0:00 slide 1	Hello and welcome, my name is Cynthia Babb and I'm part of the product marketing strategy team here at IBM. Today, we're going to be talking about managing the information lifecycle, from requirement to retirement, as part of an overall information integration in Governance Series here from IBM.
0:16 slide 2	So the next slide, what we'll be discussing today, are some of the challenges from managing the lifecycle of critical applications data. And what's at stake when this data is not effectively managed, how organizations can leverage an effective information governance approach, particularly around optimizing the lifecycle of that data. And we will cover areas around discover and defining that data, how to better develop and test, optimize, archive and access, as well as when to consolidate and retire at a price application and what to do with that data when that happens. And then I'll be talking through some customer success stories of how some of our clients have leveraged IBM solutions from InfoSphere Information Governance and how they were able to manage that data throughout its lifecycle.
1:12 slide 3	So in the next slide, you know, information can come from many places, transactional systems, operational systems, document repositories, external information sources, and in many different formats: data, content, streaming data. But wherever it comes from, there are often meaningful relationships between various sources of data. We manage all this information in our systems, integrate to build warehouses and then master the data to get single views and analyze it to make business decisions again to insure trusted, relevant and governed information. This is a supply chain of information shown here, flowing throughout the organization. However, unlike traditional supply chains, for example, a supply chain within let say a car manufacturing facility, where one windshield goes with one car, in an information supply chain, there's a many to many relationship, with information, the same bit of data about a person can come from many places. Here she may be a customer, an employee, a partner. And that information can end up in many reports, many applications, many databases. Those different systems may define information differently as well. So this makes integrating that information, ensuring it's quality and interpreting that information correctly crucial to using this information to make better decisions. Information must be turned into a trusted asset and governed to maintain the quality over it's information lifecycle.

	Underlying systems must also be cost effective, easy to maintain and perform well for the workloads that they need to handle, as information continues to grow year after year after year. So these are the needs that have driven IBM's strategies and investments in this area and that we'll be covering today.
3:12 slide 4	So on slide 4, to stay competitive, organizations need to look to improve business processes and better manage their information. But there are obviously some roadblocks along the way. So, what are some of those business risks around managing the information lifecycle? So one area is around new applications functionalities to meet the business needs of the organization may not be deployed on schedule, and this is something that organizations are often challenged by. And around these application releases, I'd like to say that software development and the airline industry have a lot in common. They're always on time until they are late. One reason for these delays is the creation and the management of test and development environments. This is often done simply by cloning production. But depending on the size of the production environment, this method can impede progress. How long does it take to create them? Refresh them? Are there developers waiting around, when their time can better used otherwise. Another area is around application defects discovered after deployment. How are you able to easily validate this test data to ensure applications are caught during that test and Q/A process. Or as much of the test and development times spent sifting through and fixing the data versus the application. Another area around poor data management challenges is around that increased operational and infrastructure costs that impact the IT budget. As data volume increase year after year, and this can be compounded by the cloning needed to support the test and development environment as I mentioned previously. Do organizations have enough disk storage? Do they have enough database licenses, for example, to create the needed non-production and production environments? How does this impact your IT staff resources? And how does this impact your system uptime for business users? So, one of the other things here is indicated by this quote here from Forrester, "many organizations today are keeping a lot of infrequently used data in they production environments, and this can have a negative impact on both performance and IT costs" as we'll see later in further slides.

| 5:30 slide 5 | So on the next slide, this is a survey done by IBM with regards to some of the data management pitfalls that some organizations experience, taking that reactive approach to managing the lifecycle of their information and taking action after the headaches begin. So here's some of the key areas that some of our clients have identified as catalysts if you will for their search of a better data management, information management approach.

One of the areas around high capital expenditures, you know reactive mode, often the solution is to add more high performance hardware to ensure that there is enough storage and to improve that declining performance that can be attributed to those ever growing volumes of data.

Another pitfall around decreased productivity, so poor application performance impacted by the business user's ability to perform their daily tasks. Perhaps batch processes are creeping into work hours. IT staff is frantically trying to tune databases and add more storage in response to this.

And this can lead to next pitfall, around missed service level agreements, which can impact a bottom line, that revenue and customer satisfaction if databases and applications are not responding as quickly as they should. And then with regards to ad hoc performance management, without a clear strategy in place to manage and optimize application performance, inefficient ad hoc fixes that were leveraged, can drain IT resources. As organizations strive to stay ahead of the competition, a more "proactive" approach to data management is needed to ensure the data is accessible and trusted. |
| 7:13 slide 6 | So on the next slide, you know there is a real organizational impact to poor data management, ineffective use of data management, if organizations are not able to immediately identify these warning signs as symptoms of uncontrolled data growth and poor data management. These are real impacts to the business so around enterprise application performing slowly, as production instances grow, so do the backups and those non-production systems that are often cloned. Managing large storage volumes can impact the IT staff productivity again the IT staff running around trying to tune databases to achieve more storage and try to put on more reactive modes solutions versus pro-active . And there is a potential liability as keeping data lying around beyond the data retention roles. And then how to access that data from those legacy and unsupported systems. And then finally, a quote that we hear often from some of our clients, every |

	time we turn around we seem to be buying more storage. Is that the appropriate fix for organizations which regards to effective data management? So, what are those requirements?
8:31 slide 7	So on the next slide, for today's discussion we're going to look at around four key areas to better managing the lifecycle of your data. And that is around Discover and Define, Develop and Test, Optimize, Archive and Access, Consolidate and Retire.
8:50 slide 8	So looking at the next slide, let's take a closer look at those requirements with regards to streamlining the management of data across its lifecycle.

So discover and define, understanding where data resides, what demands of information exists, how it's related across the enterprise and defining the policies and standards of pro-managing it.

Develop and test, creating the databases structures and their reusable database code to enhance productivity and team collaboration, efficiently creating the test and development environment that are needed, as well as protecting the sensitive data within those environment. As well effectively validating the test results and quickly and securely deploying applications that support the enterprise.

The next, optimize, archive and access. Ensuring that optimal application performance, archiving historical data to manage the data growth over its lifecycle and ensuring that business users have effective access to that archived data.

And then finally consolidate and retire. We are looking at your application portfolio, consolidating and decommissioning applications that are redundant for no longer align with the current technology but maintain access to the data for data retention role, long after the application has been retired. |
| 10:13 slide 9 | So on the next slide, let's talk about the first area, around discover and define. In order to define a governance strategy and a process to achieve your organization's goal, you need to have an understanding of what data you have within the enterprise. Without this you cannot create an effective plan to manage the data throughout its lifecycle.

So this process begins by understanding that web is information represented here across your enterprise applications and databases. So you need to understand where the data exists and what data elements there are, what relationships exists within systems, what complex relationships exists across and |

	between systems. You also need to have an understanding that complex relationships exist and where sensitive data can be located, as you are looking to manage this data throughout its lifecycle. So many organizations rely on documentation, but that is often outdated. Imagine using a map that's ten years old to be able to navigate through a new town, wouldn't you rather have the most up to date map to ensure you have the right streets and the right cross streets in order to get to where you are going. Same thing can be related back to ensuring you have the right documentation that understand the right relationship across your enterprise.
11:34 slide 10	So on the next slide, the solutions necessary for the process by which we locate and understand those data relationships, again being able to locate and inventory the databases across the enterprise. Again, you can't govern data if you don't know where it lives. So ensure your solution can help you discover and document the data entities and the databases that reside in the enterprise. Defining business objects or logical entities across heterogeneous databases & applications. Understanding how that data is related across the enterprise so you can better deploy new functionality and ensure that that logical entity is captured when archiving data. And looking at defining enterprise-standard data models, understanding transformation rules to discover data relationships across the enterprise, ensuring those relationships required for identifying sensitive data are there. And I'm looking at defining and documenting the data privacy & masking rules to ensure that sensitive data is protected throughout the enterprise. How is that data going to be used? Who should have access to it and why? And as you mask sensitive data in one table, how do you ensure all related data elements are masked with that same information, keeping that referential information intact.
13:01 slide 11	On the next slide, it's important to ensure you understand how data is related. How is it related across different tables, different

	databases, across the enterprise? And often it is not detected through a simple scan. So by leveraging discovery, you can then understand what makes up that logical entity or that complete business object, what makes up a "customer" record for example, even if it spans across different databases and systems.
13:30 slide 12	So on the next slide we get to the next pillar. So we talked about discover and define, now it's about develop and test. Now that we understand where data lives, developing and testing the applications that will manage that data effectively, we need to ensure that those applications are deployed quickly and effectively to support that trusted information. But when looking how many organizations do this today, there are areas that can negatively impact an organization's ability to keep up with the business processes and stay competitive. So let's look at some of those challenges and what some of our customers are saying here. So for example, Large US Healthcare Insurer said they didn't want to create fake or unrealistic test data. All that testing needed to be created and set up manually previously, so that took them more than a month to set up for those thirty or more accounts. So imagine the time sink associated with creating that test data to ensure that they have that ability to do the testing and development needed. Another customer needed to improve those efficiencies in development and testing environment, as well as within those production environments, being able to create realistic test environments that use much less disk space really impacted their IT budget. And then lastly, having their staff being able to implement a more efficient and cost effective task process, to shorten the time to create those multiple test environments. So again looking at how to more efficiently and effectively manage the IT staff time with regards to test and development, and also looking at the IT cost associated with managing the data throughout a test and development environment.
15:20 slide 13	So on the next slide, let's look how you can employ more effective testing and management practices to create those test and development environments. So again, for generating test data, often times organizations frequently clone their production environment. Well now you have live data in these multiple test environment. How are you ensuring that you have the storage capabilities, the storage resources, the license resources and the

	time to clone and refresh that data throughout the test and development process. And also how you are ensuring that the sensitive data within those cloned environments are protected.

A more effective way of creating test and development environments is creating more right sized sub-sets for those testing needs. Allowing testers and developers to easily sub-set, refresh and mask, the masked data to create properly sized data sets. And then after running these tests, being able to relationally compare results sets based on the new data and the actual production data to see the exact differences and this helping resolve application defects faster.

Now as I mentioned masking part of an effective testing and management strategy is the ability to protect that sensitive data within these non-production environments, so ensuring test data is masked while maintaining the referential integrity of that data, to ensure that that data transformation is appropriate to the context of that application. What I mean by that is, not only ensuring that you are protecting this sensitive data, maintaining the referential integrity throughout the test environment, but masking that data with realistic but fictional data. So to the tester and developer, it looks like real data, not excise notes, not hash marks, not fictitious data that's going to create problems. You don't want testers and developers to be testing the data, you want them to be testing the application. So ensuring that it is masked with contextually appropriate information is a key area as part of an effective testing and management strategy. |
| 17:47
slide 14 | On the next slide, so now we've discovered and defined and now we've also developed and test and deployed your application, we now need to manage the day-by-day of your data within that application.

But are you controlling your data, or is the data controlling you? Today's organizations are met with a number of primary challenges at this phase of the data lifecycle. Around application performance issues. Again, over time, data volumes will increase, your going to be getting more data, not less data year after year. So this can impact application performance. Without properly managing the data growth within those systems, companies will see the impact in the performance of their system over time. This is a particular problem when Service Level Agreements are in place that mandate set response times. So how can organizations pinpoint performance problems and ensure that those time frames are not lost. |

	Storage, license and maintenance costs. The costs of managing data spans across the initial design of the data structure throughout all the lifecycle phases. IT staff is under constant pressure to deliver more with less. A lot of major costs for managing data include not only the storage hardware costs, but the storage management costs, being able to archive, store, retrieve that data, and also looking at the costs associated with the draining IT resource allocation as they try to locate and manage those data growth issues. Again, these IT staff members should try to do more with less. So again, anything to help reduce the impact, reduce the resources associated with those costs and that storage drain can help an organization better manage that data lifecycle. And there is also risks associated with data retention and compliance. You need to safely store tat data that is no longer needed in production, but access to that data still must be retained per data retention policies. How can organizations store this data so that it's audit-ready and available for any type of e-discovery requests?
20:06 slide 15	So on the next slide, some of the requirements for more effective data management and how it is managed over it's lifetime. Your information governance strategy should incorporate the following elements. Proactive optimization of database performance. So as data volumes grow, archiving historical data out of production systems and managing that archive is critical. So it's also ensuring application independent access of that archived data, keeping the referential integrity of the data by archiving the complete business object or that logical entity is important. And then the third area around Flexible storage options based on the value of that data. For example, data that is one to two years old may be kept in the production systems. But after say 3 to 5 years old that data will be kept in the same active archive database that still provide full access to that data with mid-level performance. Then after maybe 5 to 10 years that data is stored in a flat file format that takes advantage of more cost effective storage options, etc. Having that flexibility helps organizations manage that data throughout its lifecycle.
21:23	So on the next slide, we were talking about archiving that

slide 16	historical data from the main from managing the data growth within the application. So when you look for archive solution, again, you want to ensure you have the flexible access to this data and flexible storage options. So how do you do that? Well, let's first define what we mean by archiving. It should be an intelligent process for MOVING inactive or infrequently accessed data that still has value, again remembering that Forrester's statistics there is about 75% of the data within production, critical applications, 75% of that data is often inactive or infrequently used. So, if that data can be, send it for archive. But the archive process should provide the ability to search and retrieve the data in a manner that functional users need to consume the data. Let me talk about that. So, this is a typical example of a production environment prior to archiving. Initially, both active and inactive of historical data is stored in production, taking up most of that, and that inactive data is taking up again about 75% of the space in the production server. By safely moving the inactive data to an archive, capturing that complete business object or logical entity, again they can copy of that reference data so that it makes sense for application independent access to that data later on. This data can then be stored in a variety of different environments. It can then also be easily retrieved to the application environment, it may be selectively restoring an entity from the archive. Say if this was a payroll system, and you archive past employee information, if an employee return to the organization, being able to selectively restore their contact information and put that back in the production, ensuring that your service solution that does that can be critical to the organization. And then you must also have the ability to have a universal access to this data through multiple methods, so some of the examples we have here are: IBM Mashup Center, perhaps the native application creating an archive instance of that application, through a Web search engine such as InfoSphere Optim Data Find, or through an ODBC or JDBC connection that can then be leveraged by third-party reporting tools, such as Cognos or Cristal Report
23:56	On the next slide, we've now come full-circle now, talking about

slide 17	the requirements of your data, applications, how to deploy and optimize applications effectively, and then how to access the data effectively throughout this lifecycle, whether it's for production or archive file. Now we come to what we have called the retirement phase, or to put it another way, application portfolio rationalization. So looking around your data center, hopefully it doesn't look like this old picture that is listed here, but there is likely 1 or 2, or perhaps more, systems that are merely, I'll say, "kept alive" because there might be important data stored within, but how will you access it? Have you really looked at what's in the data there? How old is that application? Is that system still supported by the manufacturer? Does anybody know what it does? Does anybody know how to access it? Could it be redundant with more current systems that your have within your application portfolio? And then, a key here is how much is this older application costing the organization, from licenses, power consumption, extended support agreements with a vendor, the data center footprint, etc? So, if the solution is to consolidate or retire the application or applications, the next big concern is what to do with the data. We don't want to simply move it all into a consolidated application, as this can cause system performance problems for that particular application. As we mentioned before, having a lot of information in production can have negative impact on that application's performance. So it may not be appropriate in the new context and could grow your data to be too large, but for business, governance and regulatory reasons, that information needs to be kept. So, the best scenario is to provide access to the data without relying on the cost of database software or servers and without relying on the original application to access it. All of that said, in most cases, we still need to get to the data, but independent of the original application.
26:10 slide 18	So in the next slide, when considering candidates for your consolidation and decommissioning projects, it is important to note that as much as 80% of historical reference data managed in application performance databases are rarely accessed, 75-80% as Forrester mentioned. However, data retention policies will apply, so you want to ensure your application retirement strategy takes into account the what, why and how long data should be kept and how it will be accessed. As you're retiring redundant applications, you may not want to move all data into a consolidated application, as again this could

	cause performance issues due to the amount of data. However, and again, you also want to look at those date retention policies often times you do need to keep it for extended lengths of time. By archiving the data and keeping that logical entity or complete business object intact, this data can be accessed independent of that retired application. By archiving the data, you will also retain that "historical snap shot" of that data for audit and e-discovery purposes. And by consolidating and retiring redundant or legacy systems, you can then improve the operational management and reduce the costs across your IT environment, including the hardware, software, network infrastructure, power, staff resources and more.
27:38 slide 19	On the next slide, the best scenario is to provide access to that data without relying on the cost of database software or servers and don't rely on the application to access it, again looking at application independent access of this data. So, for example, once data from similar business applications is consolidated and retired, a skilled database administrator can then redirect productive time toward implementing an ERP package, rather than maintaining a patchwork of databases. So there is a cost savings associated with reducing the infrastructure and maintaining archive access to that data. Another benefit is when you rationalize your infrastructure, you also reduce the complexity and therefore reduce the business risk associated with that IT infrastructure. For example, by consolidating a dozen homegrown general ledger applications into a packaged application, ERP solution, you can provide business-critical support and reduce the risk of missing key processing deadlines, such as a month-end close.
28:47 slide 20	So on this next slide, so we've talked about how to better discover and define, develop and test, optimize, archive and access that data. And we've been looking at consolidating and retiring, so then we've come full-circle from managing the information lifecycle from requirements to retirement. Let's look about what the market is been saying about how IBM has been helping our customers today to better manage the information lifecycle again from that requirements to retirement. You know, one of the things that IBM continues to try to help our customers to with a more comprehensive look at managing the

	data lifecycle from archiving solutions to testing management solutions, again being able to help organizations manage that data throughout of lifecycle. Again, these are some of the quotes in which some of the analysts and customers are saying today that really speaks to how we can help today organizations by their managed data and ensure that they are reducing the cost, the risk, the complexity and the time it takes to manage that data throughout of lifecycle.
30:00 slide 21	For the next slide, you know, in conclusion, IBM has looking to provide, you know, expertise in managing the information lifecycle for organizations today. IBM has a robust solution to manage the data throughout its lifecycle as part as your overall information governance strategy, so across both production and non-production systems. With IBM solutions and expertise, our clients are able to create and maintain trusted information, ensuring information is protected across the enterprise. I'd like to take you through a couple of customer success stories to demonstrate a little further on how we've been able to help organizations do that today.
30:45 slide 22	So in the next slide, this is an organization that recently had done an application upgrade, so they had a critical enterprise application that ran much of this mission critical system from payroll to manufacturing, to inventory, to customer service, that they needed to update to bring it up to this most current version in this case JD Edwards enterprise one application. One of the things that they needed to do was to ensure that they minimized the downtime of this business critical application. But because there were so much in their production environment, their IT staff to their business users with the recommendation of: "Look, it's going to take anywhere from 5 to 8 days of downtime for us to process this upgrade." Well, this is something that the business users, they couldn't be down, have their business critical systems down for a week. So what could they do in the meantime ? So one of the things that they looked to IBM to help them with, in leveraging their IBM InfoSphere Optim Data Growth Solution, they were able to archive historical, infrequently of inactively used data, archive that information out of their production JD Edwards system, thereby reducing the volume in their production environment. So when they did finally do the upgrade, they were able to that upgrade over the course of a 3 day week-end. So by shutting down on mid-afternoon on Friday, by Monday morning their business users were back up and running with their new

	upgraded version of JD Edwards. In this case, by archiving that data, they were able to speed the overall upgrade and data conversion time and minimize the downtime so that it did not impact the business ability to run the business each day. There was an added benefit to ensuring that not only did they not impact the business, but they also discovered that by storing that archive data on tier-2 storage, they also archived an estimated annual savings of $75,000 to $80,000 a year, just on storage cost alone.
33:11 slide 23	Now the next slide. Another area of managing data lifecycle is around speeding the test and development process for the organization and again around managing with data within those non-production environments. In this case, this company had leveraged typical practices of just cloning their production environment to just create their various test, development, UAT and training instances. They had approximately 10 or 12 test and development instances within their environment. And by simply cloning their production environment, they were managing and needed to refresh a large volume of data, which impacted their storage cost. By leveraging InfoSphere Optim Test Data Management Solution to create light size subset of that data and also and also masking that data to protect the sensitive data, they were able to reduce the sizes of their databases but almost 35 to 40%, and realizing a cross savings of about $240,000 a year by simply leveraging effective testing and management practices.
34:29 slide 24	And finally on the next slide, which refers to looking at consolidating and retiring your application portfolio. In this cases, the company had acquired not only a number of organization but also as you acquire organizations you acquire their IT infrastructures as well. So now this organization has set energy to these companies, they had several redundant instances of both legacy, customer applications, one of them again being redundant doing the same things throughout the enterprise, but also multiple versions of PeopleSoft Enterprise. So they still needed to support very large data retention policies, often times meaning to keep general electric data up to 50 years. So where did they put the data, they were keeping in production, so they were maintaining several redundant production environments, trying to manage and trying to keep up with application performance and application storage, really

extending the IT staff resources.

So, by leveraging InfoSphere Optim Data Growth Solution, they were not only able to archive historical information, but in some cases archive the data, effectively archiving the data in redundant and outdated applications. To retire those applications while still providing application independent access to that data, and so they effectively over 25 legacy applications as well as consolidated those multiple instances of PeopleSoft.

So they really reduced the infrastructure and the complexity of their application portfolio. And they were able to increase operational performance of their system as well as their IT staff to ensure that they were able to meet the service level agreements and support their business users effectively.

So in conclusion, I just want to thank you all for joining us today in talking about managing the information lifecycle from requirements to retirement. Hopefully you've gained information with regard to some of the best practices and effective practices that today's organizations are using today, to better manage their data from discovery to test and development, to archiving and to better consolidating and retire those applications across the enterprise.