Readings and References • Reading » Fluency with Information Technology Database Intro • Chapter 13, Introduction to Database Concepts • References » Access Database: Design and Programming INFO/CSE 100, Autumn 2004 • by Steve Roman, published by O'Reilly Fluency in Information Technology http://www.cs.washington.edu/100 22-Nov-2004 22-Nov-2004 cse100-20-databases © 2004 University of Washington cse100-20-databases © 2004 University of Washington 2 Why Study Databases? How to organize the data? • Some of us want to compute, but all of us want • Before relational databases (the kind we study) information ... there were only "flat files" • Much of the archived information is in tables » Structural information is difficult to express Databases enhance applications, e.g. Web » All processing of information is "special cased" Once you know how to create databases, you can use • custom programs are needed them to personal advantage » Information repeated; difficult to combine • Databases introduce interesting ideas » Changes in format of one file means all programs that ever process that file must be changed The Internet Movie Database Visited by over 28 million mayle loanes each month? Welcome to the Internet Movie Database, the biggest, best, most award-winning movie site • eg, adding ZIP codes on the plane.

22-Nov-2004

22-Nov-2004

cse100-20-databases © 2004 University of Washingtor

4

tab-delimited file example

	1100322-IL3RA-X	D001		
	1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0		IN	
	1100322-1L3RA-X	D002	G	
	1100322-1L3RA-X	D003	G	
	1100322-1L3RA-X	D004	G	
ownload of Variation Data (Single File)	1100322-1L3RA-X	D005	G	
ownoad of variation bata (Single File)	1100322-1L3RA-A	D000	G	
	1100322-1L3RA-A	D007	G	
ODIM PONTODINER FIRE	1100322-TL3RA-X	0000	۵ م	
is is a tab delivited test file in our "prettybase" format, which describes all \$49 sites	1100322-TL3RA-X	D010	N	
covered by the Seattleringer PGA. The format of this file is:	1100322-TL3RA-X	D011	N	
	1100322-IL3RA-X	D012	N	
ie format:	1100322-IL3RA-X	D013	G	
Involosome position-chromosomeret/do_bake > chroa sample tox calleretx	1100322-IL3RA-X	D014	А	
entropy of the second	1100322-IL3RA-X	D015	Ν	
ample: 74772592-00-PLAU 0001 6 T	1100322-IL3RA-X	D016	Ν	
	1100322-IL3RA-X	D033	A	
e "thromosome position" is generated from mapping to the most recent genome	1100322-IL3RA-X	D034	A	
sembly available from the UCSC Genome Assembly	1100322-IL3RA-X	D035	G	
	1100322-IL3RA-X	D036	A	
	1100322-IL3RA-X	D037	A	
	1100322-IL3RA-X	D038	G	
	1100322-IL3RA-X	D039	G	
	1100322-IL3RA-X	D040	G	

Unix termcap example

FILE FORMAT:

#
The version you are looking at may be in any of three formats: master
(terminfo with OT capabilities), stock terminfo, or termcap. You can tell
which by the format given in the header above.

The master format is accepted and generated by the terminfo tools in the # neurses suite; it differs from stock (System V-compatible) terminfo only # in that it admits a group of capabilities (prefixed `OT') equivalent to # various obsolete termcap capabilities.

ANSI capabilities are broken up into pieces, so that a terminal
implementing some ANSI subset can use many of them.
ansi+local1:\
 :do=\E[B:le=\E[D:nd=\E[C:up=\E[A:

ansi+local:\

:DO=\E[%dB:LE=\E[%dD:RI=\E[%dC:UP=\E[%dA:tc=ansi+local1: ansi+tabs:\

:bt=\E[Z:ct=\E[2g:st=\EH:ta=^I:

ansi+inittabs: $\$

:it#8:tc=ansi+tabs:

```
22-Nov-2004
```

cse100-20-databases © 2004 University of Washington

6

Library example

notice the redundancy

DBN	Title	AAID	AnName	And? he as	Public	Publisme	PubPhone	Price
- 111 - 111 - 1	0++	4	Former.	494 4 44 464 5	1	DigNoate	123-456-7590	428.93
0.09-399309-3	Enns	1.	Avetes	10-00	1	Big Hear	123-406-7890	\$28.00
0.91.335671.7	Fairle Queene	7	Spenzer	717-717-7171	1	Eig Erunt	123-456-7890	\$13.00
0.91-845678-5	Hamlet	5	Statespeare	553-355-3555	3	Alpha Press	000.008.0080	\$28.00
0-103-45678-8	lied	2	Rome	103-310-3103	1	Rig Roam	123-466-7890	\$21.00
0-12-345678-6	June Dyne	1	Aurteo	ID-DI-DU	1	Small House	71.4-008-0080	\$48.00
0.39-717117-7	King Lost	3	Statespoor	103-330-3303	1	Alpha Persu	205-518-3592	\$48.00
0.553.35353.8	Marbrib.	.5	Bulagear	353.355.3355	3	Alpha Press	999.998.9999	\$13.00
0.11-345678-8	Bdoty Dick	2	Metville	223-322-3232	3	Small House	71-6-008-0080	\$48.00
0-12-313463-3	On Likers	0	Milli	008-000-0000	1	Eig Ecan	123-456-7190	\$21.00
0-321-32153-1	Dalleon.	13	Steepy	321-321-111.1	3	Small House	71.4-008-0080	\$34.00
0.321-32153-1	Balloon.	11.	Sewogry	321-321-3323	2	Small House	71.4-008-0080	\$34.00
0.321.32132.5	Balleon.	12	Geampy	321.321.8080	3	(Imali House	71-6.008-0080	\$94.00
0.53-123456-8	Idais Street.	10	Jones	123-330-3103	3	Small House	71.4-008-0080	421.93
0-55-123456-8	Main Direct.	9	Smith.	123-122-1122	1	Small House	71.4-008-0080	421.93
0-123-40678-8	Ulyneer	6	Jogra	006-655-6585	1	Alpha Press	205-518-3092	\$34.00
1-22-233708-8	Virual Eusie	-4	Former.	444.444.4041	1	Eig Erust	123-456-7890	\$23.00

from Access Database book, Steve Roman

Relational Databases

- Information is stored in tables
 - » Tables store information about entities
 - » Entities have characteristics called attributes
 - » Each row in a table represents a single entity
 - Each row is a set of attribute values
 - Every row must be unique, identified by a key
 - » Relationships -- associations among the data values are stored

Table structure = schema Table contents = instance

22-Nov-2004

A Table in a Database

4							
sample : Eath	•						
U Last	First	JuND	Hire	Steet	Oly	State	Country
t Davalina	Mancy	0	6/1/1992	687 20th Ave E	Seattle	WA.	USA
2 Faller	Andrew	3	8/14/1992	908 W. Capital Way	Seattle	WA.	USA
3 Wacelar	Elerice .	1	4/1/1993	722 More Bay BMI	Seattle	1010	USA
4 Peacace 5 Bushapap	Showson .		10/17/1994	TR Granett Hill	Pastiano -	WHA.	15.4
5 Sullimani	Okan	2	12/12/1994	Caventry House	Seattle	WA	USA
0		0			-		
e 11 1	7 11	et et 2		Schema fo	r Exa	mpl	le table:
	1			ID	numb	er	unique number(Key)
	/			Last	text		person's last name
stance	e			First	text		person's first name
				JobCode	numb	er	current position
		/		Hire	date		first day on job
	sch	ema	L				



Redundancy in a database is Very Bad

- Not every assembly of tables is a good database
- Repeating data is a bad idea
 - » Replicated data can differ in its different locations,
 - e.g. multiple addresses can differ
 - Inconsistent data is worse than no data
 - » Keep a single copy of any data
 - if it is needed in multiple places, associate it with a key and store key rather than the data

Relationships between tables



11

"You can look it up"

- When looking for information, a single item might be the answer, but a table is more likely
 - » Which employees live in Kirkland?
 - Table of employees
 - » Who is taking INFO/CSE 100?
 - Table of students
 - » Whose mile run time $\leq 4:00$?
 - Table of runners

0 Nov 0004	and 00, 20, databases @ 2004 Linitersity of Merchinster	12
2-1101-2004	cse 100-20-databases @ 2004 Offiversity of Washington	15

Relational Algebra: Tables From Tables

- There are five basic "algebraic" operations on tables:
 - Select -- pick rows from a table
 - Project -- pick columns from a table
 - Union -- combine two tables w/like columns
 - Difference -- remove one table from another
 - Product -- create "all pairs" from two tables

From this basis, many more complicated operations can be built up

22-Nov-2004

cse100-20-databases © 2004 University of Washington

14

Select Operation

• Select creates a table from the rows of another table meeting a criterion

Select_from Example On Hire < 1993

	10	Last	Fint	JoblD	Hine	Sheet	City	State	Country
•	1	Davalino	Nancy	a	01-Mag-92	507 20th Ave E	Seattle	WA.	USA
	2	Fuller	Andrew	3	14-Aug-82	908 W. Capital Way	Seattle	WKA,	USA
	1	Wearder	Betan	1	01-Apr/93	722 Mass Bay Divi	Seattle	WA.	USA
	- 4	Peacock	Margaret	2	03-Map-83	411D Old Retmond Rd	Kirkland	VHA.	USA
	- 5	Buchasas	Steven	- 1	17-Dct-94	13 Garrett Hill	Seattle	WA.	USA
	- 6	Sulimani	Okan	2	12-Dec-94	Ceventry House	Seattle	WA.	USA
•	- 0			0					
1		14141	1.0		0				

First	Job40	Hire	Street	City	Crume.	1 martine
Mancu					104.00	Caractery
		01-bfay-92	507 20th Are E	Seattle	164	USA.
Andrew	3	14.Aug-92	908 W. Capital Way	Section	10,04	USA
	0					
	ACCESS.		D D D D D D D D D D D D D D D D D D D	andrew 3 Texapor do w. Calma way 0	Andrew 3 Texapol 200 W. Calma Way Calma 0 Texapol	C + furbal or 7

22-Nov-2004

15

Project

• Project creates a table from the columns of another table

Project Last, First From Example

	Permit								1				
	10	Last	Fint	Jubic	Hire	Street	City	State	Cos	white [
	10	araliso	Nancy	a	01-Mag-92	507 20th Ave E	Seattle	WA.	USA				
	2.5	uller	Andrew	3	14 Aug-82	908 W. Capital Way	Seattle	WKA.	USA				
	1 W	Accieber	Betan	1	01-Apr/93	722 Mass Day Divi	Seattle	WA.	USA				
	4.P	eecock.	Margaret	2	03-Map-83	411D Old Retmond Rd	Kirkland	WA.	USA				
	5.0	uchasas	Steven	3	17-Dct-94	13 Garrett Hill	Seattle	WA.	USA				
	6.0	diment	Okan	2	12-Dec-94	Ceventry House	Seattle	VHA.	Um	Description	and the second s	start Same	- itti xi
٠	0			0					P	Last.	L. Filmer		ALL N
10.0	and in	121	1.0	44 48 40	10					Lan	Ferrer		
	-								- 14	Daratino	Hancy		
									-	Fuller	ALBEN		
										Wegster	Detton		
										Peacock	Margoret		
										Buchanan	Steven		
										Sulfream	Okan		
									10			1	
ſ	'his i	s a proj	ection fr	om 9 c	limension	s to 2 dimensions				word with	,,		e t

Union

• Union combines two tables with same attributes All employees = perms UNION temps

Т	Ð	Last	Fint	Joble	Hine			Steet		City	State	Country				
•	10	Caraliso	Nanca	a	Q1-Mae	62 I	507.2	Oth Ave E	5	eattle	WA.	USA				
1	21	Fuller	Andrew	3	14-Aug	ari.	and to	1			ALC: N	1000				
1	1	Wooster	Betan	1	01-Ap	Ð	All co	найорски с О	inian Gue							10
	41	*escock	Margaret	2	03-Mag		10	Last	First	hield.	5	Hire	Storet	City	State	Count
	51	Buchasa	Steven	- 3	17-0c		- 1	Davaline	Nancy		D	5/1/1992	527 20th Ave E	Seattle	1624	USA.
]	8.3	Salimani	Oken	2	12-Dec		2	Fuller	Andrew		3	8/14/1992	908 W. Cepital	Seattle	WGL.	USA.
•	0			0			1	Wegetier	Detton		1	4/1/1990	722 Moss Bay I	Seattle	1824	USA.
	144 J	14 1 1 1	1 1	14 14 H	0		- 4	Peaceck	Margare	et in the	2	5(31993	4110 Old Redm	Kinkland	1804	USA.
	_						- 6	Buchasas	Steven		3	10/17/1994	13 Garett Hill	Seattle	16.6	USA.
		e · Table					8	Suliment	Okan		2	12/12/1994	Coventry Hoese	Seattle	1604	USA.
5	-	1.000	Gest	A SHARE	illine.		101	Soggy	Peter		D	6/1/2004	1300 20th Ave \	Seattle	1604	USA
ł	424	E.mail	Colur	- 0	Oil, hu		102	Motkan	Tavier		3	9/14/2003	100 Eartfake D	Seattle	162	USA
+		anogog	P-RW Verter	- 10	14 8 44		100	Wilshire	Bruce		1	31/1998	34 15th Are NE	Seattle	1604	USA.
+	100	WHEN PARTY	Care of		14-049 01.54v		104	Brazely	Terrys		2	3/3/2802	103 25th Ave N	Seattle	162	USA.
+	10.0	Program.	Terms	- 1	CTP Max		105	Comption	Sarah		3	11/17/1999	4034 NW 68th 1	Seattle	1804	USA
1	100	Constant	Carab		17.Alm		106	Iway	Ovid		2	1/12/1998	4502 NW 52nd	Seattle	16,24	USA
1	100	Zamera	0.4	- 2	17.144	E.	und .	10 10 1	1 10	101	# 42					
	- 0	1	2000	0	145388	-			-	-						
-		and a difference		and and and		-			_	_						
1	ан ,		100	1000	2											

Difference

- Difference (written like subtraction) removes 1 table's rows from another
 - Eastern = States WestCoast

States : Table			WestCoast : Tak	40		
Name	Capitol	Sight	Name	Capitol	Sight	
Washington	Olympia	Mt. Rainier	Washington	Olympia	Mt. Rainiar	
Oregon	Salem	Crater Lake	Oregon	Salem	Crater Lake	
California	Secremento	Golden Gate	California	Secremento	Golden Gab	
Arizona	Phoenix	Grand Canyon				
Nevada	Carson City	Las Vegas				
	E	astern : Table				
	Ĩ	Name	Capitol	Sig	jht	
	[Arizona	Phoenix	Grand Ca	anyon	
		Nevada	Carson City	Las Vega	is.	
		100 20 databasas @	2004 Liniversity of V	Vachington		

Product

• Product (written like multiplication) combines columns and pairs all rows

 $Colors = Blues \mathbf{x} Reds$



Join

• Join (written like a bow tie) combines rows if common field matches

Employee List = Perms ►⊲ JobCodes

	-	e: Table							- 20	121			
	Ð	Last	Fint	dideL	Hire	Street	City	State	Cose	All Y			
	1	Davalino	Nancy	0	01-Mag-92	507 20th Ave E	Seattle	WA.	USA				
	2	Fuller	Andrew	3	14 Aug-82	908 W. Capital Way	Seattle	V8.8.	USA				
	1	Weaster	Betan	1	01-Apr/93	722 Mass Day Divi	Seattle	WA.	USA				
	- 4	Peecock	Margaret	2	03Map-83	411D Old Retmond Rd	Kirkland	AWV b	USA				
	- 6	Buchasas	Steven	- 1	17-0ct-94	13 Garrett Hill	Seattle	WA.	USA				
	- 6	Sultream	Okan	2	12-Dec-94	Ceventry House	Seattle	VHA.	USA				
٠	- 0			a				-	-	_			l mi
F.h	****	14 1 1	1 1	1	10			1000.000					
								ID	Last	First	Jobi0	Title	Paycode
								De	aint	Nancy	D	CEO	
	. Aut	Codes : Ta	***		- 1	3		-3(W)	wateo	Derton	1	YP.	
=	-	Laber D.	1 10	4	Barrad			-4 Pe	acack	Marganet	2	Engineer	
	-	1000	0.000			-		6 Su	limani	0kan -	- 2	Engineer	
-	+		1.020					2.54	ler	Andrew:	3	Administrative	
H	÷		O Ecolomy					6 Bu	shanan	Stoven	3	Administrative	
	+		2 Crapter										
1	+		a womani	2,3,74		0					Indexts		
f			0	_		8	1.00	11	Long I	1.4	100		
A		1 - 1 - 1	5	1 1 1 1 1 1	at di								

22-Nov-2004

DB Operations

- The five DB Operations can create any table from a given set of tables
 - All modern database systems are built on these relational operations
 - Join is not primitive, but can be built from 5
 - Join, select and project are used most often
 - The operations are not usually used directly, but are used indirectly from other languages

SQL, the DB language we learn, is built on basic 5

cse100-20-databases © 2004 University of Washington

21