

به کارگیری وب‌سرویس‌ها برای مدیریت جریان‌های کار علمی در محیط گردید

سعید پارسا^۱، حسن مهدیخانی^۲

چکیده

اقبال روزافزون در به کارگیری ابزارهای اتوماسیون و مدیریت جریان کار به عنوان یک روش مهم در تعریف و اجرای فرآیندها از یک سو و از سوی دیگر توسعه کارکردهای اینترنت از صرفاً دسترسی آسان و سریع به اطلاعات به پردازش اطلاعات ما را بر آن داشته است تا به ارائه معماری و پیاده‌سازی یک سیستم مدیریت جریان کار مبتنی بر گردید مبادرت ورزیم، در این معماری از وب‌سرویس‌ها به عنوان سیستم نرم‌افزاری زیربنایی در برقراری ارتباط و تعامل میان ماسین‌های شبکه استفاده شده است و اجرای جریان‌های کاری به اتكای منابع نرم‌افزاری و سخت‌افزاری توزیع شده در گردید میسر گردیده. تحقق این امر کارکردهای مهمی نظری تسریع و تسهیل در دسترسی به سرویس‌ها، توزیع و توزین بارکاری، افزایش کارایی، قابلیت مواجه با خطأ و نظایر آن فراهم می‌آورد. در این مقاله ضمن ارایه معماری ای نوین برای سیستم‌های مدیریت جریان کار، زبانی جدید برای تعریف جریان کار و روال پیاده‌سازی آن با استفاده از گردید و وب‌سرویس‌ها معرفی و ارزیابی شده است.

کلمات کلیدی

گردید علمی، محاسبات گردیدی، سیستم‌های مدیریت جریان کار، وب‌سرویس، جریان کار علمی

Application of Web Services to Manage Scientific Workflows in Grid Environments

Saeed Parsa, Hassan Mahdikhani

Abstract

The increasing fortune in the application of automation tools and the management of workflows as an important approach to define and execute working processes from one side and the expansion of Internet applications in solely fast and easy access of information to rapid processing of information from the other side motivated us to present a new architecture and implement a workflow management system over the grid. This architecture applies web services to settle communications and interactions between distributed hardware and software resources over the grid environment. Building such a workflow engine based on web services brings many attractive characteristics such as accelerating and facilitating access to web services, good load balancing and sharing, great efficiency, higher fault tolerant ability and reliability. In this paper while presenting a novel architecture for workflow management systems, a new workflow definition language and approach to implement workflow engines, applying grid and web services, is presented and evaluated.

Keywords

Scientific Grid, Grid Computing, Workflow Management System, Web Service, Scientific Workflow

^۱ عضو هیأت علمی دانشگاه، دانشگاه علم و صنعت ایران، دانشکده مهندسی کامپیوتر، parsa@iust.ac.ir

^۲ دانشگاه علم و صنعت ایران، دانشکده مهندسی کامپیوتر، h_mahdikhani@comp.iust.ac.ir

هرچند تریانا [14] با استفاده از رابطی به نام GAT^۴ امکان ارسال جریان‌های کاری به محیط‌های گریدسرویس را نیز فراهم آورده لیکن محدودیت‌هایی از قبیل فقدان قیود کیفیت خدمات^۵ و عدم بهره‌گیری از مکانیزم‌های بازیابی و جستجوی خودکار وب‌سرویس‌ها کماکان وجود دارد که در بخش ۲ و با تمهید مولفه‌های مناسب در طرح پیشنهادی این مشکلات رفع شده است. لازم به ذکر است در بخش ۴ جدول مقایسه‌ای سیستم‌های مدیریت جریان کار گریدی با مدل پیشنهادی آمده است [۱].

از دیگر نقاط ضعف برخی سیستم‌های مدیریت جریان کار گریدی نظیر تاونا، کپلر، تریانا و گریدآنت عدم امکان به کارگیری قیود کیفیت خدمات می‌باشد در نتیجه زمانبندی جریان‌های کار بدون توجه به بارکاری شبکه، بارکاری پردازنده، حجم حافظه موردنیاز و سیستم عامل مورد استفاده صورت می‌پذیرد که این مسائل در طرح پیشنهادی همان‌گونه که در بخش ۳-۱ آمده، از میان برداشته شده است. در ادامه و در بخش ۲ مدل پیشنهادی و در بخش ۳ نکات مرتبط با پیاده‌سازی سیستم مورد بررسی قرار گرفته است و نهایتاً در بخش ۴ از دو دیدگاه نظری و عملی به ارزیابی طرح پیشنهادی پرداخته شده است.

۲- مدل پیشنهادی

در مدل مرجع پیشنهاد شده توسط ائتلاف جریان کار [12]، یک سیستم مدیریت جریان کار شامل ابزارهای تعریف جریان کار، ابزارهای مدیریتی و نظارتی، برنامه‌های کاربردی سمت مشتری، برنامه‌های فرخوانی شده و موتور جریان کار است. فرآیندهای کاری توسط ابزار تعریف جریان کار تعریف و برای اجرا در اختیار موتور جریان کار قرار می‌گیرند، وظایف موجود در جریان‌های کاری توسط عوامل انسانی یا برنامه‌ها (برنامه‌های محلی یا برنامه‌های فرخوانی شده) اجرا می‌شوند و برای نظارت و کنترل بر روند اجرای آن‌ها، ابزارهای مدیریتی خاصی پیش‌بینی شده است.

در مدل پیشنهادی، وظایف توسط وب و گریدسرویس‌های توزیع شده در گرید اجرا می‌شوند و از آنجا که رابطه‌های استاندارد و مکانیزم‌های مناسبی برای جستجوی آن‌ها پیش‌بینی شده است برنامه‌های کاربردی به راحتی امکان فرخوانی آن‌ها را دارند این امر تعریف و مدل‌سازی، ذخیره‌سازی و اجرای جریان‌های کار را نیز تسهیل می‌بخشد؛ همچنین میان‌افزار پیش‌بینی شده براساس درخواست کاربر و مشخصه‌های تعیین شده در تعریف جریان کار، معیارهای کیفی ارائه خدمات را فراهم آورده و آن را برای اجرا زمانبندی می‌کند. در نتیجه جریان کار جهت اجرا روی ماشین‌های مختلف توزیع و در روال اجرا، متناسب با کارایی ماشین‌های مقصد تغییرات مقتضی در انتخاب ماشین و حجم کار ارسالی به آن صورت می‌پذیرد؛ همچنین برای مقابله با بروز خطاهای احتمالی مکانیزم تحمل خطا پیش‌بینی شده است.

۱- مقدمه

هدف در این مقاله ارایه جزئیات و پیاده‌سازی سیستم مدیریت جریان کار برای برنامه‌های علمی بزرگ که در سطح گستردگی نیاز به استفاده از منابع گوناگون از قبیل پردازشگرها، داده‌ها، برنامه‌ها و تجهیزات علمی و آزمایشگاهی دارند، می‌باشد. محاسبات گریدی به عنوان یک بستر پراهمیت در اجرای نرم‌افزارهای علمی روی تعداد زیادی از منابع نامتجانس توزیع شده در شبکه‌ای به بزرگی اینترنت شایان توجه بوده و به جهت تاکید بر اشتراك منابع در مقیاس بزرگ و نیل به کارایی بالا از شیوه‌های مرسوم و معمول نظیر محاسبات توزیع شده، محاسبات خوش‌های و پردازش موازی متمایزند [۳].

در راستای ارائه سیستم‌های مدیریت جریان کار مبتنی بر گرید فعالیت‌های تحقیقاتی مختلفی صورت پذیرفته است. برای نمونه دگ-من [2] به عنوان یک موتور جریان کار گریدی در بسیاری پژوهش‌ها نظیر کوندورجی [2]، چیمرا [4]، پگاسوس [5] و پی‌گراید [6] مورد استفاده قرار گرفته است. یکی از ضعف‌های عمدی دگمن عدم پشتیبانی ساختارهای تکرار در تعریف و نهایتاً اجرای جریان‌های کاری است [2]. این ضعف علاوه بر دگمن در تاونا [7] و گرداس [8] نیز دیده می‌شود که با پیش‌بینی برچسب مناسبی در زبان پیشنهادی که جزیيات آن در بخش ۱-۳ آمده، برطرف شده است. علی‌رغم رفع مشکل فوق در برخی سیستم‌ها مانند گریدآنت [9]، کپلر [10] و آسکالون [11] آن‌ها خود مشکلات دیگری دارند برای نمونه گریدآنت در مدل‌سازی و تعیین محل اجرای جریان‌های کار از روش نگاشت صریح^۶ (در مقابل روش انتزاعی^۷) بهره می‌جوید؛ در بخش ۲ و در مدل پیشنهادی با ارائه مولفه‌ای جهت بازیابی سرویس‌های مستقر در گرید این نقیصه نیز برطرف شده است، با تمام این تفاسیر مشکل بزرگ سیستم‌های مدیریت جریان کار گریدی فوق عدم پشتیبانی از سکوهای نامتجانس کاری است.

از آنجایی که وب‌سرویس‌ها برقراری ارتباط و تعامل بین ایستگاه‌های کاری را مستقل از سکو، وابستگی به نرم‌افزار جانی و مبتنی بر پروتکل‌های اینترنت [12] انجام می‌دهند، در شبکه‌های گرید به عنوان عوامل ارتباطی مناسب مطرح می‌باشند. لذا از آنجایی که در سیستم‌های مدیریت جریان کار، گرددش مستندات بر روی سکوهای کاری متفاوت حائز اهمیت است، در این مقاله پیشنهاد استفاده از وب و گریدسرویس‌ها (ائلاف گلوباس^۸ با اصلاح وب‌سرویس‌ها) جهت استفاده در محیط گرید استاندارد شناخته شده خود برای معماری گرید یعنی OGSA را در قالب OGSI مطرح و مفهوم گریدسرویس را معرفی نمود [13,15] در شبکه‌های گرید برای ایجاد سیستم‌های مدیریت جریان کار به منظور انجام محاسبات علمی بزرگ [5] که در سطح گستردگی نیاز به منابع پردازشگر، داده‌ها، برنامه‌ها و تجهیزات علمی و آزمایشگاهی دارند، آمده است.

- سرویس ذخیره/بازیابی نمونه‌ها: به ازای درخواست اجرای هر وب‌سرویس، طبق مدل تولید/نمونه^۷، نمونه‌ای ایجاد و کاربر اجرای درخواست‌های خود را از طریق این نمونه محقق می‌نماید. از این‌رو می‌بایست تمهیدات لازم در زمینه‌ی ذخیره، بازیابی و بروزآوری اطلاعات مربوط به نمونه‌های در حال اجرا را پیش‌بینی نمود.

- سرویس ذخیره/بازیابی/جستجوی وب‌سرویس‌ها: عملکرد این سرویس مشابه مکانیزم‌های فهرست‌بندی و ایندکس‌گذاری است. وب‌سرویس‌ها نام و اطلاعات جزئی‌تری (نظیر عنوان سرویس، تهیه‌کننده، QoS، هزینه‌ها و دسترسی‌ها و ...) از خود را در اینباره‌ای ذخیره می‌نمایند. از این‌رو در مراجعات میان‌افزار جهت اجرای کارها و یا در پروسه‌ی تعریف جریان کار می‌توان وب‌سرویس‌های مناسب را شناسایی نمود. این سرویس می‌بایستی قابلیت ثبت، حذف و بروزرسانی اطلاعات مربوط به وب‌سرویس‌ها همچنین اجرای پرس-وجوها جهت شناسایی آن‌ها فراهم آورد.

- سرویس ذخیره/بازیابی اطلاعات مربوط به جریان‌های کاری: این سرویس مدیریت اطلاعات مربوط به دادگان جریان‌های کار نظیر تهیه‌کنندگان، محل ذخیره‌ی آن‌ها و ... را فراهم می‌آورد. به منظور درک ملموس‌تر روند کار، برای نمونه سناریوی اجرای یک جریان کار در مدل پیشنهادی در ادامه آمده است. در مرجع [۱] سناریوهای مختلف به طور کامل تشریح شده است.

سناریوی اجرای جریان کار؛ ابتدا سرویس ذخیره/بازیابی اطلاعات جریان کار فراغوانی و مکان(های) استقرار جریان کار شناسایی می‌شود، اطلاعات مربوط به جریان کار از طریق سرویس ذخیره/بازیابی تعاریف جریان کار بازیابی و جریان کار جهت اجرا در اختیار میان‌افزار قرار می‌گیرد؛ میان‌افزار سرویس ذخیره/بازیابی/جستجوی وب‌سرویس‌ها را فراغوانی و وب‌سرویس(های) مقتضی را می‌یابد سپس با فراغوانی سرویس ذخیره/بازیابی اطلاعات جریان کار، وضعیت فعلی جریان کار را به حالت اجرا تغییر می‌دهد و مناسب با درخواست‌های واصله نمونه‌هایی از وب‌سرویس‌ها ایجاد و اطلاعات مربوط به این نمونه‌ها را با فرخانی سرویس ذخیره/بازیابی نمونه‌ها ذخیره می‌نماید در ادامه میان‌افزار با همکاری نمونه‌های ایجاد شده جریان کار را اجرا می‌کند.

۳- پیاده‌سازی

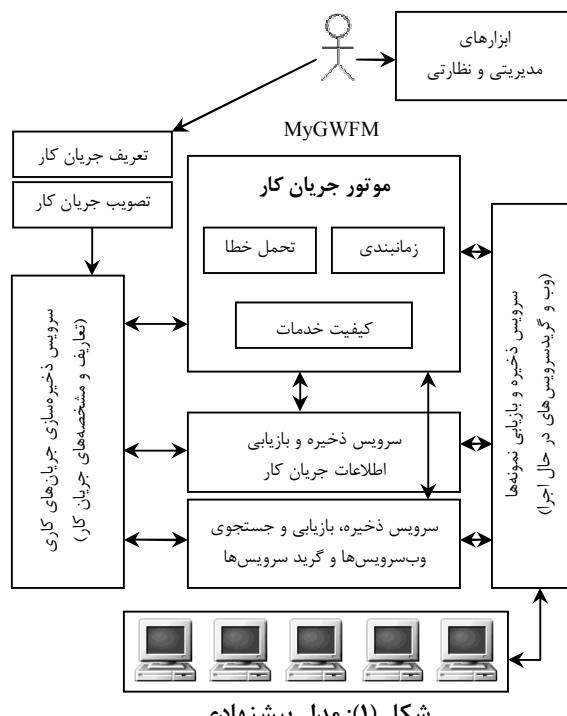
در این بخش از مقاله سه محور اصلی فعالیت‌های صورت گرفته در راستای پیاده‌سازی مدل پیشنهادی به شرح: زبان تعریف جریان کار، پیاده‌سازی موتور جریان کار و پیاده‌سازی وب‌سرویس‌ها مورد بررسی قرار می‌گیرند.

۱-۱- زبان تعریف جریان کار

تعریف وظایف، نحوه ارتباط و اجرای آن‌ها، پارامترهای ورودی و خروجی و مشخصه‌های موردنظر در اجرای هر وظیفه از مقوله‌های مهمی هستند که در زبان تعریف جریان کار می‌بایست مورد توجه قرار

میان افزار، ابزار تعریف جریان کار و ابزارهای مدیریتی و نظارتی در بخش کاربری مدل پیشنهادی قرار دارند؛ سرویس‌های ذخیره/بازیابی تعاریف جریان‌های کاری، سرویس ذخیره/بازیابی اطلاعات مربوط به جریان‌های کار (در حال اجرا)، سرویس ذخیره/بازیابی/جستجوی وب-سرویس‌ها و سرویس ذخیره/بازیابی نمونه‌ها بخش گردیدی سیستم مدیریت جریان کار علمی مبتنی بر وب‌سرویس را تشکیل می‌دهند.

شکل (۱) مدل پیشنهادی و مولفه‌های آن را نمایش می‌دهد.



شکل (۱): مدل پیشنهادی

- میان‌افزار MyGWFM^۸ حسب تعیینی که کاربر با استفاده از برچسب‌های موجود در زبان تعریف جریان کار انجام داده است به انتخاب منابع مناسب و مقتضی مبادرت می‌ورزد، کاربر می‌تواند معیارهای کیفی مدنظر خود را در اجرای جریان کار تعیین و بدین-وسیله میان‌افزار را در پیش‌بینی تمهیدات لازم در برآوردن منابع موردنظر خود، توجیه نماید. میان‌افزار با ارائه یک مکانیزم کنترل خطای متمنکز در صورت بروز خطا، آن را مجدداً زمانبندی و برای اجرا روی منابع مناسبی ارسال می‌کند.

- ابزار تعریف جریان کار؛ ضمن بهره‌گیری از کارکردهای این ابزار می‌توان به تعریف و تصویب جریان‌های کار مبادرت ورزید.

- ابزار مدیریت و نظارت؛ پس از ارسال جریان کار به منظور اطمینان از صحت و سقم همچنین کنترل و نظارت بر روند اجرا می-توان از این ابزار بهره جست.

- سرویس ذخیره/بازیابی تعاریف جریان‌های کاری؛ وب‌سرویسی است که عملیات حذف، اضافه، تغییر و بازیابی جریان‌های کاری تعریف شده را فراهم می‌آورد.

سرвис‌های پیش‌بینی شده محل(های) استقرار وب‌سرویس را یافته و وظیفه را پس از اعمال معیارهای کیفی و انتخاب ماشین مناسب جهت اجرا ارسال می‌نماید. انتخاب یک متده مشخص از میان متدهای موجود در یک وب‌سرویس توسط برچسب نام سرویس (ServiceName) و Parameters و پارامترهای ورودی آن توسط برچسب پارامترها (ParameterItem) و ParameterItem (انجام می‌شود. برچسب‌های فایل ورودی (ErrorFile)، فایل خروجی (OutputFile)، فایل خطا (InputFile) و فایل ثبت وقایع (LogFile) به ترتیب مسیر فایل ورودی، خروجی، ثبت خطاهای احتمالی و ثبت وقایع نظیر ثبت مدت زمان اجرای روال و نظایر آن را فراهم می‌آورد.

معیارهای کیفی در ارائه خدمات نیز توسط برچسب‌های حداقل حافظه موردنیاز بر حسب مگابایت (MemorySize)، حداکثر میزان فعالیت پردازنده بر حسب درصد (CPUUsage)، حداقل فرکانس کاری پردازنده بر حسب مگاهرتز (CPUSpeed)، معماری پردازنده (CPUArch) و نوع سیستم عامل (OSType) تعیین می‌گردد.

در مرجع [۱] نمونه‌های متنوعی از تعریف جریان کار به انضمام گراف‌های آن‌ها آمده است.

۲-۳- پیاده‌سازی میان‌افزار MyGWFM

میان‌افزار ضمن ارائه یک رابط گرافیکی برای انتخاب و ارسال جریان‌های کاری، سه مقوله‌ی به هم طنینده‌ی زمانبندی، تحمل خطا و کیفیت ارائه خدمات را در برنامه‌ریزی‌های اجرا مدنظر قرار می‌دهد. در صورت بروز خطا در اجرای هر یک از وظایف، میان‌افزار ضمن فرخوانی سرویس ذخیره/بازیابی/جستجوی وب‌سرویس‌ها، ماشین یا ماشین‌هایی را که وب‌سرویس درخواستی را دارند شناسایی و پس از حصول اطمینان از تحقق معیارهای کیفی و رتبه‌بندی مجدد، منبع مناسب را انتخاب و وظیفه را جهت اجرا ارسال می‌نماید. پیاده‌سازی زمانبندی به روش متمرکز صورت پذیرفته است. زمانبند پس از دریافت جریان کار و استخراج وظایف، ماشین‌های مناسب را بر حسب معیارهای تعیین شده در تعریف جریان کار رتبه‌بندی و منابع مناسب را جهت اجرا انتخاب می‌کند. در خصوص انتقال فایل و سایر نیازمندی‌های مرتبط به جهت استقرار سیستم بر اینترنت و وب‌سرویس می‌توان از پروتکل‌های وب و وب‌سرویس نظری UDDI، SOAP، HTTP و GFTP^۸ و مانند آن بهره جست. صورت نیاز از پروتکل‌های گرید نظری VS.NET 2003 و Microsoft .NET شرکت میکروسافت و مبتنی بر سکوی ۱.۱ Microsoft.NET است.

۳- پیاده‌سازی وب‌سرویس‌ها

مدل پیشنهادی بر وب و گردی‌سرویس‌های موجود روی منابع ناهمگون توزیع شده در اینترنت بنا نهاده شده است و وظایف مختلف موجود در

گیرید. هر جریان کار می‌تواند مستقیماً از تعدادی وظیفه و یا در یک ساختار سلسله‌مراتبی از تعدادی زیر جریان کار و نهایتاً وظایف درون آن‌ها تشکیل شده باشد با هدف تسهیل در تعریف جریان کار و وظایف داخل آن یک قالب مبتنی بر ایکس‌آل با دستوراتی که در ادامه آمده، پیش‌بینی شده است. قالب کلی هر جریان کار توسط زبان تعریف جریان کار پیشنهادی در شکل (۲) آمده است.

```

<GWFName Name="Grid WorkFlow Name">
<WFDate>PublishDate</WFDate>
<WFAuthor>AuthorName</WFAuthor>
...
<GSubWF Name="Grid Sub WorkFlow">
<Iterate>Number of Iteration</Iterate>
<TaskName Name="Task Name">
<TaskType>Par|Seq</TaskType>
<ServiceURL>http://WebSite/WebServiceName.ASMX</ServiceURL>
<ServiceName>MethodName</ServiceName>
<InputFile>Input File URL</InputFile>
<Parameters>
<ParameterItem>Parameter 1</ParameterItem>
<ParameterItem>Parameter 2</ParameterItem>
...
<ParameterItem>Parameter n</ParameterItem>
<Parameters>
<OutputFile>Output File URL</OutputFile>
<ErrorFile>Error File URL</ErrorFile>
<LogFile>Log File URL</LogFile>
<MemorySize>Minimum Memory Size(MB)</MemorySize>
<CPUUsage>Maximum CPU Usage(%)</CPUUsage>
<CPUSpeed>Minimum CPU Speed(MHZ)</CPUSpeed>
<CPUArch>CPU Architecture</CPUArch>
<NetworkLoad>Maximum Network Load(%)</NetworkLoad>
<OSType>Operating System</OSType>
</TaskName>
...
<TaskName> ... </TaskName>
<GSubWF>
...
<GSubWF> ... </GSubWF>

```

شکل (۲): قالب کلی تعریف جریان کار در زبان پیشنهادی

عنوان جریان کار (GWFName)، یک عنوان دلخواه برای جریان کار است. عنوان مذکور به همراه اطلاعات مربوط به تاریخ ایجاد (WFDate) و نام طراح یا طراحان جریان کار (WFAuthor) در بانک‌های اطلاعاتی ذخیره و توسط سرویس‌های پیش‌بینی شده در مدل پیشنهادی مورد بهره‌برداری قرار می‌گیرند؛ با استفاده از برچسب زیرجریان کار (GSubWF) می‌توان ضمن تعریف زیرجریان‌های کار امکان سازمان‌دهی و گروه‌بندی تعدادی وظیفه را جهت اجراهای تکراری فراهم آورد. برچسب تکرار (Iterate) تعداد دفعات اجرای یک وظیفه را معین می‌نماید. وظایف موجود در هر جریان کار توسط برچسب نام وظیفه (TaskName) تعریف و مشخصه‌های موردنظر برای آن‌ها تنظیم می‌گردد. نحوه‌ی اجرای یک وظیفه (TaskType) به یکی از دو صورت متوالی (Seq) و یا موازی (Par) تعیین می‌شود، با بررسی این مشخصه می‌توان گراف جریان کار را استخراج و در زمانبندی اجرای وظایف مورد استفاده قرار داد. اجرای هر وظیفه با اجرای یک وب‌سرویس همراه است که برچسب نشانی محل استقرار و نام وب‌سرویس (ServiceURL) آن را تعیین می‌نماید؛ چنان‌چه محل استقرار وب‌سرویس به صراحت معین گردد میان‌افزار وظیفه را روی ماشین مذکور اجرا می‌کند و در غیر این صورت با بهره‌گیری از

مشخصه‌ی [Resource] که در بخش تعریف وب‌سرویس ایجاد شده است یک WS-Resource ایجاد می‌شود. برای نمونه‌های مختلف از وب‌سرویس نیز WS-Resource های مختلف تولید می‌شود. برای ذخیره‌ی حالات WS-Resource ها در WSRF.NET هر یک از بانک‌های اطلاعاتی MS SQL Server و Xindice^{۱۴} قابل بهره‌برداری است. لازم به ذکر است سرویس‌های معرفی شده در مدل پیشنهادی ضمن مراجعه به این بانک‌های اطلاعاتی وضعیت نمونه‌های در حال اجرا را استخراج می‌نمایند [۱۱].

۴- ارزیابی

در این بخش با انتخاب موتور جریان کار کوندورجی [۲] و ارجاع یک جریان کار علمی (عملیات خوشبندی^{۱۵}) به ارزیابی مدل پیشنهادی و موتور پیاده‌سازی شده خواهیم پرداخت.

کوندورجی متشکل از دو بخش به نامهای مدیریت کارها و مدیریت منابع است. مدیریت کارها توسط ماشین ارسال^{۱۶} و مدیریت منابع توسط ماشین اجرا انجام می‌شود. کوندورجی از ابزار گلوباس^{۱۷} برای رفع نیازهای زیربنایی خود در دسترسی به منابع گردید، احرار هویت، اجرای برنامه‌ی راه دور و انتقال داده‌ها استفاده می‌کند.

۴-۱- ارزیابی نظری

در زیر برخی معیارهای نظری در ارزیابی پیاده‌سازی صورت گرفته بر مبنای مدل پیشنهادی آمده است.

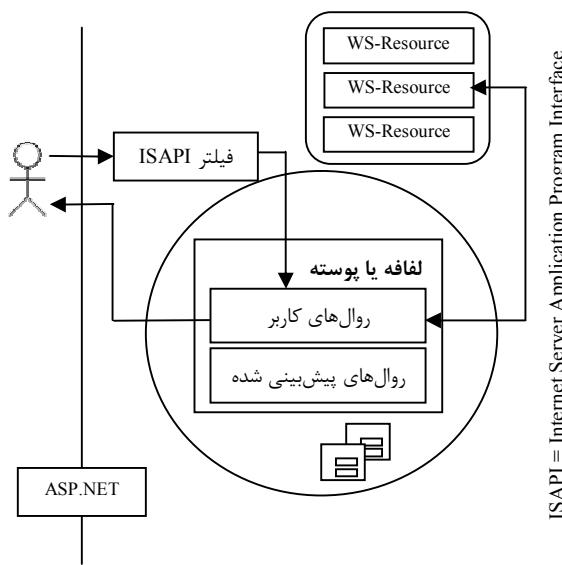
- به جهت استاندارد بودن رابطه‌ای تعاملی و استقلال از بستر محدودیتی در فرآخوانی و برقراری ارتباط با وب‌سرویس‌ها وجود ندارد؛
- این ویژگی مهم در سیستم‌های مدیریت جریان کار گردیدی از جمله کوندورجی دیده نمی‌شود.

- در مدل و زبان پیشنهادی امکان تعریف جریان‌های کار با تعداد وظایف نامحدود و متنوع وجود دارد در حالیکه این امکان در برخی سیستم‌های مدیریت جریان کار وجود ندارد، برای نمونه در کوندورجی در هر تعریف جریان کار صرفا یک وظیفه آن‌هم به صورت فایل اجرایی جهت اجرا ارسال می‌شود. در مدل پیشنهادی فایل یا سرویس منتقل نمی‌شود بلکه ماشین(هایی) که وب‌سرویس موردنظر را دارند در روال اجرا حاضر می‌شوند.

- عدمهای سیستم‌های مدیریت جریان کار گردیدی مکانیزم‌های متنوع جهت ارتقای کیفی اجرا ندارند در حالی که در مدل و زبان پیشنهادی برچسب‌های مختلفی برای این منظور پیش‌بینی شده است.
- تسهیل در انتقال فایل به دلیل استقرار سیستم بر وب و پروتکل‌های آن و استفاده از فایل‌های با قالب ایکس-ام-ال از دیگر مشخصه‌های سیستم پیشنهادی است به نحوی که برای مثال در کوندورجی به دلیل عدم وجود مکانیزم انتقال فایل (به استثنای انتقال فایل برنامه و فایل‌های استاندارد ورودی، خروجی و خطای) با محدودیت-ها و مشکلاتی مواجه هستیم.

جریان‌های کار صرفا با فرآخوانی این وب‌سرویس‌ها اجرا می‌شوند؛ همچنین جمع‌آوری اطلاعات مربوط به تمهید کیفیت خدمات با فرآخوانی تعدادی وب‌سرویس روی هر یک از منابع صورت می‌پذیرد. ابزار توسعه‌ی وب‌سرویس‌ها محصول VS.NET 2003 شرکت میکروسافت و نرم‌افزار زیربنایی [13] WSRF.NET طرح‌ریزی شده در دانشکده‌ی علوم کامپیوتو دانشگاه ویرجینیاست.

کنسرسیوم جهانی وب^۹، وب‌سرویس‌ها را مولفه‌های نرم‌افزاری فقد حالت و گذارا معرفی نموده است، بدین‌ترتیب وب‌سرویس‌های استاندارد برای به کارگیری در محیط گردید مناسب نیستند، این امر زمینه‌ی معرفی گردیدسرویس‌ها (وب‌سرویس‌های دارای وضعیت و غیرگذرا) و مشخصه‌های OGSI را فراهم آورد، چالش‌ها و اختلاف‌نظرها بین طراحان وب‌سرویس و گردیدسرویس در ژانویه ۲۰۰۴ با تشکیل یک تیم مشترک متشكل از مهندسین و طراحان IBM و انتلاف گلوباس و با معرفی WSRF پایان یافت؛ چارچوب جدید که در زمینه‌ی توسعه‌ی سیستم‌های نرم‌افزاری با قابلیت پشتیبانی گردید ارائه شده است به بازیبینی بسیاری از مفاهیم موجود در OGSI پرداخته تا از این گذار زمینه‌ی توسعه‌ی نرم‌افزارهای جدید سازگار با وب‌سرویس‌های امروزی را فراهم آورد. WSRF.NET محصول مبتنی بر مشخصه‌های WSRF و مبتنی بر سکوی.NET میکروسافت است [13]. وب‌سرویس‌های پیاده‌سازی شده در این قالب متشكل از دو بخش روال‌های پیش‌بینی شده^{۱۰} و روال‌های کاربر می‌باشند. شکل (۳) ساختار وب‌سرویس، روال درخواست و سرویس‌دهی وب‌سرویس‌های سازگار با WSRF و تحت WSRF.NET را نمایش می‌دهد.



شکل (۳): فرخوانی یک وب‌سرویس WSRF.NET

درخواست کاربر از طریق سرویس اطلاعات اینترنت^{۱۱} دریافت و در اختیار ASP.NET قرار می‌گیرد، چنان‌چه وب‌سرویس در سایت ثبت شده باشد (این امر از طریق Lafaveh^{۱۲} که هر وب‌سرویس هنگام تولید ایجاد می‌کند صورت می‌پذیرد) نمونه^{۱۳} ای از آن ایجاد و به ازای هر مشخصه‌ی [Resource] که در بخش تعریف وب‌سرویس ایجاد

In.Dat	A.Out (Cluster3)	B.Out (Cluster10)	C.Out (Cluster14)	Out.Dat
6	9	16	64	6 > 3
12	81	4	4	12 > 10
13	100	9	1	13 > 14
4	1	36	100	4 > 3
3	0	49	121	3 > 3
9	36	1	25	9 > 10
16	169	36	4	16 > 14
0	9	100	196	0 > 3

شکل (۴): نمونه‌ای از فرآیند دسته‌بندی سه خوش‌های

محیط عملیاتی برای اجرای جریان‌های کار ده کامپیوتر متصل به اینترنت با آدرس‌های IP مختلف، در یک شبکه محلی با دو سیستم عامل ویندوز (2000، XP و 2003) و Fedora Core 3 می‌باشد. پروتکل ارتباطی ماشین‌ها TCP/IP و همگی از پردازنده‌های اینتل با سرعت‌های محاسباتی مختلف و از رده‌ی پنتیوم IV بهره می‌جویند؛ همچنین خطوط ارتباطی شبکه محلی از نوع Fast Ethernet می‌باشد.

داده‌های آزمایشی برای شبیه‌سازی عملیات خوش‌بندی درون فایل In.Dat و به تعداد ۱۰۰۰۰۰ عدد صحیح ثبت چهار رقمی می‌باشد. از آن جا که دسته‌بندی داده‌ها در قالب یک خوش‌های مفهوم چندانی ندارد، کار را با دسته‌بندی دو خوش‌های آغاز و تا دسته‌بندی ده خوش‌های ادامه می‌دهیم.

تعریف جریان کار دسته‌بندی دو خوش‌های برای کوندورجی و موتور پیشنهادی به ترتیب در شکل‌های (۵) و (۶) آمده است.

```
# تعریف جریان کار مربوط به ایجاد فایل‌های میانی
executable = calcsubsqr
globuscheduler =
http://192.168.100.51:8080/jobscheduler
universe = globus
input = in.dat #ک
output = outa.dat
arguments = 2006
queue
input = in.dat#دو خوش
output = outb.dat
arguments = 1385
queue
# تعریف جریان کار مربوط به ایجاد فایل‌های نهایی
executable = calcfinal
universe = standard
output = out.dat
arguments = outa.dat outb.dat
```

شکل (۵): جریان کار دو خوش‌های در کوندورجی

- عدم وجود همیاب^{۱۸} مناسب در سیستم‌های مدیریت جریان کار گردیدی و از جمله کوندورجی باعث شده است تناسب منطقی میان درخواست‌ها و منابع برقرار نگردد در حالی که در مدل پیشنهادی میان-افزار ضمن فراخوانی وب‌سرویس‌های مستقر روی ماشین‌های گردید منابع مناسبی را جهت اجرای وظایف انتخاب می‌کند.

- از آن جا که در کوندورجی و اغلب سیستم‌های مدیریت جریان کار گردیدی فایل‌های اجرایی جهت اجرا به ماشین مقصد منتقل و در آن جا اجرا می‌شوند با مقاومت‌هایی از سوی کاربران مواجه هستیم از این رو لازم است نگرانی کاربران در حفظ حریم شخصی شان به‌نحوی مرتفع گردد که این مهم ضمن بهره‌گیری از وب‌سرویس‌ها میسر گردیده و در جلب کاربران و پیوستن آن‌ها به گردید تاثیر به‌سزایی خواهد داشت.

جدول (۱) مقایسه‌ی میان سیستم‌های مدیریت جریان کار گردیدی با سیستم پیشنهادی را نمایش می‌دهد.

۲-۴- ارزیابی عملی

به منظور ارزیابی عملی ضمن انتخاب یکی از مقوله‌های پرکاربرد علم آمار به نام دسته‌بندی یا خوش‌بندی داده‌ها جریان‌های کار را ایجاد و جهت اجرا روی دو سیستم مبتنی بر مدل پیشنهادی و کوندورجی ارسال خواهیم نمود.

در ادامه و قبل از تعریف و اجرای جریان‌های کار برای نمونه سناریوی دسته‌بندی سه خوش‌های را معرفی می‌کنیم. فرض کنید می-خواهیم داده‌های موجود در فایل In.Dat را بر حسب اعداد ۱۰، ۱۴ و ۱۶ خوش‌بندی کنیم (دسته‌بندی سه خوش‌هایی)؛ برای این منظور سه ماشین A، B و C هر کدام مربع تفاضلات داده‌های فایل ورودی را به ترتیب برای ۱۰، ۱۴ و ۱۶ محاسبه و در فایل‌های A.Out، B.Out و C.Out می‌نویسند (شکل ۴). سپس ماشین M، ضمن بررسی فایل ورودی، فایل داده‌های کلاستر شده (Out.Dat) را تولید می‌کند. روال تولید به این شکل است که ماشین M اولین سطر از فایل ورودی را خوانده، قدر مطلق تفاضل داده‌ی مفروض (عدد ۶) را با مقادیر سطر اول A.Out (عدد ۹)، B.Out (عدد ۱۶) و C.Out (عدد ۶) محاسبه می-کند سه عدد حاصل مقایسه و نتیجه‌ی کمتر بیان گر تعلق عدد ۶ به آن کلاستر است و بدین ترتیب عدد ۶ به کلاستر ۳ تعلق دارد. ستون آخر در شکل (۴) نتیجه‌ی این خوش‌بندی را نمایش می‌دهد.

همان طور که پیشتر ذکر شد، تعداد ده ماشین در اجرای جریان-های کاری حاضر بودند، از آن جا که در مراحل اولیه زمانبندی در موتور پیشنهادی امکان انتخاب ماشین‌های مناسبی (از جهت کیفی) جهت اجرای جریان کار وجود دارد مدت زمان اجرا در موتور پیشنهادی در مقایسه با کوندورجی به مراتب کمتر است زیرا میان افزار در انتخاب منابع بهتر و مناسب‌تر، دست بازنگردی دارد. این وضعیت رفتارهای خودشده را و به جهت اشتغال ماشین‌ها تغییر کرده است به نحوی که در دسته‌بندی‌های ۹، ۸ و ۱۰ خوش‌های زمان اجرای جریان کار به مراتب افزایش و به زمان اجرای کوندورجی نزدیک شده است. علت بروز این پدیده را می‌توان در عدم اقبال زمانبند در انتخاب ماشین‌های مناسب جستجو نمود، زیرا به مرور زمان که تعداد خوش‌های زیاد می‌شود ماشین‌ها به وظایف اختصاص می‌یابند و از این‌رو جامعه‌ی هدف زمانبند برای انتخاب منابع موردنظر کوچک و کوچکتر می‌شود. برای نمونه برای انتخاب نهمین ماشین جهت انجام محاسبات مربوط به خوش‌های نهم، در جامعه‌ی پیشنهادی گردید تنها دو ماشین برای ارجاع جریان کار در اختیار زمانبند می‌باشد. در نتیجه با افزایش تعداد کامپیوترها می‌توان جامعه‌ی هدف متنوع‌تری را در اختیار زمانبند گذاشت و بالطبع اجرای کیفی‌تر جریان‌های کاری را احصا نمود.

۵- نتیجه

با رشد روزافزون تکنولوژی گردید شاهد توجه بیش از پیش محافل علمی، تحقیقاتی و صنعتی به مقوله‌ی محاسبات گردیدی هستیم. گردید و محاسبات گردیدی مبتنی بر وب‌سرویس به انضمام مفاهیم مدیریت جریان کار می‌تواند ساختارهای سنتی سیستم‌های مدیریت جریان کار را با ارائه‌ی کارکردهای جالبی نظری امکان اجرای برنامه‌های پیچیده‌ی در مقیاس بزرگ، کارایی بالا، قابلیت تحمل خطای بالا و تسهیم و توزان بار کاری دستخوش تغییر قرار دهد. ما ضمن ارائه‌ی مدلی برای سیستم مدیریت جریان کار گردید علمی مبتنی بر وب‌سرویس به پیاده‌سازی و ارزیابی عملکرد آن در مقایسه با کوندورجی پرداختیم و تسریع در اجرای جریان‌های کار، تقليل و متناسب‌سازی هزینه‌ها، بهره‌گیری بهتر از منابع و سازگاری بیشتر با محیط‌های عملیاتی را مشاهده نمودیم. ما می‌کوشیم در فعالیت‌های آتی خود سیاست‌های تسهیم و تعادل بار را بهبود بخشیم و با افزایش دامنه‌ی برچسب‌های زبان پیشنهادی، کیفیت ارائه‌ی خدمات را متنوع‌تر و کارآمدتر کنیم؛ همچنین سربارهای ناشی از انتقال کار، هزینه‌های استفاده از منابع و توسعه و مقیاس‌پذیری زمانبند را در زمان اجرا مدنظر قرار دهیم.

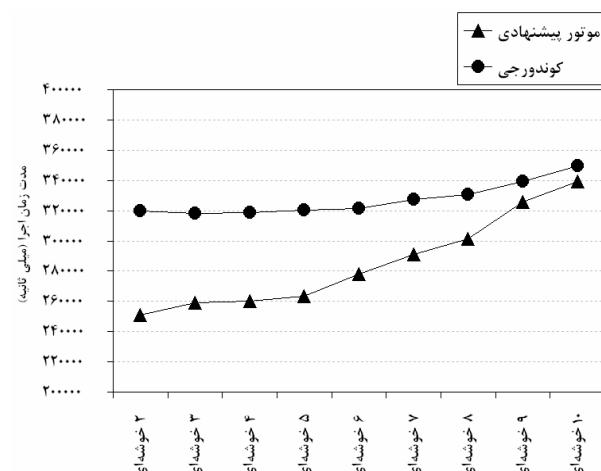
```

<GWFName Name="Clustering">
<WFDate>13850610</WFDate>
<WFAuthor>Hassan Mahdikhani</WFAuthor>
<GSubWF Name="CalculateSubSquares">
  <TaskName Name="CalculateSubSquaresA">
    <TaskType>Par</TaskType>
    <ServiceURL>clustering.asmx</ServiceURL>
    <ServiceName>CalcSubSqr</ServiceName>
    <InputFile>http://192.168.100.51/clustering/in.dat</InputFile>
    <Parameters>
      <ParameterItem>2006</ParameterItem>
    </Parameters>
    <OutputFile>http://192.168.100.51/clustering/a.dat</OutputFile>
    <ErrorFile>http://192.168.100.51/clustering/err.a.dat</ErrorFile>
    <LogFile>http://192.168.100.51/clustering/olog.dat</LogFile>
    <CPUUsage>10</CPUUsage>
    <NetworkLoad>5</NetworkLoad>
  </TaskName>
  <TaskName Name="CalculateSubSquaresB">
    <TaskType>Par</TaskType>
    <ServiceURL>clustering.asmx</ServiceURL>
    <ServiceName>CalcSubSqr</ServiceName>
    <InputFile>http://192.168.100.51/clustering/in.dat</InputFile>
    <Parameters>
      <ParameterItem>1385</ParameterItem><ParameterItem></ParameterItem>
    </Parameters>
    <OutputFile>http://192.168.100.51/clustering/b.dat</OutputFile>
    <ErrorFile>http://192.168.100.51/clustering/errb.dat</ErrorFile>
    <LogFile>http://192.168.100.51/clustering/logb.dat</LogFile>
    <CPUUsage>10</CPUUsage>
    <NetworkLoad>5</NetworkLoad>
  </TaskName>
</GSubWF>
<TaskName Name="MakeFinalClusterFile">
  <TaskType>seq</TaskType>
  <ServiceURL>http://192.168.100.51/clustering.asmx</ServiceURL>
  <ServiceName>CalcFinal</ServiceName>
  <Parameters>
    <ParameterItem>http://192.168.100.51/clustering/a.dat</ParameterItem>
    <ParameterItem>http://192.168.100.51/clustering/b.dat</ParameterItem>
  </Parameters>
  <OutputFile>http://192.168.100.51/clustering/out.dat</OutputFile>
  <ErrorFile>http://192.168.100.51/clustering/err.dat</ErrorFile>
  <LogFile>http://192.168.100.51/clustering/log.dat</LogFile>
</TaskName>
</GWFName>

```

شکل (۶): جریان کار دو خوش‌های در مدل پیشنهادی

روال تعریف جریان کار تا ده خوش‌های ادامه و میانگین نتایج حاصل از سی بار اجرا به ازا هر جریان کار روی دو موتور پیشنهادی و کوندورجی در شکل (۷) به نمایش در آمده است.



شکل (۷): نمودار مقایسه‌ای خوشبندی روی موتور پیشنهادی و کوندورجی

جدول (۱): مقایسه پروژه‌های جریان کار گردیدی با مدل پیشنهادی [۱]

نام پروژه	پیشنهادی	نمایندگی	ساختار	مدل سازی (ضمونی/صریح)	معیارهای کیفی	مجوز
کوندور جی	Condor, GT2	DAG	ض	دارد	General Public License(GPL)	
پگاسوس	Condor, GT	DAG	ض	دارد	Globus Toolkit Public License(GTPL)	
گرید آنت	Apache, Ant, GT2,3	Non-DAG	ص	دارد	GTPL	
تریانا	GAT, Web Service, GT	Non-DAG	ض	دارد	Apache Software License	
گرادرس	GT, NWS[8], MPI[8]	DAG	ض	مدت زمان اجرا	Not yet available in public	
تاورنا	Java1.4+, Web Service, BioMoby[7]	DAG	ض/اص	دارد	GNU Lesser General Public License(LGPL)	
کپلر	Java, SRB, GT	Non-DAG	ض/اص	دارد	UC Berkeley License	
گریدیاپس	GT2	DAG	ض/اص	حداقل هزینه و مدت زمان اجرا	GPL	
پیشنهادی	Web Service, WSRF	Non-DAG	ض/اص	دارد	GPL	

- [11] Fahringer, T., Jugravu, A., Pllana, S., Prodan, R., Jr, C., Truong, H., “ASKALON: a tool set for cluster and Grid computing”, Concurrency and Computation: Practice and Experience, Vol. 17, pp. 143-169, Wiley InterScience, UK, 2005.
- [12] Yang, X., Hayes, M., Usher, A., Spivack, M., “Developing Web Services in a Computational Grid Environment”, IEEE Services Computing Conference (SCC'04), pp. 600-603, 2004.
- [13] Wasson, G., Beekwilder, N., Morgan, M., Humphrey, M., “WS-ResourceFramework on .NET”, In 13th IEEE International Symposium on High Performance Distributed Computing (HPDC-1304), pp. 258-259, 2004.
- [14] Majithia, S., Shields, M., Taylor, I., Wang, I., “Triana: A Graphical Web Service Composition & Execution Toolkit”, IEEE International Conference on Web Services (ICWS'04), pp. 514-524. IEEE Computer Society, 2004.
- [15] Jin, Y., Wu, Z., Deng, S., Yu, Z., “Service-Oriented Workflow Model”, 19th IEEE International Conference on Advanced Information Networking and Applications (AINA'05), Vol. 2, pp. 484-488, 2005.
- [16] Workflow Management Coalition. Reference model and API specification: WFMC-TC00-1003, 1996, <http://www.wfmc.org/standards/docs/tc003v11.pdf>.

مراجع

- [۱] مهدیخانی، حسن، معماری یک سیستم مدیریت جریان کار مبتنی بر گرید و وب سرویس بر بستر گرید، کارشناسی ارشد، علم و صنعت ایران، ۱۵۱، اسفند ۱۳۸۴
- [2] Thain, D., Tannenbaum, T., Livny, M., *Condor and the Grid, Grid Computing: Making the Global Infrastructure a Reality*, NJ, USA, John Wiley & Sons Press, 2003.
- [3] Foster, I., Kesselman, C., Tuecke, S., “The Anatomy of the Grid: Enabling Scalable Virtual Organizations”, International Journal of Supercomputing Applications and High Performance Computing, Vol. 15, pp. 200-222, MIT Press, 2001.
- [4] Foster, I., Vöckler, J., Wilde, M., Zhao, Y., “Chimera: A Virtual Data System for Representing, Querying, and Automating Data Derivation”, 14th International IEEE Conference on Scientific and Statistical Database Management, pp. 37-46, Scotland, UK, 2002.
- [5] Deelman, E., Blythe, J., Gil, Y., Kesselman, C., Mehta, G., Patil, S., Su, M., Vahi, K., Livny, M., “Pegasus: Mapping Scientific Workflow onto the Grid”, pp. 158-165, Grids Conference 2004, Nicosia, Cyprus, 2004.
- [6] Kacsuk, P., Dózsa, G., Kovács, J., Lovas, R., Podhorszki, N., Balaton, Z., Gombás, G., “P-Grade: A Grid Programming Environment”, Journal of Grid Computing, Vol. 1, pp. 171-197, Kluwer Academic Publishers, 2003.
- [7] Oinn, T., Addis, M., Ferris, J., Marvin, D., Senger, M., GreenWood, M., Cover, T., Glover, K., Pocock, M., Wupat, A., Li, P., “Taverna: a tool for the composition and enactment of bioinformatics workflows”, Bioinformatics, Vol. 20, pp. 3045-3054, Oxford University Press, 2004.
- [8] Berman, F., Chien, A., Cooper, K., Dongarra, J., Foster, I., Gannon, D., Johnsson, L., Kennedy, K., Kesselman, C., Mellor-Crummey, J., Reed, D., Torczon, L., Wolski, R., “The GrADS Project: Software Support for High-Level Grid Application Development”, International Journal of High Performance Computing Applications, Vol. 15, pp. 327-344, SAGE Publications Inc., UK, 2001.
- [9] Laszewski, G., Amin, K., Hategan, M., Zaluzec, N., Hamton, S., Rossi A., “GridAnt: A Client-Controllable Grid Workflow System”, 37th Annual Hawaii International Conference on System Sciences, pp. 210-219, IEEE Computer Society Press, CA, USA, 2004.
- [10] Altintas, I., Berkley, C., Jaeger, E., Jones, M., Ludaescher, B., Mock, S., “Kepler: Towards a grid-enabled system for scientific workflows”, Workflow in Grid Systems Workshop in GGF10, March 2004.

زیرنویس‌ها

-
- ^۱ Concrete
 - ^۲ Abstract
 - ^۳ Globus
 - ^۴ Grid Application Toolkit
 - ^۵ QoS
 - ^۶ My Grid WorkFlow Middleware
 - ^۷ Factory/Instance
 - ^۸ Grid File Transfer Protocol
 - ^۹ W3 Consortium
 - ^{۱۰} Imported Method
 - ^{۱۱} Internet Information Services
 - ^{۱۲} Wrapper
 - ^{۱۳} Instance
 - ^{۱۴} A WAR file run over Tomcat web server
 - ^{۱۵} Clustering
 - ^{۱۶} Submit Machine
 - ^{۱۷} GT (Globus Toolkit)
 - ^{۱۸} Matchmaker