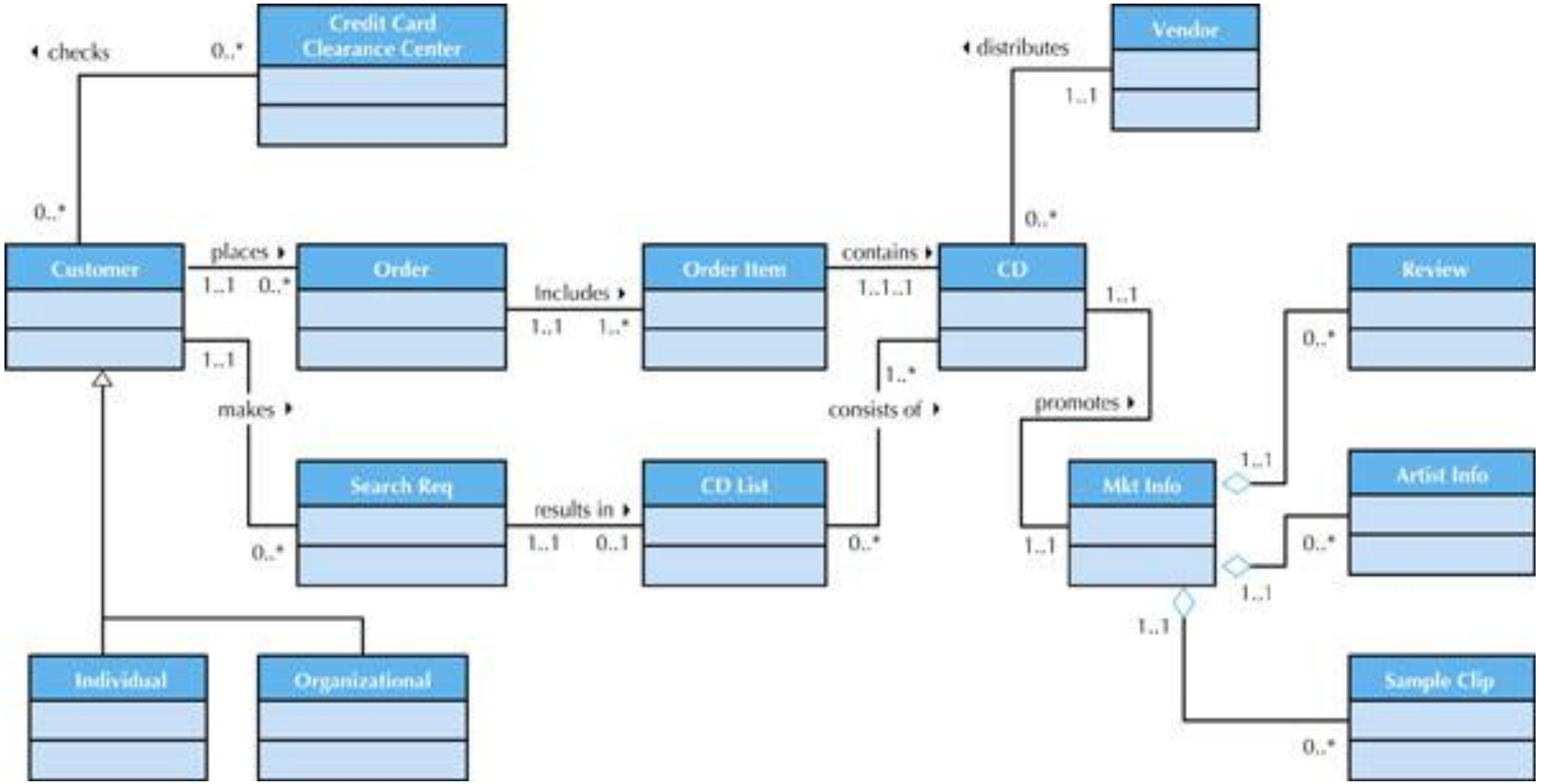
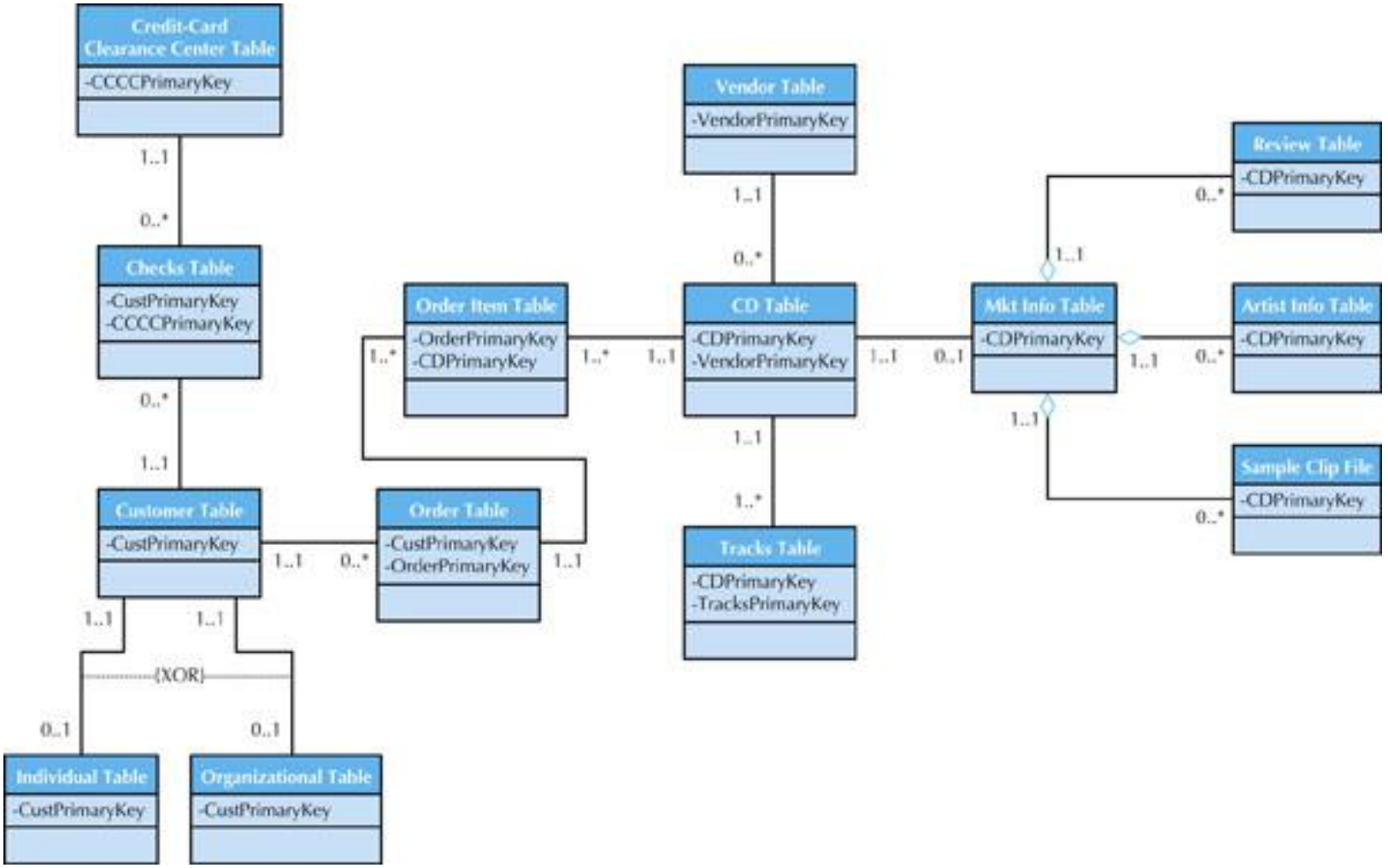
- Class diagram: the object model
- Relational DB diagram: the relational model
- E-R diagram: the projection of an object world onto the world of data

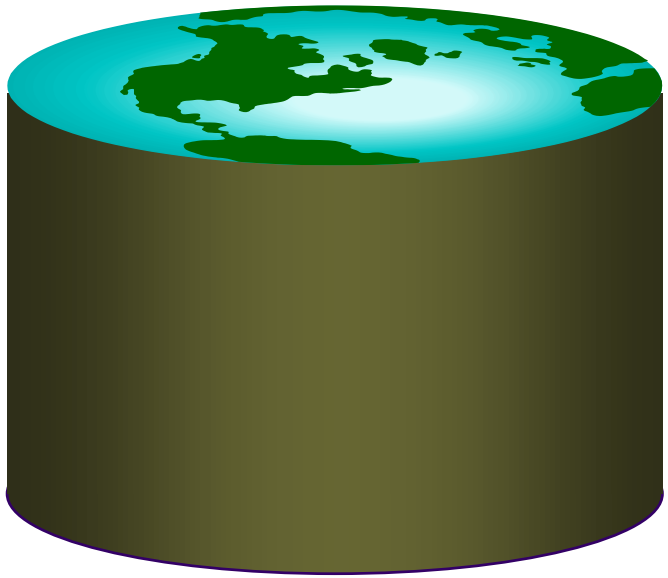
From Class diagram to RDBMS model




Credit-Card Clearance Center Table
-CCCCPrimaryKey
1..1
0..*
Checks Table
-CustPrimaryKey
-CCCCPrimaryKey
0..*
1..1
Customer Table
-CustPrimaryKey
1..1
0..*
Order Table
-CustPrimaryKey
-OrderPrimaryKey
1..1
1..1
1..1
{XOR}
0..1
0..1
Individual Table
-CustPrimaryKey
Organizational Table
-CustPrimaryKey
1..*
Order Item Table
-OrderPrimaryKey
-CDPrimaryKey
1..*
1..1
Vendor Table
-VendorPrimaryKey
1..1
0..*
CD Table
-CDPrimaryKey
-VendorPrimaryKey
1..1
1..1
1..*
Tracks Table
-CDPrimaryKey
-TracksPrimaryKey
1..1
0..1
Mkt Info Table
-CDPrimaryKey
1..1
0..*
Review Table
-CDPrimaryKey
1..1
0..*
Artist Info Table
-CDPrimaryKey
1..1
0..*
Sample Clip File
-CDPrimaryKey

A Quick Guide

Database Normalization



Topics

- Normalization – rid of abnormities
  - Business-level normalization
    - Initial design
    - 1st normal form
    - 2nd normal form
    - 3rd normal form
- Decomposition
- BCNF (Boyce Codd Normal Form)

Motivation

- How do we tell if a design is bad, e.g., *WorkOn*(*EID*, *Ename*, *PID*, *Pname*, *Hours*)?
  - This design has *redundancy*, because the name of an employee is recorded multiple times, once for each project the employee is taking.

EID	PID	Ename	Pname	Hours
1234	10	John Smith	B2B platform	10
1123	9	Tom Hanks	CRM	40
1234	9	John Smith	CRM	30
1023	10	Susan Mark	B2B platform	40

Why Redundancy is Undesirable?

- Waste disk space.
- What if we want to perform update operations to the relation
  - INSERT an new project that no employee has been assigned to it yet.
  - UPDATE the name of “John Smith” to “John L. Smith”
  - DELETE the last employee who works for a certain project

EID	PID	Ename	Pname	Hours
1234	10	John Smith	B2B platform	10
1123	9	Ben Liu	CRM	40
1234	9	John Smith	CRM	30
1023	10	Susan Sidhuk	B2B platform	40

DB Design Evaluation & Improvement

- Evaluate and improve Relation Models for design quality
  - **Purpose:** to achieve minimum redundancy in the data.
  - **Theory:** Normalization
  - **Main tool:** Functional Dependency
  - **Certification:** Normal Forms
  - **Technique:** Decompositions

Simply..

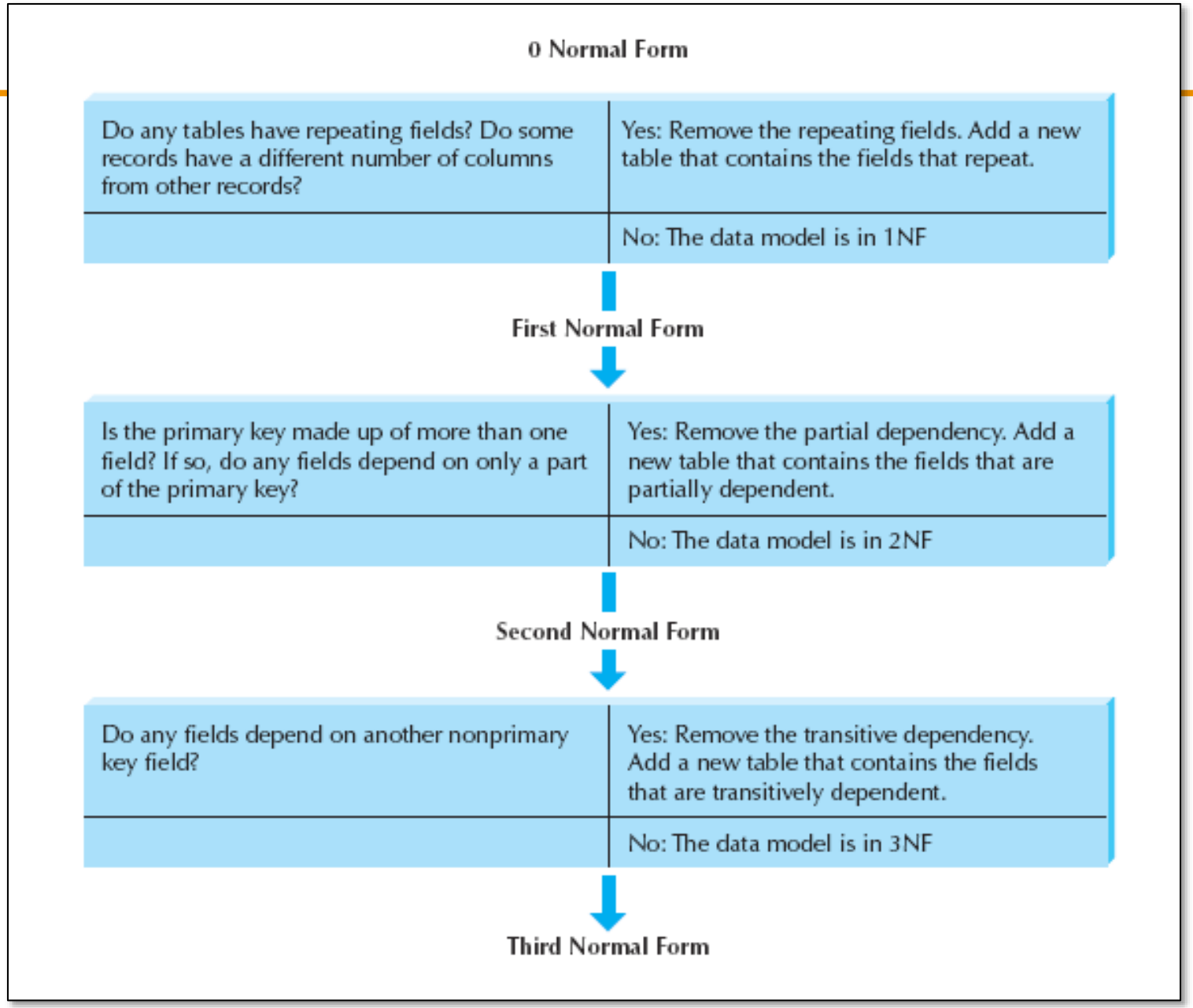
- By *normalization* theory based on *functional dependencies*, we force relational models with update anomalies, into an appropriate *normal form* using *decomposition* to produce smaller, well-structured relations, mostly to achieve *minimum redundancy* in the data.

- One Fat Relation → Multiple Smaller Relations

Normalization

- Codd (1972)
- A *normalization* is the process of decomposing unsatisfactory "bad" relations by breaking up their attributes into smaller relations

- A *normal form* is a certification that tells whether a relation schema is in a particular state


0 Normal Form
Do any tables have repeating fields? Do some records have a different number of columns from other records?
Yes: Remove the repeating fields. Add a new table that contains the fields that repeat.
No: The data model is in 1NF
First Normal Form
Is the primary key made up of more than one field? If so, do any fields depend on only a part of the primary key?
Yes: Remove the partial dependency. Add a new table that contains the fields that are partially dependent.
No: The data model is in 2NF
Second Normal Form
Do any fields depend on another nonprimary key field?
Yes: Remove the transitive dependency. Add a new table that contains the fields that are transitively dependent.
No: The data model is in 3NF
Third Normal Form

(1) 1st Normal Form

- Domain is *atomic* if its elements are considered to be indivisible units.
  - Examples of non-atomic domains
    - Set of names, **composite attributes**
    - Identification numbers like INFO421, INFO400 that can be broken up into parts
    - **Multivalued attributes**: Address, phone number
- A relational schema is in *first normal form* if the domains of all attributes are atomic.

- *Non-atomic values* complicate storage and encourage redundant (repeated) storage of data
  - E.g. Set of accounts stored with each customer, and set of owners stored with each account

Convert to 1st-NF

- Composite attribute: Address is split into several parts: a set of simple attributes.

- Multivalued attributes: move to a new table
- Example:
  R (EID, Ename, PhoneNum, email)

- Decompose R to two relations:
  - R1(EID, Ename, email)
  - R2(EID, phoneType, phoneNumber)

    - EID is FK in R2.
    - PK is all the attributes of R2.

2nd Normal Form

- An attribute A of a relation R is a ***nonprimary(nonkey) attribute*** if it is not part of any key in R, otherwise, A is a ***primary(key) attribute***.
- *Functional dependency:* Field A is *F.D. on* field B if the value of A depends on B. *Example: OrderDate is F.D. on OrderNumber.*

- A table/relation is in (general) 2nd normal form if all nonkey attributes (that are not part of the primary key) are functionally dependent on the ***entire*** primary key.
- No partial dependency

- WorkOn(*EID*, *Ename*, *email*, *PID*, *Pname*, *Hours*)

Normalize to 2nd NF

- Discover FDs
- Decompose and set up a new relation for *each partial key* with *its dependent attribute*(s).
- Make sure to keep a relation with *the original primary key* and *any attributes* that are *fully* functionally dependent on it.

Decomposition

EID	PID	Ename	email	Pname	Hours
1234	10	John Smith	jsmith@ac.com	B2B platform	10
1123	9	Ben Liu	bliu@ac.com	CRM	40
1234	9	John Smith	jsmith@ac.com	CRM	30
1023	10	Susan Sidhuk	ssidhuk@ac.com	B2B platform	40

Decomposition

Foreign key

EID	Ename	email
1234	John Smith	jsmith@ac.com
1123	Ben Liu	bliu@ac.com
1023	Susan Sidhuk	ssidhuk@ac.com

EID	PID	Pname	Hours
1234	10	B2B platform	10
1123	9	CRM	40
1234	9	CRM	30
1023	10	B2B platform	40

- Decomposition eliminates redundancy
- To get back to the original relation: Join

Decomposition

- Decomposition may be applied recursively

EID	PID	Pname	Hours
1234	10	B2B platform	10
1123	9	CRM	40
1234	9	CRM	30
1023	10	B2B platform	40

PID	Pname
10	B2B platform
9	CRM

EID	PID	Hours
1234	10	10
1123	9	40
1234	9	30
1023	10	40

3rd normal form

- Every nonkey attribute depends on the key, the whole key, and nothing but the key.

- `Customer(`**`cid`**`, cname, address,` **`salesRepNum`**`, salesRepName)`

- Split into 2 tables.

Example

- Initial design:

Student(sid, sname, schours, gpa, advid, advname, cid, cdescript, ccredit, grade)

- 1st normal form
- 2nd normal form
- 3rd normal form - BCNF

Exercise

	ItemNumber	EquipmentType	AcquisitionCost	RepairNumber	RepairDate	RepairCost
1	100	Drill Press	3500.00	2000	2011-05-05 ...	375.00
2	200	Lathe	4750.00	2100	2011-05-07 ...	255.00
3	100	Drill Press	3500.00	2200	2011-06-19 ...	178.00
4	300	Mill	27300.00	2300	2011-06-19 ...	1875.00
5	100	Drill Press	3500.00	2400	2011-07-05 ...	0.00
6	100	Drill Press	3500.00	2500	2011-08-17 ...	275.00

- EQUIPMENT_REPAIR (ItemNumber, Type, AcquisitionCost, RepairNumber, RepairDate, RepairAmount)