Origins of the data analysis process

When you decided to join this program, you proved that you are a curious person. So let's tap into your curiosity and talk about the origins of data analysis. We don't fully know when or why the first person decided to record data about people and things. But we do know it was useful because the idea is still around today!

We also know that data analysis is rooted in statistics, which has a pretty long history itself. Archaeologists mark the start of statistics in ancient Egypt with the building of the pyramids. The ancient Egyptians were masters of organizing data. They documented their calculations and theories on papyri (paper-like materials), which are now viewed as the earliest examples of spreadsheets and checklists. Today's data analysts owe a lot to those brilliant scribes, who helped create a more technical and efficient process.

It is time to enter the **data analysis life cycle**—the process of going from data to decision. Data goes through several phases as it gets created, consumed, tested, processed, and reused. With a life cycle model, all key team members can drive success by planning work both up front and at the end of the data analysis process. While the data analysis life cycle is well known among experts, there isn't a single defined structure of those phases. There might not be one single architecture that's uniformly followed by every data analysis expert, but there are some shared fundamentals in every data analysis process. This reading provides an overview of several, starting with the process that forms the foundation of the Google Data Analytics Certificate.

The process presented as part of the Google Data Analytics Certificate is one that will be valuable to you as you keep moving forward in your career:

1. **Ask**: Business Challenge/Objective/Question

2. **Prepare**: Data generation, collection, storage, and data management

3. **Process**: Data cleaning/data integrity

4. **Analyze**: Data exploration, visualization, and analysis

5. **Share**: Communicating and interpreting results

6. **Act**: Putting your insights to work to solve the problem

Understanding this process—and all of the iterations that helped make it popular—will be a big part of guiding your own analysis and your work in this program. Let's go over a few other variations of the data analysis life cycle.

EMC's data analysis life cycle

EMC Corporation's data analytics life cycle is cyclical with six steps:

1. Discovery

2. Pre-processing data

3. Model planning

4. Model building

5. Communicate results

6. Operationalize

EMC Corporation is now Dell EMC. This model, created by David Dietrich, reflects the cyclical nature of real-world projects. The phases aren't static milestones; each step connects and leads to the next, and eventually repeats. Key questions help analysts test whether they have accomplished enough to move forward and ensure that teams have spent enough time on each of the phases and don't start modeling before the data is ready. It is a little different from the data analysis life cycle this program is based on, but it has some core ideas in common: the first phase is interested in discovering and asking questions; data has to be prepared before it can be analyzed and used; and then findings should be shared and acted on.

For more information, refer to [The Genesis of EMC's Data Analytics Lifecycle](#).

SAS's iterative life cycle

An iterative life cycle was created by a company called **SAS**, a leading data analytics solutions provider. It can be used to produce repeatable, reliable, and predictive results:

1. Ask

2. Prepare

3. Explore

4. Model

5. Implement

6. Act

7. Evaluate

The SAS model emphasizes the cyclical nature of their model by visualizing it as an infinity symbol. Their life cycle has seven steps, many of which we have seen in the other models, like Ask, Prepare, Model, and Act. But this life cycle is also a little different; it includes a step after the act phase designed to help analysts evaluate their solutions and potentially return to the ask phase again.

For more information, refer to [Managing the Analytics Life Cycle for Decisions at Scale](#).

Project-based data analytics life cycle

A project-based data analytics life cycle has five simple steps:

1. Identifying the problem

2. Designing data requirements

3. Pre-processing data

4. Performing data analysis

5. Visualizing data

This data analytics project life cycle was developed by Vignesh Prajapati. It doesn't include the sixth phase, or what we have been referring to as the Act phase. However, it still covers a lot of the same steps as the life cycles we have already described. It begins with identifying the problem, preparing and processing data before analysis, and ends with data visualization.

For more information, refer to [Understanding the data analytics project life cycle](#).

Big data analytics life cycle

Authors Thomas Erl, Wajid Khattak, and Paul Buhler proposed a big data analytics life cycle in their book, **Big Data Fundamentals: Concepts, Drivers & Techniques**. Their life cycle suggests phases divided into nine steps:

1. Business case evaluation

2. Data identification

3. Data acquisition and filtering

4. Data extraction

5. Data validation and cleaning

6. Data aggregation and representation

7. Data analysis

8. Data visualization

9. Utilization of analysis results

This life cycle appears to have three or four more steps than the previous life cycle models. But in reality, they have just broken down what we have been referring to as Prepare and Process into smaller steps. It emphasizes the individual tasks required for gathering, preparing, and cleaning data before the analysis phase.

For more information, refer to [Big Data Adoption and Planning Considerations](#).

Key takeaway

From our journey to the pyramids and data in ancient Egypt to now, the way we analyze data has evolved (and continues to do so). The data analysis process is like real life architecture, there are different ways to do things but the same core ideas still appear in each model of the process. Whether you use the structure of this Google Data Analytics Certificate or one of the many other iterations you have learned about, we are here to help guide you as you continue on your data journey.