Unit 3

Fundamental of Data Acquisition, Data Organizing and Data Processing in IoT

3.1 Data Acquisition and storage

- Apply the data-acquiring and data-storage functions for IoT/M2M devices data and message.

Data Generation:

1. 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. Example are RFID or an ATM debit card.
2. 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.

Data Generation

3. 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.

4. 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.

5. Event-driven device data : A device data can generates on an event only once. Examples are 1) a device receives command from controller or monitor, and then performs actions using an actuator. 2) when an application seeks the status of a device , then the device communicates the status.

Data Acquisition

- 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.
- 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.

Data Validation

- 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 annotation.
- The applications or services depend on valid data. Then only the analytics, predictions, prescriptions, diagnosis and decisions can be acceptable.

Data Categorization for storage

- Services, business processes and business intelligence use data.
- Valid , useful and relevant data can be categorized into three categories for storage

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.
3. Online, real-time or streaming data need to be processing and the results of this processing and analysis need storage.

Assembly Software for the Events

- A device can generate events. For example, a sensor can generate an event when temperature reaches a present 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 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 and can also add date time stamp.
- Events from IoTs and logic-flows assemble using software.

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 modeled 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.

Data Center Management

- A data center is a facility which has multiple banks of computers, servers, large memory systems, high speed network and internet connectivity.
- The center 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.
- Large industrial units, banks, railways, airlines and units for whom data are the critical components use the services of data centers.
- Data centers also possess a dust free, heating, ventilation and air conditioning(HVAC), humidification and dehumidification equipment, pressurization system with a physically high secure environment.
- The manager of data center 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.

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:

1. Short rection times when the system or network is down.
2. High security standards by routinely performing system maintenance and updation.
3. Periodic system updates for state-of-the art setups.
4. Optimized performance.
5. Monitoring of all critical services, with SMS and email notifications.
6. Security of systems and protection
7. Maintaining confidentiality and privacy of data.
8. High degree of security and integrity and effective protection of data, files and databases at the organization.

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 optimized 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 object is 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 structure.
- Additional functionality in spatial databases enables efficient processing.

Spatial Storage

- 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.
- Can define a shape using the vertices.
- 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, area of polygons and other parameters.

3.2 Organizing the data

- Data can be organized in a number of ways.
- For example, objects, files, data store, database, relational database and object oriented database.

Databases

- **1. Database**
- One popular method of organizing data is a database, which is a collection of data.
- This collection is organized 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.

Databases

- **2. 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 using T-SQL.
- Object oriented database(OODB) is a collection of objects, which save the objects in objected oriented design. Examples are concept Base or Cache.

Databases

- **3. 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.

Database

- **4. 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.
- Features:

➢ 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.

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.
- 5 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, conjunctive and disjunctive normalization and semantic analysis.
3. **Optimization** which means optimizing the cost of processing.
4. **Evaluation plan**: A query-execution engine(software) takes a query-evaluation plan and executes that plan.
5. **Returning the results** of the query.

SQL

- SQL stands for Structured Query Language.
- It is a language for viewing or changing 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 language for managing the RDBMS.

3.3 Transactions, Business processes, integration and enterprise systems

Online Transactions and Processing

- OLTP means process as soon as data or events generate in real time.
- OLTP is used when requirements are available, speed, concurrency and recoverability in database for real-time data or events.
- **1) 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.

Online Transactions and Processing

- **2)Real-time Transaction 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 ATM machine transaction.
- In-memory , row-format records enable real-time transaction processing.
- Row-format means few rows and more columns.
- The CPU accesses all column in single accesses in SIMD(single instruction multiple data) streams processing.

Online Transactions and Processing

- **3) Event stream processing and complex event processing**
- Event stream processing(ESP) is a set of technologies, event processing languages, complex event processing(CEP), event visualization, event databases and event-driven middleware.
- ESP and CEP does the following:

➢ Processes takes on receiving streams of event data.

➢ Identifies the meaningful pattern from the streams.

➢ Detects relationships between multiple events.

➢ Correlates the events data

➢ Detects event hierarchies.

➢ Detects aspects such as timing, causality, subscription membership.

➢ Builds and manages the event-driven information systems.

Business Processes

- A business process consists of activities which serves a particular specific result.
- It is used when an enterprise has a number of interrelated processes which serve a particular result or goal.
- The results enable sales, planning and production.
- The Business Processes is a representation or process matrix or flowchart of a sequence of activities with interleaving decision points.
- Internet of RFIDs enables a business process called tracking of RFID labelled goods which also enables inventory control process.
- IoT/M2M enables the device's data in databases for business processes.
- Example in streetlights control and management, each group of streetlight sends data in real time through the gateways. The gateways connect to the internet. The control and management processes streetlights real time databases and group databases.

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 analyzing those results.

Architecture reference model

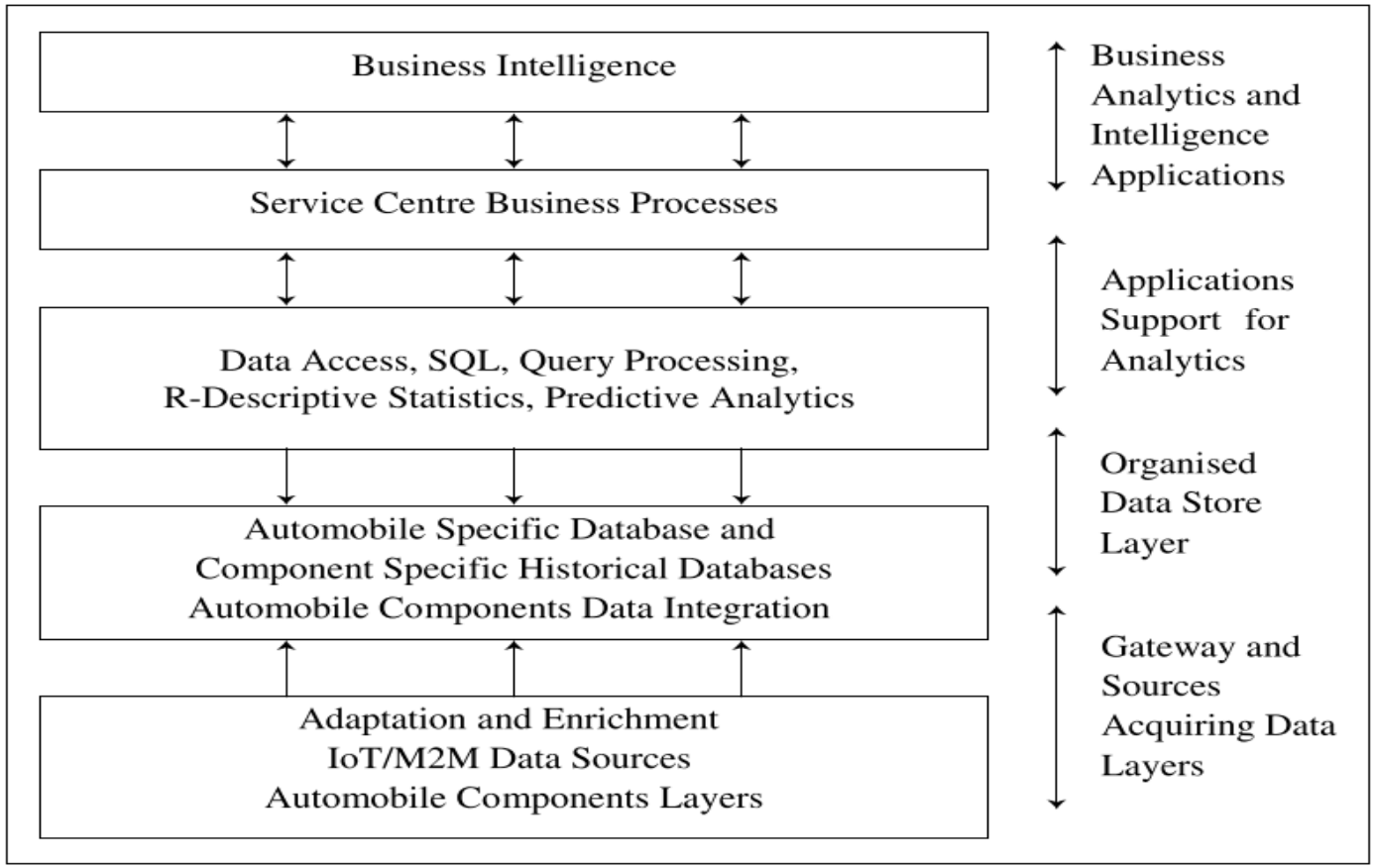

Figure 5.1 Architecture reference model for the business intelligence and business processes at ACPAMS

Distributed Business Process

- Distribution of processes reduces the complexity, communication costs, enables faster responses and smaller processing load at the central system.
- For example, streetlight project, distribution of control process for each group of lights at the gateway itself reduces complexity, communication costs, faster responses and smaller processing load at the central system.

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.
- Standardized 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
  - Collaboration and knowledge management
  - Enterprise architecture and SOA
  - E-commerce
  - External customer services
  - Supply chain automation and analytics results visualization
  - Data center optimization

Integration and Enterprise Systems

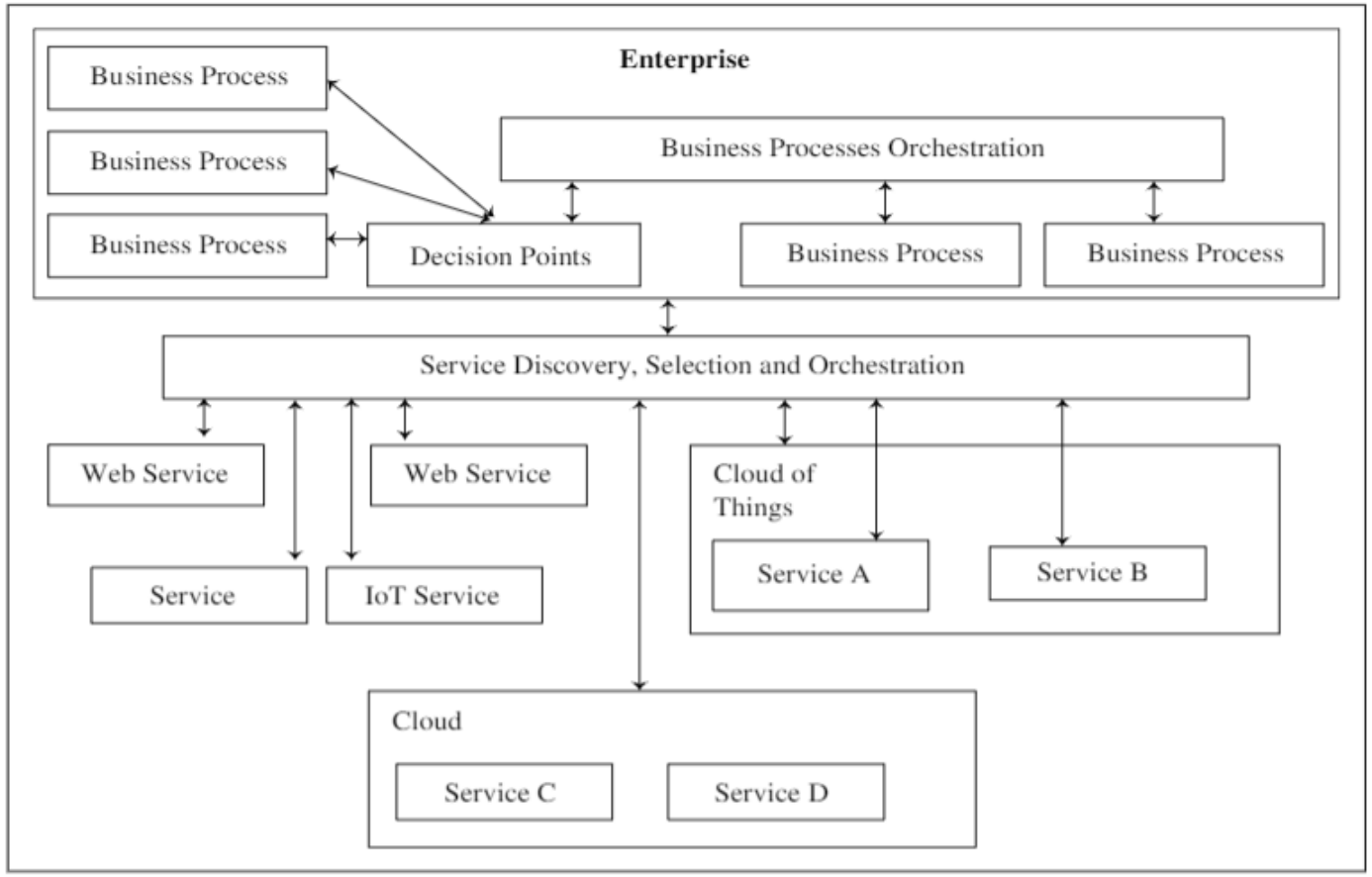

Figure 5.3 Complex applications integration architecture and SOA of cloud-based IoT services, web services, cloud services and services

Integration and Enterprise Systems

- Process orchestration means a number of business processes running in parallel and a number of processes running in sequence.
- The process matrix provides the decision points which indicate which processes should run in parallel and which in sequence.
- An SOA models the number of services and interrelationships.
- Each service initiates on receipt of messages from a process or service.
- The service discovery and selection software components select the services for application integration.
- Service orchestration software coordinates the execution of the number of services, cloud services, cloud IoT services and Web services.
- Services run in parallel and a number of processes in sequences.

3.4 Knowledge Acquiring, managing and storing Processes

Information And Knowledge

- **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.
- **Knowledge**: according to an English dictionary is sharable information and understanding about a subject or context.
- Knowledge comes from researching gives existing information.
- IoT data sources continuously generate data, which the applications or processes acquire , organize and integrates or enriches using analytics.
- Knowledge discovery tools provide the knowledge at particular point of time as more and more data is processes and analyzed .
- Knowledge is an important asset of an enterprise.

Knowledge Management

- Knowledge management (KM) is managing knowledge when the new knowledge is regularly acquired, processed and stored.
- Knowledge management also provisions for replacing the earlier gathered knowledge and managing the life cycle of stored knowledge.
- A management tool role is to create , control , use, monitor and delete.
- A KM tool has processes for discovering, using, sharing, replacing with new, creating and managing the knowledge database and information of the enterprise.

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”.
- Judgment from the knowledge of clients of a particular bank, "Operating a free dispensary will improve health of the clients”, then they will earn more and consequently bank expects to attract bigger deposits is wisdom.

Knowledge-Management Reference Architecture

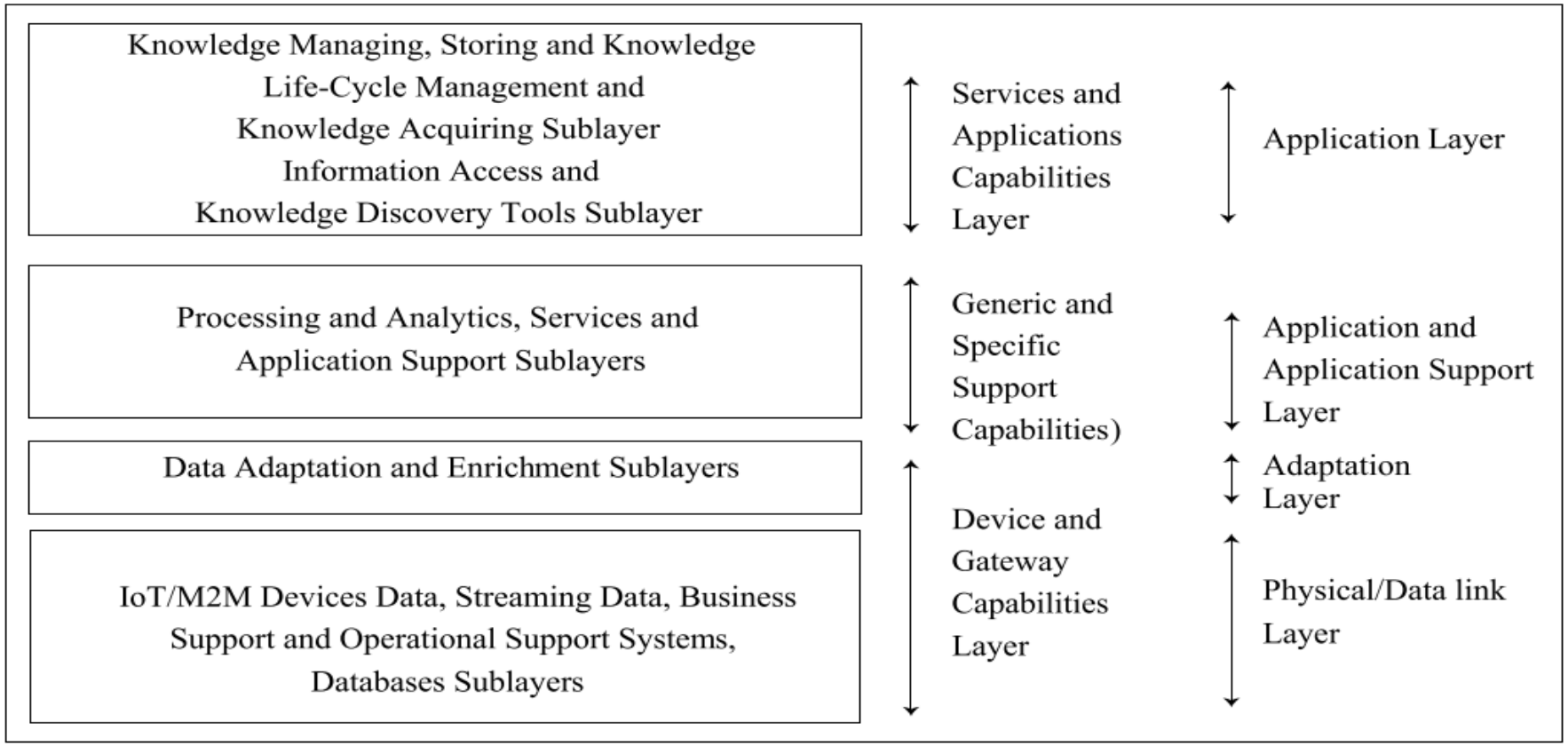

Figure 5.6 (a) A reference architecture for the knowledge management (left-hand side) and (b) Correspondence in terms of ITU-T reference model and OSI layers for IoT/M2M (middle and right-hand side)

Knowledge-Management Reference Architecture

- The lowest layer sublayers for devices data, streaming data sources which provide input for analytics and knowledge. Data bases, business support systems, operational support systems data can also be additional inputs.
- Next higher layer has data adaptation and enrichment sublayers.
- Adaptation and enrichment sublayers adapt the data from the lowest layer in appropriate forms, such as database, structured data and unstructured data so that it can be used for analytics and processing.

Knowledge-Management Reference Architecture

- Next higher layer has processing and analytics sublayers.
- These sublayers are input to information access tools and knowledge discovery tools.
- The highest layer has knowledge acquiring, managing, storing and knowledge life-cycle management.
- Sublayers for managing, storing and knowledge life-cycle management.
- Knowledge acquires from the use of information access tools and knowledge discovery tools.