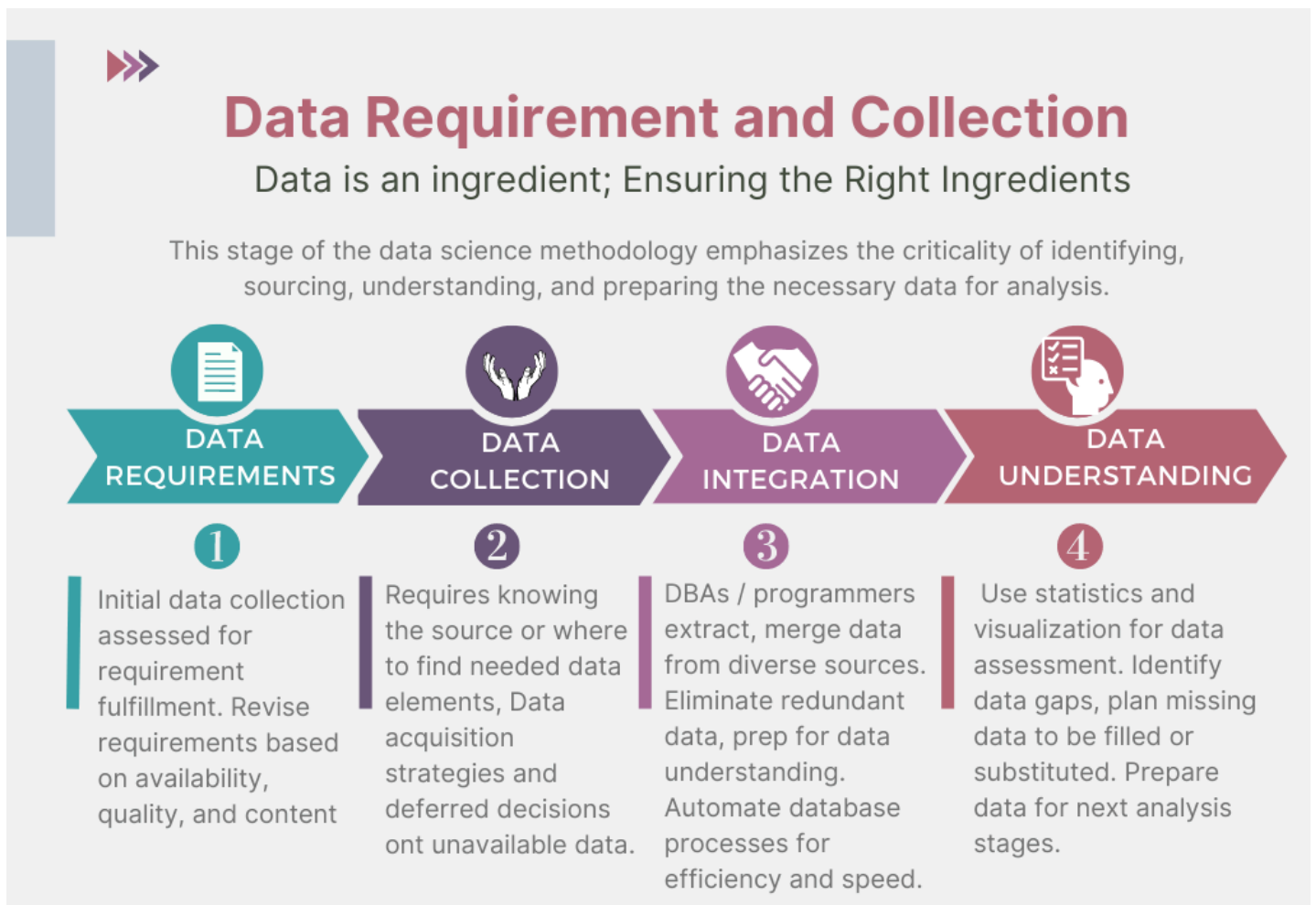# Lesson summary

Module 1 Lesson 2: From Requirements to Collection

Skills Network

Congratulations! You have completed this lesson. At this point in the course, you know:

- Data Requirements stage tasks include identifying the correct and necessary data content, data formats, and data sources for the specific analytical approach.

- During the Data Collection stage, expert data scientists meticulously revise data requirements and make critical decisions regarding the quantity and quality of data.

Data Requirement and Collection
Data is an ingredient; Ensuring the Right Ingredients

This stage of the data science methodology emphasizes the criticality of identifying, sourcing, understanding, and preparing the necessary data for analysis.

DATA REQUIREMENTS → **DATA COLLECTION** → **DATA INTEGRATION** → **DATA UNDERSTANDING**

1 Initial data collection assessed for requirement fulfillment. Revise requirements based on availability, quality, and content

2 Requires knowing the source or where to find needed data elements, Data acquisition strategies and deferred decisions ont unavailable data.

3 DBAs / programmers extract, merge data from diverse sources. Eliminate redundant data, prep for data understanding. Automate database processes for efficiency and speed.

4 Use statistics and visualization for data assessment. Identify data gaps, plan missing data to be filled or substituted. Prepare data for next analysis stages.

IBM. DATA SCIENCE METHODOLOGY Skills Network

- Data scientists apply descriptive statistics and visualization techniques to thoroughly assess the content, quality, and initial insights gained from the collected data, identify gaps, and determine if new data is needed or to substitute existing data.

Author(s)

Dr. Pooja
Patsy R. Kravitz