CPU Oversubscription in Compute Clouds

Mitch Richling

May 11, 2018   *Revision* : 1.1

TEXAS INSTRUMENTS

Outline

TEXAS INSTRUMENTS

Part I: Context

Vocabulary

- **CPU Utilization**: That thing that comes out of 'top'. ;)
- **Slot**: An allocation for a job thread to be dispatched.
- **Job Run Time**: The time a job is allocated to CPU(s)
- **Reserved Time**: Run time multiplied by slots reserved
- **CPU Time**: Quantity of CPU time consumed by job
- **Job Efficiency**: CPU time divided by reserved time

Texas Instruments

The Good Ole Days

When I was a kid...

Simulation batch jobs were single threaded and consumed precisely one, full CPU. i.e. jobs were almost 100% efficient.

The Good Ole Days

Well... OK... That was never really true,...

It *was* true that a remarkably large number of jobs were very efficient. True enough in fact that the standard practice of the day was to configure compute clusters to run one job per CPU core, and achieve nearly 100% CPU utilization.

In typical LSF speak: The *slot:core ratio* was 1.

TEXAS INSTRUMENTS

Historical Job Efficiency (Representative)
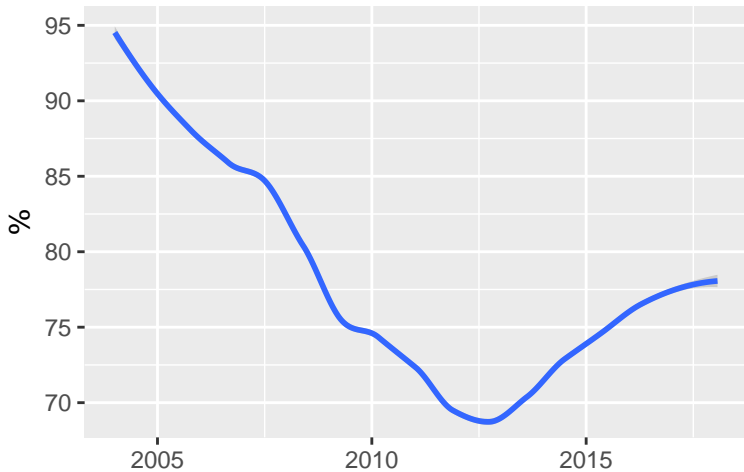
Average job efficiency is pretty low.

TEXAS INSTRUMENTS

Running with $\text{slot} : \text{core} = 1$ wastes significant resources.

Texas Instruments

Today

Is it safe to run with $slot : core > 1$?

Running with $slot : core > 1$

- Increase system utilization
- Increase throughput
- Creates a risk of decreased individual run time

Texas Instruments

- ▶ Increase system utilization
- ▶ Increase throughput
- ▶ Creates a risk of decreased individual run time

This presentation discusses techniques to quantify and limit the risks.

TEXAS INSTRUMENTS

Running with $slot : core > 1$

By quantifying and limiting the risk, we transform the question of setting higher $slot : core$ values into one of business priorities and a well formulated risk-benefit analysis.

Part II: Thought Experiments

Some Thought Experiments

Next we consider three thought experiments in which we take a simplistic point of view using averages. This line of thinking is fine for thought experiments intended to provide insight into general system behavior, but it is wholly inappropriate for decisions regarding the configuration of production compute clouds.
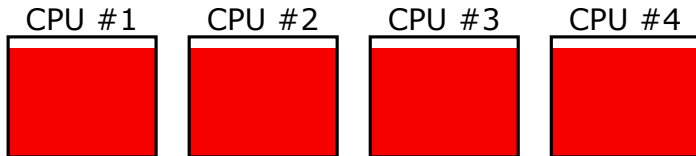
A thought experiment (Number 1)

The Scenario

Average job efficiency is 90%. Compute servers have 4 cores.

A thought experiment (Number 1)

CPU #1   CPU #2   CPU #3   CPU #4

A thought experiment (Number 1)

The Scenario
Average job efficiency is 90%. Compute servers have 4 cores.

Observations
- System utilization is capped at 90%
- 40% of a CPU is wasted

TEXAS INSTRUMENTS

A thought experiment (Number 2)

The Scenario

Average job efficiency is 80%. Compute servers have 4 cores.

A thought experiment (Number 2)

The Scenario
Average job efficiency is 80%. Compute servers have 4 cores.

Observations
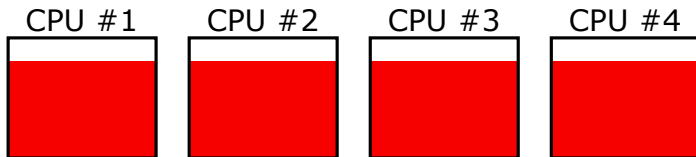- System utilization is capped at 80%
- We have 80% of a CPU wasted.
- We have room for a 5th job (i.e. 80% of a CPU)!

A thought experiment (Number 2)
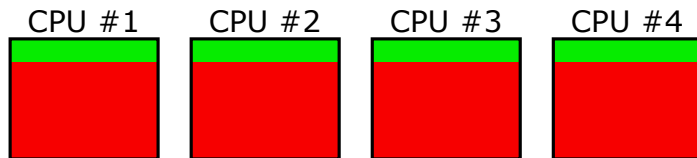
Opportunity!

If we set slot:core to 5:4, then

- ▶ Systems would be utilized at 100%
- ▶ Throughput improves by 20%
- ▶ Job run time is *not* impacted!

TEXAS INSTRUMENTS

CPU #1   CPU #2   CPU #3   CPU #4

We have room for a 5th job!

TEXAS INSTRUMENTS

A thought experiment (Number 3)

The Scenario

Average job efficiency is 90%. Compute servers have 12 cores.

A thought experiment (Number 3)

CPU #1	CPU #2	CPU #3	CPU #4

CPU #5	CPU #6	CPU #7	CPU #8

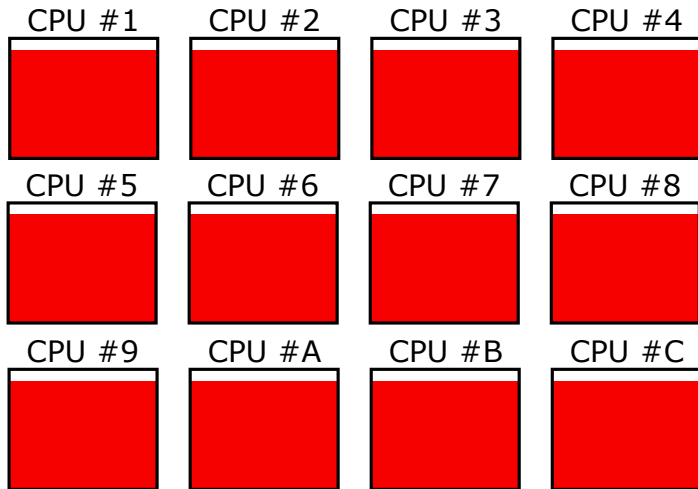| CPU #9	CPU #A	CPU #B	CPU #C

A thought experiment (Number 3)

The Scenario

Average job efficiency is 90%. Compute servers have 12 cores.

Observations

- System utilization is capped at 90%
- We have more than an *entire* CPU wasted!

A thought experiment (Number 3)

Opportunity!

If we set slot:core to 13:12, then

- Systems would be utilized at 97.5%
- 30% of a CPU is still wasted!
- Throughput improves by 7.5%
- Job run time is *not* impacted!
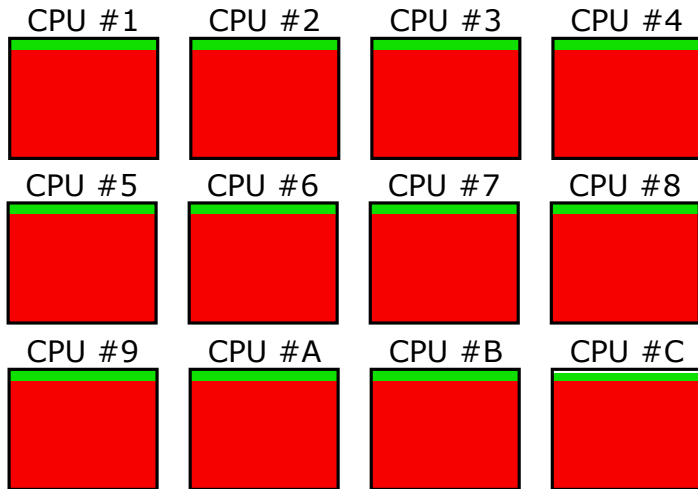
Recap

The Idea
The idea is simple: Use left over CPU capacity for more jobs

Recap

The Optimist

At 75% job efficiency, I can run 33.3% more jobs!!!

CPU #1 CPU #2 CPU #3 CPU #4

CPU #5 CPU #6 CPU #7 CPU #8

CPU #9 CPU #A CPU #B CPU #C
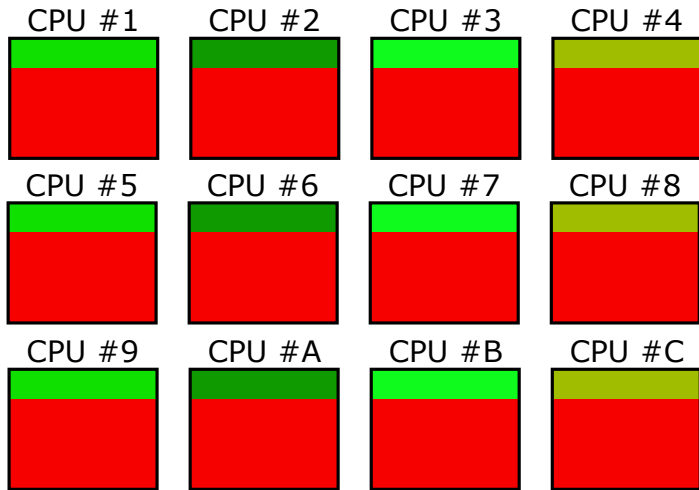
4 Extra Jobs!!

TEXAS INSTRUMENTS

Recap

The Realist

Nope.

Recap

In short, the realist recognizes the previously mentioned limitations of the simplistic approach we have taken with our thought experiments.

TEXAS INSTRUMENTS

Part III: Compute Cluster Vulgarities

Job efficiency varies as the job runs

CPU Utilization of a Single 12−way Multithreaded Job on a 12 Core Host

TEXAS INSTRUMENTS
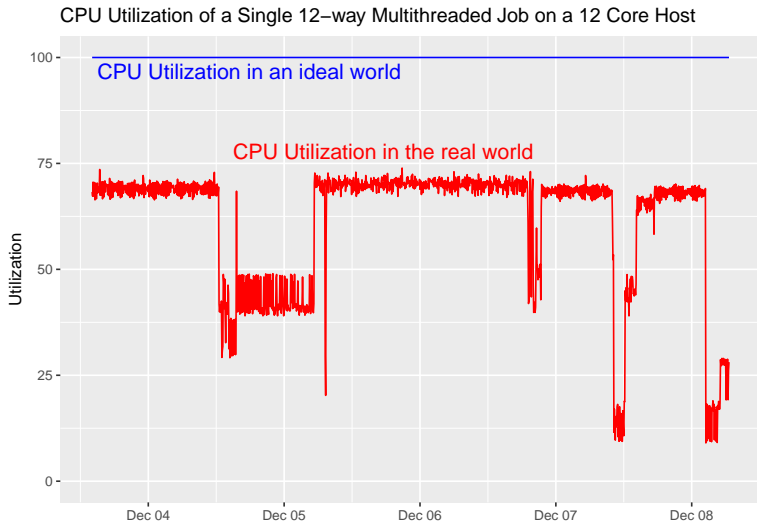
Job efficiency varies between jobs

CPU Utilization Under Mixed Job Load on a 12 Core Host

TEXAS INSTRUMENTS
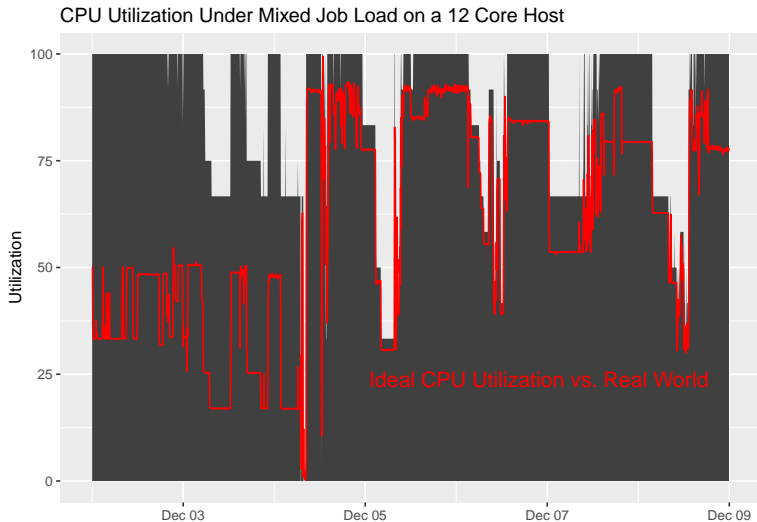
Batch scheduling job placement must be considered

Scheduling polices can..

- Lump together jobs of similar efficiency
- Pack jobs onto systems till they appear full
- Avoid systems that look full but are not
- ...

Part IV: Probability Theory

Dice!!
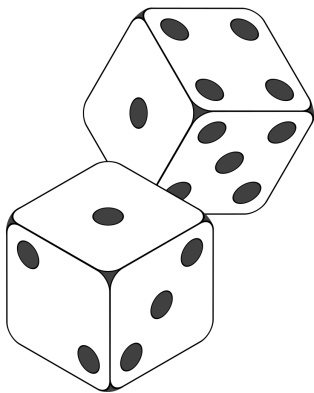
Rolling One Die

Face	Ways To Roll	Probability	
1	1	1/6	▬▬▬
2	1	1/6	▬▬▬
3	1	1/6	▬▬▬
4	1	1/6	▬▬▬
5	1	1/6	▬▬▬
6	1	1/6	▬▬▬

Rolling Two Dice

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

TEXAS INSTRUMENTS

Rolling Two Dice

	1	2	3	4	5	6
1	2	3	**4**	5	6	7
2	3	**4**	5	6	7	8
3	**4**	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

$6 \cdot 6 = 36$ different ways to roll a pair of dice.

Of those 36 we see 3 different ways to get a total of 4.

Thus the probability of rolling a 4 is $\frac{3}{36} = \frac{1}{12} = 0.083333\overline{3}$.

TEXAS INSTRUMENTS

Rolling Two Dice

Total	Ways To Roll	Probability	
2	1	1/36	▬
3	2	2/36	▬▬
4	3	3/36	▬▬▬
5	4	4/36	▬▬▬▬
6	5	5/36	▬▬▬▬▬
7	6	6/36	▬▬▬▬▬▬
8	5	5/36	▬▬▬▬▬
9	4	4/36	▬▬▬▬
10	3	3/36	▬▬▬
11	2	2/36	▬▬
12	1	1/36	▬

TEXAS INSTRUMENTS

1	1	2	3	4	5	6
1	3	4	5	6	7	8
2	4	5	6	7	8	9
3	5	6	7	8	9	10
4	6	7	8	9	10	11
5	7	8	9	10	11	12
6	8	9	10	11	12	13

2	1	2	3	4	5	6
1	4	5	6	7	8	9
2	5	6	7	8	9	10
3	6	7	8	9	10	11
4	7	8	9	10	11	12
5	8	9	10	11	12	13
6	9	10	11	12	13	14

3	1	2	3	4	5	6
1	5	6	7	8	9	10
2	6	7	8	9	10	11
3	7	8	9	10	11	12
4	8	9	10	11	12	13
5	9	10	11	12	13	14
6	10	11	12	13	14	15

4	1	2	3	4	5	6
1	6	7	8	9	10	11
2	7	8	9	10	11	12
3	8	9	10	11	12	13
4	9	10	11	12	13	14
5	10	11	12	13	14	15
6	11	12	13	14	15	16

5	1	2	3	4	5	6
1	7	8	9	10	11	12
2	8	9	10	11	12	13
3	9	10	11	12	13	14
4	10	11	12	13	14	15
5	11	12	13	14	15	16
6	12	13	14	15	16	17

6	1	2	3	4	5	6
1	8	9	10	11	12	13
2	9	10	11	12	13	14
3	10	11	12	13	14	15
4	11	12	13	14	15	16
5	12	13	14	15	16	17
6	13	14	15	16	17	18

TEXAS INSTRUMENTS

Rolling Three Dice

$6^3 = 216$ different ways to roll three dice with 6 sides.

Of those 216 we see 27 different ways to get a total of 10.

Thus the probability of rolling a 10 is $\frac{27}{216} = \frac{1}{8} = 0.125$.

TEXAS INSTRUMENTS

Rolling Three Dice

Total	Ways To Roll	Ways To Roll	Probability	
3	1	1	1/216	▪
4	1+2	3	3/216	▬
5	1+2+3	6	6/216	▬▬
6	1+2+3+4	10	10/216	▬▬▬
7	1+2+3+4+5	15	15/216	▬▬▬▬
8	1+2+3+4+6	21	21/216	▬▬▬▬▬
9	2+3+4+5+6+5	25	25/216	▬▬▬▬▬▬
10	3+4+5+6+5+4	27	27/216	▬▬▬▬▬▬▬
11	4+5+6+5+4+3	27	27/216	▬▬▬▬▬▬▬
12	5+6+5+4+3+2	25	25/216	▬▬▬▬▬▬
13	6+5+4+3+2+1	21	21/216	▬▬▬▬▬
14	1+2+3+4+5	15	15/216	▬▬▬▬
15	1+2+3+4	10	10/216	▬▬▬
16	1+2+3	6	6/216	▬▬
17	1+2	3	3/216	▬
18	1	1	1/216	▪

TEXAS INSTRUMENTS

Computational Efficiency

Rolling 2 Dice

Computing the probabilities for rolling two dice, each with six sides, required us to form a table with $6 \cdot 6 = 36$ entries. The computational complexity of this problem is $\mathcal{O}(6^2)$.

TEXAS INSTRUMENTS

Rolling 3 Dice

Computing the probabilities for rolling three dice, each with six sides, required us to form a table with $6^3 = 216$ entries. The computational complexity of this problem is $\mathcal{O}(6^3)$.

TEXAS INSTRUMENTS

Computational Efficiency

Rolling n Dice

Computing the probabilities for rolling n dice, each with six sides, would require us to form a table with 6^n entries. The computational complexity of this problem is $\mathcal{O}(6^n)$.

Texas Instruments

Computational Efficiency

What if the dice had m sides?

Computing the probabilities for rolling n dice, each with m sides, would be just about the same problem. Just more rows a columns in our table! Thus the computational complexity would become $\mathcal{O}(m^n)$.

This is very, very bad...

TEXAS INSTRUMENTS

Fundamental Probability Reminder

If the probability of E_1 is p_1, and the probability of E_2 is p_2, then

The probability of both E_1 **AND** E_2 occurring is $p_1 \cdot p_2$

The probability of E_1 **OR** E_2 occurring is $p_1 + p_2$

TEXAS INSTRUMENTS

Rolling Two Dice (again)

	$1\left\vert\left(\frac{1}{6}\right)\right.$	$2\left\vert\left(\frac{1}{6}\right)\right.$	$3\left\vert\left(\frac{1}{6}\right)\right.$	$4\left\vert\left(\frac{1}{6}\right)\right.$	$5\left\vert\left(\frac{1}{6}\right)\right.$	$6\left\vert\left(\frac{1}{6}\right)\right.$
$1\left\vert\left(\frac{1}{6}\right)\right.$	$2\left\vert\left(\frac{1}{6}\right)^2\right.$	$3\left\vert\left(\frac{1}{6}\right)^2\right.$	$4\left\vert\left(\frac{1}{6}\right)^2\right.$	$5\left\vert\left(\frac{1}{6}\right)^2\right.$	$6\left\vert\left(\frac{1}{6}\right)^2\right.$	$7\left\vert\left(\frac{1}{6}\right)^2\right.$
$2\left\vert\left(\frac{1}{6}\right)\right.$	$3\left\vert\left(\frac{1}{6}\right)^2\right.$	$4\left\vert\left(\frac{1}{6}\right)^2\right.$	$5\left\vert\left(\frac{1}{6}\right)^2\right.$	$6\left\vert\left(\frac{1}{6}\right)^2\right.$	$7\left\vert\left(\frac{1}{6}\right)^2\right.$	$8\left\vert\left(\frac{1}{6}\right)^2\right.$
$3\left\vert\left(\frac{1}{6}\right)\right.$	$4\left\vert\left(\frac{1}{6}\right)^2\right.$	$5\left\vert\left(\frac{1}{6}\right)^2\right.$	$6\left\vert\left(\frac{1}{6}\right)^2\right.$	$7\left\vert\left(\frac{1}{6}\right)^2\right.$	$8\left\vert\left(\frac{1}{6}\right)^2\right.$	$9\left\vert\left(\frac{1}{6}\right)^2\right.$
$4\left\vert\left(\frac{1}{6}\right)\right.$	$5\left\vert\left(\frac{1}{6}\right)^2\right.$	$6\left\vert\left(\frac{1}{6}\right)^2\right.$	$7\left\vert\left(\frac{1}{6}\right)^2\right.$	$8\left\vert\left(\frac{1}{6}\right)^2\right.$	$9\left\vert\left(\frac{1}{6}\right)^2\right.$	$10\left\vert\left(\frac{1}{6}\right)^2\right.$
$5\left\vert\left(\frac{1}{6}\right)\right.$	$6\left\vert\left(\frac{1}{6}\right)^2\right.$	$7\left\vert\left(\frac{1}{6}\right)^2\right.$	$8\left\vert\left(\frac{1}{6}\right)^2\right.$	$9\left\vert\left(\frac{1}{6}\right)^2\right.$	$10\left\vert\left(\frac{1}{6}\right)^2\right.$	$11\left\vert\left(\frac{1}{6}\right)^2\right.$
$6\left\vert\left(\frac{1}{6}\right)\right.$	$7\left\vert\left(\frac{1}{6}\right)^2\right.$	$8\left\vert\left(\frac{1}{6}\right)^2\right.$	$9\left\vert\left(\frac{1}{6}\right)^2\right.$	$10\left\vert\left(\frac{1}{6}\right)^2\right.$	$11\left\vert\left(\frac{1}{6}\right)^2\right.$	$12\left\vert\left(\frac{1}{6}\right)^2\right.$

Compute the probability of each roll (the product of the probabilities of each face), and then sum down diagonals (the 6 diagonal is highlighted in yellow).

TEXAS INSTRUMENTS

Rolling Two Dice (again)

	$1\,\big|\left(\frac{1}{6}\right)$	$2\,\big|\left(\frac{1}{6}\right)$	$3\,\big|\left(\frac{1}{6}\right)$	$4\,\big|\left(\frac{1}{6}\right)$	$5\,\big|\left(\frac{1}{6}\right)$	$6\,\big|\left(\frac{1}{6}\right)$
$1\,\big|\left(\frac{1}{6}\right)$	$2\,\big|\left(\frac{1}{6}\right)^2$	$3\,\big|\left(\frac{1}{6}\right)^2$	$4\,\big|\left(\frac{1}{6}\right)^2$	$5\,\big|\left(\frac{1}{6}\right)^2$	$6\,\big|\left(\frac{1}{6}\right)^2$	$7\,\big|\left(\frac{1}{6}\right)^2$
$2\,\big|\left(\frac{1}{6}\right)$	$3\,\big|\left(\frac{1}{6}\right)^2$	$4\,\big|\left(\frac{1}{6}\right)^2$	$5\,\big|\left(\frac{1}{6}\right)^2$	$6\,\big|\left(\frac{1}{6}\right)^2$	$7\,\big|\left(\frac{1}{6}\right)^2$	$8\,\big|\left(\frac{1}{6}\right)^2$
$3\,\big|\left(\frac{1}{6}\right)$	$4\,\big|\left(\frac{1}{6}\right)^2$	$5\,\big|\left(\frac{1}{6}\right)^2$	$6\,\big|\left(\frac{1}{6}\right)^2$	$7\,\big|\left(\frac{1}{6}\right)^2$	$8\,\big|\left(\frac{1}{6}\right)^2$	$9\,\big|\left(\frac{1}{6}\right)^2$
$4\,\big|\left(\frac{1}{6}\right)$	$5\,\big|\left(\frac{1}{6}\right)^2$	$6\,\big|\left(\frac{1}{6}\right)^2$	$7\,\big|\left(\frac{1}{6}\right)^2$	$8\,\big|\left(\frac{1}{6}\right)^2$	$9\,\big|\left(\frac{1}{6}\right)^2$	$10\,\big|\left(\frac{1}{6}\right)^2$
$5\,\big|\left(\frac{1}{6}\right)$	$6\,\big|\left(\frac{1}{6}\right)^2$	$7\,\big|\left(\frac{1}{6}\right)^2$	$8\,\big|\left(\frac{1}{6}\right)^2$	$9\,\big|\left(\frac{1}{6}\right)^2$	$10\,\big|\left(\frac{1}{6}\right)^2$	$11\,\big|\left(\frac{1}{6}\right)^2$
$6\,\big|\left(\frac{1}{6}\right)$	$7\,\big|\left(\frac{1}{6}\right)^2$	$8\,\big|\left(\frac{1}{6}\right)^2$	$9\,\big|\left(\frac{1}{6}\right)^2$	$10\,\big|\left(\frac{1}{6}\right)^2$	$11\,\big|\left(\frac{1}{6}\right)^2$	$12\,\big|\left(\frac{1}{6}\right)^2$

Thus the probability of rolling a 6:

$$\left(\frac{1}{6}\right)^2 + \left(\frac{1}{6}\right)^2 + \left(\frac{1}{6}\right)^2 + \left(\frac{1}{6}\right)^2 + \left(\frac{1}{6}\right)^2 = \frac{5}{36}$$

Texas Instruments

Rolling Two Dice (again)

This technique hasn't saved us any computation effort for the two dice case. It has simplified the algorithm – multiply & add instead of count, search, and add. The real power of the technique is that it easily generalizes to dice with different face counts, and to unfair dice (that is a die which has different probabilities per face).

Rolling Two Dice (again)

Now we roll one fair die with 6 sides, and one loaded die with 4.

Total	Probability	
1	1/6	▬▬
2	1/6	▬▬
3	1/6	▬▬
4	1/6	▬▬
5	1/6	▬▬
6	1/6	▬▬

Total	Probability	
1	2/15	▬
2	2/15	▬
3	9/15	▬▬▬▬▬▬
4	2/15	▬

Texas Instruments

Rolling Two Dice (again)

	$1\left\|\left(\frac{2}{15}\right)\right.$	$2\left\|\left(\frac{2}{15}\right)\right.$	$3\left\|\left(\frac{9}{15}\right)\right.$	$4\left\|\left(\frac{2}{15}\right)\right.$
$1\left\|\left(\frac{1}{6}\right)\right.$	$2\left\|\left(\frac{1}{6}\right)\left(\frac{2}{15}\right)\right.$	$3\left\|\left(\frac{1}{6}\right)\left(\frac{2}{15}\right)\right.$	$4\left\|\left(\frac{1}{6}\right)\left(\frac{9}{15}\right)\right.$	$5\left\|\left(\frac{1}{6}\right)\left(\frac{2}{15}\right)\right.$
$2\left\|\left(\frac{1}{6}\right)\right.$	$3\left\|\left(\frac{1}{6}\right)\left(\frac{2}{15}\right)\right.$	$4\left\|\left(\frac{1}{6}\right)\left(\frac{2}{15}\right)\right.$	$5\left\|\left(\frac{1}{6}\right)\left(\frac{9}{15}\right)\right.$	$6\left\|\left(\frac{1}{6}\right)\left(\frac{2}{15}\right)\right.$
$3\left\|\left(\frac{1}{6}\right)\right.$	$4\left\|\left(\frac{1}{6}\right)\left(\frac{2}{15}\right)\right.$	$5\left\|\left(\frac{1}{6}\right)\left(\frac{2}{15}\right)\right.$	$6\left\|\left(\frac{1}{6}\right)\left(\frac{9}{15}\right)\right.$	$7\left\|\left(\frac{1}{6}\right)\left(\frac{2}{15}\right)\right.$
$4\left\|\left(\frac{1}{6}\right)\right.$	$5\left\|\left(\frac{1}{6}\right)\left(\frac{2}{15}\right)\right.$	$6\left\|\left(\frac{1}{6}\right)\left(\frac{2}{15}\right)\right.$	$7\left\|\left(\frac{1}{6}\right)\left(\frac{9}{15}\right)\right.$	$8\left\|\left(\frac{1}{6}\right)\left(\frac{2}{15}\right)\right.$
$5\left\|\left(\frac{1}{6}\right)\right.$	$6\left\|\left(\frac{1}{6}\right)\left(\frac{2}{15}\right)\right.$	$7\left\|\left(\frac{1}{6}\right)\left(\frac{2}{15}\right)\right.$	$8\left\|\left(\frac{1}{6}\right)\left(\frac{9}{15}\right)\right.$	$9\left\|\left(\frac{1}{6}\right)\left(\frac{2}{15}\right)\right.$
$6\left\|\left(\frac{1}{6}\right)\right.$	$7\left\|\left(\frac{1}{6}\right)\left(\frac{2}{15}\right)\right.$	$8\left\|\left(\frac{1}{6}\right)\left(\frac{2}{15}\right)\right.$	$9\left\|\left(\frac{1}{6}\right)\left(\frac{9}{15}\right)\right.$	$10\left\|\left(\frac{1}{6}\right)\left(\frac{2}{15}\right)\right.$

Here we have a 6 sided die, and a loaded die with 4 sides!

TEXAS INSTRUMENTS

Rolling Two Dice (again)

	$1\,\big|\left(\tfrac{2}{15}\right)$	$2\,\big|\left(\tfrac{2}{15}\right)$	$3\,\big|\left(\tfrac{9}{15}\right)$	$4\,\big|\left(\tfrac{2}{15}\right)$
$1\,\big|\left(\tfrac{1}{6}\right)$	$2\,\big|\left(\tfrac{1}{45}\right)$	$3\,\big|\left(\tfrac{1}{45}\right)$	$4\,\big|\left(\tfrac{1}{10}\right)$	$5\,\big|\left(\tfrac{1}{45}\right)$
$2\,\big|\left(\tfrac{1}{6}\right)$	$3\,\big|\left(\tfrac{1}{45}\right)$	$4\,\big|\left(\tfrac{1}{45}\right)$	$5\,\big|\left(\tfrac{1}{10}\right)$	$6\,\big|\left(\tfrac{1}{45}\right)$
$3\,\big|\left(\tfrac{1}{6}\right)$	$4\,\big|\left(\tfrac{1}{45}\right)$	$5\,\big|\left(\tfrac{1}{45}\right)$	$6\,\big|\left(\tfrac{1}{10}\right)$	$7\,\big|\left(\tfrac{1}{45}\right)$
$4\,\big|\left(\tfrac{1}{6}\right)$	$5\,\big|\left(\tfrac{1}{45}\right)$	$6\,\big|\left(\tfrac{1}{45}\right)$	$7\,\big|\left(\tfrac{1}{10}\right)$	$8\,\big|\left(\tfrac{1}{45}\right)$
$5\,\big|\left(\tfrac{1}{6}\right)$	$6\,\big|\left(\tfrac{1}{45}\right)$	$7\,\big|\left(\tfrac{1}{45}\right)$	$8\,\big|\left(\tfrac{1}{10}\right)$	$9\,\big|\left(\tfrac{1}{45}\right)$
$6\,\big|\left(\tfrac{1}{6}\right)$	$7\,\big|\left(\tfrac{1}{45}\right)$	$8\,\big|\left(\tfrac{1}{45}\right)$	$9\,\big|\left(\tfrac{1}{10}\right)$	$10\,\big|\left(\tfrac{1}{45}\right)$

The probability of rolling a 6: $\frac{1}{45} + \frac{1}{45} + \frac{1}{10} + \frac{1}{45} = \frac{1}{6} = 0.16666\overline{6}$.

Rolling Two Dice (again)

Total	Probability	
2	2/90	�
3	4/90	▬
4	13/90	▬▬▬▬▬▬
5	15/90	▬▬▬▬▬▬▬
6	15/90	▬▬▬▬▬▬▬
7	15/90	▬▬▬▬▬▬▬
8	13/90	▬▬▬▬▬▬
9	11/90	▬▬▬▬▬
10	2/90	▬

This algorithm in R

d1 and d2 hold the probabilities for each die, and probs is the
resulting combined probability table.

```
1  d1      <- rep(1/6, 6);
2  d2      <- c(2/15, 2/15, 9/15, 2/15)
3  ptab    <- d1 %*% t(d2)
4  stab    <- row(ptab) + col(ptab)
5  rrange  <- min(stab):max(stab)
6  probs   <- sapply(rrange,
7                     function(r) sum(ptab[stab==r]))
```

Note: This is code is for illustration. It is super simple to understand, but is
kinda silly – like explicitly constructing the matrices..

TEXAS INSTRUMENTS

OK. That loaded dice thing was interesting, but how will that make it faster to compute the probability of throwing a whole hand full of dice?

TEXAS INSTRUMENTS

Applying What We Have Learned

Remember this? The probability table for throwing two fair dice..

Total	Ways To Roll	Probability	
2	1	1/36	▬
3	2	2/36	▬▬
4	3	3/36	▬▬▬
5	4	4/36	▬▬▬▬
6	5	5/36	▬▬▬▬▬
7	6	6/36	▬▬▬▬▬▬
8	5	5/36	▬▬▬▬▬
9	4	4/36	▬▬▬▬
10	3	3/36	▬▬▬
11	2	2/36	▬▬
12	1	1/36	▬

Applying What We Have Learned

How about we throw one fair 6 sided die, and one wonky die with the same probability rules as the pair of dice we did before?

Total	Probability	
1	1/6	▬▬▬
2	1/6	▬▬▬
3	1/6	▬▬▬
4	1/6	▬▬▬
5	1/6	▬▬▬
6	1/6	▬▬▬

Total	Ways To Roll	Probability	
2	1	1/36	▬
3	2	2/36	▬▬
4	3	3/36	▬▬▬
5	4	4/36	▬▬▬▬
6	5	5/36	▬▬▬▬▬
7	6	6/36	▬▬▬▬▬▬
8	5	5/36	▬▬▬▬▬
9	4	4/36	▬▬▬▬
10	3	3/36	▬▬▬
11	2	2/36	▬▬
12	1	1/36	▬

TEXAS INSTRUMENTS

This idea in code

Die face probabilities in `daDie`, and number of dice in `numDi`. The resulting combined probability table ends up in `probs`.

```
1  numDi  <- 15
2  daDie  <- rep(1/6, 6);
3  probs  <- daDie
4  for(i in 2:numDi) {
5    ptab   <- daDie %*% t(probs)
6    stab   <- row(ptab) + col(ptab)
7    rrange <- 1:max(stab)
8    probs  <- sapply(rrange,
9                     function(r) sum(ptab[stab==r]))
10   names(probs) <- rrange
11 }
```

Note: Again... This is code is for illustration. It is super simple to understand, but is kinda silly – like explicitly constructing the matrices..

TEXAS INSTRUMENTS

Probability Distribution For Rolling 15 Dice (fair, 6 sided dice)

TEXAS INSTRUMENTS
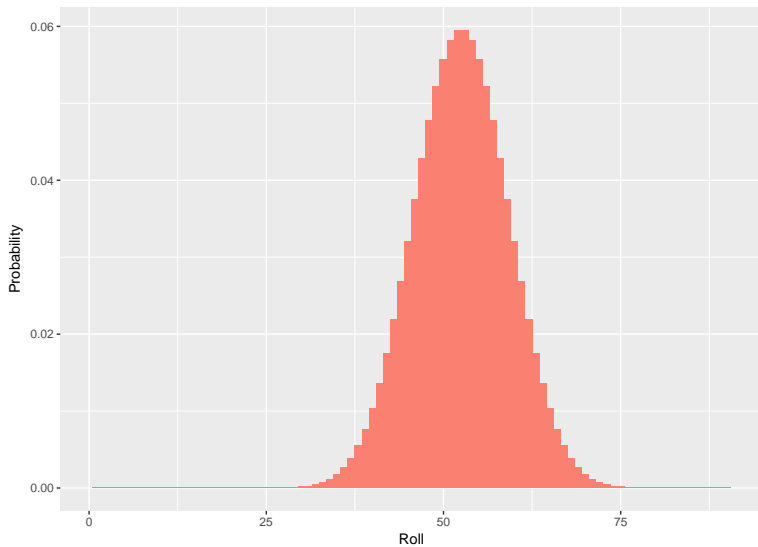
Part V: Computing Risk

Weren't we talking about slot:core ratios?

That was fun! How do I set my slot:core ratio again?

TEXAS INSTRUMENTS

Job Efficiency Histogram

TEXAS INSTRUMENTS
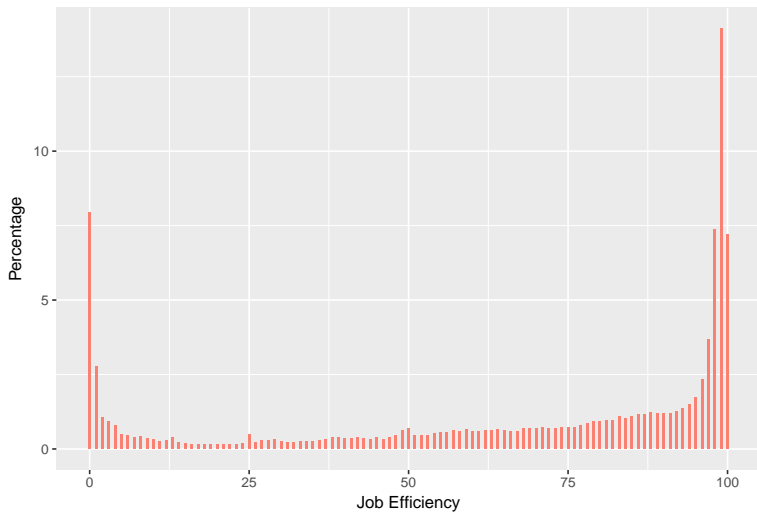
A probability problem

The Question

Randomly pick n single threaded jobs, and place them on an n core compute server. What is the utilization of that compute server?

TEXAS INSTRUMENTS

A probability problem

What kind of answer can we provide?

We can't know what the utilization will be, but we can approximate the server utilization given any n job efficiency values and report the probability of all the possibilities.

A probability problem

How?

Yep. We are going to think about that graph as if it were a loaded, 100 sided die!
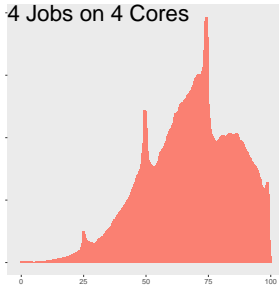
1 Job on 1 Core
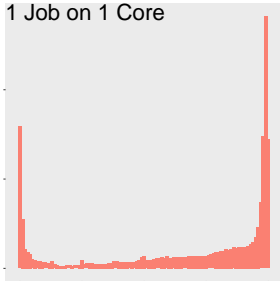
2 Jobs on 2 Cores
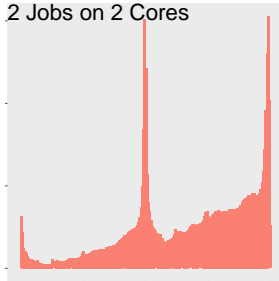
3 Jobs on 3 Cores
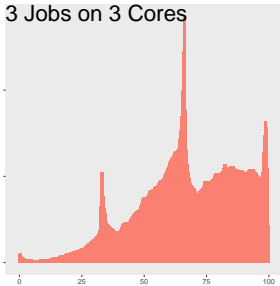
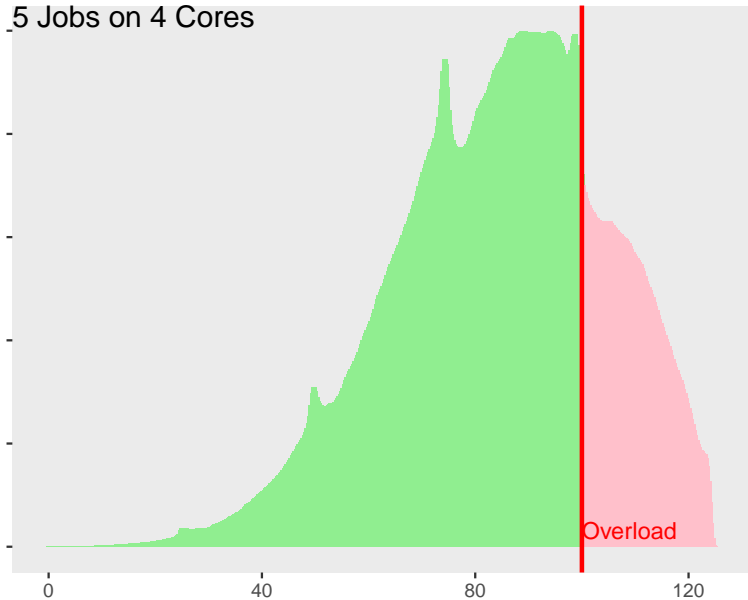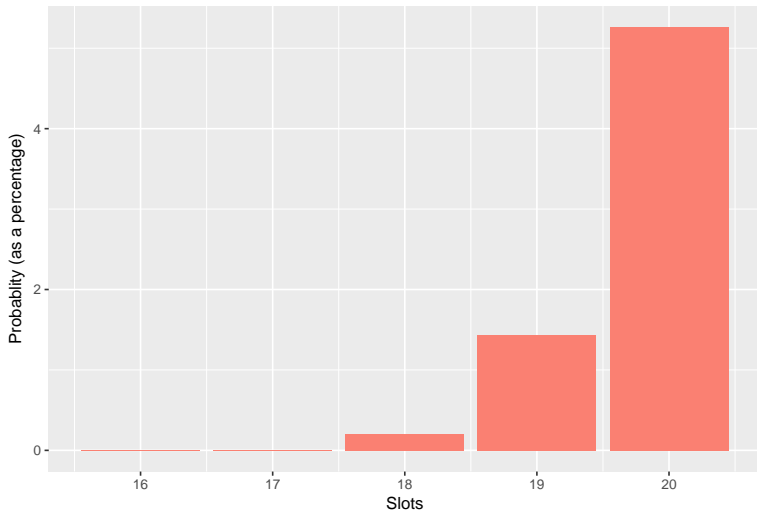4 Jobs on 4 Cores

TEXAS INSTRUMENTS

Probablity Of Overaload (16 core server)

Probability (as a percentage) vs Slots

TEXAS INSTRUMENTS

Probability is not Risk

What is risk?

The classical actuarial definition is that risk is the product of the probability of an even and the cost of that event.

$$R = p \cdot C$$

Probability is not Risk

Cost?

It can be difficult to quantify cost for many events; however, sometimes we can proxy cost with business impact measured in other terms.

Probability is not Risk

How about "runtime" as a cost?

In general if we oversubscribe a system by a particular percentage with very efficient jobs, we can expect jobs to potentially run about that same percentage slower. This isn't always true on Linux, but it is a reasonable place to start.

Probability is not Risk

How do you phrase this to leaders?

We can get a 20% improvement in throughput at a cost of around %4 of our jobs running about %15 slower.

Probability is not Risk

How do you pick a risk level?

Simply present the risk levels of the various options (perhaps with a graph like last one we looked at) to your leadership team, and ask them to select a level of risk they are comfortable with.

I just need to think about dice?

Our probably experiment was pretty simple, but it actually works for may sites. So if you are lucky, it may be all you need. For TI we had to also take into consideration

- ▶ Queue structure & LSF scheduling parameters
- ▶ Job arrival & resource requirement patterns

TEXAS INSTRUMENTS

How do I check if my model works?

One of the most effective ways is to compute a theoretical server utilization at your current slot:core ration, and then compare it to a histogram of real server utilization.

Texas Instruments

FAQ

How do I check if my model works?

Simulation! i.e. run your new cluster configuration through a cluster simulator, and simply observe the behavior. If it matches your model, then you probably have a good model.

How do I check if my model works?

Simulation! i.e. run your new cluster configuration through a cluster simulator, and simply observe the behaviour. If it matches your model, then you probably have a good model.

FAQ

Risk mitigation?

Job scheduling load thresholds based on true CPU run queue lengths and CPU utilization will go a long way towards keeping systems from getting overloaded.

TEXAS INSTRUMENTS

Thank You!