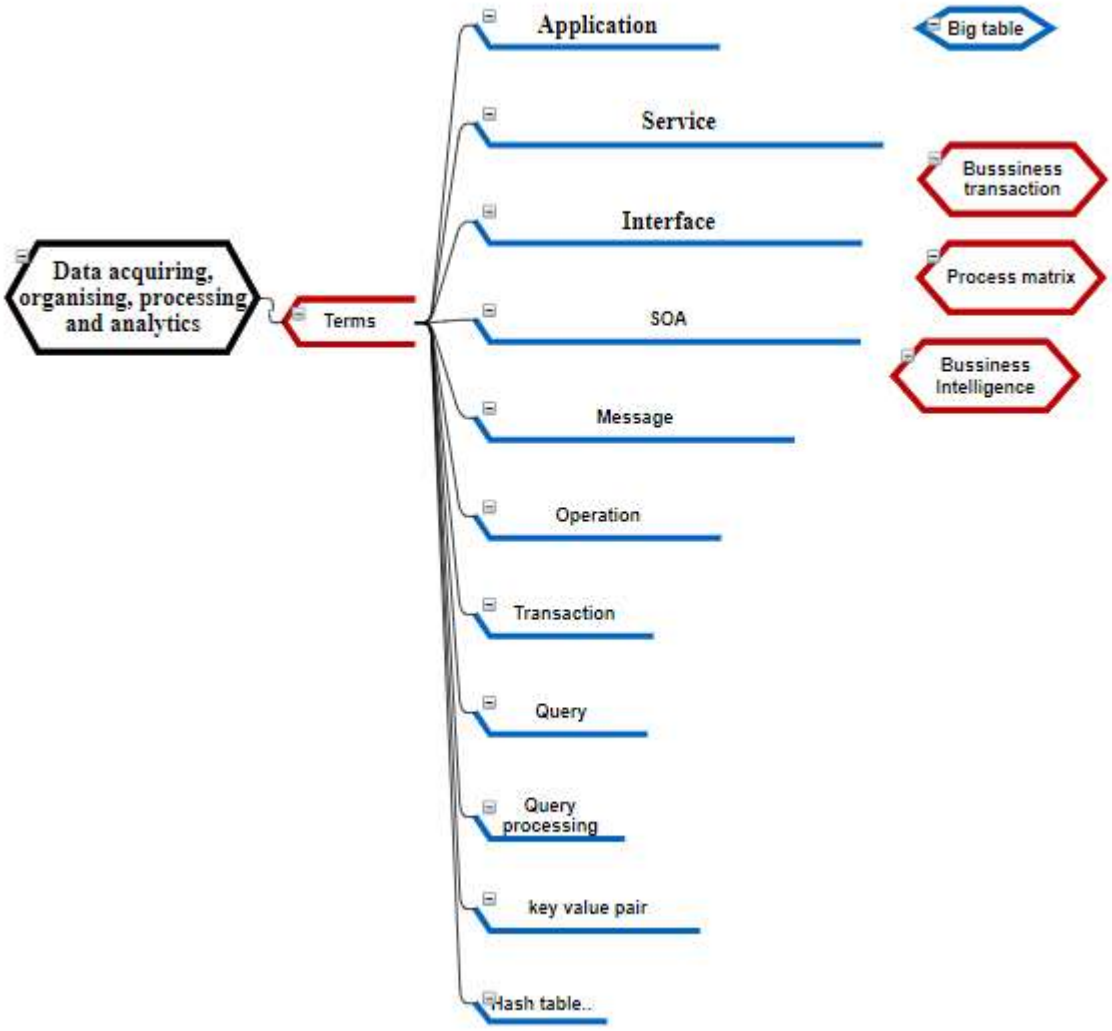
4unit iot - iot

Embedded Systems (Jawaharlal Nehru Technological University, Kakinada)

UNIT IV:

Solution framework for IoT applications: Implementation of Device integration, Data acquisition and integration, Device data storage- Unstructured data storage on cloud/local server, Authentication, authorization of devices.



Data acquiring, organising, processing and analytics
Application refers to application software or a collection of software components.
Service denotes a mechanism, which enables the provisioning of access to one or more capabilities.
An **interface** for the service provides the access to capabilities.
A **service** consists of a collection of self-contained, distinct and reusable components. It provides logically grouped and encapsulated functionalities.
Service consists of a set of related software components and their functionalities. The set is reused for one or more purposes. Usage of the set is consistent with the controls, constraints and policies which are specified in the service description for each service. A service also associates a Service Level Agreement (SLA).

Service Oriented Architecture (SOA) is a software architecture model, which consists of services, messages, operations and processes.
Message means a communicating entity or object.
Operation means action or set of actions. For example, actions during a bank transaction.
Transaction (trans + action) refers to two inter-related sets of operations or actions or instructions.
Query is a command for getting select values from a database which in return transfer the answer to the query after its processing.
Query Processing is a group of structured activities undertaken to get the results from a data store as per the query
Key Value Pair (KVP) refers to a set of two linked entities, one is the key, which is a unique identifier for a linked entity and the other is the value, which is either the entity that is identified or a pointer to the location of that entity
Hash Table (also called hash map) refers to a data structure which maps the KVPs and is used to implement an associative array (for example array of KVPs). A hash table may use an index (key) which is computed using a hash function and key maps to the value. Index is used to get or point to the desired value.
Bigtable maps two arbitrary string values into an associated arbitrary byte array. One is used as row key and the other as column key. Time stamp associates in three-dimensional mapping. Mapping is unlike a relational database but can be considered as a sparse, distributed multi-dimensional sorted map. The table can scale up to 100s to 1000s of distributed computing nodes with ease of adding more nodes
Business Transaction (BT) in database theory, refers to a (business) process that requests information from or that changes the data in a database.One operation in a BT is a command 'connect' that connects a DBMS and database, which in turn also connects with the DBMS. Similarly, BTs are processes using commands 'insert', 'delete', 'append', and 'modify'.
Process means a composition of a group of structured activities or tasks that lead to a particular goal
Process Matrix is a multi-element entity, each element of which relates a set of data or inputs to an activity (or subset of activities).
Business Process (BP) is an activity or series of activities or a collection of inter-related structured activities, tasks or processes.
Process refers to series of activities such as acquiring data from the ACVMs for the counts of each flavour sold during programmed intervals, analysing the acquired data, initiating messages for Fill service for each flavour through Fill process.
Business process refers to series of related processes, such as Fill, Collect and Display at appropriate intervals is a business process in ACVMs chain.
Business intelligence is a process which enables the ACVM business service to extract new facts and knowledge.

5.2 DATA ACQUIRING AND STORAGE

Objective

Applying the data-acquiring and data-storage functions for IoT/M2M devices data and messages

5.2.1 Data Generation

Data generates at devices that later on, transfers to the Internet through a gateway. Data

generates as follows:

● **Passive devices data**: Data generate at the device or system, following the result of interactions. A passive device does not have its own power source. An external source helps such a device to generate and send data. Examples are an RFID, or an ATM debit card. The device may or may not have an associated microcontroller, memory and transceiver. A contactless card is an example of the former and a label or barcode is the example of the latter.

● **Active devices data**: Data generates at the device or system or following the result of interactions. An active device has its own power source. Examples are active RFID, streetlight sensor or wireless sensor node. An active device also has an associated microcontroller, memory and transceiver.

● **Event data:** A device can generate data on an event only once. For example, on detection of the traffic or on dark ambient conditions, which signals the event. The event on darkness communicates a need for lighting up a group of streetlights. A system consisting of security cameras can generate data on an event of security breach or on detection of an intrusion. A waste container with associate circuit can generate data in the event of getting it filled up 90% or above. The components and devices in an automobile generate data of their performance and functioning. For example, on wearing out of a brake lining, a play in steering wheel and reduced air-conditioning is felt. The data communicates to the Internet. The communication takes place as and when the automobile reaches near a Wi-Fi access point.

● **Device real-time data**: An ATM generates data and communicates it to the server instantaneously through the Internet. This initiates and enables Online Transactions Processing (OLTP) in real time.

● **Event-driven device data**: A device data can generate on an event only once. Examples are:

(i) a device receives command from Controller or Monitor, and then performs action(s) using an actuator. When the action completes, then the device sends an acknowledgement;

(ii) When an application seeks the status of a device, then the device communicates the status.

5.2.2 Data Acquisition

Objective: Applications configure the devices for acquiring data

Data acquisition means acquiring data from IoT or M2M devices. The data communicates after the interactions with a data acquisition system (application). The application interacts and communicates with a number of devices for acquiring the needed data. The devices send data on demand or at programmed intervals. Data of devices communicate using the network, transport and security layers.

An application can configure the devices for the data when devices have configuration capability. For example, the system can configure devices to send data at defined periodic intervals. Each device configuration controls the frequency of data generation.

For example, system can configure an umbrella device to acquire weather data from the Internet weather service, once each working day in a week .

An ACVM can be configured to communicate the sales data of machine and other information, every hour. The ACVM system can be configured to communicate instantaneously in event of fault or in case requirement of a specific chocolate flavour needs the Fill service.

Application can configure sending of data after filtering or enriching at the gateway at the data-adaptation layer. The gateway in-between application and the devices can provision for one or more of the following functions—**transcoding, data management and device management**. Data management may be provisioning of the **privacy and security, and data integration, compaction and fusion.**

Device-management software provisions for device ID or address, activation, configuring (managing device parameters and settings), registering, deregistering, attaching, and detaching. The process of acquiring data from the embedded component devices in the automobiles for Automotive Components and Predictive Automotive Maintenance System (ACPAMS) application.

5.2.3 Data Validation

Objective: Data must be validated before storing. Data aggregation, adaptation and enrichment is done before communicating to the Internet

Data acquired from the devices does not mean that data are correct, meaningful or consistent. Data consistency means within expected range data or as per pattern or data not corrupted during transmission. Therefore, data needs validation checks. Data validation software do the validation checks on the acquired data.

Validation software applies logic, rules and semantic annotations. The applications or services depend on valid data. Then only the analytics, predictions, prescriptions, diagnosis and decisions can be acceptable.

Large magnitude of data is acquired from a large number of devices, especially, from machines in industrial plants or embedded components data from large number of automobiles or health devices in ICUs or wireless sensor networks, and so on. Validation software, therefore, consumes significant resources.

An appropriate strategy needs to be adopted. For example, the adopted strategy may be filtering out the invalid data at the gateway or at device itself or controlling the frequency of acquiring or cyclically scheduling the set of devices in industrial systems. Data enriches, aggregates, fuses or compacts at the adaptation layer.

5.2.4 Data Categorisation for Storage

Services, business processes and business intelligence use data. Valid, useful and relevant data can be categorised into three categories for storage—data alone, data as well as results of processing, only the results of data analytics are stored. Following are three cases for storage:

1. Data which needs to be repeatedly processed, referenced or audited in future, and therefore, data alone needs to be stored.

2. Data which needs processing only once, and the results are used at a later time using the analytics, and both the data and results of processing and analytics are stored. Advantages of this case are quick visualisation and reports generation without reprocessing. Also the data is available for reference or auditing in future.

3. Online, real-time or streaming data need to be processed and the results of this processing and analysis need storage. Data from large number of devices and sources categorises into a fourth category called Big data. Data is stored in databases at a server or in a data warehouse or on a Cloud as Big data.

5.2.5 Assembly Software for the Events

A device can generate events. For example, a sensor can generate an event when temperature reaches a preset value or falls below a threshold.

A pressure sensor in a boiler generates an event when pressure exceeds a critical value which warrants attention. Each event can be assigned an ID. A logic value sets or resets for an event state. Logic 1 refers to an event generated but not yet acted upon. Logic 0 refers to an event generated and acted upon or not yet generated. A software component in applications can assemble the events (logic value, event ID and device ID) and can also add Date time stamp. Events from IoTs and logic-flows assemble using software.

5.2.6 Data Store

A data store is a data repository of a set of objects which integrate into the store. Features of data store are:

● Objects in a data-store are modelled using Classes which are defined by the database schemas.

● A data store is a general concept. It includes data repositories such as database, relational database, flat file, spreadsheet, mail server, web server, directory services and VMware

● A data store may be distributed over multiple nodes. Apache Cassandra is an example of distributed data store.

● A data store may consist of multiple schemas or may consist of data in only one scheme. Example of only one scheme data store is a relational database.

Repository in English means a group, which can be related upon to look for required things, for special information or knowledge. For example, a repository of paintings of artists. A database is a repository of data which can be relied upon for reporting, analytics, process, knowledge discovery and intelligence. A flat file is another repository.

Flat file means a file in which the records have no structural interrelationship. Eg. spreadsheet concept. VMware uses data store to refer to a file that stores a virtual machine.

5.2.7 Data Centre Management

Objective: Data centre is meant for data storage, data security and protection

A data centre is a facility which has multiple banks of computers, servers, large memory systems, high speed network and Internet connectivity. The centre provides data security and protection using advanced tools, full data backups along with data recovery, redundant data communication connections and full system power as well as electricity supply backups.

Data centre is meant for data storage, data security and protection. Large industrial units, banks, railways, airlines and units for whom data are the critical components use the services of data centres. Data centres also possess a dust free, heating, ventilation and air conditioning (HVAC),

cooling, humidification and dehumidification equipment, pressurisation system with a physically highly secure environment.

The manager of data centre is responsible for all technical and IT issues, operations of computers and servers, data entries, data security, data quality control, network quality control and the management of the services and applications used for data processing.

5.2.8 Server Management

Server management means managing services, setup and maintenance of systems of all types associated with the server. A server needs to serve around the clock. Server management includes managing the following:

- Short reaction times when the system or network is down
- High security standards by routinely performing system maintenance and updation
- Periodic system updates for state-of-the art setups
- Optimised performance
- Monitoring of all critical services, with SMS and email notifications
- Security of systems and protection
- Maintaining confidentiality and privacy of data
- High degree of security and integrity and effective protection of data, files and databases at the organisation
- Protection of customer data or enterprise internal documents by attackers which includes spam mails, unauthorised use of the access to the server, viruses, malwares and worms
- Strict documentation and audit of all activities.

5.2.9 Spatial Storage

Consider goods with RFID tags. When goods move from one place to another, the IDs of goods as well as locations are needed in tracking or inventory control applications. Spatial storage is storage as spatial database which is optimised to store and later on receives queries from the applications.

Suppose a digital map is required for parking slots in a city. Spatial data refers to data which represents objects defined in a geometric space. Points, lines and polygons are common geometric objects which can be represented in spatial databases. Spatial database can also represent database for 3D objects, topological coverage, linear networks, triangular irregular networks and other complex structures. Additional functionality in spatial databases enables efficient processing.

Internet communication by RFIDs, ATMs, vehicles, ambulances, traffic lights, streetlights, waste containers are examples of where spatial database are used.

Spatial database functions optimally for spatial queries. A spatial database can perform typical SQL queries, such as select statements and performs a wide variety of spatial operations. Spatial database has the following features:

- Can perform geometry constructors. For example, creating new geometries

● Can define a shape using the vertices (points or nodes)

● Can perform observer functions using queries which replies specific spatial information such as location of the centre of a geometric object

● Can perform spatial measurements which mean computing distance between geometries, lengths of lines, areas of polygons and other parameters

● Can change the existing features to new ones using spatial functions and can predicate spatial relationships between geometries using true or false type queries.

5.3 Organizing the data

5.3.1 Databases Required data values are organised as database(s) so that select values can be retrieved later. Database One popular method of organising data is a database, which is a collection of data. This collection is organised into tables. A table provides a systematic way for access, management and update. A single table file is called flat file database. Each record is listed in separate row, unrelated to each other.

Relational Database

A relational database is a collection of data into multiple tables which relate to each other through special fields, called keys (primary key, foreign key and unique key). Relational databases provide flexibility. Examples of relational database are MySQL, PostGreSQL, Oracle database created using PL/SQL and Microsoft SQL server using T-SQL. Object Oriented Database (OODB) is a collection of objects, which save the objects in objected oriented design. Examples are ConceptBase or Cache.

Eg. Automatic Chocolate Vending Machine database

Table 5.1 RDBAVCM Table A—ACVMs information

Machine ID	Region	Address	Installation Date	Maintenance Schedule	Fill Service Address	Pending Request Number 1	Pending Request Number 2	Pending Request Number 3	Pending Request Number 4

Table 5.2 RDBAVCM Table B—ACVMs fill request information

Service Request Number	Machine ID	Request Receipt DateTime	Number FL1 Request	Number FL 2 Request	Number FL 3 Request	Number FL 4 Request	Number FL5 Request

Table 5.3 RDBAVCM Table C—ACVMs fill service actions

Service Request Number	Service DateTime	Number FL1 Sent	Number FL 2 Sent	Number FL 3 Sent	Number FL 4 Sent	Number FL5 Sent

Database Management System (DBMS) is a software system, which contains a set of programs specially designed for creation and management of data stored in a database. Database transactions can be performed on a database or relational database.

Atomicity, Data Consistency, Data Isolation and Durability (ACID) Rules

The database transactions must maintain the atomicity, data consistency, data isolation and durability during transactions. Let us explain these rules using Example 5.3 as follows:

Atomicity means a transaction must complete in full, treating it as indivisible. When a service request completes, then the pending request field should also be made zero.

Consistency means that data after the transactions should remain consistent. For example, sum of chocolates sent should equal the sums of sold and unsold chocolates for each flavour after the transactions on the database.

Isolation means transactions between tables 5.1 and 5.2, 5.2 and 5.3 and 5.3 and 5.1 are isolated from each other.

Durability means after completion of transactions, the previous transaction cannot be recalled. Only a new transaction can affect any change.

Distributed Database

Distributed Database (DDB) is a collection of logically interrelated databases over a computer network. Distributed DBMS means a software system that manages a distributed database. The features of a distributed database system are:

● DDB is a collection of databases which are logically related to each other.

● Cooperation exists between the databases in a transparent manner. Transparent means that each user within the system may access all of the data within all of the databases as if they were a single database.

● DDB should be 'location independent', which means the user is unaware of where the data is located, and it is possible to move the data from one physical location to another without affecting the user

Consistency, Availability and Partition-Tolerance Theorem

Consistency, Availability and Partition-Tolerance Theorem (CAP theorem) is a theorem for distributed computing systems. The theorem states that it is impossible for a distributed computer system to simultaneously provide all three of the Consistency, Availability, Partition tolerance (CAP) guarantees.5 This is due to the fact that a network failure can occur during communication among the distributed computing nodes. Partitioning of a network therefore needs to be tolerated. Hence, at all times either there will be consistency or availability.

Consistency means 'Every read receives the most recent write or an error'. When a message or data is sought the network generally issues notification of time-out or read error. During an interval of a network failure, the notification may not reach the requesting node(s).

Availability means 'Every request receives a response, without guarantee that it contains the most recent version of the information'. Due to the interval of network failure, it may happen that most recent version of message or data requested may not be available.

Partition tolerance means 'The system continues to operate despite an arbitrary number of messages being dropped by the network between the nodes'. During the interval of a network failure, the network will have two separate set of networked nodes. Since failure can always occur therefore, the partitioning needs to be tolerated.

5.3.2 Query Processing

Query means an application seeking a specific data set from a database. For example, a query at a relational database at bank server may be for the ATM transactions made in a month by a specific customer ID (Example 2.3). Other examples are: most-liked chocolate flavour in the city by children of age group 6 to 10 (Example 5.1); number of times a vehicle visited at the ACPAMS center (Example 5.2) and service was rendered with satisfaction level of 5 out of 5.

Query Processing

Query processing means using a process and getting the results of the query made from a database. The process should use a correct as well as efficient execution strategy. Five steps in processing are:

1. Parsing and translation: This step translates the query into an internal form, into a relational algebraic expression and then a Parser, which checks the syntax and verifies the relations.

2. Decomposition to complete the query process into micro-operations using the analysis (for the number of micro-operations required for the operations), conjunctive and disjunctive normalisation and semantic analysis.

3. Optimisation which means optimising the cost of processing. The cost means number of micro-operations generated in processing which is evaluated by calculating the costs of the sets of equivalent expressions.

4. Evaluation plan: A query-execution engine (software) takes a query-evaluation plan and executes that plan.

5. Returning the results of the query.

Distributed Query Processing

Distributed Query Processing means query processing operations in distributed databases on the same system or networked systems. The distributed database system has the ability to access remote sites and transmit the queries to other systems.

5.3.3 SQL

SQL stands for Structured Query Language. It is a language for viewing or changing (update, insert or append or delete) databases. It is a language for data querying, updating, inserting, appending and deleting the databases. It is a language for data access control, schema creation and modifications. It is also a language for managing the RDBMS.

SQL was originally based upon the tuple relational calculus and relational algebra. SQL can embed within other languages using SQL modules, libraries and pre-compilers. SQL features are as follows:

● Create Schema is a structure that contains descriptions of objects created by a user (base tables, views, constraints). The user can describe and define the data for a database.

● Create Catalog consists of a set of schemas that constitute the description of the database.

● Use Data Definition Language (DDL) for the commands that depict a database, including creating, altering and dropping tables and establishing constraints. The user can create and drop databases and tables, establish foreign keys, create view, stored procedure, functions in a database.

● Use Data Manipulation Language (DML) for commands that maintain and query a database. The user can manipulate (INSERT, UPDATE or SELECT the data and access data in relational database management systems.

● Use Data Control Language (DCL) for commands that control a database, including administering privileges and committing data. The user can set (grant or add or revoke) permissions on tables, procedures, and views.

5.3.4 NOSQL

NOSQL stands for No-SQL or Not Only SQL that does not integrate with applications that are based on SQL. NOSQL is used in cloud data store. NOSQL may consist of the following:

● A class of non-relational data storage systems, flexible data models and multiple schemas

● Class consisting of uninterpreted key and value or 'the big hash table'. For example in

● Class consisting of unordered keys and using the JSON. For example in PNUTS

● Class consisting of ordered keys and semi-structured data storage systems. For

examples in the BigTable, Hbase and Cassandra (used in Facebook and Apache)

● Class consisting of JSON (Section 2.3). For example in MongoDb6 which is widely used for NOSQL)

● Class consisting of name and value in the text. For example in CouchDB

● May not require a fixed table schema

● Consistency means all copies have same value like in traditional DBs.

● Availability means at least one copy available in case a partition becomes inactive or fails. For example, in web applications, the other copy in other partition is available.

● Partition means parts which are active but may not cooperate as in distributed databases.

5.3.7 Real-Time and Intelligence

Decision on real-time data is fast when query processing in live data (streaming) has low latency. Decision on historical data is fast when interactive query processing has low latency. Low latencies are obtained by various approaches: Massively Parallel Processing (MPP), in-memory databases and columnar databases.

5.4 TRANSACTIONS, BUSINESS PROCESSES, INTEGRATION AND ENTERPRISE SYSTEMS A transaction is a collection of operations that form a single logical unit. For example, a database connect, insertion, append, deletion or modification transactions. Business transactions are transactions related in some way to a business activity.

5.4.1 Online Transactions and Processing

OLTP means process as soon as data or events generate in real time. OLTP is used when requirements are availability, speed, concurrency and recoverability in databases for real-time data or events. Example 5.4 gives the uses of OLTP in the application and network domain in Internet of ATMs (ATM of a bank) connected to a bank server.

Batch Transactions Processing

Batch transactions processing means the execution of a series of transactions without user interactions. Transaction jobs are set up so they can be run to completion. Scripts, command-line arguments, control files, or job control language predefine all input parameters. Batch processing means a transaction process in batches and in an non-interactive way. When one set of transactions finish, the results are stored and a next batch is taken up. A good example is credit card transactions where the final results at the end of the month are used. Another example is chocolate purchase transactions. The final results of sell figures from ACVMs can communicate on the Internet at the end of an hour or day.

Streaming Transactions Processing

Examples of the streams are log streams, event streams and twitter streams. Query and transactions processing on streaming data need specialised frameworks. Storm from Twitter, S4 from Yahoo, SPARK streaming, HStreaming and flume are examples of frameworks for real-time streaming computation frameworks.

Interactive Transactions Processing

Interactive transactions processing means the transactions which involve continual exchange of information between the computer and a user. For example, user interactions during e-shopping and e-banking. The processing is just the opposite of batch processing.

Real-time Transactions Processing

Real-time transaction processing means that transactions process at the same time as the data arrives from the data sources and data store. An example is ATM machine transactions. In-memory, row-format records enable real-time transaction processing. Row format means few rows and more columns. The CPU accesses all columns in single accesses in SIMD (single instruction multiple data) streams processing. Event

Stream Processing and Complex Event Processing

Event Stream Processing (ESP) is a set of technologies, event processing languages, Complex Event Processing (CEP), event visualisation, event databases and event-driven middleware.

Complex Event Processing

CEP has many applications. For example, IoT event processing applications, stocks algorithmic-based trading and location-based services. A CEP application in Eclipse are used for capturing a combination of data, timing conditions and efficiently recognise the corresponding events over data streams.

5.4.2 Business Processes

A business process consists of a series of activities which serves a particular specific result. It is used when an enterprise has a number of interrelated processes which serve 176 a particular result or goal. The results enable sales, planning and production

5.4.3 Business Intelligence

Business intelligence is a process which enables a business service to extract new facts and knowledge and then undertake better decisions. The new facts and knowledge follow from the earlier results of data processing, aggregation and then analysing those results.

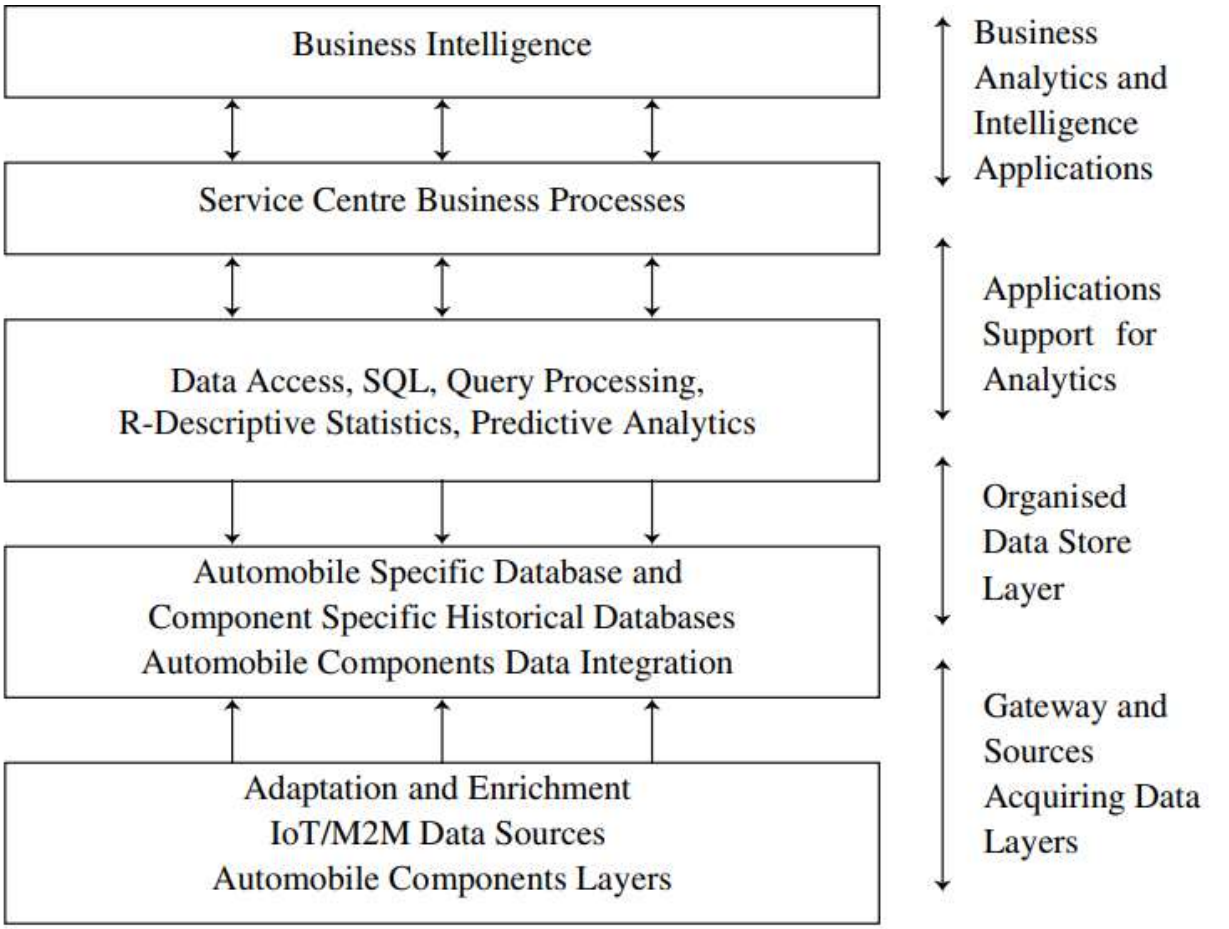


Figure 5.1 Architecture reference model for the business intelligence and business processes at ACPAMS

5.4.4 Distributed Business

Process Several times, business processes need to be distributed. Distribution of processes reduces the complexity, communication costs, enables faster responses and smaller processing load at the central system.

Distributed Business Process System (DBPS) is a collection of logically interrelated business processes in an Enterprise network. DBPS means a software system that manages the distributed BPs

DBPS is a collection of logically related BPs like DDBS. DBPS exists as cooperation between the BPs in a transparent manner. Transparent means that each user within the system may access all of the process decisions within all of the processes as if they were a single business process

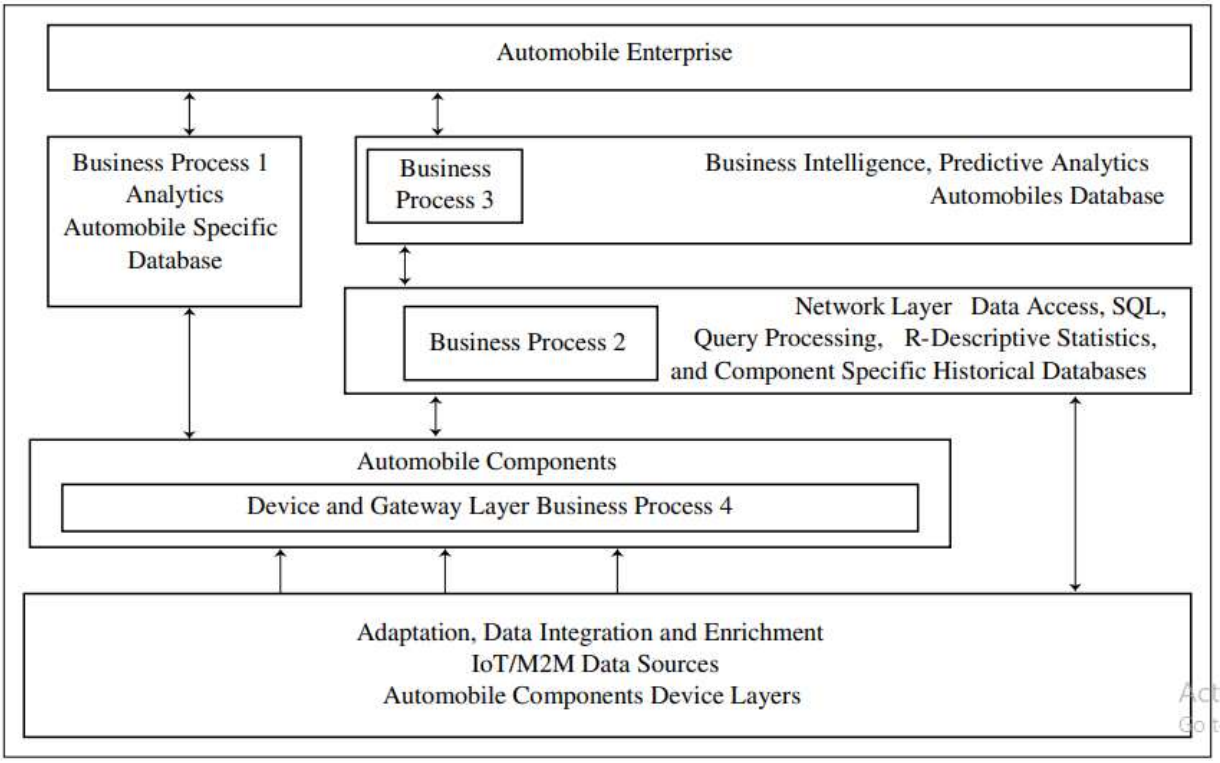


Figure 5.2 Distributed interrelated business intelligence and processes at the enterprise, network, and devices and gateway layers

5.4.5 Complex Applications Integration and Service Oriented Architecture

An enterprise has number of applications, services and processes. Heterogeneous systems have complexity when integrating them in the enterprise. Following are the standardised business processes, as defined in the Oracle application integration architecture:

- Integrating and enhancing the existing systems and processes
- Business intelligence
- Data security and integrity
- New business services and products (web services)
- Collaboration and knowledge management
- Enterprise architecture and SOA
- e-commerce ● External customer services
- Supply chain automation and analytics results visualisation
- Data centre optimisation

5.4.6 Integration and Enterprise Systems Figure 5.3 shows complex applications integration architecture and SOA of cloud-based IoT services, web services, cloud services and services

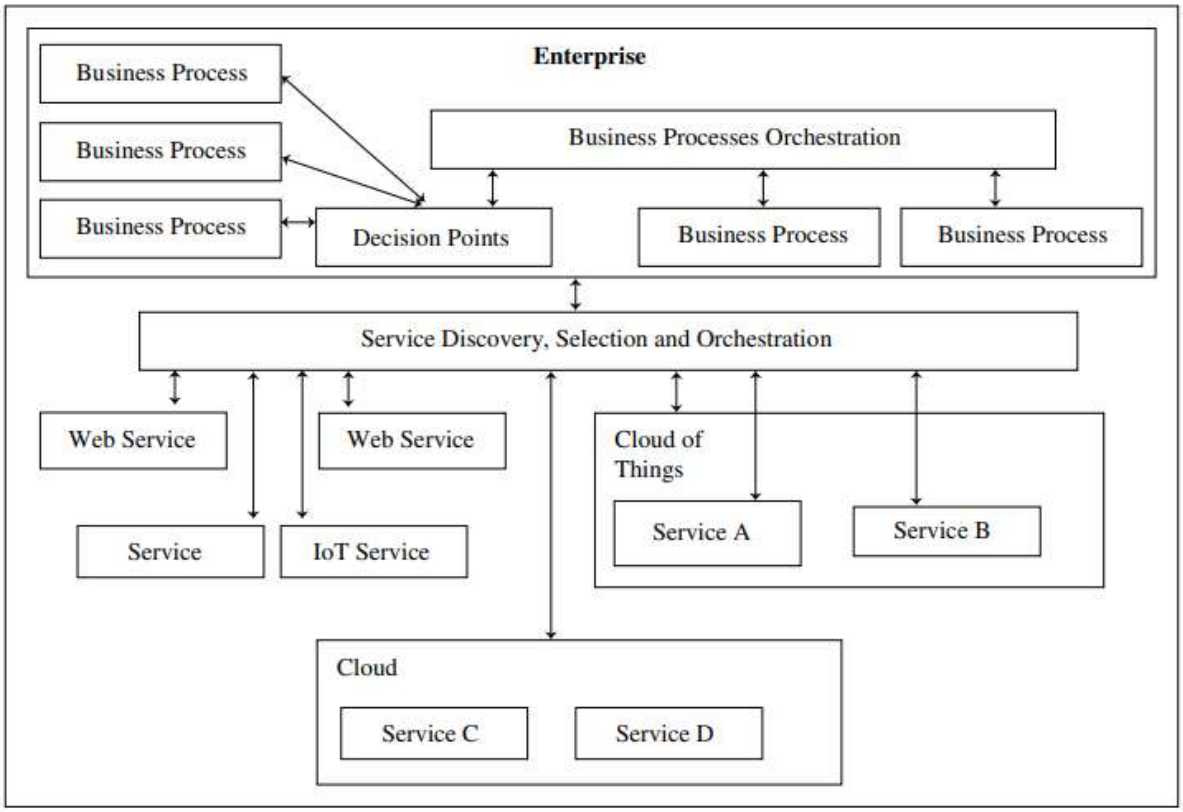


Figure 5.3 Complex applications integration architecture and SOA of cloud-based IoT services, web services, cloud services and services

5.5 ANALYTICS

Organised data after acquiring from the devices can be used for multiple purposes. Applications usually use the data of devices in two ways—for monitoring, reporting and rule-based actions

5.5.1 Analytics Phases Analytics has three phases before deriving new facts and providing business intelligence. These are:

1. Descriptive analytics enables deriving the additional value from visualisations and reports.

2. Predictive analytics is advanced analytics which enables extraction of new facts and knowledge, and then predicts or forecasts.

3. Prescriptive analytics enables derivation of the additional value and undertake better decisions for new option(s) to maximise the profits. Descriptive Analytics Descriptive analytics answers the questions about what happened in the past. Descriptive analytics means finding the aggregates, frequencies of occurrences, mean values (simple Data Acquiring, Organising, Processing and Analytics 183 or geometric averages) or variances in values or groupings using selected properties and hence applying these.

Descriptive analytics enable the following:

- Actions, such as Online Analytical Processing (OLAP) for the analytics
- Reporting or generating spreadsheets
- Visualisations or dashboard displays of the analysed results

● Creation of indicators, called key performance indicators. Descriptive Analytics Methods

● Spreadsheet-based reports and data visualisations: Results of descriptive analysis can be presented in a spreadsheet format before creating the data visuals for the user. Spreadsheet enables user visualisation of what if. For example, if sales of chocolates of specific flavour drop by 5% on specific set of ACVMs, how it will influence the profitability?

● Descriptive statistics-based reports and data visualisations:

Descriptive analysis can also use descriptive statistics. Statistical analysis means finding peak, minima, variance, probabilities, and statistical parameters. Formulae are used for the data sets to enable the data showing variations understandable.

● Data mining and machine learning methods in analytics: Data mining analysis means use of algorithms which extract hidden or unknown information or patterns from large amounts of data. Machine learning means modelling of the specific tasks.

It enables view of rollup (finer granulites data to coarse granulites data) or drill down (coarser granulites data to finer granulites data). OLAP enables obtaining summarized information and automated reports from large volume database. Results of queries are based on Metadata. Metadata is data which describes the data. Pre-storing calculated values provide consistently fast response.

OLAP uses the analysis functions which are not possible to code in SQL. The data structure is designed from the users perspective, using Spreadsheet like formulae. OLAP is a significant improvement over query systems. OLAP is an interactive system to show different summaries of multidimensional data by interactively selecting the attributes in a multidimensional data cube.7 OLAP enables analysing data in multiple dimensions in a structure called data cube. Each dimension represents a hierarchy. Each dimension has a dimension attribute which defines the dimension and summary of measure attribute. A slice of a data-cube can be viewed when values of multiple dimensions are fixed. A dice of a data-cube can be viewed with variable values in multiple dimensions. Slicing and dicing functionalities mean selecting specific values for these attributes, which are then displayed on top of the cross-tables. A slice means a data relationship in the analysed multiple dimensional data. A slice of a data relationship between two attributes can be individually visualised. For examples, monthly sales versus flavours sold at the chain of ACVMs in Example 5.1 after the analysis. A cubical dice has six faces, each face marked distinctly. Face 1 has one dot, face 2 two, and so on. Sixth face has six dots. Similarly, six different cross referenced tables can be created during OLAP for three-dimensional structure for analysing data.

Advanced Analytics:

Predictive Analytics Predictive analytics answer the question ”What will happen?” Predictive analytics is advanced analytics. The user interprets the outputs from advanced analytics using descriptive analytics methods, such as data visualisation.

Predicting trends

● Undertaking preventive maintenance from earlier models of equipment and device failure rates

Managing the campaign with integrated marketing strategy from previous studies of effect of campaigns with respect to media types, regions, targeted age group

● Predicting by identifying patterns, clusters with similar behaviour

● Predicting based on anomalous characteristics, anomaly detection The results of predictions need verifications from a domain knowledge, and view from multiple angles. Prescriptive Analytics Prescriptive analytics answers not only what is anticipated or what will happen or when it will happen, but also why it will happen based on the input from descriptive analytics and business rules.

5.5.2 Event Analytics

Events definable options are unique, non-interaction or interaction options for the events. Event analytics use event data for events tracking and event reporting. An event has the following components:

● Category–an event of chocolate purchase in ACVM example belongs to one category and event of reaching predefined threshold of sell for specific chocolate flavour which belongs to other category

● Action–sending message from ACVM on completing predefined sell is the action taken on the event

● Label (optional)

● Value (optional)–on event, messaging the number of chocolate of that flavour sold or remaining

5.5.3 In-memory Data Processing and Analytics

In-memory option of row or column formats can be selected in certain databases, for example, Oracle dual format architecture database that enables to run the real-time, adhoc, analytic queries on IoTs' data. In-memory and On-store Row Format Option (Few Rows and Many Columns) Consider the transactions of the type, ATM transactions or sales order transactions. Each row has separate record. For example, separate record for each ACVM or each bank customer or each sales order.

In-memory and On-store Column Format Option (Few Columns and More Rows)

Consider analytics of the type, monthly sales of chocolates on the ACVMs, enterprise yearly profits. Analytical workloads access few columns but scan the entire data set. Analytics therefore run faster on column format, more rows and few columns. Fast for processing needs few columns and many rows. They typically require aggregation or fusion or compaction also.

5.5.4 Real-time Analytics Management

Real-time analytics management means ensuring faster OLTP as well as OLAP. Real-time analytics works both as direct querying using an OLTP database and in a data warehouse and OLAP on queried results. Queries return fast, databases such as Oracle database provides in-memory row format option large speedups for OLTP applications and inmemory column format option for large speedups for OLAP applications.

5.5.5 Analytics using Big Data in IoT/M2M

Big data means extreme amount of data. Big data also means data of high volume, variety and velocity (3Vs) or one which also includes veracity (4Vs).

Volume means data received from number of sources of data, including data sets with sizes beyond the ability of commonly used software tools to acquire, manage and process data within a tolerable elapsed time.

Variety means structured as well as unstructured data in different formats, variety of data on which no SQL (Structured Query Language) applicable.

Velocity means data received with higher rates due to use of number of sources of data.

Veracity means variation in data quality for analytics. The analytics need the trustable data, filtered data after removing—anomalous data, non-standard and not cross referencing data.

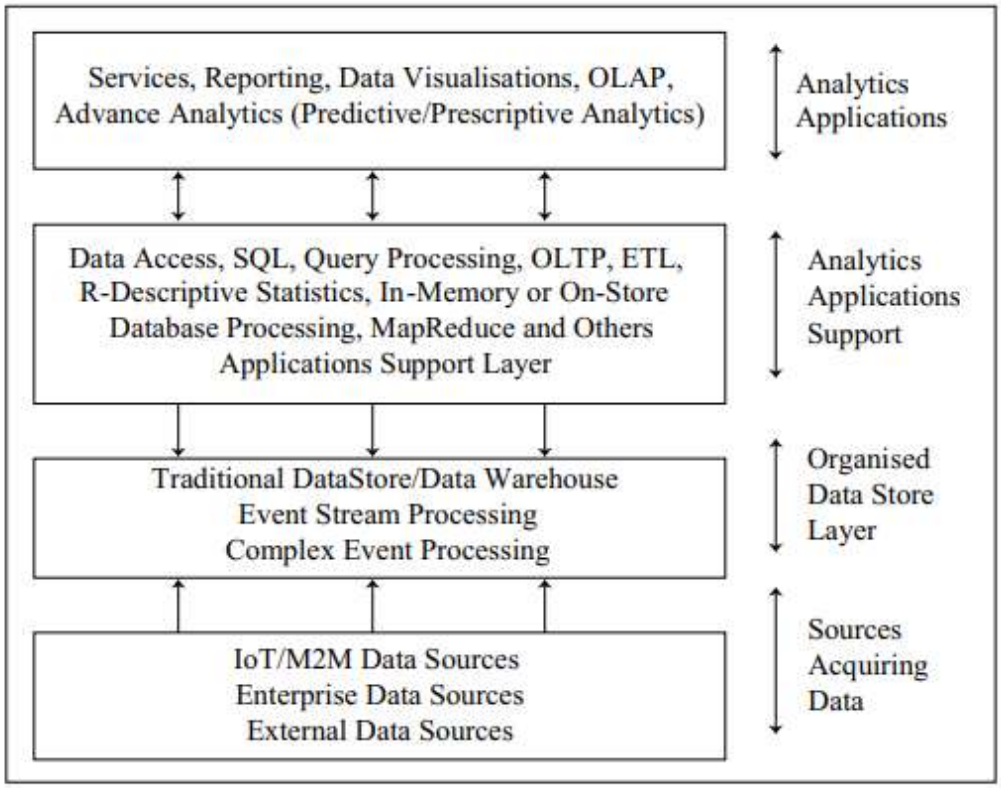


Figure 5.4 Analytics Architecture Reference Model

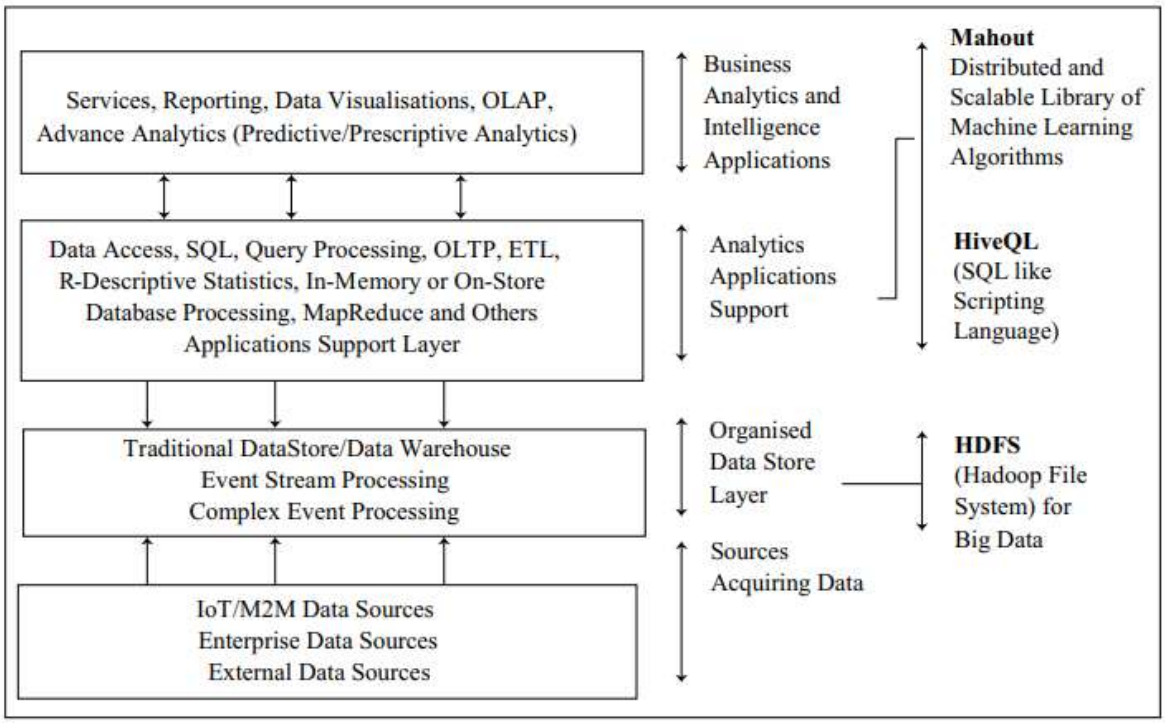


Figure 5.5 Berkeley data analytics stack architecture

5.6 KNOWLEDGE ACQUIRING, MANAGING AND STORING PROCESSES

Data Information Knowledge Wisdom (DIKW)8 forms a pyramid. Information is an enriched set of data values when considered in a context and that can be queried upon. Visualisation of data gives information. Spreadsheet gives information. Analytics gives information.

Information in the given context is an answer to the query or set of queries. The answer comes from processing the data and querying. For example, a balance sheet data is the enriched set of data values using the analytics. “Is the data in enterprise balance sheet showing consistent growth in preceding five years?” is querying of the balance sheet. The answer is the information obtained from the balance sheet. Knowledge, according to an English dictionary is sharable information and understanding about a subject or context. Information that All ACVMs filled at most times gives understanding and knowledge that “Fill service is prompt in attending to the service requests”

Knowledge Management Knowledge management9 (KM) is managing knowledge when the new knowledge is regularly acquired, processed and stored.

Wisdom Sensible and reasonable decisions are made using advanced tools which enable wise decision. Wisdom, according to an English dictionary is “Ability to use the experience and knowledge in order to make sensible and reasonable judgment and decisions”.

5.6.1 Knowledge-Management Reference Architecture

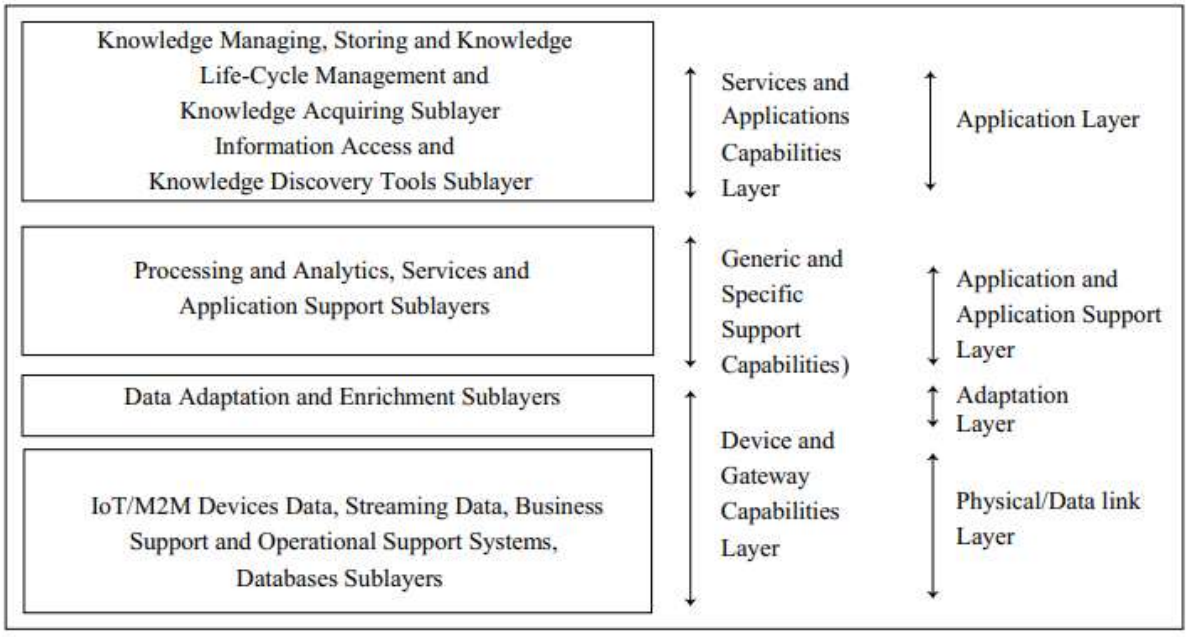


Figure 5.6 (a) A reference architecture for the knowledge management (left-hand side) and (b) Correspondence in terms of ITU-T reference model and OSI layers for IoT/M2M (middle and right-hand side)