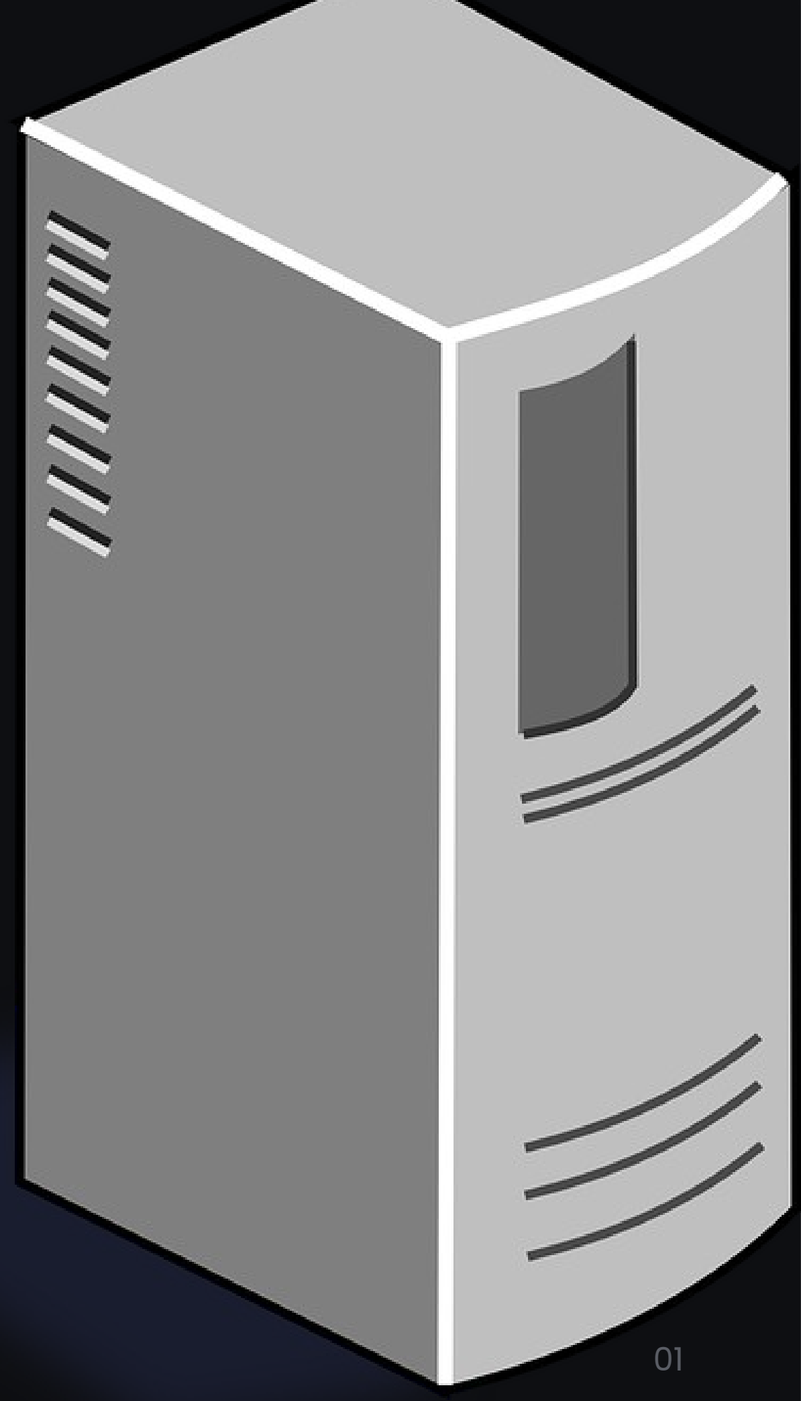
ZERBITZUA · AZPIEGITURA

Cloud GPU eta AA azpiegitura

AArako konputazio ahalmena datuak lurralde propioan ostatatuta.

Datacenter propioa A Coruñako Ciudad de las TIC-en eta zabalpen elastikoa hodei anitzetan.



cinfo

ERRONKA

Modeloak entrenatu eta zerbitzatu hodei bakar baten mende egon gabe

AAk GPU eskatzen du, eta GPUak erabakiak: non bizi den datua, zenbat balio duen eskaeraren gailurrak eta zer gertatzen den hornitzaileak prezioak edo baldintzak aldatzen dituenean.

01

Kostu aurreikusezina

Entrenamendu eta inferentzia gailurrek hodei publikoaren faktura puzten dute.

02

Datuak perimetrotik kanpo

Beren informazioa edozein jurisdikziotan prozesatu ezin duten sektore arautuak.

03

Hornitzailearekiko mendekotasuna

Hodei bakar bati lotutako arkitekturak, mugitzeko garestiak direnak.

ZER DEN

Datacenter propioa gehi zabalpen elastikoa hodei anitzetan

cinfo-k AArako bere datacenter propioa birtualizatu du A Coruñako Ciudad de las TIC-en, eta konputazio elastikoko sistema propioa du lan kargak tokiko zerbitzarietan edo edozein hodei hornitzailetan zabaldu eta kudeatzeko.

- 3.328 AMD EPYC gama altuko CPU eta 8 TB RAM memoria.
- 24 GBko 33 NVIDIA L4 GPU eta 48 GB bideo memoriako 6 NVIDIA A40 GPU.
- 114 TB disko eta 500 TB gehigarri latentzia baxuko NASean.
- Zabalpen hibrido automatizatua AWS, Azure, Google Cloud, Oracle, Akamai edo on-premise.



NOLA FUNTZIONATZEN DUEN

BERT: baliabideak bilatu eta komeni den tokian zabaldu

cinfo-k BERT garatu du, hodeiko baliabideen bilatzailea eta zabalpen elastikoaren kudeatzailea, ahalmena aurkitu eta karga hobekien ateratzen den tokian jartzen duena.

01

Karga dimentsionatzen da

Konputazio, memoria, biltegiratze eta latentzia betekizunen analisia.

02

Baliabideak aurkitzen dira

BERTek ahalmen eskuragarria bilatzen du datacenter propioan eta kontratatutako hodeietan.

03

Edukiontzitan zabaltzen da

Kapsulatzea Docker Swarm edo Kubernetes klusterretan zabalduta.

04

Ustiatu eta behatzen da

Erabilgarritasun handia, bizi-zikloaren kudeaketa eta behagarritasun aginte-koadroak.

GAITASUNAK

AA eta bideo kargetarako pentsatutako azpiegitura

- **Datacenter propioa**

  Instalazioak A Coruñako Ciudad de las TIC-en, taldearen bulegoak eraikin berean.

- **GPUak entrenamendurako**

  39 NVIDIA GPU modeloen garapenera, entrenamendura eta inferentziara bideratuta.

- **Latentzia baxuko biltegiratzea**

  114 TB diskoan eta 500 TB latentzia baxuko NASean datu multzo handietarako.

- **Hodei anitzeko zabalpena**

  Kargak hodei publikoan, pribatuan edo zerbitzari fisikoetan berdin jarrita.

- **Edukiontziak eta orkestrazioa**

  Docker Swarm eta Kubernetes erabilgarritasun handiarekin eta bizi-zikloaren kudeaketarekin.

- **CI/CD automatizatua**

  Gitetik integrazio, proba, preprodukzio eta produkziora.

- **Behagarritasuna**

  Monitorizazioa analisi aginte-koadroekin eta alerta abiarazleekin.

- **Araudia betetzea**

  Zibersegurtasuneko betetzea garapen eta ustiapen zikloaren zati gisa.

ONURAK

Ahalmena, kostuen kontrola eta datuaren subiranotasuna

01

Datuak lurralde propioan

Prozesamendua osorik gera daiteke A Coruñako datacenterrean.

02

Kostua kontrolpean

Ahalmen propioa oinarrizko kargarako eta hodei publikoa gailurretarako soilik.

03

Hornitzaile baten mendekotasunik gabe

Zabalpen sistema hodei hornitzailearekiko independentea da.

04

Benetako esperientzia GPUan

cinfo-ren bideo eta AA produktu propioetarako egunero erabiltzen den azpiegitura.

IRISMENA

Azpiegitura kontratatzeko moduak

Zerbitzua arkitektura aholkularitzatik bezeroaren plataformaren ustiapen osora doa.

Arkitektura aholkularitza

Dimentsionatzea, hodeiaren hautaketa eta zabalpen planaren diseinua.

Konputazio ahalmena

cinfo-ren GPU azpiegituraren erabilera entrenamendurako eta inferentziarako.

Zabalpen kudeatua

Plataforma edukiontzitan martxan jartzea, erabilgarritasun handiarekin.

Etengabeko ustiapena

Monitorizazioa, alertak, eguneraketak eta bizi-zikloaren kudeaketa.

Ostatatutako modeloak

LLM open source cinfo-ren hodeitik zerbitzatuta zure aplikazioetarako.

Migrazioa

Dauden kargak eskema hibrido eta hodei anitzekora eramatea.

METODOLOGIA

Dimentsionatzetik ustiapenera

01

Diagnostikoa

Egungo kargen, kostuen eta araudi betekizunen analisia.

02

Arkitektura

Eskema hibridoaren eta hodei propioaren eta publikoaren arteko karga banaketaren diseinua.

03

Zabalpena

Edukiontzitzea, orkestrazioa eta etengabeko integrazioaren kanalizazioa.

04

Ustiapena

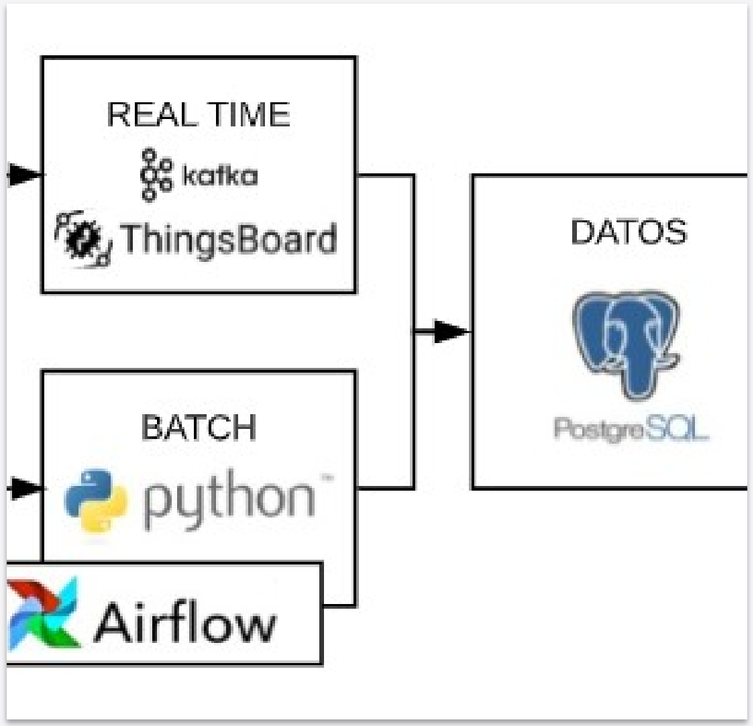
Behagarritasuna, erabilgarritasun handia eta sistemaren bilakaera produkzioan.

GALERIA

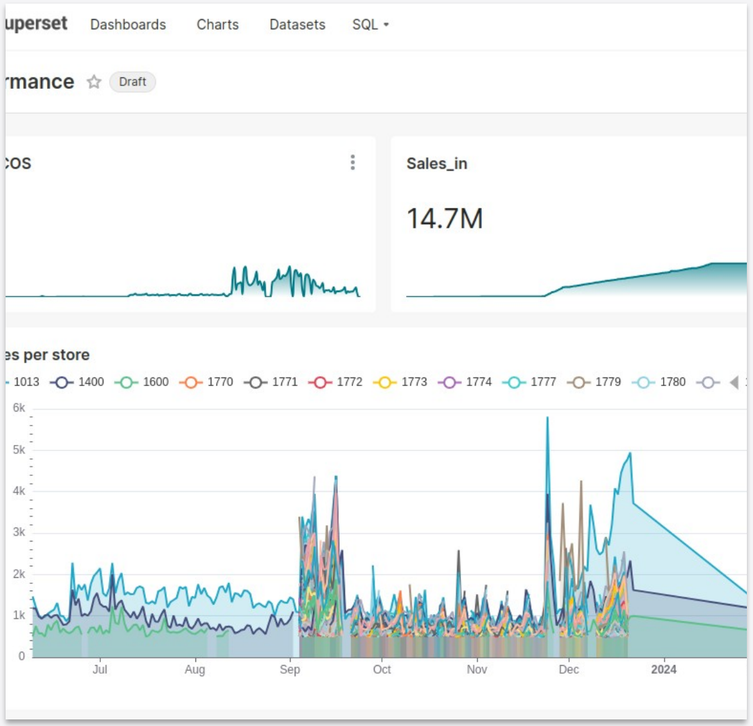
Azpiegitura eta zabalpenak



Produkzioan dagoen plataforma baten zerbitzu arkitektura.



Hodei anitzeko proiektu baten datu arkitektura.



Ustiapen eta behagarritasun aginte-koadroak.

ARRAKASTA KASUAK

Zerbitzu errealei eusten dien azpiegitura

RTVE PLAY

Egunean 7 milioi bideo baino gehiago zerbitzatuta

AA bidezko aurrebisten zerbitzua cinfo-ren azpiegituratik ematen da, CDN global bidezko banaketarekin eta eguneko milioika eskaeraren gailurrekin.

https://www.rtve.es/play/

TIIVII

Astero 750 ekitaldi ekoiztuta

Ekoizpen automatikoa, mixerra eta OTT plataformak cinfo-ren azpiegitura elastikoaren gainean exekutatzen dira, asteburuan kontzentratutako gailurrekin.

https://cinfo.es

TIIVII AI

LLM modeloak Galiziatik zerbitzatuta

cinfo-ren AA plataformak open source modeloak zerbitzatzen ditu datacenter propiotik, bezeroaren datuak adostutako ingurunetik ateratzen ez direla bermatuz.

https://cinfo.es

CINFO ZENBAKITAN

Datacenterra zenbakitan

3.328

AMD EPYC gama altuko CPU

39

NVIDIA L4 eta A40 GPU

8 TB

RAM memoria eskuragarri

614 TB

biltegiratze diskoan eta NASean

OHIKO GALDERAK

Azpiegiturari eta datuaren subiranotasunari buruzko zalantzak

Non dago fisikoki azpiegitura?

cinfo-ren datacenter propioan, A Coruñako Ciudad de las TIC-en, non talde teknikoaren bulegoak ere badauden.

Zuen hodeia nirearekin konbina dezaket?

Bai, ohiko eszenatokia da: zabalpen elastikoko sistemak kargak berdin jartzen ditu datacenter propioan, hodei publikoan edo bezeroaren zerbitzarietan.

Modeloak entrenatzeko balio du edo inferentziarako soilik?

Bietarako. GPUak egunero erabiltzen dira cinfo-ren bideo eta AA modelo propioen garapenerako, entrenamendurako eta inferentziarako.

Zer gertatzen da eskaera gailurrekin?

Ohiko eskema oinarrizko karga ahalmen propioarekin estaltzea eta gailurrak hodei publikoan xurgatzea da, eta horrek kostua kontrolpean mantentzen du.

Nork ustiatzen du plataforma?

cinfo-k ustia dezake zerbitzu kudeatu gisa, edo bezeroaren taldeak dagokion ezagutza transferentziarekin.

Zibersegurtasun betekizunak betetzen ditu?

Zibersegurtasuneko betetzea cinfo-ren garapen eta ustiapen zikloaren parte da, eta proiektu bakoitzerako dokumentatzen da.

NORENTZAT

Nork behar duen ahalmen hau

01

AA proiektuak dituzten enpresak

Entrenamendua eta inferentzia hodei publiko bakar baten mende egon gabe.

02

Sektore arautuak

Datuaren kokapena proiektuaren baldintza den kasuak.

03

Administrazio publikoak

Subiranotasun eta trazabilitate betekizunak dituzten zerbitzu digitalak.

04

Bideo plataformak

Gailur kontzentratuak eta transkodetze intentsiboko beharrak dituzten kargak.

HITZ EGIN DEZAGUN

Dimentsiona dezagun zure lan karga

Zer entrenatu edo zerbitzatu behar duzun aztertu eta hodei propioaren eta publikoaren arteko banaketa proposatzen dizugu.

Dimentsionatze bat eskatu

Datacenterra ezagutu

HARREMANA

Posta

contact@cinfo.es

Telefonoa / WhatsApp

+34 881 912 538

Weba

https://contact.cinfo.es