Wednesday 10 February 2016

So you want to send 100GB of data?

T. Charles Yun

eResearchNZ 2016

Queenstown, NZ

So you want to move a BIG data set

- What is “big”
  - Anything that is too big to send as an email attachment
  - Why not just mail a your hard drive?

- The network has changed the way people (scientists, corporate groups, individuals) interact with data
  - The “competition” is already taking advantage of the network
  - Additional funding, reduced costs, improved process, ease-of-use

- This will NOT be a technical talk (xref Ian, no lines of code) (upside: bug free) [and as it turns out, not quite true..., see corrected slide 11]

How "we" think of the network

- Line type (fiber, DSL)
- Line capacity (Gb/s)
- Packet size (jumbo packets, large MTU)
- Congestion (tcp/ip, dropped packets, packet loss)
- Host tuning (kernel, various i/o)
- Application tuning (data staging pipeline, database tuning)
- etc., etc., etc.

The Network
Congestion

https://commons.wikimedia.org/wiki/File:Motorcyclists_lane_splitting_in_Bangkok,_Thailand.jpg

Lies, Damn Lies and Statistics…

Fallacy of the station wagon

Never underestimate the bandwidth of a station wagon full of tapes hurtling down the highway.

—Tanenbaum, Andrew S. (1989). Computer Networks. New Jersey: Prentice-Hall. p. 57. ISBN 0-13-166836-6. (taken from Wikipedia)

Lies, Damn Lies and Statistics…

Imagine this scenario…

- Let's say you regularly move data between Auckland and Wellington.
  - Distance AKL to WLG: 641 km
  - Average drive speed: 80km/h

map: Google Maps

Lies, Damn Lies and Statistics…

Mazda MX6 Wagon, 2013-2014

- Mazda6 Station Wagon
- Cargo Space: ~500 Liters

http://www.drive.com.au/itcare/wagons-v-suv-comparison-test-mazda6-v-mazda-cx5-hyundai-i30-tourer-v-hyundai-ix35-holden-commodore-sportwagon-v-holden-captiva7-20140902-10eked403-litres

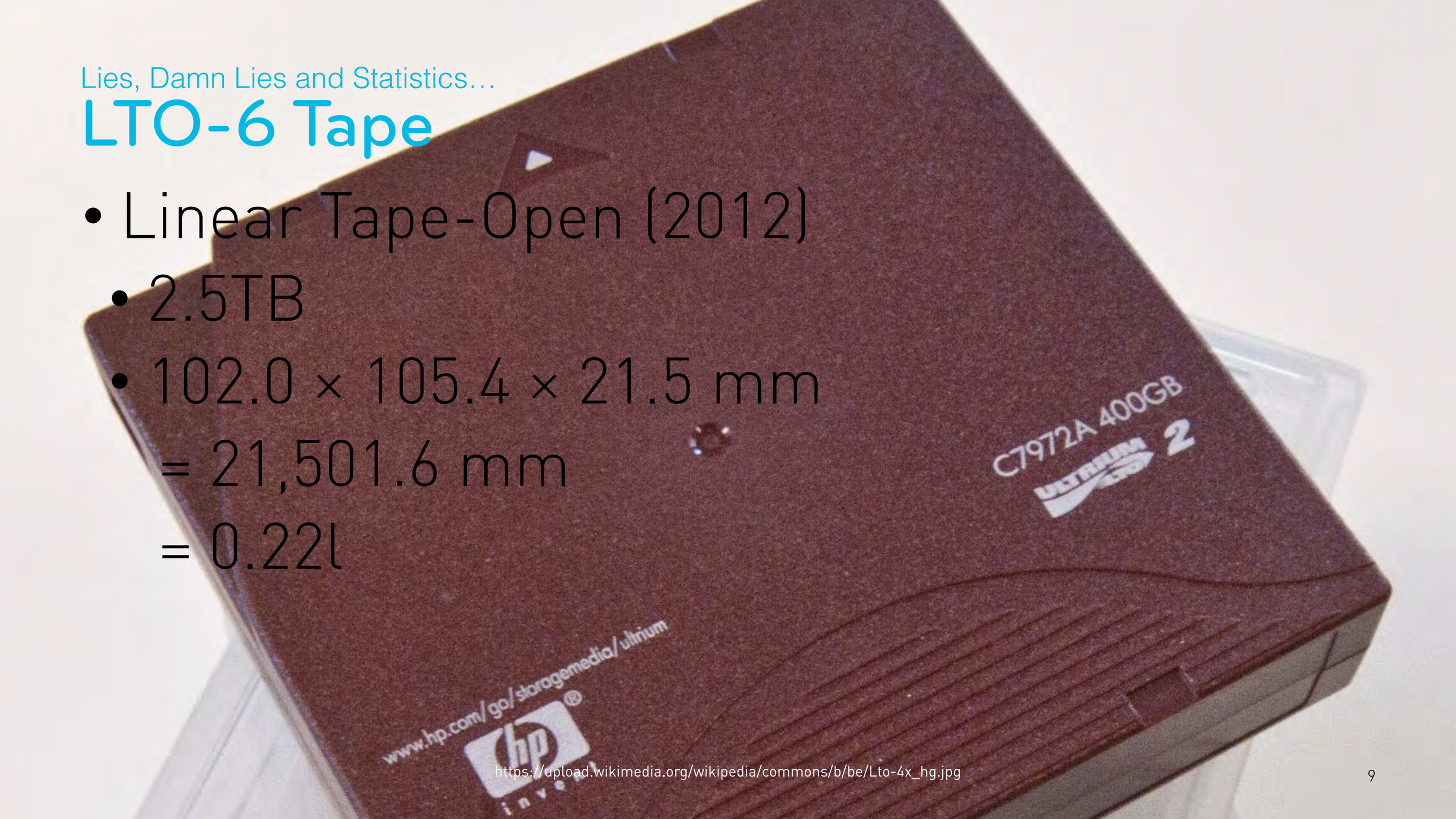
http://www.carshowroom.com.au/reviews/2012-mazda6-wagon-touring-review-and-road-test/519-litres

https://en.wikipedia.org/wiki/File:Japanese_car_accident_blur.jpg

Lies, Damn Lies and Statistics…

LTO-6 Tape

- Linear Tape-Open (2012)
  - 2.5TB
  - 102.0 × 105.4 × 21.5 mm
    = 21,501.6 mm
    = 0.22l



https://upload.wikimedia.org/wikipedia/commons/b/be/Lto-4x_hg.jpg

Lies, Damn Lies and Statistics…

Carrying Capacity

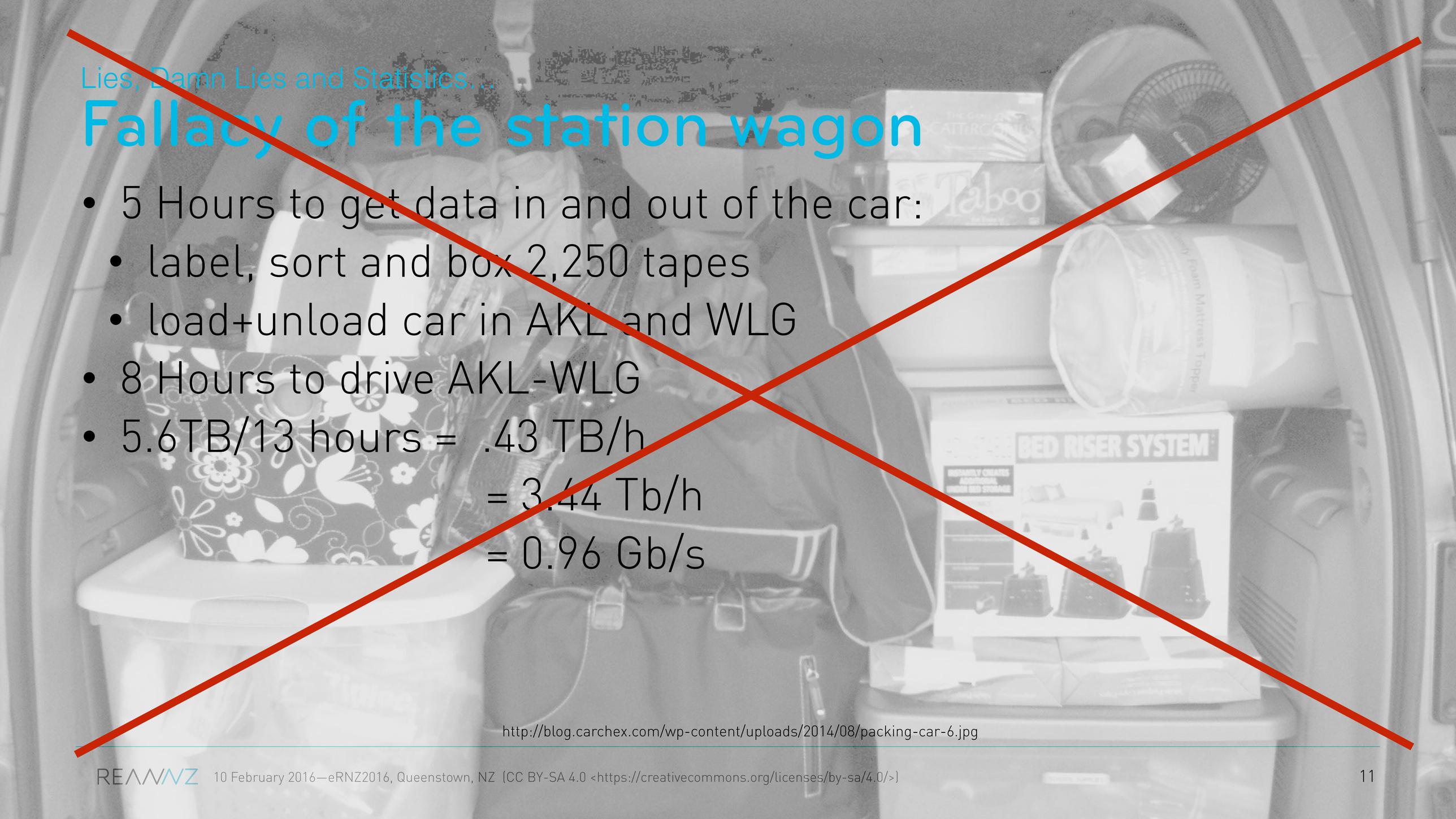
- Cargo Space: 500 Liters
- Single Tape Capacity: 2.5TB
- Single Tape Displacement:
  - 102.0 × 105.4 × 21.5 mm = 21,501.6 mm ~= 0.22l
- Tapes in Cargo:
  - 500/.22 = 2,272 ~= 2,250
- Total Data in Cargo:
  - 2,250 * 2.5TB = 5,625TB

ungraciously stolen from: http://www.wallpaperno.com/Humor/funny/minimalistic_funny_swallow_coconut_monty_python_and_the_holy_grail_1600x900_wallpaper_42922/download_1920x1080

Lies, Damn Lies and Statistics...

Fallacy of the station wagon

- 5 Hours to get data in and out of the car:
  - label, sort and box 2,250 tapes
  - load+unload car in AKL and WLG
- 8 Hours to drive AKL-WLG
- 5.6TB/13 hours = .43 TB/h
  = 3.44 Tb/h
  = 0.96 Gb/s



http://blog.carchex.com/wp-content/uploads/2014/08/packing-car-6.jpg

Fallacy of the station wagon

- 5 Hours to get data in and out of the car:
  - label, sort and box 2,250 tapes
  - load+unload car in AKL and WLG
- 8 Hours to drive AKL-WLG
- 5.6TB/13 hours = .43 TB/h
  = 3.44 Tb/h
  = 0.96 Gb/s

5.6TB? derp, that was **5,600**TB. Apologies for getting the math wrong... And belated thanks to the audience for kindly pointing out the mistake

http://blog.carchex.com/wp-content/uploads/2014/08/packing-car-6.jpg

Lies, Damn Lies and Statistics…

Fallacies: corrected, expanded, justified*

- Write data to and from all tapes (or, buying back 3 orders of magnitude error…):
  - write, label, box, read—total 1 hour
  - 2,250 tapes * 1 hours/tape = 2,250 hours
- 5 Hours to get data in and out of the car
- 8 Hours to drive AKL-WLG
- total time: 2250 + 5 + 8 = 2250 hours

- 5,600TB/2250 hours ~ 2.5 TB/h = 20 Tb/h = 5.5 Gb/s

* hopefully without errors this time around…

Lies, Damn Lies and Statistics...

Packet Loss

And remember, packet loss in the estate wagon scenario is a pretty big deal

https://en.wikipedia.org/wiki/File:Japanese_car_accident_blur.jpg

Lies, Damn Lies and Statistics…

Are you happy with "good enough"

- If you could get 10x improvement in the precision of your scientific equipment by "reading the manual", would you follow up?
  - If you could stream data continuously, would you even worry about storing files **and then** moving them?

- 1 Gb/s sounds nice
- You should be seeing 10 Gb/s
- We are planning for 100Gb/s

- Everything you need to do better is already in place

