Batch Processing technical meeting

Declared scope:

- focus on technical solutions for integrating Rosetta and ESAP

- make a proposal to be discussed with a range of stakeholders

In general:

- identify technical solutions to access computing facilities to run batch processing software

-

- identify what we mean with "batch processing software"

Batch Processing software

"Identification":

- It runs unattended, does not require human intervention

- it automatizes some computing operations

- it starts froma specific configuration user's provided

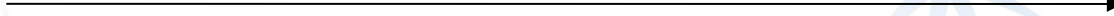# Batch Processing vs interactive

"Identification":

- Need (?) a-priori configuration

- No human intervention

- Need to monitor the "working task"

General use case

user looks for data

user looks for software

ESAP

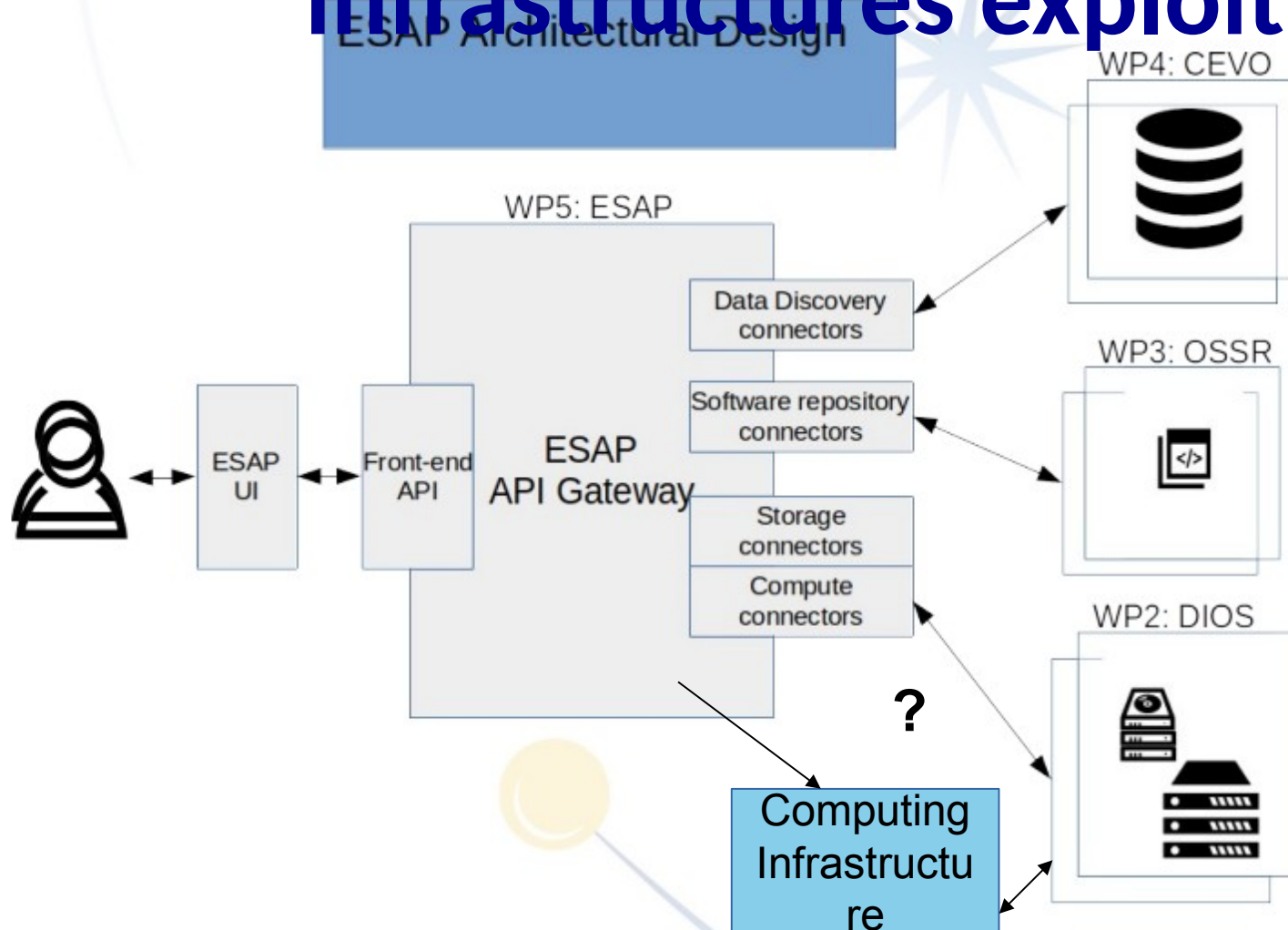**?**   user selects/config computing
resources/infrastructure

user run analysis
in the selected infrastructure

Workflow in case of HPC&HPC infrastructures exploiting

Slide @Zheng Meyer-Zhao

Focus group update

Only discussion about requiremens.

Main open questions identified:

- HPC&HTC infrastructures require Authentication&Authorization:

- a model is needed

- A Job/Task description is needed?

Authentication and authorization

Authn and Authz are not an option in HPC and HTC infrastuctures

- resource access policies

- billing policies

Authn and Authz have two levels

- platform level

- computing center level

both must be managed.

Batch Processing Technical Meeting – 27 January 2022

From: ESCAPE WP5 Progress Meeting – 26-27 October 2020

Funded by the European Union's Horizon 2020 - Grant N° 824064

Job/task description

Run analysis software in a computing infrastructure HPC or HTC requires:

- Data staging/access

- if an already existing infrastructure is exploited, data staging is needed

- A Job/Task description is needed?

Job Description Language Example

Job Type
interactive/not interactive

Analysis software

Type = "Job";

JobType = "Normal";

#JobType = "Interactive" # MPI, simple, ...

Requirements/Environment

Executable = "/bin/hostname";

Arguments = "-f";

Requirements = other.GlueCEInfoLRMSType == "PBS" &&

other.GlueCEInfoTotalCPUs > 1;

Where are data

StdOutput = "simple.out";

Transfer Protocol

StdError = "simple.err";

InputSandbox={"myscript.sh"};

#InputSandbox = {"gsiftp://example.it:5678/tmp/my_anlysis.py"

Where put resuls

9 *OutputSandbox = {"simple.out", "simple.err", "image.png"};*