ՀՀ ԳԱԱ ԻՆՖՈՐՄԱՏԻԿԱՅԻ ԵՎ ԱՎՏՈՄԱՏԱՑՄԱՆ ՊՐՈԲԼԵՄՆԵՐԻ ԻՆՍՏԻՏՈՒՏ

Հայկ Անդրանիկի Գրիգորյան

ԲԱՐՁՐ ԱՐՏԱԴՐՈՂԱԿԱՆՈՒԹՅԱՄԲ ՀԱՇՎԱՐԿՆԵՐ ԵՎ ՄԵԾ ՏՎՅԱԼՆԵՐ ՊԱՐՈՒՆԱԿՈՂ ԽՆԴԻՐՆԵՐԻ ՎԵՐԼՈՒԾՈՒԹՅԱՆ ԵՎ ՎԻԶՈՒԱԼԻԶԱՑԻԱՅԻ ՄԻՋԱՎԱՅՐԻ ՄՇԱԿՈՒՄ

Ե.13.04 - «Հաշվողական մեքենաների, համալիրների, համակարգերի և ցանցերի մաթեմատիկական և ծրագրային ապահովում» մասնագիտությամբ տեխնիկական գիտությունների թեկնածուի գիտական աստիճանի հայցման

ՍԵՂՄԱԳԻՐ

ԵՐԵՎԱՆ – 2018

ИНСТИТУТ ПРОБЛЕМ ИНФОРМАТИКИ И АВТОМАТИЗАЦИИ НАН РА

Айк Андраникович Григорян

РАЗРАБОТКА СРЕДЫ АНАЛИЗА И ВИЗУАЛИЗАЦИИ ПРОБЛЕМ С ВЫСОКОПРОИЗВОДИТЕЛЬНЫМИ ВЫЧИСЛЕНИЯМИ И БОЛЬШИМИ ДАННЫМИ

А В Т О Р Е Ф Е Р А Т

диссертации на соискание ученой степени
кандидата технических наук по специальности

04.13.04 – “Математическое и программное обеспечение вычислительных машин, комплексов, систем и сетей”

ЕРЕВАН – 2018

Ատենախոսության թեման հաստատվել է ՀՀ ԳԱԱ Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտում

Գիտական ղեկավար՝ տեխ. գիտ. թեկնածու Հ.Վ.Ասցատրյան
Պաշտոնական ընդդիմախոսներ՝

Առաջատար կազմակերպություն՝

Ատենախոսության պաշտպանությունը կայանալու է 2018թ. հունիսի **-ին, ժամը **:**-ին, ՀՀ ԳԱԱ Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտի 037 «Ինֆորմատիկա» մասնագիտական խորհրդի նիստում, հետևյալ հասցեով՝ 0014, Երևան, Պ. Սևակ 1:
Ատենախոսությանը կարելի է ծանոթանալ ՀՀ ԳԱԱ ԻԱՊԻ-ի գրադարանում:

Սեղմագիրն առաքված է 2018թ. մայիսի **-ին:

037 մասնագիտական խորհրդի գիտական
քարտուղար, ֆիզ. մաթ. գիտ. դոկտոր Հ. Գ.Սարուխանյան

Тема диссертации утверждена в Институте проблем информатики и автоматизации НАН РА

Научный руководитель: кандидат тех. наук Г.В. Асцатрян
Официальные оппоненты:

Ведущая организация:

Защита диссертации состоится **-ого июня 2018г., в **:** часов, на заседании специализированного совета 037 “Информатика” Института проблем информатики и автоматизации НАН РА, по адресу: 0014, Ереван, ул. П. Севака 1.

С диссертацией можно ознакомиться в библиотеке ИПИА НАН РА.

Автореферат разослан **-ого мая 2018г.

Ученый секретарь специализированного
совета 037, доктор физ. мат. наук А. Г.Саруханян

ԱՇԽԱՏԱՆՔԻ ԸՆԴՀԱՆՈՒՐ ԲՆՈՒԹԱԳԻՐԸ

Թեմայի արդիականությունը: Համակարգչային տեխնոլոգիաների զարգացման հետ զուգընթաց ավելանում են նաև գեներացվող տվյալները և նրանց տարաբնույթ բազմությունների (ստրուկտուրավորված և ոչ ստրուկտուրավորված) ծավալի և բարդ կառուցվածքի պատճառով անհրաժեշտություն է առաջանում կիրառել նոր մոտեցումներ՝ նրանց մշակման, պահպանման և վերլուծության համար:

Վերը նշված բազմությունները հանդիսանում են մեծ տվյալներ (Big Data), որը Doug Laney-ի կողմից[1] սահմանվել է երեք հատկանիշների միջոցով (ծավալ, արագություն, բազմազանություն). «Մեծ տվյալները մեծ ծավալով, բարձր արագությամբ և/կամ մեծ բազմազանությամբ ինֆորմացիայի բազմություն է, որում որոշումների կայացման, իմաստների հայտնաբերման և գործընթացների օպտիմալացման համար պահանջվում են մշակման նոր ձևեր»: Ժամանակի ընթացքում ավելացել են մեծ տվյալները բնութագրող ևս երկու հատկանիշներ՝ ճշմարտացիություն և արժեք: Մեծ տվյալների մշակման, պահպանման, վերլուծության, վիզուալիզացիայի և տարածման գործառույթների իրականացման համար օգտագործվում է բարձր արտադրողականությամբ տվյալների վերլուծությունը (HPDA – High Performance Data Analytics), որը միավորում է բարձր արտադրողականությամբ հաշվարկները (HPC – High Performance Computing) մեծ տվյալների հետ: HPDA խնդիրների լուծման համար գոյություն ունեցող գործիքները ապահովում են վերը նշված գործառույթների խմբից որոշակի ենթախմբերի իրականացում՝ կախված նպատակային խնդրից կամ յուրաքանչյուր գործառույթի համար գոյություն ունեցող ծրագրային լուծումները միմյանց հետ փոխկապակցելու բարդություններից: Օգտատերը գործառույթների ամբողջական ցիկլի իրականացման համար ստիպված է լինում խնդրի լուծման և արդյունքների ուսումնասիրության համար օգտագործել մեկից ավելի ծրագրային միջավայր:

Այդ իսկ պատճառով խնդիր է առաջանում մշակել այնպիսի մեթոդներ և ծրագրային միջոցներ, որոնք հնարավորություն կտան միավորել բարձր արտադրողականությամբ հաշվարկների, տվյալների պահպանման, վերլուծության, վիզուալիզացիայի և տարածման գործառույթները մեկ ընդհանուր համալիր համակարգում՝ օգտագործելով ժամանակակից

[1] Laney, D. (2001). 3D data management: Controlling data volume, velocity, and variety.

ծրագրային լուծումները և ընդունված միջազգային հիմնօրինակները: Արդյունքում օգտատերը կկարողանա մեկ ծրագրային միջավայրի օգնությամբ իրականացնել խնդրի լուծման համար անհրաժեշտ ամբողջական գործընթացը, որը կնպաստի արդյունավետության բարձրացմանն ու ժամանակի խնայողությանը:

Աշխատանքի նպատակն ու խնդիրները: Աշխատանքի հիմնական նպատակն է բարձր արտադրողականությամբ հաշվարկներ և մեծ տվյալներ պարունակող խնդիրների լուծման համար կառուցել բաց կոդով, միջազգային հիմնօրինակներին համապատասխան, տվյալների մշակման, պահպանման, վերլուծություններ և վիզուալիզացիաներ տրամադրող, մետատվյալների տարածման հնարավորությամբ համալիր համակարգ: Այդ նպատակով դրվել են հետևյալ խնդիրները.

1. Մշակել տվյալների ստացման, ձևափոխման և վերլուծության ծրագրային մեթոդներ և գրադարաններ՝ տարբեր աղբյուրներից տվյալների հավաքագրման, ոչ ստրուկտուրավորված տվյալներից անհրաժեշտ գիտելիքների հայտնաբերման համար,
2. Մշակել այլընտրանքային պահոցում տվյալների պահպանման, նրանց հիման վրա մետատվյալների ստեղծման և տարածման ավտոմատ համակարգ՝ հաշվի առնելով միջազգային հիմնօրինակները,
3. Մշակել տվյալների փոխանցման և փոխակերպման ծրագրային մեթոդներ տրամաբանական տարբեր բաղադրիչների՝ հաշվողական ռեսուրսների, տվյալների պահպանման, վերլուծությունների, վիզուալիզացիաների և տարածման, միջև փոխհարաբերությունների ստեղծման համար:

Հետազոտման օբյեկտը: Աշխատանքում հիմնական հետազոտության օբյեկտներն են.

1. Բարձր արտադրողականությամբ համակարգերը և նրանց կիրառությունները,
2. Ստրուկտուրավորված և ոչ ստրուկտուրավորված տվյալների մշակման և պահպանման եղանակները,
3. Տվյալների վերլուծության միջոցները,
4. Մեծ ծավալով տվյալների մշակման համար գոյություն ունեցող մեթոդները, վերլուծության և ինտերակտիվ վիզուալիզացիայի ծրագրային միջավայրերը,

5. Մետատվյալների մշակման մեթոդները և լուծումները:

Հետազոտման մեթոդները: Աշխատանքի մեթոդաբանության և տեսական հիմքում ընկած են բարձր արտադրողականությամբ հաշվարկների, տվյալների պահպանման, վերլուծության և վիզուալիզացիայի ոլորտների առաջատար մասնագետների աշխատանքները: Անհրաժեշտ ծրագրային միջոցների մշակման և կիրառման համար օգտագործվել են C/C++, Python, JAVA, JavaScript ծրագրավորման լեզուները և ազատ հասանելի NodeJs, Angular Js, Mondrian, Tensorflow ծրագրային միջավայրերը:

Արդյունքների գիտական նորույթը.

- Մշակվել է միջազգային հիմնօրինակներին համապատասխան բարձր արտադրողականությամբ հաշվարկների իրականացման, տվյալների պահպանման, ձևափոխման, վերլուծության, ինտերակտիվ վիզուալիզացիայի և տարածման մեթոդներ պարունակող համալիր համակարգ՝ բարձր արտադրողականությամբ հաշվարկներ և մեծ տվյալներ պարունակող խնդիրների լուծման համար:
- Օգտագործելով առաջարկվող համալիր համակարգը, մշակվել են ամպային ծառայություններ.
    1. Քվանտային ֆիզիկայում ֆոտոնիկ դիսիպատիվ համակարգերի (Photonic Dissipative System) թվային հաշվարկների և վիզուալիզացիայի համար,
    2. Տարածաշրջանում եղանակի ուսումնասիրության և կանխատեսման համար օգտագործվող թվային մեթոդների արդյունքների ուսումնասիրության և նրանց ճշգրտության վերլուծությունների համար:

Ստացված արդյունքների կիրառական նշանակությունը: Մշակված ծրագրային միջավայրը կարելի է օգտագործել բարձր արտադրողականությամբ հաշվարկներ և մեծ տվյալներ պարունակող խնդիրներում՝ տվյալների պահպանման, վերլուծության, ինտերակտիվ վիզուալիզացիայի, տարածման և արդյունավետ հաշվարկների իրականացման համար, օգտագործելով ժամանակակից հաշվողական և ծրագրային միջոցները: Միջավայրը փորձարկվել է երկու տարբեր գիտական խնդիրների համար.

1. Քվանտային ֆիզիկայում ֆոտոնիկ դիսիպատիվ համակարգերում թվային հաշվարկների և վիզուալիզացիայի համար իրականացվել է

բազմաօգտատեր ամպային ծառայություն, որը հնարավորություն է տալիս մեկ համակարգում՝ ընտրելով հետազոտության համար անհրաժեշտ մուտքային պարամետրեր, կատարել հաշվարկներ և հետազոտել արդյունքները տրամադրված վիզուալիզացիայի գործիքների միջոցով:

2. Տարածաշրջանում եղանակի ուսումնասիրության և կանխատեսման համար օգտագործվող թվային մեթոդների արդյունքների ուսումնասիրության և նրանց ճշգրտության վերլուծության համար մշակվել է ամպային ծառայություն, որը հնարավորություն է տալիս ոչ միայն ինտեգրել փաստացի առկա դիտարկման տվյալները, կանխատեսման թվային մոդելների արդյունքները, այլ նաև կատարել վերլուծություններ՝ օգտագործելով տրամադրված վիզուալիզացիայի տեսակները և վերլուծական պարամետրերը:

Ներդրումներ: Մշակված համակարգի կիրառությամբ կառուցված տարածաշրջանում եղանակի ուսումնասիրության և կանխատեսման համար նախատեսված ամպային ծառայությունը ներդրվել է Հայաստանի պետական հիդրոօդերևութաբանության և մթնոլորտային երևույթների վրա ակտիվ ներգործության ծառայությունում:

Պաշտպանության ներկայացվող հիմնական դրույթները.

- Բարձր արտադրողականությամբ հաշվարկներ և մեծ տվյալներ պարունակող խնդիրների լուծման համար մշակված միջազգային հիմնօրինակներին համապատասխան, բարձր արտադրողականությամբ հաշվարկներ, տվյալների պահպանում, վերլուծություն և վիզուալիզացիա ապահովող, մետատվյալների տարածման հնարավորությամբ համալիր համակարգը:

Համակարգի կիրառության արդյունքում կառուցված հետևյալ ամպային ծառայությունները.

- Քվանտային ֆիզիկայում ֆոտոնիկ դիսիպատիվ համակարգերում թվային հաշվարկների և վիզուալիզացիայի համար,
- Տարածաշրջանում եղանակի ուսումնասիրության և կանխատեսման համար օգտագործվող թվային մեթոդների արդյունքների ուսումնասիրության և նրանց ճշգրտության վերլուծությունների համար:

Ստացված արդյունքների ապրոբացիան: Աշխատանքի հիմնական

արդյունքներն ու դրույթները զեկուցվել և քննարկվել են Սոֆիայում կայացած «Two Years Avitohol: Advanced High Performance Computing Applications 2017» սեմինարում, Բելգրադում կայացած «VI-SEEM Climate» տարածաշրջանային վերապատրաստման ժամանակ, ինչպես նաև ԻԱՊԻ ընդհանուր սեմինարներում:

Հրապարակումներ: Ատենախոսության հիմնական դրույթները հրապարակվել են հինգ (5) գիտական աշխատություններում, որոնց ցուցակը բերված է սեղմագրի վերջում:

Աշխատանքի կառուցվածքն ու ծավալը: Ատենախոսությունը կազմված է ներածությունից, 4 գլխից, եզրահանգումից և օգտագործված գրականության ցանկից: Ատենախոսությունը շարադրված է 115 էջերում, այն պարունակում է 54 նկար և 79 գրականության հղում:

ԱՇԽԱՏԱՆՔԻ ԲՈՎԱՆԴԱԿՈՒԹՅՈՒՆԸ

Ներածության մեջ հիմնավորված են ատենախոսության արդիականությունը, ձևակերպված են աշխատանքի նպատակներն ու խնդիրները, ինչպես նաև պաշտպանության ներկայացվող հիմնական դրույթները: Ներկայացված է ստացված արդյունքների գիտական նորույթը և կիրառական նշանակությունը:

Առաջին գլխում հետազոտվել են ժամանակակից հաշվողական ենթակառուցվածքները, այդ թվում՝ բարձր արտադրողականությամբ կլաստերային հաշվողական համակարգերը և ամպային տեխնոլոգիաները, մեծ տվյալների վերլուծության մարտահրավերները և վիզուալիզացիայի կարևոր հատկությունները:

Կլաստերը հանդիսանում է բարձր արտադրողականությամբ համակարգ, որտեղ օգտագործվում են համակցված հաշվարկային հզորություններ: Ամպային տեխնոլոգիան նոր մոտեցում է, որը տրամադրում է դինամիկ ընդլայնվող և վիրտուալ ռեսուրսներ:

Տվյալների վերլուծությունը գործընթաց է, որն օգտագործվում է տվյալները ստուգելու, ֆիլտրելու և ձևափոխելու համար՝ տվյալ իրավիճակում որոշակի եզրակացության հանգելու համար: Տվյալների պեղումը ներկայացնում է հետաքրքիր և օգտակար ձևանմուշների, ինչպես նաև մեծածավալ տվյալներում փոխհարաբերությունների հայտնաբերման գործընթաց: Այն համախմբում է վիճակագրական և արհեստական

բանականության (օրինակ, նեյրոնային ցանցեր և մեքենայական ուսուցում) գործիքները տվյալների բազաների կառավարման հետ՝ մեծ չափի թվային հավաքածուների վերլուծության համար:

Տվյալները տեսանելի օբյեկտներով արտացոլելը նպաստում է տվյալները ուսումնասիրելու և հասկանալու կարողության բարձրացմանը: Գոյություն ունեն վիզուալիզացիայի ստատիկ և ինտերակտիվ գործիքներ, վերջիններս թույլ են տալիս բազմակի պարամետրերի փոփոխման միջոցով ձևափոխել վիզուալիզացիոն արտապատկերումները: Այս տեսակի վիզուալիզացիաների շրջանակներում անհրաժեշտ է հաշվի առնել ընտրության, հղման, ֆիլտրման և վերակազմավորման գործառույթները:

Երկրորդ գլխում ներկայացված են HPDA խնդիրների լուծման համար ուսումնասիրված հայտնի ծրագրային գործիքները և գրադարանները, որոնք հնարավորություն են տալիս կատարել տվյալների վերլուծություններ և ինտերակտիվ վիզուալիզացիա: Ներկայացված են նրանց առավելություններն ու թերությունները:

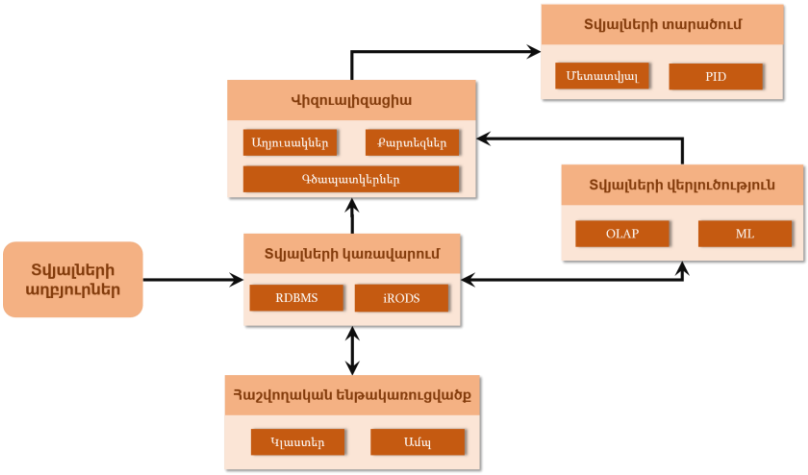
Հիմնվելով գոյություն ունեցող գործիքների ուսումնասիրության վրա, կատարել ենք դասակարգում, որը հնարավորություն է տալիս խմբավորել գործիքները՝ օգտագործելով չափանիշների ստանդարտ բազմությունը՝ դյուրին դարձնելով խնդրի յուրահատկություններին համապատասխան գործիքի ընտրությունը: Այդ գործիքները կարելի է դասակարգել ըստ իրենց տրամադրող ֆունկցիոնալության, կառուցվածքային ձևերի և աշխատանքային միջավայրերի: Ըստ աշխատանքային միջավայրի կարելի է բաժանել երկու խմբի՝ դեսկտոպ և վեբ, իսկ ըստ կառուցվածքի՝ առանց օգտագործողի միջավայրի և օգտագործողի միջավայրով:

Դեսկտոպ միջավայրերում աշխատող ծրագրային միջոցները հնարավորություն են տալիս օգտագործել լոկալ գրաֆիկական սարքավորումների հաշվողական հզորությունները, սակայն միևնույն ժամանակ իրենց բնույթով առաջացնում են բարդություններ, քանի որ պահանջում են համապատասխան օպերացիոն համակարգ և անհրաժեշտություն է առաջանում համակարգչում տեղադրել լրացուցիչ ծրագրային փաթեթներ: Այդպիսի գործիքներից են հանդիսանում օրինակ ParaView-ն և R-ը: Հրամանների տողով աշխատող գործիքները շատ հարմար են ներդրված կիրառությունների դեպքերում, երբ հնարավոր է մեկ այլ համակարգից օգտագործել տրամադրած միջոցները, սակայն հետազոտողից պահանջվում է

լրացուցիչ գիտելիքների իմացություն:

Վեբ միջավայրում աշխատող տվյալների վերլուծություն և վիզուալիզացիա ապահովող ծրագրային գործիքները հնարավորություն են տալիս օգտագործել տրամադրվող միջոցները՝ անկախ օգտագործվող սարքի ծրագրային ապահովումից: Խոսքը գնում է այնպիսի համակարգերի մասին, որոնց կառուցվածքը և բարդությունը չեն պահանջում բարդ հաշվողական ռեսուրսներ, որոնք վեբ միջավայրում հնարավոր է իրականացնել: Վեբ միջավայրում գոյություն ունեցող գործիքները բաժանվում են երկու դասի՝ գրադարաններ և պատրաստի օգտագործողի միջավայրով համակարգեր: Գրադարանները նախատեսված են որպես վիզուալիզացիա տրամադրող միջոց այլ համակարգերում կիրառությունների համար, այսինքն՝ իրենց որպես առանձին ծրագրային միջոց հնարավոր չէ օգտագործել (օր.՝ D3Js, VisJs): Իսկ օգտագործողի միջավայրով գործիքներից որոշները տվյալների բեռնման համար պահանջում են տրամադրել կոնկրետ սահմանված ձևաչափեր, սակայն ոչ բոլոր խնդիրներում է, որ վիզուալիզացիայի համար անհրաժեշտ արդյունքը ստացվում է նման ձևաչափերում (օր.՝ Quadigram, Tableua): Շատ հաճախ անհրաժեշտություն է լինում կիրառելու միջանկյալ միջոցներ, տվյալները համապատասխան ձևաչափերի փոխակերպելու համար: Որոշ ծրագրեր անվճար են տրամադրում միայն բազային հնարավորություններ, իսկ լրացուցիչ և ավելի ճկուն միջոցներ տրամադրում են միայն որոշակի գումարով:

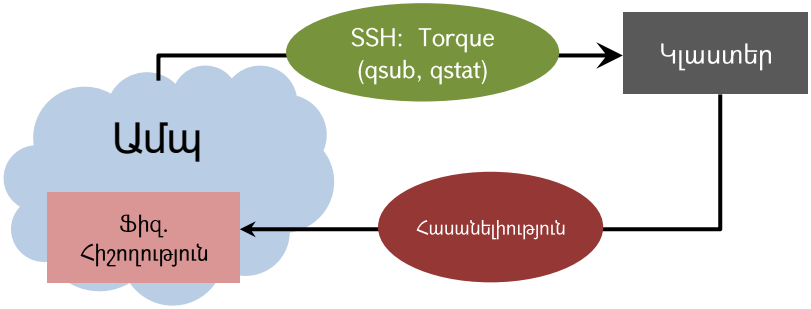
Երրորդ գլխում ներկայացված է մեր կողմից մշակված տվյալների վերլուծության և վիզուալիզացիայի ծրագրային բարդ համակարգը (Data Analytics and Visualization System - DAVS), որը հնարավորինս բավարար չափով լուծում է երկրորդ գլխում ներկայացված խնդիրները և, որի կիրառության միջոցով կարելի է կառուցել տվյալների վերլուծությունների և ինտերակտիվ վիզուալիզացիաների համակարգեր վեբ միջավայրում՝ հաշվի առնելով գիտական ոլորտի կամ առաջադրված խնդրի պահանջները:



Նկար 1. DAVS համակարգ

Մշակված DAVS համակարգը լուծում է բարձր արտադրողականությամբ հաշվարկների, մեծ ծավալով տվյալների մշակման և պահպանման խնդիրները՝ հնարավորություն տալով տվյալների վերլուծության և վիզուալիզացիայի գործիքների միջոցով ավելի արդյունավետ դարձնել տվյալների օգտագործումը և միջազգային հիմնօրինակներին համապատասխան ձևով տարածել ստացված արդյունքները (նկ. 1):

Հաշվողական ենթակառուցվածք բաղադրիչը հիմք է հանդիսանում մյուս բոլոր բաղադրիչների համար և իրականացնում է հաշվողական և ղեկավարման գործառույթ:



Նկար 2. Կլաստերային և ամպային համակարգերի միջև փոխհարաբերությունների մոդել

Բարձր արտադրողականությամբ հաշվարկների իրականացման համար օգտագործվում է կլաստերային համակարգ, իսկ սերվերին, տվյալների պահոցներին, ծրագրերին, ծառայություններին և հաշվողական ռեսուրսներին ցանցային մուտքի հասանելիության ապահովման համար՝ վիրտուալ ռեսուրսներ տրամադրող ամպային տեխնոլոգիա:

Այս երկու համակարգերը միմյանցից անկախ են և բարդություն է առաջանում նրանց միջև փոխհարաբերությունների ստեղծման: Այդ խնդիրը լուծելու համար մշակվել են ծրագրային միջոցներ, որոնց օգնությամբ ամպային համակարգից հաշվարկի համար անհրաժեշտ հարցումը կլաստերային համակարգում իրականացնելու համար օգտագործվում են Secure Shell (SSH) ցանցային արձանագրությունը և Torque հերթերի կազմակերպման համակարգի հրամանները (qsub, qstat) (նկ. 2):

Տվյալների կառավարում բաղադրիչը ապահովում է տվյալների պահպանումը, փնտրումը և հարցումների կազմակերպումը: Ոչ ստրուկտուրավորված տվյալներ պահպանելու և նրանց հետ աշխատելու բարդությունները հաղթահարելու համար մշակվել են ծրագրային միջոցներ, որոնց կիրառությունը թույլ է տալիս օգտագործել iRODS-ը (The integrated Rule-based Data management System), որը հանդիսանում է տվյալների՝ մետատվյալների օգտագործմամբ, կառավարման համակարգ:

Նկար 3. Տվյալների կառավարման բաղադրիչի աշխատանքային մոդել:

Ստրուկտուրավորված տվյալների պահպանման համար մշակված գրադարանը իրականացնում է հարաբերական տվյալների բազայի կառավարման համակարգում (Relational Database Management System - RDBMS) տվյալների ավելացման և ոչ ստրուկտուրավորված տվյալներից ստրուկտուրավորված ինֆորմացիայի հայտնաբերման գործառույթ (նկ. 3):

Տվյալների վերլուծություն բաղադրիչը նախատեսված է լուծելու մեծ ծավալով տվյալների հետ աշխատելու արդյունավետության և տվյալների օպտիմիզացիայի խնդիրները: Տվյալների բազայում մեծ ծավալի ինֆորմացիայի պարագայում հարցումների արդյունավետ իրականացման համար անհրաժեշտ է կիրառել լրացուցիչ լուծումներ: OLAP-ը (OnLine Analytical Processing) հանդիսանում է տվյալների պահոցներում բազմաչափ տվյալների

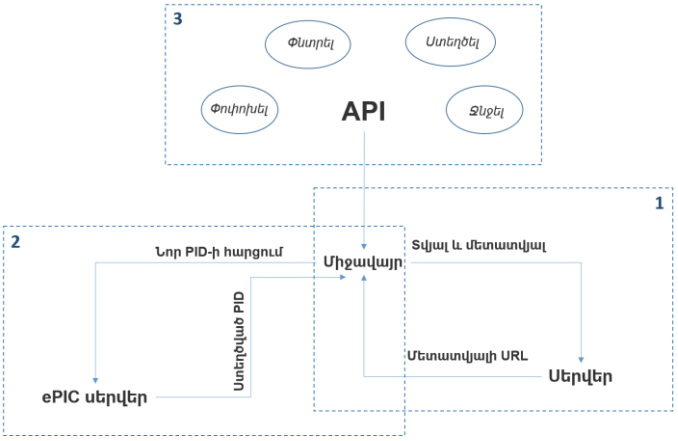
հասանելիության մոդել, որտեղ տվյալները կազմակերպող փաստերի և չափերի հավաքածուներ հանդիսացող տվյալների խորանարդների կառուցման միջոցով ստացվում է հարցումների արդյունավետ իրականացում: OLAP համակարգի կիրառման համար մշակվել է հարցումների կազմակերպման ծրագրային միջավայր, որի օգնությամբ օգտագործվում է բաց կոդով, JAVA ծրագրավորման լեզվով կառուցված Mondrian ծրագրային փաթեթը: Մշակվել է օժանդակ միջավայր տվյալների բազայի կառուցվածքի հիման վրա դյուրին կերպով տվյալների խորանարդներ և հարցումների բազմաչափ արտահայտություններ (Multidimensional Expressions - MDX) ստեղծելու համար:

Արհեստական նեյրոնային ցանցերը լայնորեն կիրառվում են ճանաչման, դասակարգման, կանխատեսման խնդիրներում: Կառուցված համակարգը հնարավորություն է տալիս տրամադրել համապատասխան ցանցային և մուտքային տվյալների կարգավորումները, և մշակված միջանկյալ գործիքների միջոցով իրականացնել նեյրոնային ցանցերի ուսուցում և կիրառում՝ օգտագործելով Google Brain թիմի կողմից մշակված բարձր արտադրողականությամբ թվային հաշվարկների TensorFlow բաց կոդով ծրագրային գրադարանի և Python ծրագրավորման լեզվի հնարավորությունները:

Վիզուալիզացիայի բաղադրիչը իրականացնում է տվյալների արտացոլում՝ տվյալների կառավարման կամ վերլուծության բաղադրիչների տվյալների հիման վրա, օգտագործվելով ինտերակտիվ վիզուալիզացիայի վեբ գրադարաններ: Մշակվել է API (Application Programming Interface), որի միջոցով օգտագործողի միջավայրը համապատասխան ձևաչափում ստանում է արտացոլման համար անհրաժեշտ տվյալները սերվերային հատվածից և, օգտագործելով վիզուալիզացիայի վեբ գրադարանները, կառուցում է տարբեր տեսակի ինտերակտիվ վիզուալիզացիաներ, սկսած պարզագույն գրաֆիկներից մինչև բարդ կառուցվածքով պատկերներ՝ 2D և 3D ձևաչափերում, և աշխարհագրական օբյեկտներ՝ քարտեզների միջոցով:

Տվյալների տարածման բաղադրիչը ապահովում է հետազոտությունների, սիմուլյացիաների արդյունքների կամ այլ տվյալների վիզուալիզացիաների տարածում: Europian Commission-ի կողմից 2013 թվականին կազմավորված Research Data Alliance (RDA) հետազոտական համայնքը սահմանել է տվյալների ազատ տարածման ստանդարտներ, որտեղ թվային տվյալների տարածման համար օգտագործվում են մշտական

նույնացուցիչներ (Persistance Identifier - PID): ePIC կոնսորցիումը տրամադրում է PID-երի տեղաբաշխման և կառավարման համակարգ, որի օգտագործմամբ մշակվել է մետատվյալների ստեղծման և տվյալների պահպանման ծրագրային միջավայր՝ հետազոտության արդյունքում ստացված արդյունքները տարածելու համար (նկ. 4):



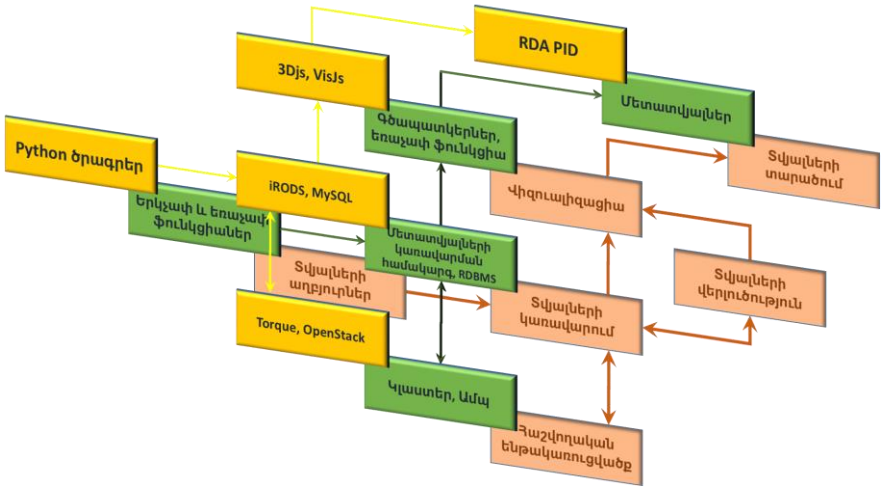
Նկար 4. Տվյալների տարածման բաղադրիչ: 1) Սերվեր 2) ePIC սերվեր 3) Կիրառման ծրագրավորման ինտերֆեյս (API)

Մշակված ծրագրային միջավայրում տվյալները իրենց համապատասխան մետատվյալներով պահվում են լոկալ տվյալների բազայում, որից հետո ստանում են իրենց լոկալ նույնացուցիչը, որի օգտագործմամբ այնուհետև հարցում է կատարվում ePIC սերվերին և նոր PID գեներացվելուց հետո այն պահվում է լոկալ բազայի համապատասխան տողում:

Չորրորդ գլխում ներկայացված է մշակված DAVS համակարգի արդյունավետության և կիրառական հնարավորությունները ուսումնասիրելու համար կատարված փորձարկումների արդյունքում ստեղծված ամպային ծառայությունները՝ երկու տարբեր գիտական ոլորտների խնդիրների լուծման համար:

Ֆոտոնիկ դիսիպատիվ համակարգերի թվային հաշվարկների և ինտերակտիվ վիզուալիզացիայի համար մշակվել է բազմաօգտատեր ամպային ծառայություն[2] (նկ. 5):

[2] Ամպային ծառայություն. http://quantum.grid.am/



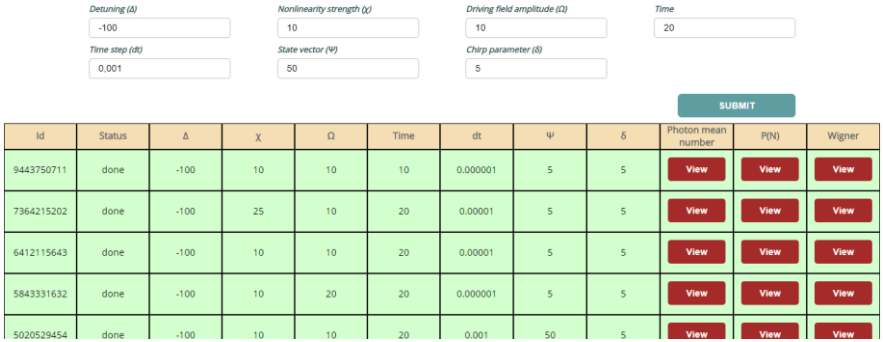
Նկար 5. Ֆոտոնիկ դիսիպատիվ համակարգերի թվային հաշվարկների և վիզուալիզացիայի միջավայրի ենթակառուցվածք:

Ներկայումս քվանտային ֆիզիկան շատ կարևոր է մի շարք գիտական ծրագրերի համար, որտեղ այլևս հնարավոր չէ անտեսել շրջակա միջավայրի փոխազդեցությունը: Նման դեպքերում քվանտային համակարգերը դիտվում են որպես բաց համակարգեր և նրանց ժամանակային էվոլյուցիան նկարագրվում է հիմնական հավասարման շրջանակներում խտության մատրիցայով: Այս քվանտային համակարգերի մոդելավորումը և վիզուալիզացիան թույլ են տալիս հետազոտողներին հաշվարկել և ուսումնասիրել այնպիսի պարամետրերի զգայունությունը, ինչպիսիք են՝ գրգռված ֆոտոնների թիվը, ֆոտոնների թվի բաշխման ֆունկցիան և Վիգներ ֆունկցիաները[3]:

Ամպային ծառայության հաշվարկների հիմքում ընկած է մշակված ECAPT (Extendible C++ Application in Photonic Technologies) գրադարանը, որում ներկայացված է բարդ քվանտային համակարգերի թվային սիմուլյացիա և մոդելավորում: Նշված գրադարանում կատարել ենք զուգահեռացում՝ արդյունավետության բարձրացման և բարձր արտադրողականությամբ համակարգերի հնարավորությունները օգտագործելու համար: Օգտագործողի

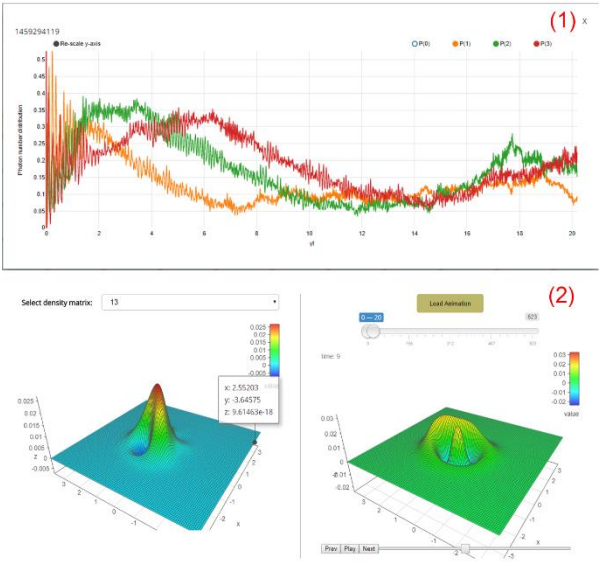
[3] Jon Brogaard. Wigner function formalism in Quantum mechanics

հարմարավետ միջավայրը հնարավորություն է տալիս իրականացնել նոր սիմուլյացիաներ, լրացնելով անհրաժեշտ դաշտերը, արդյունքում կատարել վերլուծություններ՝ հիմնվելով տրամադրած ինտերակտիվ վիզուալիզացիայի հնարավորությունների վրա (նկ. 6, 7):



Id	Status	Δ	χ	Ω	Time	dt	Ψ	δ	Photon mean number	P(N)	Wigner
9443750711	done	-100	10	10	10	0.000001	5	5	View	View	View
7364215202	done	-100	25	10	20	0.00001	5	5	View	View	View
6412115643	done	-100	10	10	20	0.00001	5	5	View	View	View
5843331632	done	-100	10	20	20	0.000001	5	5	View	View	View
5020529454	done	-100	10	10	20	0.001	50	5	View	View	View

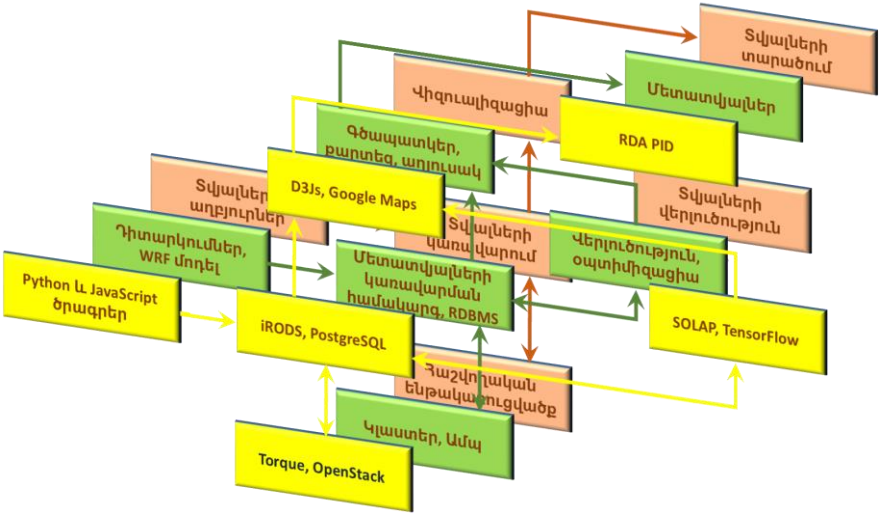
Նկար 6. Համակարգում սիմուլյացիաների կատարման և արդյունքների ուսումնասիրության վահանակ:



Նկար 7. Համակարգի վիզուալիզացիաներ. (1) Ֆոտոնների թվի բաշխում, (2) Վիգներ ֆունկցիա:

Արդյունքների վիզուալիզացիան հնարավորություն է տալիս հետազոտողներին վերլուծելու և ուսումնասիրելու համակարգերը առավել պրակտիկ և լիովին տարբեր ժամանակային պարբերությունների և ժամանակային կետերի համար: Բացի տարբեր սիմուլյացիաների համար տրամադրված պահոցից, մշակված միջավայրը դարձնում է տեսական հետազոտությունը փորձարարական և վիզուալ, որը խնայում է հետազոտության ժամանակը և կիրառվող ռեսուրսները:

Տարածաշրջանում եղանակի ուսումնասիրության և կանխատեսման համար օգտագործվող թվային մեթոդների արդյունքների ուսումնասիրության և նրանց ճշգրտության վերլուծության համար մշակվել է ամպային ծառայություն, որն իրականացնում է տվյալների հավաքագրում, պահպանում, վերլուծություն, ինտերակտիվ վիզուալիզացիաներ[4] (նկ. 8):



Նկար 8. Եղանակի ուսումնասիրության և կանխատեսման միջավայրի ենթակառուցվածք:

Հայաստանում տարածաշրջանային եղանակի ուսումնասիրությունների համար օգտագործվում են տարբեր դիտարկման աղբյուրներից ստացված տվյալներ: Օդերևութաբանական կայանները յուրաքանչյուր երեք ժամը մեկ տրամադրում են տվյալ պահի եղանակային

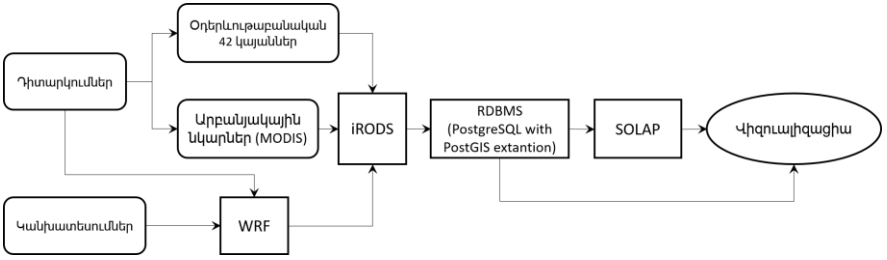
[4] Ամպային ծառայություն. http://meteo.grid.am/

տվյալները հատուկ SYNOP ֆորմատով: Քանի որ կայանների քանակը սահմանափակ է, և այն չի ծածկում Հայաստանի ամբողջ տարածքը, օգտագործվել են նաև արբանյակային նկարներ, որոնցից GrassGIS ծրագրային փաթեթի միջոցով հաշվարկվում է երկրի մակերևույթի ջերմաստիճանը (Land Surface Temperature - LST):

Եղանակի օպերատիվ կանխատեսման համար օգտագործվել է WRF թվային կանխատեսման տարածաշրջանային մոդելը: Այն կարգավորվել է Հայաստանի տարածքի համար՝ օգտագործելով երկու տիրույթ.

- D01 – Հորիզոնական 18 կմ քայլով՝ 202x202 ցանցային կետերով, 40,0 լայնության և 44,7 երկայնության կենտրոնական կետով,
- D02 – Ներդրված D1-ում, 6 կմ քայլով՝ 61x70 ցանցային կետերով:

Նախնական կանխատեսման տվյալներն են National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS)-ի կողմից տրամադրված վերլուծությունները և կանխատեսումները՝ 0,5 աստիճան հորիզոնական ցանցային չափով: Քանի որ GFS-ի կանխատեսումներում հաշվի չեն առնվում Հայաստանի բոլոր կայանների դիտարկումների տվյալները, օգտագործվում է WRF մոդելի WRFDA համակարգը, որը նախնական կանխատեսման տվյալներում ավելացնում է լրացուցիչ դիտարկման տվյալներ՝ WRF մոդելի արդյունքը լավացնելու նպատակով: Մշակվել է ծրագրային փաթեթ դիտարկման արդյունքներից LITTLE_R ֆորմատով ֆայլ կառուցելու համար, որը օգտագործվում է WRFDA համակարգում:

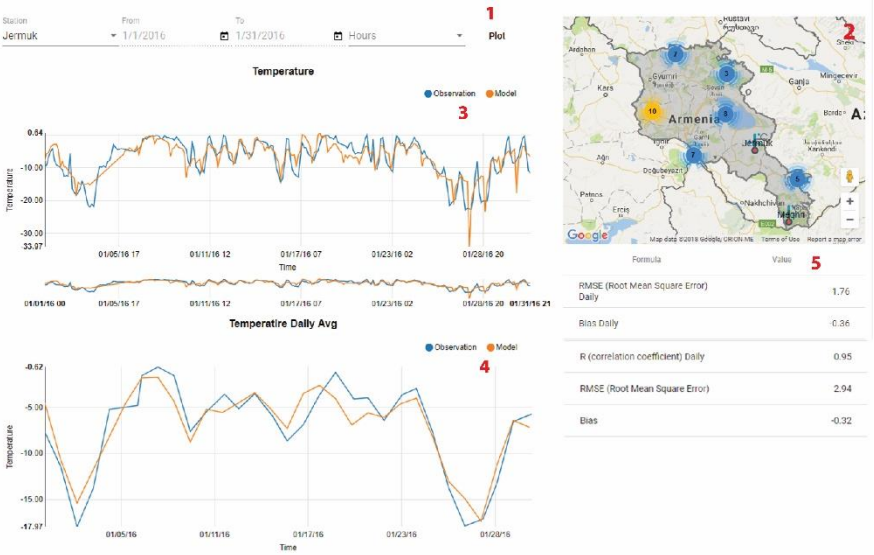


Նկար 9. Համակարգում տվյալների հոսք:

Միջավայրում տվյալների մեծ ծավալների և աշխարհագրական օբյեկտների առկայության պարագայում հարցումները արդյունավետ իրականացնելու համար օգտագործվել է Spatial OLAP (SOLAP) մոդելը, որը

հանդիսանում է OLAP-ի ընդլայնում՝ ապահովելու համար տարածական տվյալների օգտագործումը (նկ. 10):

Նկար 10. Ամպային ծառայության վահանակ. 1- հարցման դաշտ, 2 - քարտեզ, 3 – ջերնաստիճանի գծապատկեր, 4 – օրական միջին ջերնաստիճանի գծապատկեր, 5 – գործակիցների աղյուսակ:



Միջավայրը մշակվել է Հայաստանում եղանակային տվյալների ուսումնաիրության և թվային կանխատեսման մոդելների արդյունքների ճշգրտության հետազոտման համար: Օգտագործման համար դյուրին և հարմարավետ միջավայրը տրամադրում է եղանակային տվյալների վերլուծություն, վիճակագրական գործակիցների (RMSE, BIAS, R) և ինտերակտիվ վիզուալիզացիաներ (նկ. 10):

ԱՇԽԱՏԱՆՔԻ ՀԻՄՆԱԿԱՆ ԱՐԴՅՈՒՆՔՆԵՐԸ

- Մշակվել է տվյալների վերլուծության և ինտերակտիվ վիզուալիզացիայի՝ տվյալների ստացման, մշակման, ձևափոխման, պահպանման, բարձր արտադրողականությամբ հաշվարկների իրականացման, օպտիմիզացիայի և տարածման մեթոդներ պարունակող համալիր համակարգ [1-5]:

- Մշակվել է և տվյալների լոկալ պահպանման հնարավորությամբ միջավայր՝ խնդրին համապատասխան մետատվյալների ստեղծման և նրանց տարածման համար, որում հաշվի են առնված միջազգային հիմնօրինակները [1]:
- Մշակվել է բազմաօգտատեր ամպային ծառայություն քվանտային ֆիզիկայում ֆոտոնիկ դիսիպատիվ համակարգերում թվային հաշվարկների կատարման, արդյունքների վերլուծության և ինտերակտիվ վիզուալիզացիայի իրականացման համար [2]:
- Մշակվել է ամպային ծառայություն Հայաստանում եղանակի ուսումնասիրության և կանխատեսման համար օգտագործվող թվային մեթոդների արդյունքների վերլուծության և ինտերակտիվ վիզուալիզացիայի համար [4, 5]:

ՀՐԱՊԱՐԱԿՎԱԾ ԱՇԽԱՏԱՆՔՆԵՐԻ ՑԱՆԿ

1. Grigoryan H., Increasing the Visibility of Scientific Data in Armenia Using Persistent Identifiers, Mathematical Problems of Computer Science, Yerevan, Armenia, 2016, Volume 46, pp. 81-86
2. Grigoryan H., Astsatryan H., Gevorgyan T., Manukyan V., Cloud Service for Numerical Calculations and Visualizations of Photonic Dissipative Systems, Cybernetics and Information Technologies, Volume 17, No 5, Sofia, Bulgaria, 2017, pp. 89-100
3. Grigoryan H., Automation Script for WRF Model Data Assimilation System, Mathematical Problems of Computer Science, Yerevan, Armenia, 2017, Volume 48, pp. 57-63
4. Grigoryan H., Abrahamyan R., Cloud Service for Analysis and Interactive Visualization of Weather Data in Armenia, Mathematical Problems of Computer Science, Yerevan, Armenia, 2018, Volume 49, pp. 49-57
5. Astsatryan H., Grigoryan H., Gyulgyulyan E., Hakobyan A., Kocharyan A., Narsisian W., Sahakyan V., Shoukourian Yu., Mkoyan A., Abrahamyan R., Petrosian Z., Aligon J., Weather Data Visualization and Analytical Platform, Scalable Computing: Practice and Experience, Volume 19, No 2, pp 79-86, Institute e-Austria Timisoara, Romania, 2018 DOI: 10.12694/scpe.v19i2.1351

Разработка среды анализа и визуализации проблем с высокопроизводительными вычислениями и большими данными

Айк Андраникович Григорян

Совместно с развитием компьютерных технологий увеличиваются генерируемые данные и из-за сложной структуры и объема их разнообразных наборов (структурированный и неструктурированный) требуются новые подходы к их обработке, хранению и анализу.

Вышеупомянутые наборы - это большие данные (big data), которые по Doug Laney определяются тремя свойствами (объем, скорость, многообразие): «Большие данные - это набор больших объемов, высокоскоростной и/или большой многообразной информации, в котором для принятия решений, определения смысла и оптимизации процессов требуются новые подходы обработки». Со временем появились еще два свойства, определяющие большие данные: достоверность и ценность. Для обработки, хранения, анализа, визуализации и распространения больших данных используется высокопроизводительный анализ данных (HPDA - High Performance Data Analytics), который объединяет высокопроизводительные вычисления (HPC - High Performance Computing) с большими данными. Для решения проблем HPDA существующие инструменты обеспечивают реализацию определенной подгруппы из вышеуказанной группы действий в зависимости от целевой задачи или сложности взаимодействий, существующих для каждого действия программных решений. Пользователь, чтобы реализовать весь цикл действий, вынужден использовать несколько программных сред для решения проблемы и исследования результатов.

По этой причине необходимо разработать методы и программное обеспечение, которые позволят объединить действия высокопроизводительных вычислений, хранения, анализа, визуализации и распространения данных в единой комплексной системе с использованием современных программных решений и международно признанных стандартов. В результате пользователь сможет реализовать полный процесс решения проблемы с помощью одной программной среды, что повысит эффективность и сэкономит время.

Целью диссертации является создание комплексной системы с открытым исходным кодом, в соответствии с международными стандартами, с возможностью высокопроизводительных вычислений, хранения, анализа,

визуализации данных и распространения метаданных.

Научная новизна:

- Разработана комплексная система в соответствии с международными стандартами, содержащая методы выполнения высокопроизводительных вычислений, хранения, анализа, интерактивной визуализации и распространения данных, для решения задач с высокопроизводительными вычислениями и большими данными.
- С помощью предлагаемой комплексной системы были разработаны облачные сервисы:
  1. Для цифровых расчетов и визуализации фотонных диссипативных систем в квантовой физике.
  2. Для изучения результатов и анализа точности цифровых методов, используемых для исследования и прогнозирования погоды в регионе.

Области применения результатов. Разработанная программная среда может использоваться в задачах с высокопроизводительными вычислениями и большими данными, для хранения, анализа, интерактивной визуализации, распространения и реализации эффективных вычислений с использованием современных вычислительных и программных средств.

Основные результаты диссертации:

- Разработана комплексная система, содержащая методы для реализации высокопроизводительных вычислений, модификации, хранения, анализа, интерактивной визуализации и распространения данных [1-5].
- Разработана среда в соответствии с международными стандартами, адаптирующаяся к свойствам задач и с возможностью локального хранения данных, для создания и распространения метаданных [1].
- Разработан многопользовательский облачный сервис для реализации цифровых расчетов фотонных диссипативных систем в квантовой физике, анализа и интерактивной визуализации результатов [2].
- Разработан облачный сервис для анализа результатов и интерактивной визуализации численных методов, используемых для исследования и прогнозирования погоды в Армении [4, 5].

Development of systems for problems with hight performance computing and big data

Hayk Andranik Grigoryan

Together with the development of computer technologies, the generated data increases and, due to the complex structure and volume of their diverse sets (structured and unstructured), new approaches to their processing, preservation and analysis are required.

Above described data sets are called big data and Doug Laney defined them using a three-dimensional perspective: «Big data is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision-making, insight discovery and process optimization». Over time, two more properties have emerged that determine large data: veracity and value. High performance data analytics (HPDA) is used to process, store, analyze, visualize and share big data, which combines high performance computing (HPC) with large data. To solve HPDA problems, existing tools provide implementation of a certain subgroup from the above-mentioned group of actions, depending on the target task or the complexity of the interactions of existing for each action software solutions. The user, in order to implement the entire cycle of actions, is forced to use several software environments to solve the problem and study the results.

For this reason, it is necessary to develop methods and software that will combine the actions of high-performance computing, storage, analysis, visualization and data dissemination in a single integrated system using modern software solutions and internationally recognized standards. As a result, the user will be able to implement the full process of solving the problem using a single software environment, which will increase efficiency and save time.

The aim of this thesis is the development of a comprehensive system with open source code, in accordance with international standards providing high-performance calculations, data storage, analysis and visualization of data and metadata distribution.

Scientific novelty:

- Developed complex system of high-performance calculations, data storage,

conversion, analysis, interactive visualization and dissemination techniques, high-performance calculations and solutions for problems with high-performance computing and big data.

- Using the proposed complex system two cloud services have been developed:
    1. Numerical simulation and result interactive visualization of photonic dissipative systems in quantum physics.
    2. Weather research and forecasting numerical methods results investigation and accuracy analysis in the region.

Applicability of the results. Developed complex system can be used for solving problems with high-performance computing and big data. It provides abilities for data storage, analysis, interactive visualization, sharing and effective calculations using modern computing and software tools.

The main results of the thesis are:

- Developed data analysis and interactive visualization complex system which consists from different methods (API) for data collection, processing, conversion, maintenance, optimization, sharing and high performance calculation [1-5].
- Developed an environment that is flexible to the problem's specifications and has a local storage capability for the creation and sharing of metadata in which the international guidelines are taken into account [1].
- Developed a multi-user cloud service for numerical calculations of photonic dissipative systems in quantum physics and results analysis with interactive visualization [3].
- Developed a cloud service for analysis and interactive visualization of weather research and forecasting numerical method results in Armenia [4-5].