RFID Data Mining: Opportunities and Challenges

Joshua Huang

E-Business Technology Institute

The University of Hong Kong

25 August 2007

WSIETA2007
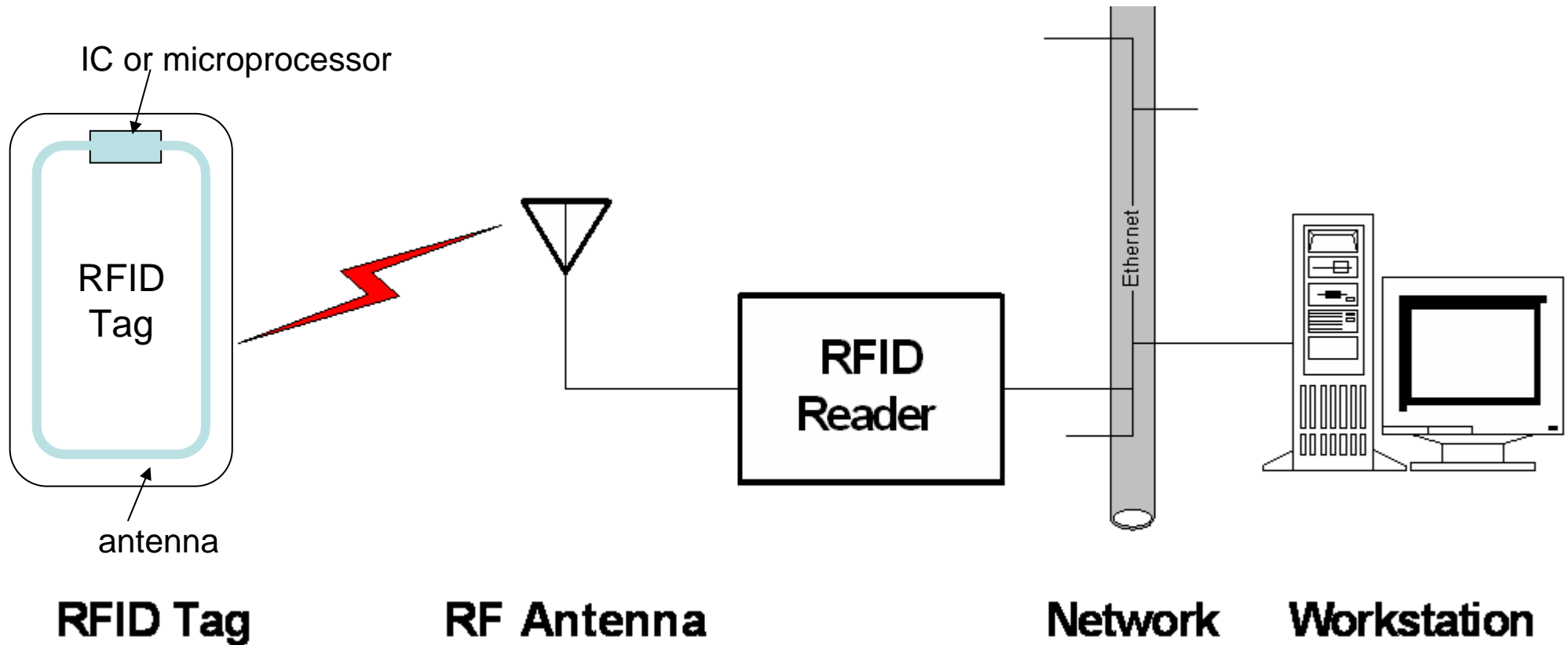
Chengdu, Sichuan, China

Agenda

- Overview of RFID Technology and Applications
- RFID in Logistics and Supply Chain Management
- Research Areas
  - **Data Collection and Processing**
  - **Data Management**
  - **Event Processing**
  - **Data Mining**
- Summary

RFID Technology Components

IC or microprocessor

RFID Tag

antenna

RFID Reader

Ethernet

RFID Tag   **RF Antenna**   **Network**   **Workstation**

Tags and Readers

- *Passive Tags*
  - Without battery.
- *Active Tags*
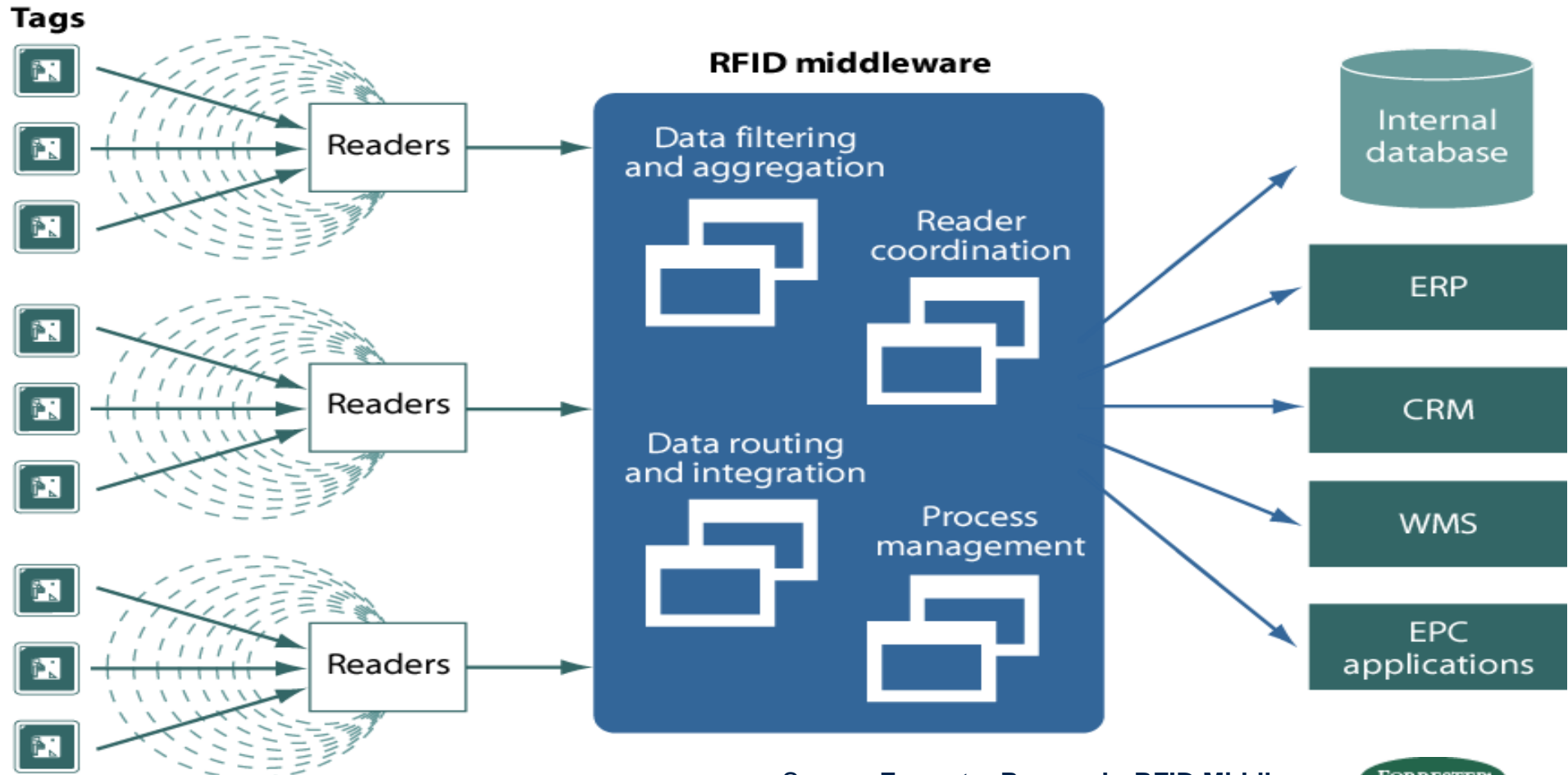 - *With battery*

2251 420

- **Readers**

"Quietized" Fully Enclosed Conveyor Tunnel
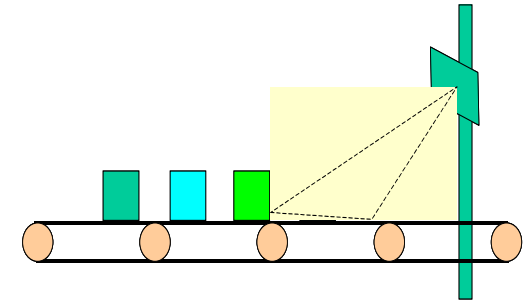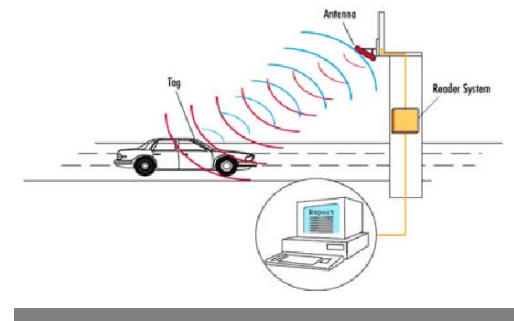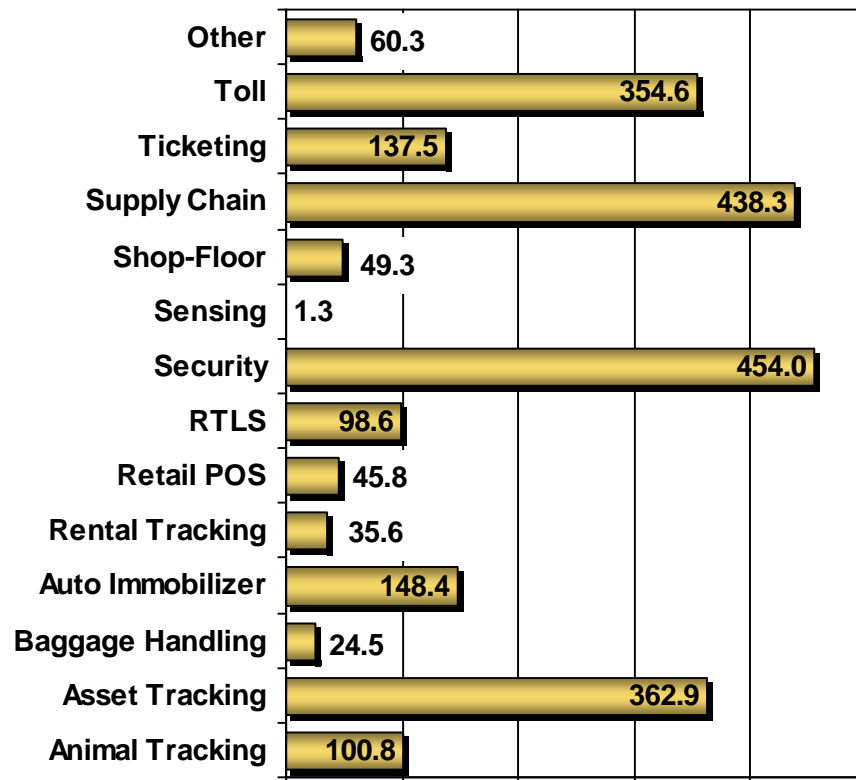
Symbol Reader

RFID middleware

Tags

Readers → RFID middleware → Internal database, ERP, CRM, WMS, EPC applications

RFID middleware:
- Data filtering and aggregation
- Reader coordination
- Data routing and integration
- Process management

Source: **Forrester Research: RFID Middleware** FORRESTER

RFID applications

Application	Value
Other	60.3
Toll	354.6
Ticketing	137.5
Supply Chain	438.3
Shop-Floor	49.3
Sensing	1.3
Security	454.0
RTLS	98.6
Retail POS	45.8
Rental Tracking	35.6
Auto Immobilizer	148.4
Baggage Handling	24.5
Asset Tracking	362.9
Animal Tracking	100.8

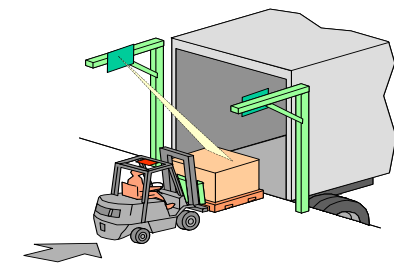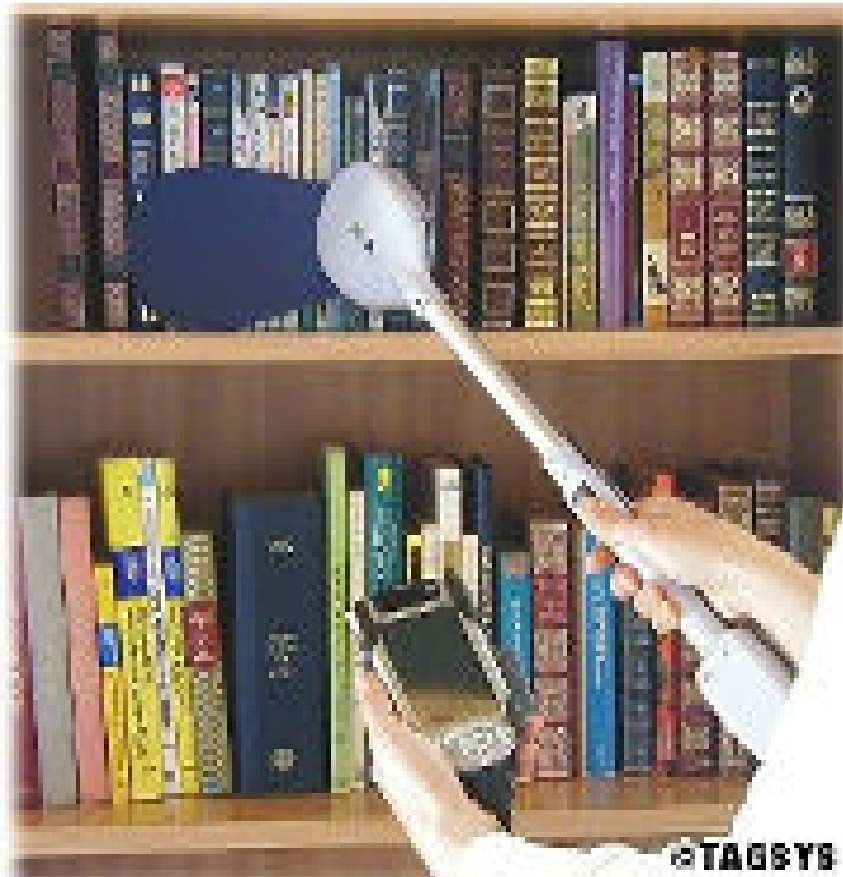**Smart Shelves**

Library Example

Book search and inventory

Security control and check out

Bridging Physical and Digital Worlds

Agenda

- Overview of RFID Technology and Applications
- RFID in Logistics and Supply Chain Management
- Research Areas
  - **Data Collection and Processing**
  - **Data Management**
  - **Event Processing**
  - **Data Mining**
- Summary

Logistics and Supply Chain Management

- **Logistics** - ...the process of planning, implementing, and controlling the efficient, effective flow and storage of goods, services, and related information from point of origin to point of consumption for the purpose of conforming to customer requirements. -- (Council of Logistics Management, http://www.clm1.org/mission.html, 12 Feb 98)

- **Supply Chain Management (SCM) - …**the oversight of materials, information, and finances moving in a process from supplier to manufacturer to wholesaler to retailer to consumer. SCM involves coordinating and integrating these flows both within and among companies.

Logistics and Supply Chain

- **Supply Chain Model**

```
                    ┌─────────────────────────────────┐
                    │  Fulfillments information flow   │
                    └─────────────────────────────────┘
```

| Customers | ↔ | Physical Distribution | ↔ | Manufac-turing | ↔ | Purchasing | ↔ | Suppliers |

```
                    ┌─────────────────────────────────┐
                    │  Requirements information flow   │
                    └─────────────────────────────────┘
```
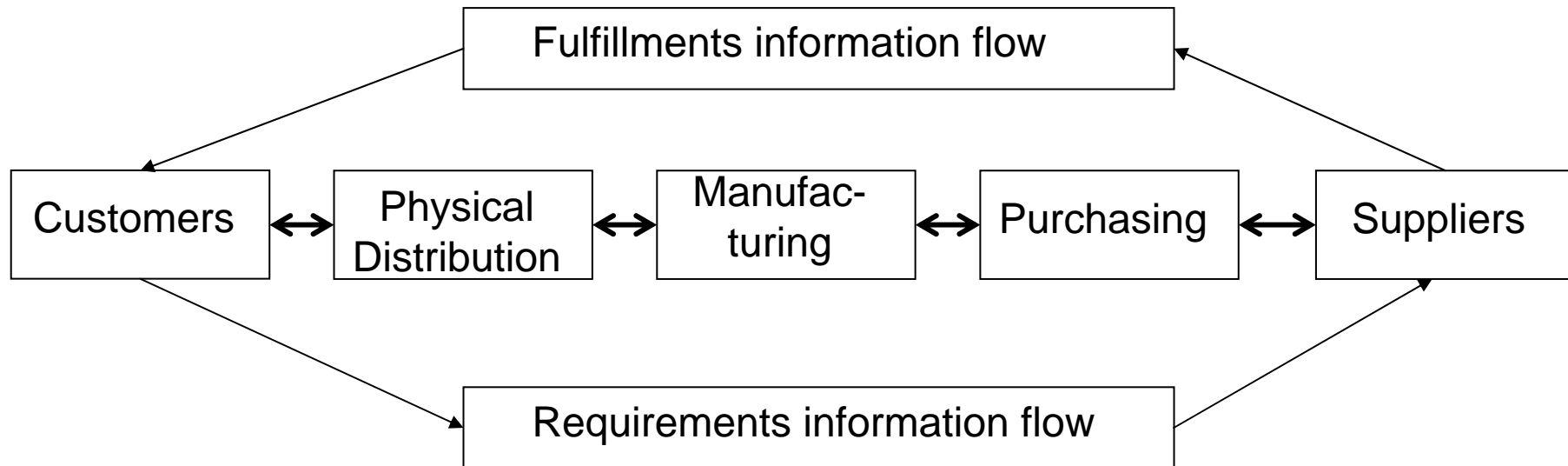
- *Supply types*
  - **One-origin/one-destination**
  - **One-origin/multiple-destinations**
  - **Multiple-origins/one-destination**
  - **Multiple-origins/multiple-destinations**

Business Objectives

- Cost reduction
- Inventory control
- Reduce lead time
- Supply chain optimization

Customers

Acknowledgement

Goods

| J+1 | → | J | → | J-1 |

Orders

● ● ●   ● ● ●
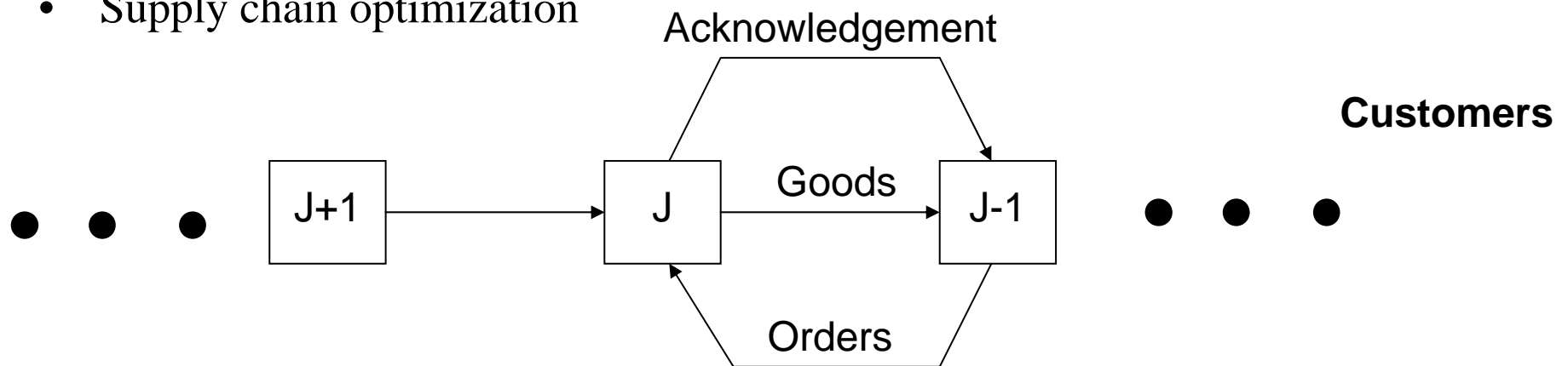
- Problems
  - **Mismatch between physical inventory and inventory database**
  - **Lack of information sharing for entire supply chain optimization**

RFID Applications in SCM

- Bridge the gaps between physical inventory and digital inventory
- Real time data collection to update inventory databases
- Enable information sharing, e.g., through EPC network
- Real time product tracking and positioning
- Enable global supply chain optimization

Food Supply Chain Example

① Radio tag placed on carton.

② Dairy ships carton to grocery store.

③ Consumer purchases tagged carton.

④ Consumer recycles milk carton.

⑤ Carton arrives at recycling center.
   Manufacturer produces replacement.

Ⓐ Manufacturer tracks product through
   wireless radio communication.
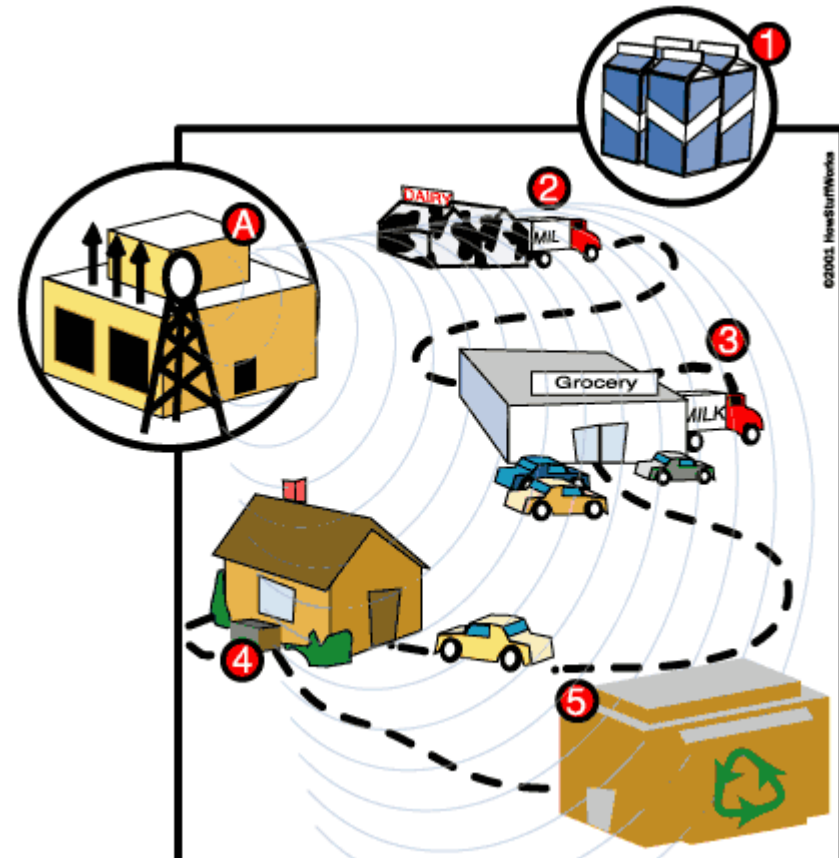
*Products are tracked through their
entire lifetime*

Source: **How Stuff Works**

- RFID enabled supply chain operation will generate massive data that needs to be stored, managed and analyzed

- *How much data?*
  - Consider a supermarket chain implementing RFID:
  - 12 bytes EPC + Reader ID + Time = 18 bytes per tag
  - Average number of tags in a neighborhood store = 700,000
  - Data generated per second = 12.6 GB
  - Data generated per day = 544 TB
  - Assuming 50 stores in the chain,
    - *data generated per day = 2720 TB*

Agenda

- Overview of RFID Technology and Applications
- RFID in Logistics and Supply Chain Management
- Research Areas
  - **Data Collection and Processing**
  - **Data Management**
  - **Event Processing**
  - **Data Mining**
- Summary

Research Areas

Data Mining

Event Processing

Data Management
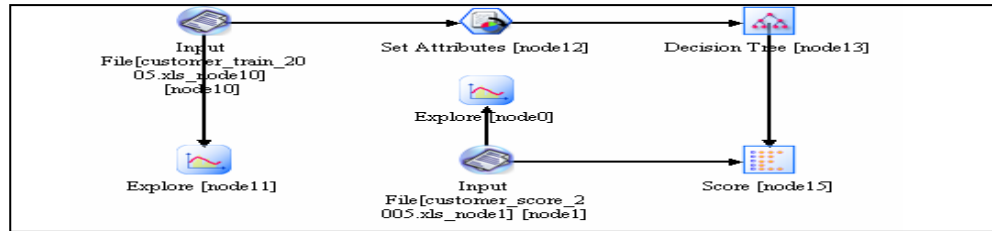
Data Collection and Processing
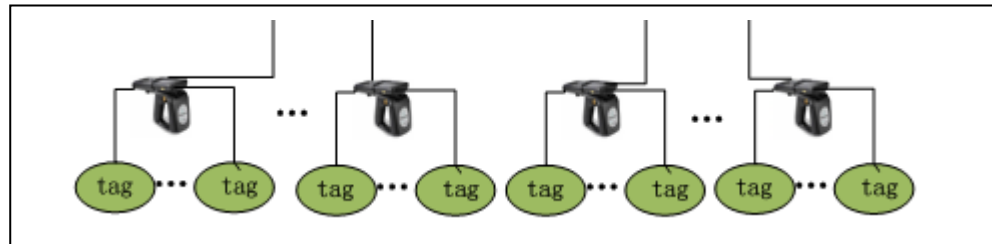
Agenda

- Overview of RFID Technology and Applications
- RFID in Logistics and Supply Chain Management
- **Research Areas**
  - **Data Collection and Processing**
  - **Data Management**
  - **Event Processing**
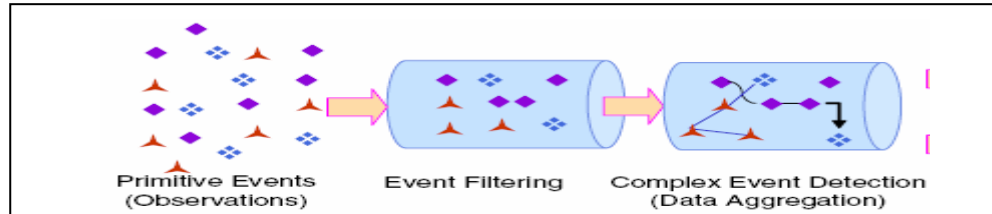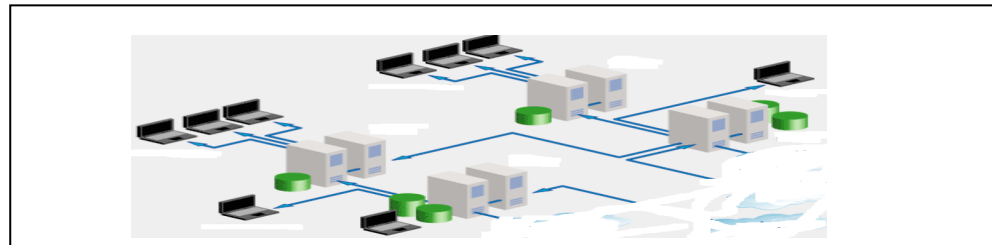 - **Data Mining**
- Summary

- **Data collection problems**
  - False readings, i.e., read unexpected tags
  - Duplicate readings, e.g., same reader reading many times, multiple readers reading the same tag
  - Missed readings, i.e., tags undetected due to RF interference, or malfunctions of tags or readers
  - High speed and large volume, e.g., many tags present to many readers

- **Solutions**
  - Multiple cycles reading to reduce missed readings
  - Data filtering and cleansing to remove duplicates and false readings
  - Data abstraction and compression
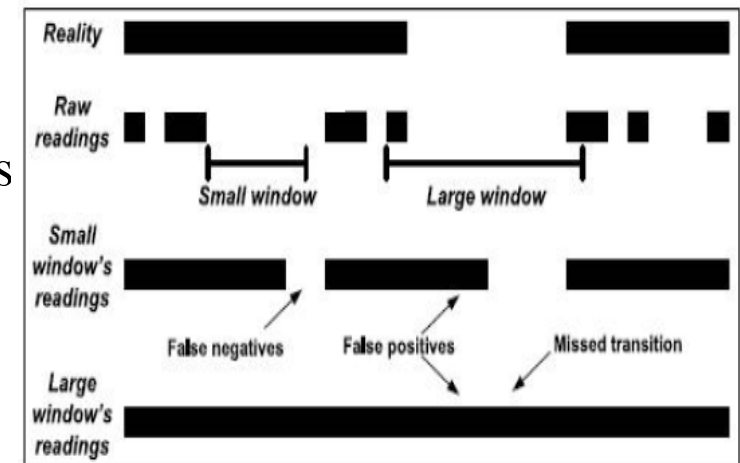
- Raw Data
  - (EPC, location, time)
  - Duplicate records due to multiple readings of a product at the same location
  - (r_1, l_1, t_1) (r_1, l_1, t_2) ... (r_1, l_1, t_{10})
- Cleansed Data: Minimal information to store and removal of raw data
  - (EPC, Location, time_in, time_out)
  - (r_1, l_1, t_1, t_{10})
- Fill-up missing records and correct wrongly-registered information through multiple data collection points (e.g., manufacturing process, different storages at large distribution centers)

Agenda

- Overview of RFID Technology and Applications
- RFID in Logistics and Supply Chain Management
- Research Areas
  - **Data Collection and Processing**
  - **Data Management**
  - **Event Processing**
  - **Data Mining**
- Summary

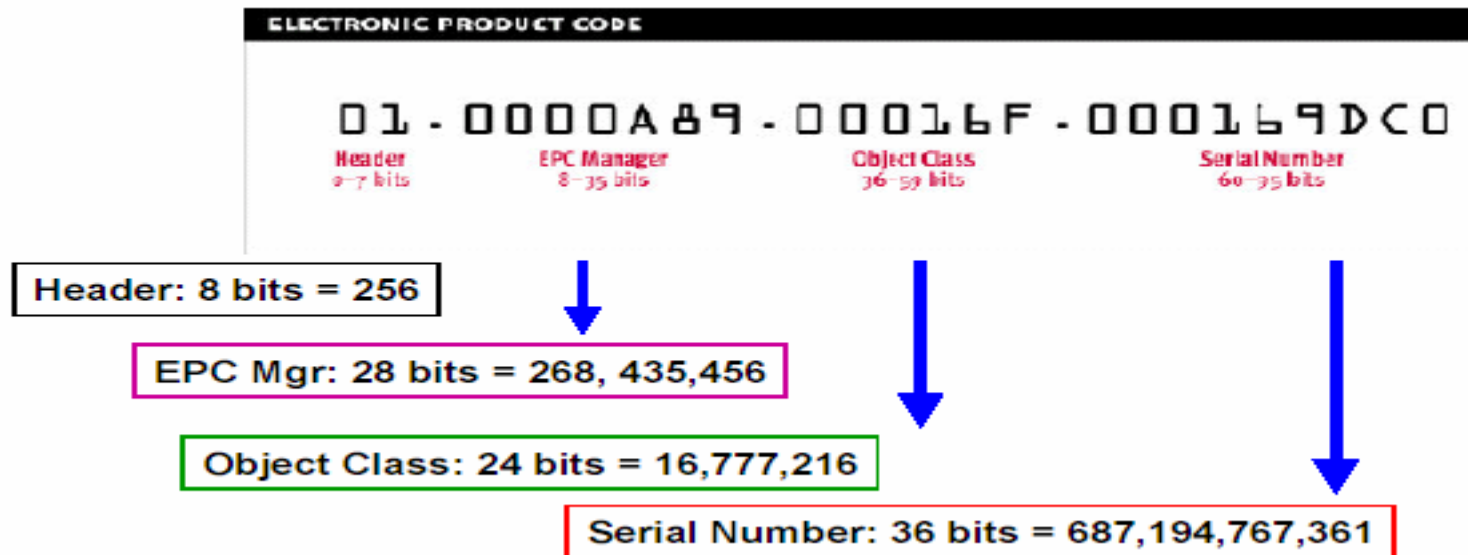- Identify an object at a specific time and a specific location
  - Three essential data elements:
    - Object identification, location and time: (ID, Loc, T)
  - Data of other attributes of identified objects are stored in a database and can be matched through ID
- A lot of ID data are collected from readers in the business process

Electronic Product Code
96 bits Standard

ELECTRONIC PRODUCT CODE

```
01 - 0000A89 - 00016F - 000169DC0
```

Header	EPC Manager	Object Class	Serial Number
0-7 bits	8-35 bits	36-59 bits	60-95 bits

Header: 8 bits = 256

EPC Mgr: 28 bits = 268, 435,456

Object Class: 24 bits = 16,777,216

Serial Number: 36 bits = 687,194,767,361

Header - Tag version number
EPC Manager   -   Manufacturer ID
Object class   - Manufacturer's product ID
Serial Number - Unit ID

With 96 bit code, 268 million companies can each categorize 16 million different products where each product category contains up to 687 billion individual units

RFID Data Modeling

- Different applications may require different data schema designs

- Data models must support data queries and analysis

- In supply chain management, modeling the process flow is important
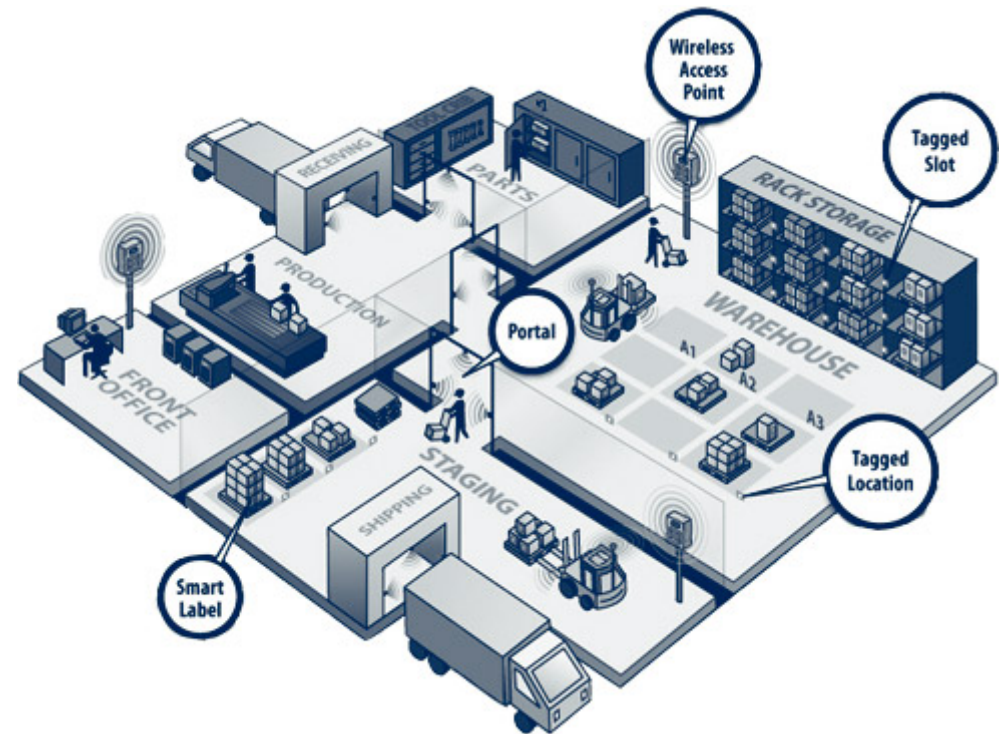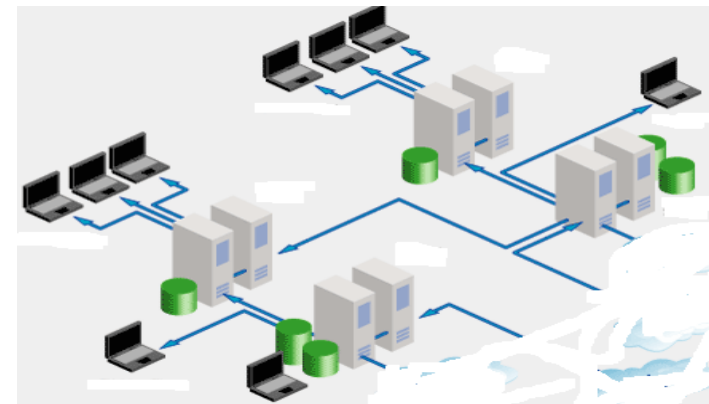

www.aeroid.co.uk/rfidbasics.html

- Indexing on EPC code
  - How to build the EPC code index to support query and analysis
- EPC code based fast query processing
  - Query related to different parts of EPC code
- Distributed data management
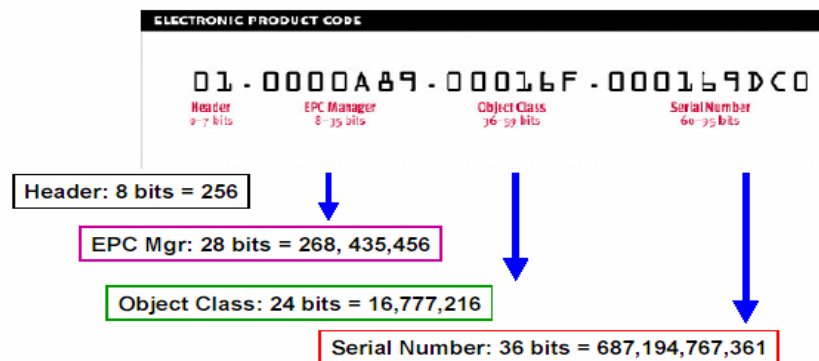
Agenda

- Overview of RFID Technology and Applications
- RFID in Logistics and Supply Chain Management
- Research Areas
  - **Data Collection and Processing**
  - **Data Management**
  - **Event Processing**
  - **Data Mining**
- Summary

RFID Event Processing

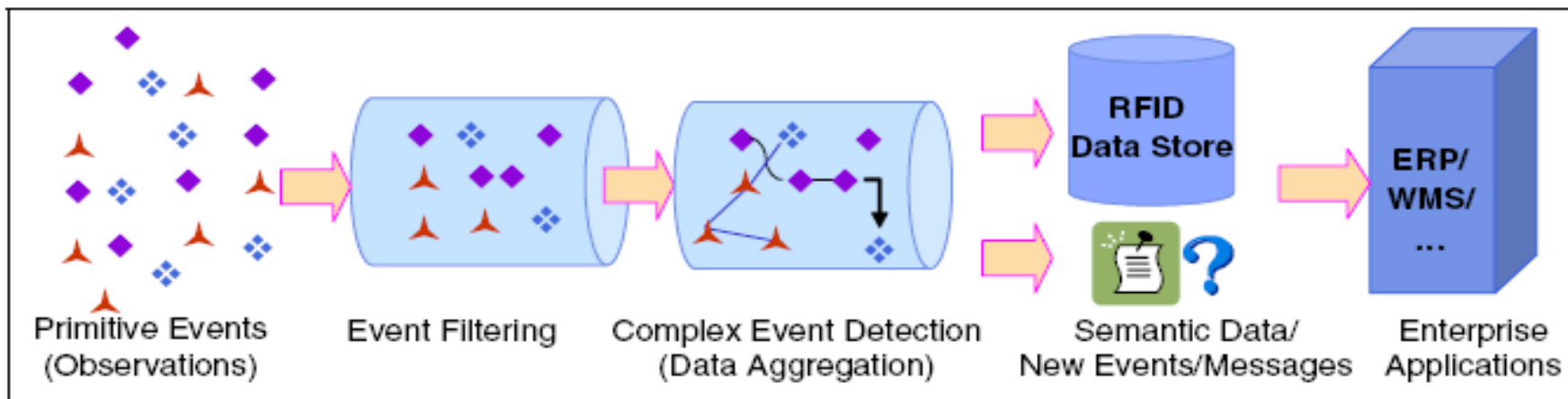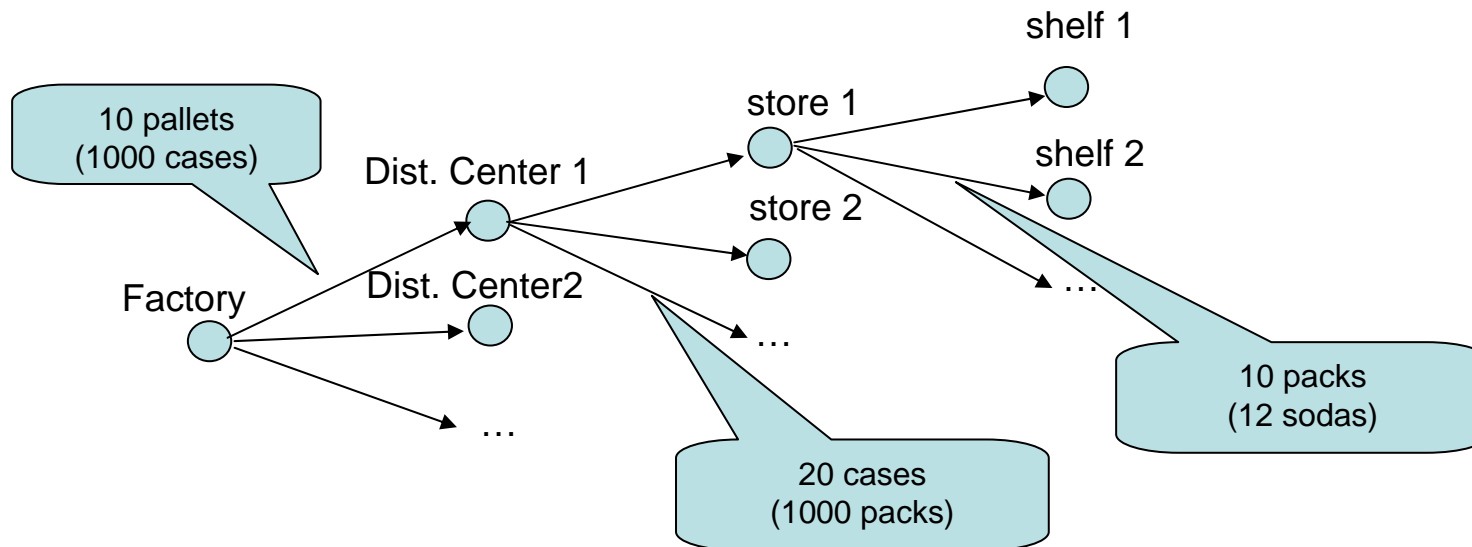- Data query and analysis are often event-oriented
- Events are defined in different ways and at different levels
- Event processing generates different events from raw RFID data

Primitive Events (Observations) → Event Filtering → Complex Event Detection (Data Aggregation) → RFID Data Store / Semantic Data/ New Events/Messages → ERP/ WMS/ ... Enterprise Applications

Grouping Data Into Different Events with GID

- Objects often move and stay together (e.g., purchase orders, shipments)
- If 1000 packs of soda stay together at the distribution center, register a single record
  - *(GID, distribution center, time_in, time_out)*
- GID is a group identifier that represents the 1000 packs that stayed together at the distribution center

shelf 1

10 pallets
(1000 cases)

store 1

Dist. Center 1

shelf 2

store 2

Factory

Dist. Center2

…

…

…

20 cases
(1000 packs)

10 packs
(12 sodas)

Grouping Table

	#EPCs	#GIDs
l1	n	1
l2, l3, l4	n	3
l5, l6, l7, l8, l9, l10	n	6
	3n	10+n
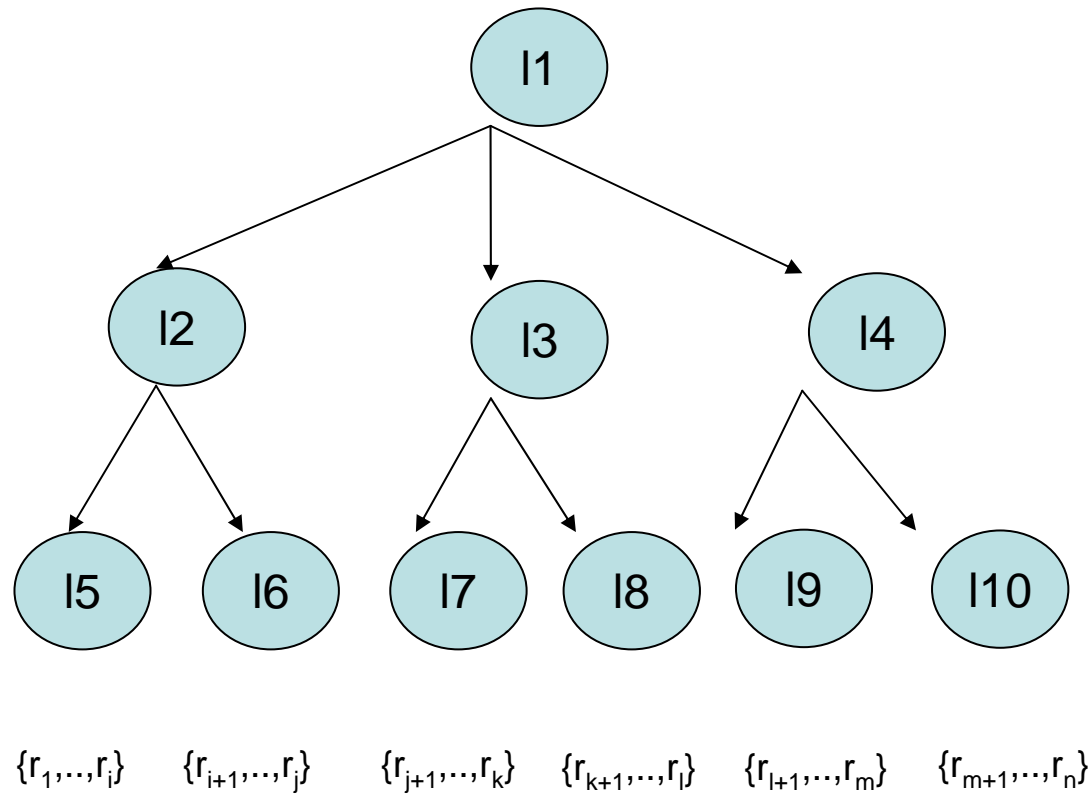
$\{r_1,...,r_i\}$ $\{r_{i+1},...,r_j\}$ $\{r_{j+1},...,r_k\}$ $\{r_{k+1},..,r_l\}$ $\{r_{l+1},...,r_m\}$ $\{r_{m+1},..,r_n\}$

Different Events Generated for Different Purposes

Store View:

| Transportation | → | backroom | → | shelf | → | checkout |

| dist. center | → | truck | → | backroom | → | shelf | → | checkout |
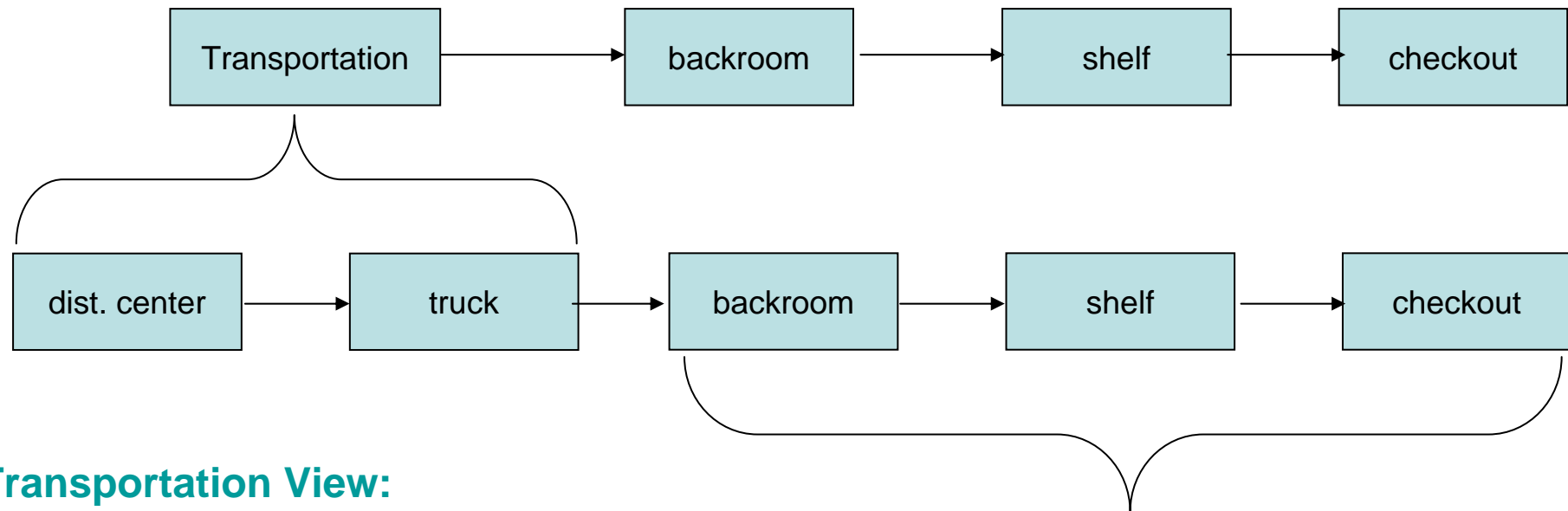
Transportation View:

| dist. center | → | truck | → | Store |

Agenda

- Overview of RFID Technology and Applications
- RFID in Logistics and Supply Chain Management
- Research Areas
  - **Data Collection and Processing**
  - **Data Management**
  - **Event Processing**
 - **Data Mining**
- Summary

Data Mining Tasks

- Classification
- Prediction
- Clustering
- Association Analysis

RFID Data Mining Research

- Data mining problems
  - RFID data cleansing
  - Event based data mining
    - *Event classification*
    - *Event prediction*
    - *Event clustering*
    - *Event association*
  - Outlier event detection and analysis
- New techniques
  - Efficient algorithms
  - Distributed data mining
- Applications
  - Supply chain analysis and optimization
  - Inventory control
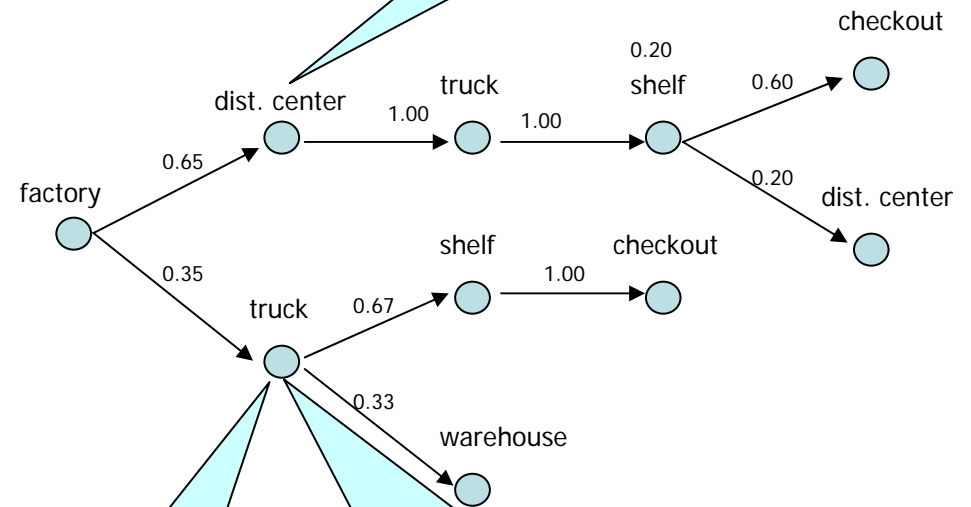 - Supply chain management
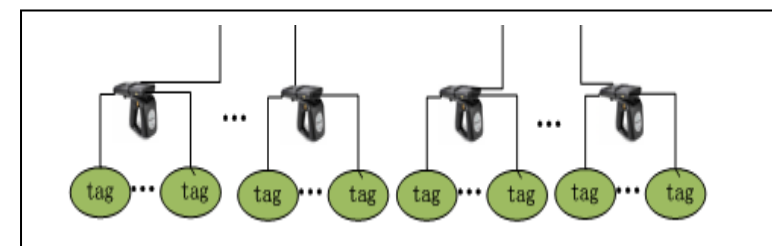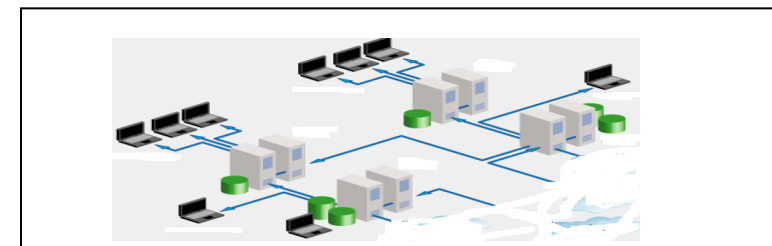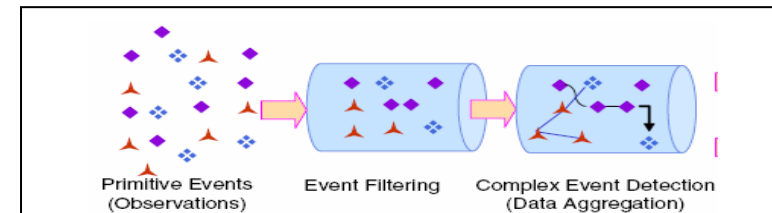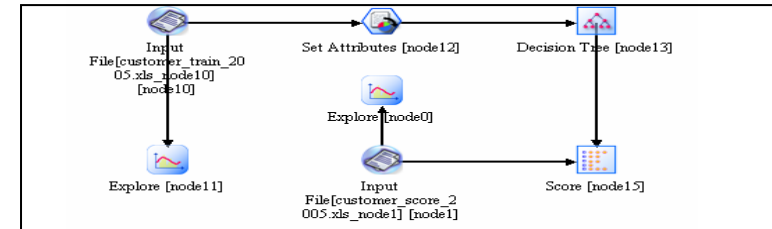
An Example

Path Event Database:

id	product	brand	path
1	tennis	nike	(f,10) (d,2) (t,1) (s,5) (c,0)
2	tennis	nike	(f,5) (d,2) (t,1) (s,10) (c,0)
3	sandals	nike	(f,10) (d,1) (t,2) (s,5) (c,0)
4	shirt	nike	(f,10) (t,1) (s,5) (c,0)
5	jacket	nike	(f,10) (t,2) (s,5) (c,1)
6	jacket	nike	(f,10) (t,1) (w,5)
7	tennis	adidas	(f,5) (d,2) (t,2) (s,20)
8	tennis	adidas	(f,5) (d,2) (t,3) (s,10) (d,5)

FlowGraph:

Duration Dist:
1: 0.2
2: 0.8
Duration Exceptions:
Given (f, 5) 1: 0.0
2: 1.0
Given (f, 10) 1: 0.5
2: 0.5

factory

dist. center   0.65   truck   1.00   shelf   1.00

0.20
shelf   0.60   checkout

0.20   dist. center

0.35   truck   0.67   shelf   1.00   checkout
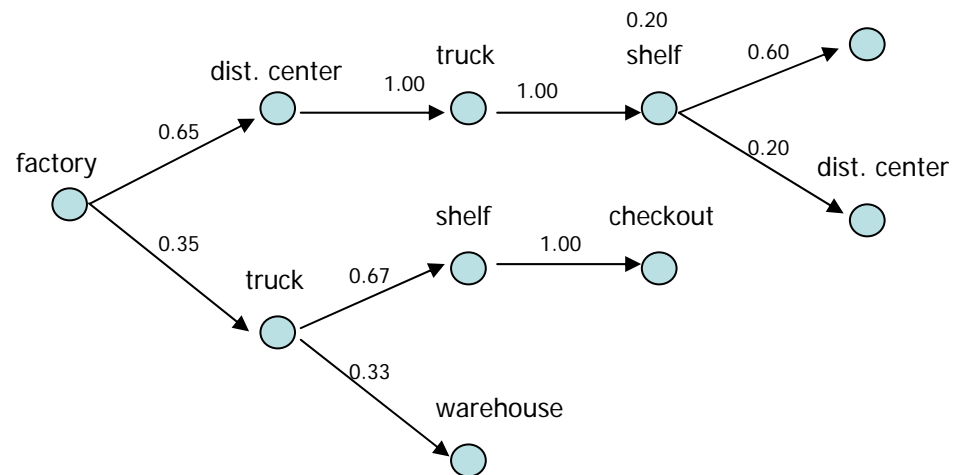
0.33   warehouse

Duration Dist:
1: 0.67
2: 0.33
Transition Dist:
shelf: 0.67
warehouse: 0.33

Transition Exceptions:
Given (t, 1) shelf: 0.5
warehouse: 0.5
Given (t, 2) shelf: 1.0
warehouse: 0.0

- Event analysis at each node
  - Distribution of durations at each node
  - Distribution of transition probabilities
  - Exceptions to duration and transition probabilities
- Path event analysis
  - Clustering and categorization
  - Classification and prediction
  - Outlier analysis

- Frequent patterns and sequential patterns can be related to event movement and paths

  – Correlation analysis of different events

  – Sequential event patterns
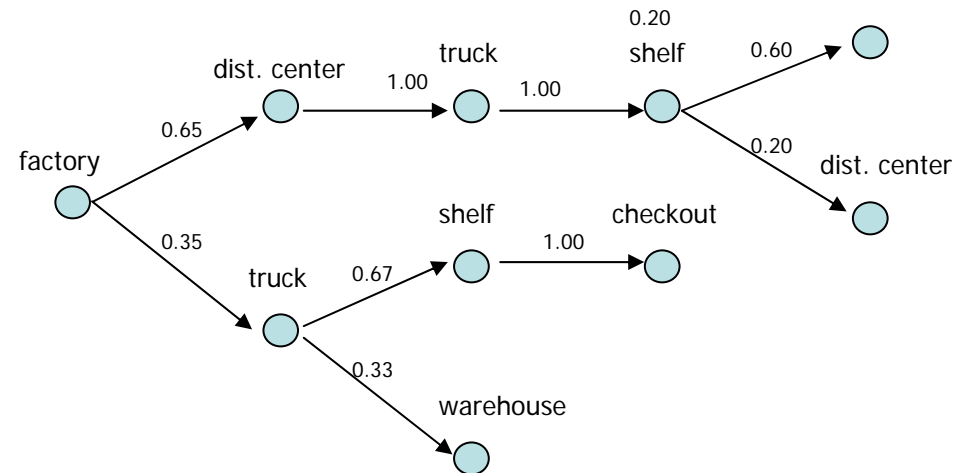
Outlier Event Analysis

- Outlier event detection (by-product of event mining)

  - Event flow analysis: Detect those not in the major flows

  - Classification: Treat outliers and normal events as different class labels

  - Cluster analysis: Identify those that are deviate substantially in major clusters

  - Trend analysis: Those not following the major trend

Agenda

- Overview of RFID Technology and Applications
- RFID in Logistics and Supply Chain Management
- Research Areas
    - **Data Collection and Processing**
    - **Data Management**
    - **Event Processing**
    - **Data Mining**
- Summary

Summary

- In general, RFID data mining is still a young, largely unexplored field, therefore, a lot of opportunities.

- RFID data mining has close links with sensor data mining, moving object data mining, work flow data mining and stream data mining (where one can find related work and techniques)

- RFID data mining research should be closely related to applications

Thank You

RFID Enabled Supply Chain

Automatic Verification and Inventory Reconciliation