

## **KATTA HAJMDAGI MA'LUMOTLAR VA ULARNING TAHLILI: TUSHUNCHALARI, TURLARI VA TEXNOLOGIYALARI**

**G'ofurov Olimjon Shuxratjon o'g'li**

Muhammad Al-Xorazmiy nomidagi Toshkent Axborot Texnologiyalari  
Universiteti kompyuter injineyringi:  
Amaliy dasturiy vositalarni loyihalash yo'nalishi II bosqich magistranti

### **ANNOTATSIYA**

Hozirgi kunda kompaniyalar to'g'ri qaror qabul qilish va o'z strategiyalarini qo'llab-quvvatlash uchun katta hajmdagi ma'lumotlarning mavjudligi muhimligini anglay boshladilar. Yangi texnologiyalar, Internet va ijtimoiy tarmoqlarning rivojlanishi bilan raqamli ma'lumotlarni ishlab chiqarish doimiy ravishda o'sib bormoqda. "Katta ma'lumotlar" atamasi xarakteristikalarini (katta hajm, turli shakllar, qayta ishlash tezligi) o'ziga xos va tobora murakkab kompyuter saqlash va tahlil vositalarini talab qiladigan kompaniyalar va shaxslar tomonidan ishlab chiqarilgan raqamli ma'lumotlarning heterojen massasini anglatadi. Ushbu maqola Big Data kontsepsiyasini, uning tushunchalari, muammolari va qo'llanilishini, shuningdek Big Data Analyticsning ahamiyatini aniqlashga mo'ljallangan.

**Kalit so'zlar:** Big data, Data analytics, Hadoop, Kletkali avtomatlar

### **KIRISH**

Ishlab chiqarilgan raqamli ma'lumotlar qisman Internetga ulangan qurilmalardan foydalanish natijasidir. Shunday qilib, smartfonlar, planshetlar va kompyuterlar o'z foydalanuvchilari haqidagi ma'lumotlarni uzatadilar. Ulangan aqlli ob'ektlar iste'molchining kundalik narsalardan foydalanishi haqida ma'lumot beradi.

Ulangan qurilmalardan tashqari ma'lumotlar turli manbalardan olinadi: demografik ma'lumotlar, iqlim ma'lumotlari, ilmiy va tibbiy ma'lumotlar, energiya iste'moli ma'lumotlari va boshqalar.

Bu ma'lumotlarning barchasi qurilmalar foydalanuvchilarining joylashuvi, ularning sayohatlari, ularning sayohatlari haqida ma'lumot beradi. qiziqishlari, ularning iste'mol qilish odatlari, bo'sh vaqtlari va loyihalari va boshqalar. Bundan tashqari, infratuzilma, mashina va apparatlardan qanday foydalanilganligi haqida ma'lumot. Internet va mobil telefon foydalanuvchilari sonining tobora ortib borishi bilan raqamli ma'lumotlar hajmi tez sur'atlar bilan o'sib bormoqda. Bugun biz Axborot

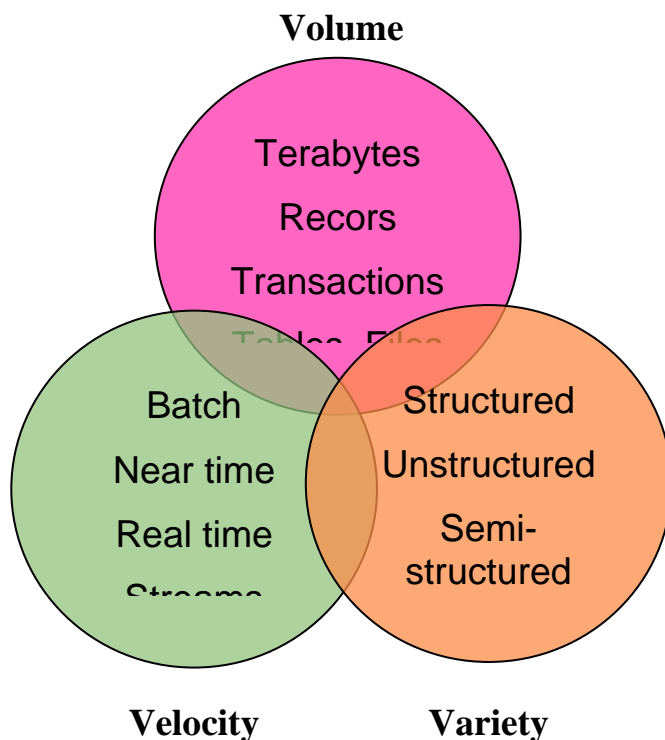
Jamiyatida yashayapmiz va bilimga asoslangan jamiyat sari intilyapmiz. Yaxshiroq bilim olish uchun bizga katta hajmdagi ma'lumotlar kerak bo'ladi. Axborot jamiyati - bu axborot iqtisodiy, madaniy va siyosiy bosqichda katta rol o'ynaydigan jamiyat. [1]

Katta hajmdagi ma'lumotlarni qayta ishlashni avtomatlar teoriyasi (automata theory) yordamida amalga oshirish qulay bo'ladi. Kletkali avtomatlar (yoki uyali avtomatlar) g'oyasi 1966 yilda fon Neumann tomonidan kiritilgan. Ba'zi tadqiqotchilar kletkali avtomatlarning ko'p jihatlarini o'rganishdi. 1D (bir o'lchovli) kletkali avtomatlar hozirgi kunga qadar turli jihatlarida o'rganilgan. Hozirgi vaqtda 2D (ikki o'lchovli) kletkali avtomatlar ishi uchun ko'p harakatlar qilindi. 1D va 2D kletkali avtomatlar foydali kontsepsiyalari tufayli fanning ko'plab sohalarida qo'llanilgan. Kletkali avtomatlar tushunchalari tabiatdagi ko'plab turli muammolarni o'rganishni osonlashtirdi. Misol uchun kompyuter fanlari, fizika, biologiya, kimyo, matematika, ijtimoiy fanlar va muhandislik kabi turli xil ilmiy sohalarda qo'llaniladigan katta miqdordagi kletkali avtomatlar ilovalari. [2]

“Katta ma'lumotlar” atamasi jamiyatimizda uzoq vaqt davomida eksponent ravishda o'sib borayotgan ma'lumotlar massasidan to'g'ri foydalanuvchini kerakli vaqtda kerakli ma'lumot bilan ta'minlaydigan texnologiyalarning rivojlanishi va ishlatilishini anglatadi. Muammo nafaqat tez sur'atlar bilan o'sib borayotgan ma'lumotlar hajmini engish, balki tobora ko'payib borayotgan heterojen formatlarni, shuningdek, tobora murakkab va o'zaro bog'liq bo'lgan ma'lumotlarni boshqarish qiyinligi.

Murakkab polimorf ob'ekt bo'lib, uning ta'rifi foydalanuvchi yoki xizmat ko'rsatuvchi provayder sifatida undan manfaatdor bo'lgan jamoalarga qarab farqlanadi. Internet gigantlari tomonidan ixtiro qilingan Big Data o'zini har bir kishiga ulkan ma'lumotlar bazalariga real vaqt rejimida kirishni ta'minlash uchun mo'ljallangan yechim sifatida taqdim etadi. Katta ma'lumotlar - bu aniq ta'riflash juda qiyin tushuncha, chunki ma'lumotlar hajmi bo'yicha katta tushunchasi bir sohadan boshqasiga farq qiladi. U texnologiyalar majmuasi bilan belgilanmaydi, aksincha, texnika va texnologiyalar toifasini belgilaydi. Bu rivojlanayotgan soha va biz ushbu yangi paradigmani qanday amalga oshirish va qiymatdan foydalanishni o'rganishga intilayotganimiz sababli, ta'rif o'zgarimoqda. [3]

Katta ma'lumotlar atamasi ulkan kattaroq ma'lumotlar to'plamlariga (hajmga) ishora qiladi; ko'proq diversifikatsiyalangan, shu jumladan tuzilgan, yarim tuzilmali va tuzilmagan (turli) ma'lumotlar va oldingiga qaraganda tezroq (tezlik) yetib borishi. Bular 3 ta:



Rasm 1. 3V konsepti

- Volume: tizimda ishlab chiqarilgan, saqlanadigan va boshqariladigan ma'lumotlar miqdorini ifodalaydi. Hajmning o'sishi ishlab chiqarilgan va saqlanadigan ma'lumotlar hajmining ko'payishi bilan izohlanadi, balki ulardan foydalanish zarurati bilan izohlanadi.

- Variety: axborot tizimi tomonidan boshqariladigan ma'lumotlar turlarini ko'paytirishni ifodalaydi. Ushbu ko'payish ushbu ma'lumotlar orasidagi bog'lanishlar va bog'lanish turlarining murakkabligiga olib keladi. Xilma-xillik, shuningdek, xom ma'lumotlar bilan bog'liq mumkin bo'lgan foydalanish bilan bog'liq.

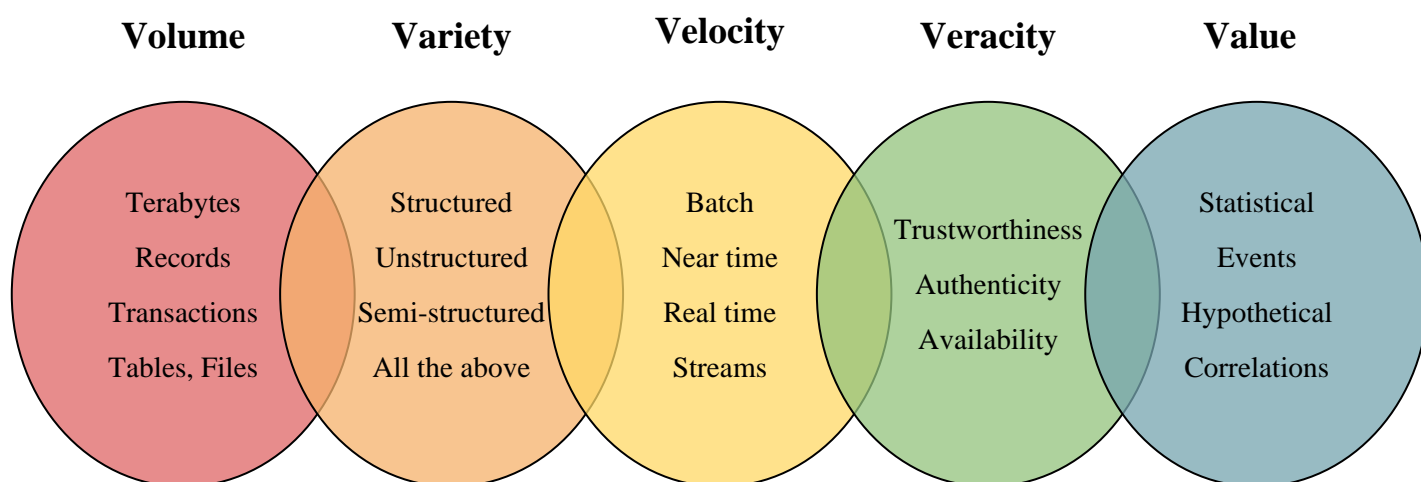
- Velocity: ma'lumotlarni yaratish, olish va almashish chastotasini ifodalaydi. Ma'lumotlar oqim orqali keladi va real vaqtda tahlil qilinishi kerak.

#### Rasm 2. 5V kosept

Ushbu klassik tavsif uchun yana ikkita "V" muhim:

- Veracity: ma'lumotlar va ma'lumotlar manbalarining sifat darajasi, aniqligi va noaniqligi.

- Value: ma'lumotlardan olingan qiymat va potentsial.



Katta ma'lumotlar odatda an'anaviy ma'lumotlar bazalari va ma'lumotlarni tahlil qilish usullarining odatiy saqlash, qayta ishlash va hisoblash imkoniyatlaridan oshib ketadigan ma'lumotlarni anglatadi. Resurs sifatida Big Data keng miqyosli ma'lumotlardan naqshlarni tahlil qilish va olish uchun qo'llanilishi mumkin bo'lgan vositalar va usullarni talab qiladi. [4] Strukturaviy ma'lumotlar tahlili manipulyatsiya qilingan ma'lumotlarning xilma-xilligi va tezligi tufayli rivojlanadi. Shuning uchun, endi ma'lumotlarni tahlil qilish va hisobotlarni ishlab chiqarishning o'zi etarli emas, ma'lumotlarning xilma-xilligi mavjud tizimlar ma'lumotlarni tahlil qilishda yordam berishga qodir bo'lishi kerakligini anglatadi. Tahlil turli xil tez o'zgaruvchan ma'lumotlar ichida ulardan foydalanishga yordam berish uchun ma'lumotlar o'rtasidagi bog'liqlikni avtomatik ravishda aniqlashdan iborat.

Big Data Analytics turli naqshlarni va boshqa foydali ma'lumotlarni topish uchun katta ma'lumotlar to'plamlarini to'plash, tartibga solish, tahlil qilish jarayonini anglatadi. Katta ma'lumotlar tahlili - bu odatdagidan farq qiladigan, murakkabroq va katta hajmdagi katta ma'lumotlar to'plamlaridan katta yashirin qiymatlarni ochish uchun integratsiyaning yangi shakllarini talab qiladigan texnologiyalar va texnikalar to'plami. U asosan yangi muammolarni yoki eski muammolarni yaxshiroq va samarali usullar bilan hal qilishga qaratilgan. [5]

Katta ma'lumotlarni tahlil qilish turlari:

a) tavsifiy tahlil

U savol berishdan iborat: nima bo'lyapti?

Bu tarixiy ma'lumotlar to'plamini yaratadigan ma'lumotlarni qayta ishlashning dastlabki bosqichidir. Ma'lumotni qazib olish usullari ma'lumotlarni tartibga soladi va tushunishni taklif qiluvchi naqshlarni ochishga yordam beradi. Ta'riflovchi tahlil kelajakdagi ehtimolliklar va tendentsiyalarni ta'minlaydi va kelajakda nima bo'lishi mumkinligi haqida fikr beradi.

b) Diagnostik tahlillar

Bu savolni berishdan iborat: Nima uchun bu sodir bo'ldi?

Diagnostik tahlil muammoning asosiy sababini izlaydi. Nima uchun sodir bo'lganligini aniqlash uchun ishlatiladi. Bu tip hodisalar va xatti-harakatlarning sabablarini topishga va tushunishga harakat qiladi.

c) Bashoratli tahlillar

Bu savolni berishdan iborat: nima bo'lishi mumkin?

U kelajakni bashorat qilish uchun o'tmishdagi ma'lumotlardan foydalanadi. Hammasi bashorat qilish bilan bog'liq. Bashoratli tahlil joriy ma'lumotlarni tahlil qilish va sodir bo'lishi mumkin bo'lgan stsenariylarni yaratish uchun ma'lumotlarni qazib olish va sun'iy intellekt kabi ko'plab usullardan foydalanadi.

d) Reseptiv tahlil

U savol berishdan iborat: nima qilish kerak?

U amalga oshirilishi kerak bo'lgan to'g'ri harakatni topishga bag'ishlangan. Ta'riflovchi tahlil tarixiy ma'lumotlarni taqdim etadi va bashoratli tahlil nima bo'lishi mumkinligini bashorat qilishga yordam beradi. Reseptiv tahlil eng yaxshi yechimni topish uchun ushbu parametrlardan foydalanadi.

Katta ma'lumotlar - bu o'n yil oldin gigant hisoblangan, saqlash va qayta ishlash imkonsiz bo'lgan ma'lumotlar to'plami. Bunday katta hajmdagi ma'lumotlarni qayta ishlash muayyan usullarni talab qiladi. Klassik ma'lumotlar bazasini boshqarish tizimi juda ko'p ma'lumotni qayta ishlay olmaydi. Hadoop ochiq kodli dasturiy ta'minot mahsuloti (yoki aniqrog'i, "dasturiy kutubxona ramkasi") bo'lib, u Apache Foundation tomonidan hamkorlikda ishlab chiqariladi va erkin tarqatiladi - samarali tarzda, bu Big Data yechimlarini yaratishni soddalashtirish uchun ishlab chiquvchilarning asboblari to'plamidir. [6]

Hadoop qayta ishlash uchun juda katta hajmdagi ma'lumotlarga ega kompaniyalar tomonidan qo'llaniladi. Ular orasida Facebook, Twitter, LinkedIn, eBay va Amazon kabi veb-gigantlar bor. Hadoop - tarqatilgan ma'lumotlarni qayta ishlash

va boshqarish tizimi. U ko‘plab komponentlarni o‘z ichiga oladi, jumladan: HDFS, YARN, Map Reduce. HDFS - bu Hadoop klasterlari bo‘ylab ma’lumotlarga yuqori samarali kirishni ta’minlovchi taqsimlangan fayl tizimi. [7] MapReduce - bu Apache Hadoop dasturiy ta’minotining asosiy komponenti. Hadoop klasterning har bir tuguni o‘z xotirasini o‘z ichiga olgan tovar kompyuter klasterlari bo‘ylab katta tuzilmagan ma’lumotlar to‘plamlarini moslashuvchan, taqsimlangan qayta ishlash imkonini beradi. MapReduce ikkita muhim funktsiyani bajaradi: U klaster yoki xarita ichidagi turli tugunlarga ishni taqsimlaydi va har bir tugun natijalarini so‘rovga yaxlit javob sifatida tartibga soladi va kamaytiradi.[8]

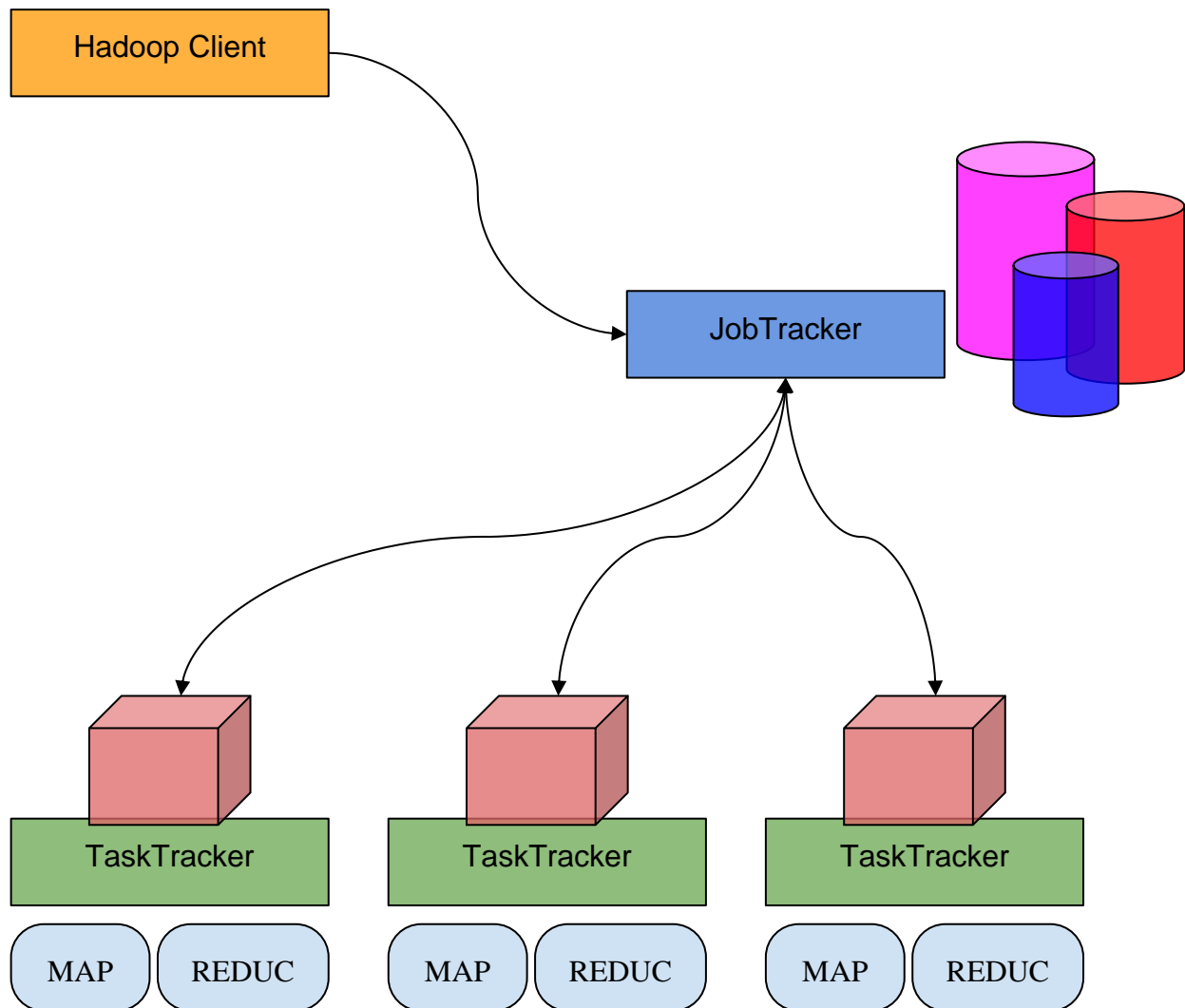
Hadoop ikkita serverga tayanadi:

JobTracker: har bir Hadoop klasterida faqat bitta JobTracker mavjud. U Xarita/Kamaytirish topshiriqlarini oladi va ularni klasterda bajarilishini tashkil qiladi. Hadoop klasterida bajarilishi uchun kodingizni yuborganingizda, JobTracker ijro rejasini tuzish uchun javobgardir. Ushbu ijro rejasiga ishlash uchun ma’lumotlarni o‘z ichiga olgan tugunlarni aniqlash, ma’lumotlarga mos keladigan tugunlarni tartibga solish, bajarilayotgan vazifalarni kuzatish va ular bajarilmasa, vazifalarni qayta ishga tushirish kiradi.[9]

TaskTracker: har bir klasterda bir nechta. Xarita/Kamaytirish ishining o‘zini bajaradi (bog‘langan kirish ma’lumotlari bilan Xarita va qisqartirish vazifasi sifatida). JobTracker serveri HDFS bilan aloqada; u Map/Reduce dasturining kirish ma’lumotlari qayerda ekanligini va chiqish ma’lumotlari qayerda saqlanishi kerakligini biladi. Shunday qilib, u tegishli ma’lumotlarga muvofiq vazifalarni taqsimlashni optimallashtirishi mumkin.

Xarita/Kamaytirish dasturini ishga tushirish uchun biz:

- Kirish ma’lumotlarini HDFS da yozish
- Dasturni klasterning JobTracker dasturiga yuboring.
- HDFS dan chiqish ma’lumotlarini olish.



Rasm 3. Hadoop arxitekturasi

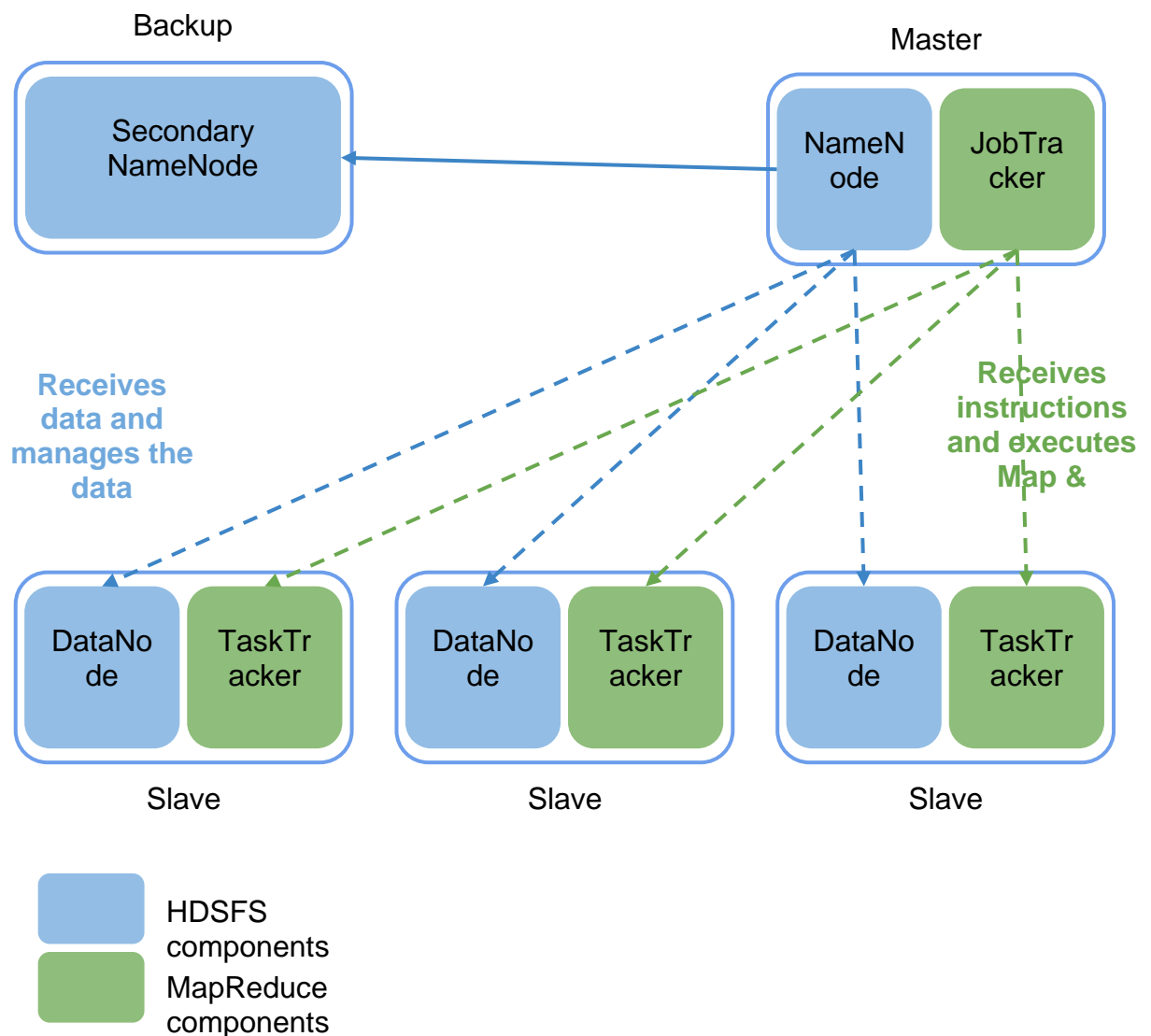
• Barcha TaskTrackers yurak urish paketlari orqali o'z holatini doimiy ravishda xabar qiladi. Agar TaskTracker bajarilmasa (yurak urishi yoki bajarilmagan vazifa), JobTracker vazifaning boshqa tugunga qayta taqsimlanishi haqida xabar beradi. HDFS ikkita serverga tayanadi:

• NameNode: klasterda yagona. U fayl nomlari va ularning xarakteristikallari haqidagi ma'lumotlarni saqlaydi. Bu slave DataNode-ni boshqaradigan HDFS ustasi.

• Ikkilamchi NameNode: Ikkilamchi NameNode HDFS klasterining holatini nazorat qiladi va NameNode tarkibidagi ma'lumotlarning "oniy tasvirlarini" oladi.

Agar NameNode muvaffaqiyatsiz bo'lsa, NameNode o'rniga ikkinchi darajali NameNode ishlatilishi mumkin.[10]

- DataNode: klaster bo'yicha bir nechta. Fayllarning mazmunini bloklarga bo'lingan holda saqlaydi (sukut bo'yicha 64 KB)



Rasm 4. Umumiy arxitektura

### XULOSA

Katta ma'lumotlar shaxsiy yoki professional maqsadlarda yangi texnologiyalardan foydalanish natijasida olingan raqamli ma'lumotlar to'plamini anglatadi. Katta ma'lumotlar tahlili - bu to'g'ri qaror qabul qilish uchun yashirin modellarni, bozor tendentsiyalarini, mijozlarning afzalliklarini va boshqa foydali



ma'lumotlarni ochish uchun ushbu ma'lumotlarni o'rganish jarayoni. Big Data Analytics - tez rivojlanayotgan texnologiya. U eng kutilmagan tarmoqlar tomonidan qabul qilindi va o'z-o'zidan sanoatga aylandi. Ammo bu ma'lumotlarni Big Data doirasida tahlil qilish ba'zida juda intruziv ko'rinadigan jarayondir.

Analitika - bu ma'lumotlar fanidir. BI qaror qabul qilish qismi bilan shug'ullanadi, Data Analytics esa savollar berish jarayonidir. Analitik vositalar kompaniya prognoz qilishlari kerak bo'lganda va kelajakda nima bo'lishini bilishni istaganda ishlatiladi, BI vositalari esa bu prognozlarni umumiy tilga aylantirishga yordam beradi. Ko'pincha Big Data Business Intelligence vorisi sifatida qabul qilinadi. Bu taqqoslash keyingi ishda muhokama qilinadi.

### **FOYDALANILGAN ADABIYOTLAR RO'YXATI: (REFERENCES)**

- [1] Perspectives on Big Data and Big Data Analytics-Database Systems Journal vol. III, no. 4/2012
- [2] Von Neumann J., The theory of self-reproducing automata, (Edited by A. W. Burks), Univ. of Illinois Press, Urbana, (1966).
- [3] The Big Data Revolution, Issues and Applications, Azzeddine Riahi, Sara Riahi-IJARCSSE, Volume 5, Issue 8
- [4] Deep learning applications and challenges in big data analytics-Najafabadi et al. Journal of Big Data (2015) 2:1 DOI 10.1186/s40537-014-0007-7
- [5] BIG DATA ANALYTICS: CHALLENGES AND APPLICATIONS FOR TEXT, AUDIO, VIDEO, AND SOCIAL MEDIA DATA-International Journal on Soft Computing, Artificial Intelligence and Applications (IJSCAI), Vol.5, No.1, February 2016
- [6] Big Data- The definitive guide to the revolution in business analytics-Fujitsu
- [7] <http://searchbusinessanalytics.techtarget.com/definition/Hadoop-Distributed-File-System-HDFS>
- [8] Choudhury, P.P., Sahoo, S., Hassan, S. S., Basu, S., Ghosh, D., Kar, D., Ghosh, Ab., Ghosh, Av., Ghosh A.K., Classification of cellular automata rules based on their properties, Int. J. of Comp. Cogn. 8, (2010), p. 50-54.
- [9] Wolfram S., Rev. Mod. Phys. 55 (3) (1983) 601-644. doi/10.1103/RevModPhys.55.601.
- [10] Akin H., Siap I., Uguz S., Structure of 2-dimensional hexagonal cellular automata, AIP Conf. Proceed., Volume 1309, (2010) p. 16-26. DOI: 10.1063/1.3525111