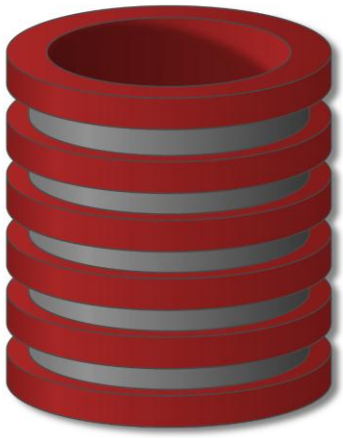


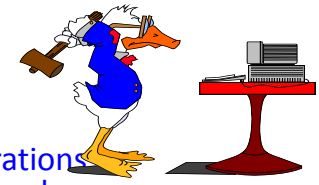
History of Databases

Imran Khan
FCS, IBA

Databases are Flower Children of the 60's

- Charles Bachmann developed the first DBMS called IDS while working at Honeywell
  - Network model where data relationships are represented as a graph
- First commercially successful DBMS developed at IBM called IMS
  - Hierarchical model where data relationships are represented as a tree
  - still in use today in IBM's and American Airline's SABRE reservation system
- Conference On DAta SYstems Languages(CODASYL) model defined
  - network model but more standardized

Problems with first DBMS'



- Access to database was through low level pointer operations
- Storage details depended on the type of data to be stored
- Adding a field to the DB required rewriting the underlying access/modification scheme
- Emphasis on records to be processed, not overall structure
- User had to know physical structure of the DB in order to query for information

Overall first DBMS' were very complex and inflexible which made life difficult when it came to adding new applications or reorganizing the data

Relational DB's to the rescue...

Edgar (Ted) Codd

- Father of The Relational Model
- Oxford-trained mathematician working for IBM @ San Jose Laboratory
- In 1970, Codd released "A Relational Model of Data for Large Shared Data Banks." This text first defined the Relational Model.
  - "It provides a means of describing data with its natural structure only--that is, without superimposing any additional structure for machine representation purposes. Accordingly, it provides a basis for a high level data language which will yield maximal independence between programs on the one hand and machine representation on the other."(Codd 1970)
- In other words the Relational Model consisted of:
  - Data independence from hardware and storage implementation
  - Automatic navigation, or a high level, nonprocedural language for accessing data. Instead of processing one record at a time, a programmer could use the language to specify single operations that would be performed across the entire data set.

Codd's 12(13) Rules

- 0. A relational DBMS must be able to manage databases entirely through its relational capabilities.
  1. Information rule-- All information in a relational database (including table and column names) is represented explicitly as values in tables.
  2. Guaranteed access--Every value in a relational database is guaranteed to be accessible by using a combination of the table name, primary key value, and column name.
  3. Systematic null value support--The DBMS provides systematic support for the treatment of null values (unknown or inapplicable data), distinct from default values, and independent of any domain.
  4. Active, online relational catalog--The description of the database and its contents is represented at the logical level as tables and can therefore be queried using the database language.
  5. Comprehensive data sublanguage--At least one supported language must have a well-defined syntax and be comprehensive. It must support data definition, manipulation, integrity rules, authorization, and transactions.
  6. View updating rule--All views that are theoretically updatable can be updated through the system.

Codd's 12(13) Rules

- 7. Set-level insertion, update, and deletion--The DBMS supports not only set-level retrievals but also set-level inserts, updates, and deletes.
  8. Physical data independence--Application programs and ad hoc programs are logically unaffected when physical access methods or storage structures are altered.
  9. Logical data independence--Application programs and ad hoc programs are logically unaffected, to the extent possible, when changes are made to the table structures.
  10. Integrity independence--The database language must be capable of defining integrity rules. They must be stored in the online catalog, and they cannot be bypassed.
  11. Distribution independence--Application programs and ad hoc requests are logically unaffected when data is first distributed or when it is redistributed.
  12. Nonsubversion rule--It must not be possible to bypass the integrity rules defined through the database language by using lower-level languages.

Codd vs. IBM

- Codd's model had an immediate impact on research, however, to become a legitimacy within the field, it had to survive at least two battles:
  - One in the technical community at large
  - One within IBM
- Within IBM
  - Conflict with existing product IMS which had been heavily invested into
  - New technology had to prove itself before replacing existing revenue producing product
  - Codd published his paper in open literature because no one at IBM (himself included) recognized its eventual impact
  - Outside technical community showed that the idea had great potential

Codd vs. IBM (Continued)

- Within IBM
  - IBM declared IMS its sole strategic product, setting up Codd and his ideas as counter to company goals
  - Codd speaks out in spite of IBM's dissatisfaction and promotes relational model to computer scientists. He arranges a public debate between himself and Charles Bachmann, who at the time was a key proponent of the CODASYL standard.
  - Debate produced further criticism from IBM for undermining its goals, but also proved his relational model as a cornerstone to the technical community.
- Finally, Two main relational prototypes emerge in the 70's
  - System R from IBM
  - Ingres from UC-Berkeley

System R

- Prototype intended to provide a high-level, nonnavigational, data-independent interface to many users simultaneously, with high integrity and robustness.
- Led to a query language called SEQUEL(Structured English Query Language) later renamed to Structured Query Language(SQL) for legal reasons. Now a standard for database access.
- Project finished with the conclusion that relational databases were a feasible commercial product
- Eventually evolved into SQL/DS which later became DB2

Ingres

- Two scientists, Michael Stonebraker and Eugene Wong at UC-Berkeley) became interested in relational databases
- Used QUEL as its query language
- Similar to System R, but based on different hardware and operating system
- Developers eventually branched off to form Ingres Corp, Sybase, MS SQL Server, Britton-Lee.

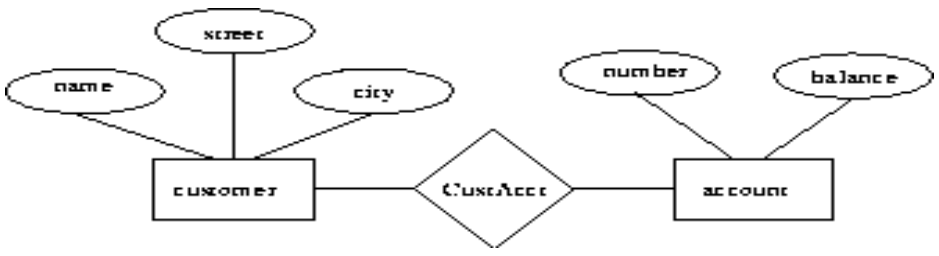
System R and Ingres inspire the development of virtually all commercial relational databases, including those from Sybase, Informix, Tandem, and even Microsoft's SQL Server

Where's Oracle!?

- Larry Ellison learned of IBM's work and founded Relational Software Inc. in 1977 in California
- Their first product was a relational database based off of IBM's System R model and SQL technology
- Released in 1979, it was the first commercial RDBMS, beating IBM to the market by 2 years.
- In the 1980's the company was renamed to Oracle Corporation and throughout the 80's new features were added and performance improved as the price of hardware came down and Oracle became the largest independent RDBMS vendor.

Entity-Relationship(ER) Models

- Proposed by Peter Chen in 1976 for database design giving an important insight into conceptual data models
- Allows the designer to concentrate on the use of data instead of the logical table structure



1980's

- Birth of IBM PC. RDBMS market begins to boom.
- SQL becomes standardized through ANSI (American National Standards Institute) and ISO (International Organization for Standardization)
- By Mid 80's it had become apparent that there were some fields(medicine, multimedia, physics) where relational databases were not practical, due to the types of data involved.
  - More flexibility was needed in how their data was represented and accessed.
- This led to research in Object Oriented Databases in which users could define their own methods of access to data and how to represent and manipulate it. This coincided with the introduction of Object Oriented Programming languages such as C++ which started to appear

1990's

- First OODBMS' start to appear from companies like Objectivity. Object Relational DBMS' hybrids also begin to appear.
- Industry shakeout begins with fewer surviving companies offering increasingly complex products at higher prices. Much of the development centers on client tools for application development such as: PowerBuilder(Sybase), Oracle Developer, Visual Basic, etc
- Development of personal/small business productivity tools such as Excel and Access from Microsoft.
- New application areas: Data warehousing and OLAP(Online Analytical Processing, a category of software tools that provides analysis of data stored in a database), internet, multimedia, etc

Late 90's-2000's

- Large investment in internet companies fuels tools-market boom for Web/Internet/DB connectors:
  - Active Server Pages, Front page, Java Servlets, JDBC, Java Beans, ColdFusion, Dream Weaver, Oracle Developer 2000, etc
- Open source projects come online with widespread use of gcc,cgi, Apache, MySQL
- Three main companies dominate in the large DB market: IBM, Microsoft, and Oracle

Sources

•INFS 614 - Section 02 -- Fall 03. Smith, Ken. Fall 2003. INFS 614 -- Section 02: Database Management. 14 Nov. 2004. <http://www.isse.gmu.edu/~kps/INFS614/>

•15 Seconds : Introduction to Relational Databases - Part 1: Theoretical Foundation. Tore Bostrup. 2004. Introduction to Relational Databases - Part 1: Theoretical Foundation. 14 Nov. 2004. <http://www.15seconds.com/Issue/020522.htm>

•NATIONAL ACADEMY PRESS. "Funding a Revolution: Government Support for Computing Research." The Rise of Relational Databases. 1999. 14 Nov. 2004. <http://www.nap.edu/readingroom/books/far/ch6.html>

•Macmillan Computer Publishing. "Teach yourself SQL in 21 Days."Day 1.2000. 14 Nov. 2004 <http://members.tripod.com/er4ebus/sql/ch01.htm>

•Oracle Tutorial - A Beginners Guide. 2002. Tutorial 1. 14 Nov. 2004.<http://www.smart-soft.co.uk/tutorial.htm>

Sources

•Marten Mickos. “Open Source Against Software Patents.” Aug 2004. 14 Nov 2004.<http://www.alwayson-network.com/comments.php?id=P5141_0_3_0_C>

•A Brief History of Databases. 2000. Founding the Future. 14 Nov. 2004. <http://wwwdb.web.cern.ch/wwwdb/aboutdbs/history/industry.html>

•Vaugh. “CPSC 343: A Sketch of Database History.” 2003. A Short Database History. 14 Nov. 2004.<http://math.hws.edu/vaughn/cpsc/343/2003/history.html>

Sources

- Thomas Connolly Book slides