

El BSC fa pública la versió 2.10 de PyCOMPSs/COMPSs i la versió 0.7.0 de la dislib



El Barcelona Supercomputing Center ofereix COMPSs a la comunitat HPC un conjunt d'eines que ajuden els desenvolupadors a programar i executar les seves aplicacions de forma eficient en infraestructures computacionals distribuïdes. En particular, PyCOMPSs es la interfície per aplicacions Python, component emprat per la convergència entre la supercomputació, intel·ligència artificial i big data.

Aquesta versió de COMPSs inclou suport per a serveis REST, tasques "request" a serveis http que permetran un nou tipus de fluxos de treball que combinen tasques tradicionals i tasques http. El mecanisme també es pot utilitzar amb tasques de Funció com a servei (FaaS) predesplegades.

Per les aplicacions en Python, PyCOMPSs inclou una nova funció d'anàlisi dels accessos a les dades que permet decidir quins paràmetres tenen una taxa de reutilització més alta i obtenir beneficis d'emmagatzemar-los a la memòria cau.

La nova versió de la dislib inclou una versió de la descomposició QR a blocs amb tres modes, regressors de random forest paral·lels, i una multiplicació de matrius que permet fer la transposta dels arguments in-situ.

El grup [Workflows and Distributed Computing](#) del Barcelona Supercomputing Center-Centro Nacional de Supercomputación (BSC) està orgullós d'anunciar el llançament de la versió 2.9 (nom en clau Jasmine), de l'entorn de programació [COMPSs](#).

Aquesta versió de COMPSs actualitza el resultat del treball de l'equip en els darrers anys en la provisió d'un conjunt d'eines que ajuden els desenvolupadors a programar i executar les seves aplicacions de forma eficient en infraestructures computacionals distribuïdes, com ara clústers, núvols i clústers gestionats amb contenidors. COMPSs és un model de programació basat en tasques capaç de millorar notablement el rendiment d'aplicacions a gran escala mitjançant la paral·lelització de manera automàtica de la seva execució.

COMPSs ha estat disponible en els últims anys pels usuaris del supercomputador MareNostrum i a la Red Española de Supercomputación (RES) i ha estat emprat en diversos projectes de recerca com ara EUBra-BIGSEA, MUG, EGI, ASCETIC, TANGO, NEXTGenIO, I-BiDaaS, mF2C i CLASS. En aquests projectes s'ha fet servir COMPSs per desenvolupar casos d'ús proporcionats per diferents comunitats de diverses disciplines com la biomedicina, l'enginyeria, la biodiversitat, la química, l'astrofísica, aplicacions financeres, telecomunicacions, processos de fabricació i ciències de la terra. Actualment també està en extensió i en ús en aplicacions en els projectes AI-SPRINT, ExaQUte, LANDSUPPORT, el BioExcel CoE, PerMedCoE, ELASTIC i en el projecte Edge Twins HPC del tipus Innovation Launchpad. També ha estat aplicat a un cas d'ús del ChEESE CoE. Una menció especial és el projecte eFlows4HPC coordinat per el grup, iniciat el gener del 2021, que té com a objectiu desenvolupar un programari per workflows on un dels components principals és l'entorn PyCOMPSs / COMPSs.

La nova versió inclou suport per a serveis REST en forma de tasques "request" a serveis http, tant per a aplicacions Python com Java. El mecanisme també es pot utilitzar amb tasques de Funció com a servei (FaaS) predeplegades. Aquest suport millora el suport anterior de COMPSs als serveis web que es limitava als serveis web SOAP i a les aplicacions Java COMPS. La sintaxi s'ha ampliat amb anotacions per a les tasques http, així com amb el suport d'execució COMPSs corresponent. Aquesta funcionalitat permet un nou tipus de fluxos de treball que poden combinar tasques tradicionals amb el codi proporcionat pel desenvolupador de l'aplicació, amb invocacions a tasques http externes.

La versió anterior 2.9 proporcionava suport per a una memòria compartida per a les tasques Python per evitar el temps de serialització i deserialització dels arguments. La versió actual amplia aquesta funcionalitat amb un mecanisme de suport a la decisió per als desenvolupadors que fa un anàlisi de quins paràmetres de les tasques s'han d'emmagatzemar a la memòria cau i quins no. Atès que l'emmagatzematge a la memòria cau comporta una sobrecàrrega addicional, el mecanisme proporciona informació sobre la reutilització dels paràmetres en termes d'accés a la memòria cau que permet al desenvolupador inferir quins es beneficiaran de ser emmagatzemats a la memòria cau.

Altres millores són l'extensió del suport per a tasques MPI amb el paràmetre `processes_per_node` que permet la distribució de processos als nodes, suport per a versions de Java superiors a 8, suport parcial per a sistemes macOS i suport de serialització d'objectes PyArrow.

A més, la versió 2.10 de COMPSs ve dotada amb altres funcionalitats menors, extensions i resolució d'errors.

COMPSs ha tingut al voltant de **1000 descàrregues** durant l'any passat i és utilitzat per uns **20 grups** en aplicacions reals. Recentment, COMPSs ha atret l'interès d'àrees com la enginyeria, el reconeixement d'imatges, la genòmica i la sismologia, àrees on s'han realitzat cursos específics i accions de divulgació.

Els paquets i la llista completa de funcions estan disponibles a la pàgina de [descarregues](#). En aquesta pàgina també es pot trobar una maquina virtual que permet provar les funcionalitats de COMPSs a través d'un tutorial pas a pas que guia a l'usuari en el desenvolupament i execució d'un conjunt d'aplicacions d'exemple.

A més del software, hi ha un conjunt de guies d'usuari i administrador, i articles publicats a conferències i revistes rellevants.

Per més informació, visiteu la nostre pàgina web: <http://www.bsc.es/compss>

El grup també anuncia la nova versió de dislib 0.7.0. La biblioteca de computació distribuïda (dislib) proporciona algorismes distribuïts preparats per utilitzar-los com a biblioteca. Fins ara, dislib s'ha centrat en algorismes d'aprenentatge automàtic i ofereix una interfície inspirada en scikit-learn. L'objectiu principal de dislib és facilitar l'execució d'algoritmes d'anàlisi de big data en plataformes distribuïdes, com ara clústers, núvols i superordinadors. Dislib s'ha implementat a fent servir el model de programació PyCOMPSs.

Dislib es basa en una estructura de dades distribuïdes, ds-array, que permet l'execució paral·lela i distribuïda dels mètodes d'aprenentatge automàtic. El codi de la biblioteca dislib s'implementa com una aplicació PyCOMPSs, on els diferents mètodes s'anoten com a tasques PyCOMPSs. En el moment de l'execució, PyCOMPSs s'encarrega de tots els aspectes de paral·lelització i distribució de dades. Tanmateix, el codi d'usuari de l'usuari final de la dislib omet els aspectes de paral·lelització i distribució, i està escrit com a simples scripts de Python, amb una interfície molt semblant a la interfície de scikit-learn. Dislib inclou mètodes d'agrupació, classificació, regressió, descomposició, selecció de models i gestió de dades. Un contracte de recerca amb FUJITSU havia finançat parcialment la biblioteca dislib i es va utilitzar per avaluar el processador A64FX. Actualment, els desenvolupaments dislib estan cofinançats en un 50% pel Fons Europeu de Desenvolupament Regional en el marc del Programa Operatiu FEDER 2014-2020 amb el suport de la Secretaria d'Universitats i Recerca, pel projecte H2020 AI-Sprint i pel projecte EuroHPC eFlows4HPC.

Des de la seva creació, dislib s'ha aplicat en casos d'ús d'astrofísica (DBSCAN, amb dades de la missió GAIA), anàlisi de resultats de dinàmiques moleculars (Daura i PCA, BioExcel CoE). En el projecte eFlows4HPC, s'està aplicant en dos casos d'ús: en informàtica urgent per a riscos naturals (regressors random forests) i en bessons digitals per a la fabricació (QR).

La nova versió 0.7.0 inclou una implementació paral·lela dels regressors random forest i la descomposició QR a blocs amb tres modes: complet, econòmic i r. La implementació de QR es basa en reflectors Householder amb l'enfocament inspirat en les rotacions de Givens proporcionant un millor paral·lelisme. La nova versió també conté una implementació de MinMax Scaler per al preprocessament de les dades d'entrada, noves funcions d'utilitat per matrius i eliminar les darreres files o columnes, i una multiplicació de matrius amb transposat de les matrius in-situ. També s'han fet diverses millores de rendiment en altres algorismes. A més d'això, totes les tasques implementades a dislib a partir d'ara poden utilitzar un nombre especificat de processadors.

Dislib 0.7.0 ve amb altres extensions i amb una nova guia d'usuari. El codi es obert i esta disponible per la seva [descarrega](#).

Barcelona Supercomputing Center - Centro Nacional de Supercomputación

Source URL (retrieved on 4 oct 2024 - 14:17): <https://www.bsc.es/ca/noticies/noticies-del-bsc/el-bsc-fa-p%C3%BAblica-la-versi%C3%B3-210-de-pycompsscompss-i-la-versi%C3%B3-070-de-la-dislib>