**Got It!**

**1. Parallel computing**

Good job! Now, one term that's commonly used in data engineering is parallel computing, sometimes also called parallel processing.

**2. Parallel computing**

Parallel computing forms the basis of almost all modern data processing tools. It is important mainly for memory concerns, but also for processing power. When big data processing tools perform a processing task, they split it up into several smaller subtasks. These subtasks are then distributed over several computers.

**3. Understanding parallel computing**

Let's look at an analogy. Let's say you're running a music merchandise shop and need to get a batch of 1000 t-shirts folded.

**4. Understanding parallel computing**

Your senior sales assistant folds 100 shirts in 15 minutes.

**5. Understanding parallel computing**

Junior sales assistants typically take 30 minutes.

1. 1 Emojis by Mohamed Hassan

**6. Understanding parallel computing**

If just one sales assistant can work at a time, it's obvious you'd have to choose the quickest one to finish the job.

**7. Understanding parallel computing**

However, if you can split the batch in 250 shirts each, having 4 junior employees working in parallel is faster;

**8. Understanding parallel computing**

they will finish in 1 hour and 15 minutes,

**9. Understanding parallel computing**

when it would take 2 hours and 30 minutes for your senior employee to finish.

**10. Benefits and risks of parallel computing**

A similar thing happens for big data processing tasks. In this case, employees would be processing units. One benefit of having multiple processing units is the extra processing power itself. Another benefit of parallel computing for big data relates to memory. Instead of needing to load all of the data in one computer's memory, you can partition the data and load the subsets into memory of different computers. That means the memory footprint per computer is relatively small. There can be some disadvantages to parallel computing though. Moving data incurs a cost. What's more, splitting a task into subtasks and merging the results of the subtasks back into one final result requires some communication between processes, which takes some additional time. So if the gains of splitting into subtasks are minimal, it may not be worth taking that risk.

**11. Understanding parallel computing**

Going back to our t-shirt folding analogy,

**12. Understanding parallel computing**

separating the t shirts into four equal piles for each of the sales assistants may take 10 minutes,

**13. Understanding parallel computing**

and collecting the four piles of folded t-shirts back together may take another 5 minutes. So it actually took them one hour and thirty minutes, instead of one hour and fifteen, to fold all the t-shirts.

**14. Parallel computing at Spotflix**

At Spotflix,

**15. Parallel computing at Spotflix**

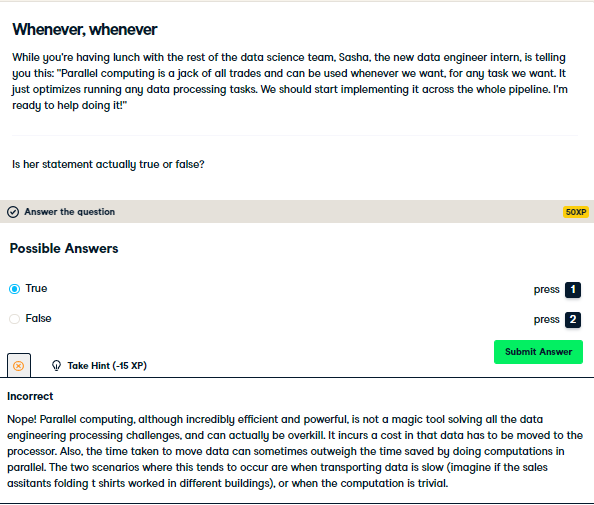
we use parallel computing to convert songs from lossless format to dot ogg. It prevents us from having to load all the new songs in one computer, and to benefit from extra processing power to run the conversion scripts.

**16. Summary**

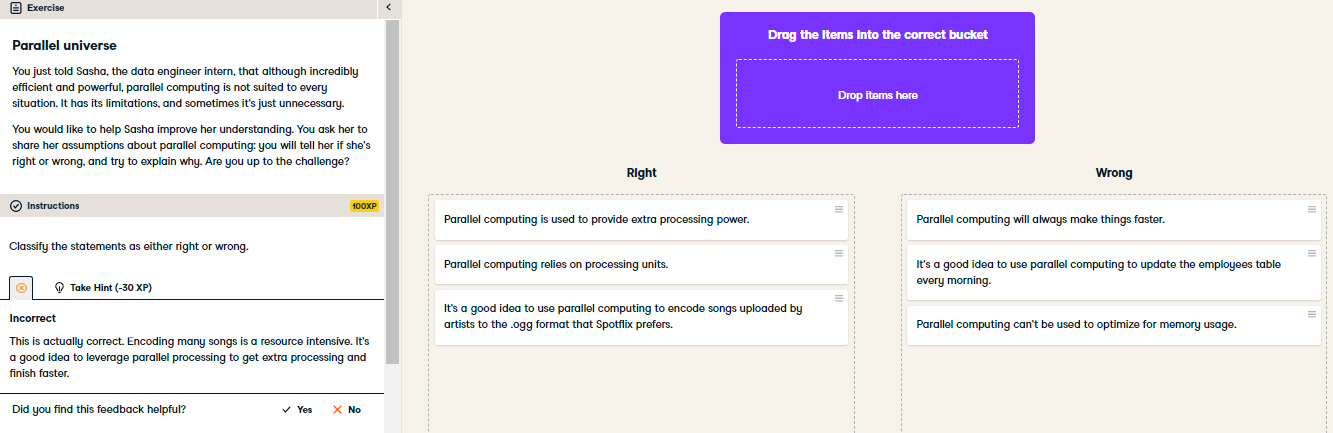
Alright, now you know the benefits and risks of parallel computing, and how it is implemented at Spotflix.

**17. Let's practice!**

Let's practice!



Yup! This statement is wrong. It's OK: interns are here to learn, and we appreciate her passion and proactiveness!



Great job helping Sasha improve. You're really getting the hang of this!