A plataforma JupytherHub da DGT

Processamento de dados de LiDAR

N.º do tema de sessão técnica: Nº. 3 - Software de dados espaciais: desenvolvimento de ferramentas, portais, aplicações web e soluções técnicas

Resumo

O Laboratório de Instrumentação e Física Experimental de Partículas (LIP) em parceria com a Direção-Geral do Território (DGT), no âmbito do Sistema de Monitorização e Ocupação do Solo (SMOS), efetuou o estudo e o desenvolvimento de uma solução para o processamento em larga escala dos dados do primeiro levantamento LiDAR integral de Portugal continental. O objetivo deste trabalho consistiu na disponibilização de meios de computação com capacidade e funcionalidades adequadas à exploração e transformação dos dados com elevado desempenho. Pretende-se desta forma evitar o descarregamento dos dados para os equipamentos dos utilizadores, proporcionando simultaneamente acesso rápido aos dados e às ferramentas de software necessárias ao seu processamento através de uma plataforma unificada.

O Centro de Dados da DGT (CDD) aloja cinco coleções de dados LiDAR, as quais incluem Modelos Digitais de Superfície e Terreno com resoluções de 2m e 50cm e nuvens de pontos em formato LAZ comprimido com uma densidade média de 10 pontos/m^2. O volume total de dados compreende aproximadamente 18 Terabytes, o que aliado ao elevado número de ficheiros dificulta o seu descarregamento e processamento em sistemas *desktop* convencionais. Para obviar este problema o LIP implementou uma plataforma JupyterHub para o processamento dos dados LiDAR.

Esta plataforma combina simplicidade de acesso com a utilização de meios computacionais em nuvem e meios de computação de alto desempenho (HPC). O processamento é efetuado em Jupyter Notebooks que são criados conforme as especificações técnicas do utilizador e o seu grau de exigência computacional. A plataforma foi integrada com o CDD e permite explorar os metadados e dados LiDAR alojados nos sistemas de armazenamento de objetos e catálogos STAC do CDD. Sendo que o CDD se encontra alojado nas infraestruturas do Centro Nacional de Computação Avançada (CNCA) em Lisboa e Vila Real, a plataforma JupyterHub foi integrada com os serviços de computação e dados do CNCA, possibilitando assim acesso local direto dos meios de computação aos dados.

Esta solução simplifica a autorização e o acesso aos dados e aos meios computacionais. Disponibiliza igualmente um ambiente de software que inclui as bibliotecas e aplicações necessárias para o processamento de dados Geoespaciais,

proporcionando um ambiente de desenvolvimento persistente e extremamente flexível. O utilizador pode selecionar o ambiente de execução para Python ou R, ambiente de desenvolvimento integrado ou linha de comando, tendo ainda a possibilidade de instalar pacotes no seu ambiente de trabalho. A plataforma suporta ainda o descarregamento remoto de ficheiros para o computador do utilizador, e a integração com repositórios Git. Foram também desenvolvidos exemplos de utilização para o descarregamento, pesquisa nos catálogos STAC, processamento e visualização dos dados.

A plataforma JupyterHub enriquece o CDD com capacidade de processamento local, e disponibiliza ao utilizador um ambiente de desenvolvimento onde o foco no código é a prioridade. A plataforma está a ser alargada para o processamento de outras coleções de dados geoespaciais alojadas no CDD e no CNCA, tais como a *cache* atualizada diariamente de imagens Sentinel-2A de Portugal, ortofotos e produtos criados pela DGT.

Palavras chave

Notebooks, JupyterHub, LiDAR, Centro de Dados da DGT, Sentinel-2, CNCA, HPC

Autores

Pedro Pinheiro
ppinheiro@lip.pt
LIP
João Martins
martinsj@lip.pt
LIP
Zacarias Benta
zacarias@lip.pt
LIP
João Pina
jpina@lip.pt
LIP
Mário David
david@lip.pt
LIP
Samuel Bernardos
samuel@lip.pt
LIP
Gonçalo Barradas
gbarradas@lip.pt
LIP
Carlos Manuel
carlos@lip.pt
LIP
Jorge Gomes
jorge@lip.pt
LIP