# Cluster and Grid Computing in Japan: Today and in 2010

Report author: Dr. Robert B. Cohen, Project Co-Director, and Fellow, Economic Strategy Institute[1]

Executive Summary: Grids Will Create an IT Revolution

We found that Japanese firms are likely to adopt grid computing in the following way:

1. The first wave is the early adoption of clusters and grids in single firms, largely during the 2002-2005 period. During this period, most firms will use clusters of computers in single facilities to capture extra compute cycles and use compute resources they already own more efficiently. Clusters will use both computers and data storage devices more efficiently.

2. The second wave of adoption is likely to occur from 2006 to 2009, when corporations begin to link various groups within the firm into a corporate or Enterprise grid. This grid will probably be characterized by a substantial expansion in the number of applications and services that can be run by utilizing grid computing. In this phase, more efficient applications are run in grids. In addition, links between grids, web services, and the semantic web are likely to add greater intelligence and sophistication to computations that rely upon grids.

   It is important to note that this phase appears likely to be very influenced by models that vendors are promoting. In Japan, many firms depend on vendors to define transitions that they can make from one system of computer use to another. During the course of this study, we found a number of firms that were delaying their adoption of Enterprise Grids but were planning to move through some of the transitional phases that vendors were planning with them so that they could adopt grids more widely.

   The most apparent change we witnessed was a number of firms going through a period of computer and data center consolidation. Vendor software made it easier to "virtualize" compute resources and data storage so that users could gain access to them regardless of where they were and this promoted the consolidation moves. For instance, some users with a large number of data storage sites consolidated their data centers into a few locations, usually four or five principal places. After this, it was easy for the equipment vendors to help these firms use these resources more efficiently through software that helped to virtualize the data stores. At the same time, the firms could achieve important economies in the use of data stores by not having to maintain as many data storage locations as they did previously and being able to reduce the number of employees they needed to support these storage centers.

   In this wave, vendors will probably employ two different Grid computing models:

   a. An Application Service Provider (ASP) model, where computer vendors offer customers a chance to run applications on computers that vendors or other firms own, so that they do not need to rely upon their own computing resources. This model is a vendor driven approach to using grids rather than an approach that is shaped by corporate users. The ASP approach can include an extensive networks of

[1] The following people are also affiliated with the project:
Satoshi Sekiguchi, Project Co-Director, and Director, Grid Technology Research Center,
National Institute of Advanced Industrial Science and Technology, Japan
Satoshi Itoh, Project Co-Director and Acting Director, Grid Technology Research Center, and Grid Diversification Team Leader, National Institute of Advanced Industrial Science and Technology, Japan
Robert Triendl, National Institute of Advanced Industrial Science and Technology, Japan

computers that encompasses millions of PCs owned by consumers or businesses that a vendor will use to form a grid computing environment.

b. An Enterprise Grid model where individual firms interconnect many parts of their corporate operations. This allows firms to rationalize their use of compute and data resources, giving different groups access to compute power or data storage wherever it resides in the corporation. This new Enterprise Grid also permits critical applications to be run across the enterprise.

3. In the <u>third wave</u> of grid adoption, Partner Grids create a “collaboration environment” or “services infrastructure” where users throughout the firm and in partner firms or firms that are linked to a corporate supply chain use Partner Grids to access the applications used in Enterprise Grids. The third wave incorporates greater management capabilities over an Enterprise Grid’s resources, allowing firms to manage knowledge, identify resources, and gain greater control over complex business processes. In this wave, significant improvements in business efficiency are likely because Partner Grids support collaboration between organizations that design and test new products and help marketing and business development groups reduce costs and open new market opportunities. Customers will also benefit from enhanced services and support. Much of the third wave is likely to be collaboration environments as well as services infrastructure. Both become more widespread after 2008. Alternatively, there is also a good chance that Japanese industries known for close ties between suppliers and large final product producers will have some features of the third wave earlier, perhaps by 2006. Industries that adopt collaboration environments very early are likely to include autos, computers, heavy industry and, perhaps, pharmaceuticals. In the third stage, grids benefit from access to a wider range of devices and the ability to exchange a broad range of information from many organizations. They also should benefit from Web services that support new interactions with supply chain members and robust networks built to support grid interactions.

That there is a “three wave” pattern of grid adoption is corroborated by industry studies of cluster and grid adoption that we prepared. We asked computer and middleware vendors, as well as the individual firms we interviewed, to provide a picture of grid adoption in their own industry and, for the vendors, industries they knew well or where they had customers. The responses permitted us to create figures depicting the likely pattern of cluster and grid computing adoption in several major industries. These figures reflect the three waves of grid adoption, beginning with clusters, moving to Enterprise Grids, and then evolving to include Partner Grids. Although some firms we interviewed mentioned that there would be exceptions to this pattern in the use of Enterprise Grids and Partner Grids, most respondents included the three waves in their answers.

The Logic of moving to Grid Computing from Clusters

Many different types of firms have begun to use clusters to speed their R&D and design processes. We were able to interview executives involved with these product development operations (i.e. design, R&D, etc) and technical computing users at a series of Japanese firms.

Since there may be some concern that executives in these positions do not have a full view of how their corporations will use clusters and grids in the coming years, questions could be raised about the validity of the conclusions this analysis draws. In order to address such questions, the project took the following steps to gain a broader picture of the likely evolution of grid computing. First, it interviewed a number of equipment and software vendors to get a broader idea of the adoption pattern that was likely to characterize early adopter firms and industries. Second, we asked questions about the evolution of cluster and grid, particularly enterprise grid use, of executives in early adopter industries where the executives felt comfortable discussing the likely pattern of adoption over the next four or five years. In some of these industries, clusters were at the heart of development processes that are major cost or innovation centers within the firm. We believe that the evolution from clusters to Enterprise Grids was likely to be described rather accurately by the executives we interviewed because they had already been involved in the planning process or in early stage discussions about expanding the firm's grid. Third, the adoption pattern that interviewees described to us made sense from an economic standpoint. The main process that they described was the following:

1. An initial "department-oriented adoption" or learning phase, where relatively small changes were made in how computers and storage devices were used in the firm. This usually involved identifying ways to complete designs or finish R&D more rapidly than was the usual standard.
2. A more substantial "corporate-wide adoption" phase that included some changes in business processes, but usually in a more limited way. For instance, a firm involved in R&D might link several different research centers and create the infrastructure required to share resources among these centers. As a consequence, R&D could be done faster and the process could draw upon a wider group of resources within the firm. These benefits were usually expected to be gained during the creation of an Enterprise Grid, although a few firms raised the prospect that some of these gains might be achieved once firms deployed Partner Grids that linked a few groups at one corporation with their counterparts at a supplier firm.
3. A broader "inter-corporate adoption" phase, where corporations extend their Enterprise Grids more widely between geographically dispersed groups within the firm. This phase also includes examples such as firms that begin to deploy Partner Grids to work more closely with their suppliers or collaborators. During this phase, collaboration becomes easier than has been true during earlier phases of IT use.

This logic for adoption implies that there is an increasing ability for adopting firms to obtain economic benefits from using clusters and grids. This logic seems to follow a pattern that consists of:

1. Initial benefits for specific groups and departments that provide for greater utility or functionality of processes that are known and common. These often include design processes, where computers are employed to estimate the parameters of a design or to test how well it works. For instance, an automaker usually used computers to estimate how new car designs might perform in crash tests, often creating a physical model that can tested to provide data on how the framework, bumpers, protective devices and other design characteristics perform when there is a crash. With clusters or grids, this testing can be analyzed more rapidly and the firm can often create a computer model of the new car design that can be put through the same tests that would be used if there was a physical model of a car.

   The benefit to the design group is that it can test many more modifications to the design in a short period of time. It can also identify shortcomings in the design and modify them before they are carried forward to the production phase and consult with suppliers about how key parts and components for the new design may need to be altered to insure that the design works properly. Thus, the use of clusters can shorten the time needed to test a new design, reducing the firm's costs to create the design because the time needed to go from design to production can be shortened. This can make it easier for an auto firm to design profitable cars than in the past when it relied upon physical crash tests and it took longer from design to actual production.

2. Second-stage benefits may be larger in scope and provide even more utility than specific departments can achieve when they are the main users of clusters and grids. If we take the case of the automaker that we cited above, when auto firms "virtualize" their compute and storage resources so that three or four design can share them, new efficiencies are added to the gains identified above. These gains are very likely due to ending the groups' use of independent groups of computers and separate databases that were not linked to each other. Thus, the economic gains in this stage are due largely to networking benefits because technology that had not been employed on a network can be "networked." As a result, special databases and separate computers are joined as a single computer or single database, making collaboration much easier. This provides additional utility for the firm. The gains are also more likely to affect a larger research or design group based in several locations rather than in just one location.

3. Third-stage benefits are likely to be even larger than those in stage two because they will affect the firm and collaboration with its partners. Most likely these benefits will be gained because the firm can exploit even greater external economies. Previously, collaborations between the firm and its partners or suppliers were not as efficient because the compute power and storage resources of each firm were physically separated from each other. With grid computing and the ability to create partner grids that link resources from one firm to another, the external economies that both firms benefit from are expanded. Thus, one would expect that there are even greater economic benefits that are joint benefits – shared between two firms – that result from Partner Grids or other means of creating collaboration between two firms.

To repeat the different types of grid computing developments that we explore in this report, we are focusing on:

1. Clusters – grid computing splits up a single application to run on many computers or processors at the same time. Clusters of different types of computers or grid computers provide the hardware to run these split up applications. This is usually a first step in the use of grids.
2. Data grids – this is a way to access geographically distributed data resources. Data grids provide a way to collaborate and share data and information resources. Many firms with dispersed research groups were the first to use data grids, such as pharmaceutical firms.
3. Enterprise Grids – provide a way to share compute and data or information resources within a single firm, usually linking together several different operations and locations within the firm.
4. Partner Grids – create a way to share compute and data resources between at least two unrelated firms, where one firm may be a supplier to another or a collaborator with another. Partner grids nearly always connect these resources beyond the boundaries of either enterprise, crossing corporate firewalls. Most firms have not yet developed plans for Partner Grids, but the Japanese history of close relationships between suppliers and final manufacturers suggests that some firms may deploy Partner Grids very soon.

We did find a few industries where the extremely close links between suppliers and final manufacturers might result in a more rapid deployment of department–to-department Partner Grids. In the case of the auto, heavy equipment and pharmaceutical industries, specific departments that have long-established relationships with each other may deploy Partner grids linking themselves to one another. This may create a series of links between corporations that have historically been parts of supply chains, with the department heads rather than the CIOs pressing for these connections. We also heard that there were industries in Japan where there might not be Enterprise Grids during this decade. This was mostly due to a concern about sharing sensitive information, particularly in the financial industry.

When we examined industry patterns for grid deployment, we were able to identify differences in the speed with which industries were expected to adopt grids. Firms in some industries were deploying clusters and grids faster than those in other industries. We categorized these two groups as:

1. "Early adopters"—the auto, computer, heavy equipment, and semiconductor industries

    a. With 45% or more of the firms in an industry expected to have clusters by 2005 and

    b. With nearly 20% or more of the firms in an industry expected to have Enterprise Grids by 2007

2. "Later adopters" – the banking and financial services, construction, distribution, and pharmaceutical industries.

a. With more than 25 percent of firms in an industry expected to deploy clusters by 2007

b. With more than 25 percent of firms in an industry likely to deploy Partner Grids by 2009.

3. *The main difference*: “later adopter” industries take two to three years longer to adopt clusters, Enterprise Grids, and Partner Grids than “early adopter” industries do.

The rapid adoption of grid computing will have a dramatic impact on costs and business processes by the end of this decade. Nearly every user firm and vendor we interviewed indicated that there would be important cost savings in the first two to three years of cluster and grid computing use. Over a longer time period, firms expected to achieve even more significant savings by using Enterprise Grids for improved internal collaboration, largely because Enterprise Grids let firms optimize compute and data resources across a corporate network. Thus, the firm can begin to use multiple data centers and compute resources as if they are a single computer and data center. When different groups are working on a single project but need to access a widespread number of databases and computers, this creates efficiencies.

In addition, when the Enterprise Grid is extended to become a Partner Grid there is “seamless integration” between a producer and its suppliers. This not only makes supply chains easier to manage, but also creates an opportunity to establish standards across firms. As an executive from a middleware vendor noted, this improves the quality of products across the supply chain because a manufacturer can help its suppliers implement Six Sigma and maintain a high level of performance. This reduces the amount spent on product maintenance and support dramatically.

We quantified the probable savings and productivity gains that specific industries would achieve over the next six years as a result of adopting grid computing. This analysis shows that:

1. Most early adopter industries will achieve significant cost savings over expected cost levels in the next six years. The pharmaceutical industry will save the most, between 11 and 17 percent of costs, although it is not one of the earliest adopters of grid technology. Two industries that are early adopters, autos and heavy equipment, will save between 7 and 10 percent of their costs by 2010 as a result of adopting grids. The semiconductor, computer, construction, and financial services industries will also reduce their costs by using grids, but not as much as the other early adopters.

2. Productivity gains due to the use of grid computing were even more substantial. We measured these gains as a percentage difference compared to productivity levels expected in 2010. For the pharmaceutical industry, productivity increases a result of grids are expected to be between 15 and 23 percent greater than forecast. For the auto and heavy equipment industries, productivity gains are expected to be 10 to 15 percent greater than forecast. For the construction industry, productivity gains are expected to be 8 to 12 percent greater than the forecast levels.

3. “Early adopter” industries will very likely increase their sales by adopting grids. Sales increases ranged from a large increase in computer industry sales, between 10 and 14.5 percent greater than expected in 2010 to gains of between 8 and 12 percent for the auto and heavy equipment industries. For “later adopter” industries, the sales gains in 2010 were more modest, ranging from between 7 and 11 percent for pharmaceuticals to between 6 and 10 percent for construction. There were even more modest gains for the semiconductor and financial services industries.

By 2010, many computer vendors will include grids as a major part of their commercial product offerings. These vendors estimated that the size of the cluster and grid computing market in Japan would be:

1. By 2006-7, between $2 billion and $4 billion (four different vendors)

2. By 2010, between $4 billion and $9 billion (two vendors).

The results for cost savings, productivity gains and sales increases at the industry level indicate that the adoption of clusters is only a part of a longer evolution to a grid computing fabric in the business world. The initial gains will be important, but will be enhanced substantially as firms establish Enterprise Grids that facilitate internal and external collaboration. Why are the latter gains so much larger? Because they begin to alter the business processes in the firm, resulting in dramatic savings. In addition, collaborative environments permit firms to manage information and processes better, so they can shorten the time it takes to get products to market.

Japanese firms may also establish these collaborative environments by relying upon computer vendors to deploy and manage them. This would differ from more traditional Enterprise Grids in the following ways. If computer vendors are Application Service Providers for large corporations, they will run many of the complex computer applications that firms use by setting up the compute and data resources. This will mean outsourcing the computations and the control over data. While this would reduce the need for IT departments and overall IT costs, it may raise issues of security and control over proprietary data that are significant for some sectors, particularly for pharmaceutical and financial firms.

If larger collaborative networks are established, they will require higher speed communications than many firms are using today. In one case, a firm that is probably using no more than 155 Mbps speeds noted that it would need about 10 Gbps speeds between corporate centers to support an Enterprise Grid. Other firms also noted a substantial increase in communications speeds would be required for collaboration. Thus, a major impact of the move to grids, particularly by 2008 or 2009, is likely to be a significant increase in the broadband connectivity between corporations or between locations within a single firm.

In sum, grid computing will change corporations in ways that we are only beginning to imagine. Besides enhancing compute power, grids are likely to become a key infrastructure for business operations, particularly for collaborations. Given the importance of relationships between suppliers and producers in Japan, the impact of collaborative environments based upon grid computing may stimulate greater efficiency and productivity gains in Japan than in nations where the supplier-producer ties are not so longstanding.

We have also estimated how grids might affect Japan’s economy. Our initial estimates show that the use of grids might:

1. Raise Japan’s GDP by 3 to 6 percent

2. Increase real disposable income by 2 to 3 percent, and

3. Increase aggregate productivity by 2 to 3 percent over levels that the economy is expected to attain by 2010.

In a broader sense, grids may spur greater use of new services and service environments to support more efficient production and service development. These environments will have four main effects. They will:

1. Support greater collaboration between parts of the corporation and between corporations.

2. Place the use of services and application based analysis at the heart of innovation within the corporation. The new environments will facilitate “service creation” and greater interaction between applications.

3. Transform the infrastructure used by corporations into one that is based upon grid computing, Web services and grid services that can prioritize the use of scarce applications and resources and create an economic rationale for computing and analysis within the firm.

4. Permit management to have greater knowledge of the processes that are key to product development, service offerings and operations. There will be far greater control over the sales and costs that contribute to the corporate “bottom line.”

Since grids permit the capital invested in businesses to become more productive, this will “deepen” the capital invested in machines and in humans. By enhancing the efficiency of existing and new investments, such a capital “deepening” should increase factor (labor and machine) productivity, particularly in firms that deploy grids. This could contribute to a new phase of economic expansion in Japan, especially in industries that have adopted grids and achieved cost savings and productivity gains like the ones that we have estimated.