Chapter Ten

Database management system

内容提要

The management services layer addresses data storage (**Databases** and mass storage technologies), retrieval (search engines), usage (data mining and machine learning), and how not to be abused (data security and privacy protection).

This chapter introduces the concepts of relational databases and the new requirements for databases in the Internet of things.

。

Review

Chapters 6-9 introduce the basic concepts and technologies of various network forms at the network construction layer and their applications in the Internet of things.

- The Internet
- Wireless broadband network (wi-fi, WiMAX)
- Wireless low speed network (Bluetooth, infrared, 802.15.4/ZigBee)
- Mobile communication network (3G)

This chapter mainly introduces the basic concepts of relational database, and briefly discusses the related theories of relational algebra, and finally discusses the characteristics and requirements of data management in the Internet of things.

Content

10.1 Origin and development of the database system

10.2 Relational database

10.3 Relational algebra

10.4 Internet of things and database

What is the database?

What are the categories and new developments of the database?

What is a database?

Database is a collection of related data stored together

- This data is structured, harmless, or unnecessarily redundant, and serves multiple applications
- The storage of the database is independent of the program that USES it
- Inserting new data, modifying and retrieving original data can be done in a common and controllable way

Application of database:

- **PC:** find files quickly
- **Company:** financial management software
- **Internet:** BBS data storage
- **Internet of things:** mass data management

Early database systems: navigational databases

- Mesh model
  - ✓ By Turing award winner Charles Bachman
  - ✓ The first database system: IDS
- The hierarchical model
  - ✓ Typical representative: IMS system of IBM

Disadvantages:

- ✓ The storage structure of data depends on the type of data
- ✓ Data is concatenated through Pointers, and you may need to traverse the entire database in order to access the desired content
- ✓ Lookup operations are expensive

RDBMS

Theoretical foundation: A Relational Model of Data for Large Shared Data Banks, published by Edgar Codd in 1970

Important thought: <u>logical composition is separated from storage structure</u>

The early days of the famous relational database system

- ✓ System R (1970, IBM)
- ✓ Ingres (university of California, Berkeley, 1973)

The current mainstream <u>relational database system</u>

ORACLE

Microsoft SQL Server 2005

MySQL

New developments in database systems

Disadvantages of relational database systems

- Lack of effective expression of real-world entities
- Lack of efficient processing of complex queries
- Lack of effective support for WEB applications

XML database (Tamino et al., Software AG, Germany) : for XML data

Spatial and temporal database (Oracle 8i Spatial) : for geographic information systems

NoSQL database (Cassandra of the Apache foundation) : for Web data

Content

10.1 Origin and development of the database system

10.2 Relational database

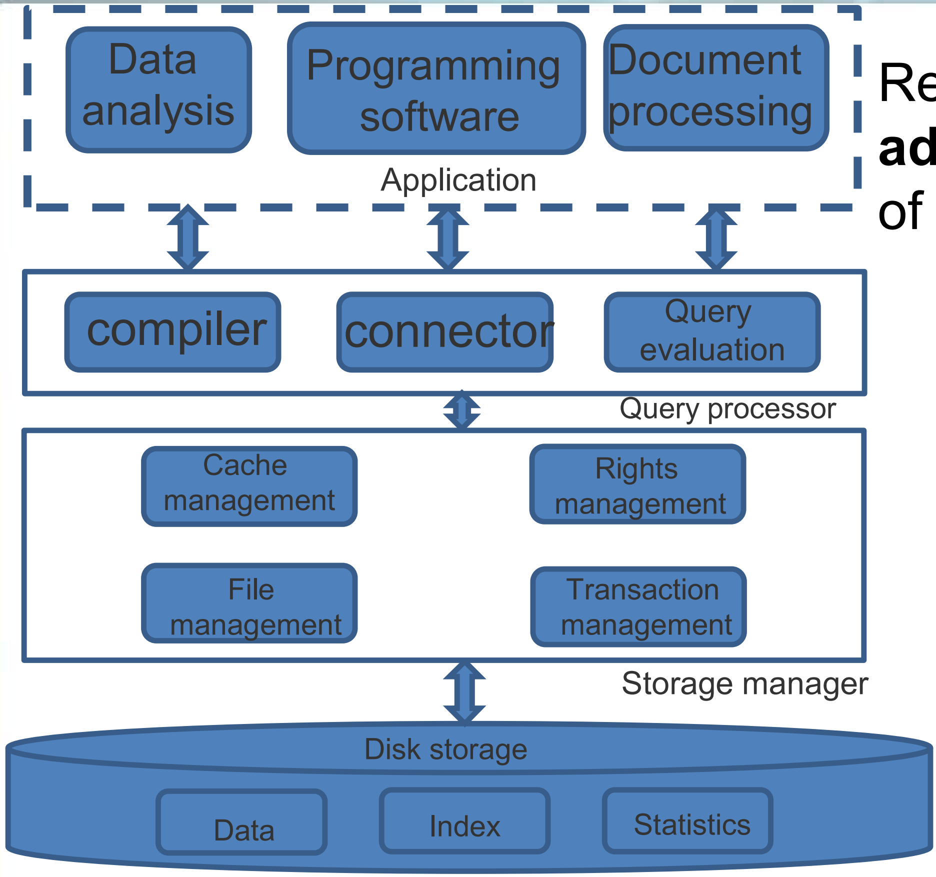
10.3 Relational algebra

10.4 Internet of things and database

What is the database?

What are the categories and new developments of the database?

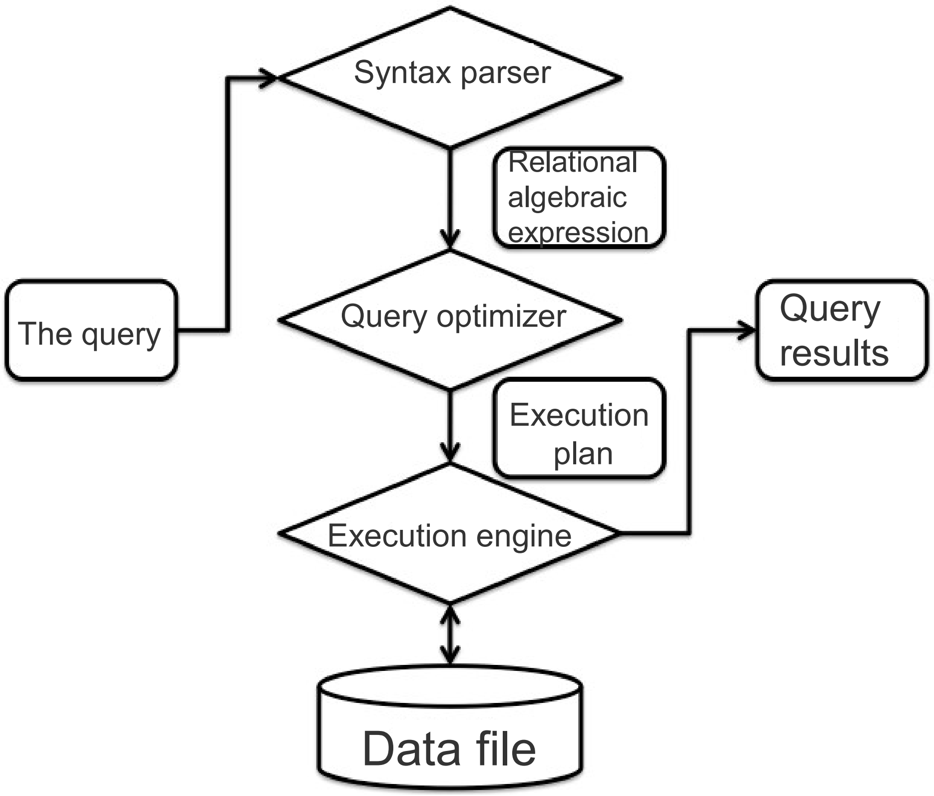
Schema diagram of a relational database



Relational databases have **advantages** over earlier architectures of navigational databases

- ✓ A high degree of data independence
- ✓ Open data semantics, data consistency, data redundancy
- ✓ Flexible custom data manipulation language

Schema diagram of a relational database



The query process has three steps

- The application sends the query statement to the database front end
- The database parses and optimizes the query statements
- The database queries the data file and returns the results as per the execution plan

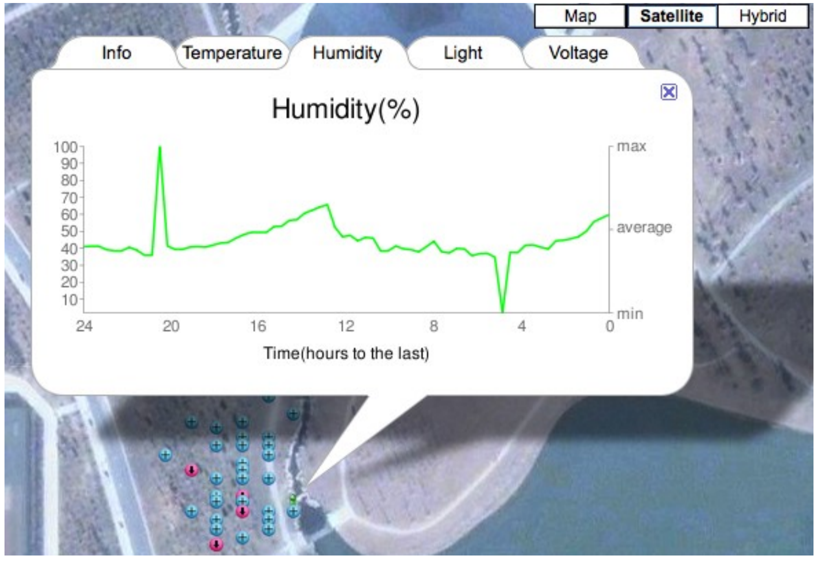
Relational database model: take greenorbs as an example

A relational database is a collection of relationships with different names

Greenorbs (http://orbsmap.greenorbs.org)

Stored data

- Basic information of sensor node (number, longitude and latitude, etc.)
- Ambient temperature, relative humidity, light intensity, sensor voltage



Each point in the diagram corresponds to a sensor deployed in the field

Relational database model

Relation

- A relationship is a table in a database
- Relational tables are only the logical organization of data, which can be physically indexed by data structures such as B+ trees
- Can express the physical quantity of the real world intuitively

sid	longitude	latitude
1	119.7227	30.2585
2	119.7224	30.2586
3	119.7225	30.2585
4	119.7224	30.2584
5	119.7220	30.2580

Relation R: Sensor position table

Meaning:
The location of sid 1
Longitude 119.7227
Latitude 30.2585

Relational database model

Attribute

- Each column of a relational table is called an attribute that describes some aspect of the data
- Each column of the table contains, and only contains one value of a property
- The value of an attribute can be of various integer, real, date-time, and other types

Attribute	Meaning	Type
Id	The unique number of sensors in the network	Int
Longitude	The longitude of the geographic location of the sensor	Float
Latitude	The latitude of the geographic location of the sensor	Float

Relational database model

Domain (Domain)

- A domain is a collection of values of the same data type
- Any properties in the table must be defined on the domain

Natural number domain (sid) — Real number field (temperature, humidity) — Time date field (Update-time)

sid	temperature	humidity	Update-time
2	24.16	85.75	9/17/2009 7:18
3	23.47	88.49	9/17/2009 6:19
5	23.57	86.01	9/17/2009 5:20
12	21.89	95.59	9/17/2009 8:20
21	28.60	68.28	9/17/2009 8:19

Relational database model

Tuple

- Each row in a relational table is called a tuple
- Tuples are the basic building blocks of relationships
- In relationships, the order in which tuples are arranged is not important

sid	longitude	latitude
1	119.7227	30.2585
2	119.7224	30.2586
3	119.7225	30.2585
4	119.7224	30.2584
5	119.7220	30.2580

Tuple

Relational database model

Degree: the number of attributes in a relational table is called degrees
Cardinality: the number of tuples contained ina relational table is called Cardinality

Degree = 3

sid	longitude	latitude
1	119.7227	30.2585
2	119.7224	30.2586
3	119.7225	30.2585
4	119.7224	30.2584
5	119.7220	30.2580

Base = 5

Relational database model

Schema

The name of the relationship and the collection of properties it contains are collectively referred to as schemas

Use "relationship name (attribute 1, attribute 2, attribute 3...) to represent patterns

The order of attributes in the schema is unordered

sid	longitude	latitude
1	119.7227	30.2585
2	119.7224	30.2586

Sensor position

sid	temperature	humidity	updatetime
2	24.16	85.75	9/17/2009 7:18
3	23.47	88.49	9/17/2009 6:19

Sensor data

Content

Relational algebra is the basis of data manipulation in relational databases.

10.3 Relational algebra

The data model of relational database not only defines the structure of database (relation, attribute, tuple, etc.), but also provides the method of querying data and modifying data. The data operation is based on the special algebraic operation of "relation algebra”

The operands of relational algebra are relational (the operands of traditional algebra are numeric constants or variables)

Operators of relational algebra fall into four main categories

- ✓ Traditional collection operators
- ✓ Special relational operators
- ✓ Comparison operator
- ✓ Logical operator

Traditional collection operators

Pay (∩)

Relation R and relation S are composed of tuples belonging to both R and S, that is, $R \cap S = \{t \mid t \in R \wedge t \in S\}$

Poor (-)

Relationship of R and S to belong to R but does not belong to S of tuples, namely $R - S = \{t \mid t \in R \wedge t \notin S\}$.

And (∪)

The relationship of R and S to belongs to R or to S of tuples, namely $R \cup S = \{t \mid t \in R \vee t \in S\}$

Generalized cartesian product (×)

The generalized cartesian product of relation R (degree n) and relation S (degree m) is a set of tuples with n+m attributes, where the first n attributes come from relation R and the last m attributes from relation S, namely $R \times S = \{t_r t_s \mid t_r \in R \wedge t_s \in S\}$

Traditional collection operators

sid	longitude	latitude
1	119.7227	30.2585
2	119.7224	30.2586
3	119.7225	30.2585
4	119.7224	30.2584
5	119.7220	30.2580

Relation R

sid	longitude	latitude
1	119.7227	30.2585
3	119.7225	30.2585
5	119.7220	30.2580
7	119.7222	30.2583
9	119.7222	30.2584

Relation S

Traditional collection operators

sid	longitude	latitude
1	119.7227	30.2585
2	119.7224	30.2586
3	119.7225	30.2585
4	119.7224	30.2584
5	119.7220	30.2580
7	119.7222	30.2583
9	119.7222	30.2584

$R \cup S$

sid	longitude	latitude
1	119.7227	30.2585
3	119.7225	30.2585
5	119.7220	30.2580

$R \cap S$

Traditional collection operators

R.sid	R.longitude	R.latitude	S.sid	S.longitude	S.latitude
1	119.7227	30.2585	2	119.7224	30.2586
1	119.7227	30.2585	4	119.7224	30.2584
3	119.7225	30.2585	2	119.7224	30.2586
3	119.7225	30.2585	4	119.7224	30.2584
5	119.7220	30.2580	2	119.7224	30.2586
5	119.7220	30.2580	4	119.7224	30.2584

Cartesian product

sid	longitude	latitude
2	119.7224	30.2586
4	119.7224	30.2584

R-S

Special relational operators

Project(**π**)

The projection operation is used to generate a new relation S from the existing relation R such that S contains some columns of R. Expression of PI A1, A2,..., the result of An(R) is the attribute columns A1,A2,...The new relationship of An S

Select**(σ)**

The selection operation is used to create a new relationship S from the existing relationship R such that S is a collection of tuples in R satisfying condition C, called sigma C (R)

Natural connection (⋈)

Connect the two relationships by having the same common attribute value

Special relational operators

θ connection

Connect the two relationships based on condition C

It can be done in two steps

- ✓ Take the cartesian product of R and S, R by S
- ✓ Select all tuples that satisfy condition C from $R \times S$

Special relational operators

$\pi_{longitude}(R)$

longitude
119.7227
119.7224
119.7225
119.7224
119.7220

$\sigma_{sid=3}(R)$

sid	longitude	latitude
3	119.7225	30.2585

Natural connection $R \bowtie T$

sid	longitude	latitude	temperature	humidity	updatetime
2	119.7224	30.2586	24.16	85.75	9/17/2009 7:18
3	119.7225	30.2585	23.47	88.49	9/17/2009 6:19
5	119.7220	30.2580	23.57	86.01	9/17/2009 5:20

Special relational operators

R1.sid	R1.longitude	R1.latitude	R2.sid	R2.longitude	R2.latitude
1	119.7227	30.2585	4	119.7224	30.2584
1	119.7227	30.2585	5	119.7220	30.2580
2	119.7224	30.2586	1	119.7227	30.2585
2	119.7224	30.2586	3	119.7225	30.2585
2	119.7224	30.2586	4	119.7224	30.2584
2	119.7224	30.2586	5	119.7220	30.2580
3	119.7225	30.2585	4	119.7224	30.2584
3	119.7225	30.2585	5	119.7220	30.2580
4	119.7224	30.2584	5	119.7220	30.2580

Make θ connection to table R, where condition C=R1. Latitude >R2. Latitude, R1 and R2 are relation R

SQL query language*

SQL(Structured Query Language) is the most widely used Query Language in modern databases. The query part of the syntax is very similar to relational algebra.

•**Select $\sigma_C(R)$：** SELECT * FROM R WHERE C;

•**Project $\pi_{A1,A2,\ldots,An}(\sigma_C(R))$：** SELECT A_1， A_2， ……， A_k FROM R WHERE C;

•**Set operation：** INTERSECT、EXCEPT、UNION

•**Union operation：** SELECT R.sid, T.temperature, T.humidity, FROM R, T WHERE R.sid = T.sid

•**Subquery：** SELECT * FROM R WHERE latitude < (SELECT latitude FROM R WHERE sid = 3);

Content

What are the characteristics of Internet of things data?

How should data be managed in the Internet of things?

Characteristics of sensor data

Magnanimity: assuming that each sensor only transmits 1K data per minute, the data volume of 1000 nodes reaches about 1.4gb per day

Polymorphism:

- ✓ Ecological monitoring system: temperature, humidity, light
- ✓ Multimedia sensor network: video, audio
- ✓ Fire navigation system: structured communication data

Relevance and semantics

- ✓ Data describing the same entity is correlated over time (temperature changes over time at the same node)
- ✓ Data describing different entities are correlated in space (temperature values of different nodes in the same area are similar)
- ✓ There is also correlation between different dimensions of the description entity (temperature and humidity measured at the same time at the same node are correlated).

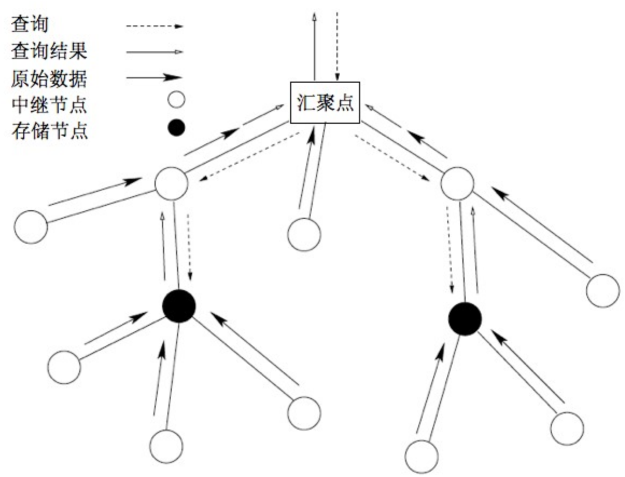
Data storage in sensor networks

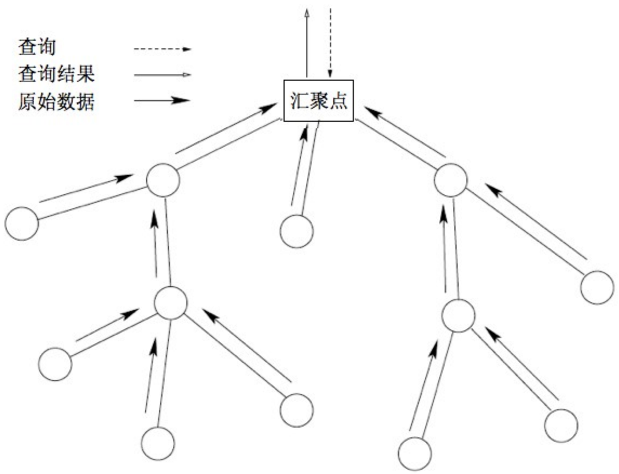
Distributed storage

- Data can be stored on a storage node
- The query is distributed to the network and returned by the storage node

Centralized storage

- All data are saved at sink end (sink point)
- The query only works at the sink end





Data query of sensor network

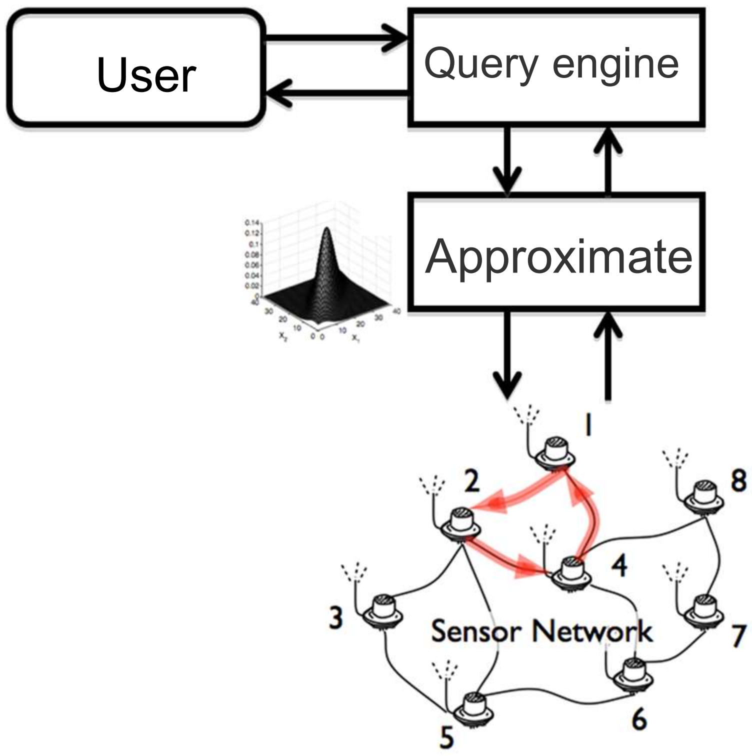
Queries are divided into **snapshot queries** and **sequential queries**.
Snapshot query features: query is not fixed, data is not fixed
Continuous query features: query fixed, data is not fixed

Approximate query technique

- Data uncertainty can reduce network communication overhead
- Model-based queries

Query optimization

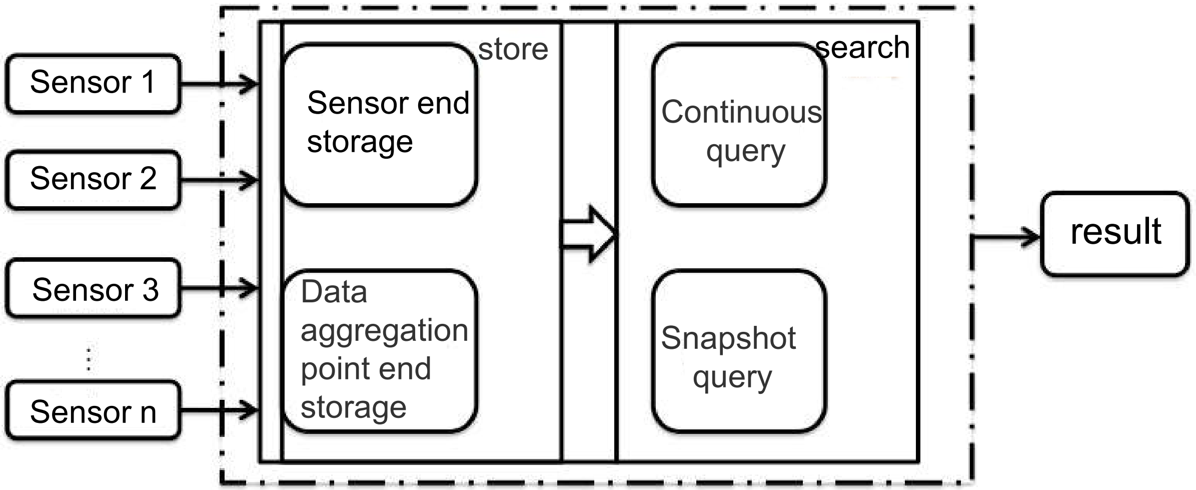
- Optimize query content for fixed queries
- The node only returns the required data
- Queries are sent only to regions that satisfy the query criteria



Data fusion of sensor networks

In the Internet, data flow flows from rich network resources to terminal equipment, and in the sensor network, data flow flows from sensor equipment to network

Data fusion, that is, how to analyze and integrate the countless data streams from different sources, is an obstacle that must be overcome by the sensor network and even the Internet of things to achieve large-scale application



Data flow management system

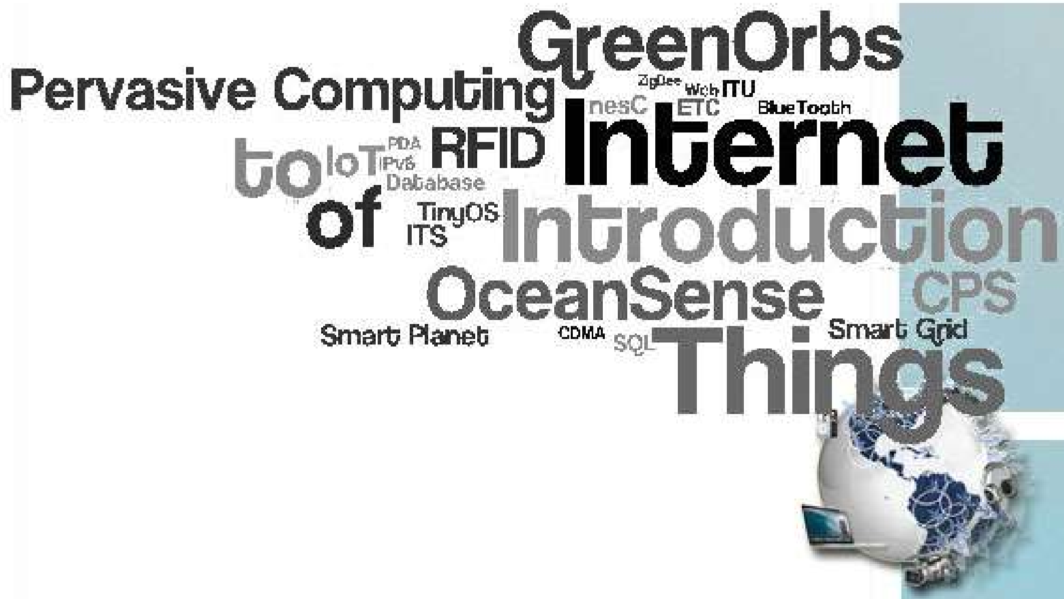
Conclusion

Review

This chapter introduces the development of database management system, mainly introduces the basic concept of relational database and the basic operation of relational algebra, and discusses the characteristics of data management in the Internet of things.

Key Points

- Understand the development of database models and illustrate the development trend of emerging databases.
- Grasp the basic concepts of relationships, attributes, domains, tuples, degrees, cardinals, patterns.
- Can write the corresponding relational algebra query expression according to the requirements.
- Grasp the characteristics of Internet of things data management (data characteristics, data query, storage and fusion methods).



Thank you!

Internet of Things