**دانشگاه تبریز**

**دانشکده علوم ریاضی**

**گروه علوم کامپیوتر**

**پایان‌نامه برای دریافت درجه کارشناسی ارشد .M.Sc**

**رشته‌ی علوم کامپیوتر- گرایش سیستم‌های هوشمند**

عنوان:

# تکثیر داده در گرید داده‌ای با استفاده از الگوریتم رنگ آمیزی گراف

استاد راهنما:

**دکتر لیلی محمدخانلی**

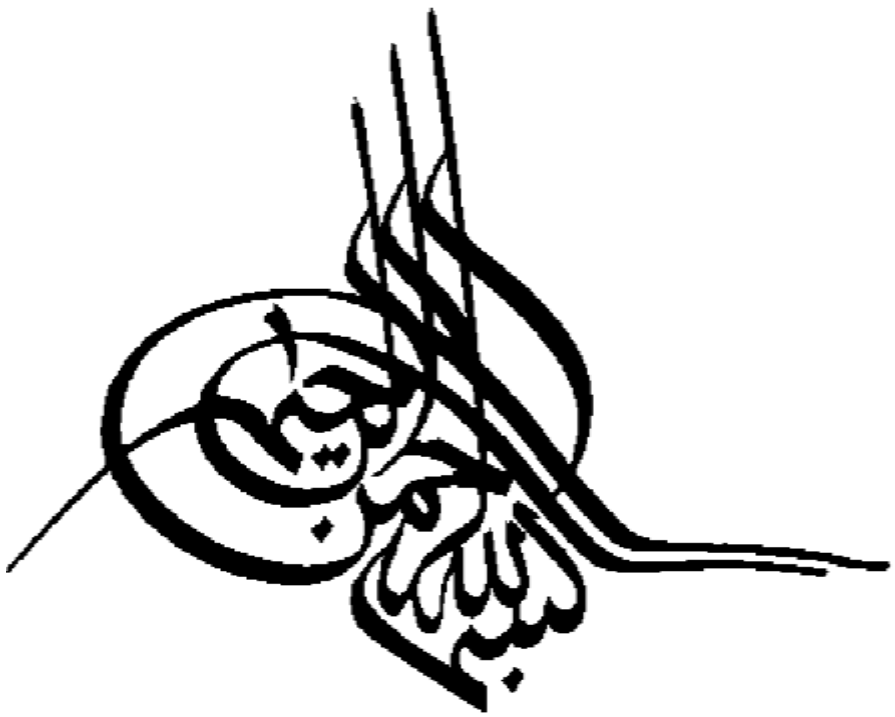
استاد مشاور:

**دکتر شهریار لطفی**

پژوهشگر:

**محمدرضا اکبری**

شهریور ۹۰



**نام خانوادگی**: اکبری **نام**: محمدرضا

**عنوان پایان نامه**: **تکثیر داده در گرید داده‌ای با استفاده از الگوریتم رنگ آمیزی گراف**

**استاد راهنما**: دکتر لیلی محمد خانلی

**استاد مشاور**: دکتر شهریار لطفی

**مقطع تحصیلی**: کارشناسی ارشد **رشته**: علوم کامپیوتر **گرایش**: سیستم‌های هوشمند **دانشگاه:** تبریز

**دانشکده**: علوم ریاضی **تارخ فارغ‌التحصیلی**: شهریور ۱۳۹۰ **تعداد صفحات**: ۷۲

**کلید واژه‌ها**: تکثیر داده، الگوریتم رنگ آمیزی گراف، الگوریتم ژنتیک، گرید داده‌ای

**چکیده**

گریدهای داده‌ای سیستم‌هایی هستند که به منظور به اشتراک گذاشتن کارآمد داده بین کاربران طراحی و پیاده‌سازی شده‌اند. کاربران این سیستم‌ها بدون آنکه از موقعیت جغرافیایی و مکانی این اطلاعات آگاه باشند، به اطلاعات دسترسی دارند. یکی از راه‌های افزایش کارآیی در گریدهای داده‌ای کپی کردن داده‌ها در قسمت‌های مختلف سیستم است. کپی کردن داده‌ها با توزیع بار شبکه، کاهش ازدهام و افزایش دسترسی پذیری باعث افزایش کارایی سیستم می‌شود.

ما برای اولین بار روش جدیدی برای تکثیرداده با استفاده از الگوریتم رنگ آمیزی گراف ارائه می‌نماییم که در آن یک گراف از داده‌های تکثیریافته مشابه از یک داده خاص داریم که می‌خواهیم آنها را در K گروه از فایل‌های تکثیر یافته در بخش‌های مجزا قرار دهیم طوری که دو داده مشابه تکثیر یافته از یک داده در یک بخش قرار نگیرند. دلیل اصلی استفاده کردن از الگوریتم رنگ‌آمیزی گراف برای حل این مسئله ایجاد زیر گراف‌های متعادل شده می‌باشد و راهکار پیاده سازی آن به دلیل غیر چند جمله‌ای سخت بودن مسئله استفاده از الگوریتم ژنتیک می‌باشد.

تقدیم به پدر و مادر عزیزم

به بهشت نمی‌روم اگر مادرم آنجا نباشد.

آمین

# فهرست مطالب صفحه

فصل اول: مقدمه

**۱-۱-مقدمه ..... ۲**

**۱-۲- اصطلاحات..... ۳**

**۱-۳- بیان مسئله..... ٤**

**۱-٤-راهکارهای گذشته..... ۱۲**

**۱-٥-اهداف پایان نامه..... ۱۳**

**۱-٦-نظریه ..... ۱٤**

**۱-۷- سازمان پایان نامه..... ۱٥**

فصل دوم:راهکارهای گذشته

**۲-۱-مقدمه..... ۱۷**

**۲-۲-مفاهیم بنیادی..... ۱۷**

**۲-۲-۱-اپراتورهای ژنتیکی..... ۱۸**

**۲-۲-۱-۱- نرخ برش..... ۱۹**

۲-۲-۱-۲- نرخ جهش..... ۲۰

۲-۲-۱-۳- تابع ارزیاب..... ۲۰

۲-۲-۱-٤- انتخاب فرضیه ها..... ۲۰

۲-۲-۱-٥- شرایط توقف الگوریتم..... ۲۱

۲-۳- شبکه عصبی مصنوعی..... ۲۳

۲-۳-۱-الگوریتم پس انتشار..... ۲٤

۲-٤-تکثیر پویا با استفاده از الگوریتمBRS ..... ۲٦

۲-٥-تکثیر با استفاده از شبکه عصبی ..... ۳۰

۲-٥-۱-انتخاب k امین همسایه نزدیک..... ۳۱

۲-٦-روش‌های کلی تکثیر در گرید داده‌ای سلسله مراتبی..... ۳۲

۲-۷-تکثیر با استفاده از PHFS..... ۳٥

۲-۸-نتیجه‌گیری..... ۳۷

فصل سوم: روش پیشنهادی

۳-۱-مقدمه..... ۳۹

۳-۲-الگوریتم پیشنهادی..... ٤۰

**۳-۲-۱-الگوریتم GBA.....۴۲**

**۳-۲-۲- الگوریتم تابع ارزیاب.....۴۳**

**۳-۲-۳-الگوریتم MRGC .....۴۴**

**۳-۲-۴- الگوریتم انتخاب چرخ رولت.....۴۸**

**۳-۲-۵-الگوریتم برش.....۴۹**

**۳-۳-نتیجه گیری.....۵۴**

فصل چهارم: ارزیابی

**۴-۱- ارزیابی.....۵۶**

**۴-۲-مراحل اجرا.....۵۹**

**۴-۳-ارزیابی کلی.....۶۵**

**۴-۴-نتیجه گیری.....۶۶**

فصل پنجم : نتیجه گیری

**نتیجه گیری.....۶۸**

**مراجع.....۷۰**

# فهرست شکل‌ها

صفحه

شکل ۲-۱-مثالی از جهش تک نقطه ای..... ۱۹

شکل ۲-۲-مکانیسم انتخاب بر اساس چرخ رولت..... ۲۰

شکل۲-۳- الگوریتم ژنتیک..... ۲۲

شکل ۲-٤- نمایش لایه ها در شبکه عصبی..... ۲۳

شکل۲-۵- تکثیربدون استفاده از روش سلسله مراتبی..... ۲۷

شکل ۲-٦-تکثیر با استفاده از BRS..... ۲۸

شکل ۳-۱- روند کلی کار..... ٤۱

شکل ۳-۲-الگوریتم پایه ژنتیک..... ٤۲

شکل۳-۲-۱- تابع ارزیاب ..... ٤۳

شکل۳-۳- الگوریتم MRGC ..... ٤۷

شکل۳-۲-۲- تابع انتخاب..... ٤۸

شکل ۳-۲-۳- تابع برش..... ۵٤

شکل ٤-١-مثالی از PHFS.....٥٦

شکل٤-٢-گراف ورودی..... ٦٠

شکل ٤-٣-مرحله اول اجرا.....٦٠

شکل ٤-٤-مرحله دوم اجرا.....٦١

شکل ٤-٥-مرحله سوم اجرا.....٦٢

شکل ٤-٦-مرحله چهارم اجرا.....٦٣

شکل ٤-٣-گراف پارتیشن شده و رنگ آمیزی شده به صورت معتبر..... ٦٤

شکل ٤-٨-راس‌های مجاور با رنگ مشابه..... ٦٥

## رسید موقت ثبت پایان نامه/رساله ( نسخه دانشجو )

**نام و نام خانوادگی دانشجو: محمدرضا اکبری**

**کد ملی:** 6259944251

**شماره دانشجویی:** 889361201

**مقطع تحصیلی: کارشناسی ارشد**

**عنوان پایان نامه: تکثیر داده در گرید داده ای با استفاده از الگوریتم رنگ آمیزی گراف**

**دانشگاه تبریز**

**دانشکده علوم ریاضی**

**استاد راهنما: لیلی محمد خانلی**

**کد رهگیری:** 2044440

**تاریخ ثبت :** 2/6/1390

این رسید جهت اطلاع دانشجو صادر گردیده و ارزش دیگری ندارد

فصل اول

# مقدمه

## ۱-۱-مقدمه

در این فصل مقدمه‌ای بر سیستم‌های توزیع شده وگرید داده‌ای را بیان می‌کنیم و در ادامه به شرح مسئله، بیان مفاهیم پایه‌ای در خصوص گرید داده‌ای می‌پردازیم و راهکارهای تکثیر داده در این خصوص را ارائه می‌نماییم. از مهمترین دلایل ایجاد سیستم‌های توزیع شده استفاده کارآمد از منابع سیستم‌های کامپیوتری می‌باشد و دلیل پیدایش این سیستم‌ها پیشرفت‌های بزرگ در زمینه‌های سخت افزاری و ارتباطات کامپیوتری می‌باشد.

در طول زمان، انواع مختلفی از سیستم‌های توزیع شده طراحی و پیاده سازی شده است. یکی از این انواع سیستم‌های توزیع شده سیستم‌های گرید است [۱]. به اشتراک گذاری منابع در گرید باید به نحوی باشد که برای کاربر دید شفافی[1] از یک سیستم کامپیوتری واحد[2] فراهم کند. به عبارت دیگر کاربر بدون توجه به مسائلی مسائلی مانند این که منبع مورد درخواست او کجا قرار دارد، به آن منبع دست پیدا کرده و از آن استفاده کند. این نوع دسترسی به منابع، بسیار شبیه به استفاده کاربران از نیروی برق است، به طوری که کاربر هنگام استفاده از نیروی برق دیگر توجهی به اینکه برقع مصرفی او در کدام نیروگاه تولید شده را ندارد. این شباهت تا اندازه‌ای است که نام این سیستم‌ها را از سیستم‌های توزیع برق گرفته‌اند.

اهمیت و مزیت اصلی این سیستم‌ها نسبت به سایر سیستم‌های توزیع شده آن است که مفهوم به اشتراک گذاری منابع را در حوزه بسیار وسیع و حتی جهانی مطرح می‌کنند [۲]. واضح است که در چنین حوزه وسیعی،

[1] transparent
[2] Single computer image

ناهمگونی منابع امری اجتناب ناپذیر خواهد بود. در واقع توسعه پذیری بالا[1] و ناهمگونی[2] زیاد منابع اشتراکی دو تفاوت اصلی سیستم‌های گرید با سایر سیستم‌های توزیع شده است.

چالش‌هایی که در سیستم گرید ممکن است با آنها مواجه شویم عبارتند از [۳]:

۱- داده‌ها باید بین علوم مختلف با منابع متفاوت و چندگانه اشتراک داده شوند.

۲- باید مطمئن شویم که داده‌ها در هر لحظه و در هر مکان در دسترس هستند.

۳- باید از نظر مقیاس پذیری و قابلیت اطمینان مطمئن باشیم.

٤- سیستم باید سیاست‌های مختلف دسترسی را پشتیبانی کند.

۵- باید امنیت داده‌ها را تضمین کند.

### ۱-۲- اصطلاحات

گرید داده‌ای: نوع خاصی از سیستم‌های گرید که برای انجام وظایف داده محور کابران و به اشتراک گذاشتن داده‌های حجیم و همکاری بین کاربران توسعه پیدا کرده‌اند.

گرید داده‌ای چند لایه[3]: نوع خاصی از گریدهای داده‌ای با معماری درختی است. در این معماری سرویس‌دهنده‌های داده‌ای در گره‌های میانی و کاربران در برگ‌ها قرار دارند.

دسترسی پذیری[4]: احتمال اینکه یک سیستم در یک لحظه زمانی درست کار کند. برای این منظور باید تمام قسمت‌ها در آن لحظه وظایف خود را به درستی انجام دهند و تمام منابع در دسترس باشند.

---

[1] Scalability
[2] Heterogeneity
[3] Multi tier data grid
[4] Availability

قابلیت اطمینان[1]: یعنی چقدر سیستم بدون خراب شدن می‌تواند پشت سر هم (بطورمداوم) کار کند.

تکثیر[2] : ایجاد و نگهداری کپی‌های یکسان از منابع سیستم در قسمت‌های مختلف سیستم. در این پایان‌نامه واژه کپی سازی مترادف با این مفهوم در نظر گرفته شده است.

الگوی درخواست کاربران[3]: مدلی از منابع داده درخواستی کاربر در طول زمان.

داده کاوی[4]: تکنیک‌هایی که برای کشف و استخراج دانش پنهان موجود در حجم انبوه داده‌ها به کار می‌رود.

تاخیر دسترسی[5]: مدت زمانی که طول می‌کشد کاربر به داده درخواستی خود دست پیدا کند.

فایل های محبوب[6]: فایل هایی که تعداد دسترسی به آنها در سیستم زیاد است.

**۱-۳- بیان مسئله**

امروزه حجم انبوهی از اطلاعات در آزمایشگاه‌های علمی و کاربردهای مهندسی تولید می‌شود. به منظور به اشتراک‌گذاری این داده‌ها وهمکاری بین دانشمندان، این داده‌ها باید در سطح شبکه‌های بزرگ توزیع شوند [۸,۷]. دلیل این امر رویکرد تازه‌ای است که امروزه در تحقیقات علمی دنبال می‌شود. در گذشته تحقیقات علمی در موسسات و دانشگاه‌های محلی و توسط دانشمندان آن موسسه یا دانشگاه انجام می‌شد، ولی امروزه به دلیل بزرگتر شدن آزمایش‌ها و پروژه‌های علمی و هزینه‌های هنگفت این پروژه‌ها، دیگر موسسات محلی قادر

[1] Reliability
[2] Replication
[3] User's access pattern
[4] Data mining
[5] Access Latency
[6] Popular files

به تامین این هزینه‌ها نیستند. این دلیل از یک طرف و از طرف دیگر نیاز به همکاری بیشتر بین دانشمندان در پروژه‌های بزرگ و پیچیده باعث شده که این پروژه‌ها در سطح بین المللی با مشارکت در هزینه‌ها و با همکاری دانشمندان کشورهای مختلف انجام شود. این امر، به اشتراک گذاری وتوزیع داده‌های تولید شده در این پروژه‌ها در شبکه‌های بزرگ را به موضوعی اجتناب ناپذیر تبدیل کرده است.

گرید داده‌ای مجموعه‌ای از صدها کامپیوتر توزیع شده جغرافیایی و منابع ذخیره‌ای که در بخش‌های مختلفی از جهان برای تسهیل اشتراک گذاشتن داده‌ها و منابع قرار گرفته را شامل می‌شود. اندازه داده‌هایی که برای دسترسی در گرید داده‌ای نیاز داریم امروزه بیشتر از ترابایت است و به زودی انتظار می‌رود که به پتابایت برسد. تضمین کارایی دسترسی به داده‌های توزیع شده بزرگ در طراحی گرید و شبکه یک مشکل جدی است. مشکل بزرگ دیگر پشتیبانی کردن از دسترسی سریع داده‌ها در گرید داده‌ای است که مقیاس پذیر بودن و تحمل شکست را در تمام سیستم گرید پیچیده می‌کند. تاخیرشبکه پوشش- وسیع[1] موضوع اصلی در گرید داده‌ای است که استراتژی برای دسترسی اطلاعات با کفایت را با دوره تاخیر زمانی پایین طراحی می‌کند. برای حل کردن مسئله، یک راهکار مناسب این است که کپی‌های محلی از داده در محل‌های مناسب ایجاد می‌کنند، سپس استفاده‌کننده محلی خیلی آسانتر و سریع‌تر به داده دسترسی خواهد داشت.

مسئله اصلی دیگر در گرید داده‌ای این است که چگونه و به چه نحو مدیریت موثر و بهینه‌ای را بر داده‌ها اعمال کنیم. از فاکتورهایی که در زمینه مدیریت موثر داده‌ها در گریدهای داده‌ای مورد تحقیق قرار می‌گیرند می‌توان به موارد زیر اشاره کرد [۹]: (به عبارت دیگر مدیریتی موثر است که باعث موارد زیر شود)

۱- افزایش کارایی

[1] wan

۲- افزایش دسترسی پذیری

۳- افزایش قابلیت اطمینان

٤- افزایش تحمل پذیری خطا

۵- توزیع مناسب بار

٦- افزایش توسعه پذیری

برای مدیریت موثر و بهینه داده‌ها در گریدهای داده‌ای، تکنیک‌ها و روش‌های مختلفی وجود دارد. یکی از این تکنیک‌ها تکرار داده (کپی‌سازی داده) [1] است. تکرار داده روشی است که در آن کپی‌های یکسانی از داده مشخص در قسمت‌های مختلف سیستم ایجاد می‌شود. تکرار داده که مورد بحث ما در این تحقیق نیز هست فقط یکی از انواع تکرار منابعی است که در سیستم‌های توزیع شده استفاده می‌شوند. از منابع دیگری که معمولا در سیستم‌های توزیع شده کپی‌سازی می‌شوند، می‌توان به موارد زیر اشاره کرد:

- منابع سخت افزاری
- سرویس‌ها و توابع
- سرویس دهنده‌ها

واضح است که ایجاد کپی‌های مختلف داده در سیستم باید مزایایی داشته باشد که تحمل هزینه اضافی برای این کار را توجیه کند. از مزایای کپی‌سازی داده می‌توان به موارد زیر اشاره کرد [۱۰]:

[1] Data replication

۱. ***افزایش دسترسی‌پذیری:*** وجود کپی‌های مختلف باعث می‌شود که اگر یک نمونه از منابع کپی شده نتوانست به در خواست کاربران پاسخ دهد منبع دیگر بتواند این کار را بکند. به عنوان مثال اگر یک نمونه از داده‌های کپی شده از بین رفت و نتوانست پاسخگوی نیاز کاربران باشد، نمونه دیگر کپی شده این کار را انجام دهد و در دسترس باشد. به عبارت خیلی ساده تر هر چه که تعداد کپی‌ها زیاد باشد، آن داده در دسترس‌تر است. کپی سازی مهمترین روش برای افزایش دسترسی‌پذیری است.

۲. ***افزایش قابلیت اطمینان***: هر یک از منابع داده ممکن است به دلایل متعددی از بین بروند، در این صورت اگر فقط یک نمونه از آن داده در سیستم موجود باشد می‌توان گفت که این از بین رفتن داده کارکرد کل سیستم را مختل خواهد کرد. با وجود کپی‌های مختلف از داده، در صورت از بین رفتن یک نمونه از آن، سایر نمونه‌ها می‌توانند در دسترس باشند و از مختل شدن کل سیستم جلوگیری کنند. به عبارت دیگر وجود کپی‌های مختلف از یک منبع احتمال اینکه همه آن منابع با هم و به طور همزمان از کار بیافتند کاهش می‌یابد و به تبع آن احتمال اینکه کل سیستم از کار افتد بسیار کاهش می‌یابد.

۳. ***افزایش کارایی***: وجود کپی‌های مختلف داده در سیستم باعث می‌شود تا با توزیع مناسب درخواست‌های داده‌ای کاربران بین سرویس دهنده‌های مختلف، از افزایش بیش از اندازه بار بر روی یک سرویس دهنده جلوگیری کنیم. زیرا با افزایش بیش از اندازه بار یک سرویس دهنده، زمان پاسخ‌دهی آن سرویس دهنده به درخواست‌های کاربران کاهش می‌یابد. این امر کارایی کل سیستم را تحت تاثیر قرار می‌دهد. فاکتور دیگری که کپی سازی داده با تاثیر گذاری بر آن بر کارایی کل سیستم تاثیر می‌گذارد، پهنای باند مصرفی است. یک منبع داده ممکن است از هر نقطه از سیستم توسط

کاربران درخواست شود. اگر فاصله کاربر درخواست کننده داده از منبع داده زیاد باشد، انتقال داده درخواست شده از منبع داده به کاربر باعث استفاده زیاد از پهنای باند می‌شود. این امر در سیستم‌هایی که با داده‌های با اندازه بزرگ سروکار دارند و در حوزه وسیع جغرافیایی توسعه یافته‌اند به مراتب مشکل آفرین‌تر خواهد بود. کپی کردن منابع داده در قسمت‌های مختلف سیستم، مخصوصاً قسمت‌هایی که تعداد درخواست‌های کاربران برای آن داده بسیار زیاد است، باعث کاهش پهنای باند مصرفی می‌شود. زیرا درخواست کاربر توسط سرویس دهنده‌ای در نزدیکی او اجابت خواهد شد و نیاز به انتقال داده از راه دور نیست.

کپی کردن داده‌ها علاوه بر موارد بالا باعث افزایش قدرت تحمل پذیری خطا و افزایش توسعه پذیری می-شود. در واقع تحمل پذیری خطا از افزایش قابلیت اطمینان سیستم ناشی می‌شود و افزایش توسعه پذیری نتیجه غیر مستقیم افزایش کارایی است.

٤. **کاهش پهنای باند مصرفی:** اگر فاصله کاربر درخواست کننده داده از منبع داده زیاد باشد، انتقال داده درخواست شده از منبع داده به کاربر باعث استفاده زیاد از پهنای باند می‌شود. این امر در سیستم‌هایی که با داده‌های با اندازه بزرگ سروکار دارند و در حوزه وسیع جغرافیایی توسعه یافته‌اند به مراتب مشکل آفرین خواهد بود. کپی کردن منابع داده در قسمت‌های مختلف سیستم، مخصوصاً قسمت‌هایی که تعداد درخواست‌های کاربران برای آن داده بسیار زیاد است، باعث کاهش پهنای باند مصرفی می‌شود. زیرا درخواست کاربر توسط سرویس دهنده‌ای در نزدیکی او اجابت خواهد شد و نیاز به انتقال داده از راه دور نیست.

۵. **افزایش قدرت تحمل پذیری خطا و افزایش توسعه پذیری:** تحمل پذیری خطا از افزایش قابلیت اطمینان سیستم ناشی می‌شود و افزایش توسعه پذیری نتیجه غیر مستقیم افزایش کارایی است.

درمقابل مزایای ذکرشده، روش کپی سازی معایبی هم دارد که استفاده از این تکنیک را محدودتر می‌کند. از معایب روش کپی سازی می‌توان به موارد زیر اشاره نمود:

۱. افزایش مصرف حافظه به دلیل تکرار در ذخیره سازی داده
۲. نیاز به بروز نگهداشتن[1] کپی‌های مختلف با هم اشاره کرد

وقتی در سیستم چندین کپی از یک منبع داده موجود باشد تغییر در یکی باید در تمام کپی‌ها منعکس شود تا این کپی‌ها با هم سازگار باشند. اگرتعداد کپی‌های یک منبع زیاد باشد مسئله بروز نگهداشتن این کپی‌ها در صورت تغییر یکی از آنها، به یکی از مسائل اصلی سیستم تبدیل می‌شود و باعث کاهش کارایی سیستم می‌شود.

ممکن است امروزه با تولید حافظه‌های پر ظرفیت، محدودیت حافظه فاکتور مهمی در بسیاری از کاربردها نباشد، ولی با توجه به این که در گریدهای داده‌ای با تعداد زیادی منبع داده‌ای پر حجم سر و کار داریم مسائل و محدودیت‌های مربوط به حافظه یکی ازمحدودیت‌های عمده به شمار می‌رود.

با توجه به گفته‌های بالا می‌توان نتیجه گرفت که ما نیاز به الگوریتم‌ها و تکنیک‌هایی داریم که با توجه به منابع موجود بهینه‌ترین پیکر بندی کپی‌ها را در سیستم مشخص کنند [۱۲,۱۱]. در واقع الگوریتم‌های مختلف

[1] Consistency

تکرار داده[1] سعی دارند تا مشخص کنند چه داده‌هایی، کی، به چه تعداد، کجاها باید کپی شود تا کارایی سیستم بهینه شود [۱۴,۱۳,۱۰]. در واقع این مسئله در سیستم‌های توزیع شده از نوع مسائل بسیار سخت[2] است، ولی الگوریتم‌های تکرار داده به دنبال یافتن بهینه ترین حالت نیستند، بلکه می‌خواهند حالتی نزدیک به بهینه را در سیستم یافته و ایجاد کنند.

الگوریتم‌ها و روش‌های کپی سازی داده را می‌توان به دو گروه ایستا و پویا تقسیم کرد.

۱. روش‌های کپی سازی ایستا[3][۱۵]: در این روش‌ها محل قرار گرفتن کپی‌ها در مرحله طراحی سیستم مشخص می‌شوند و در مرحله پیاده سازی، این کپی‌ها در محل‌های مشخص شده قرار می‌گیرند و تا آخر فعالیت سیستم محل آنها ثابت و تغییر ناپذیر است. این نوع الگوریتم‌ها بیشتر برای کاربردهایی مناسب است که الگوی در خواست کاربران تقریباً ثابت و از قبل تا حد زیادی قابل پیش بینی است. اگرچنانچه در این سیستم‌ها الگوی در خواست کاربران تغییر کند تنها راه ممکن برای سازگار کردن سیستم با این تغییرات، طراحی و پیاده سازی مجدد سیستم است. این مشکل بزرگترین عیب این روش‌ها است. بزرگترین مزیت این سیستم‌ها سادگی پیاده‌سازی آنهاست.

۲. روش‌های کپی سازی پویا[4][۱۸,۱۷,۱۶,۱۰,۷,۲,۱]: در سیستم‌هایی که الگوی درخواست کاربران ثابت نیست و در طول زمان تغییر می‌کند استفاده از روش‌های کپی‌سازی ایستا نه تنها کمک کننده نیست، بلکه ممکن است با تغییر الگوی درخواست کاربران پیکربندی فعلی به یکی از بدترین حالت‌ها تبدیل

[1] Data replication algorithm
[2] NP-hard
[3] Static replication
[4] Dynamic replication