
SEAGATE
CORTX™
Integration Challenge

Informational Webinar: CORTX Integration Challenge

April 14, 2021

Rachel Novak, Justin Woo

Seagate

AGENDA

01 CORTX 101

02 How CORTX works

03 CORTX & Integrations

04 How to set up your VM

AGENDA

05 How to build an integration

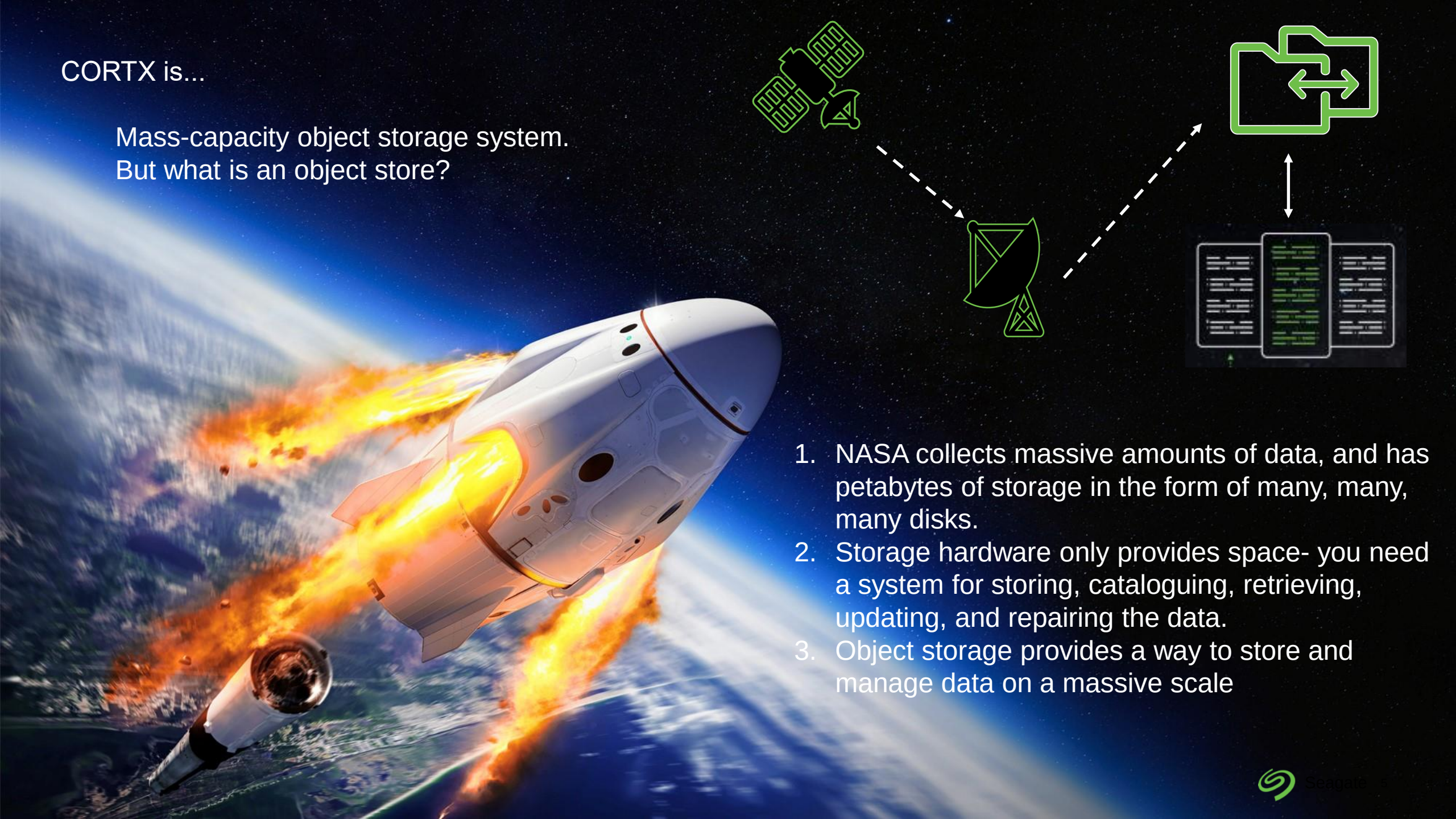
06 Live demo

07 How to submit your integration

08 Questions

CORTX is...

Mass-capacity object storage system.
But what is an object store?



1. NASA collects massive amounts of data, and has petabytes of storage in the form of many, many, many disks.
2. Storage hardware only provides space- you need a system for storing, cataloguing, retrieving, updating, and repairing the data.
3. Object storage provides a way to store and manage data on a massive scale

Key Components of object storage:

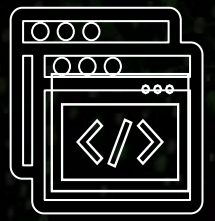
- A way to decide how and where to store data on the hardware
- How to organize the location information of that data (i.e. metadata- what is this data object and where is it)
- A way to ensure that many users can access the same data at the same time without conflicts
- Monitoring the health of the data and the system
- Mechanisms to recover from failure/errors and to ensure data isn't lost or corrupted
- A management system by which admins and users can view the health of the system and configure various behavioral aspects

Key Components of CORTX

CORTX as a platform is made up of different modules, each of which handle different functions.

- **MOTR**
  - Primary data store
  - Interacts directly with storage devices
  - Manages data storage and metadata
  - Rebuilds data after failures.

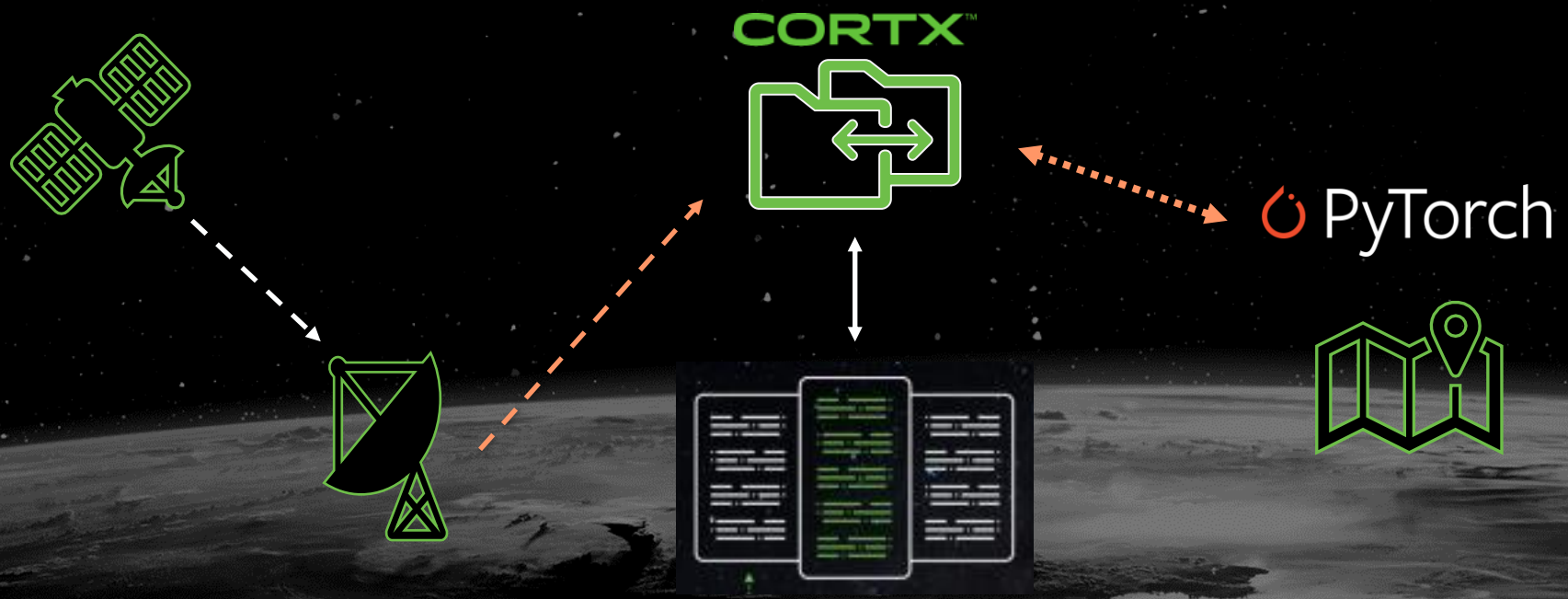
- **HA & HARE**
  - Monitor for failures
  - Tell Motr what to rebuild



- **Manager**
  - Human-friendly interface with CORTX
  - Administrative UI
  - Web GUI for monitoring & management

- **CORTX S3**
  - Translates between Motr and S3
  - Allows use of S3 API when connecting to other tools and platforms

What roles do integrations play in object storage?



1. NASA wants to analyze the data it has stored on its hardware
2. It will need to integrate its analytics tools & data collection tools with its data store
3. An integration connects a tool like PyTorch with an object store like CORTX

Selecting an Integration

1. What data is collected?
2. What needs to be done with that data?
3. Identify the tools that would need to connect to the stored data or the storage hardware.

Elasticsearch	Search & Analytics
PyTorch	Machine Learning
Apache Kafka	Stream Processing
Apache Spark	Spark Analytics
IPFS	Peer to Peer File Share
Humio	Log management
Video Loft	VsaaS
Tensorflow	Machine Learning
Prometheus	Monitoring
FIO	Benchmark

Presto	Query Engine
Greenplum	Analytics
Restic	Backup
VXG	VSaaS
dataobi	Data Migration
Grafana	Dashboard
H2O.ai	Machine Learning
Traefik	Reverse Proxy / Load Balancer
Alexa	Voice
IOR	HPC Benchmark

CORTX- How to Build an Integration

- Identify the tool you want to integrate with CORTX
- Go to their website and check out their documentation for S3 API
  (Motr integrations are a bit trickier. If you have questions, please ask in #hackathon-support-and-mentorship)
- Identify the functions for writing data to the platform (or other relevant functions, depending on what you are integrating with)
- Edit config file or write code to set up your integration
- Test! Run your integration and make sure the two components (CORTX and your integration platform) are correctly connecting
- Document it!
  - Explain what the tool does, and how it will work with CORTX.
  - Ensure you document every step, including the ones that seem obvious
  - Explain each step and be sure to mention potential failures/errors as well as documenting what success looks like.

Submitting your Integration

A submission to the CORTX Integration Challenge will be considered complete if it contains:

A functional integration between CORTX and another system, platform, or tool, which is explained and documented in:

1. A pull request to the CORTX GitHub Repository
2. A video demonstrating the set up and functionality of the integration, along with an explanation of its purpose
3. Documentation in the pull request explaining how to set up and use the integration.

The DEADLINE for project submissions in Devpost is **11:00 PM PT , Tuesday, April 27th**

Submitting your Integration

Submission instructions: https://github.com/Seagate/cortx/blob/main/doc/integrations/fhir.md

Step 1: Sign-up or sign-in to GitHub

Step 2: Fork the CORTX repository so you can make a pull request.

Step 3: In that fork, open the integration folder and create a .md Markdown file. In that file include:

- The intention behind the integration and what it does.
- Installation directions that are clear, accessible, and easily reproducible.
- A link to your integration video which includes:
  - An explanation of the integration use case - why/how it is useful or impactful in addition to demonstrating how to set up and run the integration.
  - Clear and concise directions on how to install the integration

Step 4: Make a pull request to merge your changes back to the main repository.

Step 5: Go to Devpost and submit your integration by posting a link to your pull request.

Live Demo - Integration

A good example of an integration, write up, and video can be found here:

- Splunk Integration (CORTX Github)
- Connecting CORTX to Splunk – YouTube

CORTX- Setting up your VM

There are a few ways you can get a CORTX VM to run…

Use CloudShare

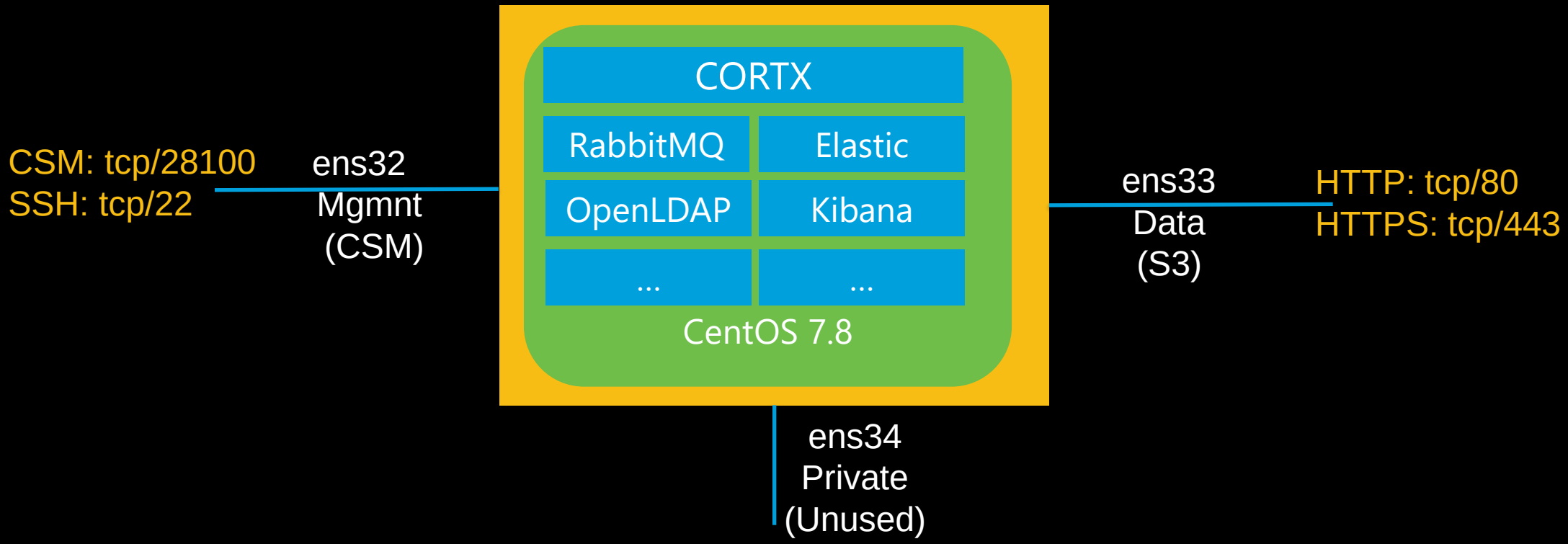
- To gain access to a pre-configured CloudShare environment, all you need to do is request access!
- https://cortx.link/cloudshare-request

Stand up your own VM

- Requirements
  - Min. 8GB of RAM
  - A VM application: Vmware, VirtualBox, etc.
- Download OVA
- Set up VM Docs:
  - CORTX VM Setup
  - Documentation Instructions for VirtualBox

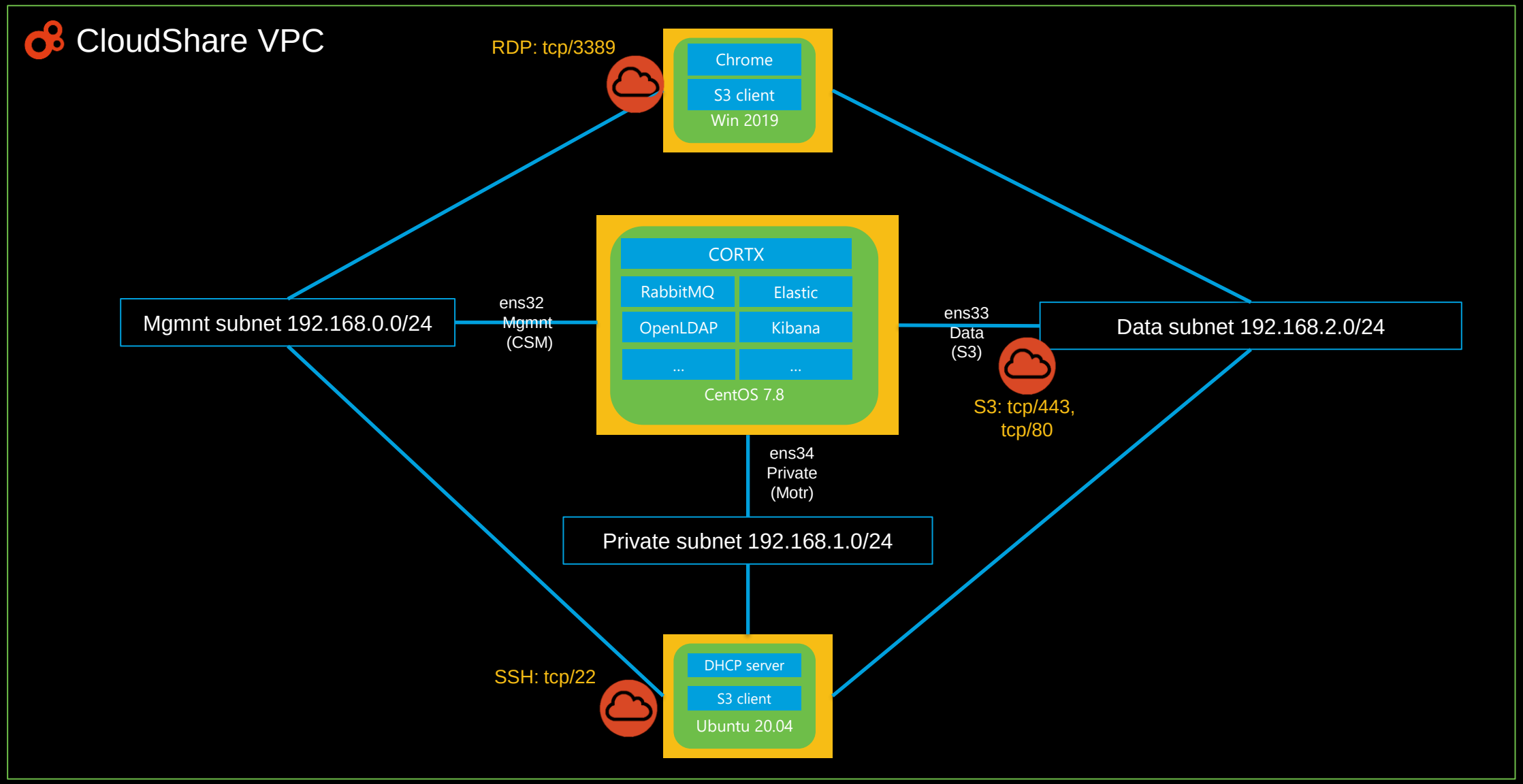
Support: Slack #**hackathon-support-and-mentorship**

CORTX VM



https://github.com/Seagate/cortx/blob/main/doc/CORTX_on_Open_Virtual_Appliance.rst

CORTX CloudShare environment



Questions?