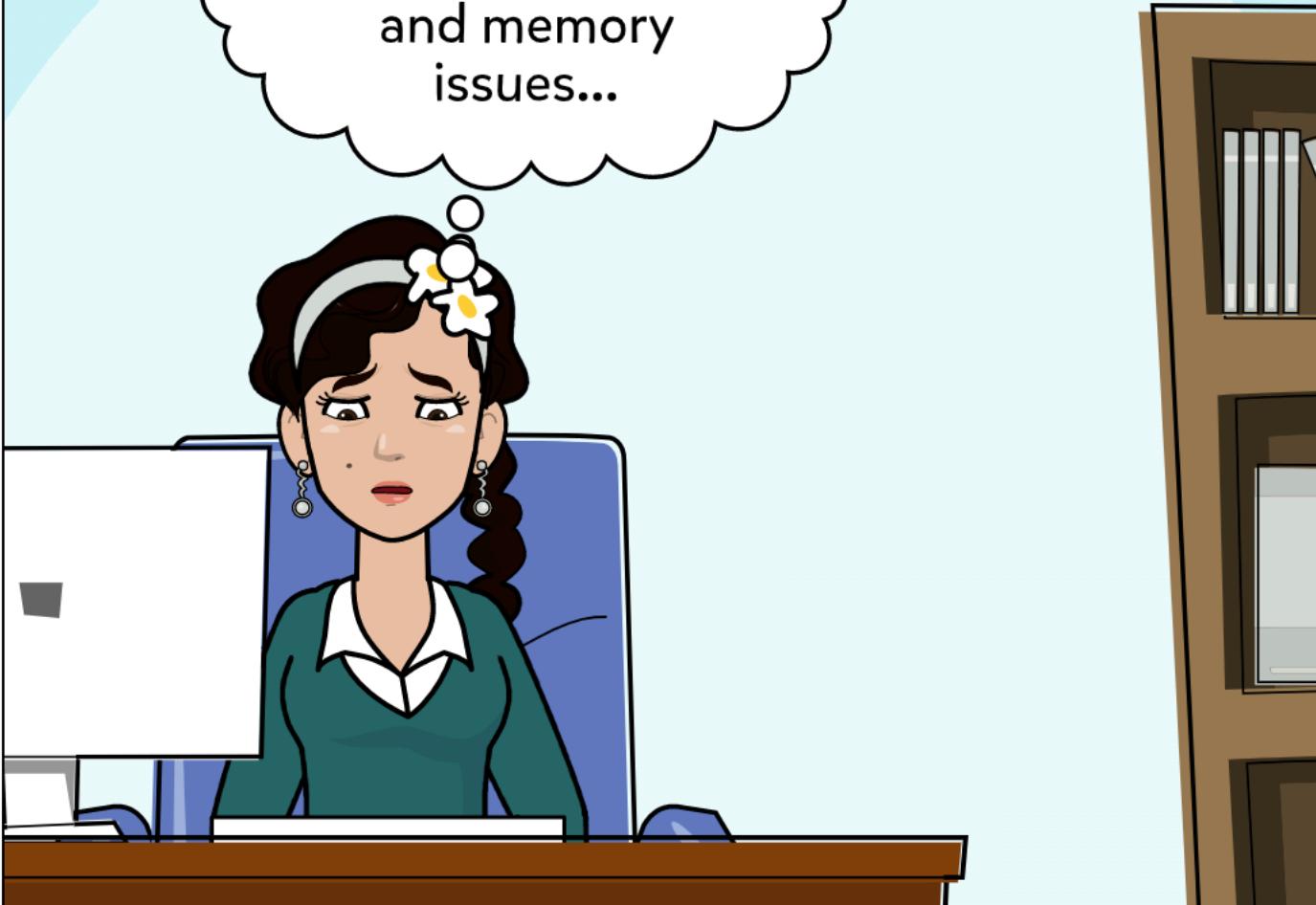


Tea-time story

Sharmili Srinivasan

Jaya is a Data Engineer... She was working on a deadline and was stuck with an issue...

Not sure
how to solve this
long-running process
and memory
issues...





Not now
grandma! Stuck
with a work to be
completed by
tonight...

Dear,
It's tea
time!



Aww.. Guess
a story will be a
good way to clear
my mind ...



Pff... As a kid,
you used to
badger me for
stories over
tea! I miss
those days...

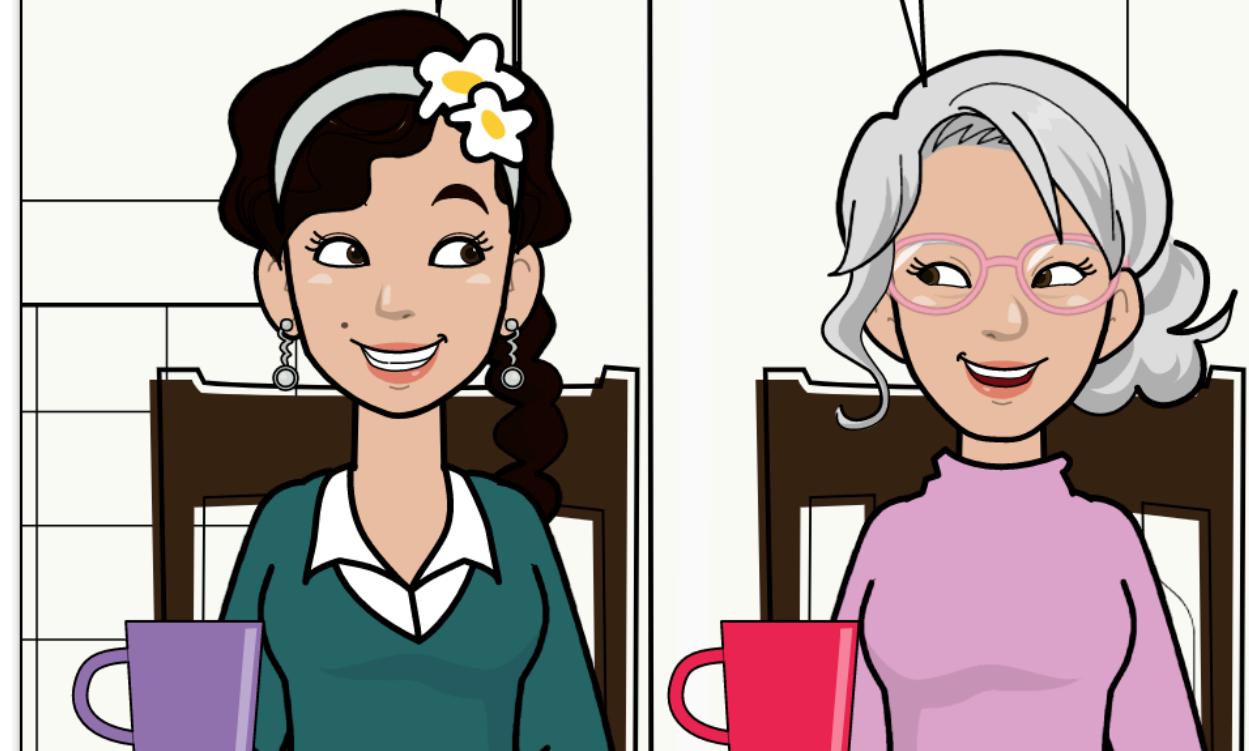
Alright,
granny... Only if
it's from my favorite
Akbar-Birbal
stories...



During tea time...

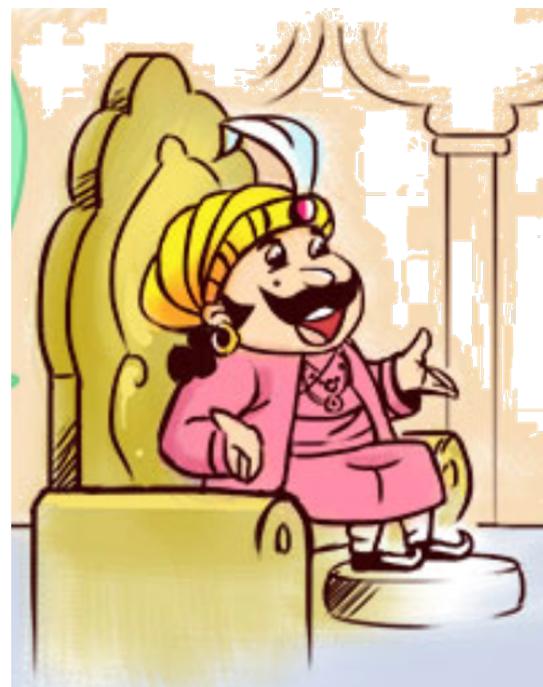
Story!
Story!

Ha!
Here you go..
Once upon a
time...



Once upon a time...

Birbal, In today's party, I would like to serve guests with 100 different sweets. Make it ready by evening!



Birbal plans to consult his friend – who is head of chain of restaurants!





Birbal, Could you
please specify
number of chefs and
stoves required?

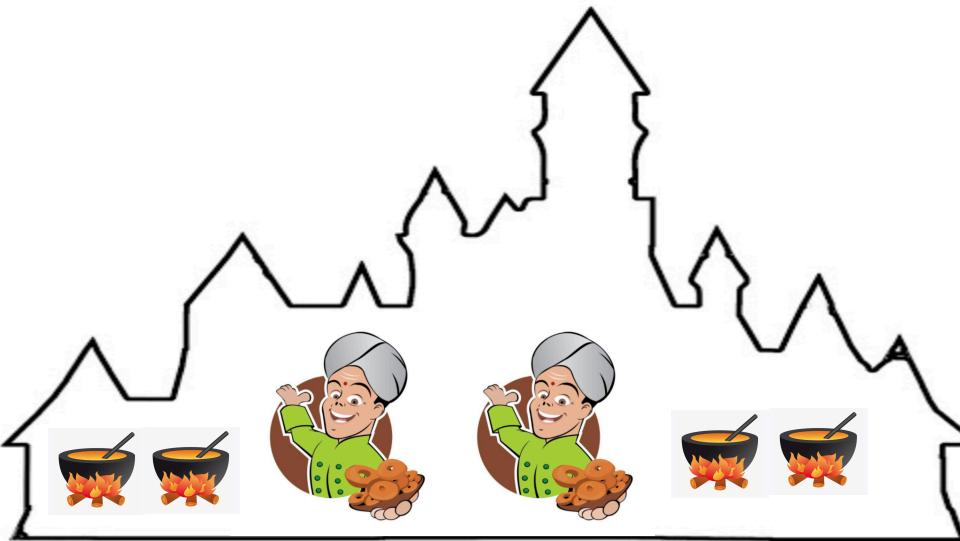
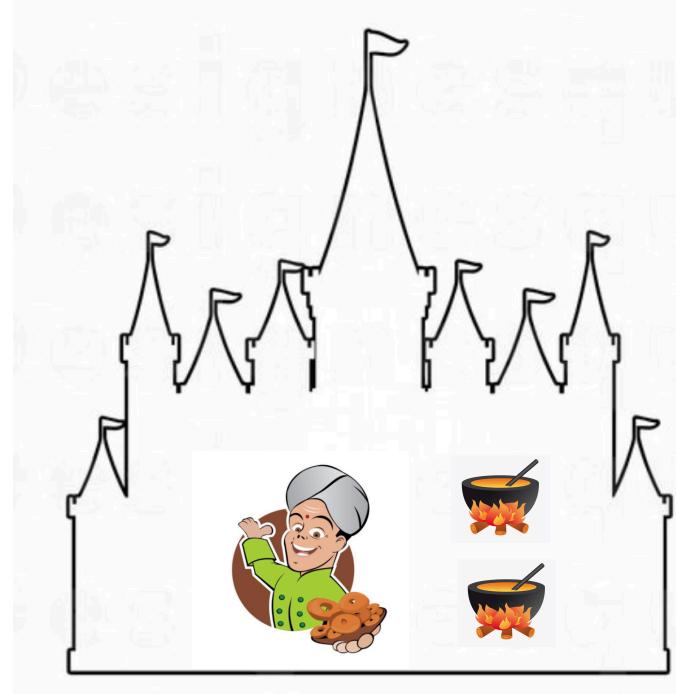
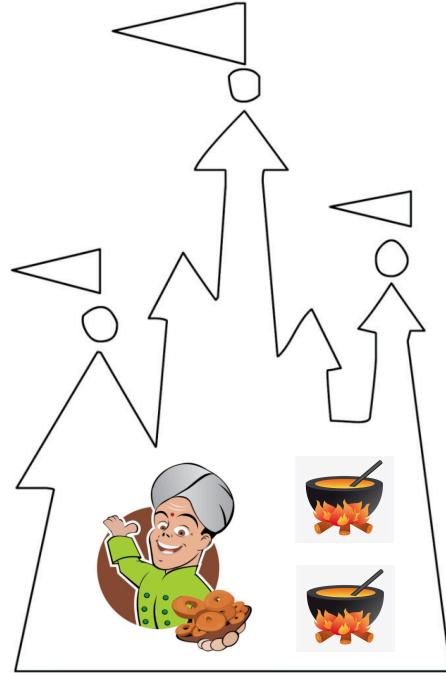
Guess 4
chefs with 2
stoves each
should
work!

Sure!
Here are the chef
details.
You must provide
them with the
recipes and raw
materials.

I can do
that!
Thanks!



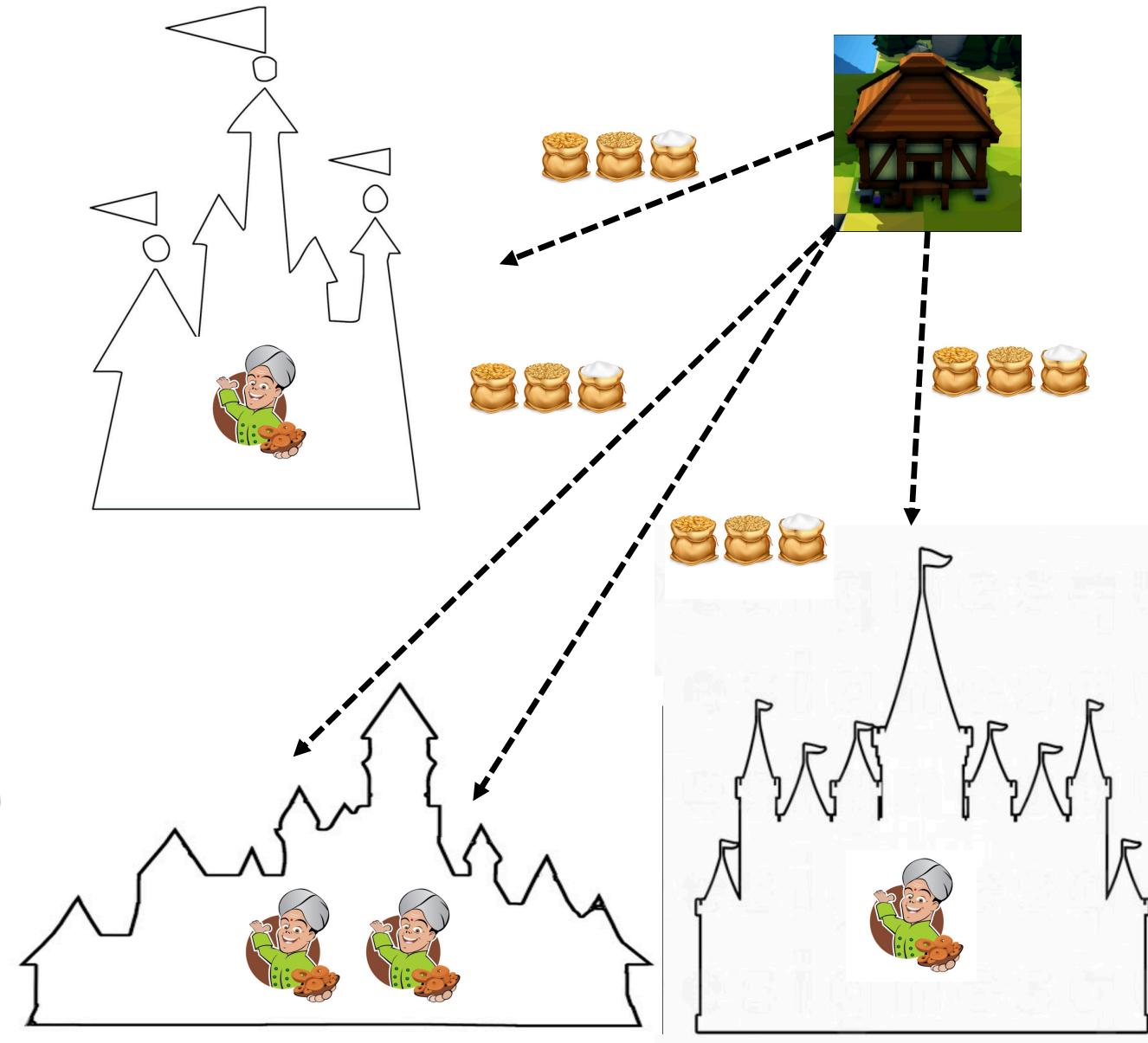
Birbal got 4 chefs from 3 different restaurants with 2 stoves per chef!



Birbal sent the location of the granary and recipe to each chef!



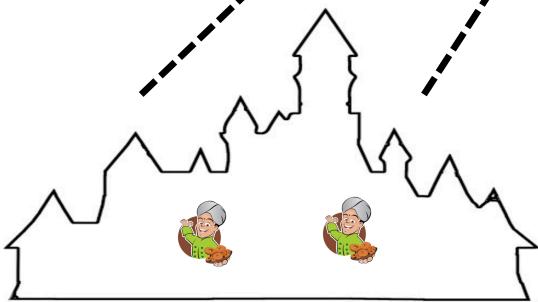
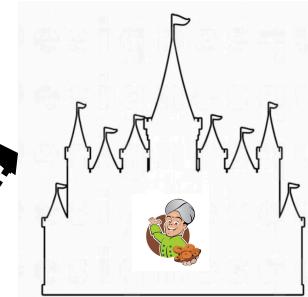
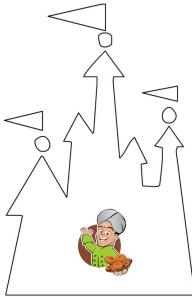
Each chef got their share of raw-materials!



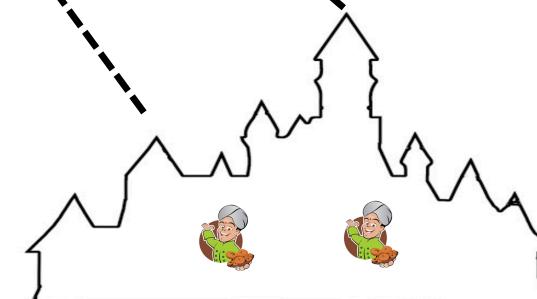
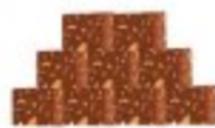
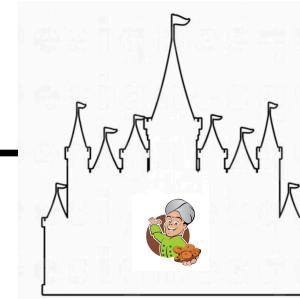
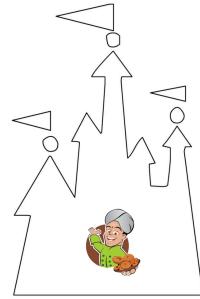
Sweets were prepared!



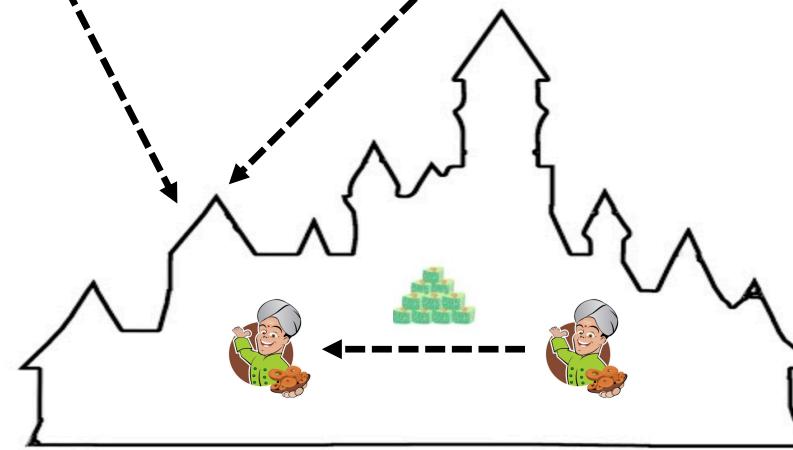
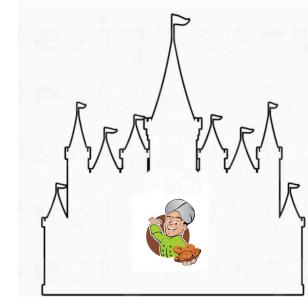
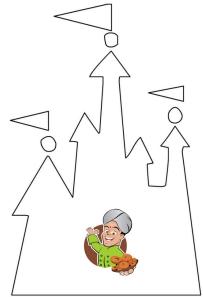
Sorted and packed!



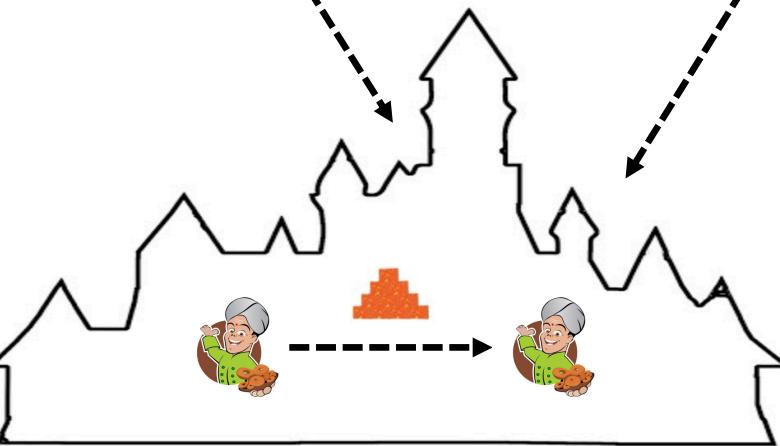
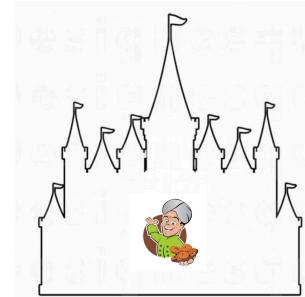
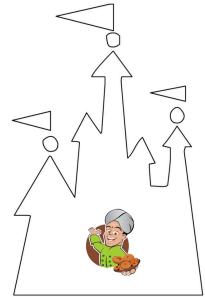
Sorted and packed!



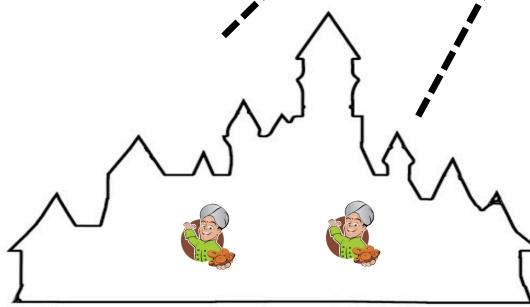
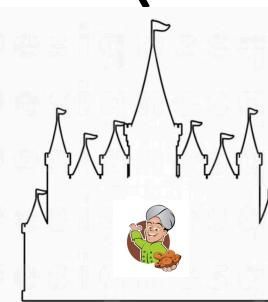
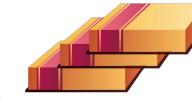
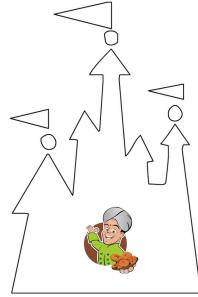
Sorted and packed!



Sorted and packed!



Sweet boxes delivered!



Akbar was very pleased with Birbal

Birbal, I am very
happy with your
service.
Here you go!

Thank you, my
lord!





And that's
how Birbal helped
Akbar with his wisdom
once again!



Thank you
grandma! You
have no idea
how this story
helps me solve a
work problem..
Yayyy!!

Not
sure I
understand! But,
happy it was
helpful!



Surprised
that a simple story
can solve a critical
problem!

What did Jaya learn?

Let's re-visit.. But now in technical jargons...

Developer has a job that cannot potentially run in a single machine

*Job: 100 different sweets
Time: By evening*

Developer

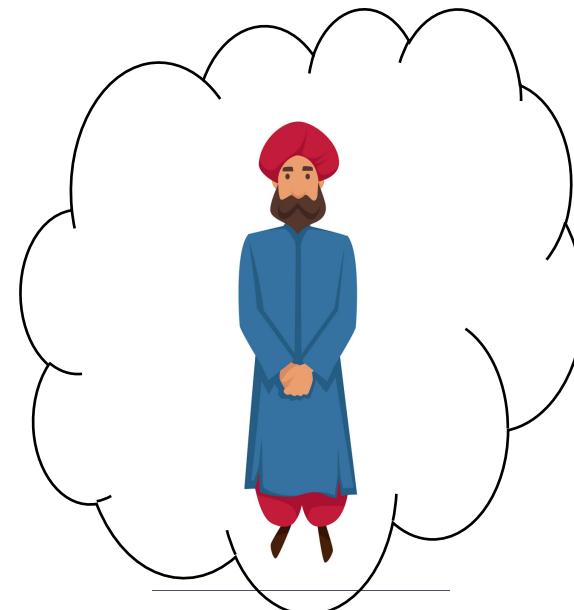


Spark Driver



Spark Driver gets help from Cluster Manager

Cluster Manager



Spark Driver



Cluster Manager appoints Application Master to manage this job

Cluster Manager



Application Master



Spark Driver negotiates for resources with Application Master

Note:

Application master is just a *coordinator*.

Resources are *owned* by Cluster Manager.

He must request the resources from Cluster Manager only!

Application Master



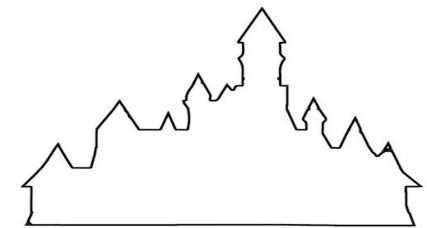
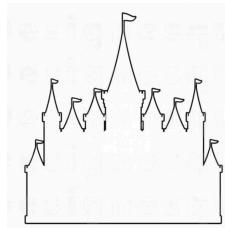
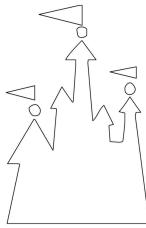
Spark Driver



Resources required:
4 chefs with 2
stoves each

Restaurant: Physical worker; Cook: Executor; Stove: Cores !

Workers



Executors



Cores



Spark Driver sends the code and data location to every executor!

Data location



Code



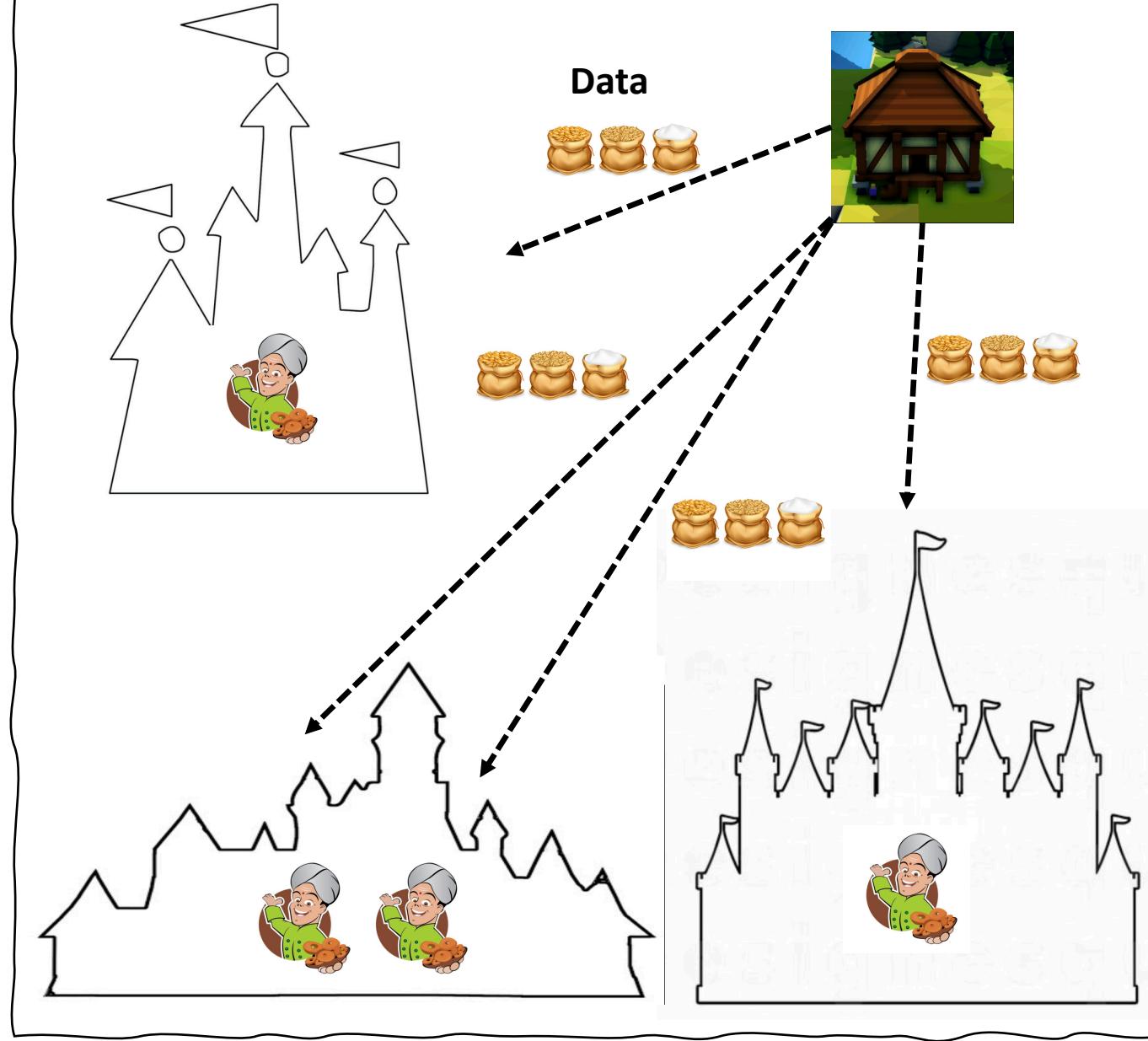
Spark Driver



Executor



Each executor reads data from data-store!



Tasks are executed in the cores of the executors

Executor



Data



Cores

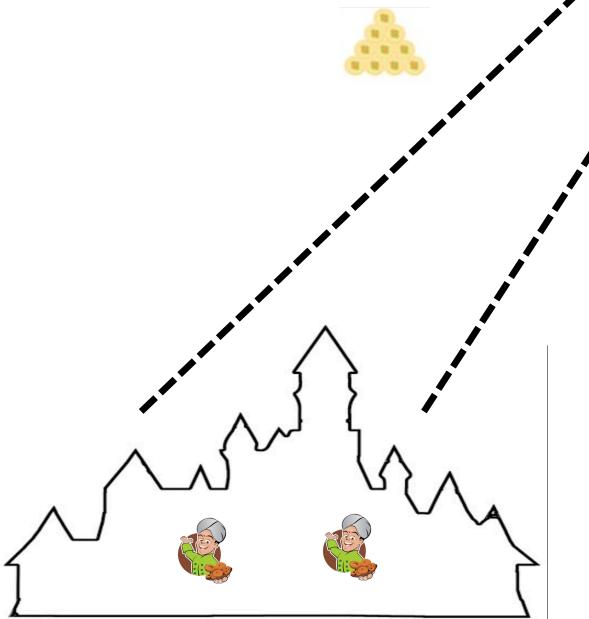
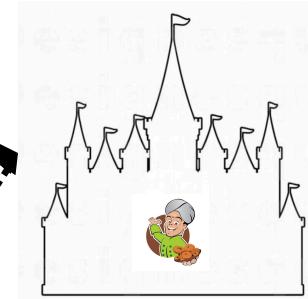
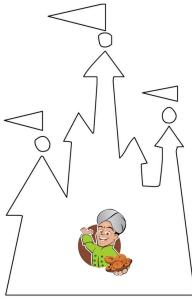
Code



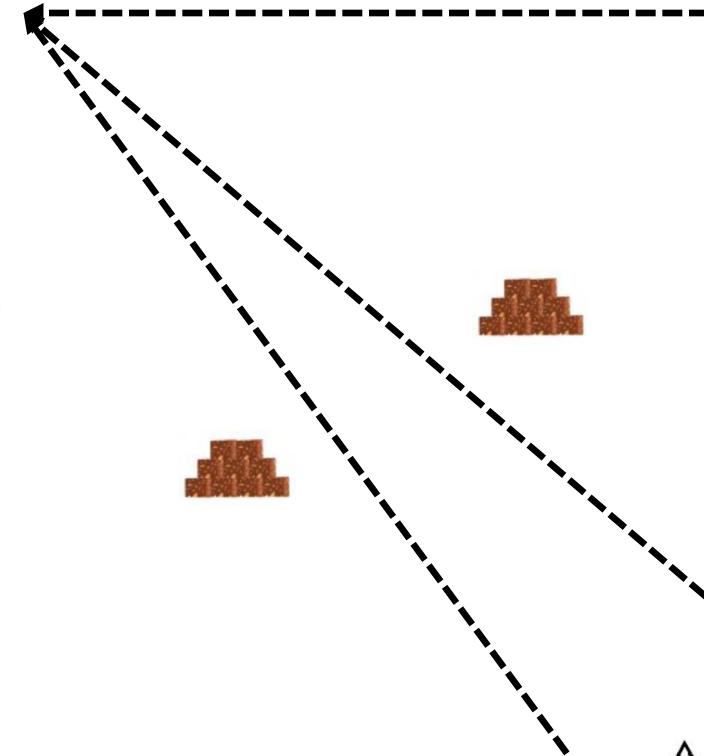
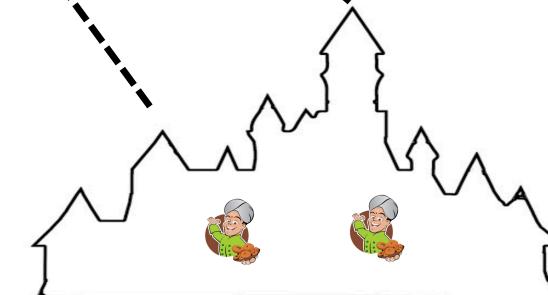
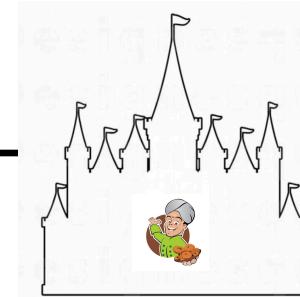
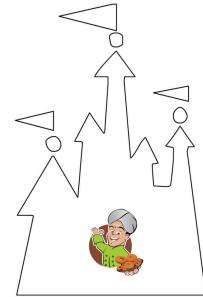
Stage outputs



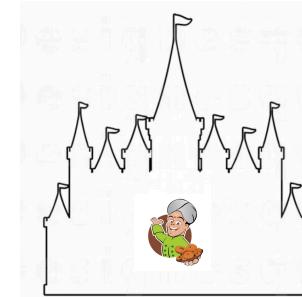
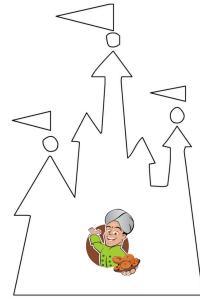
Shuffle in progress



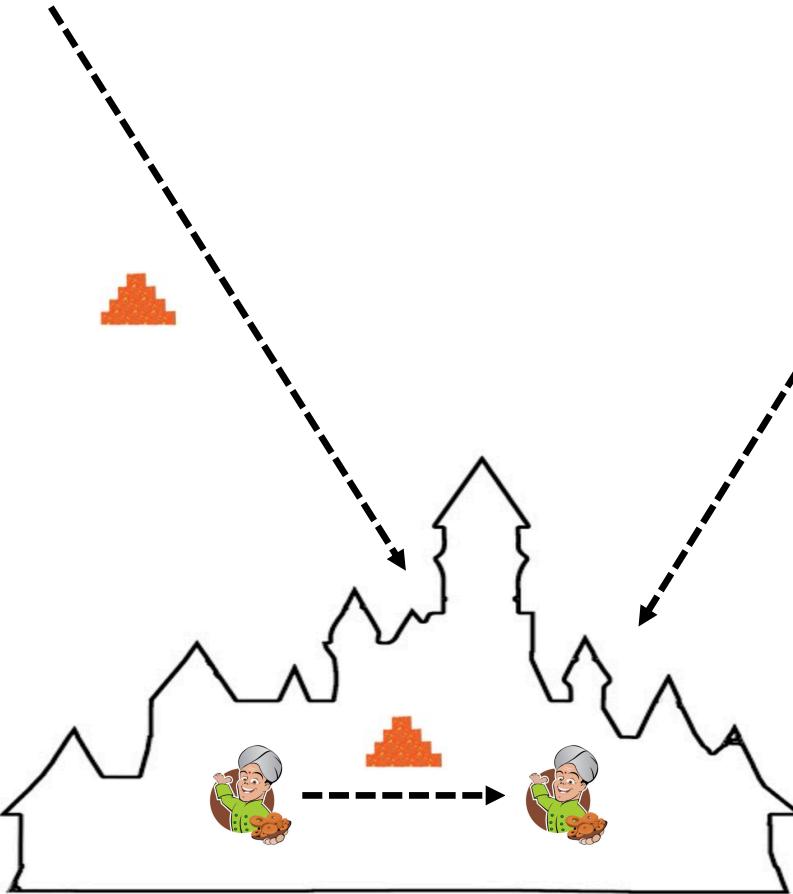
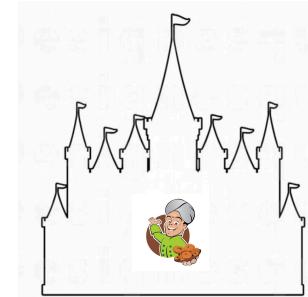
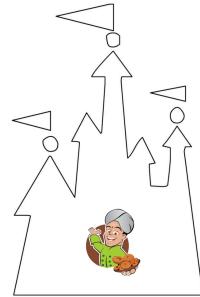
Shuffle in progress



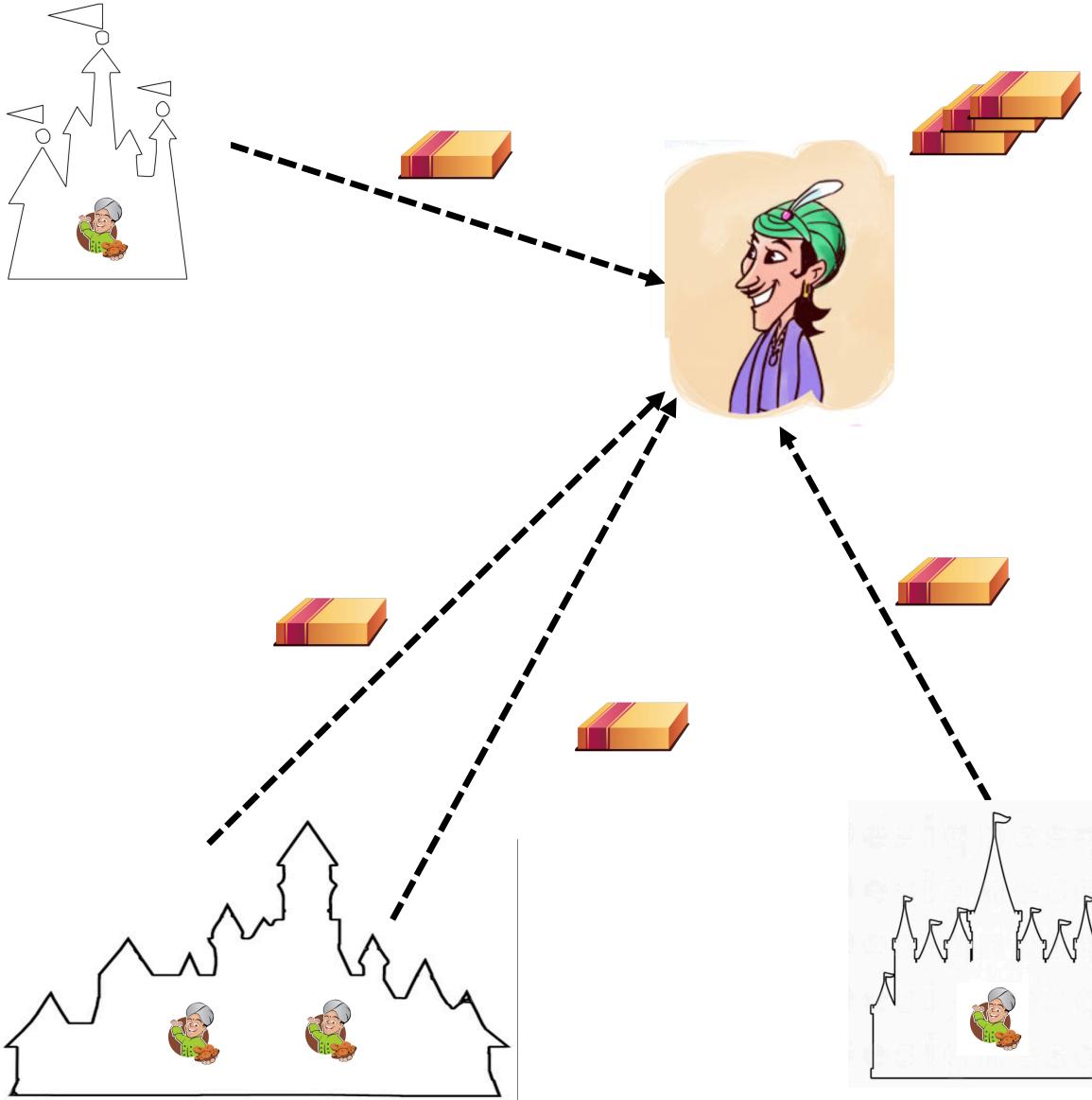
Shuffle in progress



Shuffle in progress



Action (e.g. collect()) executed!



Job Completed!



Thank you!