Distributed Database Timing Requirements - Translating Transactions per Second to Clock Accuracies

A Leading Provider of Smart, Connected and Secure Embedded Control Solutions

David Chandler

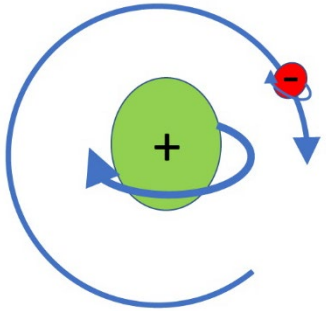
March 2023

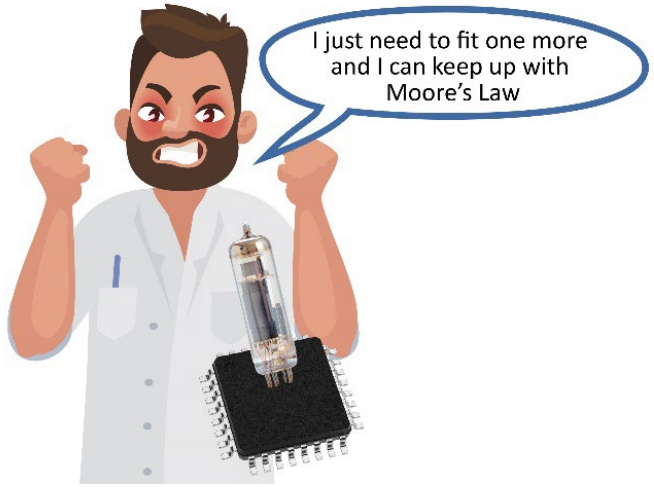
Agenda

- The need for time with distributed data
- Typical time dissemination approaches
- Distributed database examples
- Translating transactions per second to clocking requirements

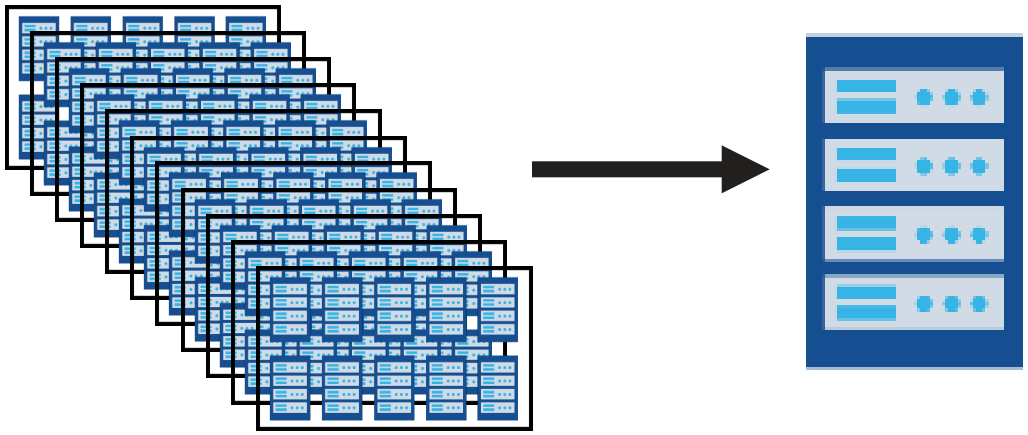
Why Distributed Databases?

- **Global computing power needed for zettabytes (10^21)/year**
  - Atom repeals Moore's Law
    - Machine speed limited by pitch (fab process)





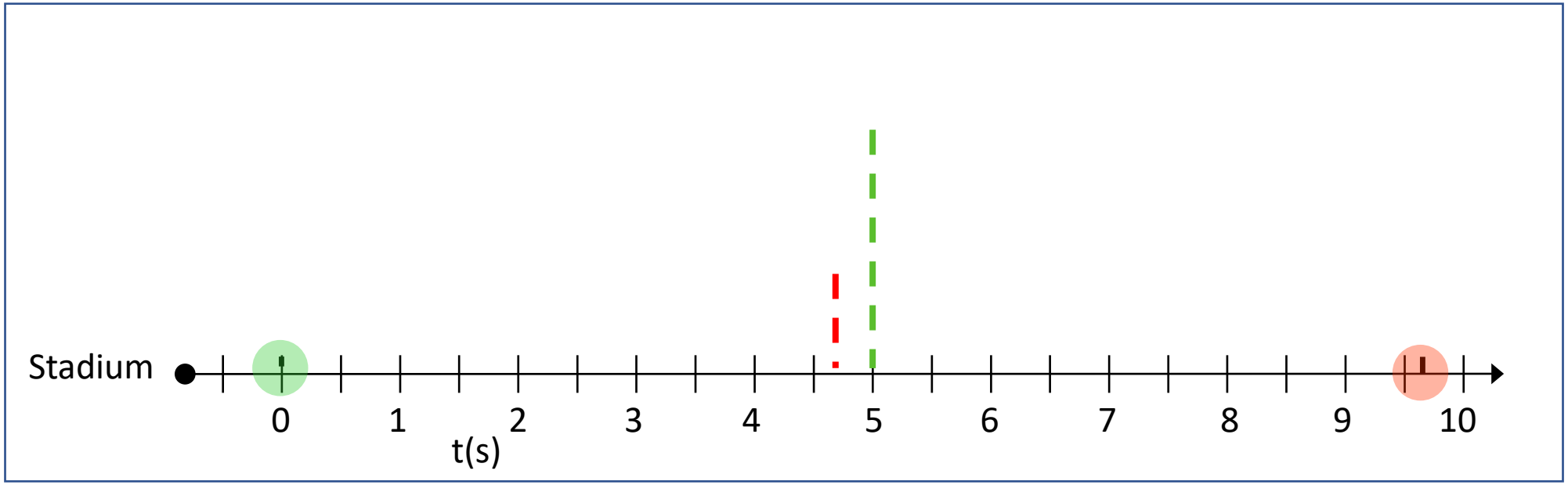
  - Horizontal scaling



MICROCHIP

Time Challenges with Multiple Machines

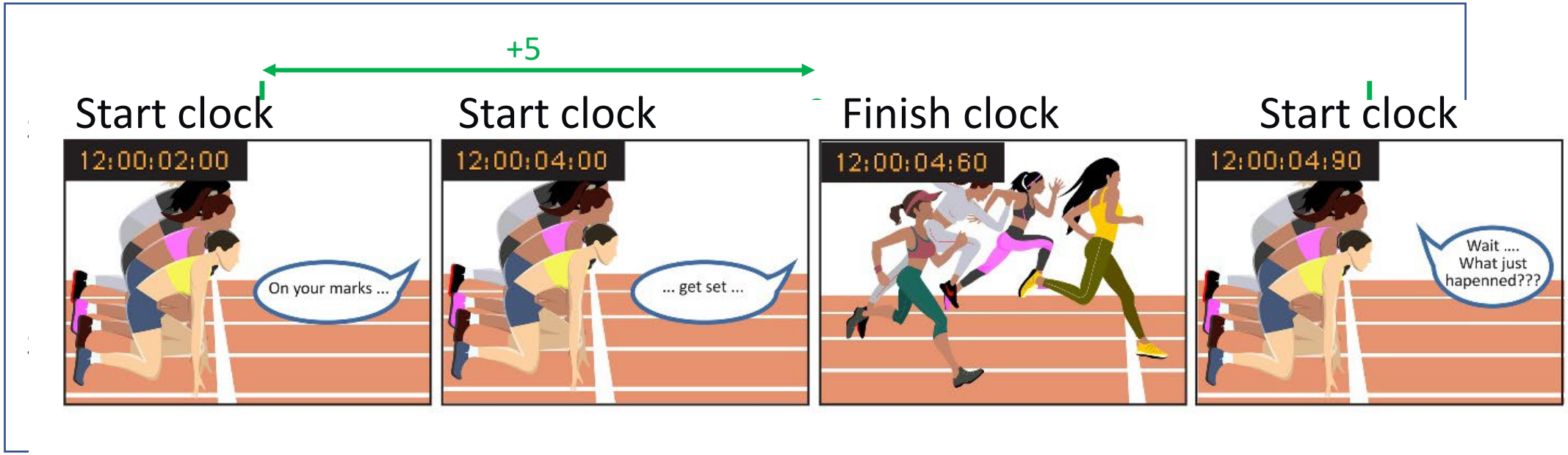
- **Some transactions are causal and must ordered correctly**
- **With an ideal time source or single machine this is trivial**
- **With clock uncertainty this can be a challenge**
- **100 m dash with start and end clock with +/- 5 second**
  - Start clock 5 seconds fast
  - Finish clock 5 seconds slow



MICROCHIP

Time Challenges with Multiple Machines

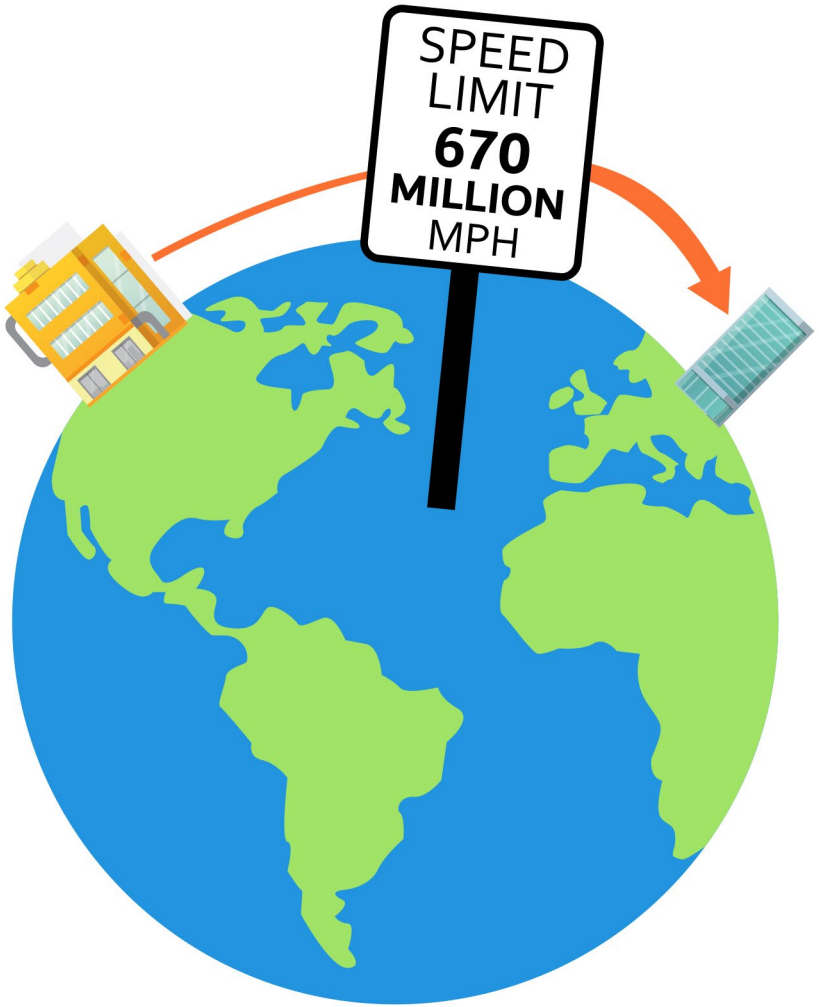
- **Some transactions are causal and must ordered correctly**
- **With an ideal time source or single machine this is trivial**
- **With clock uncertainty this can be a challenge**
- **100 m dash with start and end clock with +/- 5 second**
  - Start clock 5 seconds fast
  - Finish clock 5 seconds slow



MICROCHIP

Why Replication?

- **Users located throughout globe**
- **Speed of light limits communication speed between locations**
- **Vancouver to Dusseldorf**
  - 4900 miles
  - Speed of light – 26 ms
  - Round trip ~ 50 ms
  - Real world ~150 to 250 ms
- **Maximum 4 round trips per second**
- **Solution: replicate data at multiple locations near user**
- **Requires external time reference to synchronize**



MICROCHIP

Synchronization: External Time

Security, Capacity, Speed and Resiliency Issues

Improved Security, Capacity and Resiliency

GNSS

Public Time Server

Firewall

Data Center Time Server/Appliance

Public Internet

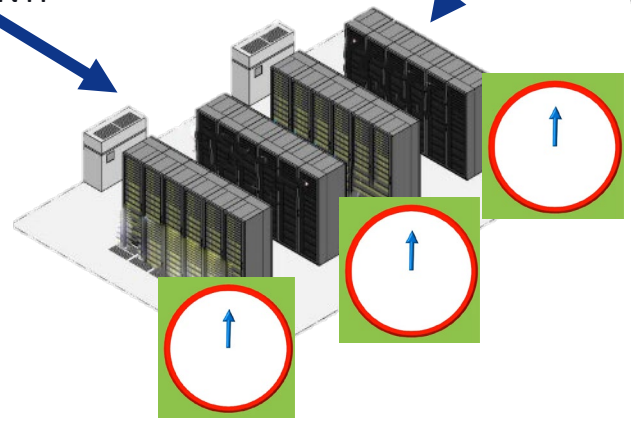
Public NTP

NTP, PTP

Atomic Clock

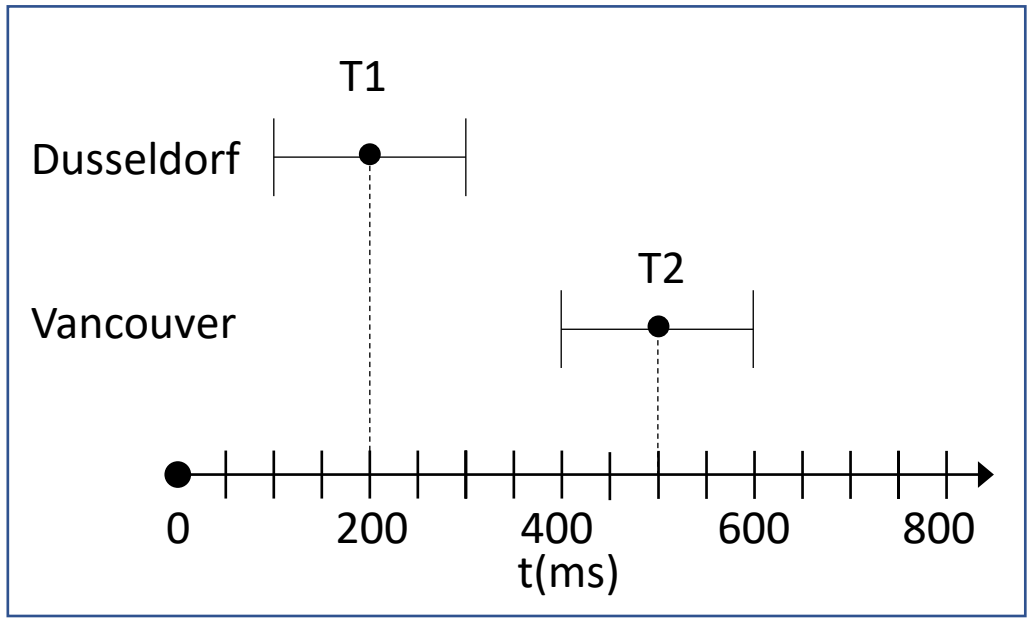
Fiber, PTP HA eLoran

Alternate Time Source

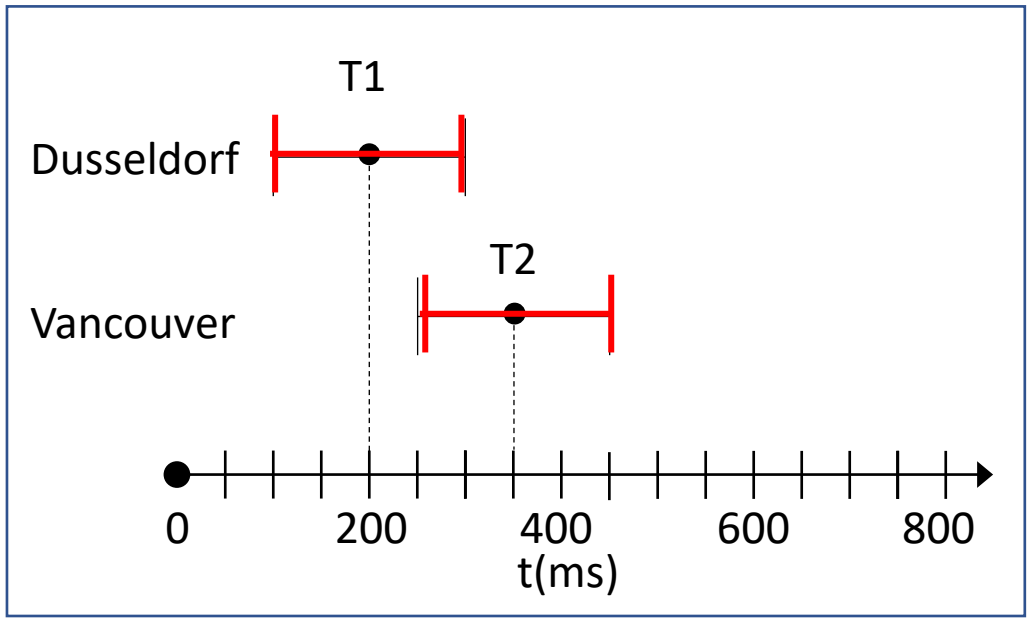


MICROCHIP

Clock with ±100 ms Uncertainty: “Time Envelopes”



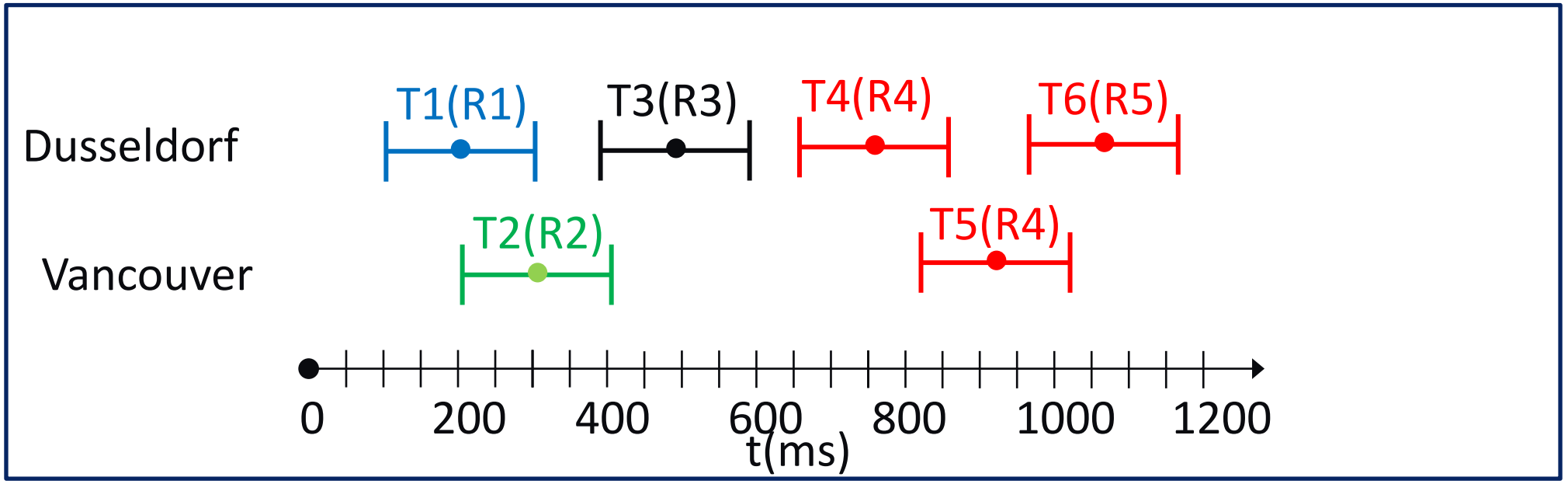
Resolvable Consistency



Irresolvable Consistency

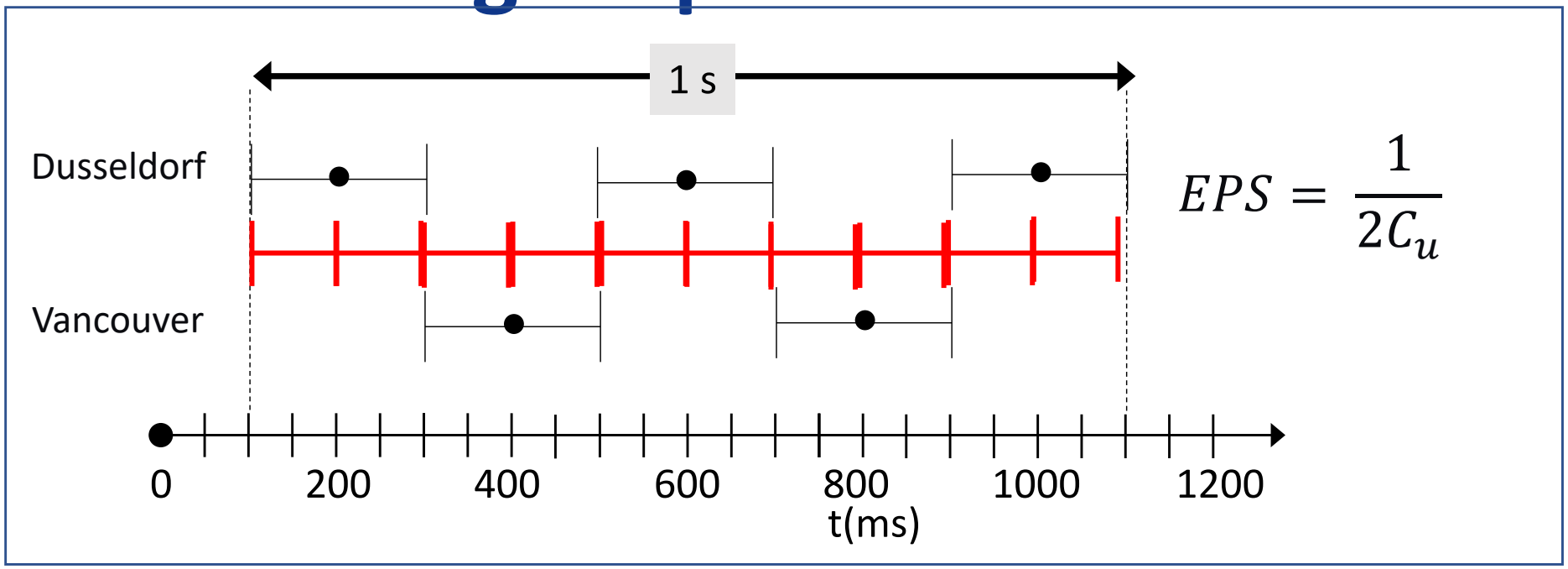
- Clock uncertainty (C_u) creates time envelope around time stamp
  - you don’t put stamps on a window
- If time envelopes overlap, at least one transaction may need to be rolled back after the fact, causing system speed issues, cost and possible corruption
- Decreasing clock uncertainty -> Higher transactions per second (TPS) and better efficiency
- Cost vs. performance trade-off

Timing Not as Critical for Unrelated Records



- If R1, R2, and R3 are unrelated, then transaction T1 through T3 overlap is acceptable
  - Example: entering names Fred, Bob, Bill
- T4 and T5 timing issue as both act on R4
  - R4=1 at 600 ms, T4=R÷2, T5=R+1 , at 1000 ms R4=1 or 1.5
- If R5 has a causal relationship with R4 then T5 and T6 are problematic due to envelope overlap
  - T6=R4 × 2, at 1200 ms R5=3, 4 or 2

Application Timing Requirements

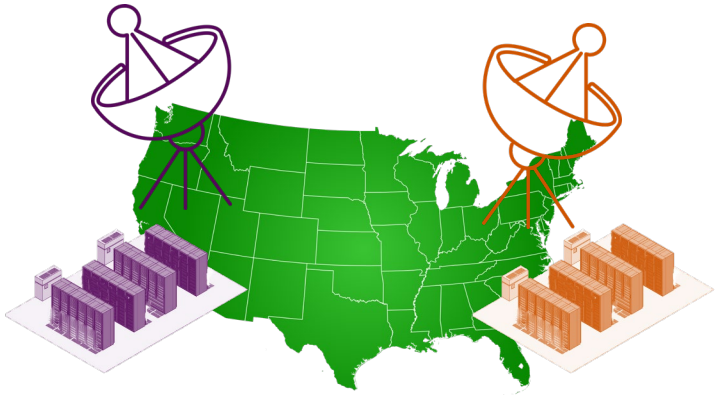


- Clock with ± 100ms uncertainty has a maximum of 5 non overlapping time envelopes per second (EPS)
- Increasing EPS will decrease probability transactions will have overlapping envelopes, but can never completely remove it
- Ideal solution: Hydrogen maser steered by Two Way Satellite Time Transfer or fiber to UTC(k) on every server

Applications: Not Just Hyperscale

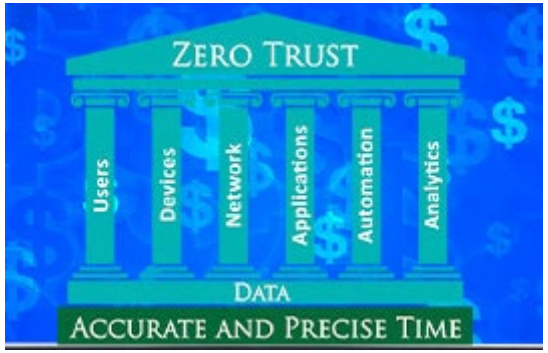
Hyperscale
Data Centers

- Precision improves efficiencies
- Reduces possibility of corruptions
- Reduces operational costs
- Timing SLA



Defense Radar
Networks

- Precision time needed to integrate sensor results



Financial
Networks

- Regulatory Requirements
- Higher precision can result in higher trade rate

MICROCHIP

Comparing TPS Needs to EPS

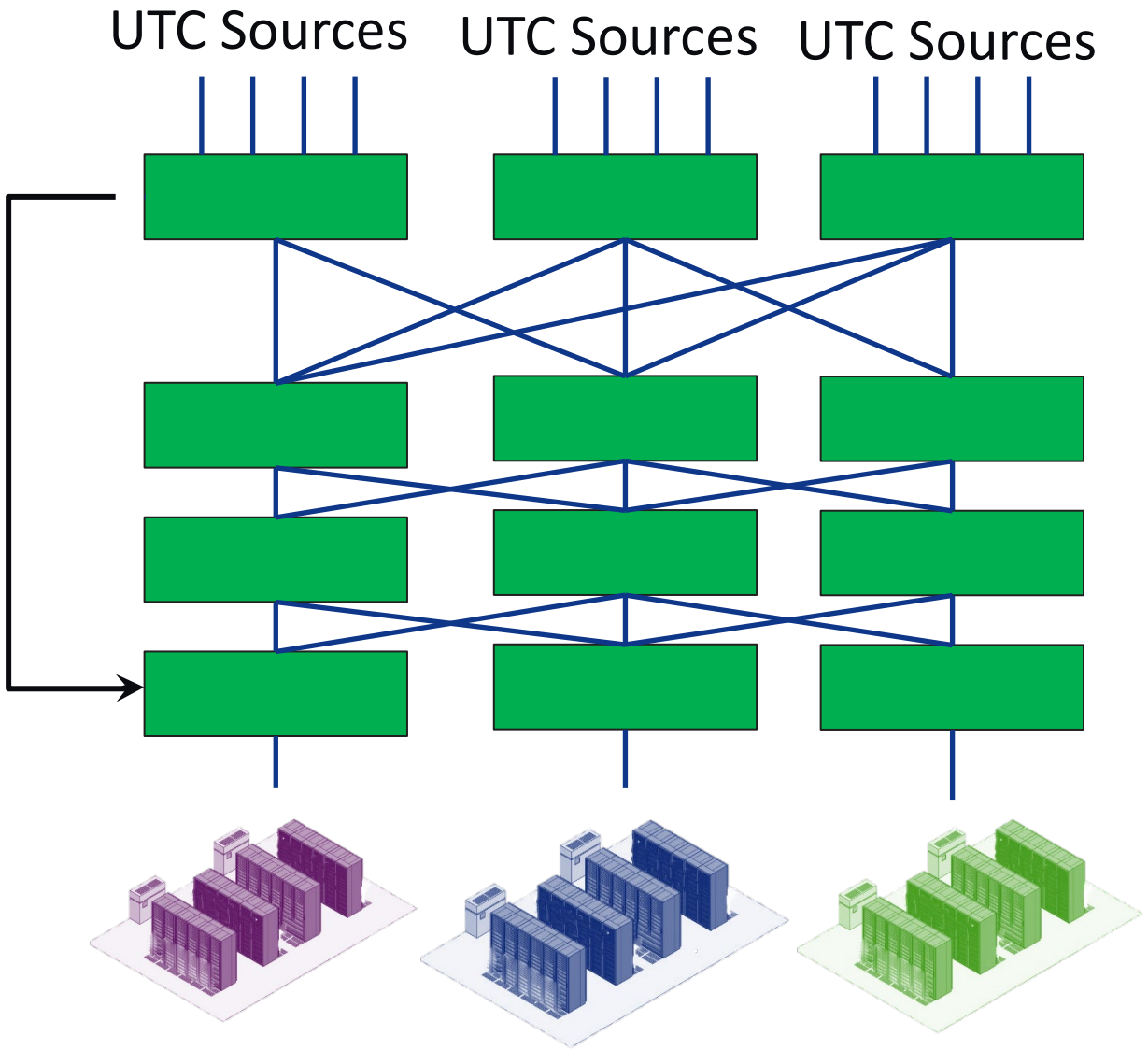
- **The brute force method would just compare TPS to EPS**
- **Weighted TPS (WTPS): scaled to reflect application requirements**
  - Criticality of Inconsistencies - S_c
  - Peak Transaction Rates - S_{pk}
  - Probability of near simultaneous write to same or causal records S_{pr}
  - Any other scale factor you want to add for your deployment (budget allowances etc.)

$$TPS \times S_c \times Spk \times Spr = WTPS < EPS = \frac{1}{2Cu}$$

Theoretical Examples

Use Case	Criticality of Unresolvable Inconsistency	Probability of Related Records Updated at Different Locations in Short Time Period	Average TPS	S_{pk}	S_c	S_{pr}	WTPS	CU (s)
Military Operations Database	**High** -Incorrectly reported data can result in loss of life	**High** -Integrated battlefield operations are updated from multiple sources	10,000	100	100	1	100,000,000	5.00E-09
Financial Trading Company	**High** -Lost trades can have siginfcant financial effects	**High** - High volume of trading requests at certain times of days and stock news creates individual targets	1,000	100	10	1	1,000,000	5.00E-07
Local Credit Union	**High** - Lost transactions can lead to regualtory and legal issues	**Medium** - Accounts need to be accessed at different locations	1,000	1	10	0.1	1,000	5.00E-04
Industrial Process Control	**Medium** - A few missing datapoints can likely be averaged out	**Low** - Data from sensors taken at regular intervals and specific locations	10,000	1	10	0.01	1,000	5.00E-04
Medical Database	**High** - Incorrectly reported data can result in loss of life	**Low -** Individual patient data seldomly transacted at differing locations simultaneously	1,000	1	10	0.01	100	5.00E-03
Social Media Photography Application	**Low** -Affects customer experience, but can be resolved by reentering	**Low** - Most photograph metadata only updated by account holder	10,000	10	0.01	0.01	10	5.00E-02
Sales Contact Database	**Medium** - Possible incorrect data stored for customers	**Low** - Updates may occur at different sites, but not likely to occur within close time frames	50	1	1.0	0.01	0.5	1

Typical Timing Architecture: Each Site



Cu

T_s

Local Atomic Clock(s)

T_h

Network Degradation

T_n

Local Oscillator

T_o

MICROCHIP

Cu Examples

G.811.1
30 ns

ePRTC

G.8273.2
5 ns

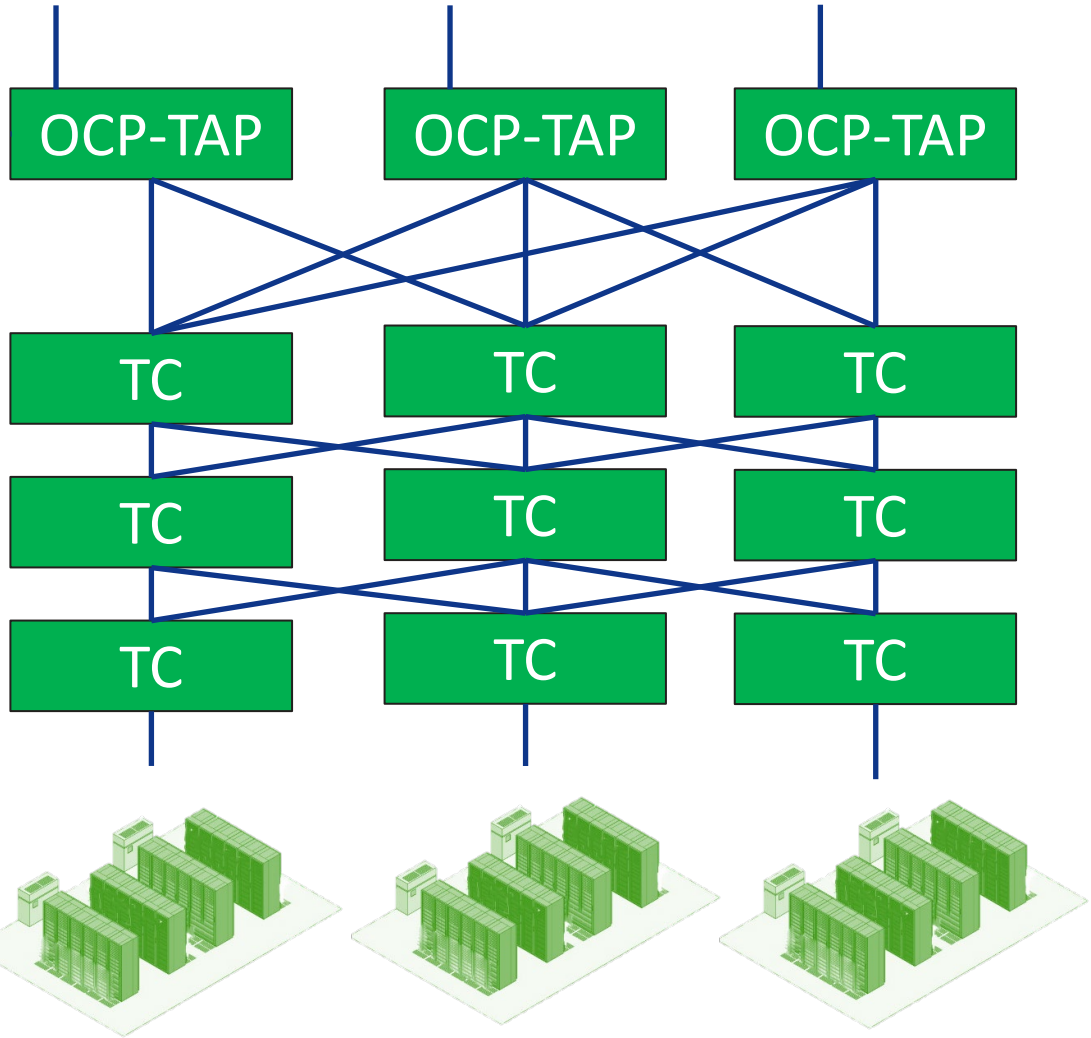
Class D Boundary Clock

G.812 type 1
10 ns

Ordinary Clock

Cu= 45ns max
(32ns if defined as sigma values)

OCP Data Center Profile

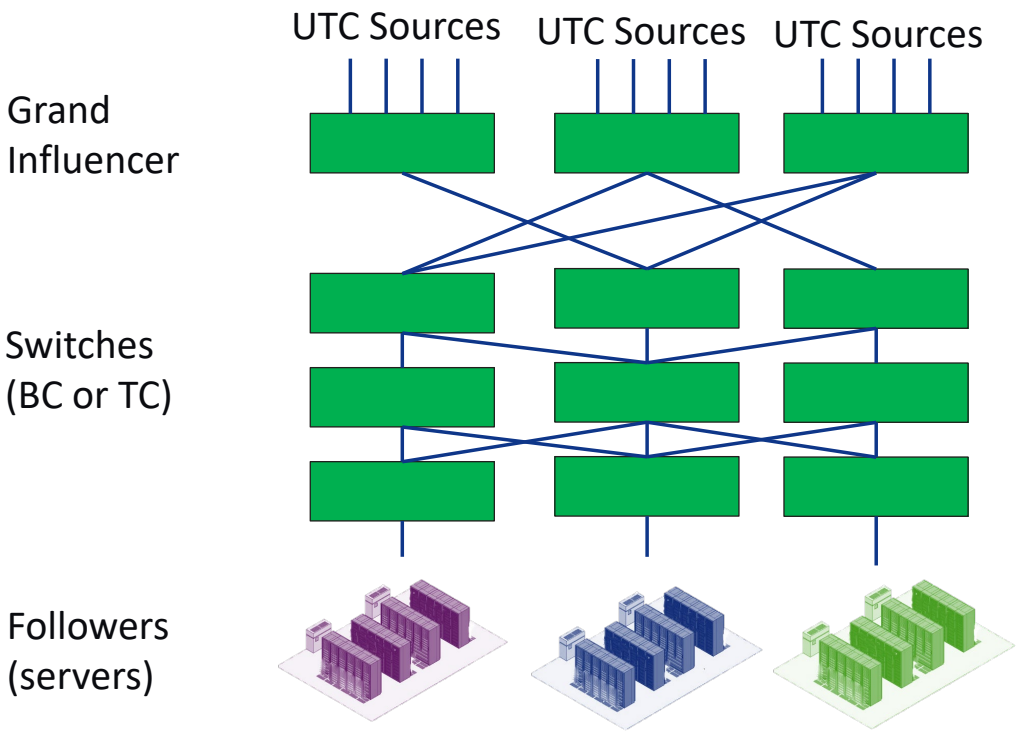


Cu= 2.5 µs

MICROCHIP

Buy Time or Make Time Decision

Typical Timing Architecture (Each Site)



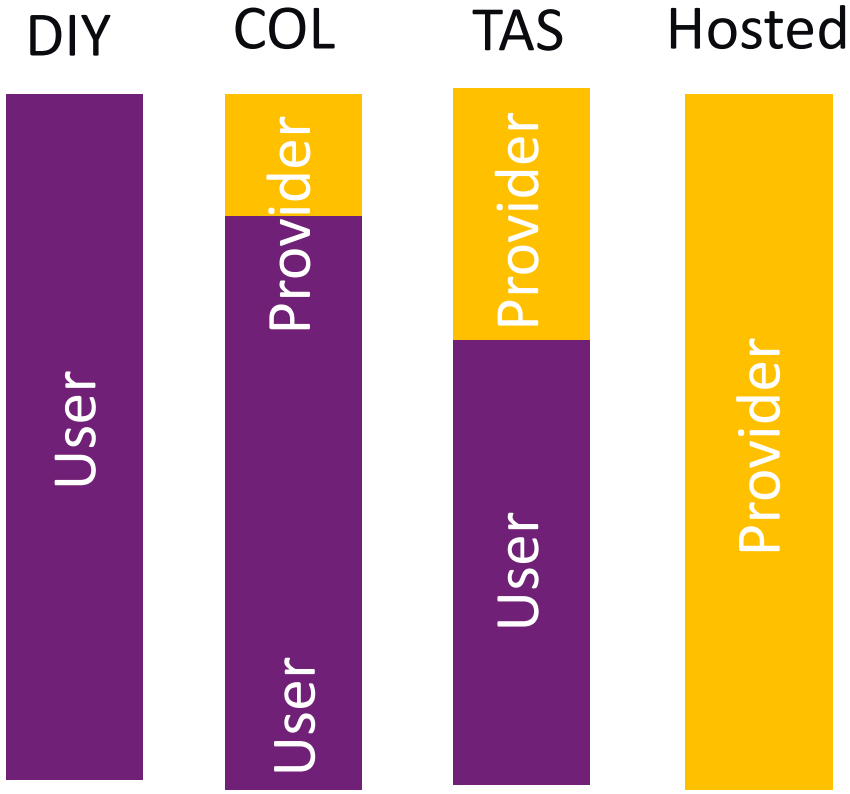
Cu

T_s

T_h

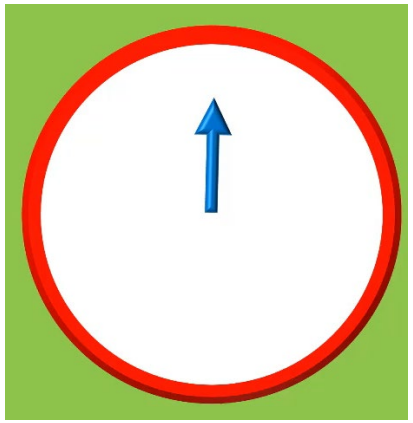
T_n

T_o



MICROCHIP

Thanks for Your Time



MICROCHIP