Thank you very much for your introduction, and let me thank Kennedy Information for inviting me to address you. In the end, it is you who take all these technologies and translate them into the kinds of offerings that bring real value to business, society and all of our personal lives.

In the last several years we have seen huge technology changes, and even more important we have seen the major impact that those technologies have had all around us.

One of the major questions that we have been asking ourselves is, where is this all heading? What are some of the major developments that will cause the next big round of changes.

In particular, given that we all associate the huge changes over the last five years with the Internet, let's focus the question a little more and ask ourselves, "Where is the Internet going and what is the next big thing that will emerge from it?"

Today the Internet is a collection of capabilities that began evolving from its inception. The first one that came on the scene was basic networking, a result of agreeing on a common set of protocols -- TCP/IP,for example -- so that systems which before could not talk to each other could now begin to communicate.

Next, the Internet became a major communications platform because we adopted a universal way of sending e-mail, just as now we are on our way to adopting a universal way for instant messaging. As a result, we now rely more and more on the Internet for the bulk of our communications.

Of course, what really brought the Internet into the mainstream -- into the world of business -- was the advent of the World Wide Web. That's when we learned that by adopting a set of relatively simple standards, we could get access to information everywhere.

By roughly the mid-nineties or so, we all realized that the connectivity and access of the Internet, when coupled with classic information technologies, could have huge value for businesses of all sorts. That's when e-business emerged as a strategy, a strategy that continues to play out in the marketplace today.

So, what's next? Where is this progression of Internet capabilities taking us? One way to address that question is to find out what's happening out there in a number of key areas: what's happening in the marketplace, what's happening with the technology, and what are some of the major cultural forces driving the Internet.

Now, in the marketplace when I talk to our customers, these are some of the main things that they tell us are very important to them.

One, they love the technology but they're concerned about cost. So sharing the technology, making its use as efficient as possible, is a prime requirement.

Technology may be getting inexpensive, but it's not that cheap. In fact, the less expensive it gets the more wonderful applications people develop to drive even more technology. So I don't think we'll ever get to a point where the demand to "make it work more efficiently" will disappear.

The second thing is that people have really accepted that the Internet is for real, and therefore they actually expect the thing to work.

A few years ago amidst all the excitement, we could always tell people, "Look, this is about the neatest thing you could ever see. So what if it's slow and goes down all the time. I mean, that "industrial-strength" technology routine is for pedestrian IT applications like, you know, ATM applications."

Well, those days are over. People expect the Internet to be exciting on top of working with total security, total availability, and total industrial-strength, as they say.

The third major requirement that everybody wants is flexibility. This is really important because, given the rate at which new technologies and applications are showing up, if every time you get a new technology or a new application you need to redesign your system, it's all over.

People expect systems that can accommodate new capabilities. They want to just plug them in, integrate them on the spot, and go on about their business.

And finally, one of the most important things people want is what I would call freedom of deployment. I think one of the main things many firms do when it comes to technology is worry about architectural deployment: “Do I want to go distributed? Do I want to go centralized? Do I run it myself? Do I outsource it to somebody else?”

And the reason people worry so much and need two-year task forces to make decisions on those things is that, once you make the decision, you are pretty well trapped.

It shouldn’t be that way. People should be free to adjust so that, if one’s architecture is too distributed, it can be readily centralized in a few weeks, or at least a few months; but not years.

Or, if you're running an in-house application and decide you are better off outsourcing it, you should be able to call up a firm and, in a matter of a few weeks, have them running it for you. It shouldn’t be a very, very big deal as it is today.

These are some of the major drivers that we hear about from our customers in the marketplace. But clearly a second major driver is technology. George Stalk talked very eloquently about what's happening with Moore's Law.

Moore’s Law is really as exciting as gets if you're a technologist. It's almost scary because it means that every year the technology will improve at a 50, 60, 70 percent compound growth rate, year in and year out.

Those are big numbers. We are all so used to them that we take them for granted. But remember, 60 percent a year means that over five years the technology is 10-times better, 10-times more different, 10-times more valuable

And it's not enough that Moore's Law predicts these straight line improvements, some technologies like storage and backbone bandwidth are improving at what we call a super-exponential rate.

Super-exponential means that when you plot it on semi log paper you don't get a straight line; the line starts curving upwards -- meaning that today it may be 57 percent; next year perhaps 62 percent; the year after, it may be 67 percent. So technology is clearly a huge driver of everything we see happening.

But, I really believe that the biggest impact on the Internet and the IT industry -- what years from now people will write books about -- is the cultural change underway now.

For example, before the Internet, if you proposed a project in which vendors had to get together and agree on standards for the benefit of their customers, everybody would roll their eyes in disbelief and say, “Why don't you go do something else?” People can still name project after project, from UNIX to OSF, that have been tried and failed.

But the Internet introduced a whole culture of standards and of collaboration by demonstrating that, if we all use the same protocol, TCP/IP, we could connect. And, if we add protocols like http and html, we could share information freely.
That changed the way we think about the future. It led to a massive movement in important areas like open source technology, like the Linux operating system. Not surprisingly, it's one of the main drivers of the next stage of the Internet.

When we put all the major forces together -- the market demand for the Internet to be an extremely efficient, reliable, flexible infrastructure; the need to absorb all the technologies coming our way; and the cultural standards that are allowing to more and more integrate all the pieces -- where does it leave us?

It leaves us with an Internet that is itself becoming a computing platform, as is already manifest in the research community. It's called the Grid and you will I'm sure hear more and more about it over the year.

Now, what does it mean for the Internet to become a computing platform? That's one of those big statements that I suspect make most people scratch their heads and wonder, "What is he telling me?"

Well, let's look at a computer. A computer is a collection of resources. It has a microprocessor, it has memory, it has storage, it has I/O devices. And it has data, files, and applications.
These are all packaged into something we call a computer. And when you turn it on, if it's your personal computer, you get a desktop that shows you the kinds of things you can do . . . these are the types of applications you can access. Essentially, that's a computer.

What does it mean for the Internet to become a computing platform? What we mean is that, when you boot up your favorite client device, you'll see a collection of resources at your disposal.

But instead of those resources being in your local computer or even in a server computer that you are connected to, those resources could now be distributed all over the Internet.

So, what you really have, in the language of technology, is a virtual computer accessible by you because you have the right security credentials.

And that virtual computer -- just like your PC -- has an aggregation of compute power, storage, applications, data, I/O devices and so on distributed all over the Internet. And that vast aggregation of resources is clearly much, much, much richer

than anything you would find in a single computer, even a very large server. That's all we mean by the Internet itself becoming a virtual computer.
Clearly achieving this is not easy. This is not the latest strategy of the moment that we'll build and leave behind six months from now.

This is what, again in the language of the technical world, we call a Grand Challenge. In fact, this is one of the grandest of challenges that we have in the IT industry. It's a challenge that the computer industry, and especially the computer sciences community, has faced for many years.

We called it different things, but basically what we were talking about was the challenge of distributed computing. And that means how do you take computing resources that are scattered all over the place and integrate them, manage them, and take advantage of them as if they were one huge, virtual computer.

Let me talk about some of the major efforts going on to make this a reality, three things in particular.

First, where are we in the Grid, where is the development going on and what is its status? And then I'll talk about the relationship of Grid computing to some incredibly important strategies that we in IBM and the industry are pursuing -- that is, the world of self-managing systems that we call Autonomic Computing and the world of e-utilities, or e-sourcing.
Let me start with the Grid. In this model of virtual computing what you are doing is taking distributed computing resources and sharing them. For example, if you are an enterprise and own lots of servers, you want to be able to share the computing power of all those servers. You want to be able to share the storage of all those servers. And, most important, you want to be able to share the data and the applications in all those servers.

Now, with them all available over the Internet, all the people that you're working with can share those same resources. In fact, the community that is developing Grid computing thinks its biggest impact will be the ability to develop virtual organizations that come together around this infrastructure and all its resources.

The people driving this effort right now are in the research community, in the NSF centers and the universities. That is the community to which attention must be paid for a very simple reason.

If you look at the people who brought about the previous incarnations of the Internet -- whether it is TCP/IP or whether it is universal e-mail, or most important, the World Wide Web -- it all came from the research community, from the university community, from the government labs. And then, over time, it moved into the wider commercial world.
So one could do a lot worse trying to figure out where the world is heading in the next few years than to look at what's hot in the research community right now. And Grid is very hot in the research community right now.

Let me give a few examples of engagements that we in IBM have had with the research communities just in the last few months.

In early August, we announced that we were going to help put together the UK national Grid. Initially the UK national Grid is taking the supercomputers at a number of major universities in the UK . . . at Oxford, Cambridge, Edinburgh, Southhampton and others . . . and linking them together over high bandwidth so that those who are allowed to use those computers no longer have to worry about where the computer is -- at Oxford, or Cambridge or Edinburgh. They simply make use of any of the resources of the Grid. And all of the resources are at their fingertips. That is clearly a far more efficient use of resources.

But over time, the UK government wants to open up these Grids to the private sector, starting probably with the more technically oriented companies -- pharmaceutical companies, chemical companies, engineering design firms, and over time, to other kinds of commercial ventures.

Then we announced a huge deal in the US, a contract the National Science Foundation awarded in August, called the Distributed Terascale Facility or TeraGrid.

This TeraGrid is a gigantic supercomputing research initiative that will have about 13 and a half teraflops worth of processing power and close to 700 terabytes of storage spread across the University of Illinois Supercomputing Center, the University of San Diego Supercomputing Center, the Argonne National Labs and Cal Tech.

All those systems will be connected via a 40-gigabit per second fiber optic pipe provided by Qwest. By the way, one of the consequences of the “infinite” bandwidth that George talked about is that suddenly you can do things -- like connecting systems that are scattered around the country -- things that were unimaginable when you didn't have 40 gigabits per second.

The TeraGrid is going to be the most powerful aggregated computing facility in the world, and it's going to be used for research across all kinds of disciplines.

Now, access to computing power and storage and networks is very important but the really big aspect of Grid computing is the ability to share data that is scattered across all those sites. After all, it is the ability to share data, and to share the applications that process that data, that leads to the virtual organization.

Let me illustrate that by talking about an announcement that we made yesterday. It’s a project that we announced with the University of Pennsylvania in the lead and working with a number of other labs and universities here and in Canada.

It will create a national repository of mammographic data stored digitally rather than on very expensive film.
By storing it digitally, it becomes accessible over the network to people who have the proper security credentials.

Most important, there will be a whole set of advanced applications developed that can compare different mammograms in time to detect features that you would probably miss if viewing the mammogram at just one point in time. That's one of the advantages of having this stuff stored digitally.

It will have advanced applications running on a tremendously powerful supercomputing facility to aid the radiologist, or any other physician, in analyzing precisely what’s going on in a particular patient.

If multiple physicians, perhaps scattered across the country need to consult and view a particular mammogram, this Grid will give them access to the information as if they were in the same room of the same hospital. That's a very good example of the kinds of virtual collaboration and virtual organization that Grid computing is unleashing.

A number of research universities are pursuing a whole set of projects around sharing life sciences data.

For example, one of the main projects being pursued at the University of San Diego -- one of the key National Science Foundation centers -- is brain mapping research, which is absolutely critical to better understand and treat Alzheimer's, schizophrenia, Parkinson's, multiple sclerosis and a whole set of neurological diseases that we just don't fully understand now.

The problem with brain mapping research is that each MRI of the brain can take up a huge, huge amount of storage. Because of that and because you can only do research if you have lots of images of different brains for comparison, you need to do this as a collaborative effort between multiple centers, each of which can share the data with the others. That's the kind of Grid that is being put together to go tackle these major problems.

Now, Grid capabilities are actually also emerging in the commercial world. But in the commercial world, we have been using a different terminology -- which I suspect will start changing over time -- to describe how we best integrate applications with each other. It's called Web services.

When all is said and done, a lot of the capabilities that a number of companies like IBM, Microsoft, Oracle and others have been bringing to the table to integrate Web services applications are very much in the spirit of the Grid initiative, in that they are focused on sharing resources over the Internet, especially applications.

We are already doing work with companies like Galileo on opening up their reservations systems to allow travel agents to integrate their applications with the applications of Galileo in the simplest way possible. That's an example of a Web services-style Grid application.

Now, if you look at what has made the Internet successful in certain areas like the transport layer, the communications layer, the content layer, it's an agreement on a set of open protocols that everybody uses. And they are, for the most part, not just open but have an open source pedigree that enables people to take the code and recompile it on their systems.

Now, in order to realize this vision of Grid computing, we essentially are climbing up what we call the stack of protocols that are part of a computer. We're defining and agreeing on the protocols that allow different computers to share data; to share files, to tell each other about their computing resources; to federate their security systems; to tell each other whether they are alive or not, so that if something has failed the system can route around it.

So there is a very big initiative that the global Grid community is undertaking to define a set of protocols. You'll see lots and lots of activities over time.

This is a very complex undertaking, but the good news is that real examples of such systems are already in action in the research world, and they are beginning to move into the commercial world.

Now, usually you get a feel for the value of a major technology when it not only fulfills its own promise, but it becomes a major enabler of other key strategies that you've been trying to implement, perhaps for years.

That's happening with Grid computing. And let me talk about two such major strategies and their relationship to Grid computing and the value they each bring to Grid computing.

The first is a strategy that we actually conceived this year called Autonomic Computing -- the effort to bring about a world of self-managing systems and a self-managing infrastructure.

Let me digress a little bit to explain why we have to do this. The answer is not very complicated. If you look at Moore's Law and all this technology coming out, you can quickly convince yourself that while the technology is wonderful, it is the cost and complexity of managing all this technology that will eventually do everybody in.

Therefore, if in addition to inventing technology, you want to figure out how to make it valuable to the marketplace, you need to address head-on the issue of managing that technology at a reasonable price.

And when I say technology, I'm including all of the software, all of the applications, all of the storage, all the pieces of the infrastructure.

There is only one answer: The technology needs to manage itself. Now, I don't mean any far out AI project; what I mean is that we need to develop the right software, the right architecture, the right mechanisms...

...so that instead of the technology behaving in its usual pedantic way and requiring a human being to do everything for it, it starts behaving more like the “intelligent” computer that we all expect it to be, and starts taking care of its own needs.

If it doesn't feel well, it does something. If somebody's attacking it, the system recognizes it and deals with the attack. If it needs more computing power, it just goes and gets it, and doesn't keep looking for human beings to step in.

We call this Autonomic Computing because of the analogy to the autonomic nervous system, which is at the heart of all organisms, including humans, and the purpose of which is to keep us alive.

It's what we usually call the primitive functions of our organism. Primitive but critical, because you'll die, unless they keep you breathing and regulate your temperature and do all kinds of things that are needed so that you can concentrate on higher functions, like giving this talk, and not worry about breathing, controlling your temperature and other little, but vital, things like that.

And we are organizing our activities in Autonomic Computing around four key areas: self-optimization, self-configuration, self-healing and self-protecting. There's quite a bit of this that we have already in our individual systems.
And that's good. But in talking to our customers they said, "You know, Irving, it's really good for you to come and tell me how great, let's say, your zSeries mainframes are because of their autonomic capabilities; what I really need is the whole end-to-end infrastructure to manage itself." It doesn't do any good to have a superb server, if everything in between is very chaotic.

So we realize we have to extend autonomic capabilities across the Internet. And to do that, you need a set of open protocols that allow all of the various systems around the Internet that are participating in this autonomic initiative to be able to collaborate and share information, identify denial of service attacks, identify any failures ... and start routing things around.

This is where Grid computing comes in, because it's the open protocols of the Grid running on every single system that enable "intelligent" management capabilities to cut across different systems. Using these Grid protocols, we can realize this vision of creating, over time, an increasingly self-managing infrastructure.

And this is clearly one of the most important areas that we need to address. It’s at least as demanding and critical as creating the telephone infrastructure of the nation. And that could not have been accomplished without moving from manual operators connecting each call to the automated switching system that finally allowed people to connect without operators. This is a very similar kind of project, a little bit more complicated but with similar kinds of objectives.

The last major aspect of Grid computing I want to talk about is its impact on the world of “utility” computing.

We've been talking in the industry for a while about computing-on-demand, storage-on-demand, applications-on-demand, and the notion of application service providers.

A lot of progress has been made. And we in IBM have had quite a number of engagements. I could talk about work we are doing with Dow Chemical and Saks in procurement and others that have adopted this notion of outsourcing individual business processes.

But the reality is that there is a long way to go, that this world of outsourcing IT and creating IT utilities has been way too difficult.

The reason it has been too difficult is a lack of standards for connecting the computing resources and applications that you're outsourcing.

When you don't have standards, every engagement is a complicated, discrete, expensive and time-consuming manual effort. But when you have standards -- such as we have for, let's say, outsourcing Web sites -- everything happens much faster.

Well, once this vision of Grid computing, based on open standards, starts to play out and enables an enterprise to link together its various computing resources and applications over their networks ... it is a small step to then link to computing resources and applications that don't happen to be owned by you, that happen to be part of a utility.

So once you have those open protocols, you get the kind of deployment flexibility that we are after. And I believe this will absolutely open up the whole on-demand world -- the world of computing utilities.

It will be just so much easier for people to make decisions on what processes and applications are core to the business and should be kept in-house, while having pretty much everything else done by a vendor. And if that provider doesn't do a good job, since it's all based on standards, it’s an easy matter to switch to somebody else.

So, to summarize, there is an incredible amount of potential in this technology of Grid computing.

I feel a little bit as if I could have been addressing you in 1994 or 1995. It’s as if back then I were saying ...”You know, there is this thing called the Internet that the research world has been making a lot of progress with. And they are doing incredible things with it. I’m not sure exactly what its impact will be, but boy, this feels like a really big deal.”

I think the feeling now is similar, that the Grid is opening up a whole set of really important capabilities we have all been after for a long while. And now of course starts the hard work, and it's the work that all of you do so well, which is to take these technologies and translate them into real business value. Thank you very much.