SAKARYA ÜNİVERSİTESİ
Bilgisayar ve Bilişim Bilimleri Fakültesi
Bilgisayar Mühendisliği Bölümü

BSM 451

Internet of Things (IoT) and Applications

BIG DATA AND CLOUD COMPUTING

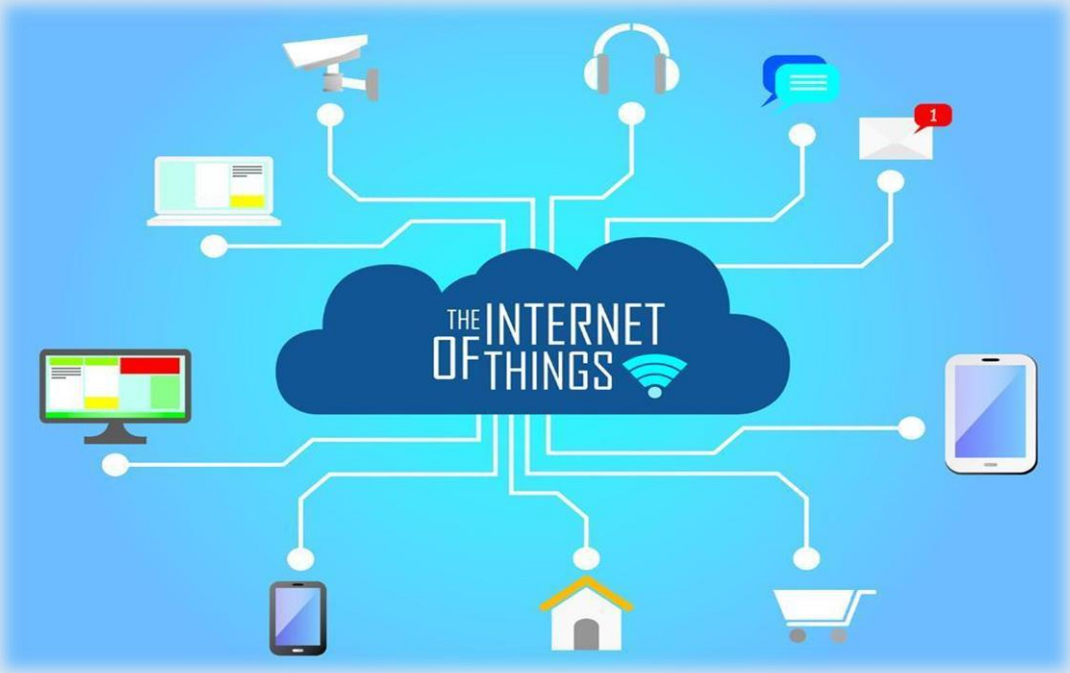
Assoc. Prof. Cüneyt BAYILMIŞ
Researcher Ünal ÇAVUŞOĞLU

Big Data and IoT

- The data that a large number of connected physical devices or sensors generate over the internet is called big data.
- Big data, conceptually, refers to the volume of data.
- Big data refers to the management and analysis of large volumes, high speed and a wide variety of data.
- IOT Data must be stored, processed, and interpreted.
- Effective and efficient disk space is required for storage.
- Big Data should be processable in an acceptable time.
- Cloud systems can be managed remotely, the size is scalable when requested.
- Therefore, data collected from IOT devices can be calculated and interpreted on Cloud systems.

Reasons for Increasing Big Data

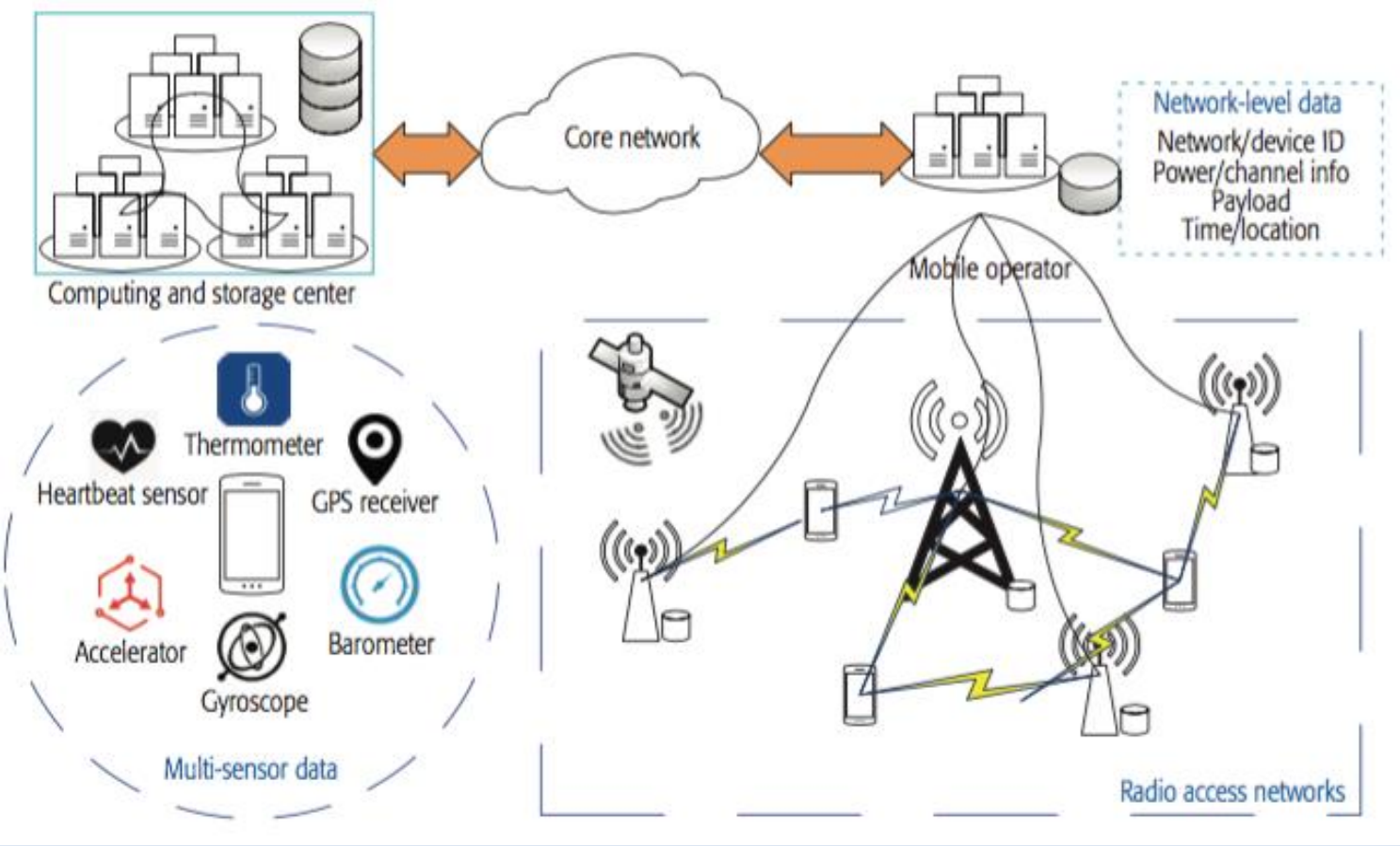
- Increased use of social networks (facebook, twitter, linkedin, etc.)
- The increase of location sensitive devices thanks to intelligent sensors,
- Increase in the number of intelligent sensors that capture and transmit information about the physical world
- In summary, as one of the causes; we can say IoT.



Mobile Big Data Source

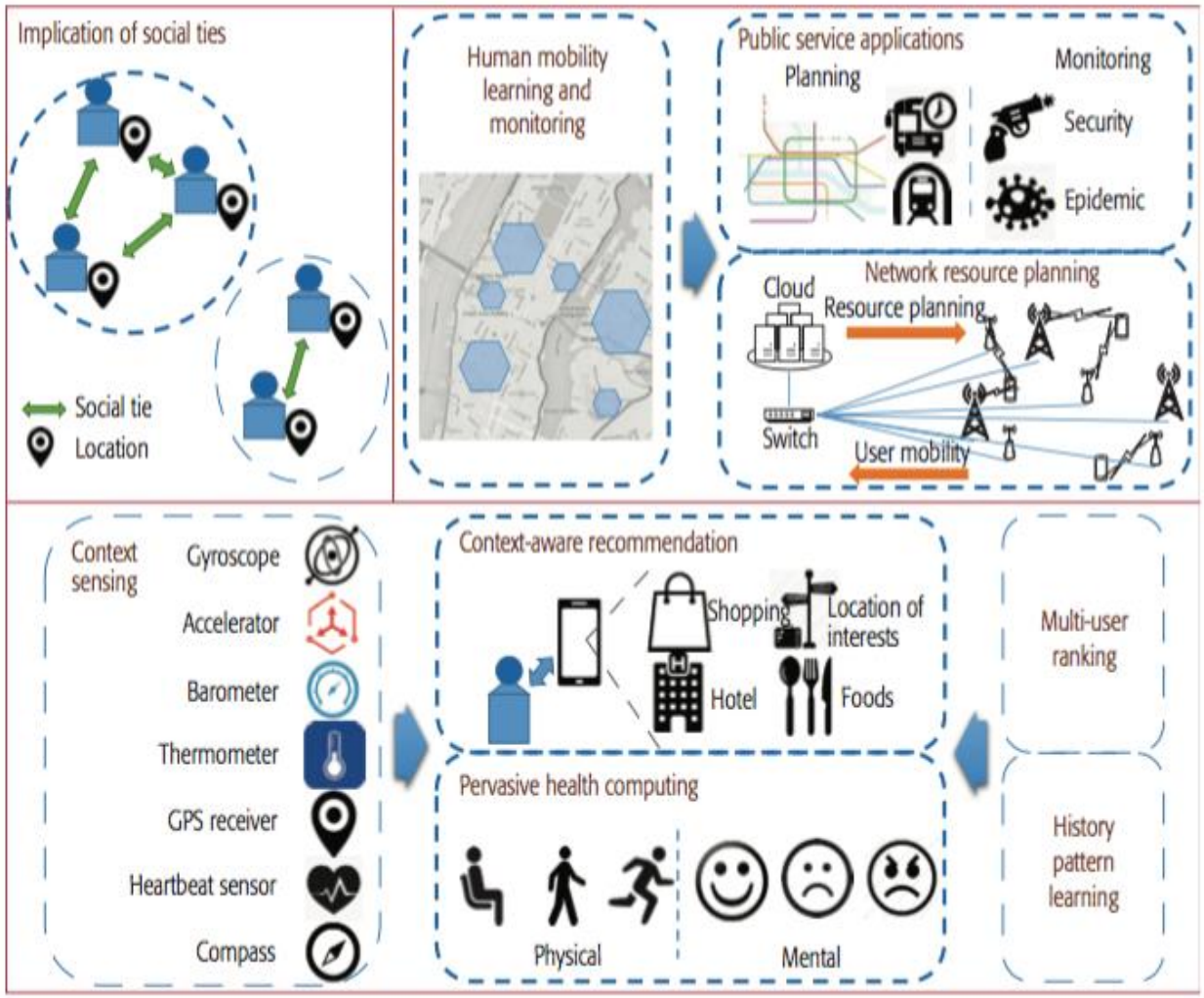
- Mobile data is obtained from mobile users (smartphone) or mobile networks (gateway, base station).
- Mobile data is divided into two classes.

- Application Data: Data collected from smart phone or embedded system sensors (GPS, accelerometer, etc.).
- Network Data: Data collected by network operators (User ID, location, device type, service type, etc.)



Kaynak: X. Cheng, L. Fang, X. Hong, and L. Yang, "Exploiting Mobile Big Data: Sources, Features, and Applications," IEEE Network, vol. 16, no. 1, pp. 72-79, 2017

Mobil Büyük Veri Uygulamaları



- Mobile data applications can be classified as individual data and community data.
- Individual services, content aware suggestions, activity recognition, etc.
- Community services include public services, smart city applications, traffic management, and so on.

Kaynak: X. Cheng, L. Fang, X. Hong, and L. Yang, "Exploiting Mobile Big Data: Sources, Features, and Applications," IEEE Network, vol. 16, no. 1, pp. 72-79, 2017

Big Data and IoT

Challenges for large data:

- Volume (Volume)

Large data, data size, deals with the data set.

- Velocity

It deals with the speed of production of the data. It deals with the analysis of the algorithms necessary for the computation of various subject lines for processing of the data.

- Variety

Expresses the structural diversity in the dataset.

- Veracity

The truth of the data is handled.

- Value

It is the value expressed as the determining characteristic of the data.

Big Data and IoT

Features	Challenges	Methods
Hacim (Volume)	Storage/Scalability	DFS (Distributed file system)
Hız (Velocity)	Fast processing	Parallel programming
Çeşitlilik (Variety)	Heterogeneous structure	NoSQL Database
Değer (Value) Doğruluk (Veracity)	Bilgi Keşfi (Knowledge Discovery) Anlambilim (Semantics) Analitik (Analytics)	Data Minning Algortihms

Big Data Analysis for IoT

- Analysis platforms for IoT

- These platforms meet the requirements of the IoT.
- They support real-time work.
- Instead of specially developed analysis programs, it is a better solution to take advantage of the platforms commonly used for the IoT.

Big Data Analysis for IoT

- Hadoop is a large data manipulation tool developed for open source Java.
- HDFS (Hadoop File System), known as a distributed data system, uses MapReduce as its processing module.
- MapReduce, developed by Google in 2004, uses data mapping and reducing functions.
- The data to be analyzed by Hadoop is stored on HDFS. In general, Hadoop runs on clusters that are created by more than one computer. This allows the distribution both tasks and data.
- The most common use cases are Facebook.

MapReduce

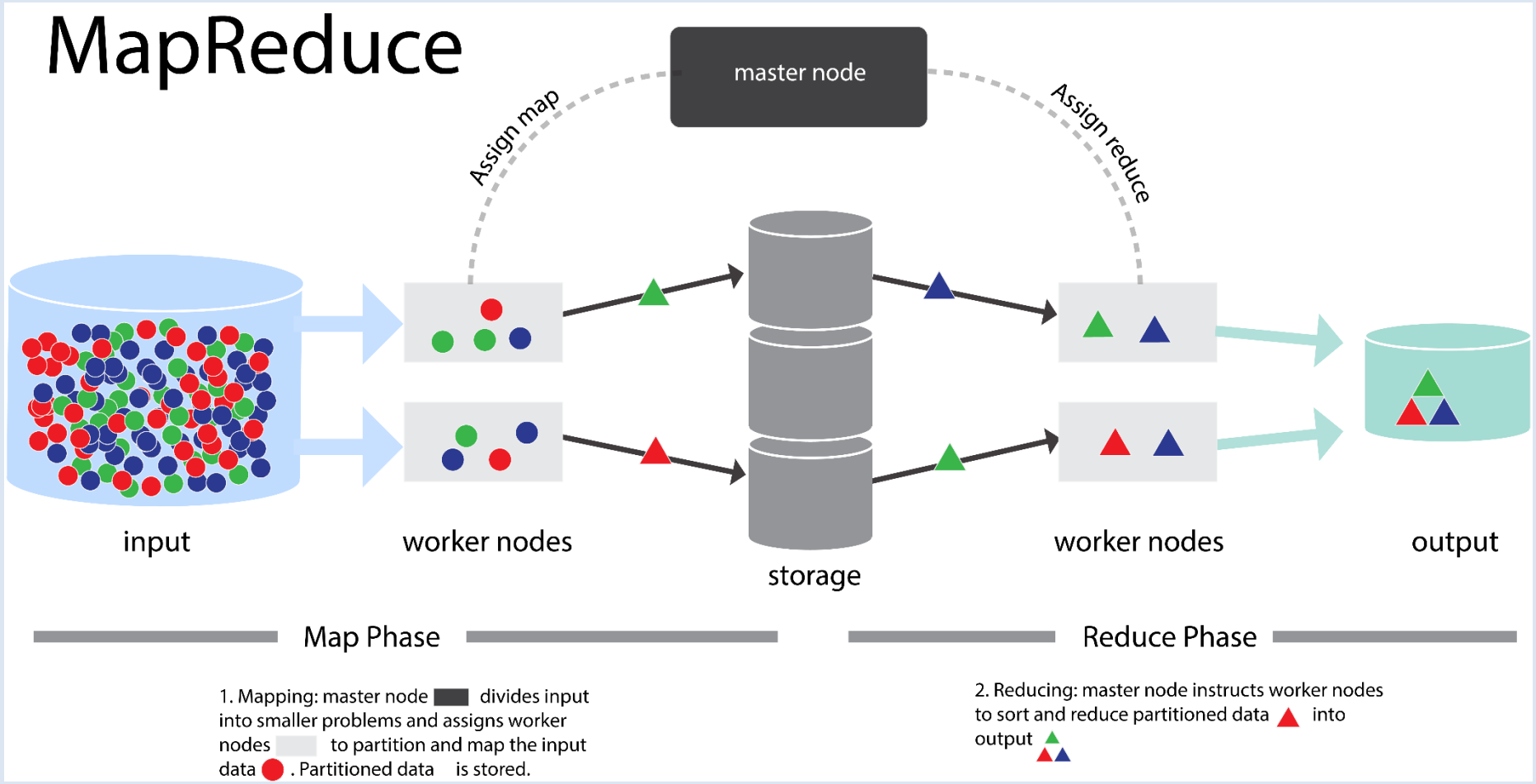
- Mapreduce is a distributed programming model aimed at large cluster processing.
- MapReduce consists of two parts:

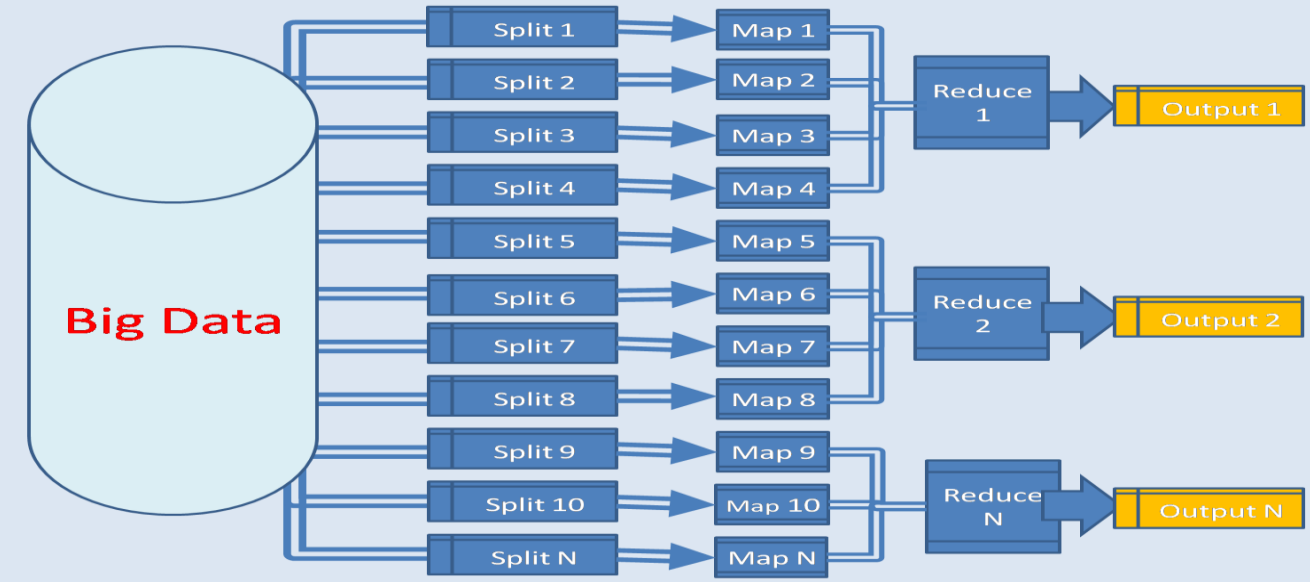
Map: Returns a result list by processing all the members of a stack with the function that they have. It divides very large data sets into two or more areas called buckets.

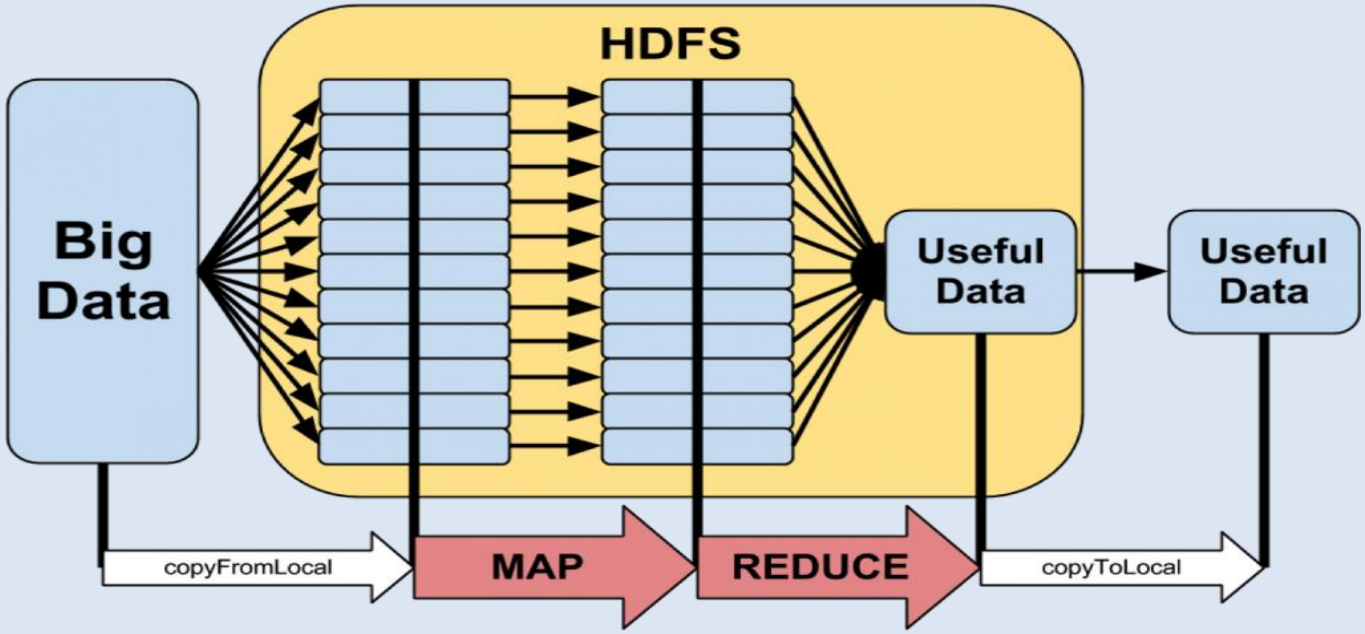
Reduce: Blends and solves the results of two or more map functions running in parallel.

- The MapReduce program can be developed in different languages.
- Google uses MapReduce to index web pages.

Big Data Analysis for IoT



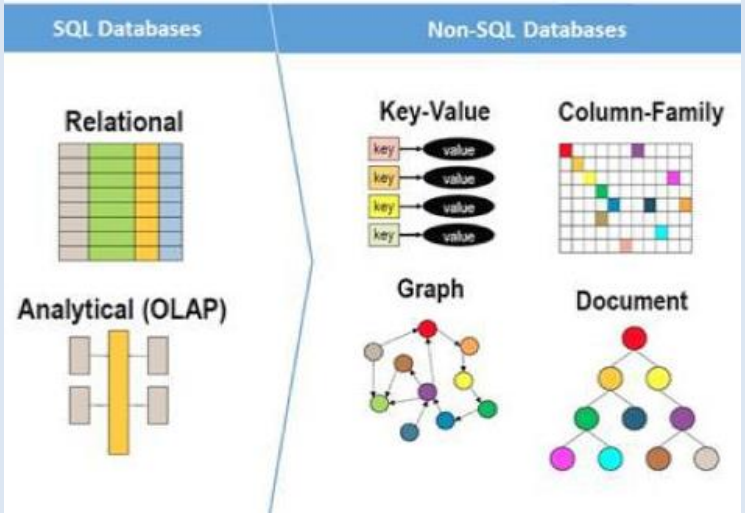

Big Data
Split 1
Split 2
Split 3
Split 4
Split 5
Split 6
Split 7
Split 8
Split 9
Split 10
Split N
Map 1
Map 2
Map 3
Map 4
Map 5
Map 6
Map 7
Map 8
Map 9
Map 10
Map N
Reduce 1
Reduce 2
Reduce N
Output 1
Output 2
Output N


HDFS
Big Data
Useful Data
Useful Data
copyFromLocal
MAP
REDUCE
copyToLocal

Big Data Analysis for IoT

NoSQL

- It is named 'Not Only SQL' because the SQL language is not used.
- Examples of systems using NOSQL are Big Table, DynamoDB, which is used by Google and Amazon.
- A life span for the data can be determined (a key can be connected), and the data that does not need to be kept very long can be held in NoSQL.
- In sites with very high data traffic, the data can be kept in columns to speed up the process of reading and writing from the database.
- The data can be stored as a document. (Mongodb)
- Graph-based databases (Graph), between users making it easier to capture links / associations (social networks).
- Hierarchical databases, geographic database types.



Big Data Analysis for IoT

- It is an open-source NoSQL database that uses a document-driven data model.
- Unlike relational databases, It is built on architecture collections and documents.
- Similar to NoSQL databases, It supports dynamic schema design that allows for documents in a collection are different areas and structures.
- Uses BSON as the format of document storage and data exchange. Binary representation JSON-like documents.
- Key Features
  - Master supports slave replication.
  - Easily store files regardless of size.
  - It can work on multiple servers.
  - It has an automatic load balancing configuration due to the data placed on the parts.

Big Data Analysis for IoT

- Apache Storm is a system used to process streaming (stream) data in real time.
- A distributed real-time for processing high volume data calculation system.
- The features:

  - High speed operation (more than one million records per node)
  - Easy to use (Standard configurations)
  - Reliable (Provides at least one processing of each data unit (tuple). messages are repeated when malfunctions occur)
  - Scalability (Parallel calculation working on a set of machines)
  - Fault tolerance (When workers die, the storm automatically restarts them. if the node dies, the worker restarts on another node)

Big Data Analysis for IoT

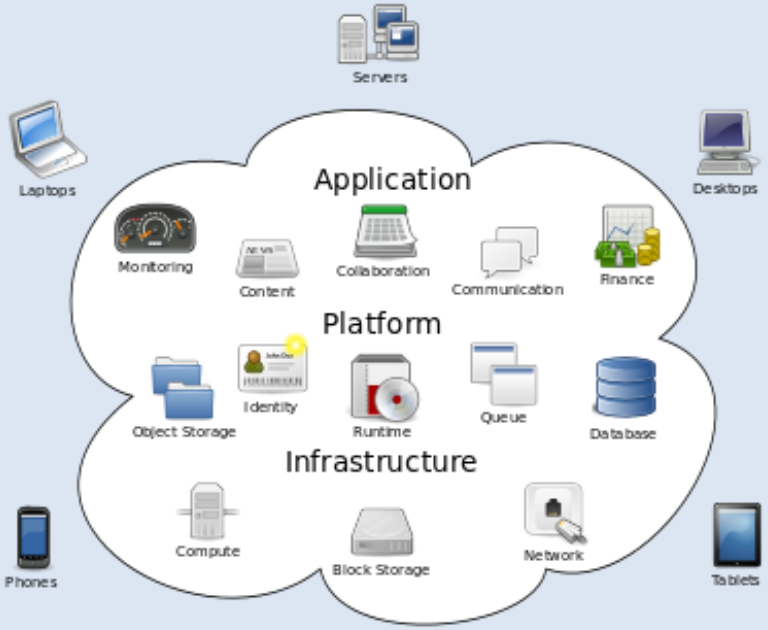
elasticsearch

- Developed in 2010, distributed, scalable, open source, enterprise-grade search engine.
- It is a NoSQL.
- The Features:

  - Fast and responsive search for large volumes of data
  - Indexing documents to warehouse
  - Non-normalized document storage, fast and direct access to data
  - Wide distributability and high scalability

Cloud Computing

- Cloud conceptually refers to a large number of device / computer-based data centers and a network system that allows users to access these data sources over the Internet.
- Cloud Computing is a technology that allows users to Access share and share information on resources (storage areas, devices, etc.) over the Internet.



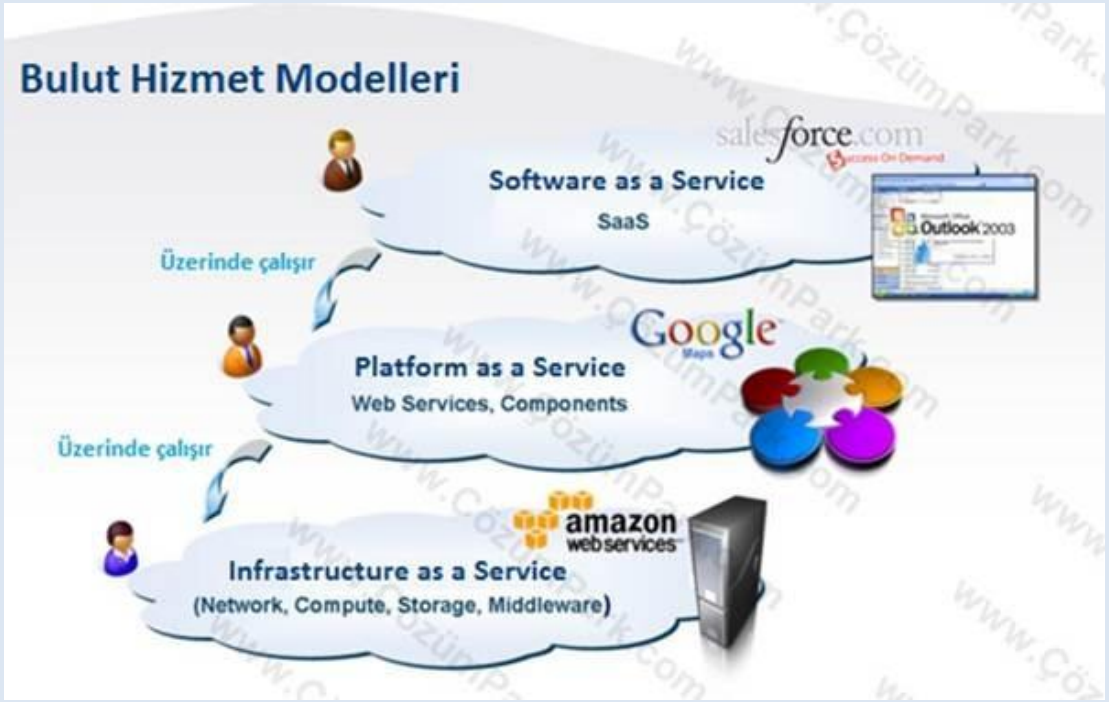
Cloud computing

Benefits of Cloud Computing

- Flexibility and Efficiency: Ease of capacity increase and decrease,
- Possibility to use anywhere with internet access,
- Security,
- Software update,
- Cost gain (low),

Cloud Computing

- It provides cloud technology, software, platform and infrastructure services.
- Physical hardware (storage areas) infrastructure, virtualization and so on.
- Data field structure, accessibility etc. for application development. (Windows Azure, Google Apps, etc.),
- As a software, providers are required to install the cloud application software and provide access services,



Big Data and Cloud Computing

Challenges encountered

Synchronization: Cloud system side devices are not able to receive data simultaneously

Standardization: Communication difficulty between devices of different manufacturers on the IOT side

Balancing: Providing balanced services for different infrastructures in cloud computing

Security: Security between cloud and devices

Management: IoT devices are separate from Cloud Computing resources and components

Reliability: Verification of the results users want

Big Data and Cloud Computing

- There are many cloud computing applications for IOT sensors and endpoints.
- They generally provide APIs for connecting to cloud systems.
- They can support protocols that sensor networks can connect. Almost all have RESTful services for IOT devices.
- The results can be exported as XML, CSV, as well as graphical interfaces.
- A warning or an alarm can be given according to the IOT data.
- The end devices can be controlled according to IOT data.

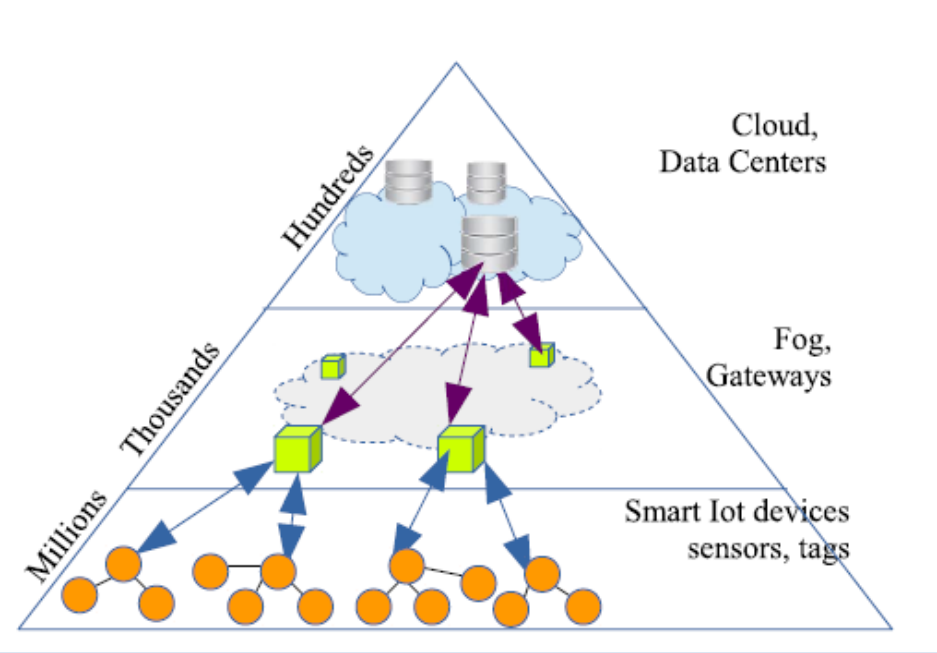
Big Data and Cloud Computing

- Platform Comparison

Platform	Gateway	Provision	Assurance	Billing	Application Protocol			
					REST	CoAP	XMPP	MQTT
Arkessa	-	+	+	-	+	-	-	+
Axeda	+	+	+	+	+	-	-	-
Etherios	+	+	+	-	+	-	-	-
LittleBits	-	-	-	-	+	-	-	-
NanoService	+	+	+	-	+	+	-	-
Nimbits	-	-	-	-	+	-	+	-
Ninja Blocks	+	-	-	-	+	-	-	-
OnePlatform	+	+	+	-	+	+	+	-
RealTime.io	+	+	-	-	+	-	-	-
SensorCloud	+	+	-	-	+	-	-	-
SmartThings	+	+	-	-	+	-	-	-
TempoDB	-	-	-	-	+	-	-	-
Thingworx	-	+	+	-	+	-	-	+
Xively	+	+	+	+	+	-	-	+

Fog Computing for IoT

- Instead of sending the data they produce to a central server for processing, the IoT devices are first analyzed at a local point, and the analysis is based on sending as much data as needed to the central server (cloud).
- Fog Informatics, Cloud Computing and IOT devices and acts as a bridge.
- Reduces response time for IOT and provides better performance.
- Reduce the use of band width.
- Increases security and privacy.



Fog Computing for IoT

The advantages of Fog on IoT:

- Location: Fog Information is closer to IOT devices.
- Distributed: Small parts can be separated and micro centers can be created.
- Scalability: Since installation is less costly than cloud systems, new micro centers can be created in limit transports.
- Reduce Device Density
- Mobile Support
- Real-Time Communication: The response time is faster than the cloud system.
- Preliminary Analysis: Data size can be reduced by making meaning.

References

- A. Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, M. Ayyash, "*Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications*", IEEE Communication Survey&Tutorials, vol. 17 (4), 2347-2376 ,2015.

- C. Tsai, C. Lai, M. Chiang, and L. T. Yang, "Data mining for Internet of Things: A survey," IEEE Commun. Surveys Tuts., vol. 16, no. 1, pp. 77– 97, 1st Quart. 2014

- TÜBİTAK Bulut Bilişim ve Büyük Veri Araştırma Laboratuvarı, b3lab.org

- X. Cheng, L. Fang, X. Hong, and L. Yang, "Exploiting Mobile Big Data: Sources, Features, and Applications," IEEE Network, vol. 16, no. 1, pp. 72-79, 2017