CSE 425: Internet of Things

Dept. of CSE BUBT| Summer 2021

Md. Hasibur Rahman

Data and Analytics for IoT and IoT Security Issues

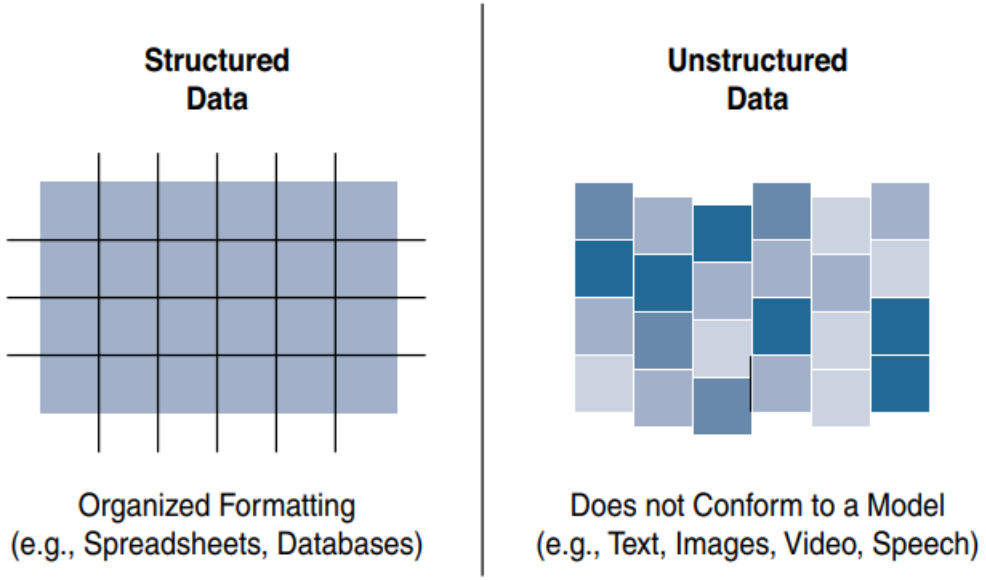
Field of data analytics from an IoT perspective

- Subject of analytics for IoT –Data
- Machine Learning: Once you have the data, what do you do with it, and how can you gain business insights from it?
- Big Data Analytics Tools and Technology: The most common technologies used in big data today, including Hadoop, NoSQL, MapReduce, and MPP.
- Edge Streaming Analytics: IoT requires that data be processed and analyzed as close to the endpoint as possible, in real-time.

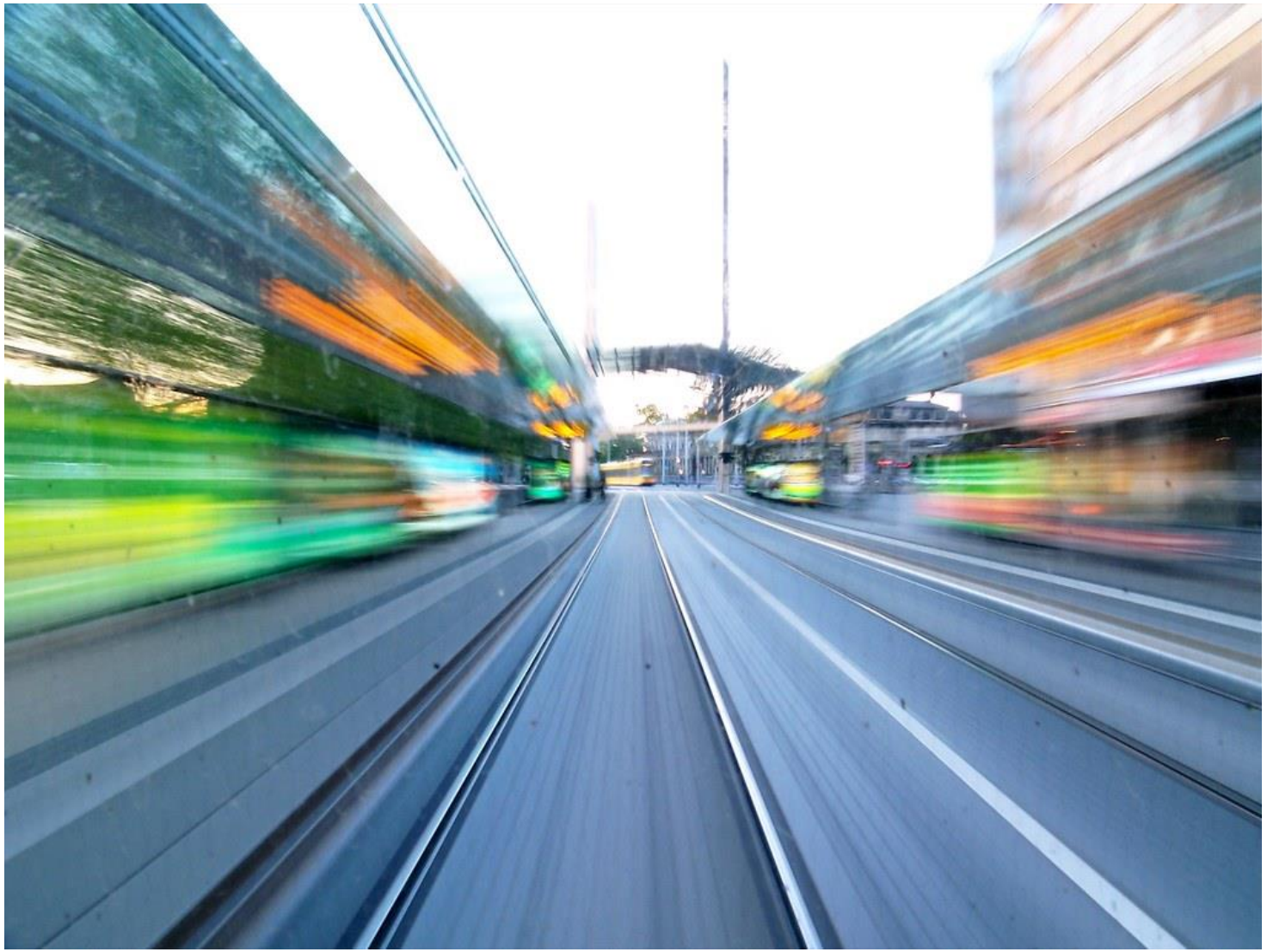
https://www.google.com/imghp?hl=EN

An Introduction to Data Analytics for IoT

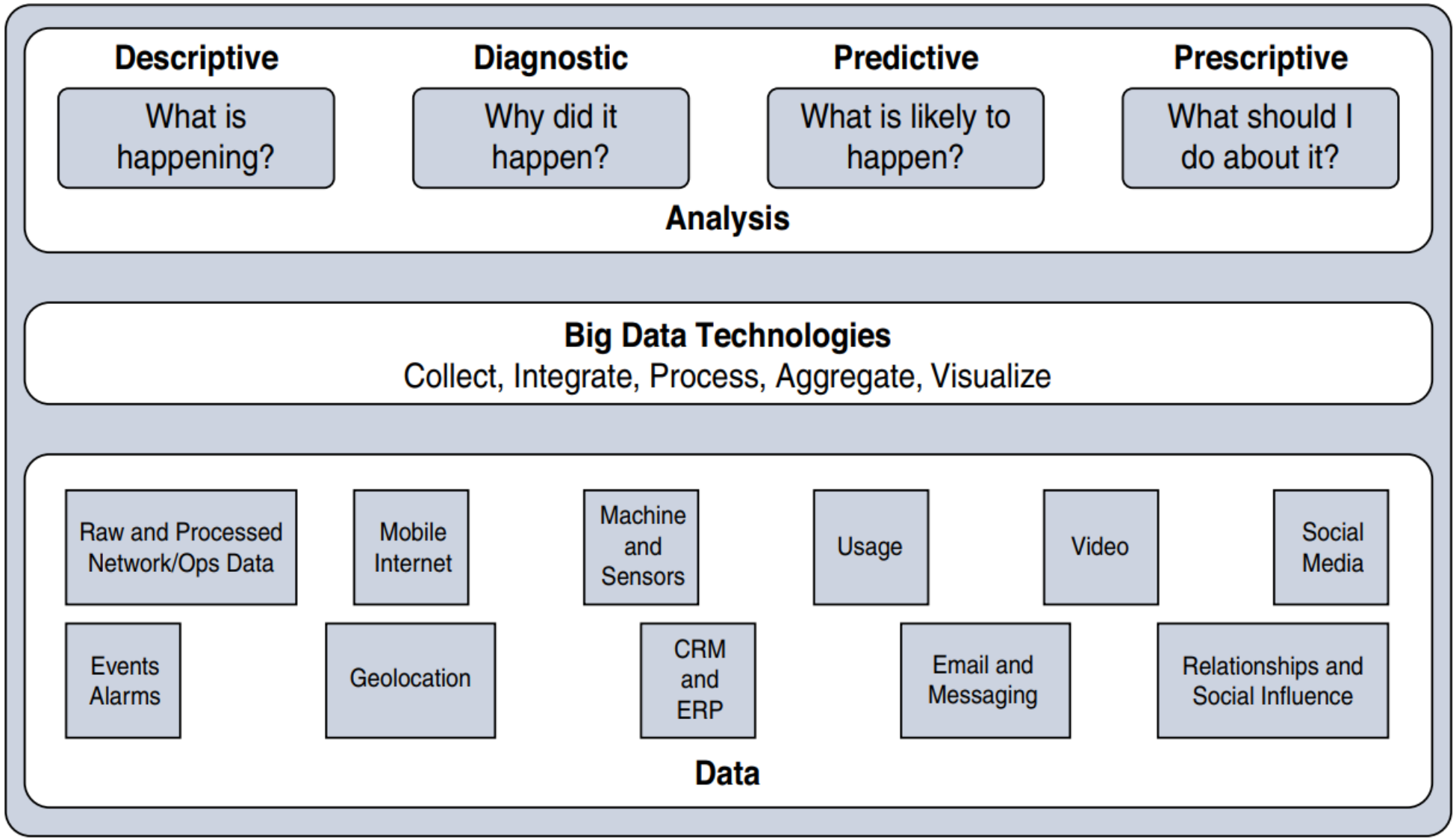
Structured Versus Unstructured Data



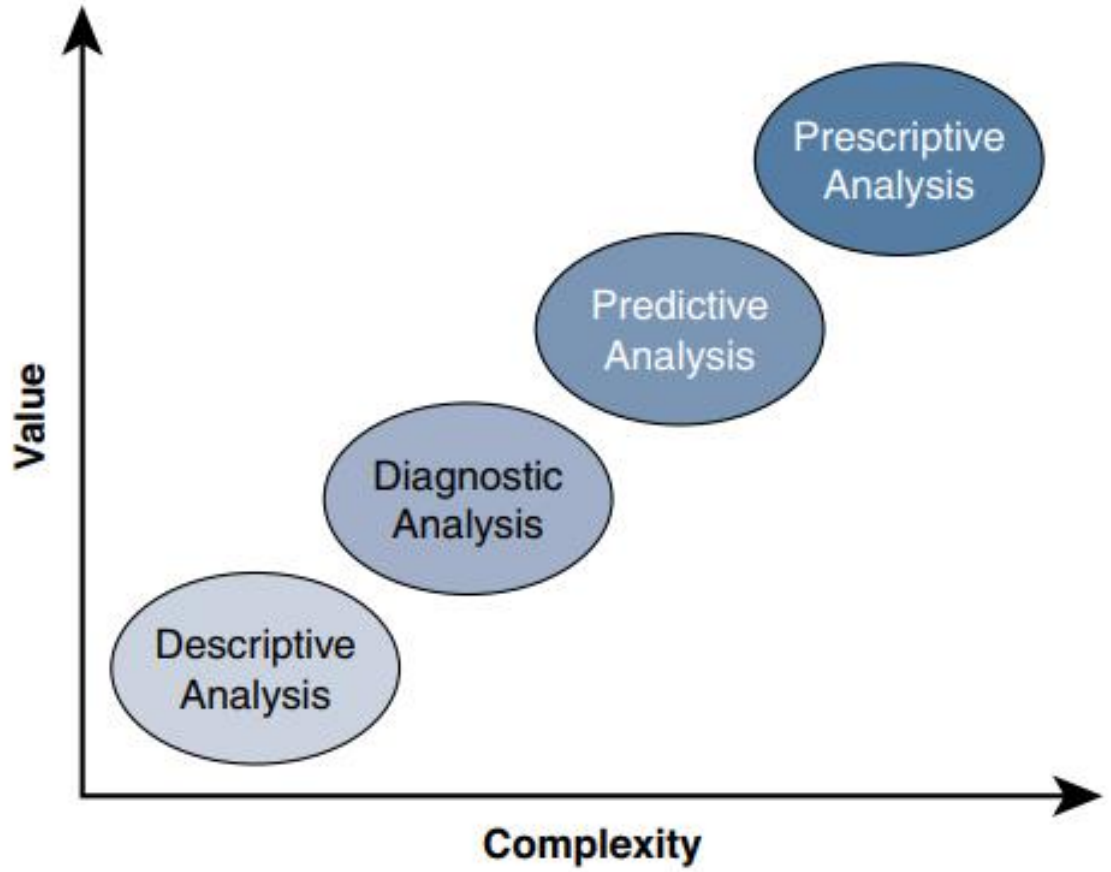
Data in Motion Versus Data at Rest



IoT Data Analytics Overview



Application of Value and Complexity Factors to the Types of Data Analysis



IoT Data Analytics Challenges

Scaling problems: Due to the large number of smart objects in most IoT networks that continually send data, relational databases can grow incredibly large very quickly. This can result in performance issues that can be costly to resolve, often requiring more hardware and architecture changes.

Volatility of data: With relational databases, it is critical that the schema be designed correctly from the beginning. Changing it later can slow or stop the database from operating. Due to the lack of flexibility, revisions to the schema must be kept at a minimum. IoT data, however, is volatile in the sense that the data model is likely to change and evolve over time. A dynamic schema is often required so that data model changes can be made daily or even hourly.

Machine Learning and IoT

Machine learning, *deep learning*, *neural networks*, and *convolutional networks* are words you have probably heard in relation to big data and IoT.

- Supervised Learning
- Unsupervised Learning
- Local learning: In this group, data is collected and processed locally, either in the sensor itself (the edge node) or in the gateway (the fog node).
- Remote learning: In this group, data is collected and sent to a central computing unit (typically the data center in a specific location or in the cloud), where it is processed.

Machine Learning and IoT

Regardless of the location where (and, therefore, the scale at which) data is processed, common applications of ML for IoT revolve around four major domains:

Monitoring: Smart objects monitor the environment where they operate. Data is processed to better understand the conditions of operations. These conditions can refer to external factors, such as air temperature, humidity, or presence of carbon dioxide in a mine, or to operational internal factors, such as the pressure of a pump, the viscosity of oil flowing in a pipe, and so on. ML can be used with monitoring to detect early failure conditions (for example, K-means deviations showing out-of-range behavior) or to better evaluate the environment (such as shape recognition for a robot automatically sorting material or picking goods in a warehouse or a supply chain).

Behavior control: Monitoring commonly works in conjunction with behavior control. When a given set of parameters reach a target threshold—defined in advance (that is, supervised) or learned dynamically through deviation from mean values (that is, unsupervised)—monitoring functions generate an alarm. This alarm can be relayed to a human, but a more efficient and more advanced system would trigger a corrective action, such as increasing the flow of fresh air in the mine tunnel, turning the robot arm, or reducing the oil pressure in the pipe.

Machine Learning and IoT

Operations optimization: Behavior control typically aims at taking corrective actions based on thresholds. However, analyzing data can also lead to changes that improve the overall process. For example, a water purification plant in a smart city can implement a system to monitor the efficiency of the purification process based on which chemical (from company A or company B) is used, at what temperature, and associated to what stirring mechanism (stirring speed and depth). Neural networks can combine multiples of such units, in one or several layers, to estimate the best chemical and stirring mix for a target air temperature. This intelligence can help the plant reduce its consumption of chemicals while still operating at the same purification efficiency level. As a result of the learning, behavior control results in different machine actions. The objective is not merely to pilot the operations but to improve the efficiency and the result of these operations.

Self-healing, self-optimizing: A fast-developing aspect of deep learning is the closed loop. ML-based monitoring triggers changes in machine behavior (the change is monitored by humans), and operations optimizations. In turn, the ML engine can be programmed to dynamically monitor and combine new parameters (randomly or semi-randomly) and automatically deduce and implement new optimizations when the results demonstrate a possible gain. The system becomes self-learning and self optimizing. It also detects new K-means deviations that result in predetection of new potential defects, allowing the system to self-heal.

Big Data Analytics Tools and Technology

- **Velocity:** *Velocity* refers to how quickly data is being collected and analyzed.
- **Variety:** Variety refers to different types of data. Often you see data categorized as structured, semi-structured, or unstructured.
- **Volume:** Volume refers to the scale of the data. Typically, this is measured from gigabytes on the very low end to petabytes or even Exabyte of data on the other extreme.

Security challenges in IoT

- **Inconsistent security standards**
- **Low processing power**
- **Legacy assets**
- **Lack of awareness in the users**
- **Botnet Attacks**
- **Lack of encryption**
- **Firmware updates Missing**
- **Rogue and Counterfeit IoT devices**