TRUFA

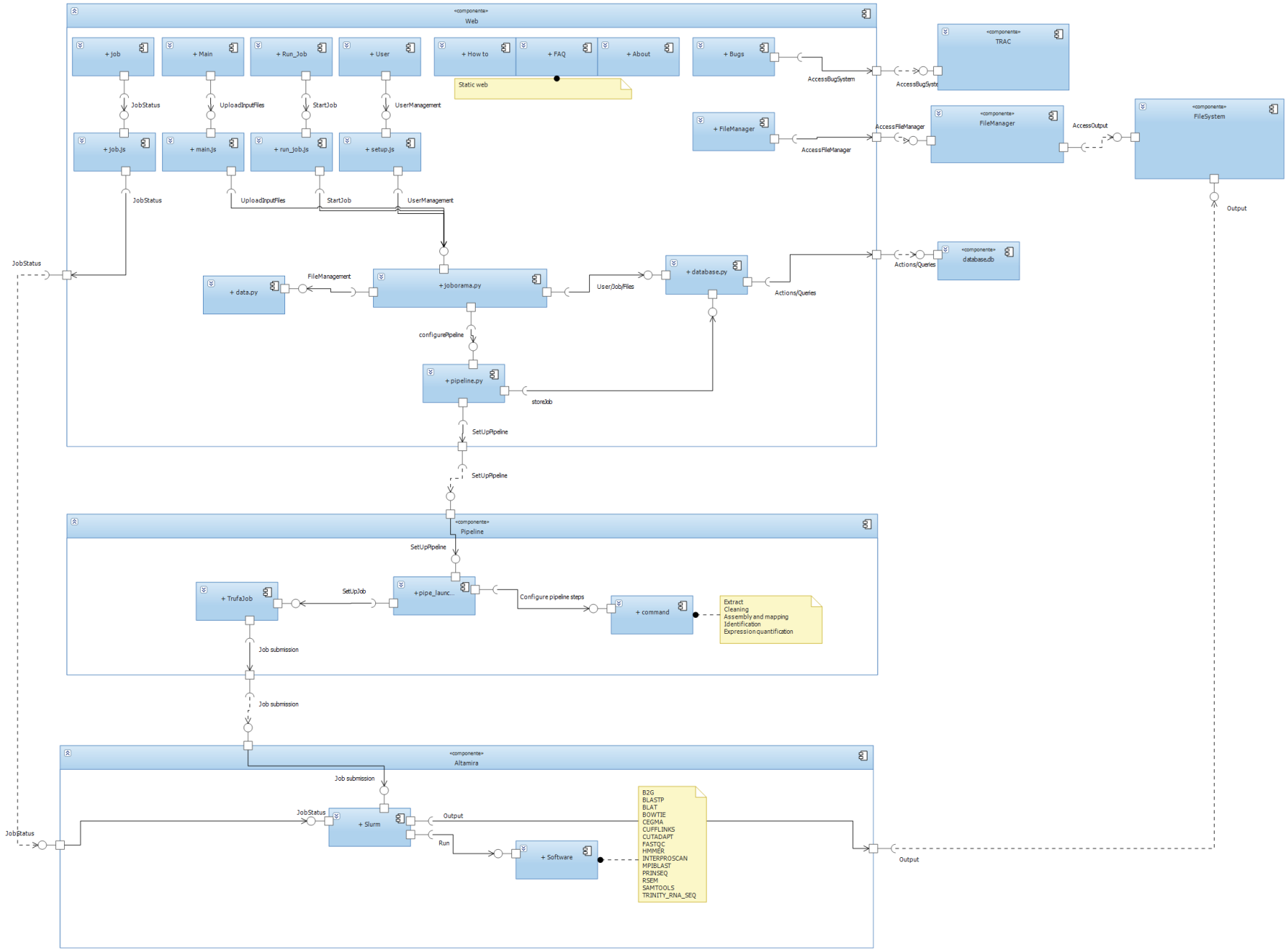
TRANSCRIPTOMES USER-FRIENDLY ANALYSIS

Fernando Aguilar Gómez (aguilarf@ifca.unican.es)

EGI-Lifewatch CC Meeting - Santander

WHAT IS TRUFA?

- **Application of next-generation sequencing (NGS) methods for transcriptome analysis (RNA-seq) has become increasingly accessible in recent years.**
- **TRUFA is an open informatics platform offering a web-based interface that generates the outputs commonly used in de novo RNA-seq analysis and comparative transcriptomics.**
- **The platform permits analyzing RNA-seq data in a fast, robust, and user-friendly manner.**
- **Currently there is a version at IFCA over Altamira Supercomputer.**

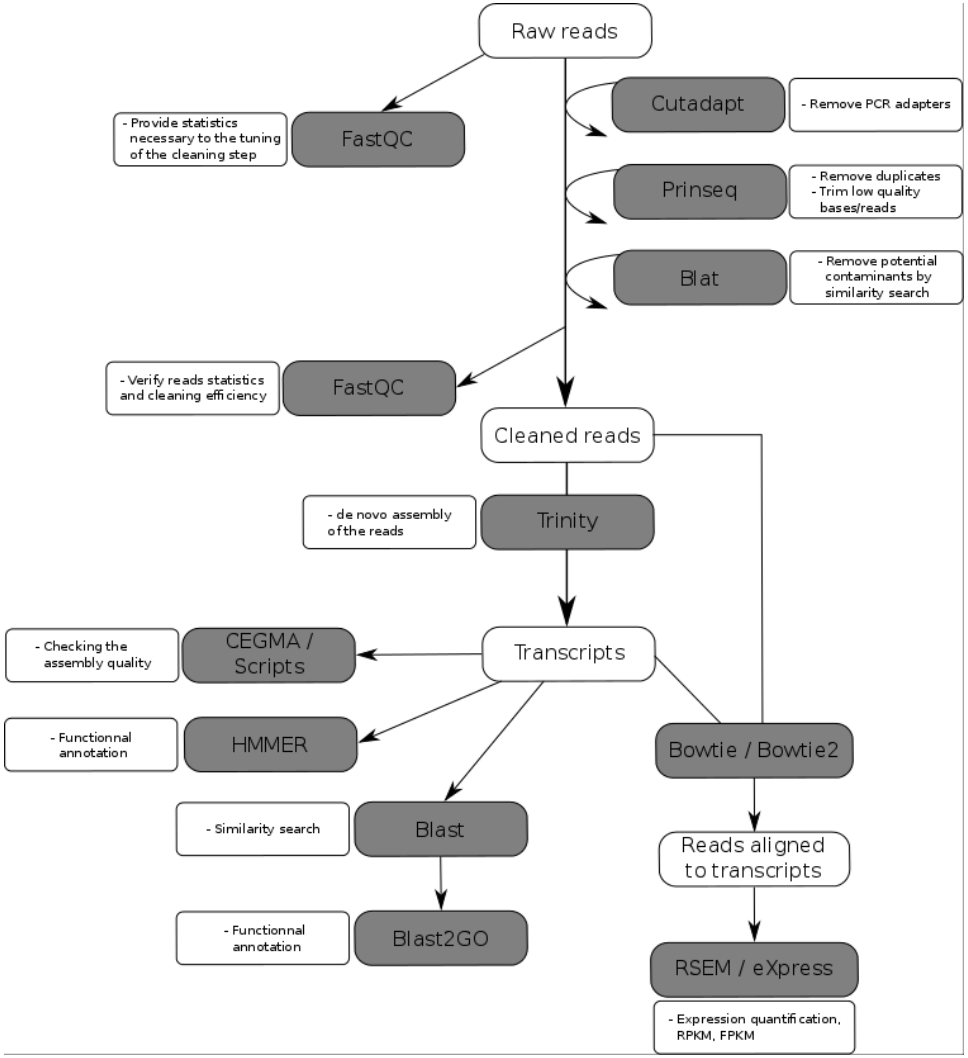

«componente»
Web
+ Job
+ Main
+ Run_Job
+ User
+ How to
+ FAQ
+ About
+ Bugs
Static web
JobStatus
UploadInputFiles
StartJob
UserManagement
+ job.js
+ main.js
+ run_job.js
+ setup.js
«componente»
TRAC
AccessBugSystem
+ FileManager
AccessFileManager
«componente»
FileManager
AccessOutput
«componente»
FileSystem
Output
+ data.py
FileManagement
+ joborama.py
User/JobFiles
+ database.py
Actions/Queries
«componente»
database.db
configurePipeline
+ pipeline.py
storeJob
SetUpPipeline
«componente»
Pipeline
+ TrufaJob
SetUpJob
+ pipe_launc...
Configure pipeline steps
+ command
Extract
Cleaning
Assembly and mapping
Identification
Expression quantification
Job submission
«componente»
Altamira
+ Slurm
Run
+ Software
B2G
BLASTP
BLAT
BOWTIE
CEGMA
CUFFLINKS
CUTADAPT
FASTQC
HMMER
INTERPROSCAN
MPIBLAST
PRINSEQ
RSEM
SAMTOOLS
TRINITY_RNA_SEQ

TRUFA EXPERIENCE

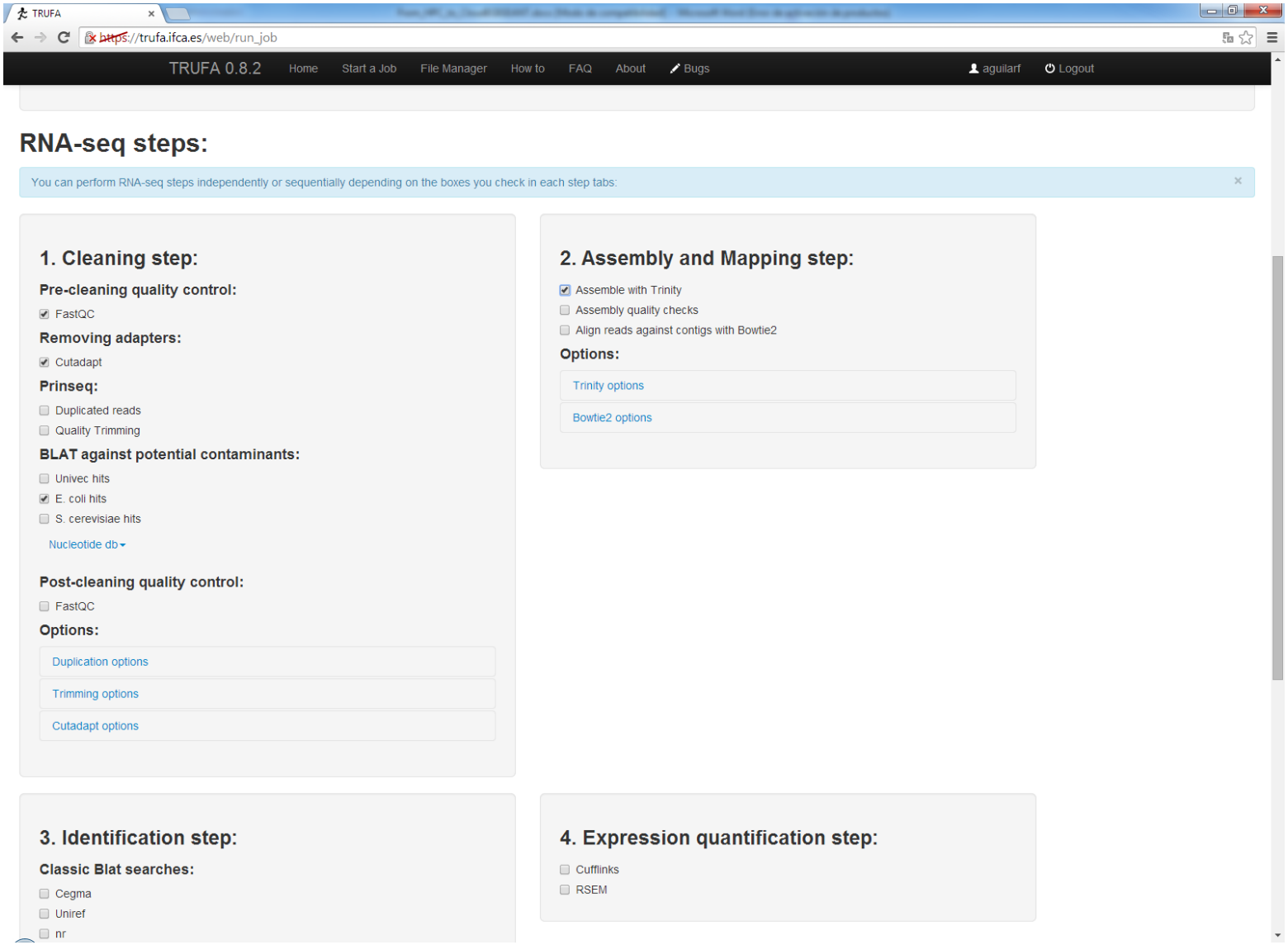
- **The experience with local users: 1-2 Million hours and 10 TB are only starting numbers to produce a few papers in the area. Requests from: Brazil, China, US, Argentina, India and of course Europe.**

Software	Number of CPUs	Type
B2G	16	identify
BLASTP	16	assembly/mapping
BLAT	64	cleaning, identify
BOWTIE	16	assembly/mapping
CEGMA	16	assembly/mapping
CUFFLINKS	16	expression
CUTADAPT	16	cleaning
FASTQC	2	cleaning
HMMER	96	identify
INTERPROSCAN	64	identify
MPIBLAST	96	identify
PRINSEQ	16	cleaning
RSEM	1	expression
SAMTOOLS	1	assembly/mapping
TRINITY_RNA_SEQ	16	assembly/mapping

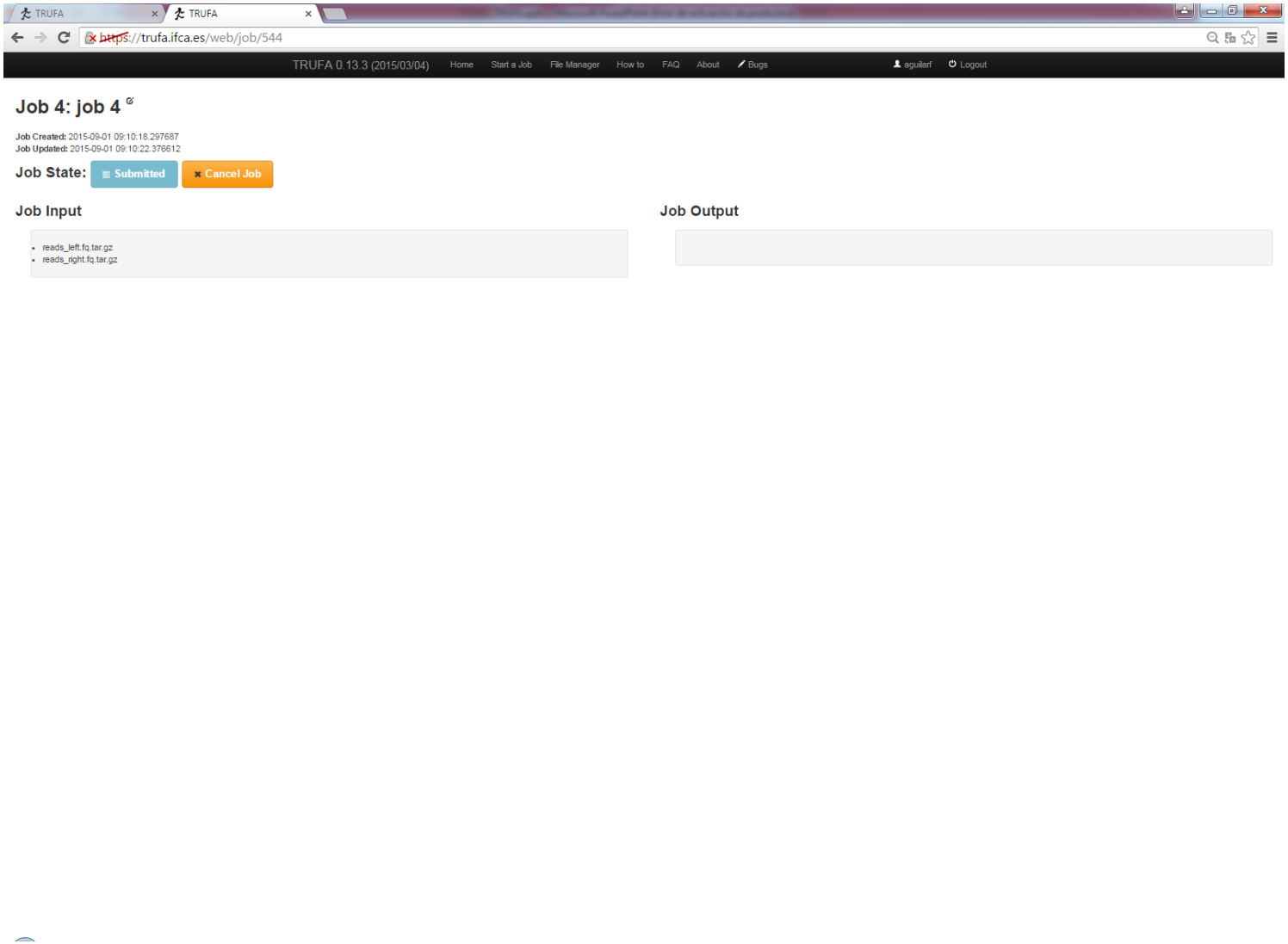
PIPELINE


Raw reads
- Provide statistics necessary to the tuning of the cleaning step
FastQC
Cutadapt
- Remove PCR adapters
Prinseq
- Remove duplicates
- Trim low quality bases/reads
Blat
- Remove potential contaminants by similarity search
- Verify reads statistics and cleaning efficiency
FastQC
Cleaned reads
- de novo assembly of the reads
Trinity
Transcripts
- Checking the assembly quality
CEGMA / Scripts
- Functionnal annotation
HMMER
- Similarity search
Blast
- Functionnal annotation
Blast2GO
Bowtie / Bowtie2
Reads aligned to transcripts
RSEM / eXpress
- Expression quantification, RPKM, FPKM

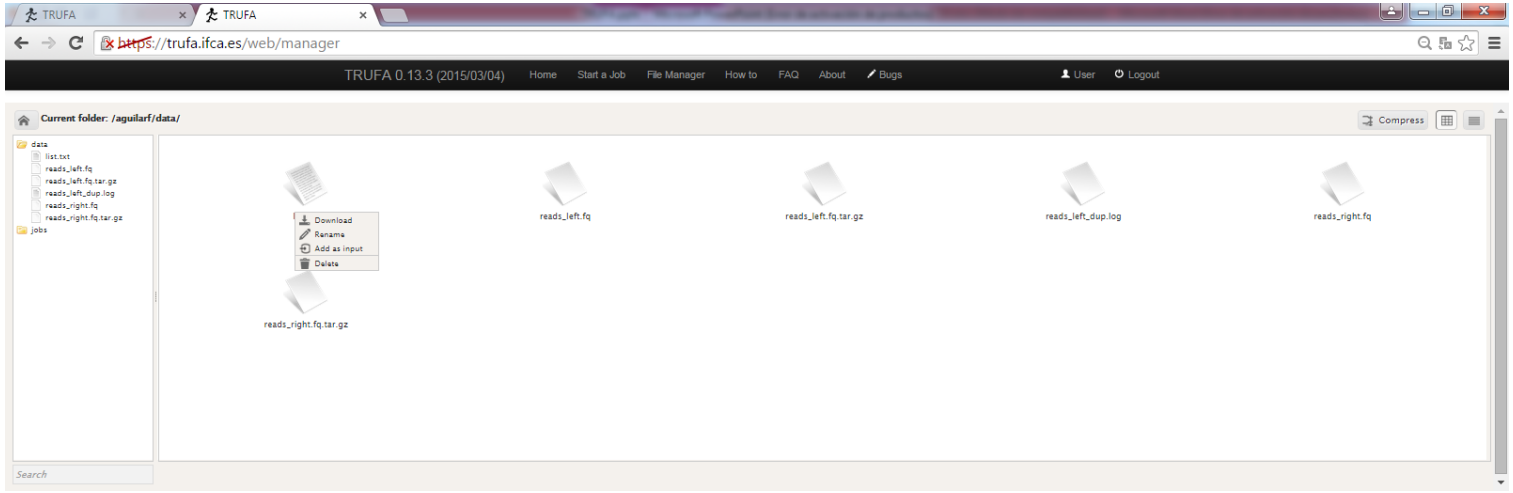
TRUFA WEB

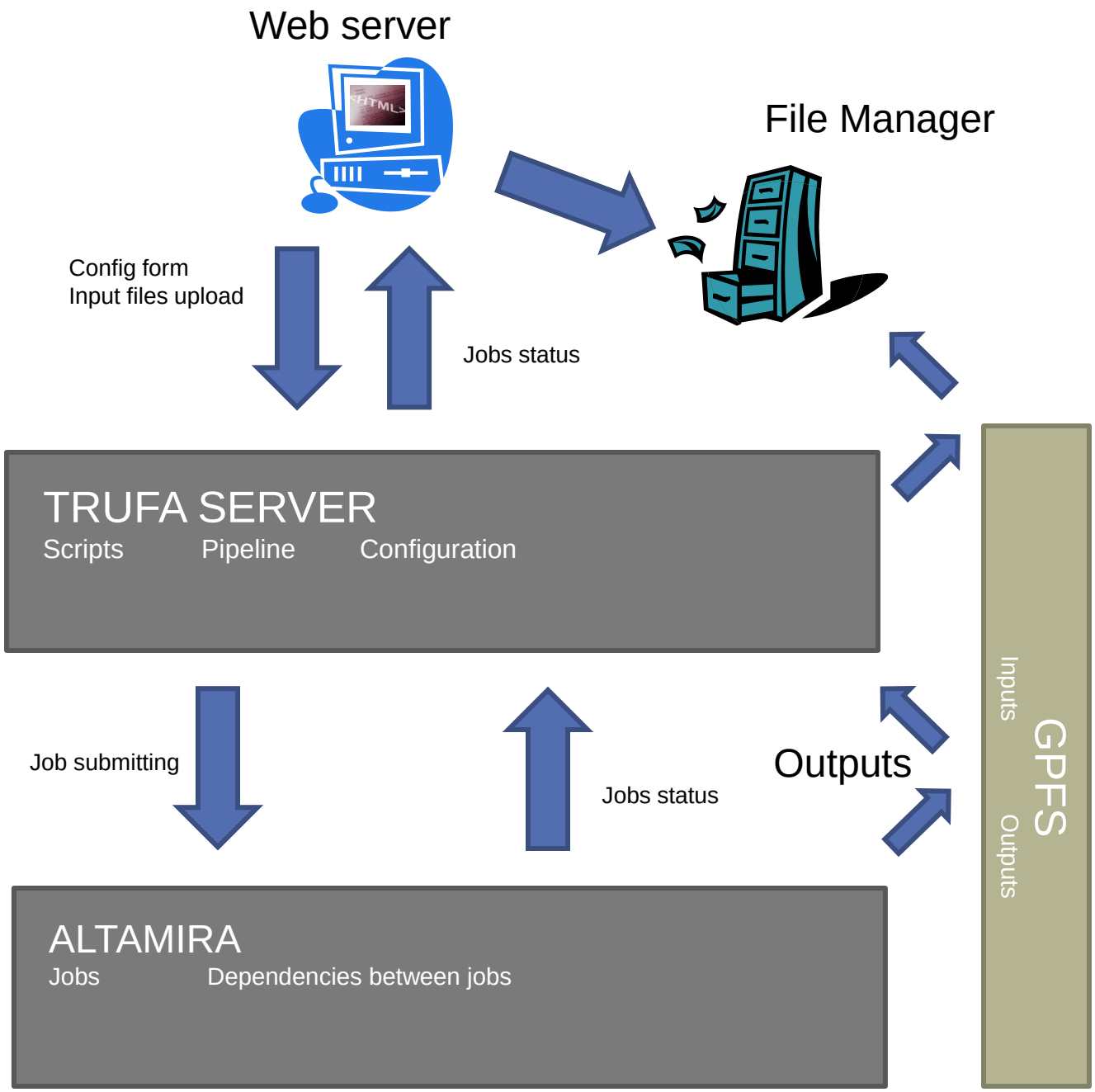

TRUFA
https://trufa.ifca.es/web/run_job
TRUFA 0.8.2
Home
Start a Job
File Manager
How to
FAQ
About
Bugs
aguilarf
Logout
RNA-seq steps:
You can perform RNA-seq steps independently or sequentially depending on the boxes you check in each step tabs:
1. Cleaning step:
Pre-cleaning quality control:
FastQC
Removing adapters:
Cutadapt
Prinseq:
Duplicated reads
Quality Trimming
BLAT against potential contaminants:
Univec hits
E. coli hits
S. cerevisiae hits
Nucleotide db
Post-cleaning quality control:
FastQC
Options:
Duplication options
Trimming options
Cutadapt options
2. Assembly and Mapping step:
Assemble with Trinity
Assembly quality checks
Align reads against contigs with Bowtie2
Options:
Trinity options
Bowtie2 options
3. Identification step:
Classic Blat searches:
Cegma
Uniref
nr
4. Expression quantification step:
Cufflinks
RSEM

TRUFA WEB


TRUFA
TRUFA
https://trufa.ifca.es/web/job/544
TRUFA 0.13.3 (2015/03/04)
Home
Start a Job
File Manager
How to
FAQ
About
Bugs
aguilarf
Logout
Job 4: job 4
Job Created: 2015-09-01 09:10:18.297687
Job Updated: 2015-09-01 09:10:22.376612
Job State:
Submitted
Cancel Job
Job Input
reads_left.fq.tar.gz
reads_right.fq.tar.gz
Job Output

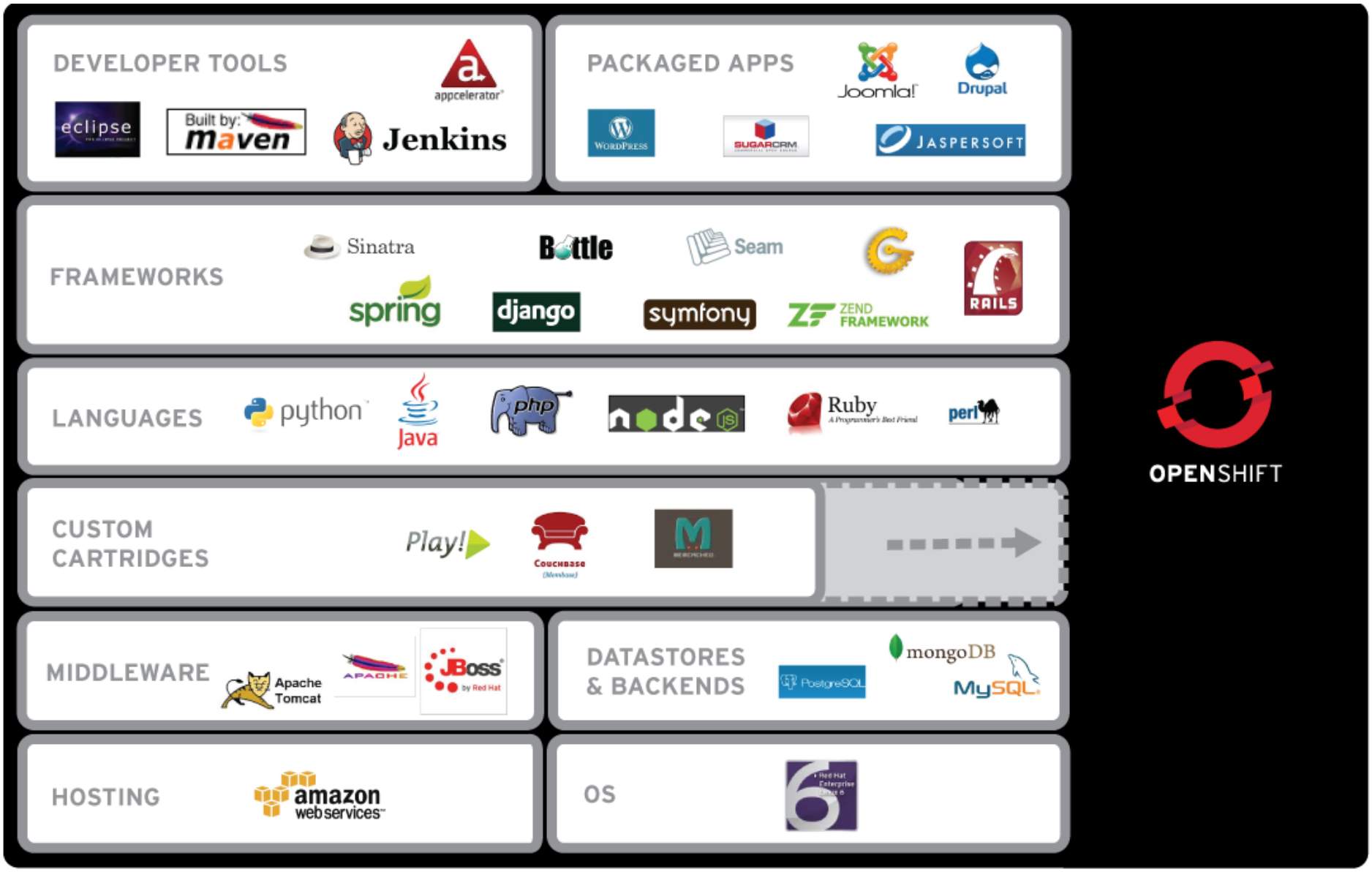
TRUFA WEB

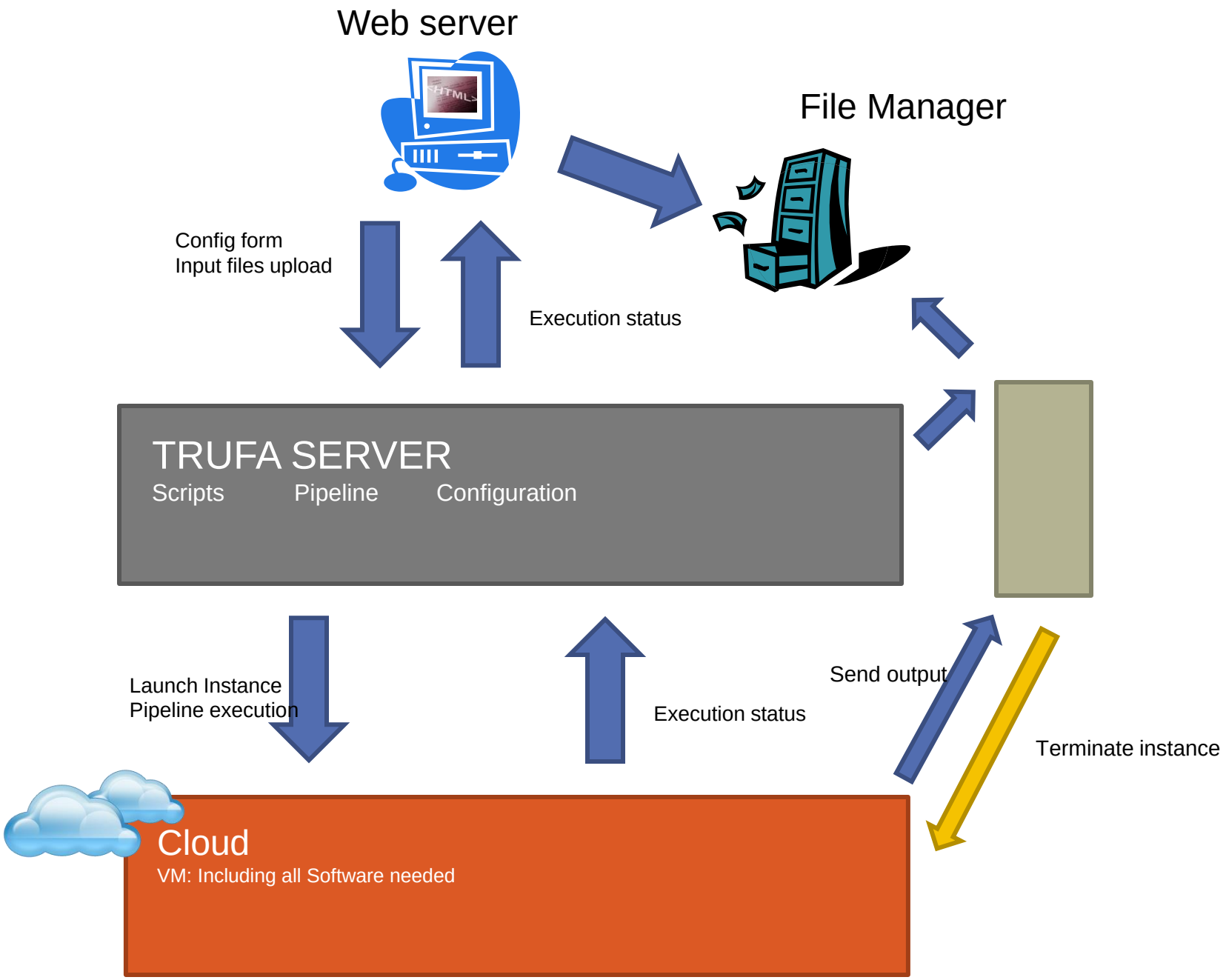

TRUFA
TRUFA
https://trufa.ifca.es/web/manager
TRUFA 0.13.3 (2015/03/04)
Home
Start a Job
File Manager
How to
FAQ
About
Bugs
User
Logout
Current folder: /aguilarf/data/
Compress
data
list.txt
reads_left.fq
reads_left.fq.tar.gz
reads_left_dup.log
reads_right.fq
reads_right.fq.tar.gz
jobs
Download
Rename
Add as input
Delete
reads_left.fq
reads_left.fq.tar.gz
reads_left_dup.log
reads_right.fq
reads_right.fq.tar.gz
Search


Web server
File Manager
Config form
Input files upload
Jobs status
TRUFA SERVER
Scripts
Pipeline
Configuration
Job submitting
Jobs status
Outputs
GPFS
Inputs
Outputs
ALTAMIRA
Jobs
Dependencies between jobs

OPENSHIFT

- **Is a PaaS tool for container managing and deploying.**
- **Broker: is the orchestrate mechanism for all the activities of the application management platform.**
  - Manage the login of the users.
  - Update the DNS dynamically.
  - Provide information about the status of the application.
  - Orchestrate the application.
  - Services and operations.
- **Nodes: host that executes applications.**
  - A node contains gears that contain applications.
  - Each node is defined according to a series of gears.
  - Represent the portions of CPU, memory and space in disk available for a single application.


DEVELOPER TOOLS
appcelerator
eclipse
Built by: maven
Jenkins
PACKAGED APPS
Joomla!
Drupal
WordPress
SUGARCRM
JASPERSOFT
FRAMEWORKS
Sinatra
Bottle
Seam
spring
django
symfony
ZEND FRAMEWORK
RAILS
LANGUAGES
python
Java
php
node.js
Ruby
perl
CUSTOM CARTRIDGES
Play!
Couchbase
MIDDLEWARE
Apache Tomcat
APACHE
JBoss by Red Hat
DATASTORES & BACKENDS
PostgreSQL
mongoDB
MySQL
HOSTING
amazon web services
OS
OPENSHIFT


Web server
File Manager
Config form
Input files upload
Execution status
TRUFA SERVER
Scripts
Pipeline
Configuration
Launch Instance
Pipeline execution
Execution status
Send output
Terminate instance
Cloud
VM: Including all Software needed