Name : جهاد يونس عرفة ابراهيم

ID : 275

Group: 2

Sec:12

* big data project have 4 webpage

1. Home

<div id="home">

            <p>The term big data has been in use since the 1990s, with some giving credit to John Mashey for

                popularizing the term.

                Big data usually includes data sets with sizes beyond the ability of commonly used software tools to

                capture, curate, manage, and process data within a tolerable elapsed time.

                Big data philosophy encompasses unstructured, semi-structured and structured data, however the main

                focus is on unstructured data. <br>

                Big data is a field that treats ways to analyze, systematically extract information from, or otherwise

                deal with data sets that are too large or complex to be dealt with by traditional data-processing

                application software. Data with many fields (columns) offer greater statistical power, while data with

                higher complexity (more attributes or columns) may lead to a higher false discovery rate.<br>

                Big data analysis challenges include capturing data, data storage, data analysis, search, sharing,

                transfer, visualization, querying, updating, information privacy, and data source. Big data was

                originally associated with three key concepts: volume, variety, and velocity.[3] The analysis of big

                data presents challenges in sampling, and thus previously allowing for only observations and sampling.

                Therefore, big data often includes data with sizes that exceed the capacity of traditional software to

                process within an acceptable time and value.<br>

                Current usage of the term big data tends to refer to the use of predictive analytics, user behavior

                analytics, or certain other advanced data analytics methods that extract value from big data, and seldom

                to a particular size of data set. "There is little doubt that the quantities of data now available are

                indeed large, but that's not the most relevant characteristic of this new data ecosystem."[4] Analysis

                of data sets can find new correlations to "spot business trends, prevent diseases, combat crime and so

                on".[5] Scientists, business executives, medical practitioners, advertising and governments alike

                regularly meet difficulties with large data-sets in areas including Internet searches, fintech,

                healthcare analytics, geographic information systems, urban informatics, and business informatics.

                Scientists encounter limitations in e-Science work, including meteorology, genomics,[6] connectomics,

                complex physics simulations, biology, and environmental research<br>

                The size and number of available data sets has grown rapidly as data is collected by devices such as

                mobile devices, cheap and numerous information-sensing Internet of things devices, aerial (remote

                sensing), software logs, cameras, microphones, radio-frequency identification (RFID) readers and

                wireless sensor networks.[8][9] The world's technological per-capita capacity to store information has

                roughly doubled every 40 months since the 1980s;[10] as of 2012, every day 2.5 exabytes (2.5×260 bytes)

                of data are generated.[11] Based on an IDC report prediction, the global data volume was predicted to

                grow exponentially from 4.4 zettabytes to 44 zettabytes between 2013 and 2020. By 2025, IDC predicts

                there will be 163 zettabytes of data.[12] One question for large enterprises is determining who should

                own big-data initiatives that affect the entire organization<br>

                Relational database management systems and desktop statistical software packages used to visualize data

                often have difficulty processing and analyzing big data. The processing and analysis of big data may

                require "massively parallel software running on tens, hundreds, or even thousands of servers".[14] What

                qualifies as "big data" varies depending on the capabilities of those analyzing it and their tools.

                Furthermore, expanding capabilities make big data a moving target. "For some organizations, facing

                hundreds of gigabytes of data for the first time may trigger a need to reconsider data management

                options. For others, it may take tens or hundreds of terabytes before data size becomes a significant

                consideration<br>

            </p>

        </div>

1. Variety

 <div id="Variety">

            <p>

                The type and nature of the data. The earlier technologies like RDBMSs were capable to handle structured

                data efficiently and effectively. However, the change in type and nature from structured to

                semi-structured or unstructured challenged the existing tools and technologies.<br>

                The big data technologies evolved with the prime intention to capture, store, and process the

                semi-structured and unstructured (variety) data generated with high speed (velocity), <br>

                The speed at which the data is generated and processed to meet the demands and challenges that lie in

                the path of growth and development. Big data is often available in real-time. Compared to small data,

                big data is produced more continually. Two kinds of velocity related to big data are the frequency of

                generation and the frequency of handling, recording, and publishing.<br>

                The truthfulness or reliability of the data, which refers to the data quality and the data value.[30]

                Big data must not only be large in size, but also must be reliable in order to achieve value in the

                analysis of it. The data quality of captured data can vary greatly, affecting an accurate analysis.<br>

                The worth in information that can be achieved by the processing and analysis of large datasets. Value

                also can be measured by an assessment of the other qualities of big data.[32] Value may also represent

                the profitability of information that is retrieved from the analysis of big data.<br>

                The characteristic of the changing formats, structure, or sources of big data. Big data can include

                structured, unstructured, or combinations of structured and unstructured data. Big data analysis may

                integrate raw data from multiple sources. The processing of raw data may also involve transformations of

                unstructured data to structured data.<br>

                Whether the entire system (i.e., {\textstyle n}{\textstyle n}=all) is captured or recorded or not. Big

                data may or may not include all the available data from sources.

            </p>

        </div>

1. Architecture

<div id="Architecture">

            <p>Big data repositories have existed in many forms, often built by corporations with a special need.

                Commercial vendors historically offered parallel database management systems for big data beginning in

                the 1990s. For many years, WinterCorp published the largest database report.

                In 2004, Google published a paper on a process called MapReduce that uses a similar architecture. The

                MapReduce concept provides a parallel processing model, and an associated implementation was released to

                process huge amounts of data. With MapReduce, queries are split and distributed across parallel nodes

                and processed in parallel (the "map" step). The results are then gathered and delivered (the "reduce"

                step). The framework was very successful,[37] so others wanted to replicate the algorithm. Therefore, an

                implementation of the MapReduce framework was adopted by an Apache open-source project named

                "Hadoop".[38] Apache Spark was developed in 2012 in response to limitations in the MapReduce paradigm,

                as it adds the ability to set up many operations (not just map followed by reducing).<br>

                MIKE2.0 is an open approach to information management that acknowledges the need for revisions due to

                big data implications identified in an article titled "Big Data Solution Offering".[39] The methodology

                addresses handling big data in terms of useful permutations of data sources, complexity in

                interrelationships, and difficulty in deleting (or modifying) individual records.<br>

                Studies in 2012 showed that a multiple-layer architecture was one option to address the issues that big

                data presents. A distributed parallel architecture distributes data across multiple servers; these

                parallel execution environments can dramatically improve data processing speeds. This type of

                architecture inserts data into a parallel DBMS, which implements the use of MapReduce and Hadoop

                frameworks. This type of framework looks to make the processing power transparent to the end-user by

                using a front-end application server.<br>

                The data lake allows an organization to shift its focus from centralized control to a shared model to

                respond to the changing dynamics of information management. This enables quick segregation of data into

                the data lake, thereby reducing the overhead time

            </p>

            <!-- img of Architecture big data -->

            <img src="https://lh3.googleusercontent.com/proxy/8J87vn6sY9hJTS4FzfrjMHiJAtRFisw5vyjFAxXtzYkXZ2KOWmMOPKqQRrJAeTKFeuuqbMkEvUpcLSEIvLwlfuz1nrXclLBIKzllzNJq8CTWcKgUhsNaizdVXkdAEMew5sEiBKUvpCt0nJ7ovEw0MgNRLcQPZf9VjI3ZP\_Yiwbl9Lunh\_nn7xYhfwmWXeitdYoAC9LwpCt12Ofgd"

                alt="Architecture">

        </div>

1. Polarities

 <div id="polarities">

            <table>

                <caption>polarities of big data</caption>

                <tr>

                    <th scope="col">core dimentially </th>

                    <th scope="col">core Assumbtion</th>

                </tr>

                <tr>

                    <th scope="row">temporal dimentially</th>

                    <td>direct of time </td>

                    <td>volicity</td>

                </tr>

                <tr>

                    <th scope="row"> factual dimentially </th>

                    <td>scape</td>

                    <td>reality </td>

                    <td>risk</td>

                </tr>

            </table>

        </div>

And this All Pages

<!DOCTYPE html>

<html lang="en">

<head>

    <meta charset="UTF-8">

    <meta name="viewport" content="width=device-width, initial-scale=1.0">

    <!-- font awsome -->

    <title>Big data</title>

</head>

<body>

    <header>

        <p class="logo">Big Data</p>

        <!-- list of anchor -->

        <ul>

            <li><a href="#home">home</a> </li>

            <li><a href="#Variety">Variety</a></li>

            <li><a href="#Architecture">Architecture</a></li>

            <li><a href="#Application">Application</a></li>

            <li><a href="#polarities">polarities </a></li>

        </ul>

    </header>

    <main>

        <div id="home">

            <p>The term big data has been in use since the 1990s, with some giving credit to John Mashey for

                popularizing the term.

                Big data usually includes data sets with sizes beyond the ability of commonly used software tools to

                capture, curate, manage, and process data within a tolerable elapsed time.

                Big data philosophy encompasses unstructured, semi-structured and structured data, however the main

                focus is on unstructured data. <br>

                Big data is a field that treats ways to analyze, systematically extract information from, or otherwise

                deal with data sets that are too large or complex to be dealt with by traditional data-processing

                application software. Data with many fields (columns) offer greater statistical power, while data with

                higher complexity (more attributes or columns) may lead to a higher false discovery rate.<br>

                Big data analysis challenges include capturing data, data storage, data analysis, search, sharing,

                transfer, visualization, querying, updating, information privacy, and data source. Big data was

                originally associated with three key concepts: volume, variety, and velocity.[3] The analysis of big

                data presents challenges in sampling, and thus previously allowing for only observations and sampling.

                Therefore, big data often includes data with sizes that exceed the capacity of traditional software to

                process within an acceptable time and value.<br>

                Current usage of the term big data tends to refer to the use of predictive analytics, user behavior

                analytics, or certain other advanced data analytics methods that extract value from big data, and seldom

                to a particular size of data set. "There is little doubt that the quantities of data now available are

                indeed large, but that's not the most relevant characteristic of this new data ecosystem."[4] Analysis

                of data sets can find new correlations to "spot business trends, prevent diseases, combat crime and so

                on".[5] Scientists, business executives, medical practitioners, advertising and governments alike

                regularly meet difficulties with large data-sets in areas including Internet searches, fintech,

                healthcare analytics, geographic information systems, urban informatics, and business informatics.

                Scientists encounter limitations in e-Science work, including meteorology, genomics,[6] connectomics,

                complex physics simulations, biology, and environmental research<br>

                The size and number of available data sets has grown rapidly as data is collected by devices such as

                mobile devices, cheap and numerous information-sensing Internet of things devices, aerial (remote

                sensing), software logs, cameras, microphones, radio-frequency identification (RFID) readers and

                wireless sensor networks.[8][9] The world's technological per-capita capacity to store information has

                roughly doubled every 40 months since the 1980s;[10] as of 2012, every day 2.5 exabytes (2.5×260 bytes)

                of data are generated.[11] Based on an IDC report prediction, the global data volume was predicted to

                grow exponentially from 4.4 zettabytes to 44 zettabytes between 2013 and 2020. By 2025, IDC predicts

                there will be 163 zettabytes of data.[12] One question for large enterprises is determining who should

                own big-data initiatives that affect the entire organization<br>

                Relational database management systems and desktop statistical software packages used to visualize data

                often have difficulty processing and analyzing big data. The processing and analysis of big data may

                require "massively parallel software running on tens, hundreds, or even thousands of servers".[14] What

                qualifies as "big data" varies depending on the capabilities of those analyzing it and their tools.

                Furthermore, expanding capabilities make big data a moving target. "For some organizations, facing

                hundreds of gigabytes of data for the first time may trigger a need to reconsider data management

                options. For others, it may take tens or hundreds of terabytes before data size becomes a significant

                consideration<br>

            </p>

        </div>

        <div id="Variety">

            <p>

                The type and nature of the data. The earlier technologies like RDBMSs were capable to handle structured

                data efficiently and effectively. However, the change in type and nature from structured to

                semi-structured or unstructured challenged the existing tools and technologies.<br>

                The big data technologies evolved with the prime intention to capture, store, and process the

                semi-structured and unstructured (variety) data generated with high speed (velocity), <br>

                The speed at which the data is generated and processed to meet the demands and challenges that lie in

                the path of growth and development. Big data is often available in real-time. Compared to small data,

                big data is produced more continually. Two kinds of velocity related to big data are the frequency of

                generation and the frequency of handling, recording, and publishing.<br>

                The truthfulness or reliability of the data, which refers to the data quality and the data value.[30]

                Big data must not only be large in size, but also must be reliable in order to achieve value in the

                analysis of it. The data quality of captured data can vary greatly, affecting an accurate analysis.<br>

                The worth in information that can be achieved by the processing and analysis of large datasets. Value

                also can be measured by an assessment of the other qualities of big data.[32] Value may also represent

                the profitability of information that is retrieved from the analysis of big data.<br>

                The characteristic of the changing formats, structure, or sources of big data. Big data can include

                structured, unstructured, or combinations of structured and unstructured data. Big data analysis may

                integrate raw data from multiple sources. The processing of raw data may also involve transformations of

                unstructured data to structured data.<br>

                Whether the entire system (i.e., {\textstyle n}{\textstyle n}=all) is captured or recorded or not. Big

                data may or may not include all the available data from sources.

            </p>

        </div>

        <div id="Architecture">

            <p>Big data repositories have existed in many forms, often built by corporations with a special need.

                Commercial vendors historically offered parallel database management systems for big data beginning in

                the 1990s. For many years, WinterCorp published the largest database report.

                In 2004, Google published a paper on a process called MapReduce that uses a similar architecture. The

                MapReduce concept provides a parallel processing model, and an associated implementation was released to

                process huge amounts of data. With MapReduce, queries are split and distributed across parallel nodes

                and processed in parallel (the "map" step). The results are then gathered and delivered (the "reduce"

                step). The framework was very successful,[37] so others wanted to replicate the algorithm. Therefore, an

                implementation of the MapReduce framework was adopted by an Apache open-source project named

                "Hadoop".[38] Apache Spark was developed in 2012 in response to limitations in the MapReduce paradigm,

                as it adds the ability to set up many operations (not just map followed by reducing).<br>

                MIKE2.0 is an open approach to information management that acknowledges the need for revisions due to

                big data implications identified in an article titled "Big Data Solution Offering".[39] The methodology

                addresses handling big data in terms of useful permutations of data sources, complexity in

                interrelationships, and difficulty in deleting (or modifying) individual records.<br>

                Studies in 2012 showed that a multiple-layer architecture was one option to address the issues that big

                data presents. A distributed parallel architecture distributes data across multiple servers; these

                parallel execution environments can dramatically improve data processing speeds. This type of

                architecture inserts data into a parallel DBMS, which implements the use of MapReduce and Hadoop

                frameworks. This type of framework looks to make the processing power transparent to the end-user by

                using a front-end application server.<br>

                The data lake allows an organization to shift its focus from centralized control to a shared model to

                respond to the changing dynamics of information management. This enables quick segregation of data into

                the data lake, thereby reducing the overhead time

            </p>

            <!-- img of Architecture big data -->

            <img src="https://lh3.googleusercontent.com/proxy/8J87vn6sY9hJTS4FzfrjMHiJAtRFisw5vyjFAxXtzYkXZ2KOWmMOPKqQRrJAeTKFeuuqbMkEvUpcLSEIvLwlfuz1nrXclLBIKzllzNJq8CTWcKgUhsNaizdVXkdAEMew5sEiBKUvpCt0nJ7ovEw0MgNRLcQPZf9VjI3ZP\_Yiwbl9Lunh\_nn7xYhfwmWXeitdYoAC9LwpCt12Ofgd"

                alt="Architecture">

        </div>

        <div id="Application">

            <p>Big data has increased the demand of information management specialists so much so that Software AG,

                Oracle Corporation, IBM, Microsoft, SAP, EMC, HP, and Dell have spent more than $15 billion on software

                firms specializing in data management and analytics. In 2010, this industry was worth more than $100

                billion and was growing at almost 10 percent a year: about twice as fast as the software business as a

                whole.<br>

                The use and adoption of big data within governmental processes allows efficiencies in terms of cost,

                productivity, and innovation,[55] but does not come without its flaws. Data analysis often requires

                multiple parts of government (central and local) to work in collaboration and create new and innovative

                processes to deliver the desired outcome. A common government organization that makes use of big data is

                the National Security Administration (NSA), who monitor the activities of the Internet constantly in

                search for potential patterns of suspicious or illegal activities their system may pick up.<br>

                Research on the effective usage of information and communication technologies for development (also

                known as "ICT4D") suggests that big data technology can make important contributions but also present

                unique challenges to international development.[56][57] Advancements in big data analysis offer

                cost-effective opportunities to improve decision-making in critical development areas such as health

                care, <br>

                management.[58][59][60] Additionally, user-generated data offers new opportunities to give the unheard a

                voice.[61] However, longstanding challenges for developing regions such as inadequate technological

                infrastructure and economic and human resource scarcity exacerbate existing concerns with big data such

                as privacy, imperfect methodology, and interoperability issues.[58] The challenge of "big data for

                development"[58] is currently evolving toward the application of this data through machine learning,

                known as "artificial intelligence for development (AI4D).<br>

                Research on the effective usage of information and communication technologies for development (also

                known as "ICT4D") suggests that big data technology can make important contributions but also present

                unique challenges to international development.

                Advancements in big data analysis offer cost-effective opportunities to improve decision-making in

                critical development areas such as health care, employment, economic productivity, crime, security, and

                natural disaster and resource management.[58][59][60] Additionally, user-generated data offers new

                opportunities to give the unheard a voice

                However, longstanding challenges for developing regions such as inadequate technological infrastructure

                and economic and human resource scarcity exacerbate existing concerns with big data such as privacy,

                imperfect methodology, and interoperability issues.[58] The challenge of "big data for development"

                is currently evolving toward the application of this data through machine learning, known as "artificial

                intelligence for development (AI4D).

            </p>

        </div>

        <!-- table for polarities of big data -->

        <div id="polarities">

            <table>

                <caption>polarities of big data</caption>

                <tr>

                    <th scope="col">core dimentially </th>

                    <th scope="col">core Assumbtion</th>

                </tr>

                <tr>

                    <th scope="row">temporal dimentially</th>

                    <td>direct of time </td>

                    <td>volicity</td>

                </tr>

                <tr>

                    <th scope="row"> factual dimentially </th>

                    <td>scape</td>

                    <td>reality </td>

                    <td>risk</td>

                </tr>

            </table>

        </div>

    </main>

</body>

</html>