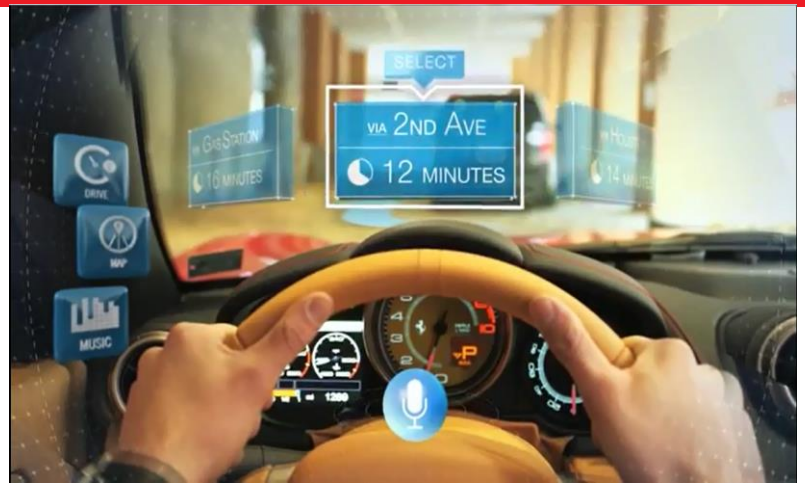
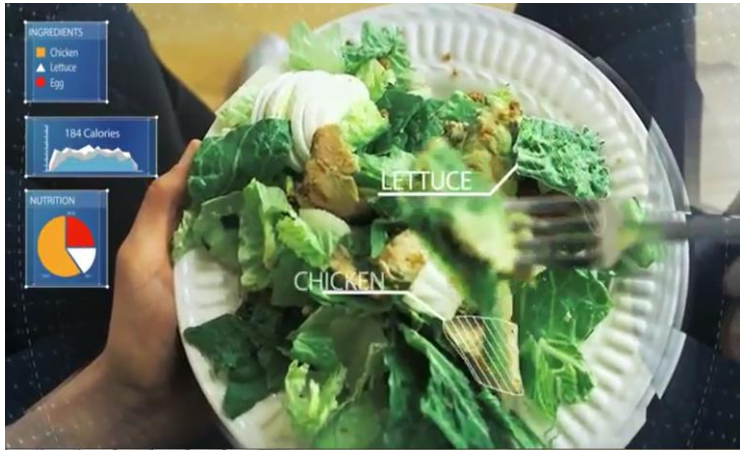
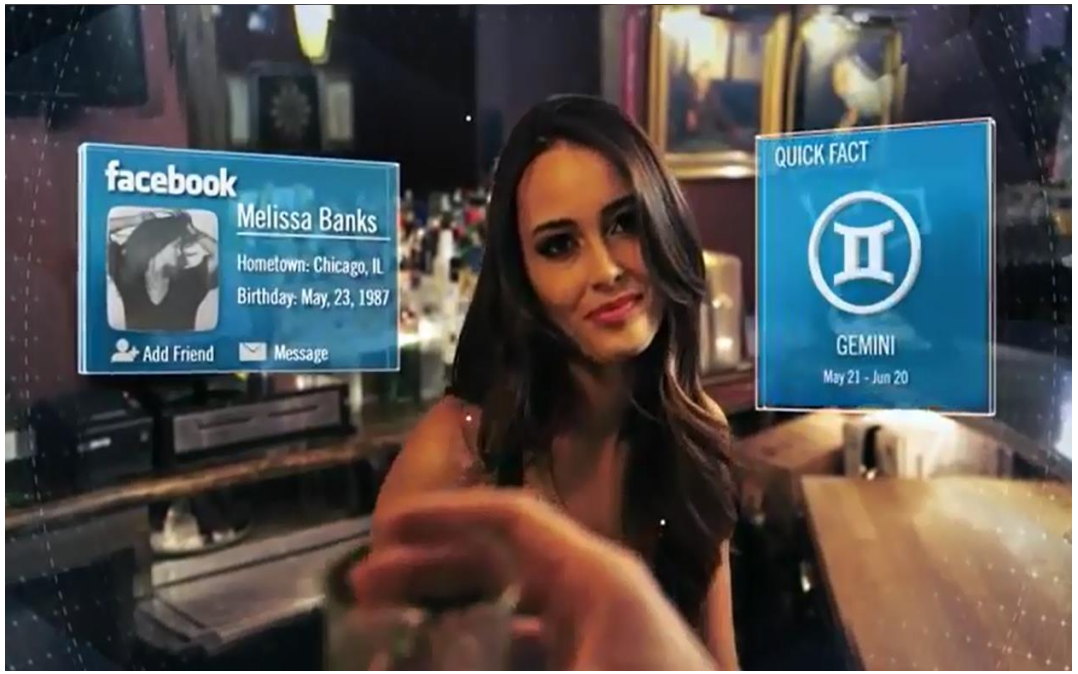
Data management

The Importance of Effective Data Management

Griffith UNIVERSITY







- Read/Write speed of HDD can range: **80 – 160 MB/s**
- A typical SSD R/W speed: **200 MB/s to 550 MB/s**

The Importance of Effective Data Management

- So, just storing of data is NOT a major issue today –
  - In 2008, $600 disk drive could store all world's music!

- Effective management of data is a major issue that involves
  - organise,
  - store,
  - Retrieval & process,
  - & update.

What method to use?

Let's do a simple Exercise

- Read the scenario below and then answer the question.
  - *A local video shop manually processes video loans and returns, and keeps their records in paper format. The owners of the shop are considering computerising their business operation and record keeping.*

How to store: Option 1

- The video shop can store as an One-dimensional text data

1. *David Low has rented "Avatar" on 23/8/22 at 1pm.*
2. *Patrick Vidgen has rented "Titanic" on 19/8/22 at 2pm.*
3. *John Campbell has rented "Rebirth" on 1/8/22 at 4pm.*
4. *…*
5. *…*

…

101. Patrick Vidgen returned "Titanic" on 17/9/22 at 10am.

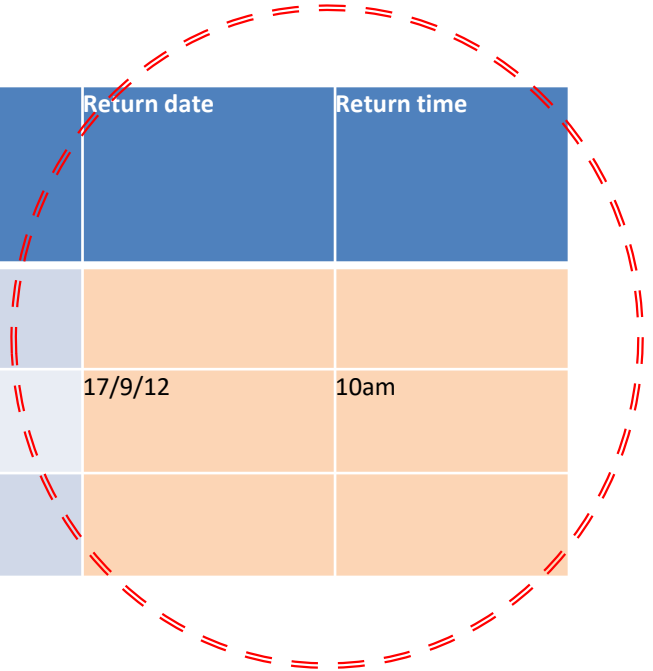
How to store: Option 2

- The video shop can also store as a Two-dimensional table

Customer First Name	Customer Last Name	DVD Title	Borrowing Date	Borrowing Time
David	Low	Avatar	23/8/22	1pm
Patrick	Vidgen	Titanic	19/8/22	2pm
John	Campbell	Rebirth	1/8/22	4pm

Which Option is More Rational and Why?

Customer First Name	Customer Last Name	DVD Title	Borrowing Date	Borrowing Time	Return date	Return time
David	Low	Avatar	23/8/12	1pm		
Patrick	Vidgen	Titanic	19/8/12	2pm	17/9/12	10am
John	Campbell	Rebirth	1/8/12	4pm		



- Answer: Option 2 is better
- Can be used **Spreadsheet** (MS Excel) to record & store

Data Redundancy, Inconsistency & Security

- **Redundancy:** duplication of data or storing the same information in multiple places.

CUSTOMER FILE

C_NAME	C_PHONE	C_ADDRESS	C_ZIP	A_NAME	A_PHONE	TP	AMT	REN
Alfred A. Ramas	615-844-2573	218 Fork Rd., Babs, TN	36123	Leah F. Hahn	615-882-1244	T1	100.00	05-Apr-2016
Leona K. Dunne	713-894-1238	Box 12A, Fox, KY	25246	Alex B. Alby	713-228-1249	T1	250.00	16-Jun-2016
Kathy W. Smith	615-894-2285	125 Oak Ln, Babs, TN	36123	Leah F. Hahn	615-882-2144	S2	150.00	29-Jan-2017
Paul F. Olowski	615-894-2180	217 Lee Ln., Babs, TN	36123	Leah F. Hahn	615-882-1244	S1	300.00	14-Oct-2016
Myron Orlando	615-222-1672	Box 111, New, TN	36155	Alex B. Alby	713-228-1249	T1	100.00	28-Dec-2016
Amy B. O'Brian	713-442-3381	387 Troll Dr., Fox, KY	25246	John T. Okon	615-123-5589	T2	850.00	22-Sep-2016
James G. Brown	615-297-1228	21 Tye Rd., Nash, TN	37118	Leah F. Hahn	615-882-1244	S1	120.00	25-Mar-2017
George Williams	615-290-2556	155 Maple, Nash, TN	37119	John T. Okon	615-123-5589	S1	250.00	17-Jul-2016
Anne G. Farriss	713-382-7185	2119 Elm, Crew, KY	25432	Alex B. Alby	713-228-1249	T2	100.00	03-Dec-2016
Olette K. Smith	615-297-3809	2782 Main, Nash, TN	37118	John T. Okon	615-123-5589	S2	500.00	14-Mar-2017

- **Inconsistency:** one of the primary problems with redundant information. In spread sheet, update to data could use a search/replace; may not work if people are not consistent with data entry. Example: Mohammad or Muhammad or Mohammed or Mohamed? Which one is used? Typos may also happen!

- **Security** – an important consideration; who should have access to what information? Who should be able to make changes to data? At the moment the security is limited to what is available within the spreadsheet software. May be able to put a password on the file, or lock certain fields from being able to be changes. However, there are limitations to what can be implemented.

Data Redundancy, Inconsistency & Security

- **Anomalies** are introduced into the data when inserting new data, updating old data or deleting data.
  - Insertion anomaly – e.g., addition of a new agent to customer table requires dummy customer entry into the table.
  - Deletion anomaly – e.g., deletion of customer data requires deletion of agent data too.
  - Update anomaly – e.g., updating an agent's phone number requires update in more than one entries.

Update anomaly!

CUSTOMER FILE

C_NAME	C_PHONE	C_ADDRESS	C_ZIP	A_NAME	A_PHONE	TP	AMT	REN
Alfred A. Ramas	615-844-2573	218 Fork Rd., Babs, TN	36123	Leah F. Hahn	615-882-1244	T1	100.00	05-Apr-2016
Leona K. Dunne	713-894-1238	Box 12A, Fox, KY	25246	Alex B. Alby	713-228-1249	T1	250.00	16-Jun-2016
Kathy W. Smith	615-894-2285	125 Oak Ln, Babs, TN	36123	Leah F. Hahn	615-882-2144	S2	150.00	29-Jan-2017
Paul F. Olowski	615-894-2180	217 Lee Ln., Babs, TN	36123	Leah F. Hahn	615-882-1244	S1	300.00	14-Oct-2016
Myron Orlando	615-222-1672	Box 111, New, TN	36155	Alex B. Alby	713-228-1249	T1	100.00	28-Dec-2016
Amy B. O'Brian	713-442-3381	387 Troll Dr., Fox, KY	25246	John T. Okon	615-123-5589	T2	850.00	22-Sep-2016
James G. Brown	615-297-1228	21 Tye Rd., Nash, TN	37118	Leah F. Hahn	615-882-1244	S1	120.00	25-Mar-2017
George Williams	615-290-2556	155 Maple, Nash, TN	37119	John T. Okon	615-123-5589	S1	250.00	17-Jul-2016
Anne G. Farriss	713-382-7185	2119 Elm, Crew, KY	25432	Alex B. Alby	713-228-1249	T2	100.00	03-Dec-2016
Olette K. Smith	615-297-3809	2782 Main, Nash, TN	37118	John T. Okon	615-123-5589	S2	500.00	14-Mar-2017
NULL	NULL	NULL	NULL	Mohammad	Ex55032	NULL	NULL	NULL

Deletion anomaly!

Insertion anomaly!

- **Solution**:
  - What is the solution?
  - Ans: use a separate Agent file and Customer file only have agent IDs to refer to Agent table!

AGENT FILE

A_NAME	A_PHONE	A_ADDRESS	ZIP	HIRED	YTD_PAY	YTD_FIT	YTD_FICA	YTD_SLS	DEP
Alex B. Alby	713-228-1249	123 Toll, Nash, TN	37119	01-Nov-2000	26566.24	6641.56	2125.30	132737.75	3
Leah F. Hahn	615-882-1244	334 Main, Fox, KY	25246	23-May-1986	32213.78	8053.44	2577.10	138967.35	0
John T. Okon	615-123-5589	452 Elm, New, TN	36155	15-Jun-2005	23198.29	5799.57	1855.86	127093.45	2

Thank you