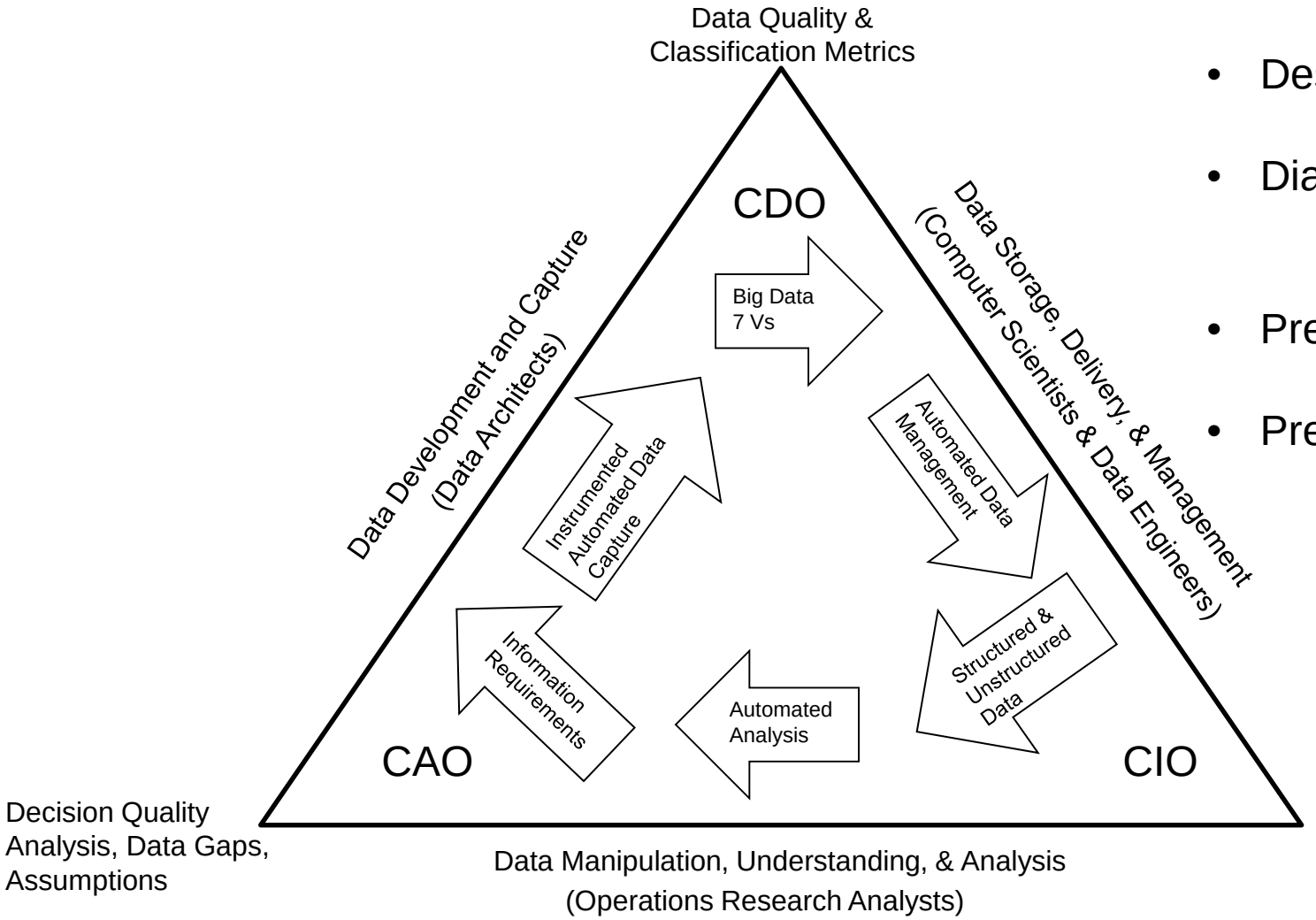
Data Analytics Introduction

Darryl K. Ahner, PhD, P.E.

Air Force Data Analytics Ecosystem (Core functionals)



- Descriptive Analytics (visualization, trend analysis, summary statistics)
- Diagnostic Analytics (examines data or content to answer the question “Why did it happen?)”
- Predictive Analytics (pattern recognition, data relationships, forecasting)
- Prescriptive Analytics (informs decisions, balances constraints, optimizes objectives)

Definitions

- **Data Analytics is the process of extracting information from data. It involves multiple stages including establishing a data set; preparing the data for processing; applying statistical, heuristics, or prescriptive models to include machine learning; identifying key findings and creating reports. The goal of data analytics is to find actionable insights that can inform decision making.**
  - ***Focus on the application of extracting meaningful information from collected data that can be acted upon***
- **Data science is a multi-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data. Data science is a "concept to unify statistics, data analysis, machine learning and their related methods" in order to "understand and analyze actual phenomena" with data. It employs techniques and theories drawn from many fields within the context of mathematics, statistics, computer science, and information science.**
  - ***Theory and algorithm development focused (including machine learning)***

Note: Consistent with other academic definitions found in backup slides

Inter-relationship of Offerings

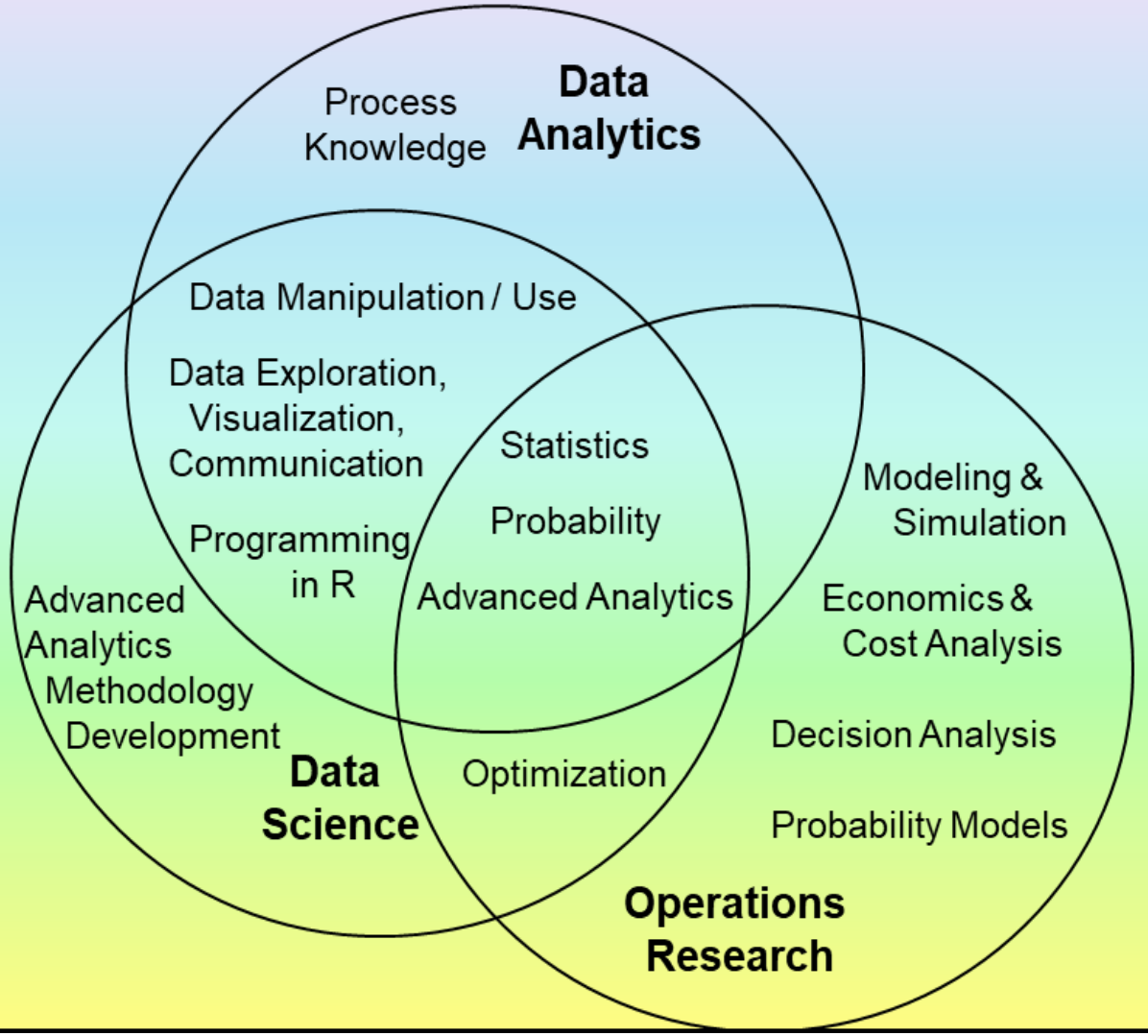
Descriptive Analytics

Diagnostic Analytics

Data Production, Assessment, Manipulation, and Cleaning

Predictive Analytics

Prescriptive Analytics

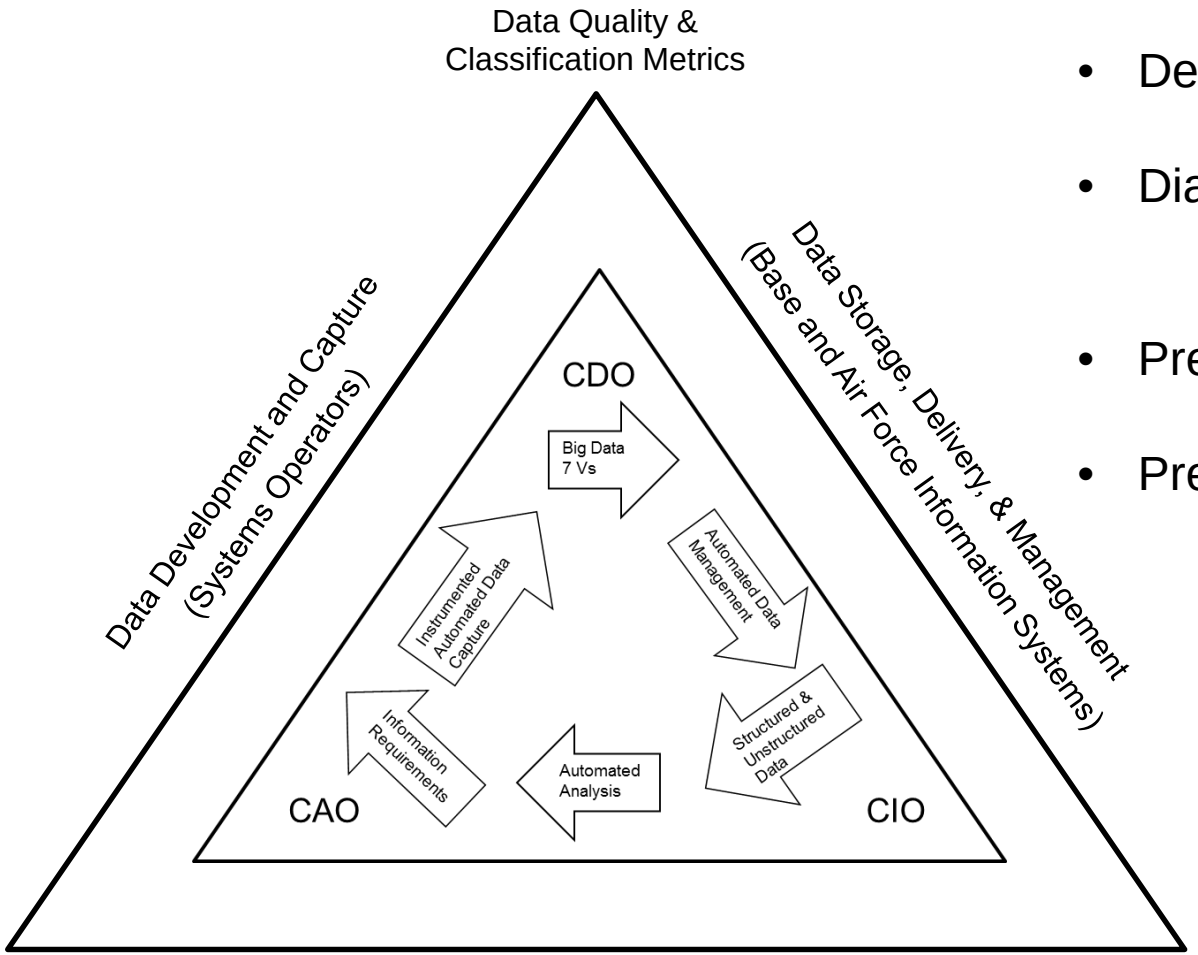


Terms and Definitions

- **Data scientists are big data wranglers. They take an enormous mass of messy data points (unstructured and structured) and use their formidable skills in math, statistics and programming to clean, manage and organize them. Then they apply all their analytic powers – industry knowledge, contextual understanding, skepticism of existing assumptions – to uncover hidden solutions to business challenges.**
- **Data engineers build massive reservoirs for big data. They develop, construct, test and maintain architectures such as databases and large-scale data processing systems. Once continuous pipelines are installed to – and from – these huge "pools" of filtered information, data scientists can pull relevant data sets for their analyses.**
- **Data architects create blueprints for data management systems. After assessing a company's potential data sources (internal and external), architects design a plan to integrate, centralize, protect and maintain them. This allows employees to access critical information in the right place, at the right time.**
- **Data analysts collect, process and perform statistical analyses of data. Their skills may not be as advanced as data scientists (e.g. they may not be able to create new algorithms), but their goals are the same – to discover how data can be used to answer questions and solve problems.**

Source: https://www.mastersindatascience.org/careers/

Air Force Data Analytics Ecosystem (User Community)



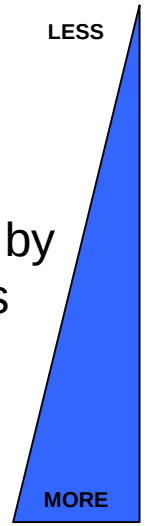
Data Manipulation, Understanding, & Analysis
(Career Agnostic Data Analysts)
(21x, 14x, 32x, 38x, 0140, 0810, 1515, etc.)

- Descriptive Analytics (visualization, trend analysis, summary statistics)
- Diagnostic Analytics (examines data or content to answer the question "Why did it happen?)"
- Predictive Analytics (pattern recognition, data relationships, forecasting)
- Prescriptive Analytics (informs decisions, balances constraints, optimizes objectives)

Types of Air Force Analysis

Types of Air Force Analysis

Approached most effectively by Data Analytics



- Resource Allocation Analysis
- Operational and Campaign Analysis
- Weapon Systems Performance Analysis
- Process Analysis (personnel, accessions, logistics, supply chain)
- Analytical Intel Analysis (as contrasted to subject matter expert analysis conducted through tradecraft)

Approached most effectively by Operations Research

Introduction

Introduction

- Three developments spurred recent explosive growth in the use of analytical methods in business applications:
- First development: Technological to use data to improve the efficiency and profitability of their operations, better understand their customers, price their products more effectively, and gain a competitive advantage.
- Second development: Ongoing research has resulted in numerous methodological developments
  - Advances in computational approaches to effectively handle and explore massive amounts of data.
  - Faster algorithms for optimization and simulation.
  - More effective approaches for visualizing data.
- Third development: an explosion in computing power and storage capability

DECISION MAKING

Decision Making (1 of 4)

- Leaders' responsibility:

  To make strategic, tactical, or operational decisions.
- **Strategic decisions**:
  - Involve higher-level issues concerned with the overall direction of the organization.
  - Define the organization's overall goals and aspirations for the future.

Decision Making (2 of 4)

- **Tactical decisions:**
  - Concern how the organization should achieve the goals and objectives set by its strategy.
  - Are usually the responsibility of midlevel management.
- **Operational decisions:**
  - Affect how the firm is run from day to day.
  - Are the domain of operations managers, who are the closest to the customer.

Decision Making (3 of 4)

Decision making process:

1. Understand the system to be analyzed*
2. Identify and define the problem.
3. Determine the criteria that will be used to evaluate alternative solutions.
4. Determine the set of alternative solutions.
5. Evaluate the alternatives.
6. Choose an alternative.

Decision Making (4 of 4)

Common approaches to making decisions include:

- Tradition.
- Intuition.
- Rules of thumb.
- Using the relevant data available to gain insights and reveal counterintuitive findings.

BUSINESS ANALYTICS DEFINED

Business Analytics Defined (1 of 2)

Business analytics:

- Scientific process of transforming data into insight for making better decisions.
- Used for data-driven or fact-based decision making, which is often seen as more objective than other alternatives for decision making.

Business Analytics Defined (2 of 2)

Tools of business analytics can aid decision making by:

- Creating insights from data.
- Improving our ability to more accurately forecast for planning.
- Helping us quantify risk.
- Yielding better alternatives through analysis and optimization.

Volume
Velocity **BIG DATA**
Variety
Veracity + Variability, Visualization, and Value.

Big Data (1 of 5)

- **Big data**: Any set of data that is too large or too complex to be handled by standard data-processing techniques and typical desktop software.
  - Volume.
  - Velocity.
  - Variety.
  - Veracity.
  - Variability.*
  - Visualization.*
  - Value*

Big Data (2 of 5)

The 7 Vs of Big Data

- **Volume:** Volume is how much data we have – what used to be measured in Gigabytes is now measured in Zettabytes (ZB) or even Yottabytes (YB). The IoT (Internet of Things) is creating exponential growth in data**.**
- **Velocity**: Velocity is the speed in which data is accessible. I remember the days of nightly batches, now if it's not real-time it's usually not fast enough.
- **Variety**: Variety describes one of the biggest challenges of big data. It can be unstructured and it can include so many different types of data from XML to video to SMS. Organizing the data in a meaningful way is no simple task, especially when the data itself changes rapidly.
- **Variability**: Variability is different from variety. A coffee shop may offer 6 different blends of coffee, but if you get the same blend every day and it tastes different every day, that is variability. The same is true of data, if the meaning is constantly changing it can have a huge impact on your data homogenization.

Big Data (3 of 5)

The 7 Vs of Big Data

- **Veracity**: Veracity is all about making sure the data is accurate, which requires processes to keep the bad data from accumulating in your systems. The simplest example is contacts that enter your marketing automation system with false names and inaccurate contact information. How many times have you seen Mickey Mouse in your database? It's the classic "garbage in, garbage out" challenge.
- **Visualization**: Visualization is critical in today's world. Using charts and graphs to visualize large amounts of complex data is much more effective in conveying meaning than spreadsheets and reports chock-full of numbers and formulas.
- **Value**: Value is the end game. After addressing volume, velocity, variety, variability, veracity, and visualization – which takes a lot of time, effort and resources – you want to be sure your organization is getting value from the data.

Big Data (4 of 5)

- Represents opportunities.
- Presents challenges in terms of data storage and processing, security, and available analytical talent.
- The 7 Vs have led to new technologies:
  - **Hadoop:** An open-source programming environment that supports big data processing through distributed storage and processing on clusters of computers.
  - **MapReduce:** A programming model used within Hadoop that performs two major steps: the map step and the reduce step.

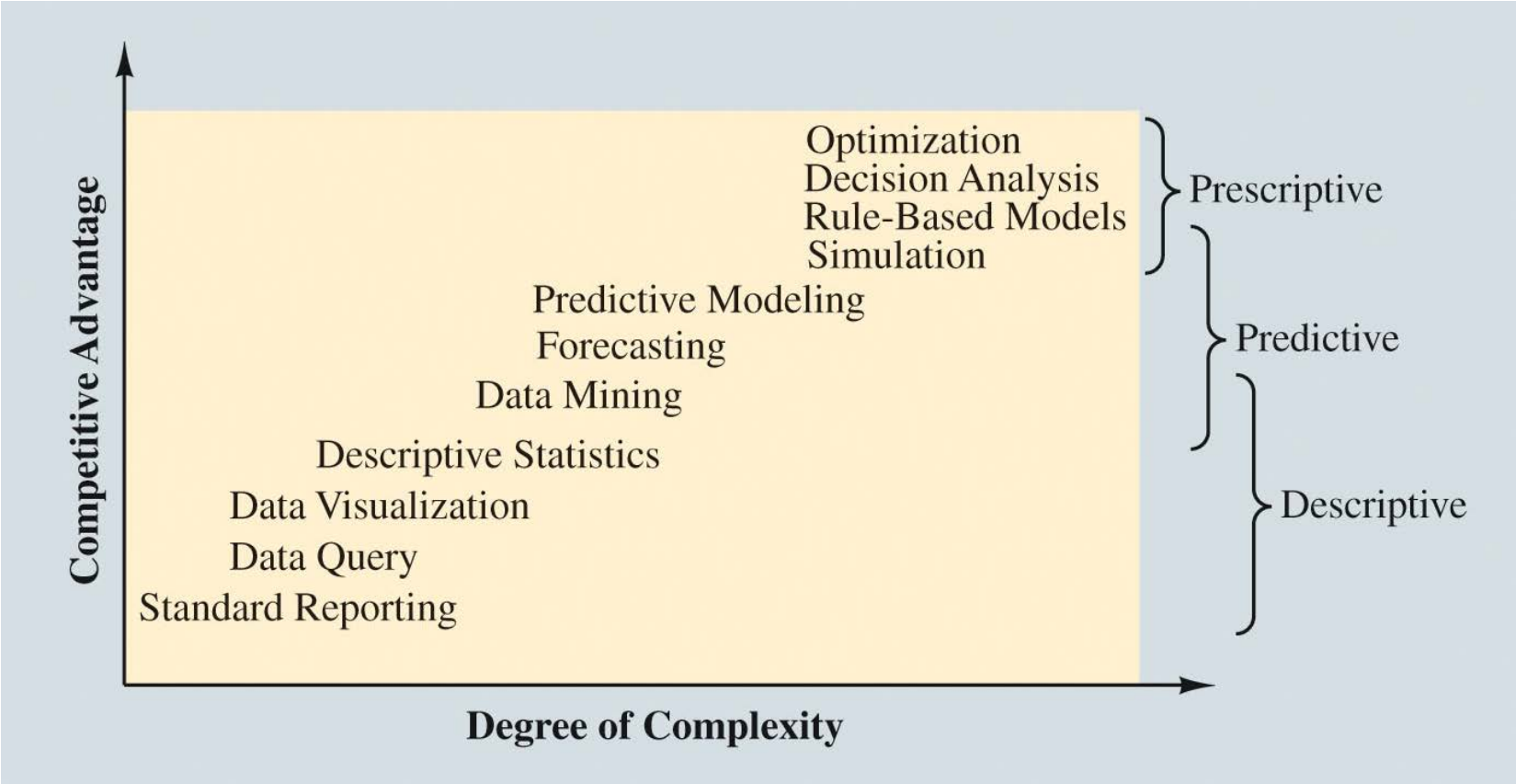
Big Data (Slide 5 of 5)

- **Data security**, the protection of stored data from destructive forces or unauthorized users, is of critical importance.
- The complexities of the 7 Vs have increased the demand for analysts, but a shortage of qualified analysts has made hiring more challenging.
- More companies are searching for **data scientists**, who know how to process and analyze massive amounts of data.
- The **Internet of Things (IoT)** is the technology that allows data, collected from sensors in all types of machines, to be sent over the Internet to repositories where it can be stored and analyzed.

BUSINESS ANALYTICS IN PRACTICE

Business Analytics in Practice

Figure 1.2: The Spectrum of Business Analytics



Source: Adapted from SAS

Business Analytics in Practice

- Predictive and prescriptive analytics are sometimes referred to as **advanced analytics.**

Financial Analytics:

- Use of predictive models to:
  - Forecast financial performance.
  - Assess the risk of investment portfolios and projects.
  - Construct financial instruments such as derivatives.
  - Construct optimal portfolios of investments.
  - Allocate assets.
  - Create optimal capital budgeting plans.
- Simulation is also often used to assess risk in the financial sector.

Business Analytics in Practice

Military Analytics:

- As there is a greater focus on data and tools become more assessable, analytics becomes more accepted.
- Operations (the hope for AI and ML)
- Logistics (focus beyond effectiveness to efficiency while maintaining effectiveness)
- Intelligence operations (PED)
- Acquisition