qərarlar qəbul etməsinə şərait yaratmaqdır. Bu, iri həcmli tranzaksiya verilənlərini analiz edə bilən verilənlər üzrə alimlərin (data scientists) cəlb edilməsi yolu ilə həyata keçirilir. Bu da öz növbəsində *verilənlər üzrə alimin* və *verilənlər elminin* meydana gəlməsinə zərurət yaradır.

*Verilənlər elmi* – müxtəlif elementlərdən ibarətdir və bir çox sahələrə aid üsulların və nəzəriyyələrin üzərində qurulur [15]. Bu üsullar riyaziyyat, riyazi statistika, verilənlər mühəndisliyi, obrazların tanınması, təkmilləşdirilmiş texnologiyalar (*advanced computing*), vizuallaşdırma, qeyri-müəyyənliyin modelləşdirilməsi, verilənlər anbarı (*data warehouse*), yüksək məhsuldarlıqlı texnologiyalardan (*high performance computing*) ibarətdir (şəkil 1).

*Verilənlər elminin məqsədi* – bütün əlçatar və müvafiq verilənlərdən istifadə edərək, qeyri-ekspertlərin asanlıqla başa düşə biləcəyi formada fikirlər əldə etməkdir. Verilənlər elmini icra edən şəxs *verilənlər üzrə alim* adlandırılır. Verilənlər üzrə alimin analitika sahəsində böyük bacarıqları olmalıdır və bəzi əlaqədar sahələrin ekspertləri vasitəsi ilə verilənlərlə bağlı mürəkkəb problemləri həll edə bilməlidir.

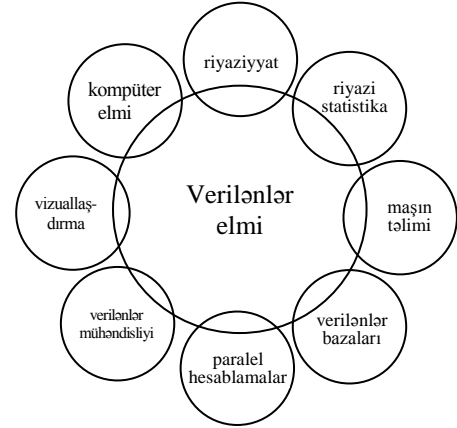
*Bulud texnologiyaları* – paralel və paylanmış sistemin bir növüdür, qarşılıqlı əlaqəli və virtuallaşmış kompüterlər çoxluğundan ibarətdir, bu kompüterlər dinamik yaradılır və müştərilərə servis provayderi ilə müştəri arasında razılışma vasitəsi ilə qurulan servis səviyyəsi müqavilələri əsasında bir və ya bir neçə vahid hesablama resursları şəklində təqdim olunur [16].

### Big data analitika üzrə tədqiqatların analizi

Günümüzə qədər mövcud olan verilənlər analitikası mürəkkəb hesablama apararaq elmi ixtiralar üçün biliklər aşkarlamaqla, öz potensialını artırıq maliyyə, administrativ, elm sahələrində qərar qəbulunun dəstəklənməsi istiqamətində sübut etmişdir. Sənaye və elmi cəmiyyət böyük verilənlərin meydana gətirdiyi problemlərlə qarşılaşmaq məqsədi ilə ənənəvi analitika metodlarında paylanmış fayl sistemi (*Hadoop*), *MapReduce* proqramlaşdırma modeli və digər başqa texnologiyaları tətbiq etməklə, alqoritmlərdə paralelləşmə əldə edərək big data analitikasını formalaşdırırlar.

Hazırda elmi cəmiyyətin böyük diqqətini cəlb edən məsələlərdən biri klasterləşmənin paralellik xüsusiyyəti almasına nail olmaqdır. Paralel klasterləşmə böyük verilənlərin klasterləşməsi məsələsində ən effektiv üsul hesab olunur. Paralel klasterləşməni bir-birindən asılı olmayan hesablama prosedurlarını bir neçə hesablama prosessorunda paralel şəkildə yerinə yetirməklə həyata keçirirlər. [17]-də müxtəlif paralelləşmə strategiyalarının tədqiqi aparılır, böyük verilənlərin klasterləşməsi üçün PIC (*Power iteration clustering*) adlı alqoritmin paralelləşdirilməsi üçün yanaşma verilir. DBCURE-MR sıxlığa əsaslanan klasterləşmə alqoritm təklif edilmişdir [18]. Bu alqoritm müxtəlif sıxlıqlı klasterləri tapmağa imkan verir və paralelləşmə xüsusiyyətinə malik olması üçün *MapReduce* modelinə əsaslanır. Sıxlığa əsaslanan ənənəvi alqoritmlər hər bir klasteri bir-bir tapır, lakin DBCURE-MR alqoritm bir neçə klasteri birlikdə paralel olaraq tapır.

Ənənəvi OLAP (*On-Line Analytical Processing*) texnologiyası böyük verilənlərin meydana gətirdiyi problemləri dəf edə bilmir. Bu səbəbdən [19]-də böyük verilənlər üçün Haolap adlı OLAP sistemi yaradılmışdır. Haolap (*Hadoop based OLAP*) çoxölçülü OLAP



Şəkil. Verilənlər elmi üçün tələb olunan bacarıqlar

texnologiyalı sistemdir, Hadoop paylanmış fayl sistemi və *MapReduce* alqoritminə əsaslanır. Bu sistem obyektləri və ölçmələri yadda saxlamaq üçün xüsusi çoxölçülü modeldən istifadə edir. *MapReduce* alqoritminin hesabına sistem OLAP əməliyyatlarını böyük miqyaslı verilənlər üzərində həyata keçirə bilir.

Əksər sahələrdə qraf strukturu o qədər böyük qurulur ki, onların emalı xüsusi üsullar, əsasən də paralelləşmə xüsusiyyətinə malik üsullar tələb edir. Böyük qrafların paralel emalı prosesini sınaqdan keçirmək üçün Hadoop mühitində əlçatar olan *MapReduce* (MR), *MapReduce2* (MR2) və *Bulk Synchronous Parallel* (BSP) adlı mövcud metodların empirik müqayisəsi aparılmışdır. [20]. Nəticə olaraq *MapReduce* modeli ilə müqayisədə qrafların problemlərini aradan qaldırmaqda ən yaxşı model BSP seçilmişdir. *MapReduce* şəbəkə halına salınmış verilənlərin emalı zamanı bir sıra problemlərlə qarşılaşır. Bu problemləri dəf etmək məqsədi ilə *Google* şirkəti *Pregel* adlı sistem yaratmışdır. *Pregel* sistemi qrafları buludlarda emal etmək üçün BSP modelindən istifadə edir [21].

[22] verilənlər çoxluğunun referatlaşdırılması (*data set summarization*) məsələsini sürətləndirmək məqsədi ilə əsas komponentlərin analizi (*Principal Component Analysis, PCA*) modelinin paralel hesablanması qaydalarını tədqiq edir. PCA böyük ölçülü verilənlərin ölçüsünün azaldılması və onlardan əlamət çıxarılmasına imkan verən geniş yayılmış üsuldur. Bu tədqiqat işində ardıcılıq xüsusiyyətinə xas olan mövcud metodun paralel alqoritm halına salınmasına cəhd olunur və o, referatlaşdırma metrikalarından istifadə edərək, böyük verilənlər çoxluğunda PCA-nı bir keçidə hesablaya bilir. Burada, həmçinin alqoritmın VBİS (verilənlər bazasının idarəetmə sistemləri) ilə inteqrasiyası məsələsinə də baxılmışdır.

Verilənlərin analizini paralel şəkildə emal etməyə imkan verən sistemlərdən biri də *Ophidia* sistemidir [23]. *Ophidia* elektron elm sahəsində big data analitika rolunda çıxış edir. Sistemdə iqlim dəyişməsi, astrofizika, bioinformatika və s. sahələrə aid iri həcmli çoxölçülü elmi verilənləri idarə etmək üçün verilənləri paralel analiz edən komponent, daxili saxlanma və verilənlərin iyerarxik təşkili üslubu nəzərdə tutulmuşdur.

Böyük verilənlərin əsas xarakteristikalarını açıqlayan HACE (*Heterogeneous, Autonomous, Complex, Evolving*) adlı teorem və onun verilənlərin intellektual analizi (data mining) baxımından emalı prosesini əks etdirən model verilmişdir [24]. HACE teoreminə görə böyük verilənlər, onların mənbəyinin qeyri-bircins (heterogeneous) olması səbəbindən nəhəngdir, paylanmış və mərkəzləşməmiş idarəetmə mexanizminə əsaslandığına görə avtonomdur, verilənlər arasında əlaqələrə görə mürəkkəb və inkişaf edəndir. Bu xarakteristikalar böyük verilənlərdən yararlı biliklər əldə edilməsinə ciddi maneə törədir. Buna görə də bu maneələri aradan qaldırmaq məqsədi ilə müəlliflər tərəfindən HACE xarakteristikalarını nəzərə alan böyük verilənlərin data mining baxımından emalı prosesini əks etdirən model verilmişdir.

[25]-də iqlimin öyrənilməsi prosesində visual analitika yanaşmasının istifadəsinin iqlim sahəsində toplanan böyük verilənləri daha effektiv analiz etməyə imkan verdiyi iddia edilir. Burada EDEN (*Exploratory Data analysis ENvironment*) adlı vizual analiz sisteminin yer sistemini simulyasiya edən verilənlər çoxluğunun analizi prosesinə tətbiq olunması məsələsinə baxılmışdır.

Böyük verilənlərin meydana gəlməsi ilə böyük diqqət çəkən texnologiyalardan biri də mətnlərin intellektual analizi (text mining) texnologiyasıdır. [26]-də mətn analitikası və text mining texnologiyasının hüquqi, biznes analitikası və təhlükəsizlik kimi müxtəlif tərəflərinin böyük verilənlər baxımından dərin analizi aparılmışdır.

[27]-də böyük verilənlərin mövcud problemlərinin, metodlarının və texnologiyalarının geniş analizi aparılmışdır. [28]-də paralel və paylanmış sistemlərin big data analitikaya tətbiqi məsələlərinin müasir vəziyyəti araşdırılmışdır.

Son zamanlar grid və bulud texnologiyaları ilə big data analitika arasında asılılığın artması da aşkar şəkildə müşahidə olunur. Belə ki, dünyanın qabaqcıl "*Future Generation*

*Computer Systems*” adlı jurnalının 2014-cü il xüsusi buraxılışına [29] böyük verilənlər arxitekturu, böyük verilənlərin emalı sistemləri, böyük verilənlərin idarə edilməsi, böyük verilənlərin tətbiqləri və modelləşdirilməsi, *MapReduce* optimallaşdırma, resursların paylanması, resursların monitorinqi, resursların enerjiyə qənaət olunmaqla istifadə olunması kimi məsələləri əhatə edən işlər daxil edilmişdir. Bu məqalələr sahəyə aid yeni ideyaları və mövcud metodları əhatə etməklə yanaşı, həmçinin gələcək tədqiqatları da şərh edir.

Böyük verilənlərin analizində bulud texnologiyalarının əhəmiyyətini ön plana çəkən tədbirlərdən biri də 17-18 oktyabr 2013-cü ildə dünyanın *British Petroleum, Shell, Anadarko, Chevron* kimi nəhəng neft şirkətlərinin verilənlər üzrə aparıcı alimlərini bir araya yığan “Big data analytics for oil&gas” adlı konfransı olmuşdur. Konfransda verilənlər üzrə alimlər xüsusən seysmoloji proseslərdə Hadoop ekosistemindən istifadə qaydaları haqqında, daha böyük verilənlərin analizində bulud texnologiyalarının əhəmiyyətindən məruzələr təqdim etmişlər. Bundan əlavə, 19-20 may 2015-ci il tarixində “*Big Data Energy Innovation Summit*” adlı sammiti keçirilmişdir. Sammitin əsas mövzuları böyük verilənlərin enerji sektoruna verəcəyi töhfələrdən bəhs edir.

Bu sahədə də müəyyən tədqiqat işləri aparılmışdır. Qərar qəbulunu dəstəkləyən servis-yönümlü sistem yaratmaq üçün (*service-oriented decision support systems, SODSS*) konseptual model təklif edilmişdir [8]. Burada SODSS arxitekturu 3 laydan ibarət təşkil olunmuşdur: IaaS, SaaS və business process (BP). Arxitekturaya görə, istifadəçilərə əməliyyat sistemləri, informasiya saxlanları, onlayn analitik emal, kompüter vasitələri servis şəklində təqdim edilə bilər. Arxitekturanın verilənlər xidməti (*Data as a Service, DaaS*) blokunda verilənlərin istənilən məkanda yerləşdirilə bilməsi və emal üsullarının 24 saat ərzində əlçatar olması nəzərdə tutulur. İnformasiya xidməti (*Information as a service, IaaS*) bloku informasiyaya sürətli girişi təmin edir. Analitika xidməti blokunda bulud texnologiyaları analitik məsələlərin həllində istifadə olunur. Bu xidməti bəzən çevik analitika (*Agile Analytics*) da adlandırırlar. Miqyaslılıq və xərclərin az olması çevik analitikanın əsas üstünlüklərindəndir.

Elmi araşdırma sistemlərinin buludlarda icra edilməsi, modelləşdirilməsi və monitorinqi sahəsindəki təcrübəni bölüşmək məqsədi ilə Kanadanın Ottava şəhərində 2012-ci ildə “Bulud Texnologiyaları və Elmi Tətbiqlər” (*Cloud Computing and Scientific Applications*) adlı tədbir keçirilmişdir. Burada məruzələr əsasən elmi tədqiqat sistemlərinin imkanlarını buludların istifadəsi ilə genişləndirməyə cəhd edən ekspertlər tərəfindən təqdim edilirdi. Tədbirin planının əsas mövzusu “Buludlarda böyük verilənlər”, “Buludlarda elmi hesablamalar”, “Buludlarda sosial hesablamalar” və s. kimi başlıqlar təşkil edirdi. Böyük verilənlərin analizində buludlardan istifadə zamanı alimlər bir sıra ciddi problemlərlə qarşılaşırlar. Bu problemlərlə daha ətraflı ədəbiyyat [9]-da tanış olmaq olar.

Bulud texnologiyalarının big data analitikanın səmərəliliyinin artırılmasında ciddi rolu olduğu [10]-da geniş müzakirə edilmişdir. Burada bulud texnologiyaları mühitində böyük verilənlərin meydana gəlməsinin əhəmiyyəti, böyük verilənlərlə bulud texnologiyaları, böyük verilənlərin saxlanma sistemləri və Hadoop texnologiyası arasında olan əlaqələr müzakirə olunur. Böyük verilənlərin buludlarda istifadəsi zamanı miqyaslılıq, əlçatarlıq, verilənlərin bütövlüyü, verilənlərin ötürülməsi, verilənlərin keyfiyyəti, verilənlərin heterogenliyi, gizlilik, verilənlərin hüquqi və tənzimləmə məsələləri, verilənlərin idarə edilməsi ilə bağlı elmi-tədqiqat problemləri dərinlənən araşdırılır.

### **Big data analizi üsulları**

Big data analizi üsulları riyazi statistika, verilənlərin intellektual analizi, maşın təlimi, neyron şəbəkələri, sosial şəbəkə analizi, siqnalların emalı, obrazların tanınması, optimallaşdırma və vizuallaşdırma üsulları kimi istiqamətləri əhatə edir (şəkil 2) [27]:

• *Optimallaşdırma üsulları.* Kəmiyyət qiymətləndirilməsi tələb olunan məsələlərin həllində tətbiq olunur.

• *Riyazi statistika.* Verilənlərin toplanması, təşkili və interpretasiyasından bəhs edən elmdir. Statistik üsullar müxtəlif obyektlər arasında korrelyasiya münasibətlərini və səbəb-nəticə əlaqələrini müəyyən etmək üçün istifadə olunur. Statistik üsullar vasitəsi ilə ədədi şərhlər də aparmaq mümkündür. Lakin standart statistik üsullar böyük verilənlərin idarə edilməsi məsələlərində yetərli hesab olunmurlar. Bu səbəbdən əksər tədqiqatçılar bu tip verilənləri emal etmək üçün klassik üsulların genişləndirilməsi istiqamətində yanaşmalar təklif edirlər, ya da tamamilə yeni üsullar işləyib hazırlayırlar.

• *Data mining.* Verilənlərdən yararlı informasiya (obraz) əldə etməyə imkan verən üsulların toplusudur. Bu tip üsullara klaster analizi, klassifikasiya, reqressiya və assosiativ qaydaların öyrənilməsi üsulları aid edilir. Verilənlərin intellektual analizi maşın təlimi və riyazi statistika üsullarından istifadə edir. Verilənlərin ənənəvi intellektual analizi alqoritmlərinin tətbiqi ilə böyük verilənləri analiz etmək olduqca mürəkkəb məsələdir. Buna görə, bu sahədə də böyük verilənləri emal etmək üçün mövcud metodları genişləndirməyə cəhd edirlər.

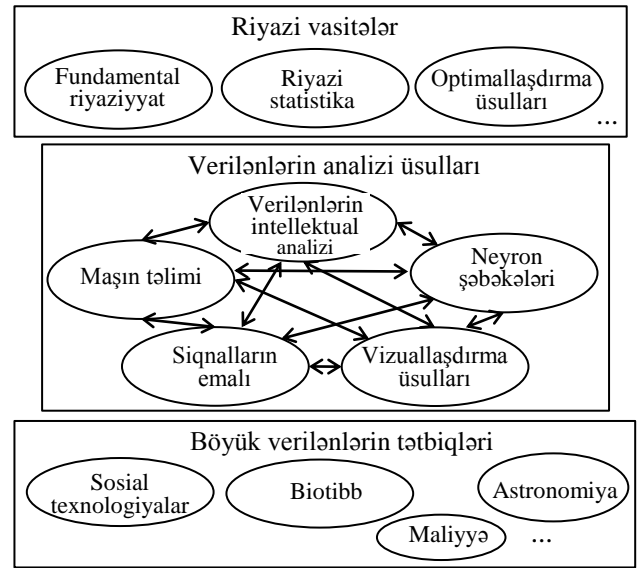
• *Maşın təlimi.* Süni intellektin mühüm bölməsidir, məqsədi empirik verilənlər əsasında davranışı aşkarlamağa xidmət edən alqoritmlər formalaşdırmaqdır. Maşın təliminin ən bariz xüsusiyyəti onun bilikləri aşkarlaması və avtomatik şəkildə intellektual qərarlar qəbul etməsidir. Böyük verilənlərin öhdəsindən gəlmək üçün hər iki tip: həm müəllimlə öyrənmə (supervised learning) və həm də müəllimsiz öyrənmə (unsupervised learning) maşın təlimi alqoritmlərini genişləndirməyə cəhd edirlər.

• *Vizuallaşdırma üsulları.* Verilənləri qavramaq məqsədi ilə cədvəlləri, şəkilləri, diaqramları və başqa intuitiv görüntüləri yaratmaq üçün istifadə edilən üsullardır. Böyük verilənlər həcm (volume), sürət (velocity) və müxtəliflik (variety) (bunu 3V də adlandırırlar) kimi problemlərə malik olduqları üçün [30, 31], onların vizuallaşdırılması kiçik ölçülü ənənəvi relyasion verilənlər çoxluğunun vizuallaşdırılması kimi asan olmur. Ənənəvi vizuallaşdırma üsullarının genişlənməsi istiqamətində bir sıra işlər aparılmışdır. Lakin bu vasitələr böyük verilənlərin emalı üçün yetərli hesab olunmur. Bu səbəbdən əksər tədqiqatçılar böyük verilənləri vizuallaşdırmaq üçün xüsusiyyətlərin (əlamətlərin) çıxarılması və həndəsi modelləşdirmədən istifadə etməklə ilk öncə onların ölçüsünü əhəmiyyətli dərəcədə azaltmağa çalışırlar

• *Sosial şəbəkə analizi.* Müasir sosiologiyanın əsas üsullarından biridir, sosial münasibətləri şəbəkə nəzəriyyəsi anlamında təsvir edir. Sosial şəbəkəni qovşaqlar və onlar arasındakı əlaqələr təşkil edir.

• *Neyron şəbəkələri.* Neyron şəbəkələri verilənlərin saxlandca yerləşdirilməsinə zərurət qoymadan onlayn təlimlənmən alqoritmləri özündə cəmləşdirir. Neyron şəbəkələri həm axın təşkil edən verilənlərin və həm də saxlandakı böyük verilənlərin emalında yararlıdırlar. Neyron şəbəkə alqoritmləri paralel hesablamaları çox yüksək dəqiqliklə həyata keçirir, belə ki, bu imkan digər maşın təlimi texnologiyalarında nəzərə alınmamışdır.

• *Siqnalların emalı.* Müasir dövrdə elektron sensor qurğularından istənilən məkanda istifadə olunması siqnalların emalı nəzəriyyəsini böyük verilənlərin vacib bir sahəsinə çevirmişdir. *Siqnalların emalı* siqnal şəklində təyin olunmuş fiziki, simvolik və ya abstrakt formatlarda mövcud



Şəkil 2. Big data analizi üsulları

olan informasiyanın emalı və ya ötürülməsi nəzəriyyəsini, proqram tətbiqlərini, alqoritmlərini əhatə edən texnologiyadır. Siqnalları təsvir etmək, modelləşdirmək, analiz etmək, sintez etmək, aşkarlamaq, çıxarmaq, məhkəmə ekspertizasını aparmaq üçün riyazi, statistik, hesablama, evristik və linqvistik üsullardan istifadə edir. Burada müxtəlif qaynaqlar verilənləri kəsilməz olaraq real vaxt rejimində generasiya etdiyindən, verilənlərin analizi keçmiş yazılara müraciət etmədən “anındaca” həyata keçirilməlidir. Bu səbəbdən əsas komponentlər analizi (principal component analysis), lüğət təlimi (dictionary learning), sıxılma üçün nümunələr (compressive sampling) və s. kimi mövcud siqnal emalı analitikalarının bugünkü geniş miqyaslı verilənlər rejiminə uyğunlaşdırılması istiqamətində ciddi tədqiqatlar aparılır.

Yuxarıda sadalanan metodlar adətən verilənlərin aid olduğu spesifik sahəyə xidmət edir. Hər hansı predmet sahəsinə aid verilənlər analitikası texnologiyası üzrə ekspertlər sahə haqqında tələb olunan bilik sahəsinə malik olmadıqda verilənləri analiz prosesində ciddi problemlərlə qarşılaşırlar. Bundan əlavə, verilənlərin analizi üçün müvafiq aparat və proqram təminatı resurslarının seçilməsi hətta ekspert qismində çıxış edən verilənlər üzrə alimlər üçün də çətinlik yarada bilər.

### Cloud computing və Big Data konvergensiyası

Big data və cloud computing hazırda dünya təşkilatlarının ən birinci önəm verdiyi informasiya texnologiyaları təşəbbüsləridir [7]. Big data texnologiyasını, adətən, cloud computing texnologiyaları ilə əlaqələndirirlər. Çünki bu texnologiya böyük verilənlər çoxluğunun paylanmış klasterlərdə saxlanmasını təmin etmək üçün Hadoop və müxtəlif qaynaqlardan götürülmüş verilənlərin birləşdirilməsini və emalını təmin etmək üçün MapReduce kimi platformaların olmasına ehtiyac duyur.

*MapReduce* böyük həcmdə verilənləri aşağı səviyyəli kompüterlərdə paralel emal etməyə imkan verən miqyaslanma xüsusiyyətinə malik imtinalara dayanıqlı sistemdir [32]. MapReduce ilk dəfə Google şirkəti tərəfindən 2004-cü ildə yaradılmışdır. Perspektivli imkanlarına görə həm elmi, həm də sənaye təşkilatlarının böyük diqqətini cəlb etmiş və geniş tətbiq olunmuşdur. Bulud texnologiyaları kontekstində infrastruktur resurslarını sorğu əsasında təqdim etmək mümkün olduğundan, MapReduce sistemi daha yaxşı miqyaslanma bilən və səmərəli hesab olunur. Sadəlik, miqyaslılıq, imtinalara dayanıqlılıq MapReduce yanaşmasının üç əsas qabarıq xüsusiyyətidir. Bu səbəbdən böyük verilənləri emal etmək üçün, məsələn, Amazon EMR kimi MapReduce xidmətlərindən istifadə olunması təşkilat və müəssisələrə əlverişli və sərfəlidir.

*Amazon EMR.* Amazon Elastik *MapReduce* (*Amazon Elastic MapReduce, Amazon EMR*) – böyük həcmdə verilənləri sürətli və az xərc sərf etməklə emal etməyə imkan verən veb servisdır. Amazon EMR böyük verilənlərin emal prosesini öz tərkibinə Hadoop sistemini daxil etməklə sadələşdirir. Burada Hadoop sisteminin funksiyası böyük həcmdə verilənləri *Amazon EC2* elementlərinə paylamaqdır. *Amazon EMR* böyük verilənlərin aid olduğu loq yazıların analizi, veb-indeksləşmə, verilənlər anbarının təşkili, maşın təlimi, maliyyə analizi, elmi tədqiqatların simulyasiyası, bioinformatika kimi sahələri təhlükəsiz və etibarlı tərzdə idarə edə bilər.

*Hadoop.* Açıq kodlu proqram təminatı platformasıdır, böyük verilənlər çoxluğunun paylanmış hesablama mühitində emalını təmin edir. Bu emal vasitəsinin təyinatı verilənləri paralel şəkildə işləyən balanslaşdırılmış maşınlar klasteri arasında paylamaqdır.

*Hadoop Distributed File System (HDFS).* Hadoop platformasının fayl sistemi komponentidir. HDFS-in təyinatı fayl sisteminin metaverilənlərini və proqram tətbiqinin verilənlərini paylanmış şəkildə saxlamaqdır [33].

Böyük verilənlərin emalında bulud infrastrukturundan istifadənin səbəbləri aşağıdakılardır [7]:

- *Böyük verilənlərin analizinə sərf olunan investisiyaların kifayət qədər böyük olması.* Bu buludlar kimi effektiv və az xərc tələb edən infrastrukturun olmasına ehtiyac yaradır.

- *Böyük verilənlərin həm daxili və həm də xarici infraqururlarda yerləşdirilmiş olması.* Adətən təşkilatlar özlərinin daha həssas məlumatlarını daxili infraqururda, az əhəmiyyətli məlumatlarını isə üçüncü tərəf infraqururda yerləşdirirlər. Bulud texnologiyaları həm daxili, həm də xarici infraqururlarda yerləşdirilmiş verilənlərin analizini təmin edə bilər.

• *Böyük verilənlərdən biliklərin çıxarılması üçün analiz xidmətlərinin olması zərurəti.* Analitikadan xidmət kimi istifadə olunması İT büdcəsinin səmərəli idarə olunmasını təmin edən yeganə vasitə kimi qiymətləndirilir.

### **Analitika xidməti**

Buludda analitika xidmətinə, adətən, təşkilatın analitika sahəsində bacarıqları olmadığı halda ehtiyac yaranır. AaaS həm kiçik, həm də böyük təşkilatların istifadəsinə yararlı olan maraqlı konsepsiyadır.

Buludda böyük verilənlər analitikası həllərini həyata keçirmək üçün proqram təminatı xidməti (*Software as a Service, SaaS*), platforma xidməti (*Platform as a Service, PaaS*), infrastruktur xidməti (*Infrastructure as a Service, IaaS*) modellərindən istifadə edilə bilər [34].

1. *Analitika proqram təminatı xidməti (data analytics software as a service).* Analitika proqramlarını son istifadəçiyə İnternet xidməti şəklində təqdim edir.

2. *Analitika platforması xidməti (data analytics platform as a service).* İstifadəçilərin şəxsi analitika proqramlarının yaradılmasını təmin etmək üçün onlara platforma təqdim edir. Böyük həcmli verilənlər platforması xidmətinə misal olaraq, *Google* təşkilatının təqdim etdiyi *BigQuery* xidmətini göstərmək olar. *BigQuery* istifadəçilərin böyük həcmli verilənləri analiz etmək üçün *Google* təşkilatının nəhəng kompüter və saxlanma imkanlarından istifadə etməsinə və real zamanda qərarlar qəbul etməsinə şərait yaradır. *BigML* digər böyük həcmli verilənlər platforması xidmətidir. *BigML* bulud-yönlü maşın təlimi (*machine learning*) xidməti təqdim edir. Böyük həcmli verilənlər platformaları, adətən, çox sayda modullardan ibarət olur, məsələn, məsələnin spesifikasiyasının analizi bloku, verilənlərin saxlanması və idarə edilməsi bloku, aşkarlama, vizuallaşdırma və s. [35].

3. *Analitika infrastrukturu xidməti (data analytics infrastructure as a service).* İstifadəçilərə virtuallaşmış resurslar təqdim edir, bu resurslardan proqram istehsalçıları öz analitika proqramlarını icra etmək üçün kompüter infrastrukturu qismində istifadə edirlər.

İstənilən AaaS aşağıdakı prinsiplər əsasında qurulmalıdır [11]:

- məsələnin yerinə yetirilməsi üçün addımlar ardıcılığı iyerarxik təşkil olunmalıdır. İyerarxiyaya verilənlərin strukturu, analitik modellərin spesifikasiyası, analitik proqramın konfigurasiyası, məsələnin yerinə yetirilmə addımları daxil edilməlidir;
- ardıcıl, paralel, iterativ və selektiv axınları dəstəkləməlidir;
- verilənlər və idarəetmə axınlarını dəstəkləməlidir;
- kimliyi təsdiqləmək üçün lazım olan məhdudiyətlər ətraflı təsvir olunmalıdır.
- servis səviyyəsi müqaviləsi əsasında AaaS üçün xidmət keyfiyyəti təqdim olunmalıdır;
- məsələlər axınını səmərəli icra etmək üçün miqyaslı bulud resurslarından şəffaf şəkildə istifadə olunmalıdır;
- analitik və vizuallaşdırma modellərinin qurulmasında və məsələlər axınının icra edilməsində konkret sahəyə köməklik göstərmək üçün tövsiyə sistemlərini dəstəkləməlidir;
- servislər təqdim etdikdə istifadəçinin roluna və məşğuliyyət aspektlərinə diqqət etməlidir;
- çox yığcam məsələlər axınını geniş qrafik təsvir etməlidir;
- verilənlərin yerdəyişməsinə minimal yol verməlidir;
- verilənlərin məxfiliyini və təhlükəsizliyini təmin etməlidir.

AaaS-ın istifadəçiləri aşağıdakı şəxslər ola bilər:

❖ *elmi analitiklər* – analitik proqramlar və metodologiyalar haqqında biliyi olan şəxslərdir. Bu şəxslərin verilənlərin aid olduğu sahə haqqında biliyi ola və olmaya bilər, məsələn, *data mining* texnologiyası üzrə ekspertlər;

❖ *verilənlərin aid olduğu sahənin ekspertləri (domain experts)* – bu şəxslərin verilənlərin aid olduğu sahə haqqında biliyi var, verilənlərin müxtəlif qiymətlərinin mənasını başa düşürlər,



adətən, analitiklərə təlimat verirlər və alınmış nəticələrdən qərarlar çıxarmağa köməklik edirlər. Belə şəxslər, məsələn, həkimlər, meteoroloqlar ola bilər;

❖ *praktiklər (practitioners)* – verilənlərin dəyərini praktik başa düşür və məsələlərin əvvəlcədən təyin olunmuş axını son istifadəçi qismində icra edir. Tibb bacıları bu tip şəxslərə misal göstərilə bilər;

❖ *administratorlar* – ikinci dərəcəli çıxarış həyata keçirərək biznes dəyərləri əldə edir və bunun əsasında inzibati tədbirlər görürlər. Bu şəxslərə satış və ya işçi heyət administratorları misal göstərilə bilər;

❖ *menecerlər (managers)* – şirkətin məqsədlərinə əsasən strateji planlaşdırma aparır və buna uyğun olaraq xüsusi informasiyanı müxtəlif aspektlərdən sorğu edirlər.

AaaS xidmətinin aşağıdakı imkanları vardır [7]:

❖ etibarlı qaynaqlardan strukturlaşmış və ya strukturlaşmamış verilənlərin əldə olunması və çıxarılması;

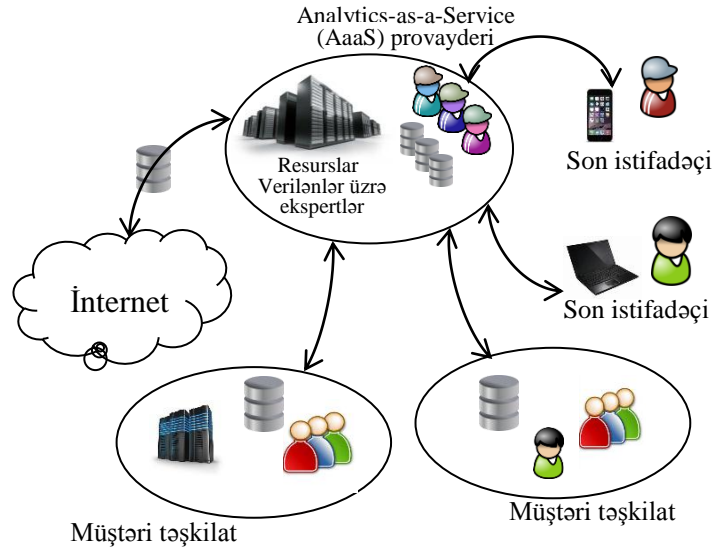
❖ verilənlərin siyasi təlimatlar altında idarə edilməsi və nəzarətin təşkili;

❖ lazımi məkana lazımi vaxtda lazımi informasiyanın çatdırılması üçün verilənlərin inteqrasiyasının, analizinin, çevrilməsinin və vizuallaşdırılmasının həyata keçirilməsi.

İstifadəçilərə AaaS xidməti göstərməyə cəhd edən aparıcı təşkilatlardan birincisi IBM olmuşdur. IBM təşkilatına məxsus AaaS platforması Şəkil 3-də təsvir edilmişdir. Bu platformanın məqsədi ilk növbədə təşkilatların özəl analitikaya sərf etdikləri maliyyə yükünü azaltmaqdır.

Bu platforma istifadəçilərə öz verilənlərini (strukturlaşmış və strukturlaşmamış) analiz olunmaq məqsədi ilə üçüncü tərəflərin infrastrukturuna yükləməsinə imkan verir.

AaaS xidmətləri sahəsində fəaliyyət göstərən digər aparıcı təşkilat SAS (*Statistical Analysis System*) təşkilatıdır. SAS təşkilatı istifadəçilərə proqnozlaşdırma analitikasına əsaslanan AaaS xidməti təqdim edir.



Şəkil 3. Analitika xidməti

### Neft və qaz sənayesində böyük verilənlər analitikasının tətbiqinin müasir vəziyyətinin analizi

Neft-qaz kəşfiyyatı və hasilatı üçün aşağıdakı problemlər müəyyən edilmişdir [36]:

➤ qazma:

- qərarların real zamanda qəbul olunmasının avtomatlaşdırılmasına zərurətin olması;
- qazma işləri başa çatdıqdan sonra real zaman verilənləri arxivindən biliklərin aşkarlanmasına zərurətin olması.

➤ neft hasilatı (*production*):

- neft quyularındakı problemlər ciddi səviyyəyə çatmamışdan onların aşkarlanmasına zərurətin olması;
- çıxış nəticələrinin keyfiyyətini maksimallaşdırmaq üçün sürətli optimallaşdırma zərurətin olması.

➤ verilənlərin idarə edilməsi (*data management*):

- verilənlərin astronomik həddə çatmasının böyük verilənlər üsullarına zərurət yaratması;
- analizin real zamanda aparılmasına verilənlərin ölçüsünün maneə törətməsi;

- real zaman verilənlərinin qorunması.

Böyük verilənlər analitikasını sadalanan problem sahələrinə tətbiq etməklə neft-qaz kəşfiyyatı və hasilatının effektivliyini artırmaq mümkündür.

Böyük verilənlər texnologiyası neft-qaz çıxarılması prosesinin bütün mərhələlərinə (kəşfiyyat, hasilat, emal, pərakəndə satış) tətbiq oluna bilər [37]:

- axtarışların gücləndirilməsi (*enhancing searches*). Müəssisə verilənlərinin real zamanlı hasilat verilənləri ilə birləşməsinə təmin etməklə işçi qrupa yeni ideyalar vermək;
- ərazinin qiymətləndirilməsi və yeni perspektivlərin yaradılması. Coğrafi məlumatlardan, xəbər yayımlarından və başqa sindikativ informasiya (*syndicated information*) mənbələrindən istifadə etməklə rəqabətli kəşfiyyat yaratmaq;
- mühəndis tədqiqatların təkmilləşdirilməsi (*improving engineering studies*). Mürəkkəb geoloji modellərdən istifadə etməklə quyularda müfəssəl texniki tədqiqatlar aparmaqla, ticari perspektivləri erkən və aşağı risklə müəyyən etmək;
- geoloji tədqiqatların optimallaşdırılması (*optimizing subsurface understanding*).
- Neft-qaz sənayesi sahəsində əldə edilən verilənlərdən böyük verilənlər analitikası vasitəsi ilə biliklərin çıxarılması bu sektorun təşkilatlarına aşağıdakı imkanları verə bilər [36]:
- planlaşdırma, kəşfiyyat, hasilat və neft yataqları yaratmaqla rəqabətə davamlı olmaq;
- idarəetmə və proqnozlaşdırmanın hesabına neft hasilatını maksimallaşdırmaq;
- neft-qazın ilk dəfə çıxarılmasına sərf olunan vaxtın azaldılması, əməliyyatlara olan xərclərin azaldılması və neft quyularının həyat dövrü ərzində məhsuldarlığının artırılması;
- doğru informasiyanın işçi qüvvəsinə birbaşa və avtomatik əlçatarlığının vaxtında təmin edilməsi.

Böyük verilənlərə investisiyaların yatırılması böyük dönüş (*big return on investment, ROI*) kimi qiymətləndirilir.

[38]-da neft-qaz çıxarılması və hasilatı verilənlərinin idarə edilməsi üçün böyük verilənlər texnologiyasına əsaslanan Hadoop əsaslı həll təklif edilmişdir. Həll arxitekturu neft yataqlarında toplanan verilənlərin böyük verilənlər analitikası vasitəsi ilə emalı prosesini əks etdirir. Arxitektura verilənlərin əldə olunması, verilənlərin saxlanılma yerləşdirilməsi, verilənlərin aqreqasiyası, verilənlər analitikası, verilənlərin vizuallaşdırılması kimi bloklardan ibarət olaraq 5 mərhələdə təşkil olunmuşdur. Burada ilk öncə müxtəlif quyulardan neft hasilatı verilənlərinin toplanması həyata keçirilir, sonra toplanmış böyük verilənlər Hadoop paylanmış fayl sistemində saxlanılır, sonrakı mərhələdə paylanmış verilənlər müxtəlif bazalara (*Hive/Hbase, NoSQL*) aqreqasiya olunaraq analiz olunmağa başlayır. Verilənlər analitikası mərhələsində qazma obrazlarının tapılması, neft quyularının loq fayllarındakı müxtəlif parametrlərə əsaslanan litologiya kontentini əldə etmək məqsədi ilə analiz həyata keçirilir. Verilənlərin vizuallaşdırılması mərhələsində analitika mərhələsindəki çıxış nəticələri biznes-analizi sisteminə ötürülür, burada qiymətləndirmə cədvəli formalaşdırılır və bu cədvəl əsasında verilənlərin vizuallaşdırılması həyata keçirilərək qərarlar qəbul olunur.

Böyük verilənlər texnologiyası neft sənayesinin aşağıdakı sahələrində əməliyyatların keyfiyyətini yüksəldə bilər [39]:

#### ***kəşfiyyat və istifadə:***

- *izlərin müəyyən edilməsi* – əvvəllər gözdən yayınmış seysmik iz siqnaturalarını müəyyən etmək üçün seysmik kəşfiyyat zamanı toplanan böyük verilənlər çoxluğunda təkmilləşdirilmiş analitikadan (məsələn, obrazların tanınması) istifadə olunması;
- *kəşfiyyat səylərinin gücləndirilməsi* – real zamanda toplanmış neft hasilatı verilənləri ilə müəssisədə toplanmış verilənləri birləşdirməklə əməliyyat qruplarına yeni biliklərin təqdim edilməsi;
- *yeni perspektivlərin yaradılması* – coğrafi verilənlərə, neft-qaz məlumatlarına tətbiq olunmuş analitikadan istifadə etməklə rəqabətli kəşfiyyat yaratmaq.

#### ***qazma və tamamlama:***

- *qazmanın uğurlu olmasının proqnozlaşdırılması* – məhdud verilənlər əsasında monitoring və

xəbərdarlıq etməklə yanaşı, anomalionaları müəyyən etmək və ya qazmanın uğurlu olması ehtimalını proqnozlaşdırmaq üçün qazma haqqında böyük həcmli verilənlərə real zamanda müraciət olunması.

**hasilat və əməliyyatlar:**

- *məhsuldarlığın proqnozlaşdırılması* – on minlərlə neft quyusunda məhsuldarlığın proqnozlaşdırılması. Köhnə quyularda proqnozların nəticəsi əvvəlcədən təyin olunmuş məhsuldarlıq həddini ödəmədikdə, onların təcili bərpa olunması üçün xəbərdarlıq olunur;

- *genişləndirilmiş neft hasilatı* – Neft yataqlarından çıxarıla bilən xam neftin miqdarını artırmaq üçün müxtəlif üsulların istifadəsini nəzərdə tutur. Burada müxtəlif tipli böyük verilənlərin analizində istifadə olunan analitikanın xam neftin miqdarını artırmaq məqsədi ilə eyni zamanda seysmoloji, qazma və hasilat verilənlərinin analizində də istifadə olunmasının mümkünlüyü nəzərdə tutulur;

- *önləyici texniki xidmət* – qurğuda texniki xidmətin aparılmasının zəruri olduğu müddəti proqnozlaşdırın üsuldur. Təzyiq, həcm və temperatur birlikdə analiz edilir, qurğunun sıradan çıxdığı əvvəlki tarix ilə müqayisə edilir, təkmilləşdirilmiş analitika tətbiq olunaraq potensial uğursuzluqlar proqnozlaşdırılır.

Böyük verilənlərin neft və qaz sənayesində istifadəsinin aşağıdakı üstünlükləri vardır [40]:

- *əməliyyatların təkmilləşməsi* – strukturlaşmış, strukturlaşmamış və real zaman verilənlərini özündə birləşdirən vahid idarəetmə platformasından kollektiv bilik əldə etmək. Fəaliyyətdə qeyri-məhsuldar vaxtın (*non productive time*, NPT), real zamanlı risk idarəetmə mexanizminin olması hesabına normativ aktların qiymətinin azaldılması.

- *neft hasilatında yüksək sürətin əldə olunması* – verilənlərin “darboğazlılığı”nı (*data bottlenecks*) minimallaşdırmaqla neft hasilatı vaxtını sürətləndirmək;

- *aktivlərin yaxşılaşdırılması* – qurğunun istismar müddətini artırmaq və qurğunun istismar tələblərini proqnozlaşdırmaq;

- *neft və qaz üçün vahid ontologiyanın yaradılması.*

**Neft-qaz sənayesi üçün Big Data Analitika platformaları**

Neft-qaz sənayesində böyük verilənlərin analizini həyata keçirən proqramlardan biri *IBM* təşkilatına məxsus *InfoSphere BigInsights* proqramıdır [41]. *InfoSphere BigInsights* Hadoop üzərində qurulmuş platformadır, neft-qaz sahəsinə aid iri həcmli və müxtəlif tipli verilənlərin toplanmasını, emalını, analizini və idarə edilməsini həyata keçirir. *InfoSphere BigInsights* proqramının tərkibinə *InfoSphere Streams*, *InfoSphere Data Explorer*, *IBM PureData System for Analytics*, *IBM PureData System for Operational Analytics* kimi proqramlar daxildir. Onlar ayrı-ayrılıqda avtonom proqram kimi fəaliyyət göstərə bilmir, yalnız *InfoSphere BigInsights* proqramının idarəçiliyi altında icra olunurlar. Bu proqramların hər birinin özünəməxsus funksiyaları vardır. Təşkilatın analitika sahəsində xüsusi bulud həlli *Blue Insight* adlanır. *Blue Insight* 100 müxtəlif bazadan verilənləri toplaya bilir və petabaytdan çox həcmdə verilənin üzərində analitik əməliyyatlar aparır. *IBM-in “Blue Insight”* adlı bulud infrastrukturunu xarici müştərilərə “*IBM Smart Analytics Cloud*” adı altında təqdim olunur.

Böyük verilənlər analitikası həllərinin yaradılması ilə məşğul olan təşkilatlardan biri də *Hitachi* şirkətidir. Şirkət müştərilərə beynəlxalq səviyyədə böyük verilənlərin imkanlarına söykənən innovativ həllər təqdim etmək məqsədi ilə 2013-cü ildə İnnovativ Analitika üçün Beynəlxalq Mərkəz (*Hitachi Global Center for Innovative Analytics, HGC-IA*) təchiz etmişdir. Mərkəzin strukturunu böyük verilənlər analitikası həlləri yaratmağa xidmət edən proqram təminatı toplusu, verilənlər üzrə alimlər, memarlar, məsləhətçilər təşkil edir. Mərkəzin böyük verilənlər sahəsində Beynəlxalq İnnovasiya Laboratoriyası (*Global Big Data Innovation Lab, GBDIL*) adlı elmi-tədqiqat bölməsi dünyada böyük verilənlərlə bağlı tədqiqat fəaliyyətlərinin koordinasiyası ilə məşğul olur.

GBDIL laboratoriyası böyük verilənlər həllərinin yaradılmasında digər şirkətlərə təlimat qismində çıxış edə bilən ümumi arxitektura təklif etmişdir. Arxitekturanı böyük verilənlərin saxlanması, idarə edilməsi, analizi və vizuallaşdırılmasını həyata keçirən komponentlər təşkil edir.

Təklif edilən arxitektura əsasında yaradılmış həllər həm vahid qovşaqla, həm də böyük klasterdə icra oluna bilər. Arxitektur yanaşmadan başqa, HGC-IA və GBDIL laboratoriyası kommunikasiya, neft-qaz, səhiyyə və başqa sənaye sahələri üçün müştəri həllərinin yaradılması ilə də məşğul olur. [42]-də *Hitachi* təşkilatının neft-qaz sənayesi üçün böyük verilənlər analitikasının sxematik təsviri verilmişdir. *Hitachi*-nin neft-qaz texnologiyaları sahəsinə olan marağını gücləndirmək məqsədi ilə “*US Big Data Lab*” laboratoriyası Şimali Dakota Universitetinin nəzdində fəaliyyət göstərən Enerji və Ətraf mühit tədqiqatlar mərkəzi (*Energy&Environmental Research Center, EERC*) ilə əməkdaşlıq yaratmışdır. EERC enerji və ətraf mühit texnologiyalarının yaradılması ilə məşğul olan müəssisədir. *Hitachi* və *EERC* tədqiqat istiqamətlərini ilk öncə operatorların *Bakken* mədəni (Bakken Field) yaratdıqları zaman qarşılaşdıqları mürəkkəb problemlərin həllinə yönəlmişlər. Şirkətin “*Bakken Production Optimization*” adlı proqramı 2 əsas məqsədə xidmət edir:

- *Bakken* mədəniyyətinin geologiyasını yaxşı analiz etməklə, neft quyusunun məhsuldarlığını optimallaşdırmaq;
- qazma işlərinin səmərəliliyini artırmaq və su, qaz və başqa üsullardan istifadə etməklə, qazmanın yerüstü təsirlərini azaltmaq.

Proqram həlli yuxarıdakı məqsədləri böyük verilənlərdən bilik əldə etməklə həyata keçirir.

Neft-qaz sənayesində eksponensial şəkildə artan verilənlərin idarə edilməsinə xidmət edən sistemlərdən biri də *Hitachi* (*Hitachi data systems*) və Intel şirkətlərinin birgə səyi nəticəsində yaranmış *Lustre* sistemidir. *Lustre* adı “*Linux*” və “*cluster*” adlarının birləşməsindən əmələ gəlmişdir. *Lustre* kompüter klasterləri üçün nəzərdə tutulmuş yüksək məhsuldarlıqlı fayl sistemidir. *Lustre* on minlərlə kompüter qovşağında, petabayt yaddaşa malik saxlanmada icra oluna bilər və saniyədə yüzlərlə qıqabayt veriləni emal edir. Sistem müxtəlif mənbələrdən götürülmüş bütün tip verilənləri emal edə bilər [43].

*iGATE* korporasiyası neft quyularının loq analitikası (*Oil Well Log Analytics*) adlı sistem yaratmağı planlaşdırır [40]. Bu sistem neft kəşfiyyatı və hasilatı verilənlərini vahid şəkildə birləşdirilmiş platformada inteqrasiya etməyə xidmət edir. Sistemin *iGATE* korporasiyasının bulud infrastrukturundan istifadə etməsi planlaşdırılır. Burada müxtəlif mənbələrdən (strukturlaşmış, strukturlaşmamış, real vaxt) götürülmüş verilənlərin indekslənməsi, saxlanması, təmizlənməsi, klasterləşdirilməsi, miqrasiyası, standartlaşdırılması və analizi Hadoop ekosistemi və R analitikası (revolution analytics) ilə təchiz edilmiş platformada həyata keçirilir.

*Microsoft* şirkəti neft-qaz sənayesində böyük verilənlərin emalı üçün yeni texnologiyaların istifadə qaydalarını əhatə edən təlimat hazırlamışdır [37]. Neft şirkətlərinə böyük verilənlər həlləri yaratmasında təlimat rolunda çıxış edə bilən MURA (*Microsoft Upstream Reference Architecture*) adlı arxitektura təklif edir.

Neft-qaz sənayesində texniki maneələrin, verilənlərin təhlükəsizlik problemlərinin ekstremal həddə olması bu sektorda toplanan böyük verilənlərin ümumi buludlara köçürülməsi prosesini çətinləşdirir. Bu səbəbdən [44]-də neft sektorunda xüsusi və hibrid bulud həllərindən istifadə edilməsinin zəruri olduğu əsaslandırılır.

### **Dünyanın iri neft-qaz şirkətlərinin big data sahəsində təcrübəsi**

Neft şirkətləri nə qədər çox seysmik verilənlərə malik olsalar, neft və təbii qaz zonalarını bir o qədər yaxşı tapmaq imkanı əldə etmiş olurlar.

Neft və qaz sənayesində böyük verilənlər texnologiyasının tətbiqi hələ ki, eksperimental mərhələdədir [45]. Böyük verilənlər texnologiyasından az sayda neft şirkəti istifadə edir [46].

2010-cu ildə “*Shell*” neft şirkəti dünya üzrə paylanmış 3 və regional 400 verilənlər mərkəzinin müqabilində böyük enerji xərcləri ilə rastlaşmışdır. Bu vəziyyətdən çıxış yolunu şirkət buluddan istifadədə görmüşdür və 2010-cu ildə *Amazon* təşkilatının təqdim etdiyi Virtual Xüsusi Bulud xidmətindən istifadə edərək özünün bulud infrastrukturunu yaratmışdır. Quyularda neftin aşkarlanmasını həyata keçirən seysmik sensorların topladığı coğrafi verilənlərin kütləvi hal alması qərar qəbul etmə prosesini olduqca çətinləşdirmişdir. Bu problemi aradan qaldırmaq

üçün şirkət minlərlə neft quyularından toplanmış böyük verilənləri emal etmək üçün Amazon təşkilatının bulud infrastrukturunun təqdim etdiyi Hadoop sistemindən istifadə edir. Şevron neft şirkəti də seysmik məlumatların emalında Hadoop texnologiyasından istifadə edir.

*Cloudera* şirkəti tərəfindən yaradılmış *Seismic Hadoop* layihəsi mövcuddur. Bu layihə seysmik verilənlərin Hadoop klasterində saxlanması və emalı məqsədi ilə yaradılmışdır.

Neft-qaz sənayesi üçün böyük verilənlər həllərinin yaradılması sahəsində ixtisaslaşmış təşkilatlardan biri də *PointCross* şirkətidir. *PointCross* texnologiya və xidmətlər üzrə qlobal əhəmiyyətli şirkətdir. O, qazma, seysmik və istehsal verilənləri üçün korporativ proqram həllərinin yaradılması və tətbiqi ilə məşğul olur.

Proqnozlara görə, yaxın bir neçə ildən sonra dünyanın əksər verilənləri Hadoop sistemi vasitəsi ilə emal olunacaq.

### Neft-qaz sənayesində böyük verilənlərin problemləri

Neft-qaz sənayesində böyük verilənlərin aşağıdakı problemləri göstərilir [40] :

- quyuların istifadəsi dövründə müxtəlif mənbələrdəki verilənlərin (strukturlaşdırılmış, strukturlaşmamış, real-zaman) həcmnin eksponensial artımı;
- neft-qaz təşkilatları kəşfiyyat və hasilat verilənlərinin idarə edilməsinə, müxtəlif mənbələrdəki, adətən, bir-biri ilə uzlaşmayan verilənlər axınının emalına çox xərc sərf edir;
- istifadəçilərin tələblərinə çevik və effektiv cavab vermək üçün verilənlərin istifadəsində çətinlik;
- geoloqlar və geofiziklər qrupunun verilənləri interpretasiya etmək və qərarlar qəbul etmək üçün müxtəlif proqram təminatı məhsullarının qarışığından istifadə etməsi;
- hər bir verilənlər klasterinə daxil edilmiş problem-yönümlü məlumatların böyük həcmdə olması.

### Təklif və tövsiyələr

Neft-qaz sənayesi üçün Big Data analitikanın *Cloud Computing* platformasında AaaS xidməti kimi reallaşdırılması zamanı aşağıdakıların təmin olunması zəruri olardı:

1. buludların təqdim etdiyi platformalardan, proqram tətbiqlərindən və infrastruktur resurslarından minimal riskə yol verməklə istifadə olunması;
2. neft-qaz hasilatı verilənlərini buludlarda yerləşdirərkən bulud provayderlərinin təhlükəsizlik risklərinə qarşı nə dərəcədə dayanıqlı olduğunu aydınlaşdırılması;
3. ayrı-ayrı məkanlarda yerləşən verilənlərin səmərəli inteqrasiyasının təşkili;
4. virtual buludlarda yerləşdirilmiş strukturlaşmış və strukturlaşmamış neft hasilatı verilənlərinin inteqrasiyasını və onların real vaxt rejimində analizini təşkil edə bilən vasitələrin yaradılması;
5. müştərinin verilənlər bazası və analitik həllərinin AaaS mühiti ilə inteqrasiyasının təşkili;
6. neft-qaz hasilatında axın təşkil edən verilənlərin davamlı analizini həyata keçirə bilən analitik vasitələrin yaradılması;
7. mövcud data mining alqoritmlərini bulud-yönümlü vasitələrə çevirmək və onlardan neft-qaz hasilatında toplanan böyük həcmdə verilənlərdən biliklərin aşkarlanması üçün istifadə etmək;
8. buludların dəstəklədiyi bugünkü analitik həllərdə neft-qaz hasilatını monitorinq etmək, trafikə vəziyyətini izləmək və baş verən hadisələri aşkarlamaq üçün real vaxt rejimində baş verən hadisələrdən davamlı analitik qərarlar hasil edən funksiyanın yaradılması.
9. verilənlərlə informasiya axını arasında koordinasiya təşkil etmək;
10. əsas emal prosesi və aktivləri əldə etmədən, onlar üzərində idarəetmənin genişləndirilməsi;
11. qərar qəbuluna sərf olunan vaxtın minimal olmasına çalışmaq;
12. verilənlərin, informasiyanın və analitik modellərin təhlükəsizliyini təmin etmək;

13. qərar qəbulu sistemlərinin effektiv və səmərəli idarəçilik modellərini yaratmaq;
14. biznes əməliyyatlarında baş verən dəyişikliklər barədə xəbərdarlıq etmək və anındaca cavab vermək.

## Nəticə

Hal-hazırda əksər neft şirkətləri sahib olduqları verilənlərdən və verilənlərin keyfiyyətli toplanmasını, saxlanmasını, idarə və analiz edilməsini həyata keçirən idarəetmə xidmətlərindən əhəmiyyətli dərəcədə asılıdırlar. Neft-qaz sənayesində böyük verilənlərin meydana gətirdiyi problemləri aradan qaldırmaq üçün hazırda bir çox şirkətlərdə böyük verilənlər strategiyası işlənməyə başlanmışdır.

Neft-qaz hasilatı sahəsində verilənlərin analizi sistemlərinin qurulma metodologiyası proqram tətbiqlərinin yaradılması ideyasına əsaslanır və servise yönəlik xüsusiyyətə malik deyil. Bu sistemlərdə proqram tətbiqlərinin quraşdırılması, konfigurasiya və idarə edilməsi istifadəçinin öz üzərinə düşür. Bundan əlavə, bu sistemlər istifadəçinin dinamik dəyişə bilən tələblərini ödəyə bilmir. Bu səbəbdən bulud texnologiyalarının təqdim etdiyi AaaS xidmətindən istifadə olunması neft-qaz sənayesində toplanan böyük verilənlərin emal prosesinin dinamikliyini əhəmiyyətli dərəcədə artırmış olar. Qeyd edək ki, böyük verilənlərin emalında bulud texnologiyalarından istifadə olunması təşkilatın verilənlərinin ümumi həcmi optimallaşdırmağa da imkan verə bilər.

Təqdim olunan məqalədə neft-qaz sənayesində big data analitikanın cloud computing platformasında istifadəsinin mövcud vəziyyəti araşdırılmış, bu sahədə mövcud olan bir sıra analitik proqramların və həllərin analizi aparılmışdır. Bundan əlavə, neft-qaz sənayesində toplanan böyük verilənlərin AaaS xidməti vasitəsi ilə effektiv analizinin həyata keçirilməsinin təmin edilməsi istiqamətində bir sıra təklif və tövsiyələr verilmişdir.

## Ədəbiyyat

1. Cuzzocrea A., Song I.Y., Davis K.C. Analytics over large-scale multidimensional data: the big data revolution! / Proc. of the ACM 14th International Workshop on Data Warehousing and OLAP, 2011, pp.101–104.
2. Chen C.L. Zhang C.Y. Data-intensive applications, challenges, techniques and technologies: A survey on big data // Information Science, 2014, vol.275, pp.314–327.
3. Obama Administration Unveils “Big data” initiative: Announces \$200 million in new R&D investments, 2012, 4 p.
4. [www.technologyreview.com/news/427876/big-oil-goes-mining-for-big-data/](http://www.technologyreview.com/news/427876/big-oil-goes-mining-for-big-data/)
5. Manyika J., Chui M., Brown B., Bughin J., Dobbs R., Roxburgh C., Byers A.H. Big Data: The next frontier for innovation, competition, and productivity, 2011, 156 p.
6. Brulé M., Tapping the power of Big Data for the oil and gas industry, IBM Software White Paper for Petroleum Industry, 2013, 8 p.
7. Big Data in the Cloud: Converging technologies, Intel IT Center, 2015, 12 p.
8. Demirkan H., Delen D. Leveraging the capabilities of service-oriented decision support systems: putting analytics and big data in cloud // Decision Support Systems, 2013, vol.55, pp.412–421.
9. Pandey S., Nepal S., Cloud computing and scientific applications – big data, scalable analytics and beyond // Future Generation Computer Systems, 2013, vol.29, pp.1774–1776.
10. Hashem I.A., Yaqoob I., Anuar N., Mokhtar S., Gani A., Khan S.U. The rise of “big data” on cloud computing: Review and open research issues // Information Systems, 2015, vol.47, pp.98–115.
11. Zulkernine F., Bauer M., Abounaga A. Towards cloud-based analytics-as-a-service (CLAaaS) for big data analytics in the cloud / Proc. of the IEEE International Congress on Big Data, 2013, pp.62–69.
12. What is advanced analytics? [www-01.ibm.com/software/data/infosphere/what-is-advanced-](http://www-01.ibm.com/software/data/infosphere/what-is-advanced-)

analytics/

13. Russom P. Big data analytics, TDWI research, 2011, 35 p.
14. Febowitz J. Big data in upstream oil and gas, IDC energy insights, 2013, 45 p.
15. Liebowitz J. Business analytics: an introduction, 2013, 288 p.
16. Buyya R., Yeo C.S., Venugopal S., Broberg J., Brandic I. Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility // *Future Generation Computer Systems*, 2009, vol.25, no.6, pp.599–616.
17. Yan W., Brahmakshatriya U., Xue Y., Gilder M., Wise B. p-PIC: Parallel power iteration clustering for big data // *Journal of Parallel and Distributed Computing*, 2012, vol.73, no.3, pp.352–359.
18. Kim Y., Shim K., Kim M., Lee J.S. DBCURE-MR: an efficient density-based clustering algorithm for large data using MapReduce // *Information Systems*, 2014, vol.42, pp.15–35.
19. Song J., Guo C., Wang Z., Zhang Y., Yu G., Pierson J. HaoLap: Aa Hadoop based OLAP system for big data // *Journal of Systems and Software*, 2015, vol.102, pp.167–181.
20. Kajdanowicz T., Kazienko P., Indyk W. Parallel processing of large graphs // *Future Generation Computer Systems*, 2014, vol.32, pp.324–337.
21. Malewicz G., Austern M., Bik A., Dehnert J., Horn I., Leiser N., Czajkowski G. Pregel: A system for large-scale graph processing / *Proc. of the International Conference on Management of Data*, 2010, pp.135–146.
22. Ordonez C., Mohanam N., Garcia A.C. PCA for large data sets with parallel data summarization // *Distributed and Parallel Databases*, 2014, vol.32, no.3, pp.377–403.
23. Fiore S., D'Anca A., Palazzo C., Foster I., Williams D.N., Aloisio G. Ophidia: Toward big data analytics for eScience / *Proc. of the International Conference on Computational Science*, 2013, pp.5–7.
24. Wu X., Zhu X., Wu G., Ding W., Data mining with big data // *IEEE Transactions on Knowledge and Data Engineering*, 2014, vol.26, no.1, pp.97–107.
25. Steed C.A., Ricciuto D.M., Shipman G., Smith B., Thornton P.E., Wang D., Williams D.N. Big data visual analytics for earth system simulation analysis // *Computers & Geosciences*, 2013, vol.61, pp.71–82.
26. Truyens M., Eecke P.V. Legal aspects of text mining // *Computer Law & Security Review*, 2014, vol.30, no.2, pp.153–170.
27. Chen C.L., Zhang C.Y. Data-intensive applications, challenges, techniques and technologies: a survey on big data // *Information Science*, 2014, vol.275, pp.314–327.
28. Kambatla K, Kollias G, Kumar V., Grama A. Trends in big data analytics // *Journal of Parallel and Distributed Computing*, 2014, vol.74, no.7, pp.2561–2573.
29. Hsu C., Li G., Niu W., Batten L., Dorrnsoro B., Danoy G., Bouvry P., Katz D.S., Zhang Z. Intelligent big data processing // *Future Generation Computer Systems*, 2014, vol.36, 452 pp.
30. Əliquliyev R.M., Hacırahimova M.Ş. "Big data" fenomeni: Problemlər və imkanlar // *İnformasiya Texnologiyaları Problemləri*, 2014, №2, s.3–16.
31. Əliquliyev R.M., Hacırahimova M.Ş. "Big data" texnologiyaları / *Elektron dövlət quruculuğu problemləri 1-ci Respublika elmi-praktiki konfransının materialları*, 2014, s.124–127.
32. Dean J, Ghemawat S. MapReduce: a flexible data processing tool // *Communications of the ACM*, 2010, vol.53, no.1, pp.72–77.
33. Shvachko K., Hairong K., Radia S., Chansler R. The Hadoop distributed file system / *Proc. of the IEEE 26th Symposium on Mass Storage Systems and Technologies*, 2010, pp.1–10.
34. Talia D. Clouds for scalable big data analytics // *Computer*, 2013, vol 46, no.5, pp.98–101.
35. Zheng Z., Zhu J., Lyu M.R. Service-generated big data and big data-as-a-service: An overview / *IEEE 2nd International Congress on Big Data*, 2013, pp.403–410.
36. Sangvai P. Impact of big data in oil and gas industry / *Proc. 10th Biennial international Conference & Exposition*, 2013, pp.439–440.



37. Hems A., Soofi A., Perez E. Drilling for new business value. How innovative oil and gas companies are using big data to outmaneuver the competition, A Microsoft White Paper, 2013, 12 p.
38. Taneja P., Wate P. Big Data enabled digital oil field / Computer Society of India Communications, 2013, pp.18–20.
39. Baaziz A., Quoniam L. Big data in upstream oil and gas, IDC energy insights, How to use Big Data technologies to optimize operations in Upstream Petroleum Industry // International Journal of Innovation, 2013, vol.1, no.1, pp.1–9.
40. Big Data for the oil and gas industry, Issue 5/4, TechConnect, 6 p.
41. Tapping the power of big data for the oil and gas industry, IBM Software, 2013, 8 p.
42. Dayal U. Akatsu M, Gupta C. Expanding global big data solutions with innovative analytics // Hitachi Review, 2014, vol.63, no.6, pp.333–339.
43. Big data in oil and gas: how to tap its full potential, Hitachi, WebTech Q&A Session, 2013, 45 p.
44. Perrons R.K., Hems A. Cloud computing in the upstream oil & gas industry: a proposed way forward // Energy Policy, 2013, pp.732–737.
45. Febowitz J., The big deal about big data in upstream oil and gas. Paper & Presentation, IDC Energy Insights, 2012.
46. Nicholson R., Big data in the Oil & Gas Industry, IDC Energy Insights, 2012.

#### UOT 004.9

**Алыгулиев Рамиз М.<sup>1</sup>, Имамвердиев Ядигар Н.<sup>2</sup>, Абдуллаева Фаргана Дж.<sup>3</sup>**

Институт Информационных Технологий НАНА, Баку, Азербайджан

<sup>1</sup>[a.ramiz@science.az](mailto:a.ramiz@science.az), <sup>2</sup>[yadigar@lan.ab.az](mailto:yadigar@lan.ab.az), <sup>3</sup>[farqana@iit.ab.az](mailto:farqana@iit.ab.az)

#### **Исследование возможностей big data аналитики как Analytics-as-a-Service в платформе cloud computing для нефтегазовой промышленности**

Увеличение объема данных, собранных в нефтегазовой промышленности, привело к серьезным проблемам в этом секторе. В статье исследуются проблемы, создаваемые в нефтегазовой промышленности большими данными, и современное состояние применения big data аналитики в этой сфере. Рассматриваются big data платформы для нефтегазовой промышленности, производимые крупными организациями, и опыт крупнейших мировых нефтегазовых компаний в области big data аналитики. Исследуются методы анализа больших данных и цели использования cloud computing для big data аналитики. Предлагается ряд предложений и рекомендаций для реализации в нефтегазовой промышленности big data аналитики как Analytics-as-a-Service.

**Ключевые слова:** *big data аналитика, OLAP, cloud computing, Analytics-as-a-Service, разведка и добыча нефти и газа, Hadoop, MapReduce, наука о данных.*

**Ramiz M. Alguliyev<sup>1</sup>, Yadigar N. Imamverdiyev<sup>2</sup>, Farqana J. Abdullayeva<sup>3</sup>**

Institute of Information Technology of ANAS, Baku, Azerbaijan

<sup>1</sup>[a.ramiz@science.az](mailto:a.ramiz@science.az), <sup>2</sup>[yadigar@lan.ab.az](mailto:yadigar@lan.ab.az), <sup>3</sup>[farqana@iit.ab.az](mailto:farqana@iit.ab.az)

#### **Investigation of the opportunities of big data analytics as Analytics-as-a-Service in cloud computing for oil and gas industry**

The increase in the volume of data generated in oil and gas industry has led to serious problems in this sector. In this paper, the challenges posed by big data in oil and gas industry, current state of implementation of big data analytics in this sphere are analyzed. The platforms of Big data analytics produced by large organizations for oil and gas industry and the experience of the largest oil and gas companies of the world in big data analytics are studied. Big data analysis techniques and the goals of using cloud computing for big data analytics are analyzed. Suggestions and recommendations for the realization of big data analytics as Analytics-as-a-Service in the oil and gas industry are given.

**Keywords:** *Big data analytics, OLAP, cloud computing, Analytics-as-a-Service, oil and gas exploration and production, Hadoop, MapReduce, data science.*