



ALGORITHMS AND SOFTWARE FOR PARALLEL PROCESSING OF REMOTE SENSING ARCHIVE DATA IN A DISTRIBUTED COMPUTING ENVIRONMENT

A. K. Utebergenova

Assistant, University of Innovation Technologies

<https://doi.org/10.5281/zenodo.19105725>

ARTICLE INFO

Received: 12th March 2026

Accepted: 18th March 2026

Online: 19th March 2026

KEYWORDS

Remote sensing, parallel computing, distributed systems, Big Data, cluster systems, Apache Spark, MPI, algorithms.

ABSTRACT

This article examines the issues of operational processing of large volumes of archival data (Big Data) from remote sensing of the Earth (RSE). Methods for optimizing time and resources using parallel algorithms in distributed computing systems, as well as the capabilities of modern software platforms, are analyzed.

YERNI MASOFADAN ZONDLASH ARXIV MA'LUMOTLARINI TAQSIMLANGAN HISOBLASH MUHITIDA PARALLEL QAYTA ISHLASH ALGORITMLARI VA PROGRAMMAVIY TA'MINOTI

Utebergenova A. K.

Assistent, Innovatsion texnologiyalar universiteti

<https://doi.org/10.5281/zenodo.19105725>

ARTICLE INFO

Received: 12th March 2026

Accepted: 18th March 2026

Online: 19th March 2026

KEYWORDS

Yerni masofadan zondlash, parallel hisoblashlar, taqsimlangan tizimlar, Big Data, klasterli tizimlar, Apache Spark, MPI, algoritmlar.

ABSTRACT

Ushbu maqolada yerni masofadan zondlash (YMZ) arxiv ma'lumotlarining katta hajmini (Big Data) operativ qayta ishlash masalalari o'rganiladi. Taqsimlangan hisoblash tizimlarida parallel algoritmlarni qo'llash orqali vaqt va resurslarni optimallashtirish usullari, shuningdek, zamonaviy dasturiy platformalarning imkoniyatlari tahlil qilinadi.

So'nggi o'n yillikda kosmik texnologiyalarning rivojlanishi natijasida yerni masofadan zondlash (YMZ) ma'lumotlarining hajmi va sifati keskin ortdi. Sun'iy yo'ldoshlar har kuni yer yuzining turli nuqtalaridan gigabaytlab, hatto terabaytlab vizual va spektral ma'lumotlarni uzatib turadi. Bu ma'lumotlar arxivlarda to'planib, uzoq

yillik monitoring uchun zarur manba bo'lib xizmat qiladi. Biroq, to'plangan "uch o'lchovli" (fazoviy, spektral va vaqt) ma'lumotlar arxivini an'anaviy bir protsessorli kompyuterlarda qayta ishlash ko'p vaqt talab etadi. Bu esa tezkor qaror qabul qilish jarayonini sekinlashtiradi. Bu muammodan chiqish uchun taqsimlangan hisoblash tizimlari



(distributed computing) va parallel algoritmlarni qo'llash bugungi kunning eng dolzarb masalasi hisoblanadi.

1. YMZ ma'lumotlarining strukturasi va qayta ishlash bosqichlari

YMZ ma'lumotlari ko'pincha raster ko'rinishidagi ko'p qavatli tasvirlar bo'lib, ular har bir pikselida yer yuzidagi ob'ektlarning qaytish xususiyatlarini aks ettiradi. Arxiv ma'lumotlarini qayta ishlash quyidagi asosiy bosqichlardan iborat:

1. **Oldindan qayta ishlash:** Atmosfera va geometric xatoliklarni tuzatish.
2. **Raqamli tahlil (Feature Extraction):** Tasvirlardagi turg'un ob'ektlarni (yo'l, daryo, imorat) ajratib olish.
3. Indeksli tahlil (**Index Analysis**): NDVI (vegetatsiya), NDWI (suv) kabi spektral indeksni hisoblash
4. **Dinamik monitoring (Change Detection):** Arxivdagi eski tasvirlar bilan yangilarini solishtirish orqali ekologik o'zgarishlarni aniqlash.

2. taqsimlangan hisoblash muhitining arxitekturasi

Taqsimlangan hisoblash muhiti bir-biri bilan yuqori tezlikdagi tarmoq orqali bog'langan bir nechta hisoblash tugunlaridan (node) iborat. Bu yerda katta topshiriq kichikroq bo'laklarga bo'linib, har bir tugunda parallel ravishda bajariladi.

Parallel qayta ishlashning ikki asosiy modeli bor:

1. **Klasterli sistemalar:** Bir korxonada yoki institut tarkibidagi lokal serverlar yig'indisi.
2. **Bulutli hisoblashlar (Cloud Computing):** Amazon AWS, Google

Cloud yoki Microsoft Azure kabi uzoqdagi platformalar.

3. **Parallel qayta ishlash algoritmlari** YMZ ma'lumotlarini qayta ishlashda eng samarali algoritmlar – bu ma'lumotlarni domenlarga bo'lish (Domain Decomposition) prinsipiga asoslangan.

3.1. MapReduce Algoritmi

MapReduce – kata ma'lumotlar qo'plamini qayta ishlashning eng taniqli modeli. U ikki bosqichdan iborat:

1. **Map (Kartaga tushirish):** Katta rasm arxivini kichkina bloklarga bo'ladi va klaster tugunlariga tarqatadi. Har bir tugun o'zidagi blokni parallel qayta ishlaydi.
2. **Reduce (Biriktirish):** Tugunlarda orinlangan natijalar yig'iladi va umumiy bir butun karta yoki statistic ma'lumot shakillantiradi.

3.2. Parallel Segmentatsiya Algoritmi

Ushbu algoritmda tasvirdagi obyektlarni klassifikatsiyalash uchun neyron tarmoqlari parallel ravishda ishlatiladi. Tasvir bir necha "tile" (plitka) ko'rinishidagi bo'laklarga bo'linib, protsessorlar o'rtasida teng taqsimlanadi.

4. Programmaviy taminot va texnologiyalar

Hozirgi vaqtda YMZ ma'lumotlarini parallel qayta ishlash uchun quyidagi dasturiy yechimlar keng qo'llanilmoqda.

Apache Spark: "In-memory" (operativ xotirada) hisoblash texnologiyasi bo'lib, YMZ algoritmlarini o'n marta tezroq bajarish imkonini beradi. Spark tarkibidagi Spark SQL va MLlib (mashinali o'qitish) tasvirlarni tahlil qilish uchun barcha vositalarga ega.

MPI (Message Passing Interface): Eng an'anaviy va quvvatli parallel dastur



yoziq standarti. Bu ko'pincha superkompyuterlarda matematik modellashtirish uchun qo'llaniladi.

GDAL (Geospatial Data Abstraction Library): Raster ma'lumotlari bilan ishlovchi eng asosiy kutubxona bo'lib, u parallel jarayonlarni qo'llab-quvvatlaydi.

Python Dask: Python tili asosan katta massivlarni parallel qayta ishlash uchun qulay platforma.

5. Hisoblash samaradorligini baholash

Parallel algoritmlarning samaradorligi ular qancha yaqshi "masshtablanishiga" (scaling) bog'lik. Agar P - procesorlar soni bo'lsa, ideal holatda hisoblash vaqti P barobar kamayishi kerak. Lekin, tarmoqdagi ma'lumot almashish tugunlarni sinxronlash vaqti sababli betob kishi xatoliklar bolishi mumkin.

Amdal qonuni (Amdahl's Law) b'yicha parallellashning shegarasi bor:

$$S = \frac{1}{(1-f) + \frac{f}{P}}$$

Bunda f - algoritmnin parallelashga keledigan qismi. Bu formula orqali biz arxiv ma'lumotlarini qayta ishlashda qancha procesor qollanish eng optimal ekenligini aniqlaymiz.

6. Taqsimlangan Klaster Sistemasiing Strukturaviy Sxemasi

Taqsimlangan hisoblashlarda klaster arxitekturasi YMZ ma'lumotlarini qayta ishlash tezligiga bevosita tasir qiladi. Quyidagi tablicada standart klaster sistemasining tarkibi va har bir bo'limining vazifasi keltirilgan:

1-tablica. YMZ ma'lumotlarini qayta ishlash klasterining strukturasi

Komponent nomi	Vazifasi	Asosiy xizmati
Master Node (Boshqaruvchi tugun)	Resurslarni boshqarish	Topshiriqlarni taqsimlash, tugunlarning holatini kuzatish.
Worker Nodes (Ishchi tugunlar)	Parallel hisoblash	Ma'lumot bloklarini (tiles) parallel qayta ishlash.
Distributed Storage (HDFS / S3)	Ma'lumatlarni saqlash	Arxivdagi katta rasmlarni taqsimlangan holda saqlash.
Cluster Manager (YARN / Mesos)	Resurslar ro'yxati	Procesor (CPU/GPU) va xotira (RAM) resurslarini ajratish.
High-Speed Network (InfiniBand)	Ma'lumot almashish	Tugunlar o'rtasidagi bog'lanish tezligining yuqori bo'lishini tamiinlash.



**7. YMZ Arxivlarini Qayta Ishlash
Algoritmlarining taqqoslangan tahlili**

YMZ ma'lumotlari bilan ishlashda algoritmnining turiga qarab parallellash usullari farq qiladi.

**2-tablica. Algoritmlarning
parallellash imkoniyatlari**

Algoritm turi	Parallellash usuli	Qollaniladigan texnologiya	Effektliligi
Piksel bo'yicha tanlash	Data Parallelism	Apache Spark, Dask	Juda yuqori
Obyektga yo'naltirilgan (OBIA)	Domain Decomposition	MPI, OpenMP	O'rta
Neyron tarmoqlar (CNN)	Model/Data Parallelism	TensorFlow, PyTorch	Yuqori (GPU kerak)
Atmosferaviy korrekciya	Task Parallelism	Multiprocessing	Yuqori

**8. Programmaliq Tamiynatting
Salistirmali Kórsetkishleri**

Hozirgi vaqtdagi taniqli programmaviy platformalarning YMZ

arxivlari bilan ishlashdagi xossalari quyidagicha:

3-tablica. Programmaviy taminotlarning texnik tahlili

Platforma	Qollash tili	Hisoblash modeli	Arxivlar bilan ishlash
Apache Spark	Scala, Python, Java	In-memory RDD	Streaming va Batch processing
MPI	C++, Fortran	Message Passing	Yuqori mahsulotli klasterler (HPC)
Google Earth Engine	JavaScript, Python	Cloud-based	Global arxivlarga (Landsat/Sentinel) bevosita kirish



Platforma	Qollash tili	Hisoblash modeli	Arxivlar bilan ishlash
Python Dask	Python	Task Graphs	NumPy va Pandas massivlarini parallellash

9. Hisoblash Modelini Optimallashtirish

Arxiv ma'lumotlari bilan ishlashda eng kata muammo - bu "I/O bottleneck" ya'ni ma'lumotlarni oqish va yozish tezligining pastligi. Buni yechish uchun ma'lumotlar Cloud Optimized GeoTIFF (COG) yoki Zarr formatlarida saqlanishi taklif qilinadi.

Parallel hisoblashning asosiy ko'rsatkichi - **Speedup** (S) va **Efficiency** (E):

1. **Speedup** (S): Bir procesordagi vaqtning P procesordagi vaqtga nisbati:
 $S = T_1 / T_P$.

2. **Efficiency** (E): Procesorlardan qanchalik samarali foydalanilganligi:
 $E = S / P$.

YMZ arxiv ma'lumotlarini qayta ishlashda taqsimlangan sistemalarni qollanish faqat vaqtni tejap qoymay, balki avval mumkin bolmagan global monitoring topshiriqlarini bajarishga imkon beradi.

References:

1. Chandra, R. "Parallel Programming in C++ with MPI", 2019.
2. Gorelick, N. et al. "Google Earth Engine: Planetary-scale geospatial analysis for everyone", 2017.
3. Karau, H. "Learning Spark: Lightning-Fast Big Data Analysis", 2015.
4. Jolshibaev, K. "Jerdi araliqtan zondlawda Big Data texnologiyalari", 2023.