Internet of Things

Programming

Week 5 – Data Management

Anas Dawod

adawod@swin.edu.au

Swinburne University of Technology

March 2023

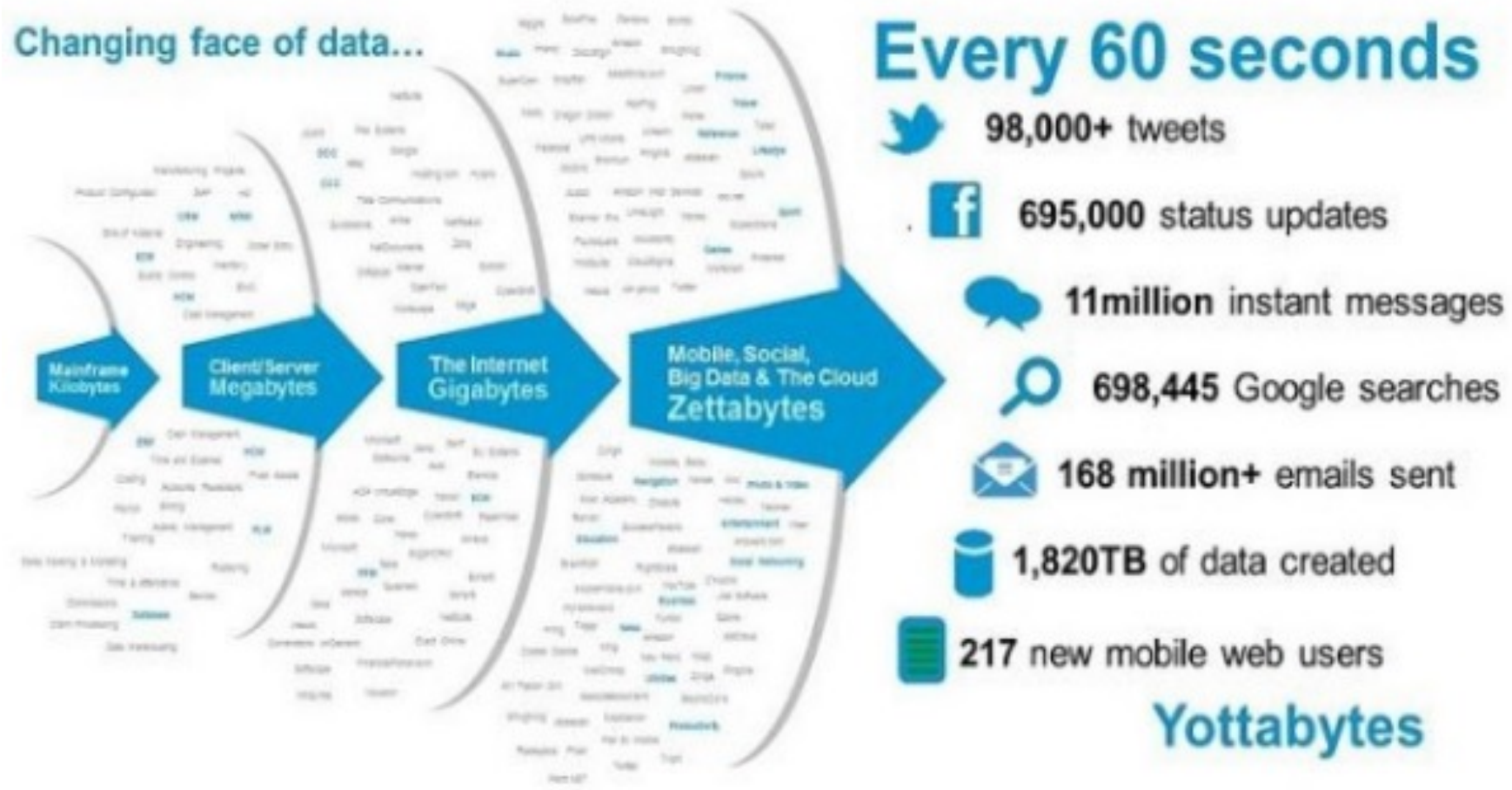
Announcements

- Survey Paper Assignment Due date – 31/03 @5:00pm.
- Individual Practical Assignment will be available next week.

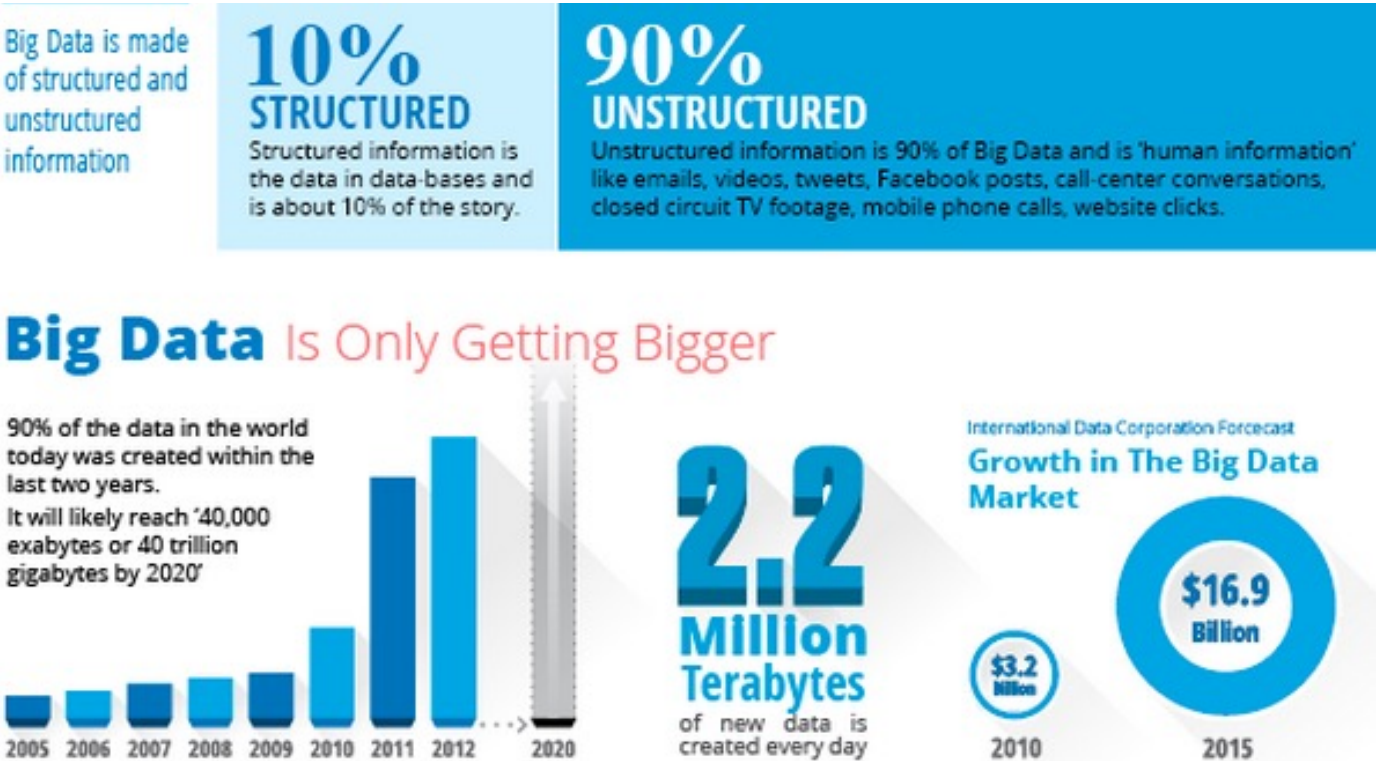
Questions…?

"From the dawn of civilization to 2003, five exabytes of data were created. The same amount was created in the last two days.“

Eric Schmidt (Former Google CEO)



https://www.slideshare.net/asertseminar/big-data-34369979



http://hrboss.com/hiringboss/articles/big-data-infographic

Activities

Conversation

Manufacturing

Smart Cars

Social Media

Sensors

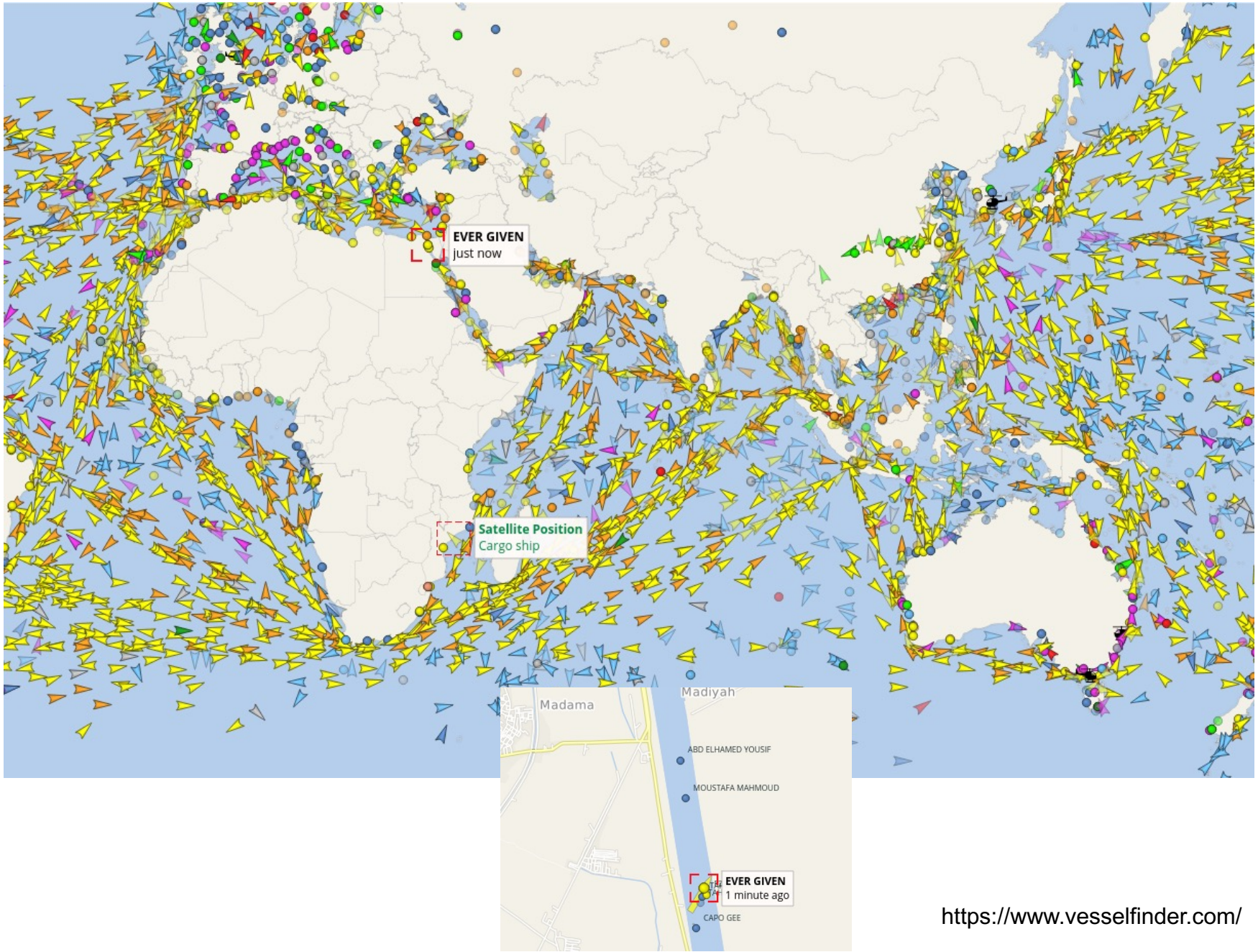
Military

Monitoring

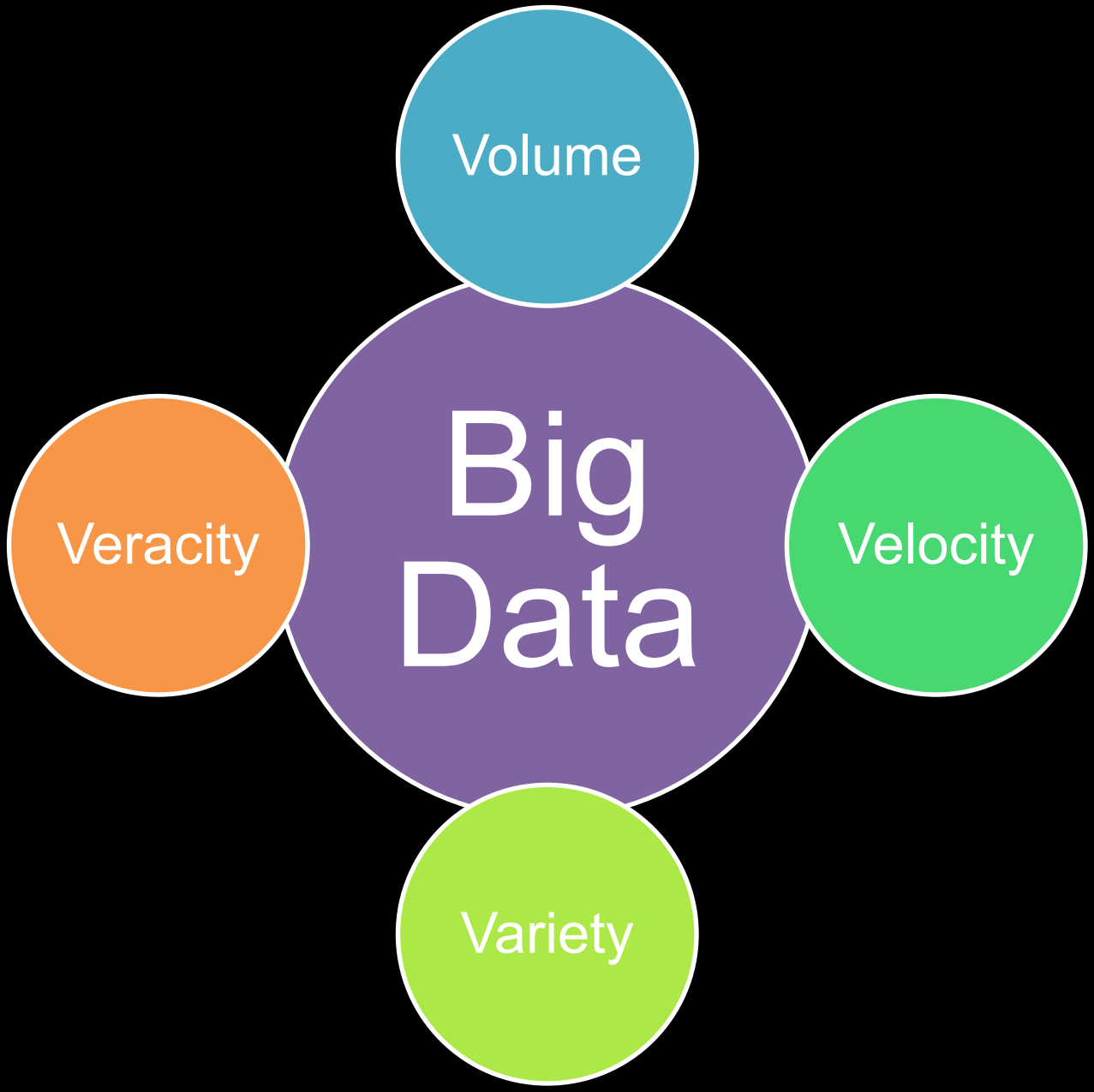
Videos

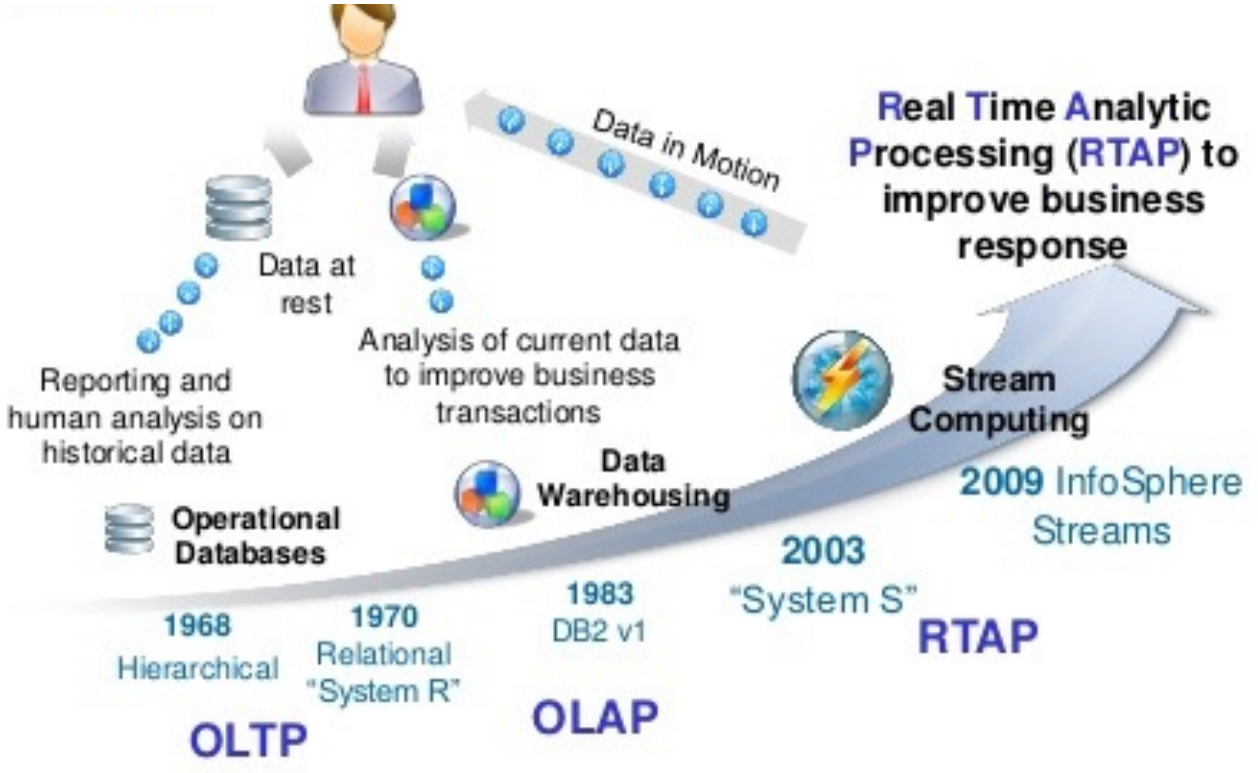
Internet of Things

Photos



https://www.vesselfinder.com/


Volume
Veracity
Big Data
Velocity
Variety



OLTP: Online Transaction Processing (DBMSs)
OLAP: Online Analytical Processing (Data Warehousing)
RTAP: Real-Time Analytics Processing (Big Data Architecture & technology)

http://faculty.bcitbusiness.ca/kevinw/4800/Lecture_Slides/Yallouz_Lectures/2016_BUSA_4800_Big_Data_Lecture.pdf

The Model Has Changed…

- **The Model of Generating/Consuming Data has Changed**

Old Model: Few companies are generating data, all others are consuming data

New Model: all of us are generating data, and all of us are consuming data

http://faculty.bcitbusiness.ca/kevinw/4800/Lecture_Slides/Yallouz_Lectures/2016_BUSA_4800_Big_Data_Lecture.pdf

Turning Big Data into Value

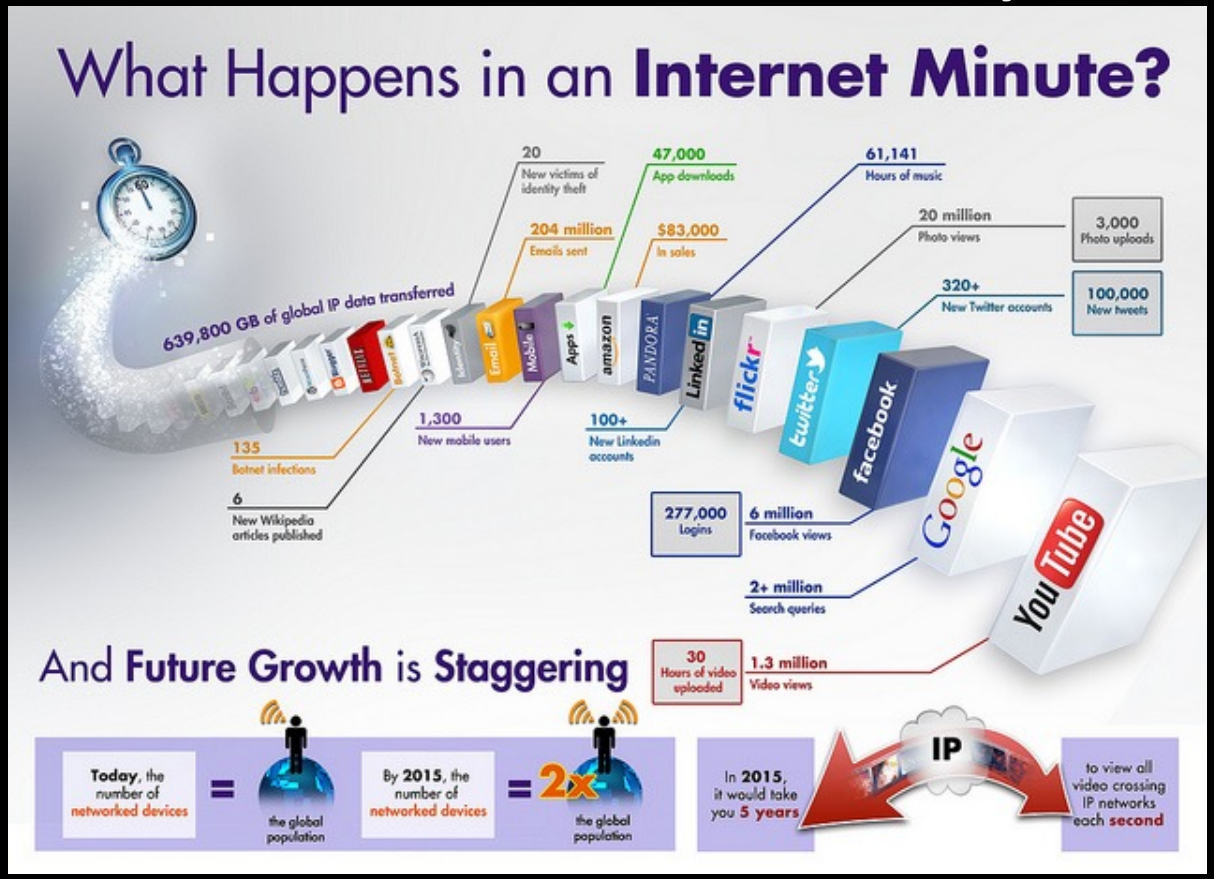
small data

big data

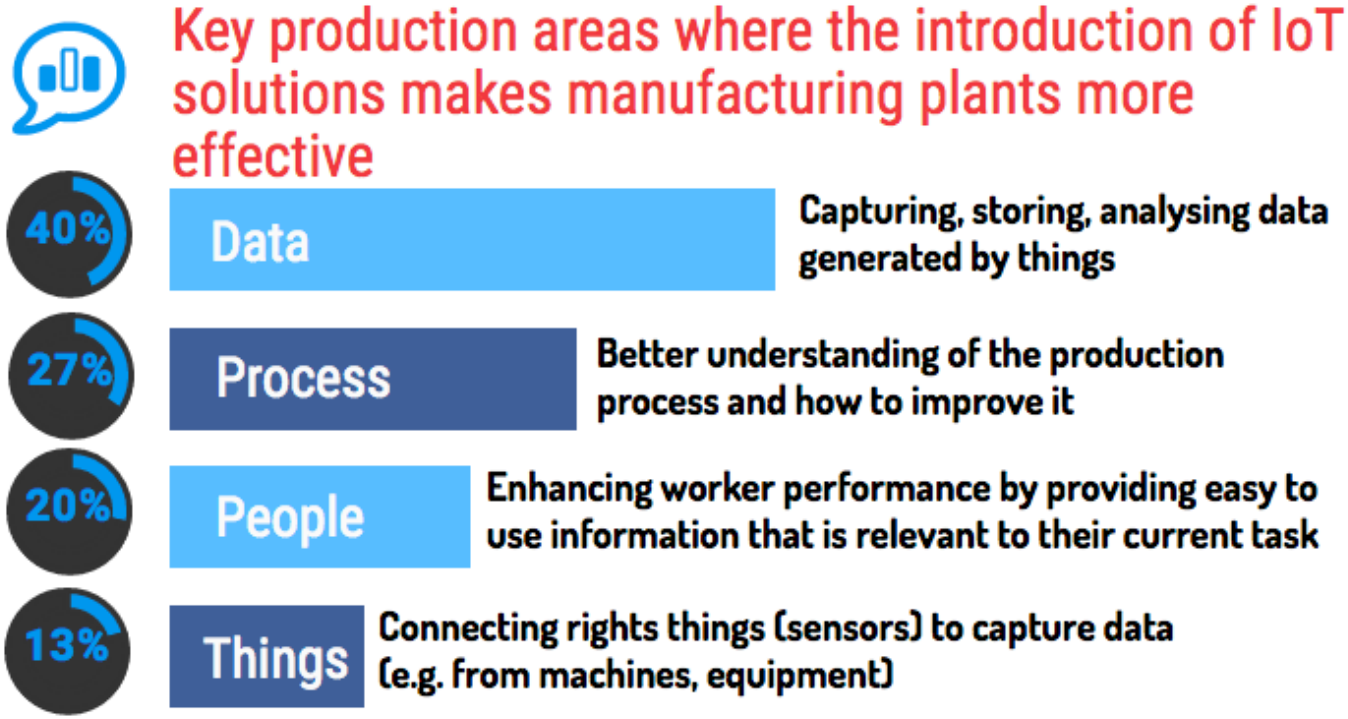
“Not everything that can be counted counts, and not everything that counts can be counted. ”

— William Bruce Cameron (1963), Informal Sociology: a casual introduction to sociological thinking

0.5% of All Data is Analysed

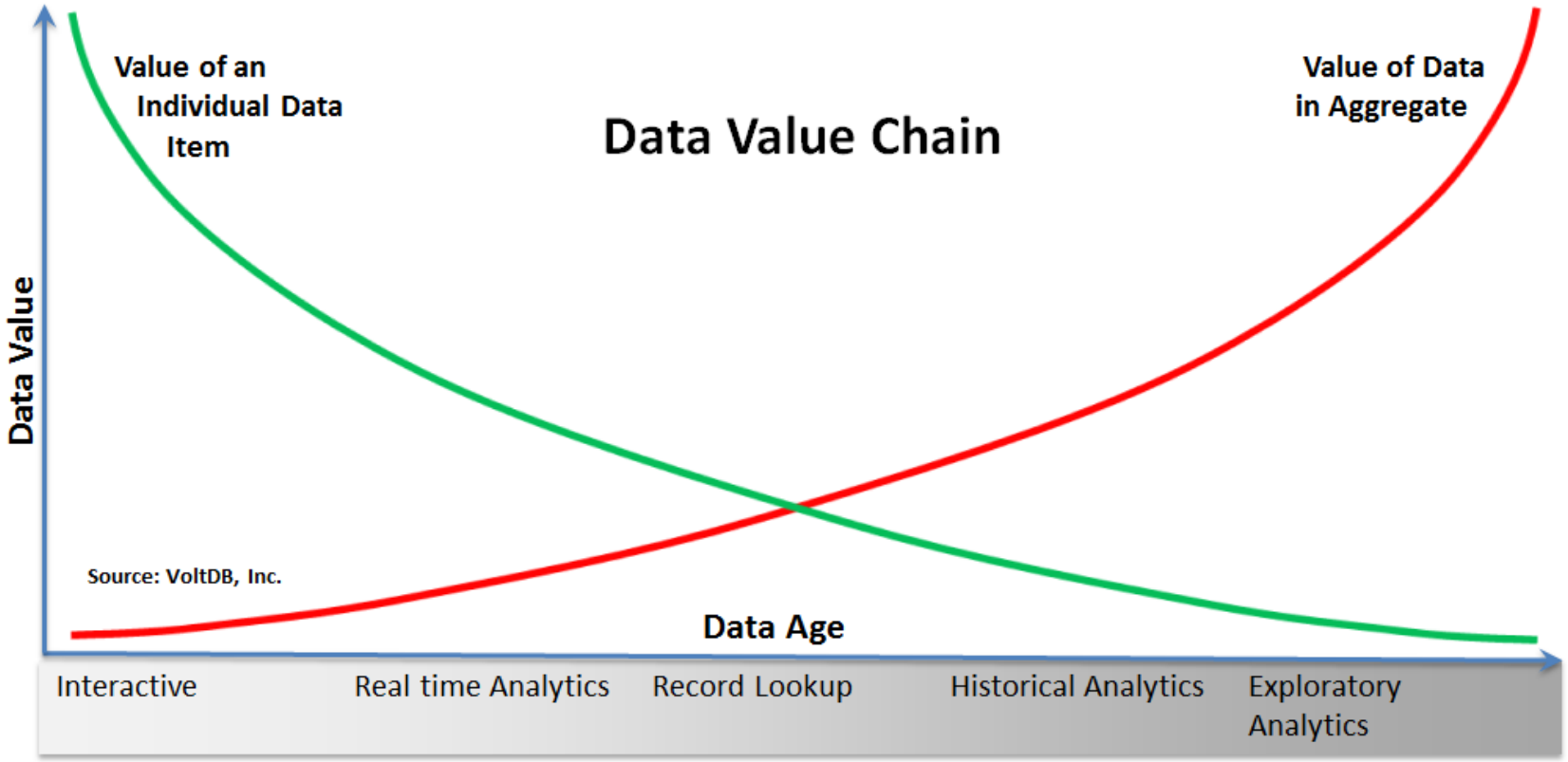


MIT Technology Review - Antonio Regalado - 2013



CISCO, 2016, http://www.cisco.com/c/dam/en_us/solutions/trends/iot/docs/iot-data-analytics-white-paper.PDF (2016)

IoT Data


Value of an Individual Data Item
Data Value Chain
Value of Data in Aggregate
Data Value
Source: VoltDB, Inc.
Data Age
Interactive
Real time Analytics
Record Lookup
Historical Analytics
Exploratory Analytics

Data Management in IoT… Why do we need it?

IoT data can be in structured, semi-structured or unstructured formats.

Structured data:

- is represented using some model or schema, can associated with a database, is represented as tabular representation (like a spreadsheet), can be easily formatted, queried, and processed to use in decision-making.

Unstructured data:

- Does not follow a model or schema, mostly exists in raw form.

Semi-structured data:

- Hybrid of structured and unstructured data and share the characters of both.

Mahanthappa, S., Chandavarkar, B.R. (2021). Data Formats and Its Research Challenges in IoT: A Survey. In: Suma, V., Bouhmala, N., Wang, H. (eds) Evolutionary Computing and Mobile Sustainable Networks. Lecture Notes on Data Engineering and Communications Technologies, vol 53. Springer, Singapore. https://doi.org/10.1007/978-981-15-5258-8_47

Data Management in IoT… Why do we need it?

For IoT to handle a high volume of data, the following operations need to be performed:

- Handle heterogeneous data,
- Prepare data for the analysis,
- Aggregate, integrate, and keep track of data origin,
- Preserve integrity and privacy of the data,
- Choose a storage that can balance between performance, reliability, flexibility, and cost.

Also considering that the high volume of IoT data is processed at a high rate. The processing of the data should occur closer to the event environment to avoid delay and loss of data.

Mahanthappa, S., Chandavarkar, B.R. (2021). Data Formats and Its Research Challenges in IoT: A Survey. In: Suma, V., Bouhmala, N., Wang, H. (eds) Evolutionary Computing and Mobile Sustainable Networks. Lecture Notes on Data Engineering and Communications Technologies, vol 53. Springer, Singapore. https://doi.org/10.1007/978-981-15-5258-8_47

Data Management in IoT… Why do we need it?

Most IoT data exists in semi-structured or unstructured formats. Therefore, we need a data management system to connect to all of the interconnected sensors/devices and adhere to the various protocols so the data from those systems can be efficiently recorded, stored, transmitted, and analysed.

IoT Data Formats

Major data formats generated by IoT sensors and applications

- Raw Data (Text, Binary) – simple and less detailed
- XML
- JSON

  - Specialised encoding more details (device status, meta data, captured data)

- Others...

XML

- XML stands for eXtensible Markup Language
- XML is a markup language much like HTML
- XML was designed to store and transport data
- XML was designed to be self-descriptive
- XML is a W3C* Recommendation

```
<note>
  <to>Tove</to>
  <from>Jani</from>
  <heading>Reminder</heading>
  <body>Don't forget me this
weekend!</body>
</note>
```

*W3C: The World Wide Web Consortium is the main international standards organisation for the World Wide Web

JSON

- JavaScript Object Notation (JSON)
- Open-standard file format that uses human-readable text to transmit data
- JSON is a language-independent data format

```json
{
  "firstName": "John",
  "lastName": "Smith",
  "isAlive": true,
  "age": 27,
  "address": {
    "streetAddress": "21 2nd Street",
    "city": "New York",
    "state": "NY",
    "postalCode": "10021-3100"
  },
  "phoneNumbers": [
    {
      "type": "home",
      "number": "212 555-1234"
    },
    {
      "type": "office",
      "number": "646 555-4567"
    },
    {
      "type": "mobile",
      "number": "123 456-7890"
    }
  ],
  "children": [],
  "spouse": null
}
```

JSON vs. XML

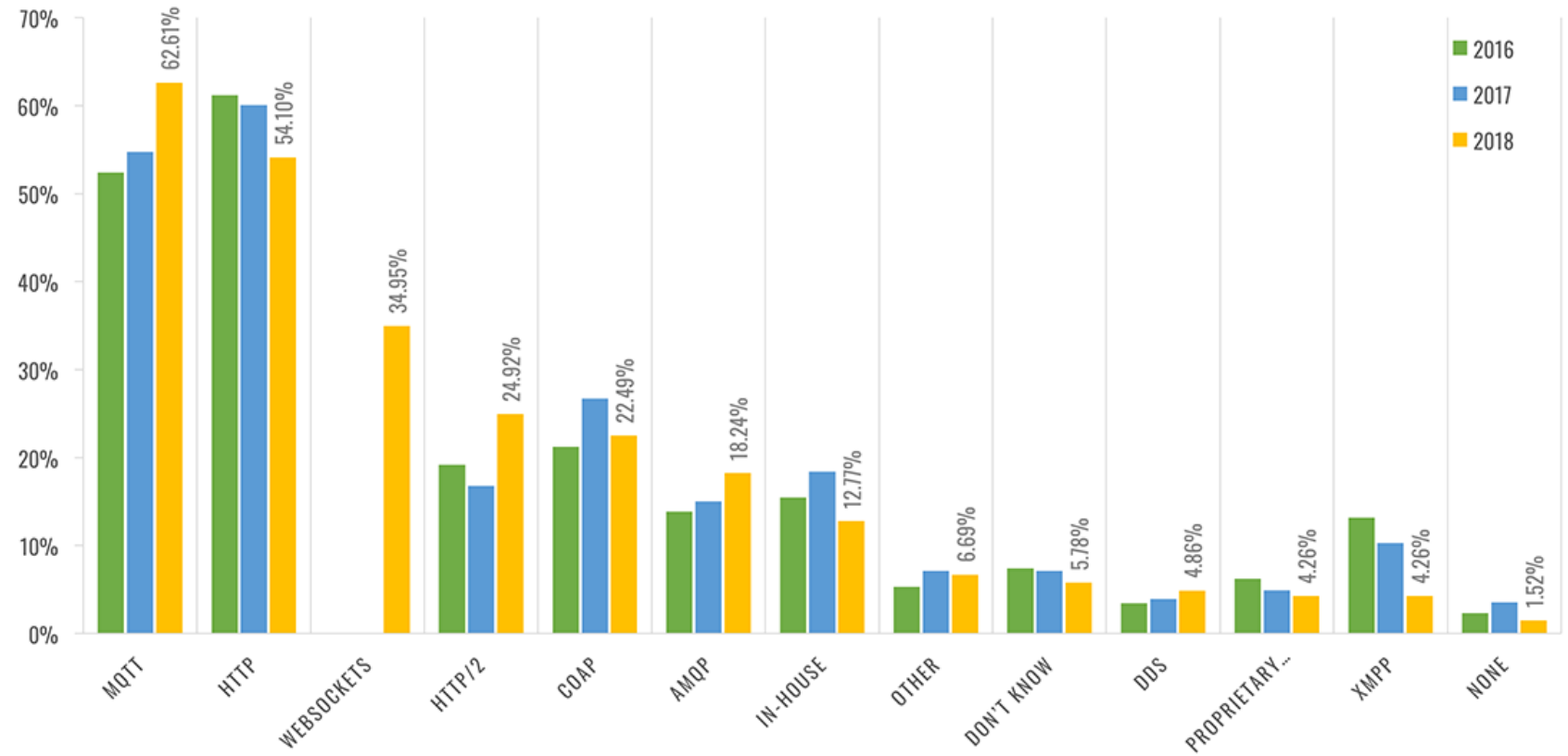
- Both JSON and XML are "self describing" (human readable)
- Both JSON and XML are hierarchical (values within values)
- Both JSON and XML can be parsed and used by lots of programming languages
- XML is much more difficult to parse than JSON.

```
{"widget": {
    "debug": "on",
    "window": {
        "title": "Sample Konfabulator Widget",
        "name": "main_window",
        "width": 500,
        "height": 500
    },
    "image": {
        "src": "Images/Sun.png",
        "name": "sun1",
        "hOffset": 250,
        "vOffset": 250,
        "alignment": "center"
    },
    "text": {
        "data": "Click Here",
        "size": 36,
        "style": "bold",
        "name": "text1",
        "hOffset": 250,
        "vOffset": 100,
        "alignment": "center",
        "onMouseUp": "sun1.opacity = (sun1.opacity / 100) * 90;"
    }
}}
```

```
<widget>
    <debug>on</debug>
    <window title="Sample Konfabulator Widget">
        <name>main_window</name>
        <width>500</width>
        <height>500</height>
    </window>
    <image src="Images/Sun.png" name="sun1">
        <hOffset>250</hOffset>
        <vOffset>250</vOffset>
        <alignment>center</alignment>
    </image>
    <text data="Click Here" size="36" style="bold">
        <name>text1</name>
        <hOffset>250</hOffset>
        <vOffset>100</vOffset>
        <alignment>center</alignment>
        <onMouseUp>
            sun1.opacity = (sun1.opacity / 100) * 90;
        </onMouseUp>
    </text>
</widget>
```

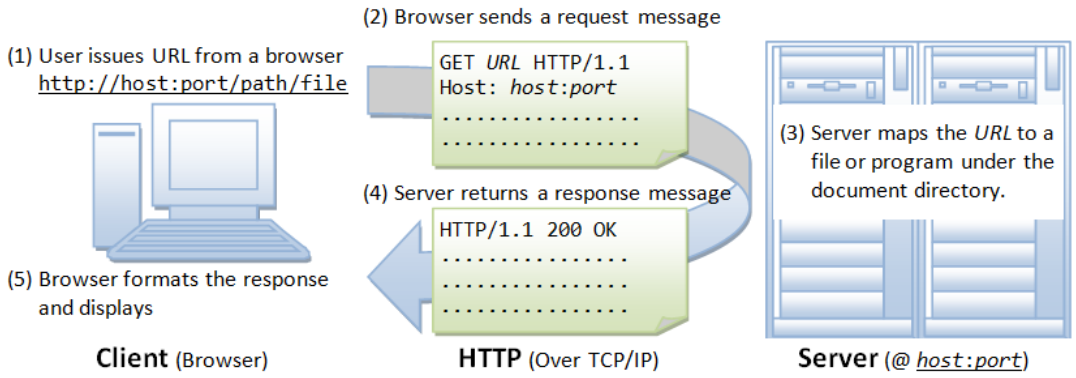
w3schools.com and json.org

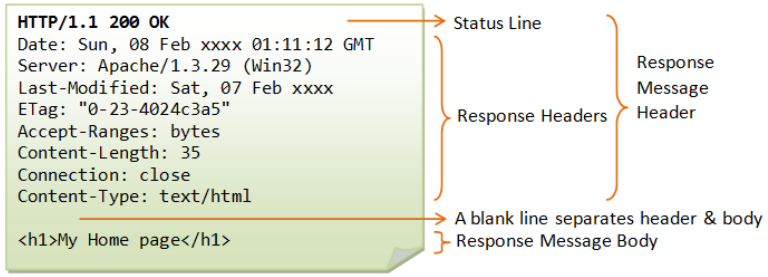
IoT Transferring Data



HTTP

- Hypertext Transfer Protocol (HTTP)
- HTTP works as a request-response protocol between a client and server.
- HTTP messages are composed of textual information encoded in ASCII, and span over multiple lines





Websockets

- A computer communications protocol, providing full-duplex communication channels over a single TCP connection.
- The WebSocket protocol was standardized by the IETF as RFC 6455 in 2011
- WebSockets provide a persistent connection between a client and server that both parties can use to start sending data at any time.

CoAP

- Sensors Network Nodes are usually constrained platforms (low energy, minimise network traffic use, etc.)
- Constrained Application Protocol (CoAP) is a specialised Internet Application Protocol for constrained devices
- CoAP is designed to easily translate to HTTP for simplified integration with the web
- Low overhead protocol designed for environments like low-end microcontroller boards
- Client/Server connection

https://opensourceforu.com/2016/09/coap-get-started-with-iot-protocols/

MQTT

- Message queuing telemetry protocol
- Initially developed by IBM and Eurotech
- A lightweight message queuing and transport protocol
- Low overhead (2 bytes)
- Run on TCP
- Publish Subscribe Model
- Machine to Machine communication
- Ideal for Internet of Things

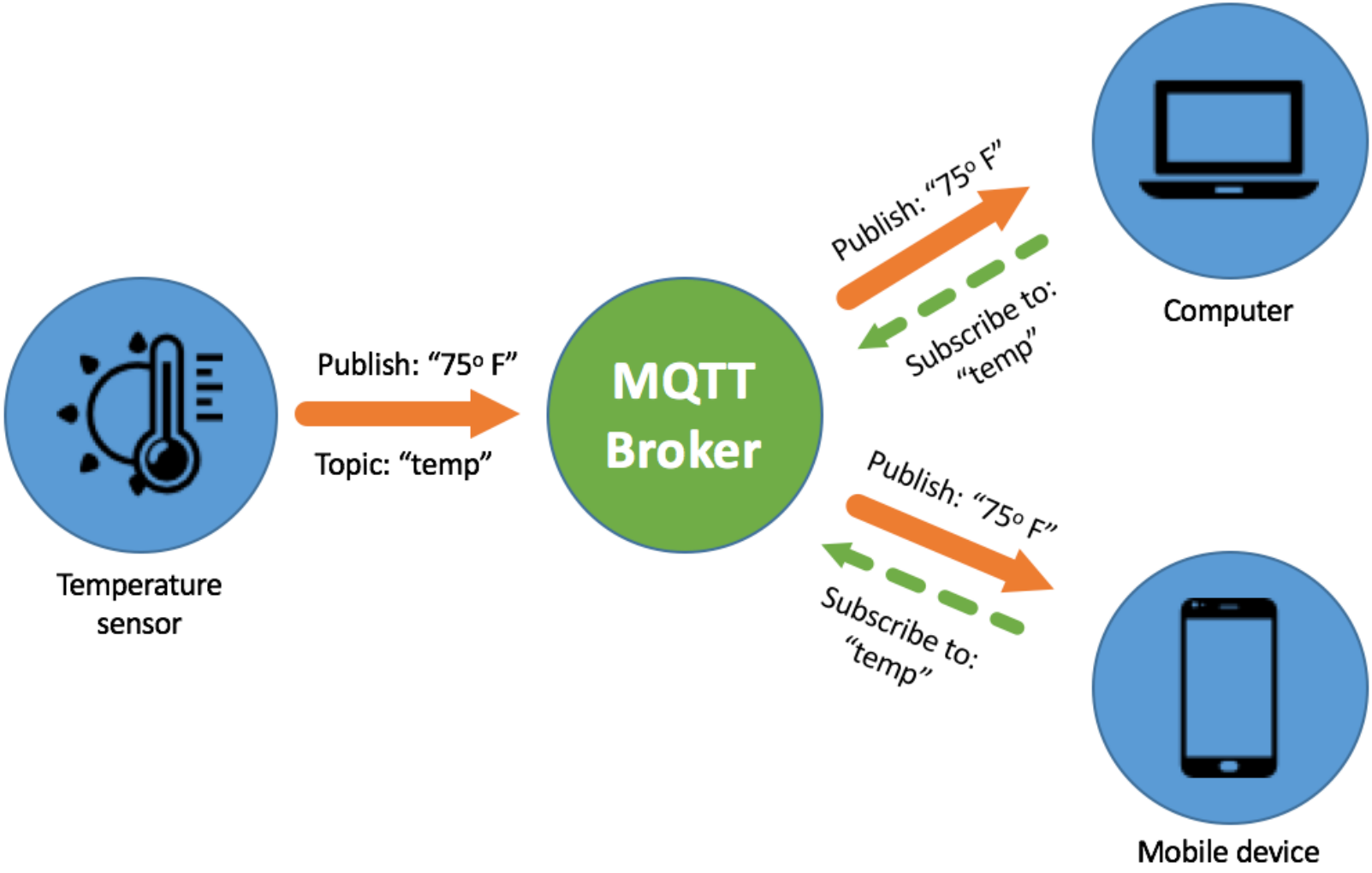
Publish Subscribe pattern

The sender of the messages (publisher) does not program the message to be sent directly to specific receiver (subscriber)

The sender publish the message in a topic without knowing which subscribers, if any, there may be.

The receiver subscribes to a specific topic

- Publisher defines the topic
- Loose coupling
- Scalable



http://www.innorobix.com/mqtt

MQTT Topic

Topics are represented with strings separated by slashes

Home/Bedroom_1/Light_1

Home/Bedroom_1/Light_2

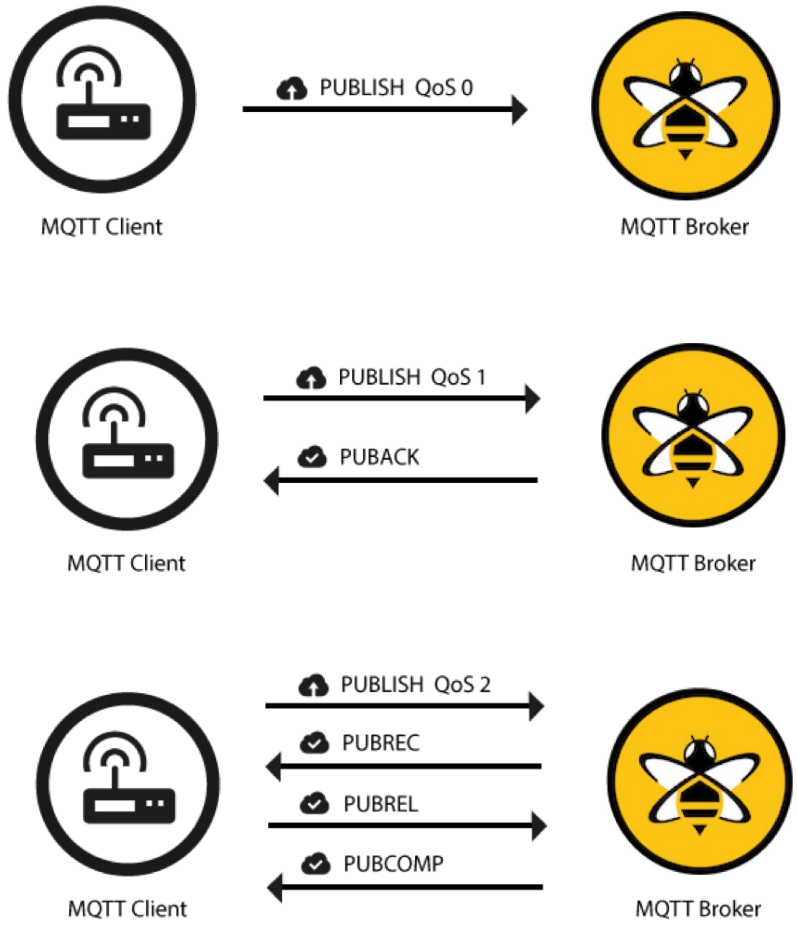
Home/Bedroom_2/Light_1

Home/Bedroom_3/Light_1

MQTT brokers

Quality of Service

- A sender and receiver agrees on QoS level
- There are 3 QoS Levels
  - QoS0 : At most once
    - Sender sends the message and doesn't wait for acknowledge
  - QoS1: At least once
    - Sender sends the message and waits for the acknowledge
  - Qos2: Exactly once
    - Sender sends the message and waits for the double acknowledge



CoAP vs MQTT

- MQTT and CoAP are both useful as IoT protocols
- MQTT is a many-to-many communication protocol for passing messages between multiple clients through a central broker.
- CoAP is, primarily, a one-to-one protocol for transferring state information between client and server.
- MQTT provides no support for labelling messages with types or other metadata to help clients understand it. Clients must know the message formats up-front.

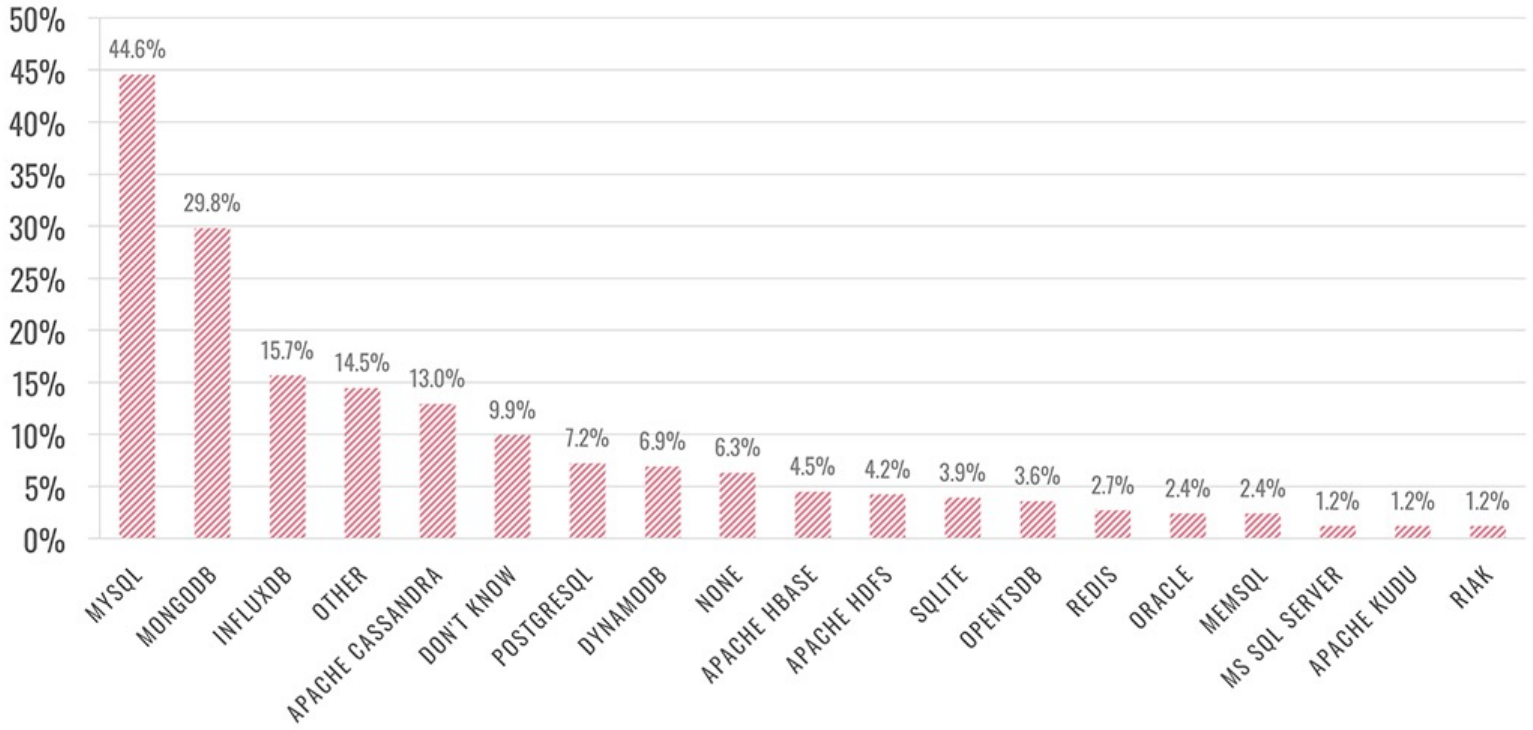
https://www.eclipse.org/community/eclipse_newsletter/2014/february/

Questions…?

IoT Storing Data

IoT Databases

Which of the following database technologies do you use in your IoT solution?



SQL

- Structured Query Language
- We have Schema
- Data distributed across multiple tables
- Horizontal scaling is difficult
- Vertical scaling is possible
- SQL databases are good fit for the complex query
- SQL databases are primarily called as Relational Databases (RDBMS)
- SQL database examples: MySql, Oracle, Sqlite, Postgres and MariaDB
- Allows you to write queries

Table Structure

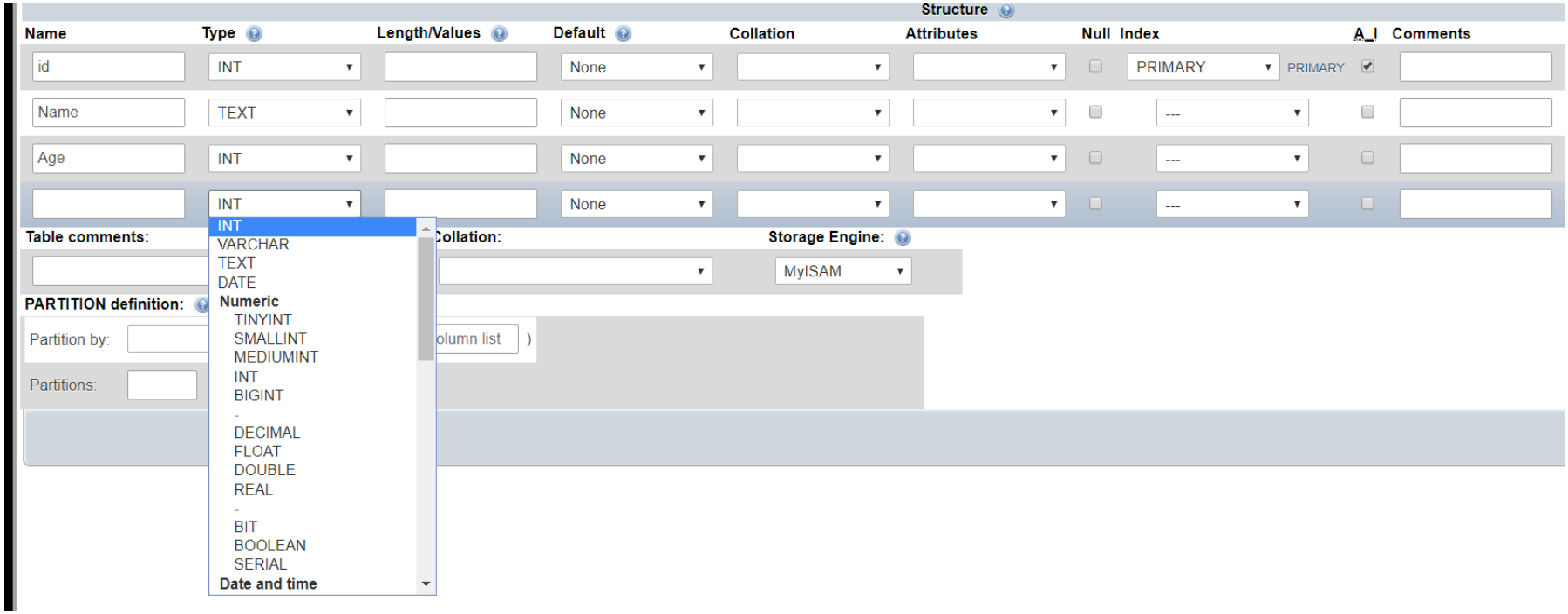


Table Structure

test.sql

```
##  Creating DB

CREATE DATABASE IF NOT EXISTS IoT_Programming;
USE IoT_Programming;


##  Creating Tables

CREATE TABLE IF NOT EXISTS IoT_Programming.NameList (
    id int NOT NULL AUTO_INCREMENT,
    name varchar(50) NOT NULL,
    age int NOT NULL,
    PRIMARY KEY (id)
);
```

```
mysql> source test.sql
Query OK, 1 row affected (0.01 sec)

Database changed
Query OK, 0 rows affected (0.03 sec)
```

```
mysql> use IoT_Programming;
Database changed
mysql> show tables;
+---------------------------+
| Tables_in_IoT_Programming |
+---------------------------+
| NameList                  |
+---------------------------+
1 row in set (0.00 sec)

mysql> describe NameList;
+-------+-------------+------+-----+---------+----------------+
| Field | Type        | Null | Key | Default | Extra          |
+-------+-------------+------+-----+---------+----------------+
| id    | int         | NO   | PRI | NULL    | auto_increment |
| name  | varchar(50) | NO   |     | NULL    |                |
| age   | int         | NO   |     | NULL    |                |
+-------+-------------+------+-----+---------+----------------+
3 rows in set (0.00 sec)
```

SQL Syntax

```
mysql> INSERT INTO IoT_Programming.NameList (id, name, age)  VALUES (NULL, "IoT Programming", 1);
Query OK, 1 row affected (0.01 sec)

mysql> SELECT * FROM NameList;
+----+-----------------+-----+
| id | name            | age |
+----+-----------------+-----+
|  1 | IoT Programming |   1 |
+----+-----------------+-----+
1 row in set (0.00 sec)

mysql> DELETE FROM NameList WHERE id=1;
Query OK, 1 row affected (0.01 sec)

mysql> SELECT * FROM NameList;
Empty set (0.00 sec)
```

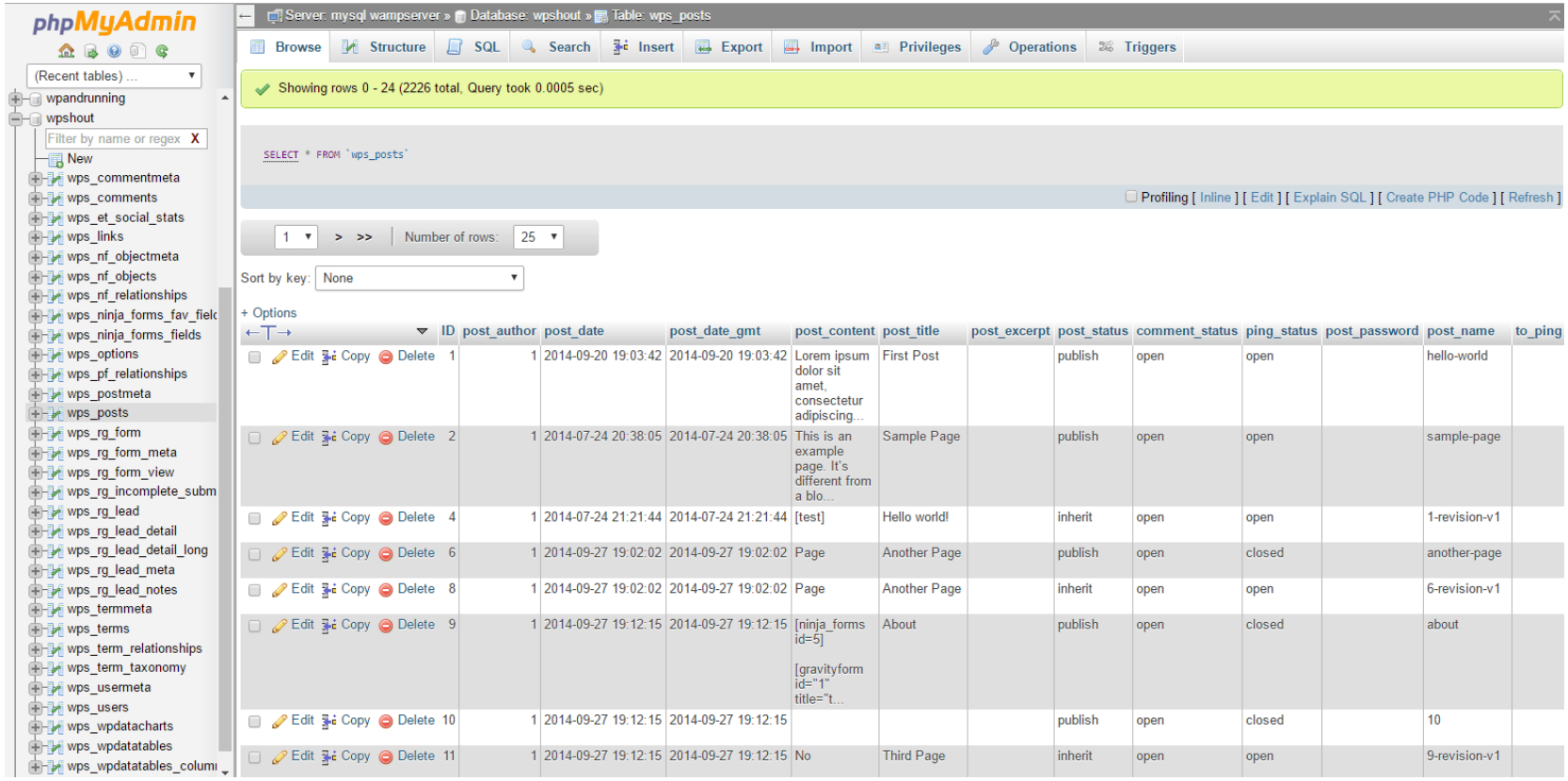
MariaDB (MySQL fork)

- Data uses schema
- Open source
- Free
- Used by Google, Facebook, Twitter, YouTube (MySQL)
- MariaDB 10.1 is now the default mysql server in Debian 9 "Stretch"
- Raspberry Pi OS is based on Debian
- Installation on Raspberry Pi
  - sudo apt-get install mariadb-server

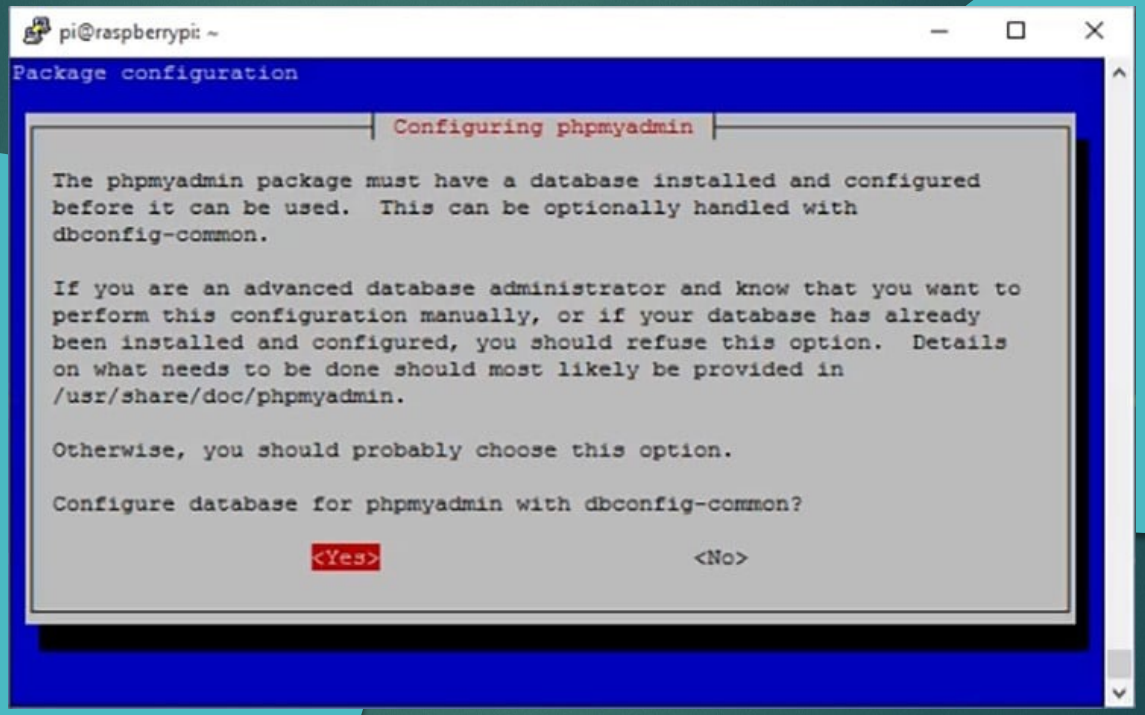
Week 6 Tutorial

- Installing MariaDB
- Granting privileges
- Create database with a table
- Write python script to read data from serial bus
- And store it in the database

phpMyAdmin


phpMyAdmin
(Recent tables) ...
wpandrunning
wpshout
Filter by name or regex
New
wps_commentmeta
wps_comments
wps_et_social_stats
wps_links
wps_nf_objectmeta
wps_nf_objects
wps_nf_relationships
wps_ninja_forms_fav_field
wps_ninja_forms_fields
wps_options
wps_pf_relationships
wps_postmeta
wps_posts
wps_rg_form
wps_rg_form_meta
wps_rg_form_view
wps_rg_incomplete_subm
wps_rg_lead
wps_rg_lead_detail
wps_rg_lead_detail_long
wps_rg_lead_meta
wps_rg_lead_notes
wps_termmeta
wps_terms
wps_term_relationships
wps_term_taxonomy
wps_usermeta
wps_users
wps_wpdatacharts
wps_wpdatatables
wps_wpdatatables_columi
Server: mysql wampserver » Database: wpshout » Table: wps_posts
Browse
Structure
SQL
Search
Insert
Export
Import
Privileges
Operations
Triggers
Showing rows 0 - 24 (2226 total. Query took 0.0005 sec)
SELECT * FROM `wps_posts`
Profiling [Inline] [Edit] [Explain SQL] [Create PHP Code] [Refresh]
1 > >> Number of rows: 25
Sort by key: None
+ Options
ID post_author post_date post_date_gmt post_content post_title post_excerpt post_status comment_status ping_status post_password post_name to_ping
Edit Copy Delete 1 1 2014-09-20 19:03:42 2014-09-20 19:03:42 Lorem ipsum dolor sit amet, consectetur adipiscing... First Post publish open open hello-world
Edit Copy Delete 2 1 2014-07-24 20:38:05 2014-07-24 20:38:05 This is an example page. It's different from a blo... Sample Page publish open open sample-page
Edit Copy Delete 4 1 2014-07-24 21:21:44 2014-07-24 21:21:44 [test] Hello world! inherit open open 1-revision-v1
Edit Copy Delete 6 1 2014-09-27 19:02:02 2014-09-27 19:02:02 Page Another Page publish open closed another-page
Edit Copy Delete 8 1 2014-09-27 19:02:02 2014-09-27 19:02:02 Page Another Page inherit open open 6-revision-v1
Edit Copy Delete 9 1 2014-09-27 19:12:15 2014-09-27 19:12:15 [ninja_forms id=5] [gravityform id="1" title="t... About publish open closed about
Edit Copy Delete 10 1 2014-09-27 19:12:15 2014-09-27 19:12:15 publish open closed 10
Edit Copy Delete 11 1 2014-09-27 19:12:15 2014-09-27 19:12:15 No Third Page inherit open open 9-revision-v1

- sudo mysql -u root –p
- GRANT ALL PRIVILEGES ON mydb.* TO pi@localhost IDENTIFIED BY raspberry;
- sudo apt-get install phpmyadmin



- sudo nano /etc/apache2/apache2.conf
- Add “Include /etc/phpmyadmin/apache.conf”

MongoDB

- Most popular NoSQL database
- Humongous (built to store lots of data)
- No schema
- No / few relations

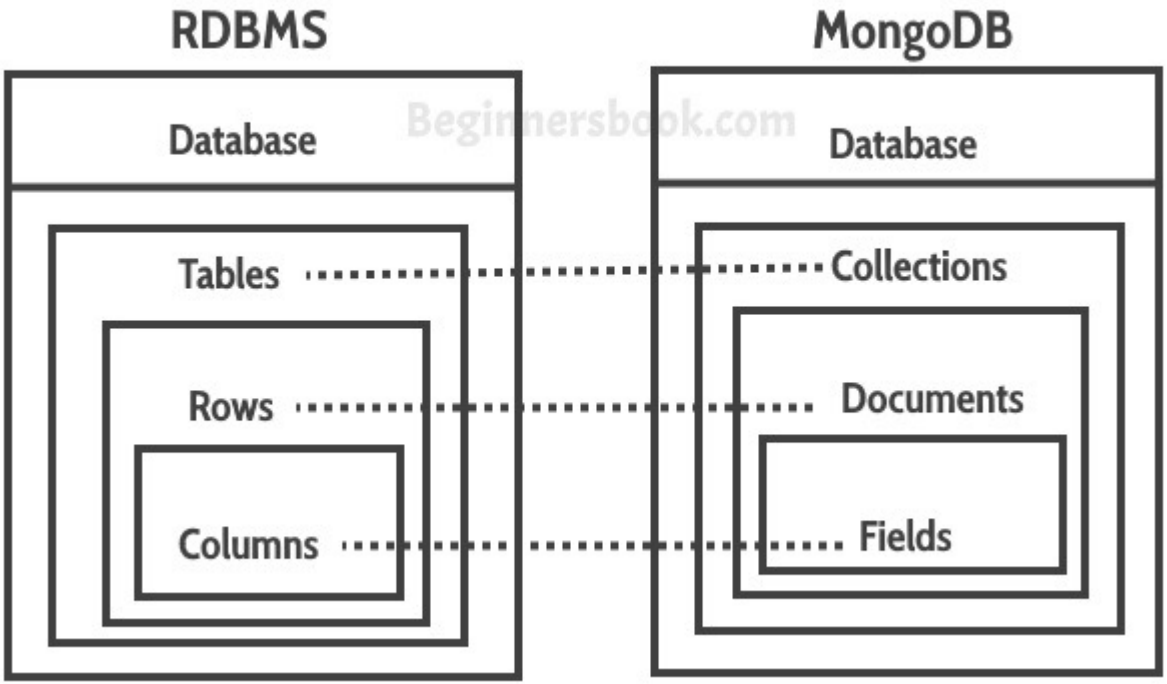
```
db.createUser({
user : '<userName>',
pwd : '<password>',
            roles : [ { role : '<roleName>', db : '<dbName>' } |
'<roleName>', …]
})
```

NoSQL

- Schema-less
- SQL databases are table based databases whereas NoSQL databases are document based, key-value pairs, graph databases or wide-column stores.
- No / few relations
- Data is typically merged
- Both horizontal and vertical scaling is possible
- Great performance for read/write requests
- MongoDB, BigTable, Redis, RavenDb, Cassandra, Hbase, Neo4j and CouchDb

SQL vs NoSQL

- Depends on kind of application and data
- Flexibility and predictable layout
- Relations are changing



https://beginnersbook.com/2017/09/mapping-relational-databases-to-mongodb/

Relational Database

Student_Id	Student_Name	Age	College
1001	Chaitanya	30	Beginnersbook
1002	Steve	29	Beginnersbook
1003	Negan	28	Beginnersbook

MongoDB

```
{
 "_id": ObjectId("....."),
 "Student_Id": 1001,
 "Student_Name": "Chaitanya",
 "Age": 30,
 "College": "Beginnersbook"
}
{
 "_id": ObjectId("....."),
 "Student_Id": 1002,
 "Student_Name": "Steve",
 "Age": 29,
 "College": "Beginnersbook"
}
{
 "_id": ObjectId("....."),
 "Student_Id": 1003,
 "Student_Name": "Negan",
 "Age": 28,
 "College": "Beginnersbook"
}
```

https://beginnersbook.com/2017/09/mapping-relational-databases-to-mongodb/

Questions…?