AICPA CITP Credential Examination Series

Topic: Data Analysis & Reporting Infrastructures/Platforms

Presenters: Maria Pollieri and Adam Chaikin

Maria: Hello and welcome to the AICPA CITP Credential Examination series. This podcast will assist you in preparing for the examination specific to the Information Management - Data Management topic. By way of introductions, my name is Maria Pollieri and I am an AICPA CITP credential holder based out of Toronto, Canada, and I'm employed by Oracle Corporation. I am joined today by my peer and colleague, Adam Chaikin, who works with Thought2Execution.

We are pleased to share with you our insights around the topic of data management. More specifically we will be discussing the following five subject areas. Number one, types of infrastructure and platforms that are typically employed. Number two, data preparation and manipulation. Number three, data analysis, functions, tools, and approaches including business intelligence and analytics. Number four is information traceability for source and transformation and finally, number five is information quality.

Now let's begin. Subject area number one, types of infrastructure platforms that are typically employed. Before we can discuss data management we need to introduce you to the underlying technology or in colloquial terms, the "tech stack" that supports applications, which is where data is typically captured and maintained. At the very bottom of the tech stack is infrastructure, which typically consists of capabilities such as compute power, memory and storage. And infrastructure is typically the hardware required to support data management.

Infrastructure is evaluated based upon its ability to deliver high performance and high availability in a cost-effective manner. The infrastructure layer is where one would go to, to manage load balancing workloads and monitor data processing throughput. The next layer in the tech stack is the platform which typically consists of "base" software applications that are the underpinnings of more advanced applications. Here is where you will find the application development tool set to manage things like databases, integration, and security. This software layer also typically houses big data and analytic warehouses which we will be discussing shortly.

As far as deployment there are two deployment methods currently available today. The first one is infrastructure and platform that are available for purchase on premise which means that organizations procure the hardware and software which they own by way of a software license. This entitles them to the usage of the hardware and software in perpetuity and typically carries a recurring maintenance fee. Infrastructure and platforms are also available for rent as a subscription.

The acronyms IAAS and PAAS are used to refer to these subscriptions "as-a-service." These services are deployed via sharing computer processing resources leveraging the ubiquity of the internet. This is referred to as cloud computing and is the second deployment method. That concludes subject area number one.

AICPA CITP Credential Examination Series

Next, we'll move on to subject area number two which is data preparation and manipulation. Let's begin with the definition of data management according to the Data Management Association. Data management is the development, execution and supervision of plans, policies, programs and practices that control, protect, deliver and enhance the value of data and information assets.

Data are just facts. In IT processes data are generally presented as content in a field. Data becomes information when they are structured to provide context and meaning. Management of data generally focuses on the defining of the data element and how it's structured, stored and moved. Management of information is more concerned with the security, accuracy, completeness and timeliness of multiple pieces of data. That concludes subject area number two.

Moving on to subject area number three. Data analysis, functions, tools, and approaches including business intelligence and analytics. There are basically 10 primary data management functions. Let's review them. Number one is data governance. This is the exercise of authority, control and shared decision-making which consists of planning, monitoring, and enforcement over the management of the data sets.

The objectives of data governance are to enable better decision-making, to reduce operational friction, to protect the needs of the stakeholders, to train management and staff to adopt common approaches to data issues, to build standard repeatable processes, to reduce costs and increase effectiveness through the coordination of efforts and to ensure transparency of processes. Finally, the principles of data governance are integrity, transparency, auditability, accountability, stewardship, checks and balances, standardization and change management.

Number two is data architecture management. The development and maintenance of enterprise data architecture within the context of all enterprise architecture and its connections with the application system solutions and projects that implement enterprise architecture. The third is data development. The data focused activities within the system development lifecycle including data modeling and data requirements analysis, design, implementation and maintenance of databases and data related solution components.

Number four is data operations management. This is the planning, control, and support for structured data assets across the data lifecycle from creation and acquisition through to archival and purge.

Number five is data security management. This is the planning, implementation and control activities to ensure privacy and confidentiality and to prevent unauthorized and inappropriate data access, creation or change.

Number six is reference and master data management. This is the planning, implementation and control activities to ensure consistency of contextual data values with a golden version of these data values. Number seven is data warehousing and business intelligence management. This is the planning, implementation and control processes to provide decision support data and support knowledge workers engaged in reporting, query and analysis.

AICPA CITP Credential Examination Series

Number eight is document and content management. This is the planning, implementation and control activities to store, protect and access data found within electronic files and physical records including text, graphics, image, audio and video. Number nine is metadata management which is the planning, implementation and control activities to enable easy access to high quality integrated metadata.

Number 10 is data quality management. This is the planning, implementation and control activities that apply quality data management techniques to measure, assess, improve and ensure the fitness of data for use. The accounting profession can provide leadership in organizations by helping identify what data are important and how data is captured and stored to allow it to become actionable. We call this "insight to action" and even more valuable when it becomes predictive of future outcomes based upon historical trends.

The definition of predictive analytics is to use facts about the past to make inferences and then use these inferences to make assertions about the future. There is an implied trust by decision makers when choosing to leverage a prediction and inherent risk which needs to be managed, something the accounting profession is well equipped at managing.

Now is a good time to take a closer look at business intelligence. Business intelligence is about two things. It's about informed decision making and performance optimization. It supports activities from the strategic level down to the operational level.

Let me explain. Business intelligence normally starts off as a more passive opportunistic activity concentrated on identifying high-level problems whereby senior managers are expected to take action on one problem. When it proliferates down to the middle managers detailed information and timely analysis is needed and a reactive to proactive cultural shift starts taking place. Next, BI activities start becoming integrated into tactical planning and operations management. The focus now shifts to process effectiveness within and among departments.

Most modern analytical applications have Microsoft Excel-like functionality, making it easy for a non-techie to develop reports and wizard-type step by step configurations supporting the setup. This setup can consist of alerts as well, so that we can perform most of these tasks by ourselves. In short, properly managed business intelligence programs can fundamentally change the way a company looks at itself and makes decisions.

Okay, let's recap. Accountants are already familiar with applying many of these principles above to the financial data that they work with in a regular basis. Becoming involved in a data management and/or government initiative provides the opportunity to apply these principles into other parts of the organization. That concludes subject area number three. Next, I'm going to pass off to my colleague Adam who's going to discuss the remaining two topics of information traceability and information quality.

Adam: Thank You, Maria. Today I'll be speaking about information traceability and information quality, within the realm of information traceability I'll first be talking about source traceability. Source traceability is the ability to identify the source of a piece of data within your organization, within the realm of data quality management this is

important so that you can understand the reliability of a source of data and also understand where to find it if a downstream process is impacted by it.

I very often refer to Wikipedia for very succinct definitions, where I found this one that I think is valuable. Traceability is the ability to verify the history, location or application of an item by means of documented recorded identification. Consider the concept of a single source of truth, where is information sourced from?

Recently, I had the opportunity to work with a cooperative that provided retirement, health and life insurance benefits to their members. The cooperative had a highly trained and skilled IT department that was trying to work on a common data management project. They had separate and specialized systems to handle the following benefits: a pension plan which was a defined benefit plan, a 401k plan which was a defined contribution plan, a health care plan, a pharmacy plan and a life and disability insurance plan.

Each of these systems had a different enrollment process and that process captured a number of redundant pieces of information, which is not uncommon in the financial insurance industry. Through the enrollment process data like employee address and phone number were captured. However, over time, the members of the cooperative moved, got married, in some cases got divorced, they had children and they also retired. When evaluating the source of information within the cooperative that provided the latest address and phone number, the data traceability was an important factor.

In certain cases, data can be augmented to include a tag or traceability identifier. In other cases, a unique aspect of the data may be used to identify its uniqueness. In ERP systems, also referred to by Maria as the tech stack, traceability implies the use of a unique piece of data, for example, order date or time or a serialized sequence number, which can be traced to the transaction flow of applications and programs so the source can be easily identified.

Transformation traceability is the ability to trace data back through a transformation process that would allow you to identify the single source of truth. Very often companies will use third-party applications for reporting, in doing so they may extract data from a system or ledger, map the data and then use the transformed data for another company purpose.

Consider for example a company that uses a system called Hyperion Essbase for financial reporting. During the month-end cycle, the data is extracted from the general ledger and mapped to a summary format for reporting purposes. In the reported results, there is a single line item that is called 'cash and cash equivalents.' However, in the general ledger, there are 14 accounts that are used to manage the various cash accounts for payroll, accounts payable, deposits, petty cash and certificates of deposit.

The ability to trace the petty cash account from the GL to the reported results in the financial statements is transformation traceability. In the case of our example, the petty cash account was consolidated with other accounts, exported to another system and then summarized in a different fashion. Now, on to information quality. Has this ever happened to you? Have you received a piece of mail that was forwarded to you

from an address that you had three addresses ago? Have you wondered why there are discrepancies in the consolidated statements of a foreign subsidiary?

In the first case, poor information quality led to a company spending time and money trying to advertise to you when their subscriber information lacked good overall quality. This information lacked the accuracy that was needed to successfully get this to you the first time and relied on the postal service to forward it to you.

In the second case, the foreign currency exchange rates were not calculated to enough digits and the inaccuracy, therefore, triggered small but persistent changes in the values that were converted. Information quality assurance is the process to guarantee confidence that particular information meets some context specific quality requirements. Information quality relates to the reliability of the information that you have captured in your systems, we have often heard the phrase, "garbage in means garbage out," poor information quality can be particularly costly to an organization.

To recap, we just covered two areas, information traceability, and information quality. On behalf of the AICPA Information Management & Technology Assurance Division, this is Adam Chaikin. On behalf of Maria and myself, I'd like to thank you for tuning in for this CITP exam series podcast on the topic of data analysis and reporting infrastructures.

This is one in a series of podcasts that the AICPA's IMTA division is pleased to offer around a variety of topics of importance for the CITP exam. Be sure to check out other podcasts in this series on topics that include data backup and recovery, information lifecycle management, service organizational controls, the COSO model framework, internal audit, PCI compliance and HIPAA compliance.

Disclaimer