دانشگاه شیخ بهایی

دانشکده مهندسی

پایان‌نامه کارشناسی ارشد رشته مهندسی کامپیوتر - نرم‌افزار

ارائه روشی بهبودیافته برای تکرار داده سلسله مراتبی پویا در سیستم‌های گرید داده‌ای

پژوهشگر

شیدا دیانی دردشتی

استاد راهنما

دکتر محمدرضا خیام باشی

شهریور ۱۳۹۲

دانشگاه شیخ بهایی

دانشکده مهندسی

پایان‌نامه کارشناسی ارشد رشته مهندسی کامپیوتر - نرم‌افزار

ارائه روشی بهبودیافته برای تکرار داده سلسله مراتبی پویا در سیستم‌های گرید داده‌ای

پژوهشگر

شیدا دیانی دردشتی

استاد راهنما

دکتر محمدرضا خیام باشی

شهریور ۱۳۹۲

باسمه تعالی

دانشگـــــــــاه شیـــــــــخ بهائی

دانشکده فنی مهندسی

گروه کامپیوتر و فناوری اطلاعات

پایان‌نامه کارشناسی ارشد رشته مهندسی کامپیوتر – نرم افزار خانم شیدا دیانی دردشتی

تحت عنوان

ارائه روشی بهبودیافته برای تکرار داده سلسله مراتبی پویا در سیستم‌های گرید داده‌ای

در تاریخ **۱۳۹۲/۷/۳۰** توسط هیأت داوران زیر بررسی و با درجه ..... به تصویب نهائی رسید.

۱- استاد راهنمای پایان‌نامه	**دکتر محمدرضا خیّام‌باشی**	با مرتبه علمی	**استادیار**	امضاء
۲- استاد داور داخل گروه	**دکتر مهدی باطنی**	با مرتبه علمی	**استادیار**	امضاء
۳- استاد داور خارج گروه	**دکتر فرامرز صافی**	با مرتبه علمی	**استادیار**	امضاء

دکتر سید محمدحسن فیض

مدیر تحصیلات تکمیلی

باسمه تعالی

اقرارنامه

اینجانب **شیدا دیانی دردشتی** به شماره دانشجویی **۹۰۲۵۱۱۰** دانشجوی رشته **مهندسی کامپیوتر- نرم‌افزار** که پایان‌نامه خود تحت عنوان:

ارائه روشی بهبودیافته برای تکرار داده سلسله مراتبی پویا در سیستم‌های گرید داده‌ای

را نوشته و برای دفاع آماده کرده‌ام اعلام می‌نمایم که محتوا و نوشته‌های این پایانامه متعلق به خودم بوده و هیچ قسمت از آن به طور مستقیم یا غیرمستقیم کپی و یا برگرفته از کار دیگران خارج از ضوابط متعارف نگارش پایان‌نامه نمی‌باشد. اینجانب نیز آگاهم که در صورتی که خلاف موضوع فوق‌الذکر در هر زمان و به هر طریق اثبات گردد از کلیه امتیازات مکتسبه از این پایان‌نامه محروم و ملزم به پذیرش عواقب و مجازات حقوقی ناشی از آن می‌باشم.

نام و نام‌خانوادگی امضا

هرچه در زندگی ام نگریستم

تکیه گاهی استوارتر و سرپناهی پرصلابت تر از پدر نیافتم

وهرچه در کتاب های لغت جستجو کردم

کلمه ای زیباتر از مادر نیافتم تا با آن سنگ صبور زندگی ام را مخاطب قرار دهم

تقدیم به دو جزء جدایی ناپذیر زندگی ام

پدرم ، مادرم

با تشکر و قدردانی از

استاد فرزانه و دانشمند جناب آقای دکتر محمدرضا خیام باشی، به پاس

حمایت ها و راهنمایی های ارزنده ایشان.

چکیده

پردازش گرید، یکی از انواع پردازش توزیع‌شده می‌باشد که در آن منابع مختلف کامپیوتری توسط سازمان‌های مختلف به اشتراک گذاشته می‌شوند تا یک کامپیوتر مجازی با قدرت زیاد ایجاد شود و برای حل مسائل پیچیده داده‌ای و محاسباتی آن سازمان‌ها به کار گرفته شود. گرید امروزه به عنوان یک فناوری حیاتی برای اجرای برنامه‌های با کارایی بالا مطرح شده است و با پیشرفت فن‌آوری، پروتکل‌ها و بسترهای آن بی‌تردید در آینده استفاده از آن فراگیرتر خواهد شد. منابع به اشتراک گذاشته شده در گرید، عموماً ناهمگن بوده و به صورت جغرافیایی توزیع‌شده هستند. تکرار داده از معروف‌ترین استراتژی‌هایی است که در سیستم‌های توزیع‌شده به منظور بهبود دستیابی به داده و بدست آوردن سطح بالایی از دسترس‌پذیری، تحمل‌پذیری خرابی و قابلیت اطمینان سیستم استفاده می‌شود. به همین منظور از این روش برای مدیریت داده‌ها در سیستم‌های گرید داده‌ای نیز استفاده می‌شود. استراتژی‌های تکرار داده، وابسته به این مسئله می‌باشند که تکرار کجا، چه زمانی، چگونه ساخته شده و چگونه از بین برود.

امروزه با وجود تمام تحقیقات صورت گرفته در زمینه تکرار پویای داده‌ها در محیط‌های گرید، هنوز چالش‌های زیادی از قبیل بالا بودن زمان اجرای کارها، تاخیر زیاد در دسترسی به فایل‌ها، محدود بودن تعداد و ظرفیت منابع ذخیره‌سازی در دسترس و عدم استفاده ترکیبی از استراتژی‌های تکرار و زمان‌بندی مطرح می‌باشند.

در این پژوهش هدف ارائه استراتژی تکرار داده جدیدی به نام RDT[1] می‌باشد. اغلب الگوریتم‌های ارائه شده، در هر مرحله فقط یک تکرار از فایل ایجاد می‌کنند؛ در صورتی که ممکن است تعداد درخواست‌ها به آن فایل به حدی زیاد باشد که عدم ایجاد تکرار به تعداد مناسب سبب کاهش کارایی و افزایش مصرف پهنای‌باند شود. بنابراین، تعیین یک حد آستانه پویا جهت مشخص کردن مناسب‌ترین تعداد تکرار برای فایل تقاضا شده با توجه به شرایط موجود در محیط گرید، موجب کاهش زمان اجرا و افزایش کارایی خواهد شد. هم‌چنین، تکرار فایل در تمامی سایت‌هایی که درخواست فایل را داده‌اند ایجاد نمی‌شود، بلکه به کمک استراتژی پیشنهادی این پژوهش، تکرار فقط در سایت-هایی که به صورت پویا مناسب شناخته می‌شوند ایجاد خواهد شد و در نتیجه هزینه ذخیره سازی و میانگین زمان اجرای کار کاهش خواهد یافت. استراتژی تکرار داده‌ای که در این پژوهش ارائه خواهد شد، با استفاده از الگوریتم زمان‌بندی پیشنهادی، کارایی قابل قبولی خواهد داشت و قابل استفاده در محیط‌های واقعی گرید می‌باشد. در نهایت روش ارائه شده در یک محیط شبیه‌ساز گرید با پارامترهای مشابه گریدهای واقعی مورد بررسی و ارزیابی قرار خواهد گرفت.

کلمات کلیدی:

سیستم‌های توزیع‌شده[2]، گرید داده‌ای[3]، تکرار داده[4]، زمان‌بندی[5].

[1] Replication with Dynamic Threshold
[2] Distributed systems
[3] Data grid
[4] Data replication
[5] Scheduling

فهرست مطالب

۱. فصل اول - معرفی و آشنایی با مفاهیم اولیه ۱

۱-۱- مقدمه ۲

۱-۲- طرح مسئله ۴

۱-۳- اهداف و کاربرد تحقیق ۷

۱-۳-۱- کاربرد در سیستم‌های گرید ۸

۱-۳-۲- کاربرد در تکرار داده‌ها ۹

۱-۴- اهمیت و ارزش تحقیق ۹

۱-۵- فرضیات و مراحل انجام تحقیق ۱۰

۱-۶- ساختار پایان‌نامه ۱۲

۱-۷- جمع‌بندی ۱۲

۲. فصل دوم - ادبیات موضوع ۱۳

۲-۱- مقدمه ۱۴

۲-۲- سیستم‌های توزیع‌شده ۱۴

۲-۲-۱- سیستم‌های اطلاعاتی توزیع شده ۱۷

۲-۲-۲- سیستم‌های توکار توزیع‌شده ۱۷

۲-۲-۳- سیستم‌های محاسباتی توزیع‌شده ۱۷

۲-۳- پردازش گرید ۱۸
۲-۴- گرید داده‌ای ۲۳
۲-۵- تکرار داده‌ها ۲۳
۲-۵-۱- مزایا و معایب ۲۴
۲-۵-۲- پارامترهای بهینه سازی در اگوریتم‌های تکرار ۲۶
۲-۵-۳- دسته‌بندی استراتژی‌های تکرار داده ۲۷
۲-۵-۳-۱- تکرار داده ایستا ۲۷
۲-۵-۳-۲- تکرار داده پویا ۲۷
۲-۵-۳-۳- مقایسه دو روش تکرار داده ۲۸
۲-۶- زمان‌بندی کارها در گرید ۲۸
۲-۷- جمع‌بندی ۲۹

۳. فصل سوم - مروری بر کارهای انجام شده ۳۰

۳-۱- مقدمه ۳۱
۳-۲- مروری بر استراتژی‌های تکرار داده ۳۱
۳-۳- جمع‌بندی ۴۰

۴. فصل چهارم - ارائه راه حل‌های پیشنهادی ۴۲

۴-۱- مقدمه ۴۳
۴-۲- داده‌های استاندارد و ساختار توپولوژی گرید در شبیه‌ساز ۴۳
۴-۳- ایجاد تکرار جدید ۴۶
۴-۴- استراتژی تکرار داده DHRT ۴۶

۴-۴-۱- الگوریتم انتخاب تکرار ۴۶
۴-۴-۲- الگوریتم تصمیم‌گیری یا مکان‌یابی تکرار ۴۹
۴-۴-۳- الگوریتم حذف و جای‌گزینی تکرار ۵۲
۴-۵- استراتژی تکرار داده RDT ۵۴
۴-۵-۱- الگوریتم انتخاب تکرار ۵۴
۴-۵-۲- الگوریتم تصمیم‌گیری یا مکان‌یابی تکرار ۵۷
۴-۵-۳- الگوریتم حذف و جای‌گزینی تکرار ۵۷
۴-۶- الگوریتم زمان‌بندی کارها ۵۹
۴-۷- جمع‌بندی ۶۲

۵. فصل پنجم - پیاده‌سازی و ارزیابی نتایج ۶۳

۵-۱- مقدمه ۶۴
۵-۲- شبیه‌ساز اپتورسیم ۶۴
۵-۳- پیاده‌سازی الگوریتم‌های تکرار پیشنهادی ۶۷
۵-۳-۱- الگوریتم انتخاب تکرار ۶۷
۵-۳-۲- الگوریتم مکان‌یابی تکرار ۶۷
۵-۳-۳- الگوریتم جای‌گزینی تکرار ۶۸
۵-۴- پیاده‌سازی الگوریتم زمان‌بندی پیشنهادی ۶۸
۵-۵- پارامترهای شبیه‌سازی ۶۹
۵-۶- مرحله اول: ارزیابی استراتژی تکرار داده پیشنهادی DHRT ۷۱
۵-۷- مرحله دوم: ارزیابی استراتژی تکرار داده پیشنهادی RDT ۷۴
۵-۸- مرحله سوم: مقایسه سه استراتژی DHR، DHRT و RDT ۷۷

۵-۹- مرحله چهارم: ارزیابی الگوریتم زمان‌بندی پیشنهادی QAQ ۸۱

۵-۱۰- مرحله پنجم: درصد بهبود نتایج استراتژی‌های پیشنهادی ۸۵

۵-۱۱- جمع‌بندی ۸۶

۶. فصل ششم- نتیجه‌گیری و کارهای آتی ۸۹

۶-۱- نتیجه‌گیری ۹۰

۶-۲- پیشنهادات و کارهای آتی ۹۲

فهرست مراجع ۹۴

فهرست اشکال

شکل ۱-۱- حوزه پژوهش (الف) ۶

شکل ۱-۲- حوزه پژوهش (ب) ۷

شکل ۱-۳- جایگاه استراتژی تکرار و الگوریتم زمانبندی پیشنهادی در سیستم گرید..... ۸

شکل ۱-۴- مراحل انجام پروژه..... ۱۱

شکل ۲-۱- اجزای تشکیل‌دهنده یک سیستم گرید ۱۹

شکل ۲-۲- معماری لایه‌ای گرید ۲۱

شکل ۴-۱- ساختار شبکه داده‌های استاندارد CMS..... ۴۴

شکل ۴-۲- توپولوژی گرید مورد استفاده..... ۴۵

شکل ۴-۳- شبه کد الگوریتم انتخاب تکرار استراتژی پیشنهادی DHRT ۴۸

شکل ۴-۴- نمودار گردشی الگوریتم انتخاب تکرار استراتژی پیشنهادی DHRT ۴۸

شکل ۴-۵- شبه کد الگوریتم مکان‌یابی تکرار استراتژی پیشنهادی DHRT..... ۵۰

شکل ۴-۶- نمودار گردشی الگوریتم مکان‌یابی تکرار استراتژی پیشنهادی DHRT..... ۵۱

شکل ۴-۷- شبه کد الگوریتم حذف و جای‌گزینی تکرار استراتژی پیشنهادی DHRT ۵۳

شکل ۴-۸- نمودار گردشی الگوریتم حذف و جای‌گزینی تکرار استراتژی پیشنهادی DHRT ۵۴

شکل ۴-۹- شبه کد الگوریتم انتخاب تکرار استراتژی پیشنهادی RDT ۵۶

شکل ۴-۱۰- نمودار گردشی الگوریتم انتخاب تکرار استراتژی پیشنهادی RDT..... ۵۶

شکل ۴-۱۱- شبه کد الگوریتم حذف و جای‌گزینی تکرار استراتژی پیشنهادی RDT ۵۸

شکل ۴-۱۲- نمودار گردشی الگوریتم حذف و جای‌گزینی تکرار استراتژی پیشنهادی RDT ۵۹

شکل ۴-۱۳- نمودار گردشی الگوریتم زمان‌بندی پیشنهادی ۶۱

شکل ۵-۱- ساختار شبیه‌ساز اپتورسیم ۶۵

شکل ۵-۲- میانگین زمان اجرای کار استراتژی پیشنهادی DHRT در مقایسه با استراتژی‌های تکرار مختلف ۷۲

شکل ۵-۳- درصد استفاده از عنصر ذخیره‌سازی استراتژی پیشنهادی DHRT در مقایسه با استراتژی‌های تکرار مختلف ۷۳

شکل ۵-۴- میزان استفاده موثر از شبکه استراتژی پیشنهادی DHRT در مقایسه با استراتژی‌های تکرار مختلف ۷۴

شکل ۵-۵- میانگین زمان اجرای کار استراتژی پیشنهادی RDT در مقایسه با استراتژی‌های تکرار مختلف ۷۵

شکل ۵-۶- درصد استفاده از عنصر ذخیره‌سازی استراتژی پیشنهادی RDT در مقایسه با استراتژی‌های تکرار مختلف ۷۶

شکل ۵-۷- میزان استفاده موثر از شبکه استراتژی پیشنهادی RDT در مقایسه با استراتژی‌های تکرار مختلف ۷۷

شکل ۵-۸- مقایسه میانگین زمان اجرای کار سه استراتژی DHR، DHRT و RDT ۷۹

شکل ۵-۹- مقایسه درصد استفاده از عنصر ذخیره‌سازی سه استراتژی DHR، DHRT و RDT ۷۹

شکل ۵-۱۰- مقایسه میزان استفاده موثر از شبکه در سه استراتژی DHR، DHRT و RDT ۸۰

شکل ۵-۱۱- مقایسه میانگین زمان اجرای کار برای تعداد مختلف کار در سه استراتژی DHR، DHRT و RDT ۸۰

شکل ۵-۱۲- میانگین زمان اجرای کار برای الگوریتم‌های تکرار و زمان‌بندی مختلف ۸۲

شکل ۵-۱۳- میانگین زمان اجرای کار زمان‌بندهای مختلف برای استراتژی تکرار داده LRU ۸۳

شکل ۵-۱۴- میانگین زمان اجرای کار زمان‌بندهای مختلف برای استراتژی تکرار داده پیشنهادی RDT ۸۴

شکل ۵-۱۵- میانگین زمان اجرای کار زمان‌بندهای مختلف برای استراتژی بدون تکرار ۸۵

فهرست جداول

جدول ۵-۱- پیکربندی پهنای باند ۶۹

جدول ۵-۲- پارامترهای شبیه‌سازی ۶۹

جدول ۵-۳- مقایسه استراتژی‌های DHR، DHRT و RDT ۷۸

جدول (۵-۴): درصد بهبود استراتژی‌های تکرار DHRT و RDT در حالت بهینه ۸۵

جدول (۵-۵): درصد بهبود استراتژی‌های تکرار DHRT و RDT در حالت میانگین ۸۶

جدول (۵-۶): درصد بهبود الگوریتم زمانبندی QAQ در حالت بهینه ۸۶

جدول (۵-۷): درصد بهبود الگوریتم زمانبندی QAQ در حالت میانگین ۸۶

فصل اول

معرفی و آشنایی با مفاهیم اولیه

فصل اول

معرفی و آشنایی با مفاهیم اولیه

۱-۱- مقدمه

پردازش گرید[1] یکی از انواع پردازش توزیع‌شده[2] می‌باشد که در آن منابع مختلف کامپیوتری توسط سازمان‌های مختلف به اشتراک گذارده می‌شوند تا یک کامپیوتر مجازی بزرگ با قدرت زیاد ایجاد شود و برای حل مسائل بزرگ و پیچیده داده‌ای و محاسباتی آن سازمان‌ها به کار گرفته شود. سیستم‌های گرید با توجه به نیازهای کاربران می‌توانند در انواع مختلف مورداستفاده قرار گیرند. اغلب این سیستم‌ها بر اساس نوع کاری که انجام می‌دهند، طبقه‌بندی می‌شوند. از سه نوع اصلی این سیستم می‌توان از موارد زیر نام برد [۱]:

۱) گرید محاسباتی[3]
۲) گرید جوینده منابع[4]
۳) گرید داده‌ای[5]

تحقیقات علمی در زمینه‌های مختلفی مانند فیزیک، شیمی، پزشکی و غیره، حجم زیادی از داده را تولید می‌کنند؛ که لازم است این داده‌ها به منظور دست‌یابی سایر محققان در سراسر دنیا به اشتراک گذاشته شوند. واضح است که مدیریت این حجم انبوه داده به صورت متمرکز[6]، به علت تاخیر دسترسی[7] و بارگذاری[8] بسیار زیاد روی سرور مرکزی بی‌نتیجه

[1] Grid computing
[2] Distributed computing
[3] Computational grid
[4] Scavenging grid
[5] Data grid
[6] Centralised
[7] Access latency
[8] Load

خواهد بود. برای حل این مشکل استفاده از تکنولوژی‌های گرید پیشنهاد شده است. گرید داده، توزیع چنین داده‌هایی را در مکان‌های جغرافیایی مختلف تسهیل می‌کند و تکنیکی کارآمد و موثر، برای مدیریت داده[1]های توزیع شده با حجم بالا و منابع داده‌ای مشترک می‌باشد [۲، ۳]. این نوع از گریدها، موظفند اطلاعات را ذخیره کنند و آن‌ها را در اختیار کاربران قرار دهند. کاربران این سیستم‌ها بدون آن که از موقعیت جغرافیایی و مکانی این اطلاعات آگاه باشند، به اطلاعات دسترسی دارند. هم‌چنین بر خلاف گریدهای محاسباتی، دستگاه‌های متصل به سیستم نیاز به قدرت زیاد ندارند و فقط مسئول به اشتراک گذاشتن اطلاعات هستند [۱].

تکرار داده‌ها[2] یک تکنیک مهم برای مدیریت داده‌های انبوه به صورت توزیع شده می‌باشد. ایده کلی تکرار داده‌ها، قرار دادن تکرارها در مکان‌های مختلف می‌باشد؛ در حالتی که چندین تکرار از فایلی خاص در سایت‌های مختلف وجود دارد. در گریدهای داده‌ای اصولا تکرارهای متعددی از هر داده نگه‌داری می‌شود و کاربران برای دسترسی به اطلاعات به در دسترس‌ترین، نزدیک‌ترین و سریع‌ترین تکرار هدایت می‌شوند. برای مدیریت گریدهای داده‌ای مکانیزم‌های مختلفی تاکنون پیشنهاد شده است. یکی از مهم‌ترین مسائلی که در مدیریت گریدهای داده‌ای مطرح است، چگونگی برخورد سیستم با تکرارها است. این‌که از کدام فایل‌ها تکرار ساخته شود، در چه زمانی از داده‌ها تکرار ایجاد شود و این تکرار در کجای سیستم قرار داده شود تا بازدهی کلی گرید داده‌ای بالا برود، از جمله مسائلی است که پیرامون آن‌ها بحث، تبادل نظر و تحقیقات وسیعی در جریان است [۲].

انتخاب بهترین تکرار، زمان اجرای کار[3] را بهبود می‌بخشد. هم‌چنین زمان‌بندی[4] موثر در محیط گرید، با اجرای کار در جایی که به داده‌های مورد نیازش دسترسی سریع‌تر داشته باشد، باعث کاهش انتقال داده از طریق اینترنت شده و زمان اجرای کارها را بهبود می‌بخشد [۴].

تا کنون الگوریتم‌های زیادی برای تکرار داده‌ها ارائه شده است، که به دو دسته کلی زیر تقسیم می‌شوند:

۱) تکرار ایستا[5]: در این الگوریتم‌ها، سیاست تکرار از ابتدا به صورت ثابت و ایستا مشخص می‌گردد و در واقع جزئی از پیکربندی سیستم محسوب می‌شود. با تغییر توپولوژی شبکه گرید داده‌ای و یا تغییر الگوی درخواست‌ها تغییری در این سیاست داده نخواهد شد. بنابراین بازدهی سیستم به شدت پایین می‌آید و ممکن است از منابع به شکل مناسب استفاده نشود [۵].

[1] Data management
[2] Data Replication
[3] Job execution time
[4] scheduling
[5] Static Replication

۲) تکرار پویا[1]:در الگوریتم‌هایی که از تکرار پویا استفاده می‌کنند، با توجه به تغییر الگوهای دسترسی و موارد متغیر دیگر در گرید داده‌ای مورد نظر، عملیات تکرار فایل به صورت اتوماتیک و در صورت نیاز انجام خواهد شد و تکرار ایجاد شده در جای مناسب قرار می‌گیرد، تا فرکانس دسترسی به آن بهینه باشد و باعث بالا رفتن هزینه نگردد [۵].

۱-۲- طرح مسأله

در [۴]، یک استراتژی برای تکرار داده پویای سلسله مراتبی به نام DHR[2] ارائه شده است. این الگوریتم از نوع الگوریتم‌های تکرار پویا است و برای ساختار شبکه سلسله مراتبی سه سطحی ارائه شده است. این الگوریتم تکرار فایل‌ها را در سایت‌های مناسب قرار می‌دهد. بهترین سایت برای هر تکرار، سایتی است که بالاترین تعداد دسترسی به آن فایل مشخص را داشته است. استراتژی DHR، از سه الگوریتم زیر تشکیل می‌شود:

۱. الگوریتم انتخاب تکرار[3]: به طور کلی ممکن است چندین تکرار در سایت‌های مختلف در دسترس باشند. در این صورت الگوریتم تکرار به کار رفته در استراتژی DHR، سایتی که کم‌ترین تعداد درخواست‌ها را داشته باشد، انتخاب می‌کند [۴].

۲. الگوریتم مکان‌یابی یا تصمیم‌گیری تکرار[4]: در استراتژی DHR، زمانی‌که عملیات تکرار رخ می‌دهد، بر اساس محلیت زمانی و جغرافیایی، تکرار ایجاد شده در بهترین سایت قرار داده می‌شود. برای انتخاب بهترین سایت، استراتژی DHR یک لیست مرتب شده از تمامی المان‌های ذخیره سازی[5] موجود در ناحیه که آن فایل خاص را تقاضا کرده‌اند تهیه می‌کند. سپس تکرار فایل مورد نظر در سایتی که شماره آن در بالای لیست مرتب شده قرار دارد، ذخیره خواهد شد [۴].

۳. الگوریتم جای‌گزینی تکرار[6]: اگر فضای کافی برای تکرار کردن فایل درخواست شده وجود نداشته باشد به کمک روش LRU[7] برخی از فایل‌ها برای حذف انتخاب می‌شوند و تا زمانی‌که فضای کافی ایجاد شود این روند ادامه می-یابد [۴].

اکنون با وجود تمام تحقیقات صورت گرفته در زمینه تکرار پویای داده‌ها در محیط‌های گرید، هنوز چالش‌های زیر مطرح می‌باشد:

- بالا بودن زمان اجرای کارها در سایت های گرید [۳]

[1] Dynamic Replication
[2] Dynamic Hierarchical Replication
[3] Replica selection algorithm
[4] Replica decision algorithm
[5] Storage Element
[6] Replica replacement algorithm
[7] Least Recently Used