HITACHI
Inspire the Next

Testing in minimum time with ivy

"seen enough & stop"

August 9, 2016
Allart Ian Vogelesang
ian.vogelesang@hds.com

Hitachi Data Systems

What does it mean – a “valid” measurement

HITACHI
Inspire the Next

- Valid results are *repeatable*
  - If you run the test again, within specified +/- experimental error, you will get the same result again.
  - Valid, repeatable, results are for steady-state conditions
- If the workload / subsystem are not steady-state, you can’t make a valid measurement.
  - After imposing a workload on the subsystem
    1. we need to wait for the behaviour to settle down into a steady state, waiting for initial transient conditions to settle down, then
    2. depending on how much "noise" there is in what is being measured, measure for long enough to obtain a valid measurement to a specified plus/minus accuracy.

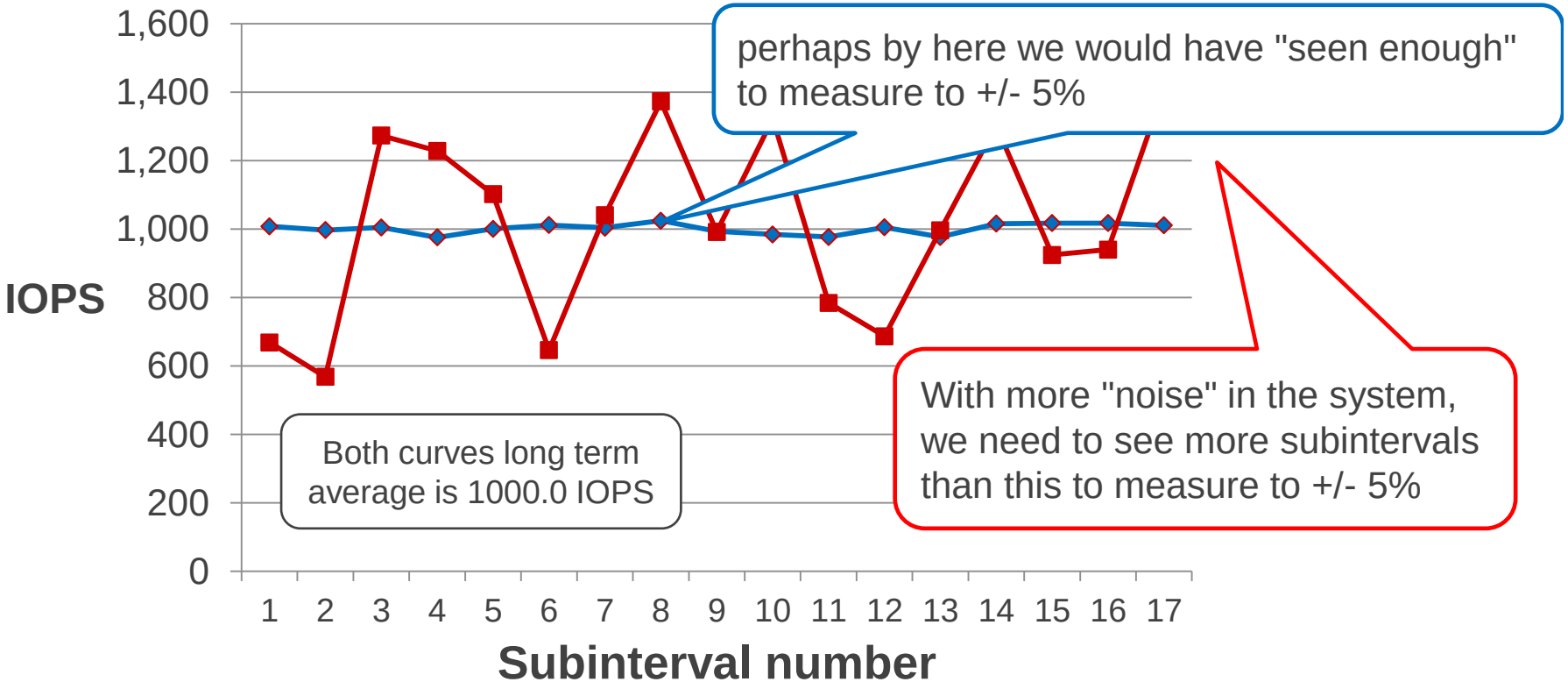
An ivy test is a sequence of "subintervals"

HITACHI
Inspire the Next

- The default is **`subinterval_seconds = 5`**
- At end of subinterval, workload thread measurement data sent to ivymaster.
- ivymaster decides to continue the test for another subinterval, or to stop.
- For most workloads, subsystem response is very stable, so you can make an accurate measurement quite quickly.
- For some workloads where behaviour is more dynamic, either due to initial transient behaviour or else instability, we need to test longer to get an accurate measurement

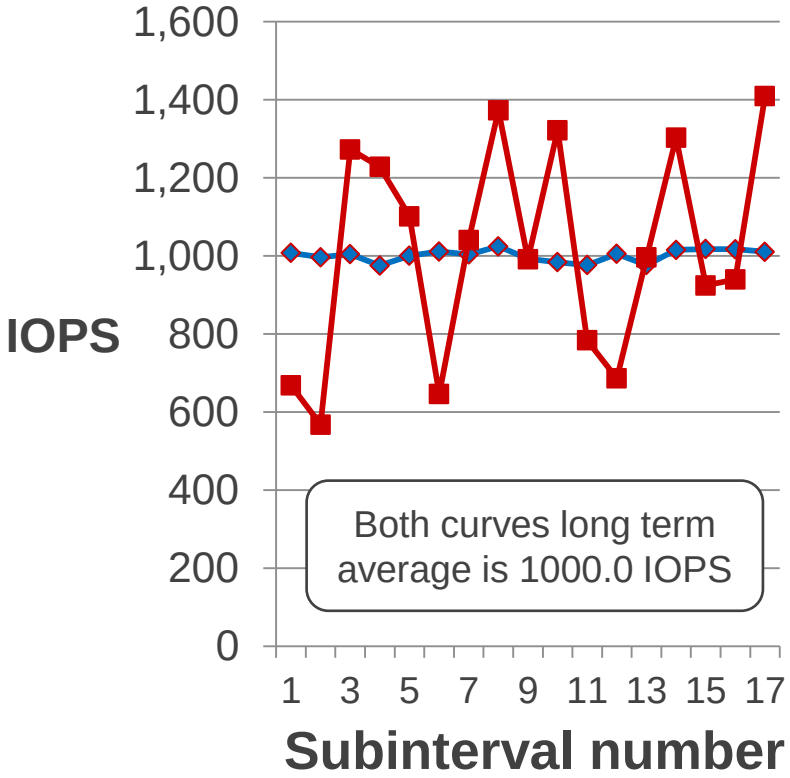
measure=IOPS, accuracy_plus_minus=5%

HITACHI
Inspire the Next



measure=IOPS, accuracy_plus_minus=5%

HITACHI
Inspire the Next



- The ivy "`measure`" feature extends the test until enough measurements have been made to be statistically confident that the average of the by-subinterval values is within a specified `accuracy_plus_minus` variation from what the long term average would be continuing to test indefinitely.

`measure` parameters

HITACHI
Inspire the Next

- `accuracy_plus_minus = "5%"`
  - Default is "`5%`".
- `confidence = "95%"`
  - How confident you need to be that your measurement falls within the specified plus or minus range around the long term average that you would get measuring forever.
  - Default is "`95%`"
  - Ivy has a menu of 11 specific pre-loaded confidence values that you pick from.
    - 50%, 60%, 70%, 80%, 90%, 95%, 98%, 99%, 99.5%, 99.8%, and 99.9%

Standard statistical formulas used

HITACHI
Inspire the Next

- It is both convenient and effective to use the standard statistical formulas using the `accuracy_plus_minus` and `confidence` parameters
  - http://en.wikipedia.org/wiki/Student%27s_t-distribution (sorry, for the math literate)
- Limitation – the standard formulas don't exactly apply to ivy
  - The formulas are for samples drawn at random from a large population.
  - The behaviour / measurement for one subinterval is related to those for preceding / succeeding subintervals, thus "gaming the system".
- Rule of thumb to correct for this
  - Specify an `accuracy_plus_minus` parameter for twice the required accuracy, and you'll be safe.
    - Say `accuracy_plus_minus = 5%` to get repeatability within 10%.

`measure` parameter "shorthand" presets

HITACHI
Inspire the Next

- `measure = IOPS`
  `measure = MB_per_second`
  `measure = service_time_seconds`
  `measure = response_time_seconds`

- These are "shorthand"
  - See "ivy programmer's reference" for more detail

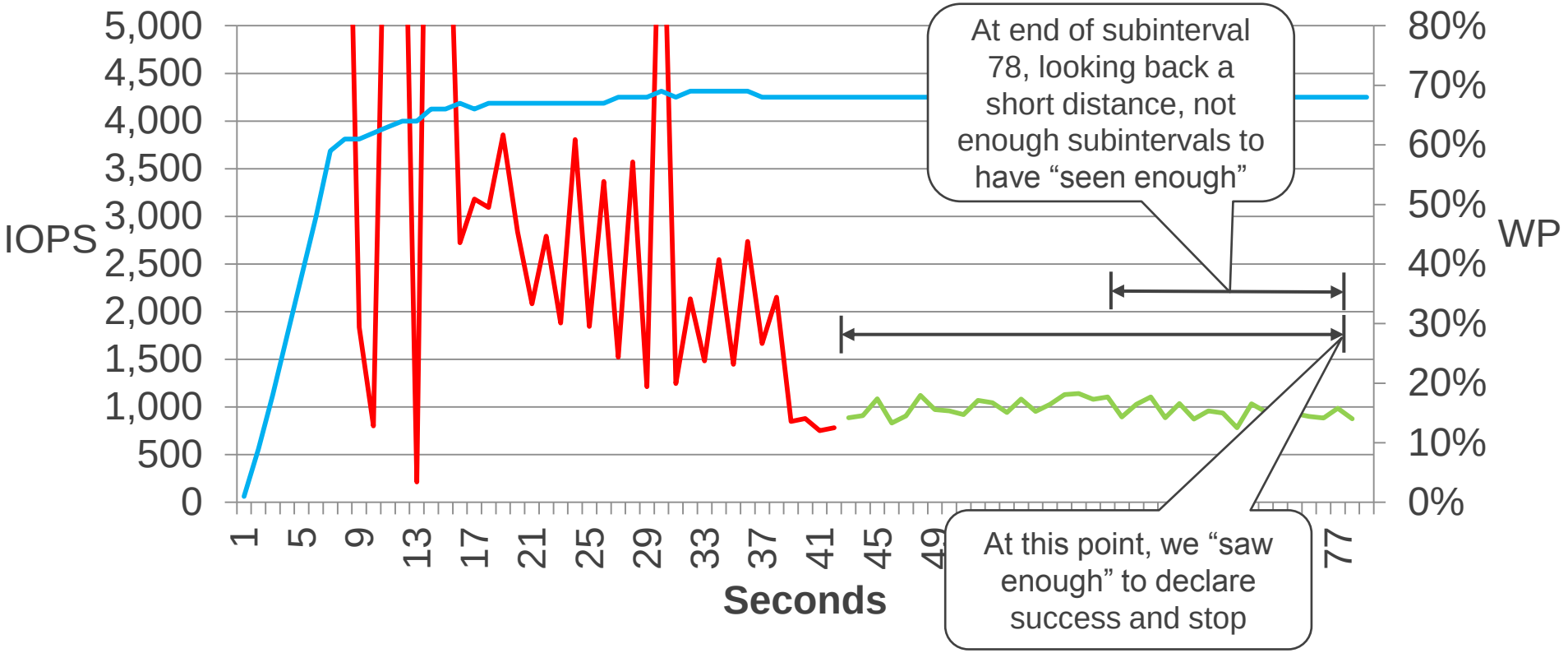
Initial study - 60% test time reduction, better quality

HITACHI
Inspire the Next

- Instead of running for a fixed 6 minutes for each step of the several hundred combinations / permutations of a standard scalability test sequence, when a variable step run time using `measure` was used
  - Overall test run time decreased by 60%
    - The vast majority of the time workloads settled down quickly and were then stable.
  - Quality was improved
    - In some rare cases, there were test steps that needed to run for 20 minutes to get an accurate measurement.
- The use of the `measure` feature in ivy substantially reduces test time, while at the same time improving quality.

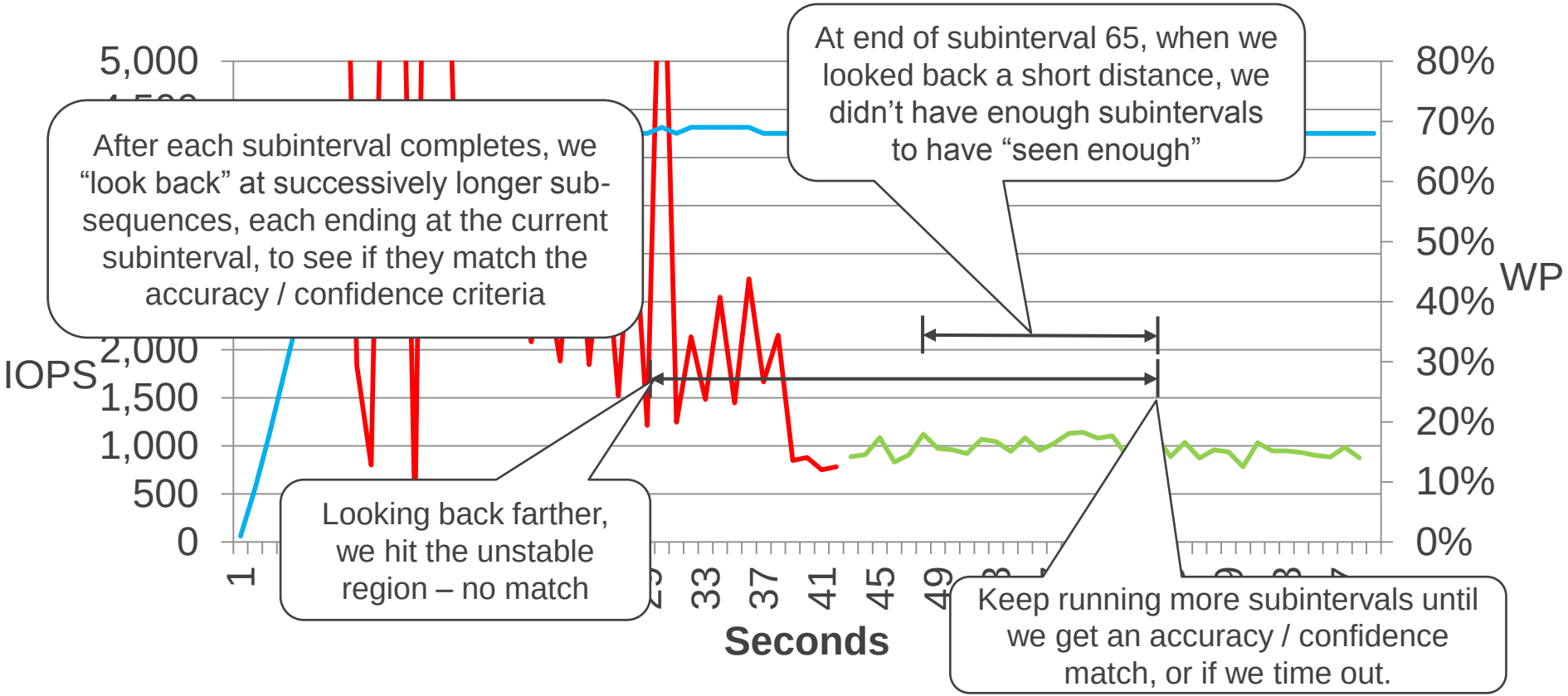
`measure` automatically rejects initial transients

HITACHI
Inspire the Next



How it works - example of "have not seen enough”

HITACHI
Inspire the Next



warmup_seconds, measure_seconds, subinterval_seconds

HITACHI
Inspire the Next

- warmup_seconds – default = 5
- measure_seconds – default = 60
- Without measure, the warmup and measurement periods are fixed.
- When measure is used, these warmup_seconds and measurement_seconds values represent minimum periods during automatic detection of a valid measurement.
- subinterval_seconds default is 5
  - Don't make this shorter, but if you are running really long test steps, you could use longer subintervals to reduce the volume of csv output.

Conclusion

HITACHI
Inspire the Next

- The `measure` feature is convenient and effective to run tests for the minimum necessary number of subintervals
  - In one example, overall test time was reduced by 60% while measurement quality was improved.
- Standard small sample set math is effective with a 2x fudge factor.
  - Say `accuracy_plus_minus=5%` and the rule of thumb is you will safely get a measurement repeatable to within 10%.

HITACHI
Inspire the Next

Questions and Discussion

Hitachi Data Systems

HITACHI
Inspire the Next

Thank You

Hitachi Data Systems