Tutorial 7 — Data Mining and Transactions

Richard Wong

`rk2wong@edu.uwaterloo.ca`

Department of Electrical and Computer Engineering
University of Waterloo

March 11, 2018

What is a data warehouse, and why is it useful to have one?

Suppose we are trying to predict the value *Wait*.
Between attributes *Pat* and *Type*, which is better to split a decision node on?

Example	Attributes										Target
	Alt	Bar	Fri	Hun	Pat	$Price$	$Rain$	Res	$Type$	Est	$Wait$
X_1	T	F	F	T	Some	\$\$\$	F	T	French	0–10	T
X_2	T	F	F	T	Full	\$	F	F	Thai	30–60	F
X_3	F	T	F	F	Some	\$	F	F	Burger	0–10	T
X_4	T	F	T	T	Full	\$	F	F	Thai	10–30	T
X_5	T	F	T	F	Full	\$\$\$	F	T	French	>60	F
X_6	F	T	F	T	Some	\$\$	T	T	Italian	0–10	T
X_7	F	T	F	F	None	\$	T	F	Burger	0–10	F
X_8	F	F	F	T	Some	\$\$	T	T	Thai	0–10	T
X_9	F	T	T	F	Full	\$	T	F	Burger	>60	F
X_{10}	T	T	T	T	Full	\$\$\$	F	T	Italian	10–30	F
X_{11}	F	F	F	F	None	\$	F	F	Thai	0–10	F
X_{12}	T	T	T	T	Full	\$	F	F	Burger	30–60	T

What are the ACID transaction properties, and what can a database do to ensure each one?

Distinguish between the following:

1. a serial schedule
2. a serializable schedule
3. a conflict-serializable schedule

Is the following schedule conflict-serializable?
If not, what can the database transaction manager do to make the schedule conflict-serializable?

T1	r(x)		w(y)		
T2		r(y)			r(x)
T3				w(x)	r(x)