IOT with Big Data

H.M.H.T.Hikkaduwa

What is Big Data?

- Collection of larger volume of data used in machine language and advance applications and growing exponentially with the time.
- Big data is a combination of structured, semi structured and unstructured data collected by organizations that can be mined for information and used in machine learning projects, predictive modeling and other advanced analytics applications.
- There was large quantity of data and complex data.
- Ea.
  - Social Media. (Face book generates 500+terabytes of data.)
  - Single jet engine. (10 + of terabytes of data)

Importance of Big Data

- Companies use big data in their systems to improve operations, provide better customer service, create personalized marketing campaigns and take other actions that, ultimately, can increase revenue and profits. Businesses that use it effectively hold a potential competitive advantage over those that don't because they're able to make faster and more informed business decisions.
- In the energy industry, big data helps oil and gas companies identify potential drilling locations and monitor pipeline operations; likewise, utilities use it to track electrical grids.
- Financial services firms use big data systems for risk management and real-time analysis of market data.
- Manufacturers and transportation companies rely on big data to manage their supply chains and optimize delivery routes.
- Other government uses include emergency response, crime prevention and smart city initiatives.

Types of Big Data

- Structured data.
- Un structured data.
- Semi structured data.

Structured data

- Data that is the easiest to search and organize, because it is usually contained in rows and columns and its elements can be mapped into fixed pre-defined fields, is known as structured data.
- Think about what data you might store in an Excel spreadsheet and you have an example of structured data. Structured data can follow a data model a database designer creates - think of sales records by region, by product or by customer.
- In structured data, entities can be grouped together to form relations (‘customers’ that are also ‘satisfied with the service). This makes structured data easy to store, analyze and search and until recently was the only data easily usable for businesses. Today, most estimate structured data accounts for less than 20 percent of all data.

Structured data

- The data that can be processed, stored and access in fixed format we are call structured data. These data are highly organized and upload neatly into the relational databases.
- Relatively simple to enter, store, analyze the data and strictly defined in terms of field name and type. All data which can be stored in database in SQL. Have relational fields and mapped in to pre design fields.

Un structured data.

- A much bigger percentage of all the data is our world is unstructured data. Unstructured data is data that cannot be contained in a row-column database and doesn't have an associated data model. Think of the text of an email message. The lack of structure made unstructured data more difficult to search, manage and analyze, which is why companies have widely discarded unstructured data, until the recent proliferation of artificial intelligence and machine learning algorithms made it easier to process.

Un structured data.

- Type of data that is which is not organized in a predefined manner or not have a predefine data model.
- Unstructured data has it own internal data structure but not neat like structured data.
- Ea.

  Search out put from the Google.

Semi structured data

- Does not reside in a relational database but have some organization properties that make it easier to analyze.
- mix between both of them. The type of data defined as semi-structured data has some defining or consistent characteristics but doesn't conform to a structure as rigid as is expected with a relational database. Therefore, there are some organizational properties such as semantic tags or metadata to make it easier to organize, but there's still fluidity in the data.
- Email messages are a good example. While the actual content is unstructured, it does contain structured data such as name and email address of sender and recipient, time sent, etc. Another example is a digital photograph.

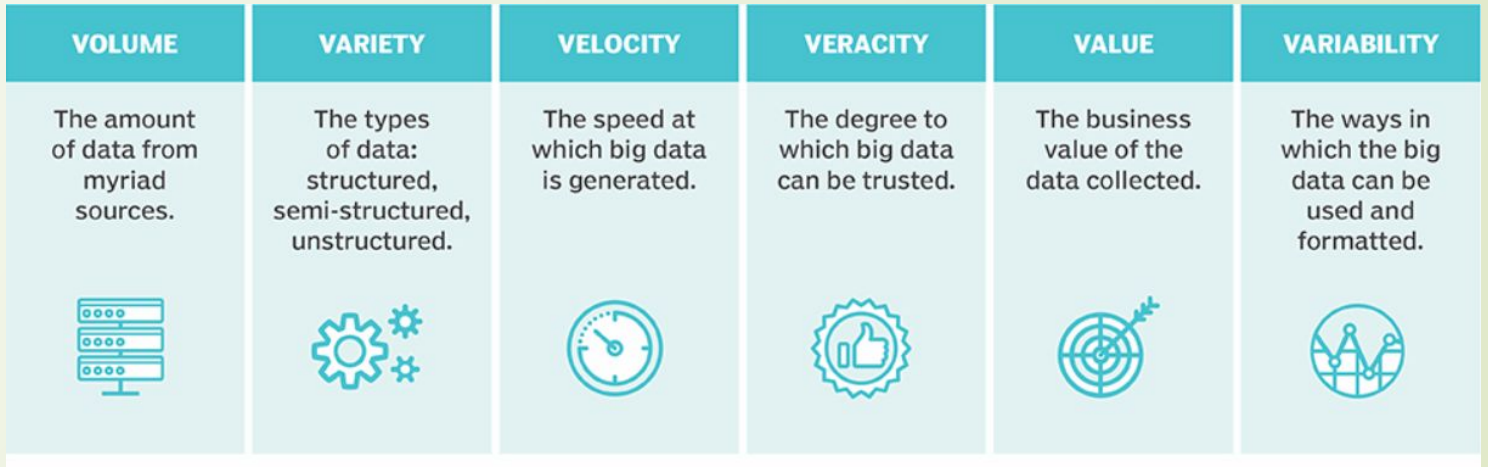
- XML, JSON documents and also NOSQL databases also semi structured in property.

Comparison

Properties	Structured data	Semi-structured data	Unstructured data
Technology	It is based on Relational database table	It is based on XML/RDF(Resource Description Framework).	It is based on character and binary data
Transaction management	Matured transaction and various concurrency techniques	Transaction is adapted from DBMS not matured	No transaction management and no concurrency
Version management	Versioning over tuples,row,tables	Versioning over tuples or graph is possible	Versioned as a whole
Flexibility	It is schema dependent and less flexible	It is more flexible than structured data but less flexible than unstructured data	It is more flexible and there is absence of schema
Scalability	It is very difficult to scale DB schema	It's scaling is simpler than structured data	It is more scalable.
Robustness	Very robust	New technology, not very spread	–
Query performance	Structured query allow complex joining	Queries over anonymous nodes are possible	Only textual queries are possible

The V’s of Big Data

- Velocity.
- Volume.
- Value.
- Varity.
- Veracity.



Velocity

- Refers to the speed with which the data is generated, collected and analyzed. Ea. -: Number of videos, twitter messages, photos, video clips.
- Big data allow us now to analyze the data while it is being generated without ever put it into database
- In real time we can do the analyzing of the data..

Volume

- Define the huge amount of data that is produce each day by the company. Generation of data is large and complex. No longer can be save and analyze using conventional data processing units.
- We use distributed systems to part of the data stored in different locations and brought together by software,
- Ea. Facebook alone there are 10 billion messages and 350 millions pictures are upload a day.

Value

- May companies recently established there own data platforms. They have data pools and invest lots of money in infrastructure. They need to generate a business value from their investments.
- Having endless data one thing, unless it can turned to values its loose less.
- he most important part of embarking on a big data initiative is to understand the costs and benefits of collecting and analyzing the data to ensure that ultimately the data that is reaped can be monetized.

Varity

- Define the diversity of data types and data sources. 80 present of the worlds data must be unstructured and does not show any indication of relationships.
- Big Data such algorithms, data is able to be sorted in a structured manner and examined for relationships. Data does not always comprise only conventional datasets, but also images, videos and speech recordings.

Veracity

- Veracity means quality or trust worthiness of the data. How accurate the all the data.
- Ea. Twitter post with hash tags.
- GPS data will be drift in urban area. Also Signals are lost.
- For above situation how we can enhance the veracity?

Relationship between IOT and Big Data

- Big data related with larger quantities of data and IOT was the one of easiest way to collect the larger set of data.
- Another factor was that IOT was real time and provide productivity, also improvement of data safety and working routine.
- The connected objects work well and different fields that make convenient to extract data.

IOT and Big Data Benefits

- A fast-growing network of connected sensors around the world, the IOT is making a colossal contribution to the body of big data. Because it depends on machines to gather data, rather than humans, the IOT creates huge data sets very quickly. Often, organizations benefit most from data if they can access, analyze, and interpret it in real-time, as it's flowing into a central platform.

- Real-time analytics depends on machine-driven tools like AI, machine learning, and deep learning. The insights they uncover then bounce back to make an impact on IOT devices, users, and organizational decisions.

- Here's an illustration: traffic lights and traffic cameras are connected to a smart city's IOT network. One camera reveals regular traffic jams at an interchange around rush hour each day, causing traffic to back up onto the interstate. AI tools analyze the data and suggest a solution: increasing the frequency and duration of green lights for traffic exiting the interstate during that period of time each day. Since the traffic lights are also connected to the network, city leaders can easily make that adjustment and cause traffic to flow more smoothly during peak hours.

Impact on integration.

- Customer engagement.

The integration in the application has a huge role in the engagement of customer and client. This has made it possible for companies to work with the advancement as per the strategies that take it to the next level. The motive to do so is by making it fast and easy in terms of size.

- Customized promotion.

he IoT App Development Company can incorporate it amazingly well to ensure that the system or dynamic nature is maintained. This can be achieved with the help of customer experience that will be automatically customized for product promotion. The customer will only receive updates or ads which are preferred by them.

- Mobile experience.

The advancement in technology has taken us to a new format of mobile experience in today's world. It is becoming essential to connect IoT and big data in such a way that the mobile application can enhance the experience of the customer and work upon the different aspects such as mobile payments.

IOT and Big Data Benefit Industries

- Healthcare
- As telehealth and telemedicine continue to expand, connected monitors will become common components of the healthcare system. These devices enable providers to track patients' readings remotely—from connected blood pressure meters, glucometers, heart monitors, just to name a few. As the connected sensors collect and aggregate data, AI and machine learning tools can identify trends and warning signs, alert patients and providers, and potentially save lives.

IOT and Big Data Benefit Industries

- IoT devices in the transportation sector include asset trackers, telematics systems, traffic control systems, surveillance, and remote monitoring systems. These devices generate tremendous amounts of disparate data that can be difficult to combine and interpret but AI and machine learning are helping.
- For example, a shipping company might have a thousand trucks traversing the country's interstates at a given moment. Inside the truck, containers are equipped with asset trackers, while the vehicle itself has an onboard telematics system. Using a centralized cloud platform, the company can view all this information on a single IoT dashboard, along with similar data from the rest of the fleet. AI and machine learning tools can identify trends and notify drivers of potential problems, such as upcoming vehicle maintenance needs or potential traffic backups.

IOT and Big Data Benefit Industries

- Manufacturing

The industrial IoT (IIoT) is a rapidly expanding segment of the wider IoT, and it's generating massive amounts of data. Putting that data together can be tricky, since the typical factory encompasses information from a conglomeration of legacy machines, cutting edge sensors, and manually entered data. IoT gateways equipped with edge analytics can help to streamline data flowing from different sources—often in different formats—and perform initial analyses and data cleansing before sending it to the cloud. At the enterprise cloud level, the factory data undergo more in-depth analyses. The organization's data science team, equipped with AI and machine learning tools, can then interpret the data and make recommendations for process adjustments that can increase efficiencies.

IOT and Big Data Benefit Industries

- Agriculture
- On a smart farm, hundreds or thousands of IoT sensors are deployed throughout an agricultural facility—placed far out in a field, inside a remote water tank to monitor usage, or even to track grazing animals. The sensors generate a large body of data on soil conditions, weather patterns, irrigation availability, and other factors. Taken together, the data can provide insights that drive precision agriculture, which takes a focused, individualized approach to cultivation at a particular site. For example, variations between soil conditions in fields on a single farm can be noted and offset with different fertilizing methods.

Key requirements (IOT analytics)

- How fast you need result?

  How fast need result from data gathered and design changes vary according to each scenario.

For example if you are dealing with stock market it is matter with the seconds. Who ever by the profitable shares win the situation.

Speed must be come in several levels.

A few hour -: Send your data to a data pool and do the map reduce. Use Hadoop.

A few second -: Send data into a stream processing system. (Apache Storm)

A few milliseconds -: Send data to system like complex event processing.

Key requirements (IOT analytics)

- How much data should keep?

In this stage you have to decide number of data to keep and in what format.

This is a trade off between the cost and potential data value and associated risk.

Most important factor was that data is the most valued assert for industries.

For this question we must categorize options as below.

- Keep all the data and save data to the data lake.
- Keep data in streaming fashion.
- Keep summarize version of data.

Data hindsight? Insight? Foresight?

- **Hindsight.**

  understanding of an event or situation only after it has happened. Hindsight is possible with aggregations and applied statistics. You can aggregate data by different groups and compare those results using statistical techniques, such as confidence intervals and statistical tests.

Insight.

- Insight is about gathering a deep understanding of something or someone.
- Insight can be gathered from many sources whether internal or external.

Data hindsight? Insight? Foresight?

- **Foresight.**

  Foresight is the ability to predict what will happen in the future. For businesses, prediction isn’t easy.

Insight and foresight would require machine learning and data mining. This includes finding patterns, modeling current behavior, predicting future outcomes, and detecting anomalies.

Time series processing.

- data are collected via sensors over time. Hence, they are time series data, and often most readings are auto correlated, e.g. a temperature reading is often highly affected by the earlier time step's reading.
- Recurrent Neural Networks (RNN) has shown promising results with time series data. However, widely used big data frameworks, such as Apache Spark and Hadoop, do not support these models yet.

Anomaly detections

- IoT use cases like predictive maintenance, health warnings, finding plug points that consumes too much power, optimizations, etc. depend on detecting anomalies. Anomaly detection poses several challenges.
- Lack of training data – most use cases would not have training data, and unsupervised techniques, such as clustering, should used.
- Class imbalance – Even when training data is available, often there will be a few dozen anomalies that exist among millions of data points. This problem is generally handled by building an ensemble of models where each model is trained with anomalous observations and resampled data from regular observation.
- Click and explore – after detecting anomalies, it must be understood in context and vetted by humans. Tools, therefore, are required to show those anomalies in context and enable operators to explore data further starting from the anomalies. For example, if an anomaly in a turbine is detected, it is useful to see that anomaly within regular data before and after. Also the anomaly able to study previous similar cases.

Links

- What is Big Data and Why is it Important? (techtarget.com)
- What's The Difference Between Structured, Semi-Structured And Unstructured Data? (forbes.com)

Thank You.