MAPREDUCE MODELI VA UNING MAPPER, REDUCER, COMBINER KOMPONENTLARI

Madg‘oziyev Farrux Farhodjon o‘g‘li
Farg‘ona davlat texnika universiteti, Telekommunikatsiya yo‘nalishi, 630-23 guruh talabasi
madgoziyevfarrux@gmail.com
Dalibekov Lochinbek
Ilmiy rahbar, Farg‘ona davlat texnika universiteti, Telekommunikatsiya va kiberxavfsizlik kafedrasi o‘qituvchisi
dalibekov86@gmail.com

Annotatsiya: Mazkur tezisda katta hajmdagi ma’lumotlarni taqsimlangan muhitda qayta ishlash uchun mo‘ljallangan MapReduce dasturlash modeli, uning ishlash tamoyili hamda Mapper, Reducer va Combiner komponentlarining vazifalari yoritilgan. Combiner yordamida tarmoq trafigini kamaytirish imkoniyatlari tahlil qilingan.

Kalit so‘zlar: MapReduce, Mapper, Reducer, Combiner, Hadoop, Big Data, taqsimlangan hisoblash, Shuffle.

Kirish

Bugungi kunda ma’lumotlar hajmi tez o‘sib bormoqda va ularni bitta kompyuterda qayta ishlash samarasiz bo‘lib qolmoqda. Bu muammoni hal qilish uchun 2004-yilda Google mutaxassislari J. Dean va S. Ghemawat MapReduce modelini taklif qilishdi [1]. Model ma’lumotlarni klasterdagi ko‘plab tugunlarda parallel qayta ishlash imkonini beradi, dasturchi esa parallellashtirish, nosozliklarga chidamlilik va ma’lumotlarni taqsimlash kabi murakkab jarayonlar bilan shug‘ullanmaydi. Model Apache Hadoop platformasida ochiq kodli tarzda amalga oshirilgan [2].

Asosiy qism

MapReduce modelida ma’lumotlar **(kalit, qiymat)** juftliklari ko‘rinishida ifodalanadi va qayta ishlash uch bosqichdan iborat: Map, Shuffle & Sort hamda Reduce.

Mapper kirish ma’lumotlarining har bir yozuvini alohida qayta ishlaydi va oraliq (kalit, qiymat) juftliklarini hosil qiladi: *map(k1, v1) → list(k2, v2)*. Kirish ma’lumotlari

bo‘laklarga (split) ajratiladi va har bir bo‘lak alohida Mapper tomonidan, odatda ma’lumot saqlanayotgan tugunning o‘zida qayta ishlanadi. Bu tarmoq yuklamasini kamaytiradi.

Shuffle & Sort bosqichida Mapperlar chiqargan oraliq natijalar kalit bo‘yicha guruhlanadi, saralanadi va tegishli Reducerlarga uzatiladi. Aynan shu bosqich tarmoq resurslarini eng ko‘p talab qiladi.

Reducer bir xil kalitga ega barcha qiymatlarni qabul qilib, ularni jamlaydi yoki boshqa usulda birlashtiradi: *reduce(k2, list(v2)) → list(k3, v3)*. Natija odatda taqsimlangan fayl tizimiga (HDFS) yoziladi.

Combiner — Mapper chiqishida ishlaydigan ixtiyoriy “mini-Reducer” bo‘lib, har bir tugun ichida oraliq natijalarni dastlabki jamlaydi. Masalan, so‘zlarni sanash masalasida Mapper (“data”, 1) juftligini 1000 marta chiqarsa, Combiner ularni (“data”, 1000) ga birlashtiradi. Natijada Shuffle bosqichida uzatiladigan ma’lumot hajmi sezilarli kamayadi. Combiner faqat assotsiativ va kommutativ amallar (yig‘indi, maksimum, minimum) uchun to‘g‘ri ishlaydi, o‘rtacha qiymatni hisoblashda esa to‘g‘ridan-to‘g‘ri qo‘llab bo‘lmaydi. Framework Combinerni bir necha marta chaqirishi yoki umuman chaqirmasligi mumkin, shuning uchun algoritm natijasi unga bog‘liq bo‘lmasligi kerak [3].

Afzalliklari: gorizontal kengayish imkoniyati, nosozliklarga chidamlilik (vazifa boshqa tugunda qayta bajariladi), dasturlash soddaligi. **Kamchiliklari:** oraliq natijalarni diskka yozish sababli iterativ algoritmlar (masalan, mashinaviy o‘qitish) uchun sekinligi; shu sababli bunday vazifalarda Apache Spark kabi vositalar afzal ko‘riladi.

Xulosa

MapReduce katta hajmdagi ma’lumotlarni qayta ishlashning soddaligi va ishonchliligi bilan ajralib turuvchi fundamental modeldir. Mapper ma’lumotlarni dastlabki qayta ishlaydi, Reducer yakuniy jamlashni bajaradi, Combiner esa tarmoq trafigini kamaytirib, umumiy samaradorlikni oshiradi. Uch komponentning to‘g‘ri loyihalanishi tizim unumdorligini belgilaydi.

Foydalanilgan adabiyotlar

1. Dean J., Ghemawat S. MapReduce: Simplified Data Processing on Large Clusters // OSDI’04, 2004.

2. White T. Hadoop: The Definitive Guide. — O’Reilly Media, 4th ed., 2015.

3. Lin J., Dyer C. Data-Intensive Text Processing with MapReduce. — Morgan & Claypool, 2010.

4. Apache Hadoop rasmiy hujjatlari: https://hadoop.apache.org/docs/

