





دانشگاه پیام نور

مرکز تهران

دانشکده فنی و مهندسی

گروه علمی مهندسی کامپیوتر و فناوری اطلاعات

پایان نامه برای دریافت درجه کارشناسی ارشد

در رشته مهندسی کامپیوتر - نرم افزار

عنوان پایان نامه:

الگوریتم تکرار داده آبخاری چند تایی بر اساس کمینه

سازی بار کاری در گرید داده

استاد راهنما: دکتر احمد فراهی

استاد مشاور: دکتر داود کریم زادگان

نگارش: کامران فرجی

زمستان ۱۳۹۰

تاریخ
شماره
پیوست



دانشگاه شهروخ سامنپور
دانشگاه پیام نور استان تهران



جمهوری اسلامی ایران
وزارت علوم، تحقیقات و فناوری

مرکز شمیرانات

تصویب نامه

پایان نامه کارشناسی ارشد رشته مهندسی کامپیوتر (نرم افزار)

تحت عنوان:

"الگوریتم تکرار داده آبخاری چند تایی بر اساس بهینه سازی بار کاری در گرید داده"

تاریخ دفاع: ۹۰/۱۱/۱۹ ساعت: ۱۴/۳۰ - ۱۳

نمره: ۱۸۱ درجه ارزشیابی: عالی

هیات داوران:

داوران	نام و نام خانوادگی	مرتبۀ علمی	امضاء
استاد راهنما	دکتر احمد فراهی	استادیار	
استاد مشاور	دکتر داود کریم زادگان مقدم	استادیار	
استاد داور	دکتر مهدی جوانمرد	استادیار	
نماینده گروه	دکتر محمد هادی معتمد		

تهران - بزرگراه ارتش - انستشهای
بسلوار شهید مسزدی (اوشان)
خیابان شهید پیروز شفیعی
خیابان باران - خیابان باران دوم
دانشگاه پیام نور مرکز شمیرانات
تلفن: ۲۲۱۹۵۳۰۲-۴
دورنگار: ۲۲۴۸۴۸۳۴
www.shemiranat.tpnu.ac.ir
shemiranat@tpnu.ac.ir

تقدیم به :

مادر و همسر عزیزم

تشکر و قدردانی

اول از خداوند بزرگ سپاسگزارم که همواره بهترین‌ها را برایم خواسته است. بر خود واجب می‌دانم که از زحمات بی‌دریغ، تلاش بی‌وقفه و راهنمایی‌های راه‌گشای استاد ارجمند جناب آقای دکتر فراهی در طول این پروژه تشکر و قدردانی به عمل آورم. همچنین از زحمات استاد محترم جناب آقای دکتر کریم زادگان که با راهنمایی‌های خود راه‌گشای اینجانب بوده‌اند، تشکر می‌نمایم. از همسر عزیزم، به خاطر تمام سختی‌هایی که در طول این تحقیق ناگزیر بر ایشان گذشت و بدون هرگونه شکوه‌ای همواره یاور و حامی من بوده‌اند، قدردانی می‌نمایم. امیدوارم بتوانم قدردان زحمات ایشان باشم.

در پایان از کلیه عزیزانی که به هر شیوه ممکن در انجام این پروژه کمک و یاری رسانده‌اند، تشکر می‌نمایم.

کامران فرجی

چکیده

با توجه به نیاز روز افزون برای دسترسی به داده‌های غیر متمرکز^۱ و طیف وسیع حجم آن‌ها در فواصل دور و نزدیک نوع خاصی از گرید محاسباتی بنام گرید داده مورد مطالعه قرار گرفته است. گرید داده با نگرش توزیع شده ناگزیر به اعمال قواعدی در سیستم‌های توزیع شده برای مدیریت داده‌های خود است. یکی از مهم‌ترین مسائل چالش برانگیز در گرید داده، تکرار داده می‌باشد. در تکرار داده پارامترهای زمانی مانند زمان تکرار، محل تکرار و الگوریتمی مانند روش تکرار از اهمیت خاصی برخوردار هستند.

در این رساله گرید داده، الزامات مدیریت داده‌ای آن و نمونه‌هایی از روش‌های تکرار و تفاوت بین آن‌ها در کارایی ارائه شده است. سپس به بیان یک الگوریتم جدید بنام الگوریتم تکرار داده آبشاری چند تایی بر اساس کمینه سازی بار کاری در گرید داده پرداخته شده است. تمرکز اصلی الگوریتم ارائه شده به روی روش تکرار و محل تکرار بر اساس بار کاری گرید داده در زمان مورد نظر است. بر اساس فرضیه‌ها و تعاریف انجام شده شبه کد^۲ الگوریتم مورد نظر به همراه توضیحات و یک مثال ساده ارائه شده است. همچنین الگوریتم مورد نظر در یک شبیه ساز^۳ با دو الگوریتم پایه‌ای دیگر بنام‌های "بهترین سرویس گیرنده" و "بدون تکرار" مقایسه شده است. عوامل مقایسه در بین سه الگوریتم مورد نظر عبارتند از میزان زمان پاسخ و میزان استفاده از حافظه در گرید داده. شبیه ساز برای الگوهای تصادفی در یک گرید داده با در نظر گرفتن فاصله بین سایت‌ها، الگوی دسترسی، اندازه فایل‌های مورد انتقال، اجرا و نتایج حاصل از آن طبق جداول و نمودارهای مرتبط مقایسه شده است. نتایج حاصل از بررسی الگوریتم پیشنهادی با دو الگوریتم مورد نظر به صورت میانگین بهبود نسبی را نسبت به دو الگوریتم دیگر در دو حوزه میزان زمان پاسخ و میزان استفاده از حافظه را نشان می‌دهد.

کلمات کلیدی

گرید محاسباتی، گرید داده، تکرار داده، بار کاری، تکرار آبشاری چند تایی، کمینه سازی.

¹ Non Central Data Access

² Pseudo Code

³ Simulator

- ۱۹ ۵-۲-۲ پردازش تراکنش
- ۱۹ ۶-۲-۲ همگام سازی داده
- ۲۱ ۷-۲-۲ تصدیق، کنترل دسترسی، حسابرسی
- ۲۲ ۸-۲-۲ مدیریت دسترسی داده و مدیریت ذخیره سازها
- ۲۳ ۹-۲-۲ یکپارچگی داده
- ۲۳ ۳-۲-۲ سرویس‌های فراداده در گرید داده
- ۲۴ ۱-۳-۲ انواع فراداده
- ۲۵ ۲-۳-۲ سرویس‌های فراداده
- ۲۶ ۴-۲ تکرار داده
- ۲۸ ۵-۲ انتقال موثر داده
- ۳۰ ۶-۲ جمع بندی
- ۳- پیشینه تحقیق بررسی چند استراتژی تکرار ۳۳**
- ۳۴ ۱-۳ بدون تکرار
- ۳۴ ۲-۳ بهترین سرویس گیرنده
- ۳۵ ۳-۳ تکرار آبخاری
- ۳۶ ۴-۳ نهانگاه ساده
- ۳۶ ۵-۳ نهانگاه و تکرار آبخاری
- ۳۷ ۶-۳ انتشار سریع
- ۳۷ ۷-۳ تکرار پویا بر اساس محبوبیت
- ۴۰ ۸-۳ جمع بندی
- ۴- الگوریتم پیشنهادی ۴۳**

۴۵	۱-۴ فرضیه‌ها و تعاریف
۴۶	۲-۴ نشان گذاری های مورد استفاده
۴۸	۳-۴ الگوریتم تکرار داده آبخاری چند تایی بر اساس کمینه سازی بار کاری در گرید داده .
۵۶	۱-۳-۴ بیان نحوه اجرای الگوریتم پیشنهادی توسط یک مثال
۶۲	۵-ارزیابی الگوریتم پیشنهادی
۶۲	۱-۵ تعاریف پیاده سازی و محیط شبیه سازی
۶۴	۲-۵ متوسط زمان پاسخ
۶۸	۱-۲-۵ مقایسه زمان پاسخ
۶۹	۳-۵ میزان حافظه مصرفی
۷۳	۴-۵ جمع بندی
۷۵	۶-جمع بندی و پیشنهادها
۷۵	۱-۶ فعالیت‌های انجام شده در تحقیق
۷۶	۲-۶ یافته‌های تحقیق
۷۶	۳-۶ پیشنهادها
۷۸	۷-منابع و مراجع
۸۱	۸-واژه‌نامه انگلیسی به فارسی
۸۶	۹-واژه نامه فارسی به انگلیسی

فهرست جداول

- جدول ۱-۴ نحوه محاسبه پیدا کردن گره ای با بیشترین هزینه در گرید داده شکل ۱-۴ ۵۷
- جدول ۲-۴ اندازه فایل‌های مورد ارجاع در مثال دوم ۵۹
- جدول ۳-۴ نحوه محاسبه پیدا کردن گره ای با بیشترین هزینه در گرید داده شکل ۲-۴ ۵۹
- جدول ۱-۵ میزان زمان پاسخ در شبیه سازی گرید داده ۶۵
- جدول ۲-۵: نتایج شبیه سازی در خصوص میزان استفاده از حافظه در یک گرید داده ۷۱

فهرست اشکال

- شکل ۳-۱. یک مثال از روش بهترین سرویس گیرنده..... ۳۵
- شکل ۳-۲. یک مثال از روش بهترین سرویس گیرنده ۳۶
- شکل ۳-۳. یک مثال از روش انتشار سریع ۳۷
- شکل ۳-۴. جمع تعداد دسترس‌یها برای یک فایل مشخص ۳۸
- شکل ۴-۱: یک مثال از نمایش لحظه‌ای گرید داده بدون در نظر گرفتن سایز فایلها ۵۶
- شکل ۴-۲: یک مثال از نمایش لحظه‌ای گرید داده با در نظر گرفتن سایز فایلها ۵۹
- شکل ۵-۱: نمودار مقایسه زمان پاسخ در گرید داده به عمق پنج برای سه الگوریتم مورد مقایسه ۶۷
- شکل ۵-۲: نمودار مقایسه زمان پاسخ در گرید داده به عمق شش برای سه الگوریتم مورد مقایسه ۶۸
- شکل ۵-۳: نمودار مقایسه زمان پاسخ در گرید داده به عمق هفت برای سه الگوریتم مورد مقایسه ۶۸
- شکل ۵-۴: نمودار مقایسه حافظه مصرفی در گرید داده به عمق پنج برای دو الگوریتم مورد مقایسه ۷۲
- شکل ۵-۵: نمودار مقایسه حافظه مصرفی در گرید داده به عمق شش برای دو الگوریتم مورد مقایسه ۷۳
- شکل ۵-۶: نمودار مقایسه حافظه مصرفی در گرید داده به عمق هفت برای دو الگوریتم مورد مقایسه ۷۳

فهرست علائم اختصاری

ACL	Access Control List
BC	Best Client
CR	Cascading Replication (CR)
DAM	Data Access Management
DBMS	Data Base Management System
DD	Data Discovery
DDMS	Distributed Database Management System (DDBMS)
DFS	Distributed File System
DG	Data Grid
DGW	Data Grid WorkLoad
DNA	Deoxyribonucleic Acid
DO	Date Object
DR	Data Replication
FS	Fast Spread
FTP	File Transfer protocol
GC	Grid Computing
GFTP	Grid File Transfer Protocol
GSI	Grid Security Interface
HTC	High Throughput Computing
MC	Multi Cascading
MCRBW	Multi Cascading Replication Based On Minimizing _ Work Load In Data Grid
NR	No replication
RLI	Replica Location Index
RMS	Replication Management Service
RSS	Replica Selection Service
SQL	Structural Query Language
SRB	Storage Resource Broker
VO	Virtual Organization

فصل اول

مقدمه

۱- مقدمه

مقدمه

یک گزید داده عبارت است از مجموعه‌ای از داده‌ها به صورت توزیع شده^۱ که دسترسی به داده‌ها، هم‌زمانی^۲، و هماهنگی بین داده‌های توزیع شده^۳ را مدیریت می‌نماید. از تفاوت‌های اصلی بین گزید داده و مدیر پایگاه داده توزیع شده وجود یک محیط کاملاً غیریکنواخت و ناهمگن^۴ در سکوی ارتباطی گزید داده می‌باشد. همچنین کنترل عملیات مرتبط به داده‌ها در گزید داده بر خلاف مدیر پایگاه داده^۵ توزیع شده، به صورت غیر متمرکز انجام می‌پذیرد. تکرار داده^۶، همگام سازی داده، یکپارچگی داده، بار کاری گزید داده^۷ و انتقال موثر داده از مباحث اصلی مرتبط با گزید داده می‌باشد. ایده اصلی تکرار در گزید داده با استفاده از حافظه نهان^۸ در سمت سرویس گیرنده‌ها ناشی شده است. یک تکرار داده در سمت سرویس گیرنده برای بالا بردن کارایی و یا دسترسی پذیری مؤثرتر برای رسیدن به اهداف تعیین شده در محیط گزید داده می‌باشد.

¹ Distributed

² Synchronization

³ Distributed data

⁴ Heterogeneous

⁵ Data Base Management System

⁶ Replication

⁷ Data Grid WorkLoad

⁸ Cache Memory

یکی از مباحث مهم در گرید داده تکرار داده در سناریوی اجرا می‌باشد مختصراً تکرار داده به معنای ایجاد یک کپی از داده اصلی و ذخیره سازی آن در یک گره یا نزدیک آن گره در محیط شبکه یا گرید داده در مکانی است که سربار انتقال اطلاعات و هزینه محاسبات به روی آن داده کاهش یابد. در فصل دوم خواهیم دید که گزینش سایت مقصد برای تکرار داده می‌تواند بر اساس پارامترهای مختلفی در سیستم ارزیابی شود.

در خصوص گزینش سایت مقصد توسط فرایند مدیر تکرار، الگوریتم‌های مختلفی ارائه شده است. هر کدام از این روش‌ها بر اساس پیش فرض‌ها و پارامترهای خاصی مانند تعداد درخواست‌های انجام شده برای یک موجودیت داده^۱، یا صرف نظر کردن از ترافیک شبکه، الگوریتم مورد نظر را ارائه داده‌اند. در خصوص این الگوریتم‌ها و روش انتخاب سایت مقصد توسط آن‌ها در فصول بعدی بحث خواهد شد.

با توجه به اینکه هدف اصلی از الگوریتم‌های مورد بحث در این رساله در تکرار داده، بالا بردن سرعت دستیابی یا کوتاه‌تر کردن زمان پاسخ می‌باشد، در این رساله دو شاخص مجموع زمان پاسخ و مجموع میزان استفاده از حافظه مورد بررسی قرار گرفته است.

جهت دستیابی به ساختار مناسب و منسجم برای تحقیق در ادامه فصل اصول و مبانی که در این تحقیق مورد نظر می‌باشد، ارائه می‌شود.

¹ Date Object

۱-۱ تعریف مسئله و سوالات اصلی تحقیق

گردد داده برای مدیریت داده‌ها در یک محیط توزیع شده و غیر همگن مورد استفاده قرار می‌گیرد و محیط پیاده سازی ارتباطی آن مانند یک سیستم مدیریت پایگاه داده توزیع شده، یک شبکه توزیع شده فیزیکی می‌باشد. در چنین محیط فیزیکی توزیع شده‌ای، یکی از مسایل مهم و دغدغه اصلی مدیریت داده در گزید داده، مدیریت تکرار داده می‌باشد. به صورت غیر رسمی تکرار داده دسترسی موثر تر با هزینه کمتر و یا حل مشکل گلوگاه در سرویس‌های درخواستی را در سناریوی درخواست‌های کاربر فراهم می‌کند. وظیفه اصلی مدیریت تکرار پیدا کردن محل تکرار بهینه یا کمینه از بین تعدادی از سایت‌های موجود می‌باشد. در این انتخاب باید سرعت دسترسی به داده‌ها، ممانعت از انتقال بیهوده و حجیم داده‌ها وضعیت فیزیکی و پارامترهای انتقال داده در شبکه مورد نظر مورد بررسی قرار گیرد.

الگوریتم مورد نظر به صورت خلاصه عبارت است از:

«روشی جدید جهت انتشار و تکرار داده در یک مسیر از گره تکرار کننده به یک یا چند گره تکرار شونده در گزید داده بر اساس بررسی و نقش پارامترهای میزان فاصله بین سایت‌ها یا پهنای باند، تعداد درخواست‌ها و اندازه فایل برای یک موجودیت داده‌ای در یک زمان مشخص با در نظر گرفتن کمینه سازی بارکاری کل گزید داده.»

بنابراین در پایان این تحقیق ما به سوالات زیر پاسخ خواهیم داد :

۱. محل تکرار بر اساس تعداد درخواست و فاصله بین دو سایت سرویس دهنده^۱ و سرویس گیرنده^۲ و همچنین یک ویژگی کلی برای کل گزید داده بنام بار کاری^۳ چگونه تعیین می‌شود؟

۲. آیا حجم موجودیت‌های داده‌ای در انتخاب سایت مقصد موثر می‌باشد؟

¹ Server

² Client

³ WorkLoad

۳. آیا فاصله یا پهنای باند شبکه فیزیکی و ارتباطی می‌تواند در تعیین سایت مقصد برای تکرار داده موثر باشد؟
۴. آیا تعداد درخواست‌های قبلی می‌تواند به عنوان برآوردی در درخواست‌های آتی کاربر باشد؟
۵. ارتباط بین تعداد درخواست‌ها، پهنای باند شبکه، حجم فایل درخواستی با تکرار داده در گرید داده چگونه ارزیابی می‌شود؟
۶. آیا این الگوریتم نسبت به دو الگوریتم پایه‌ای دیگر از کارایی مناسبی در دو حوزه زمان دسترسی و حافظه برخوردار هست؟

۲-۱ سوابق و ضرورت انجام تحقیق

محل تکرار و پیدا کردن محل مناسب برای تکرار داده یکی از جنبه‌های مهم و تحقیق برانگیز در مدیریت تکرار داده‌ها در گرید داده محسوب می‌شود. در این خصوص الگوریتم‌های زیادی برای تکرار داده ارائه شده است. تکرار در گرید داده به دو صورت پویا و ایستا انجام می‌شود. الگوریتم‌های ایستا در خصوص تغییرات درخواست کاربران و تغییرات محیطی واکنش مناسبی از خود نشان نمی‌دهند. برخی از روش‌های تکرار پویا عبارتند از:

روش بهترین سرویس گیرنده:^۱ یکی از روش‌ها و مکانیزم‌های پویا برای انجام عمل تکرار است که در آن هر گره یک پیشینه از تعداد درخواست‌های مورد نظر و گره درخواست دهنده را نگهداری می‌کند. در زمان‌های مشخص هر گره نسبت به بررسی تعداد درخواست‌ها بررسی و اگر تعداد درخواست‌ها از یک حد آستانه بزرگ‌تر بود بهترین سرویس گیرنده برای تکرار انتخاب می‌شود. بهترین سرویس گیرنده سرویس گیرنده‌ای است که بیشترین درخواست را برای فایل مورد نظر یا موجودیت داده‌ای مورد نظر در مجموعه داشته است (جورتمن^۲، ۱۹۹۵).

¹ Best client

² Gwertzman

تکرار آبخاری:^۱ یکی از روش‌ها و مکانیزم‌های پویا برای انجام عمل تکرار است. در این روش هنگامی که تعداد درخواست برای یک فایل یا موجودیت داده‌ای در ریشه (که به عنوان فایل سرور اصلی است) از یک حد آستانه بیشتر شد عمل تکرار داده به صورت آبخاری در مسیری از ریشه به بهترین سرویس گیرنده انجام می‌پذیرد. به عبارت دیگر سایت منتخب یک جد از بهترین سرویس گیرنده است. این کار می‌تواند از ریشه شروع و با تکرار الگوریتم تا رسیدن به بهترین سرویس گیرنده ادامه یابد (هونگ و همکاران^۲, ۲۰۰۸).

انتشار سریع:^۳ یکی از روش‌ها و مکانیزم‌های پویا برای انجام عمل تکرار در گرید داده است. در این روش هنگامی که یک سرویس گیرنده برای یک فایل یا موجودیت داده‌ای درخواستی را اعلام می‌کند، عمل تکرار از ریشه تا بهترین سرویس گیرنده انجام می‌پذیرد (نگوین و همکاران^۴, ۲۰۰۸).

تکرار پویا بر اساس محبوبیت:^۵ یک روش ایجاد تکرار پویا برای کاهش زمان دسترسی بر اساس تکرار فایل‌هایی با محبوبیت بیشتر در گره‌های یک گرید داده است. در این روش در مرحله اول از برگ‌های درخت نسبت به محاسبه تعداد درخواست‌ها تا رسیدن به ریشه میزان مجموع درخواست‌ها برای یک فایل در هر سطح مشخص می‌شود و در مرحله دوم از ریشه به سمت فرزندها حرکت و اگر تعداد درخواست‌ها برای فرزند یک گره از یک حد آستانه کمتر یا فرزند آخرین گره بود یک عمل تکرار در گره انجام می‌پذیرد (شورفازامن^۶ و همکاران, ۲۰۰۸).

¹ Cascade Replication

² Hong

³ Fast Spread

⁴ Nguyen

⁵ Popularity-Driven Dynamic Replica or Popularity Based Replica Placement (PBRP)

⁶ Shorfuzzaman

۳-۱ فرضیه‌ها

- ویژگی کوتاه‌ترین فاصله در انتخاب سایت مقصد برای تکرار داده موثر است.
- تعداد درخواست‌های قبلی برای یک موجودیت داده‌ای در گزینش سایت مقصد برای تکرار داده در گرید داده موثر می‌باشد.
- اندازه فایل مورد انتقال از سایت مبدأ به سایت مقصد در گزینش سایت مقصد برای تکرار موثر می‌باشد.
- یک شاخص کلی و اصلی در انتخاب سایت مقصد در گرید داده بار کاری مجموعه می‌باشد. این بار کاری مستقیماً از جمع فاصله ضرب در تعداد درخواست‌های قبلی و اندازه فایل بدست می‌آید.
- میزان زمان پاسخ^۱، و حجم استفاده شده در حافظه جانبی در سناریوی اجرای^۲ گرید داده در مقایسه با دو الگوریتم دیگر کمتر می‌باشد.

۴-۱ اهداف تحقیق

- هدف اصلی این تحقیق ارایه یک الگوریتم برای افزایش کارایی گرید داده در دستیابی به اطلاعات مورد نیاز می‌باشد. در کنار این هدف اصلی یک شبیه ساز پایه برای مقایسه الگوریتم‌های مختلف گرید داده بر اساس میزان استفاده از حافظه جانبی و زمان دسترسی ایجاد می‌شود. برای رسیدن به هدف اصلی، موضوعات زیر را دنبال می‌کنیم:
- استفاده از ویژگی‌هایی برای تکرار داده که محاسبات پیچیده‌ای نداشته باشند و کل گرید داده را پوشش دهد.
- پیشنهاد الگوریتم مورد نظر و پیاده سازی آن.
- شبیه سازی سناریوی اجرای گرید داده در درخواست‌های متوالی به صورت تصادفی و مقایسه کارایی الگوریتم پیشنهادی با دو الگوریتم پایه‌ای دیگر.

¹ Response Time

² Execution Scenario

۵-۱ روش تحقیق

جهت انجام تحقیق، مطالعه و بررسی تحقیقات قبلی در زمینه گرید داده، ویژگی‌های آن و الگوریتم‌های تکرار صورت می‌گیرد. برای این منظور کتب، پایان نامه‌های داخلی و خارجی و مقالات معتبر داخلی و بین‌المللی مطالعه می‌شوند. در ارزیابی کارهای مرتبط ارتباط بین پارامترهای مختلف در گزینش سایت مقصد در نظر گرفته می‌شود و سعی می‌شود که بسط متناسبی در خصوص این ارتباطات در الگوریتم پیشنهادی دیده شود. بعد از پیاده سازی سیستم پیشنهادی به صورت شبه کد، به علت عدم امکان استفاده از محیط واقعی برای اجرا و ارزیابی الگوریتم ارایه شده و همچنین فقدان شبیه ساز مناسب برای تطبیق با خصوصیات مورد نیاز این الگوریتم، نسبت به پیاده سازی الگوریتم مورد نظر در یک شبیه ساز اقدام خواهد شد. همچنین این شبیه ساز منطق دو الگوریتم مورد مقایسه دیگر را در خود پیاده سازی خواهد کرد. معیارهای مقایسه بین سه الگوریتم مورد نظر میزان زمان اجرای سناریوی اجرا و میزان استفاده از حافظه در نظر گرفته خواهند شد و برای این کار شبیه ساز برای الگوهای دسترسی مختلف و عمق‌های متفاوت در درخت گرید داده اجرا و نتایج حاصل از آن ثبت و مورد بررسی قرار می‌گیرد.

۶-۱ جنبه نوآوری طرح

در تکرار داده پویا پارامترهای زمانی مانند زمان تکرار، محل تکرار و الگوریتمی، مانند روش تکرار از اهمیت خاصی برخوردار هستند. در این تحقیق به بیان یک الگوریتم جدید پویا بنام «الگوریتم تکرار داده آبشاری چند تایی بر اساس کمینه سازی بار کاری در گرید داده» خواهیم پرداخت. تمرکز اصلی الگوریتم مورد نظر به روی روش تکرار و محل تکرار بر اساس بار کاری گرید داده در زمان مورد نظر است.

این الگوریتم بر خلاف روش بهترین سرویس گیرنده علاوه بر تعداد درخواست قبلی برای یک موجودیت داده‌ای، فاصله بین سایت‌های درخواست دهنده را به عنوان معیار گزینش سایت مقصد برای تکرار داده ارایه می‌دهد. همچنین این الگوریتم بر خلاف تکرار داده آبشاری نسبت به گزینش سایت خاصی در مسیر ریشه به سرویس گیرنده اقدام می‌کند. این الگوریتم علاوه بر پارامتر تعداد درخواست و فاصله بین سایت‌ها از پارامتر سومی بنام اندازه فایل مورد نظر برای انتقال کمینه داده