

ATLAS Event Index - Supervisor

Jornadas Tecnológicas 2025 – IFIC

Roberto Uzum

ATLAS Event Index

- Graduado en Ingeniería Telemática - UV
- Contratado por proyecto: ATLAS TIER 2
- IP de Proyecto:
 - Santiago González de la Hoz
 - José Enrique García Navarro
- Event Index
 - Javier Sánchez Martínez
 - Roberto Uzum



ATLAS Event Index

¿Dónde se ubica?

- Forma parte de la infraestructura de *Computing* del experimento **ATLAS** en el CERN.
- Se integra con sistemas como **AMI** (ATLAS Metadata Interface) y **Rucio** (gestión de datasets y almacenamiento distribuido).

¿Qué hace?

- Es un **catálogo global** de todos los eventos producidos en ATLAS (colisiones reales y simulaciones Monte Carlo).
- Indexa metadatos de cada evento: dataset, run, número de evento, ubicación de los ficheros, etc.

¿Cuál es su propósito?

- Permitir **buscar y localizar rápidamente eventos** en exabytes de datos distribuidos mundialmente.
- Facilitar el análisis cruzado entre datasets, runs y condiciones.
- Reducir la necesidad de acceder directamente a los archivos pesados.

¿De qué manera?

- Usando tecnologías de indexación distribuida y bases de datos escalables.

Data Collection - Event Index

- Proceso Event Index:
 - **Data Collection**
 - Recolecta y consolida información de datasets
 - Integra datos reales y simulados
 - Indexa metadatos en la base central del Event Index
 - Data Storage and Query Service
 - System Monitoring and Functional Tests
 - Event Picking Services

Event Index - Supervisor

- Herramienta que coordina e integra la ingestión y procesamiento de datos:
- Funciones principales para el Event Index:
- Aplicación backend:
 - Control de flujo de trabajos
 - Verificación de consistencia
 - Gestión de colas de tareas
- Aplicación Web desarrollada en Flask:
 - Monitorización y diagnóstico
 - Interacción, análisis y declaración de tareas obsoletas

Supervisor – Ecosistema

STOMP: Recepción de mensajes sobre nuevos datasets; registro de tareas en el Supervisor.

Rucio: Sistema de gestión y localización de datasets distribuidos. Nos ayuda a validar la consistencia de los datos a lo largo del proceso.

Tier0: Indexado de datos de eventos reales.

Panda: Indexado de datos reales y de eventos simulados en Monte Carlo.

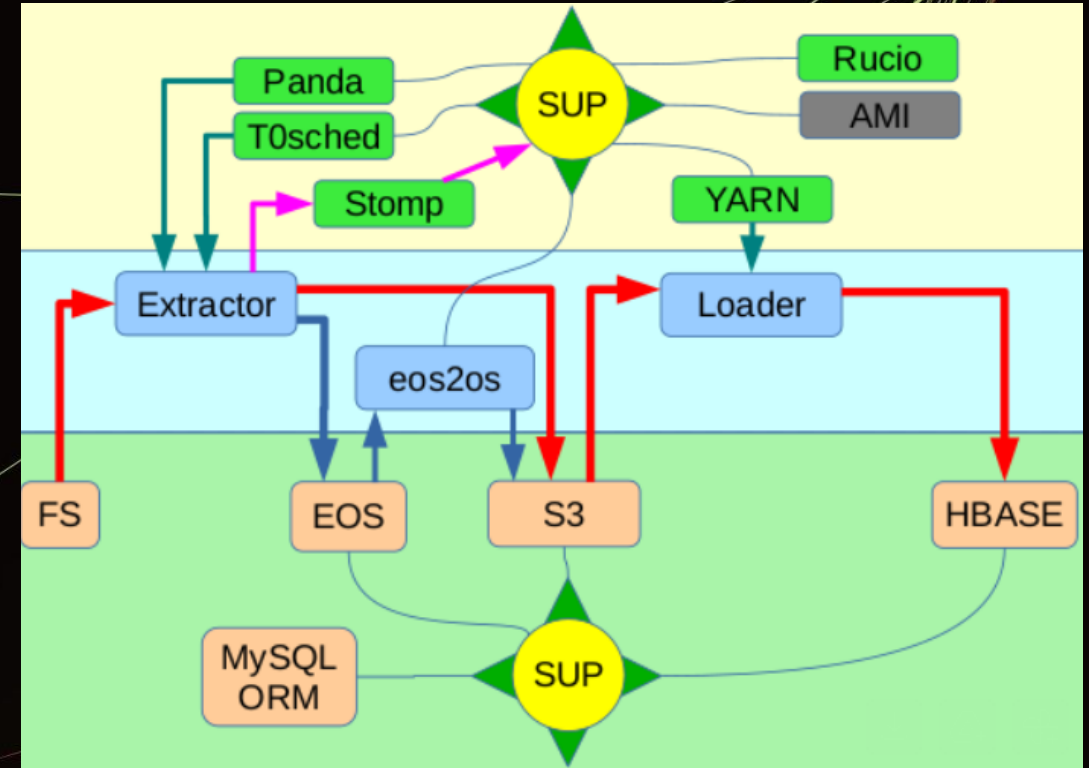
S3: Almacenamiento temporal de los metadatos indexados.

EOS: Almacenamiento alternativo de contingencia de s3.

Hadoop: Base de datos NoSQL. Consultas mediante

Phoenix.

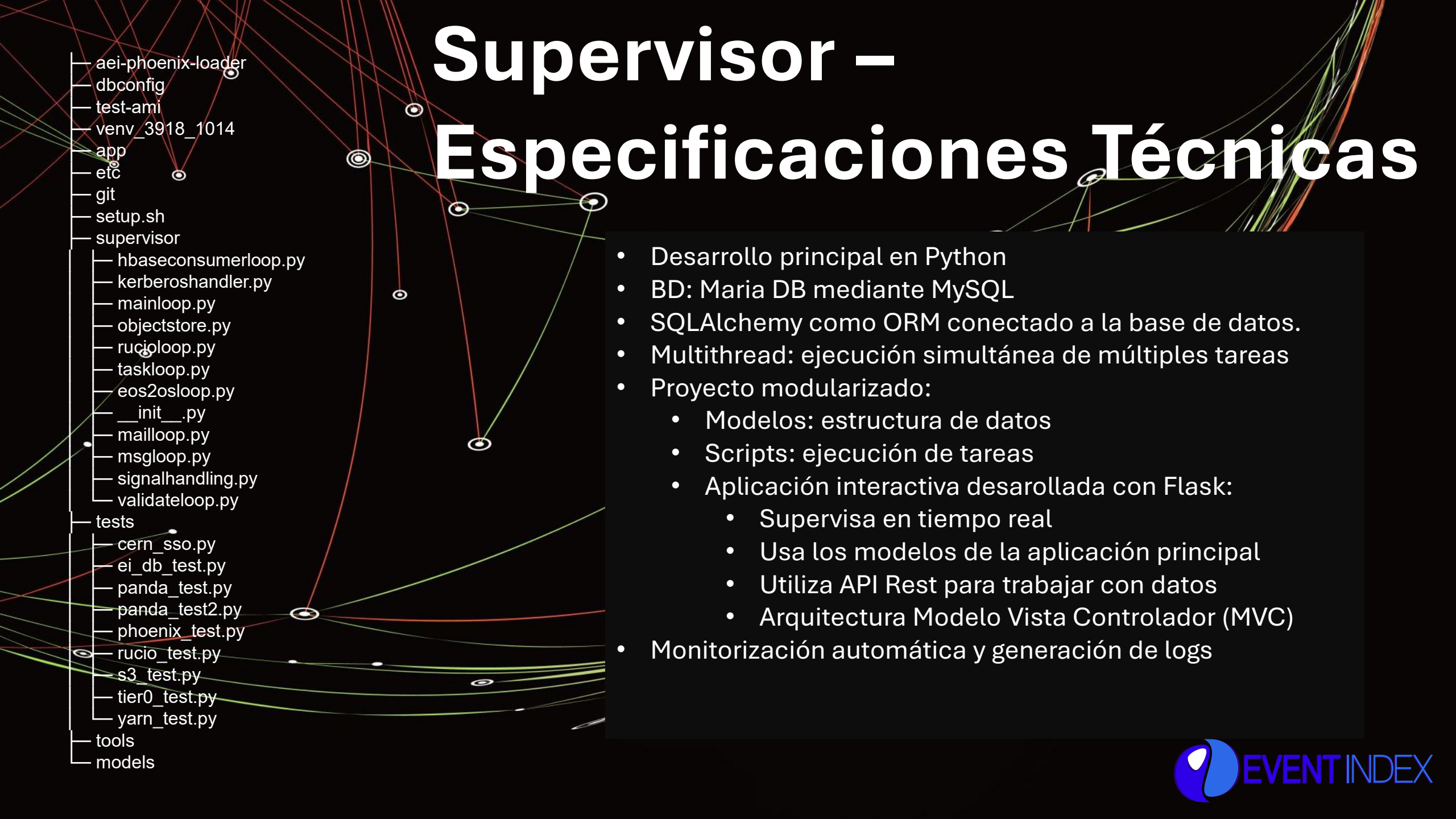
AMI: Catálogo de metadatos de datasets de ATLAS. En proceso.



El Supervisor orquesta todo el flujo: mensajes → tareas → validación → ejecución → almacenamiento.

Supervisor – Especificaciones Técnicas

- Desarrollo principal en Python
- BD: Maria DB mediante MySQL
- SQLAlchemy como ORM conectado a la base de datos.
- Multithread: ejecución simultánea de múltiples tareas
- Proyecto modularizado:
 - Modelos: estructura de datos
 - Scripts: ejecución de tareas
 - Aplicación interactiva desarrollada con Flask:
 - Supervisa en tiempo real
 - Usa los modelos de la aplicación principal
 - Utiliza API Rest para trabajar con datos
 - Arquitectura Modelo Vista Controlador (MVC)
- Monitorización automática y generación de logs



```
graph LR; aei-phoenix-loader --- dbconfig; aei-phoenix-loader --- test-ami; aei-phoenix-loader --- venv_3918_1014; aei-phoenix-loader --- app; aei-phoenix-loader --- etc; aei-phoenix-loader --- git; aei-phoenix-loader --- setup.sh; aei-phoenix-loader --- supervisor; supervisor --- hbaseconsumerloop.py; supervisor --- kerberoshandler.py; supervisor --- mainloop.py; supervisor --- objectstore.py; supervisor --- rucio.py; supervisor --- taskloop.py; supervisor --- eos2osloop.py; supervisor --- __init__.py; supervisor --- mailloop.py; supervisor --- msgloop.py; supervisor --- signalhandling.py; supervisor --- validate.py; supervisor --- tests; tests --- cern_sso.py; tests --- ei_db_test.py; tests --- panda_test.py; tests --- panda_test2.py; tests --- phoenix_test.py; tests --- rucio_test.py; tests --- s3_test.py; tests --- tier0_test.py; tests --- yarn_test.py; supervisor --- tools; supervisor --- models;
```

aei-phoenix-loader
dbconfig
test-ami
venv_3918_1014
app
etc
git
setup.sh
supervisor
hbaseconsumerloop.py
kerberoshandler.py
mainloop.py
objectstore.py
rucio.py
taskloop.py
eos2osloop.py
__init__.py
mailloop.py
msgloop.py
signalhandling.py
validate.py
tests
cern_sso.py
ei_db_test.py
panda_test.py
panda_test2.py
phoenix_test.py
rucio_test.py
s3_test.py
tier0_test.py
yarn_test.py
tools
models

Supervisor - GUI

Tier 0 Panda Active Rucio Validation Phoenix eos2OS Brokers Consumers Health

Panda Virtual Tasks

Filters

Show 25 entries Search within results

Id	Container	Creation	El. State	Files	Events	Jobs
857655	mc23_13p6TeV:mc23_13p6TeV.569187.MGPy8EG_A14NNPDF23LO_C1N2N1_LOD_1200_10ns.merge.EVNT.e8599_e8586	2025-09-16 21:52:20	EL_RUNNING	4 / 7	10,000 / 70,000	1 / 1
857654	mc23_13p6TeV:mc23_13p6TeV.543616.MGPy8EG_A14N23LO_SqSq_qN1_RPVUDD_1000_100.deriv.DAOD_PHYSLITE.e8544_e8528_a910_s4114_r14932_r14663_p6697	2025-09-16 21:08:17	EL_VALID	1 / 1	10,000 / 10,000	1 / 1

Panda Virtual Task #857654

mc23_13p6TeV:mc23_13p6TeV.543616.MGPy8EG_A14N23LO_SqSq_qN1_RPVUDD_1000_100.deriv.DAOD_PHYSLITE.e8544_e8528_a910_s4114_r14932_r14663_p6697

Panda Tasks

JEDI Task Id	Dataset	Task State	Creation	El. State	Files	Events	Jobs
46441445	mc23_13p6TeV:mc23_13p6TeV.543616.MGPy8EG_A14N23LO_SqSq_qN1_RPVUDD_1000_100.deriv.DAOD_PHYSLITE.e8544_e8528_a910_s4114_r14932_r14663_p6697_s446239333_00	spinning done	2025-09-16 18:35:05	EL_LOADED	1 / 1	10,000 / 10,000	1 / 1

This container has 1 datasets

Times

Creation	Start	State Change	Modification	End

Panda Task #46443388

mc23_13p6TeV:mc23_13p6TeV.569187.MGPy8EG_A14NNPDF23LO_C1N2N1_LOD_1200_10ns.merge.EVNT.e8599_e8586_tid46404675_00

Task

Name: mc_evind.569187.MGPy8EG_A14NNPDF23LO_C1N2N1_LOD_1200_10ns.eventIndex.e8599_e8586_j77

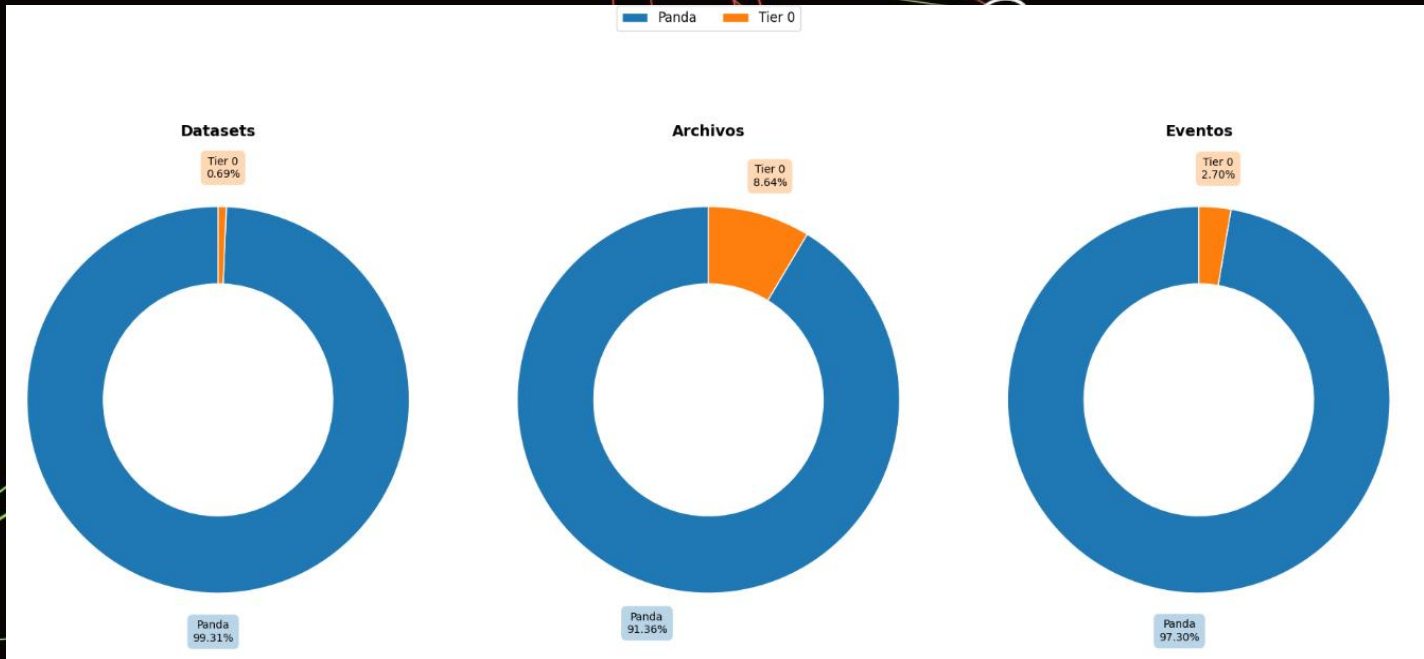
Technical Container: group.proj.evind.mc.EventIndex.OpenEnded.EVNT.i77

Id	El Id	Virtual Task	Request id	Status	Type	Username
46443388	1119152	857655	64490	done	prod	sgonzale

Task Times

Creation	Start	State Change	Modification	End
2025-09-16 21:41:43	2025-09-16 21:41:43	2025-09-16 21:52:39	2025-09-16 21:52:39	2025-09-16 21:52:39

Supervisor – Volumen de Datos

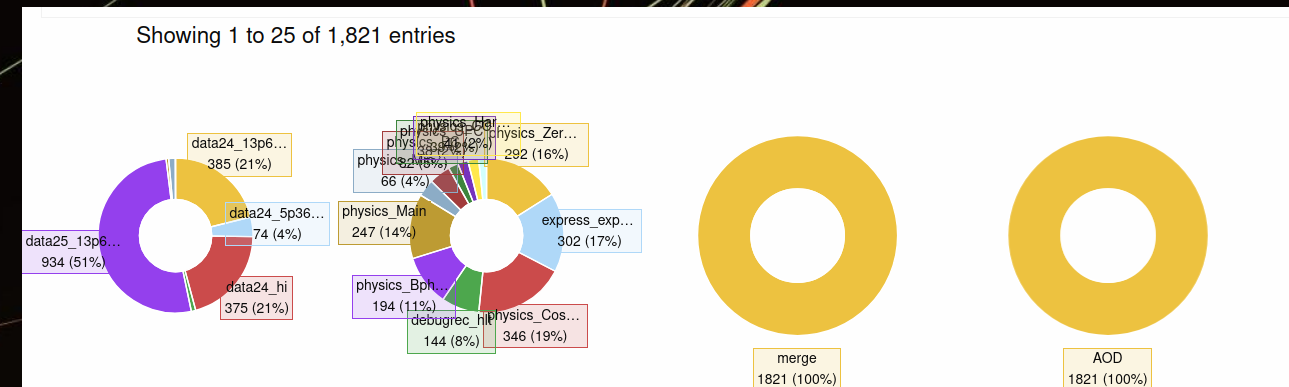
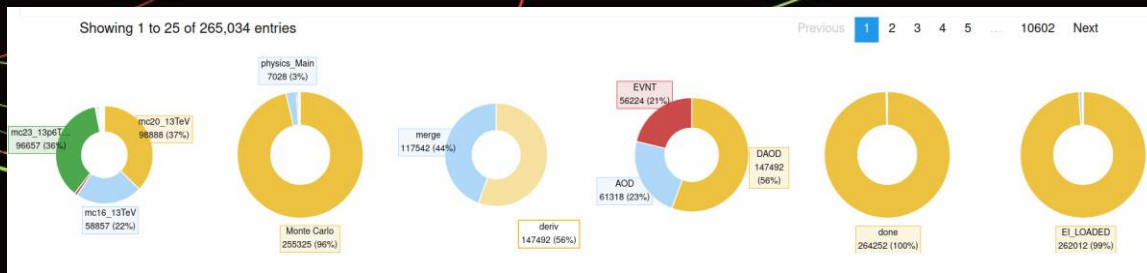
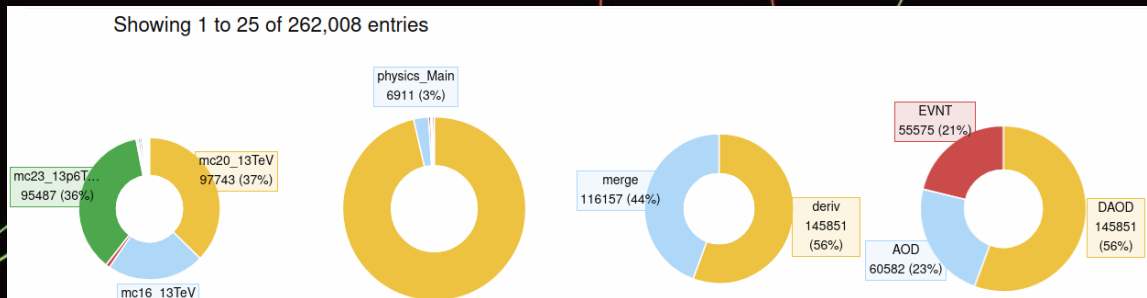


	Panda	Tier 0
Datasets	265,034	1,837
Files	17,488,816	1,652,945
Events	835 Billon	23 Billon

Supervisor – Volumen de Datos 2

Panda

Tier0



Supervisor – Trabajo Continuo

- **Automatización de procesos (tareas a mano)**
- **Monitorización y resolución de incidencias**
 - Validación automática de tareas
 - Intervención humana cuando fallan tareas, declarar tareas obsoletas.
 - Coordinación con otros equipos que nos dan soporte
- **Migración de sistemas**
 - Python 2 → Python 3
 - Actualización sintáctica y reorquestación de librerías
- **Desarrollo y nuevas funcionalidades**
 - Test unitarios para cada modulo
 - Módulo de integración con AMI
 - Filtrado y selección de candidatos
 - Attacheo de datasets a Technical Containers en Rucio



**GRACIAS POR
VUESTRA ATENCION**