




Building the Flomics bioinformatics infrastructure from the ground up


Cristina Tuñí, André Lopes, Pedro Ferreira, Lluc Cabus, Joao Curado & Julien Lagarde
Flomics Biotech, S.L. Barcelona

Floxxxics


Flomics is a biotechnology company operating in the field of genomics, bioinformatics and NGS. We are developing a novel multicancer screening solution to be implemented in the clinical practice. Flomics also provides a complete range of personalized genomics services, with a fully compliant molecular biology laboratory and specialized bioinformatics department.



Flomics uses Nextflow pipelines to carry out R&D work on biomarker discovery, usually using already existing curated pipelines from nf-core such as **rnaseq**, **ampliseq**, **viralrecon**, and **rnafusion** pipelines. Flomics also forks them systematically to fit our needs, offering those pipelines as SaaS (software as a service) to our clients.



Companies offering GaaS and BaaS (Genomics and Bioinformatics as a service) can benefit from using a cloud computing provider, as opposed to an on-site infrastructure, in terms of costs and maintenance.
Flomics cloud computing infrastructure is divided into Baas and R&D

**(Baas)**

16S Metagenomics pipeline
This pipeline takes any NGS amplicon-based 16S data for metagenomics analysis. It generates taxonomic composition tables at and above genus ranks, and alpha and beta diversity analysis and statistics.
2 Credits per sample
Read more
Select

COVID-19 NGS data analysis
It takes any kind of COVID-19 NGS raw data (FASTQ format) and generates interpretable information. It includes QC, read alignment, variant calling and interpretation.
1 Credit per sample
Read more
Select

Flomics RNAseq pipeline
This pipeline performs the first level of analysis needed for any RNA-seq experiment: Quality Control, read mapping, gene/transcript quantification, etc.
1 Credit per sample
Read more
Select

SARS-CoV-2 waste waters
Detect and identify SARS-CoV-2 variants in wastewater samples.
1 Credit per sample
Read more
Select

Currently available pipelines on Stratus web server

Monitoring & Logging
CloudWatch

CI/CD - DevOps
Code Build
Code Pipeline
Code Deploy

Elastic Beanstalk (User End Service Platform)
Internet Gateway
VPC
Gateway Load Balancer
EC2
T2g


Loosely Coupled Architecture Components
Lambda
Simple Queue Service
Storage / Systems of Record
RDS Database
S3 Storage Service

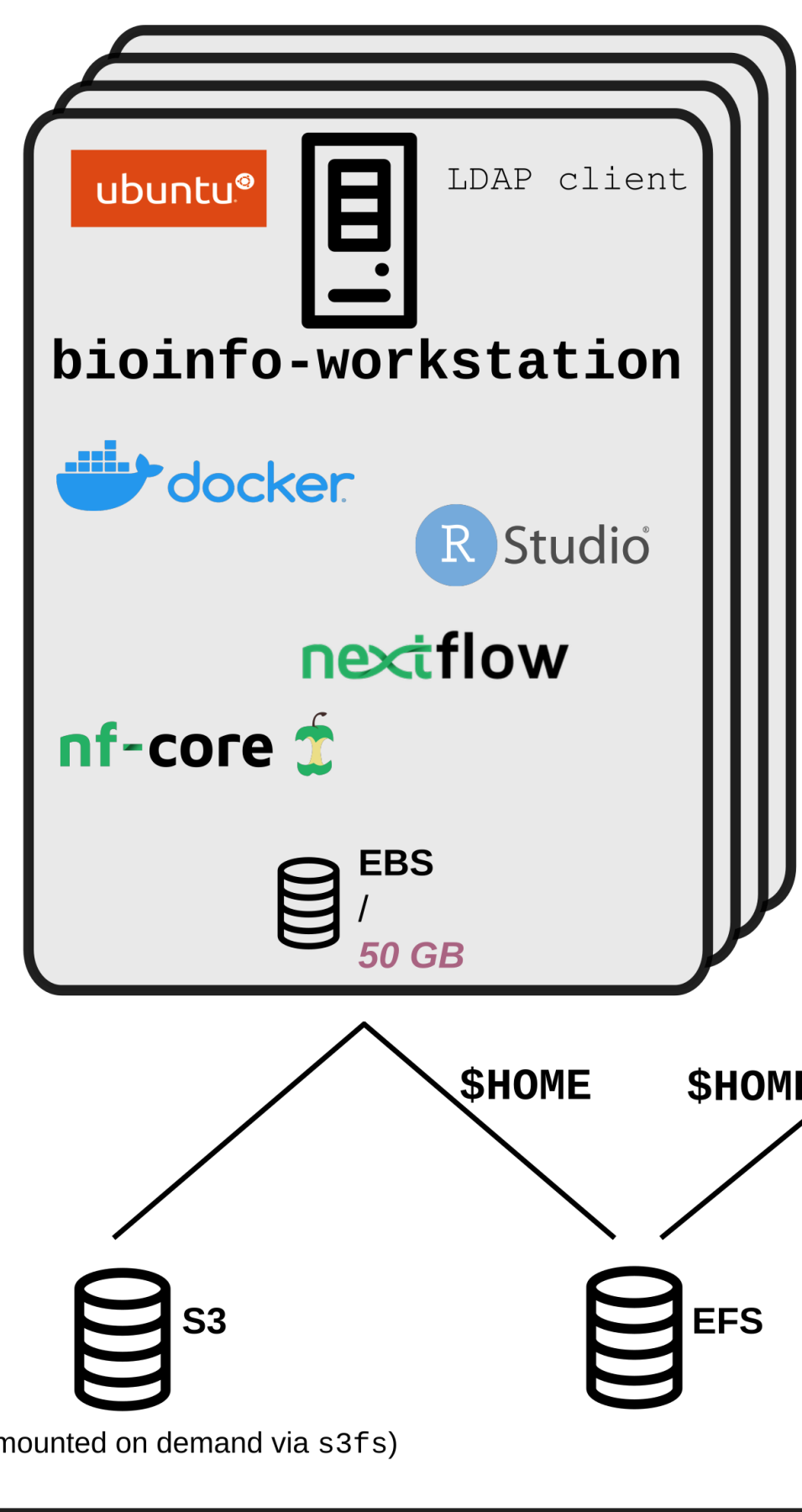
NGS Compute
NextFlow EC2 Orchestrator
AWS Batch
EC2 Individual Contributors Computing

Stratus Infrastructure

R&D bioinformatics infrastructure

Flomics' bioinformatics cloud infrastructure





ubuntu®

LDAP client

bioinfo-workstation

docker

R Studio

nextflow

nf-core

EBS / 50 GB

ssh server

EBS / 8 GB

S3

EFS

(mounted on demand via s3fs)


ssh

ssh






Flomics is hiring!!

- Bioinformatics Senior Scientist
- Software developer


More info at <https://www.flomics.com>



Solved issues

Problem:  Amazon EC2 Spot Instances Out of disk space	Solution:  Amazon Elastic Block Store Autoscaling
Problem:  Amazon S3 Storage space	Solution:  S3 Intelligent-Tiering & nextflow clean -f -k
Problem:  Batch Sub-optimal nf-core queue usage	Solution: Dedicated process-specific Batch queues

Remaining issues




Key

Value

Owner = DbAdmin

Stack = Test



Owner = DbAdmin

Stack = Production

Support for tagging AWS resources
Starting on release 22.10 of Nextflow

CPU

Cache

Main Memory

Word Transfer

Block Transfer

Nextflow clean and manual removal of workdir remove both cached files and run info