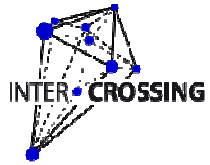


Nispero

Marina Manrique

2013-08-28

Era7 bioinformatics

ohnosequences!

content

- What is this?
- Architecture and players
- Nispero + Statika
- Why Nispero is cool for NGS data analysis?
- Hands-on Nispero

What is this?

loquats

INTER•CROSSING
Era7 bioinformatics
ohnosequences!

What is this?

A component to scale independent tasks

Basic building block to implement distributed systems

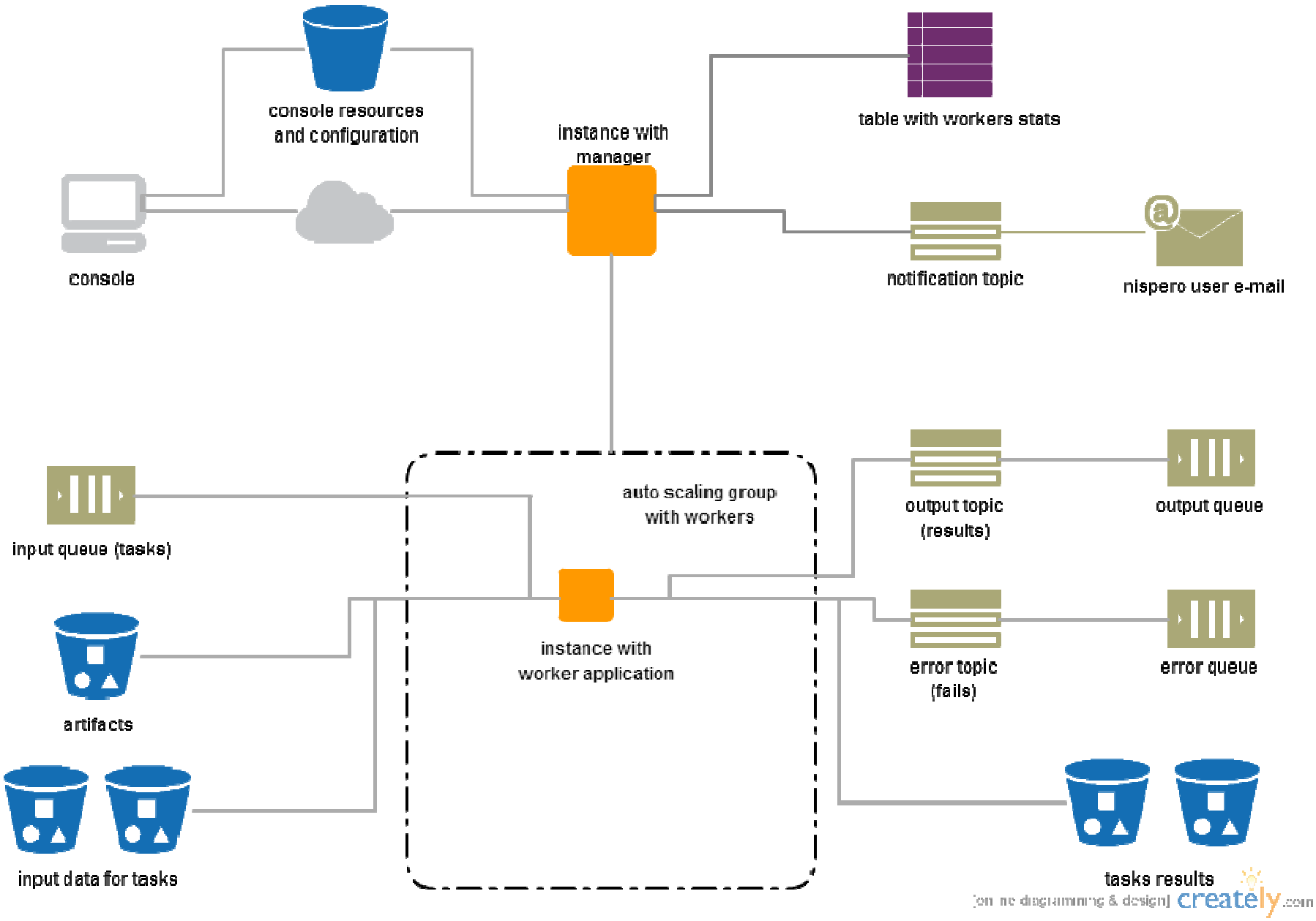
Era7 bioinformatics

ohnosequences!

Architecture and players

Era7 bioinformatics

ohnosequences!


console resources
and configuration
instance with
manager
table with workers stats
console
notification topic
nispero user e-mail
auto scaling group
with workers
input queue (tasks)
output topic
(results)
output queue
instance with
worker application
artifacts
error topic
(fails)
error queue
input data for tasks
tasks results
Online diagramming & design creately.com

Nispero + Statika

You need to configure the workers!!

Era7 bioinformatics

ohnosequences!

Why Nispero is cool for NGS data analysis?

Era7 bioinformatics

ohnosequences!

Why Nispero is cool for NGS data analysis?

parallel automatic tasks
horizontal scaling

Era7 bioinformatics
ohnosequences!

Why Nispero is cool for NGS data analysis?

parallel automatic tasks
horizontal scaling

- Easy to use
- Easy to reuse
- Robust

Era7 bioinformatics
ohnosequences!

Why Nispero is cool for NGS data analysis?

usual parallel tasks in NGS data analysis

Era7 bioinformatics

ohnosequences!

Why Nispero is cool for NGS data analysis?

usual parallel tasks in NGS data analysis

- QA
- Reads Preprocessing: trimming, filtering
- Blast

Why Nispero is cool for NGS data analysis?

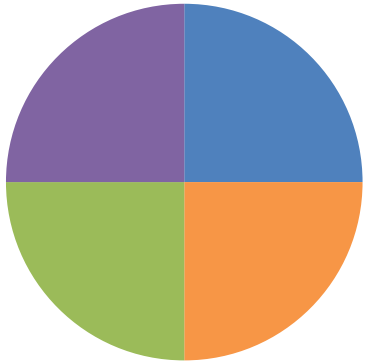
Samples are parallel

Reads are parallel

Era7 bioinformatics

ohnosequences!

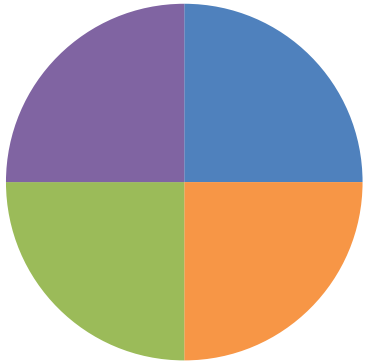
Why Nispero is cool for NGS data analysis?



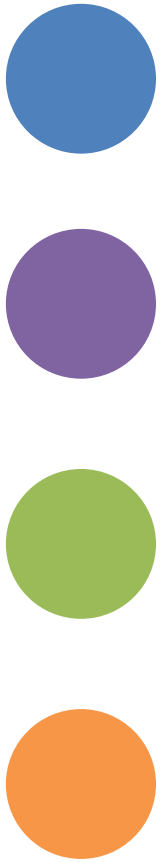
100Gb
X min

Era7 bioinformatics
ohnosequences!

Why Nispero is cool for NGS data analysis?

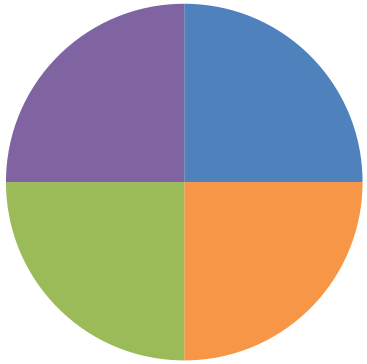


100Gb
X min

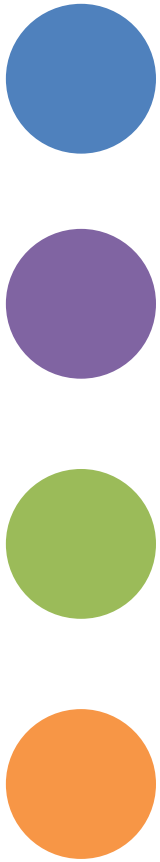


X/4 min

Why Nispero is cool for NGS data analysis?



100Gb
X min



Samples

Reads

~X/16 min

Era7 bioinformatics
ohnosequences!

Why Nispero is cool for NGS data analysis?

horizontal scaling

Era7 bioinformatics

ohnosequences!

Why Nispero is cool for NGS data analysis?

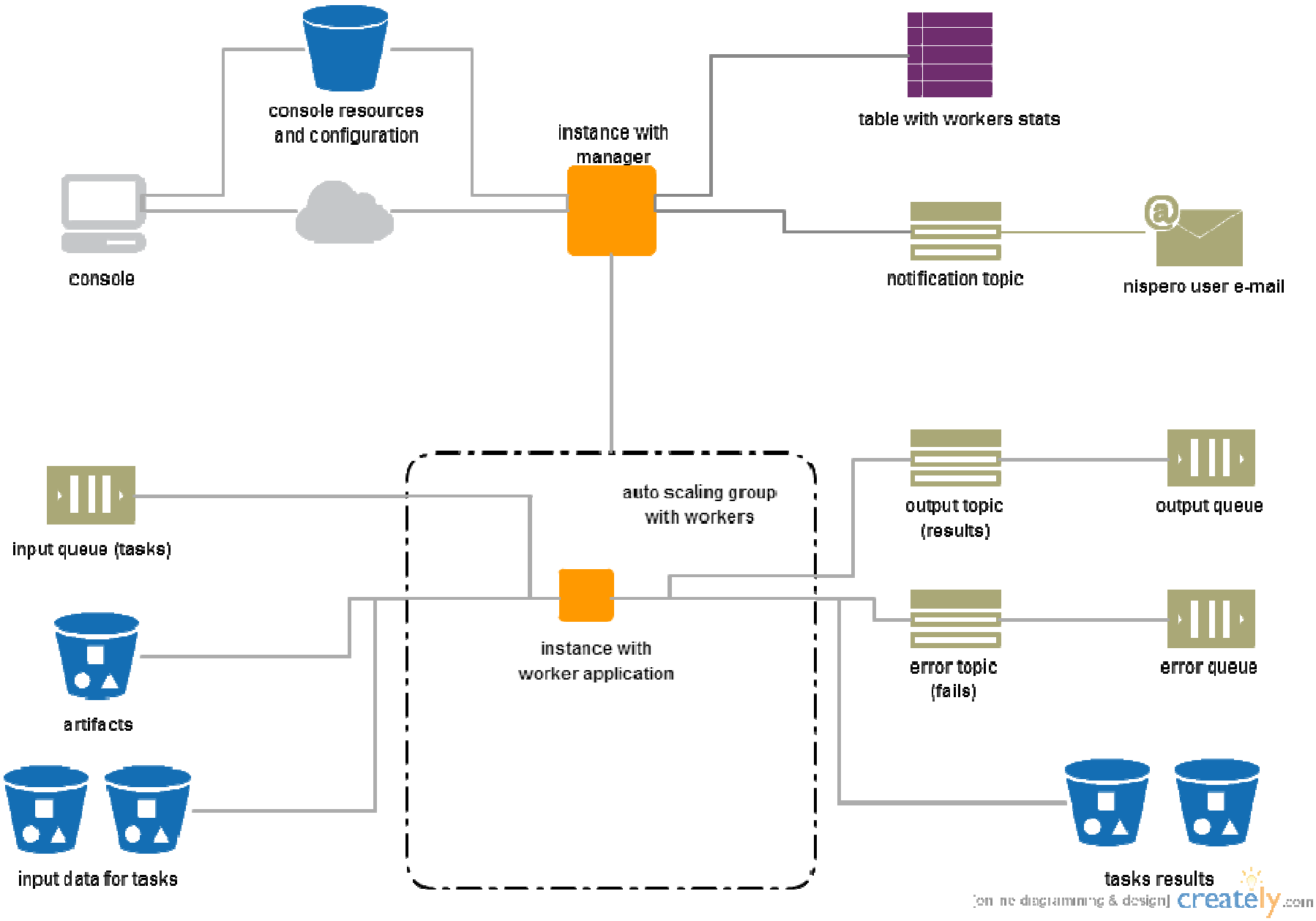
horizontal scaling

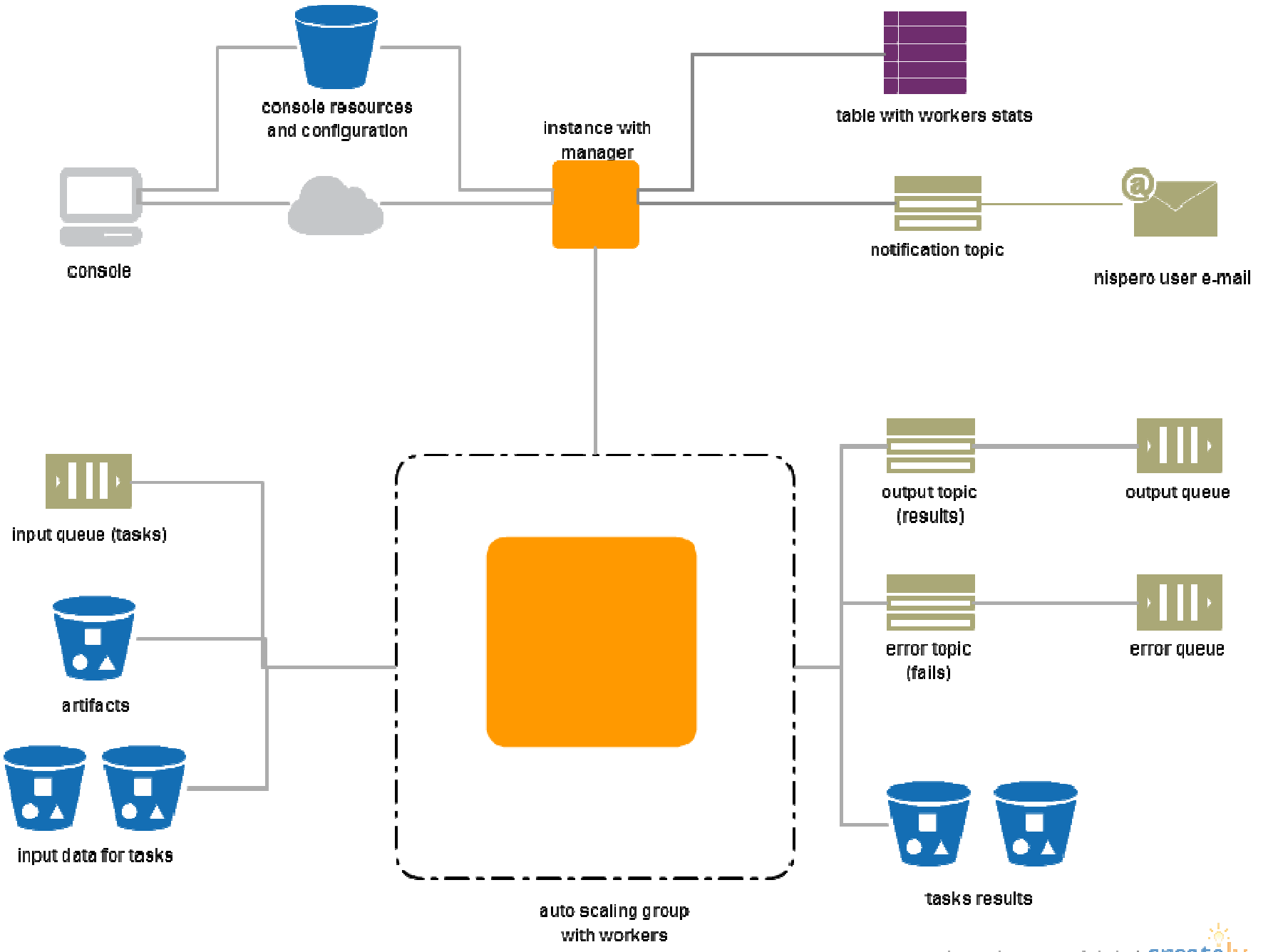
limited number of different instances types

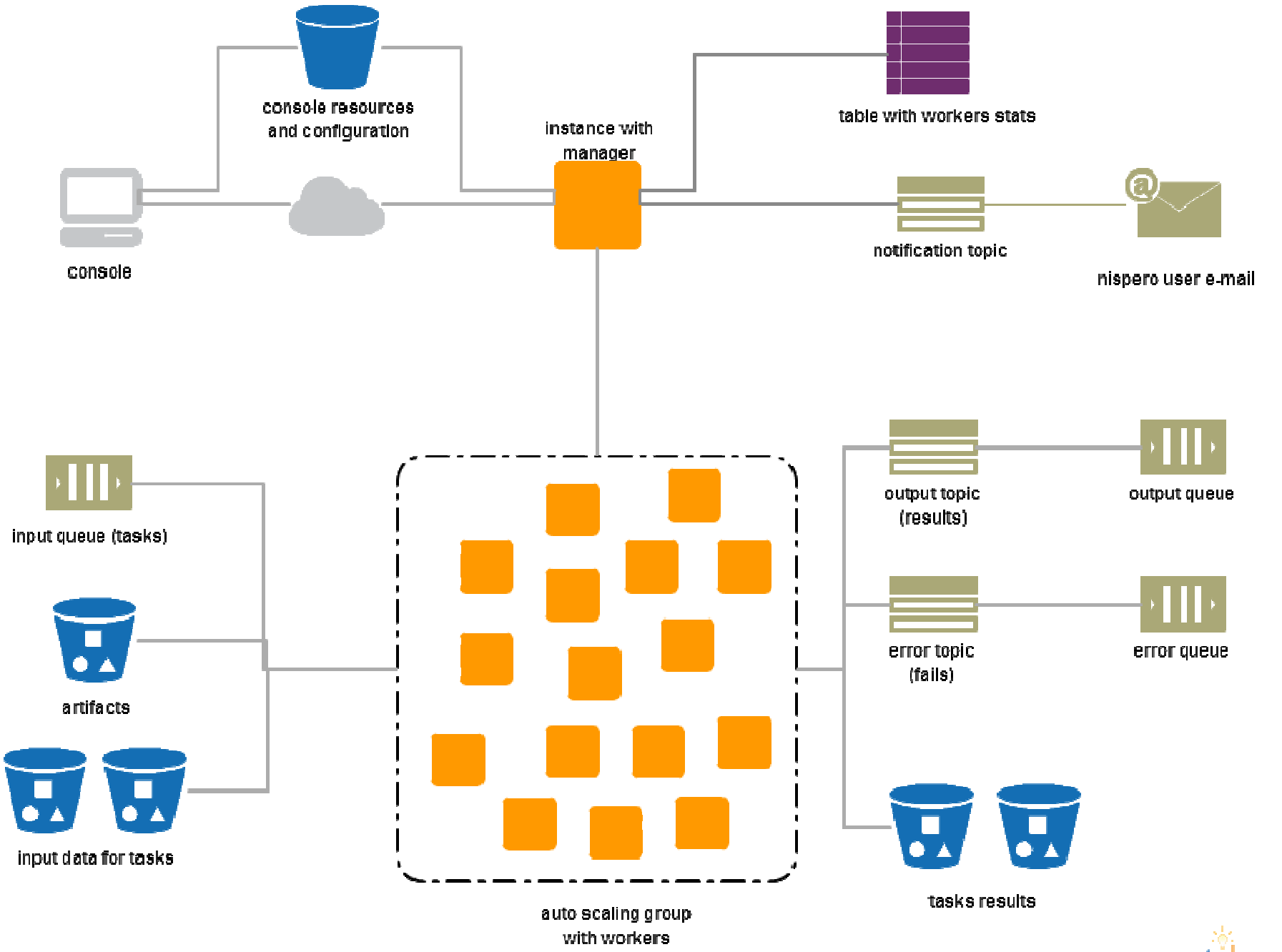
Why Nispero is cool for NGS data analysis?

horizontal scaling

limited number of different instances types : cr1.8xlarge


console resources
and configuration
instance with
manager
table with workers stats
console
notification topic
nispero user e-mail
auto scaling group
with workers
input queue (tasks)
output topic
(results)
output queue
instance with
worker application
artifacts
error topic
(fails)
error queue
input data for tasks
tasks results
Online diagramming & design creately.com


console resources
and configuration
instance with
manager
table with workers stats
console
notification topic
nispero user e-mail
input queue (tasks)
output topic
(results)
output queue
artifacts
error topic
(fails)
error queue
input data for tasks
tasks results
auto scaling group
with workers
creately.com



[online diagramming & design] creately.com

Hands-on Nispero

Now it’s your turn

You’re going to try to use Nispero to run two parallel blast

Era7 bioinformatics

ohnosequences!

Hands-on Nispero

Now it’s your turn

You’re going to try to use Nispero to run two parallel blast

And Kim please support the others but let them try by themselves :)

Hands-on Nispero

How to run a Nispero

1. Set up the environment
2. Prepare the tasks file
3. Prepare the scripts
4. Download
5. Set up the configuration file
6. Publish the folder
7. Run Nispero
8. Terminate

nispero-usage.md

Era7 bioinformatics

ohnosequences!

Hands-on Nispero

1. Set up the environment

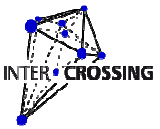
aws-linux-env-setup.md

Standard Linux Amazon m2.xlarge Spot request

God mode + right key pair

Hands-on Nispero

2. Prepare the tasks file




```
"id": "task1",
  "inputObjects":{
    "database":{
      "bucket":"team1-resources",
      "key":"input/NC_000913.fna"
    },
    "query":{
      "bucket":"team1-resources",
      "key":"input/E-coli-rna.frn"
    }
  },
  "outputObjects":{
    "results":{
      "bucket":"team1-resources",
      "key":"results/ecoli-blastresults.txt"
    }
  }
```

Era7 bioinformatics

ohnosequences!

Hands-on Nispero

2. Prepare the tasks file

Put the tasks file in a bucket in S3

Hands-on Nispero

3. Prepare the scripts

- Workers configuration (Statika?)

- Run the task (Blast)

Hands-on Nispero

4. Download

4.1 Connect to the instance you launched at 1.

4.2 Download Nispero to the instance https://github.com/ohnosequences/cloud-ngs-course/blob/master/nispero-task/nispero-usage.md#download

Name + mail

Hands-on Nispero

5. Set up the configuration.scala file

5.1 Use VIM to edit the config file

5.2 https://github.com/ohnosequences/cloud-ngs-course/blob/master/nispero-task/nispero-usage.md#setup-your-configuration

Hands-on Nispero

6. Publish

https://github.com/ohnosequences/cloud-ngs-course/blob/master/nispero-task/nispero-usage.md#publish

7. Run

https://github.com/ohnosequences/cloud-ngs-course/blob/master/nispero-task/nispero-usage.md#run

Hands-on Nispero

8. Terminate / Undeploy

AGPL
Free Software
3
Free as in Freedom