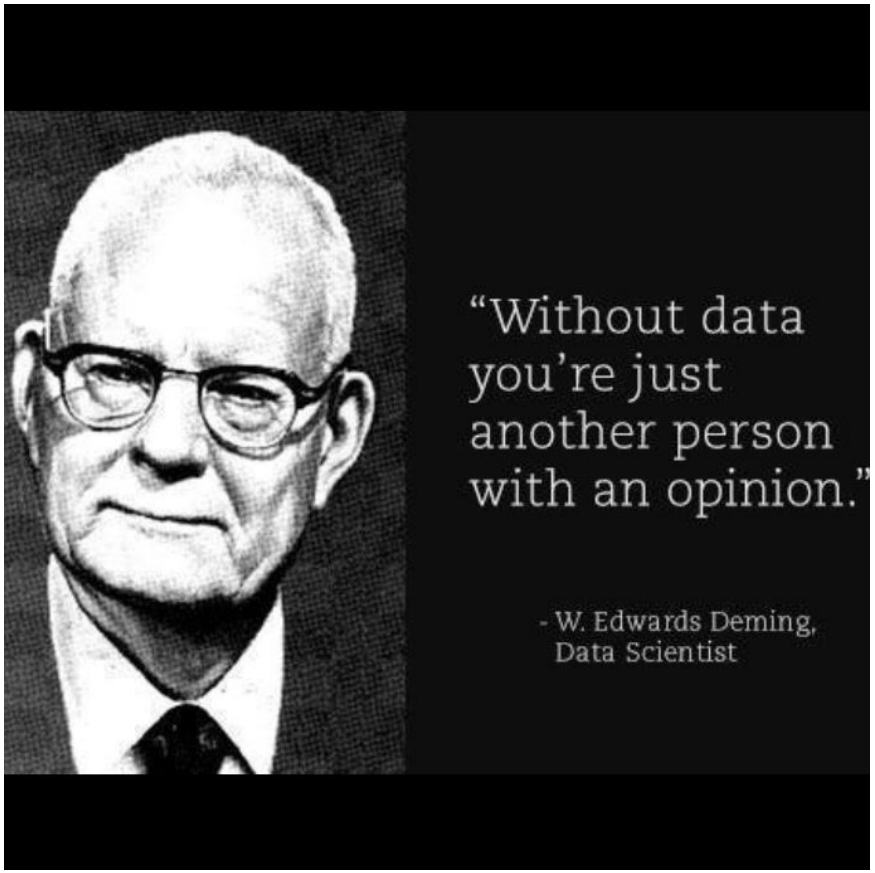
Data Analytics Architecture

NIST Concepts and Definitions
NIST.SP.1500-{1,2}



Data Analytics Architecture

- What **attributes** define Big Data solutions?
- How is Big Data **different** from traditional data environments and related applications?
- What are the **essential characteristics** of Big Data environments?
- How do these environments **integrate** with currently deployed architectures?
- What are the central scientific, technological, and standardization **challenges** that need to be addressed to accelerate the deployment of robust Big Data solutions?

Data Analytics Architecture

- **Big Data** is a term often used to describe the large amount of data in the networked, digitized, sensor-laden, information-driven world...
- …data can **overwhelm** traditional technical approaches, and the **growth** of data is outpacing scientific and technological advances in data analytics

Data Analytics Architecture

- Big Data also refers to the **inability** of traditional data architectures to efficiently handle the new datasets. Some characteristics of Big Data that force new architectures are:
- **Volume** (i.e., the size of the dataset);
- **Variety** (i.e., data from multiple repositories, domains, or types);
- **Velocity** (i.e., rate of flow); and
- **Variability** (i.e., the change in other characteristics).

Data Analytics Architecture

- Big Data characteristics dictate the overall design of a Big Data system, resulting in different data system architectures which must exploit
  - extreme parallelism
  - ubiquitous caching & locality
  - state-of-the-art service/device speed

Data Analytics Architecture

- The **growth rates** for data volumes, speeds, and complexity are **outpacing** scientific and technological advances in data analytics, management, transport, and data user spheres.
- The growth rates of data volumes are considered **faster than Moore's Law**, with data volumes more than doubling every eighteen months.
- One significant shift is in the amount of **unstructured data**. Historically, structured data has typically been the focus of most enterprise analytics, and has been handled through the use of the **relational** data model
- Non-relational models, frequently referred to as **NoSQL**, refer to logical data models that do not follow relational algebra for the storage and manipulation of data

Data Analytics Architecture

- **Data science** combines various technologies, techniques, and theories from various fields, mostly related to computer science and statistics, to obtain **actionable knowledge** from data
- Data-intensive science, shortened to data science, refers to the conduct of data analysis as an empirical science, learning directly from data itself...open-ended analysis often without preconceived hypotheses (sometimes referred to as discovery or data exploration)

Data Analytics Architecture

- A **data scientist** is a practitioner who has sufficient knowledge in the overlapping regimes of business needs, domain knowledge, analytical skills, and software and systems engineering to manage the end-to-end data processes in the data life cycle

Data Analytics Architecture

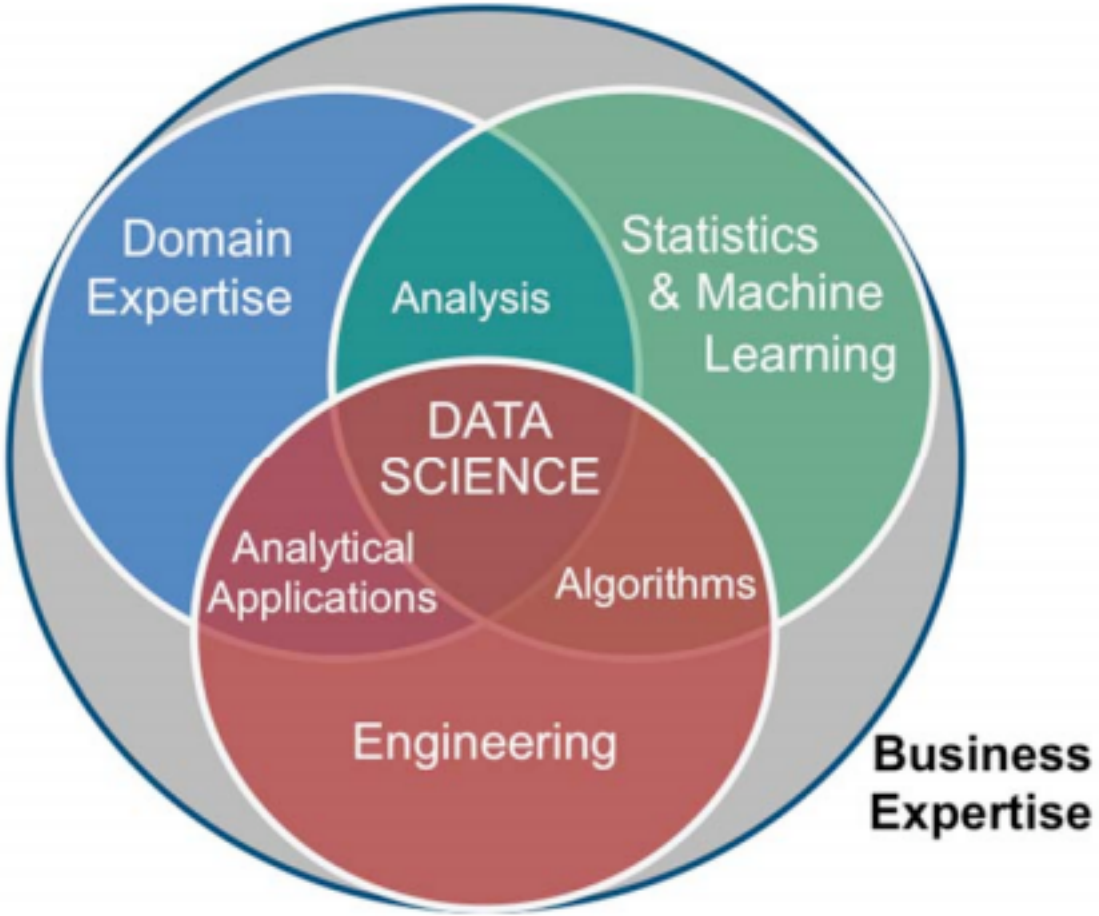


Figure 1: Skills Needed in Data Science

Data Analytics Architecture

- The **data life cycle** consists of the following four stages:

- **Collection**: This stage gathers and stores data in its original form (i.e., raw data.).
  - [we've done a **lot** of collecting…the rest, not so much!]
- **Preparation**: This stage involves the collection of processes that convert raw data into cleansed, organized information.
- **Analysis**: This stage involves the techniques that produce synthesized **knowledge** from organized information.
- **Action**: This stage involves processes that **use** the synthesized knowledge to generate **value** for the enterprise.

Data Analytics Architecture

- NIST.SP.1500-1.pdf:
  - Review Appendix A: Terms & Definitions
  - Review Appendix B: Abbreviations & Acronyms

Data Analytics Architecture

How to talk about Big Data to different audiences

- For **managers**, the terms will distinguish the **categorization of techniques** needed to understand this changing field.
- For **procurement officers**, it will provide the framework for discussing organizational needs and distinguishing among offered approaches.
- For **marketers**, it will provide the means to promote Big Data solutions and innovations.
- For the **technical community**, it will provide a common language to better differentiate Big Data's specific HW/SW offerings.

Data Analytics Architecture

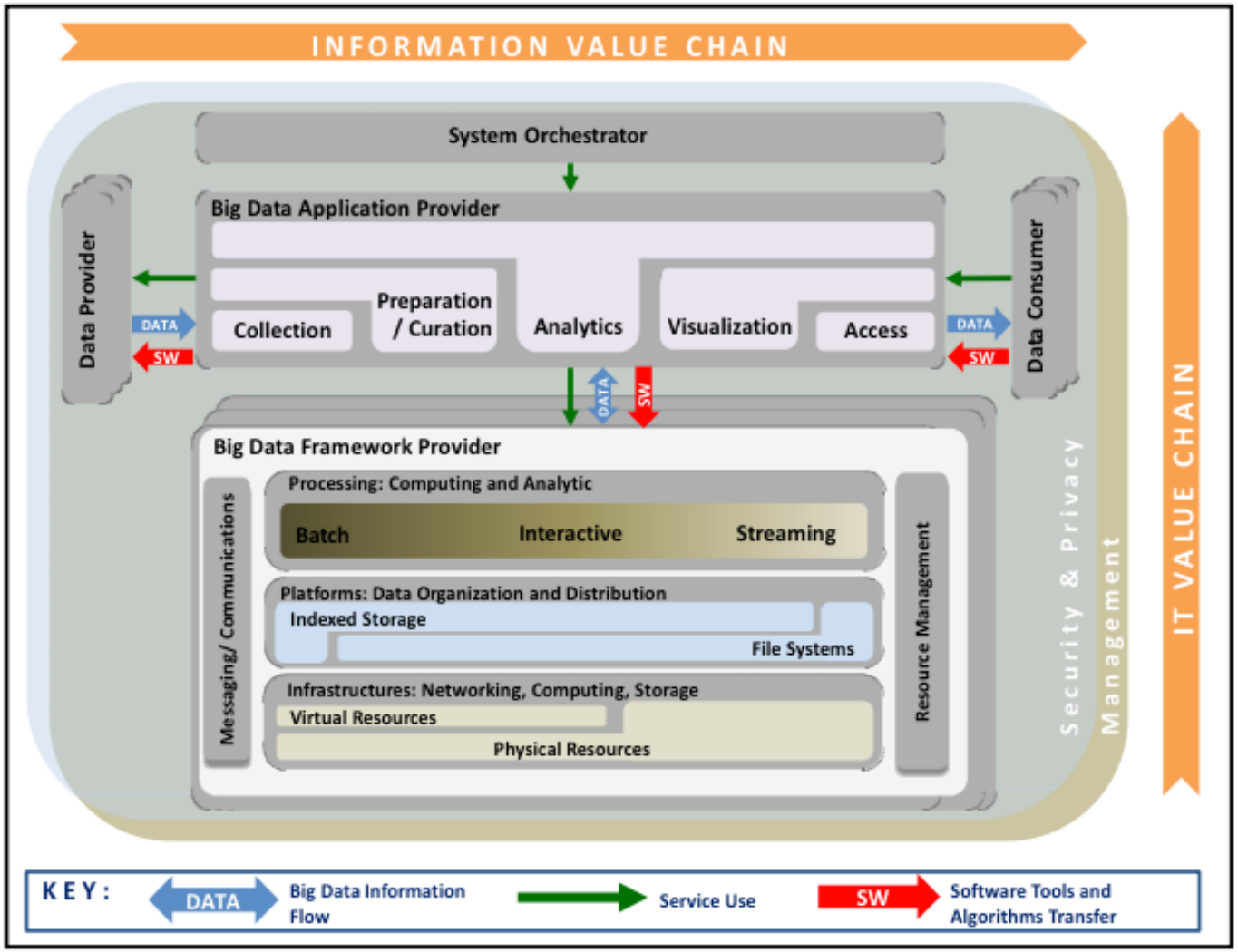


Figure 1: NIST Big Data Reference Architecture

Data Analytics Architecture

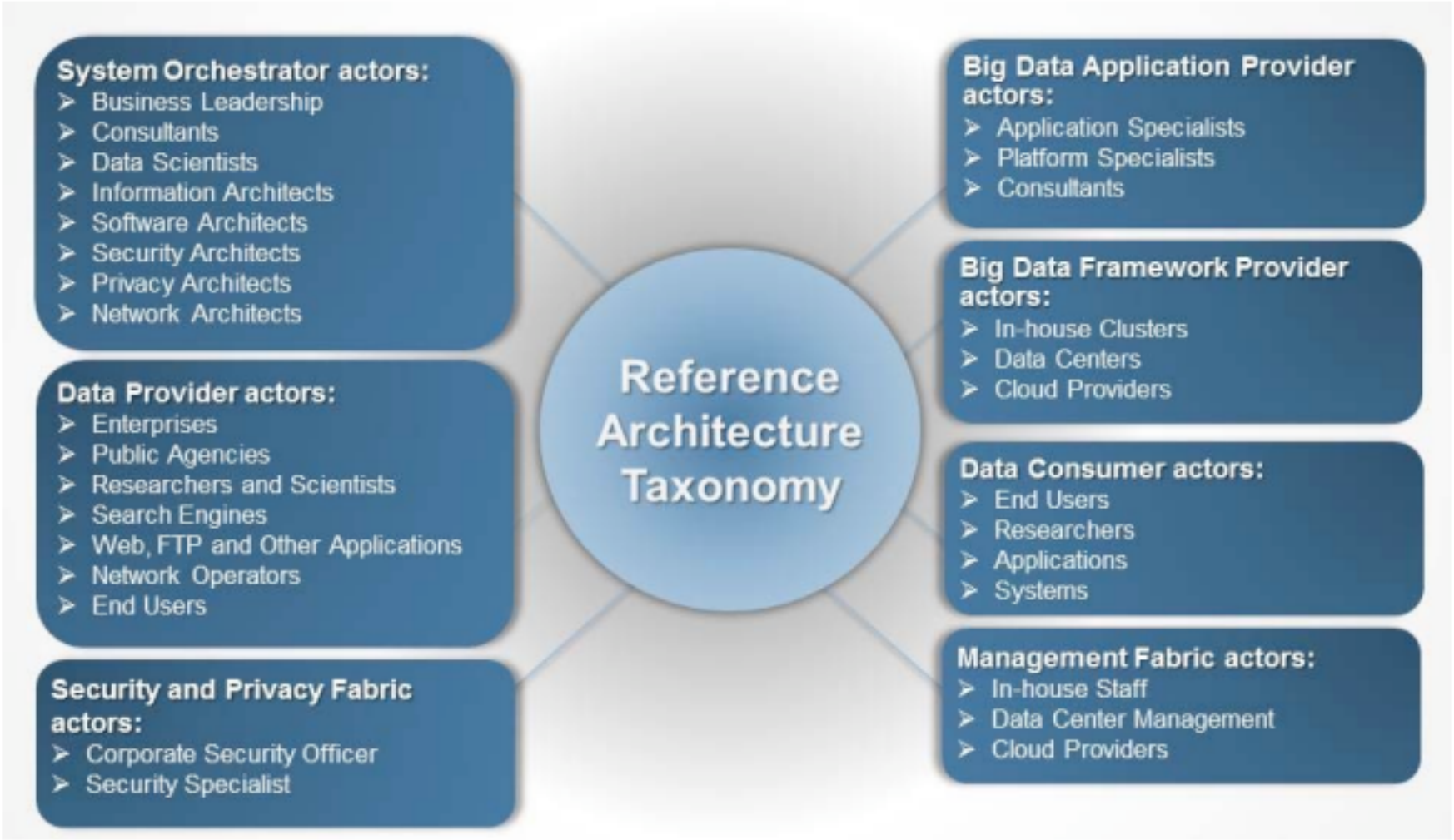


Figure 2: Roles and a Sampling of Actors in the NBDRA Taxonomy

Data Analytics Architecture

- NIST.SP.1500-2.pdf
  - Review Appendix B: Abbreviations & Acronyms