

[Inici](#) > El BSC fa pública la versió 2.8 de COMPSs i la versió 0.6.0 de la dislib

El BSC fa pública la versió 2.8 de COMPSs i la versió 0.6.0 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; i la dislib, una llibreria de machine learning paral·lelitzada amb PyCOMPSs.



Aquesta versió inclou una extensió per millorar la programabilitat de les reduccions definides per l'usuari. Amb aquesta addició, la reducció es pot descriure fàcilment amb una anotació a les tasques, el runtime és capaç de descomposar la tasca en diverses tasques que es poden executar en paral·lel i que tenen en compte la localitat de les dades a nivell de node, reduint les comunicacions inter-node.

S'ha afegit una nova anotació de tasques @container que permet l'orquestració d'aplicacions PyCOMPSs que inclouen tasques desplegadas en contenidors. COMPSs dona suport als entorns de contenidors Docker i Singulairy.

La nova versió de la dislib inclou Daura, l'algorisme de clusterització més utilitzat per la comunitat de dinàmica molecular; un PCA amb SVD paral·lel; i nous mètodes bàsics com la multiplicació de matrius o el producte Kronecker.

El grup [Workflows and Distributed Computing team](#) del Barcelona Supercomputing Center-Centro Nacional de Supercomputación (BSC) està orgullós d'anunciar el llançament de la versió 2.8 (nom en clau Iris), 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 mF2C. 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 i ciències de la terra. Actualment també està en extensió i en ús en aplicacions en els projectes ExaQute, LANDSUPPORT, el BioExcel CoE, CLASS, ELASTIC, i l'EXPERTISE ETN, així com en un contracte de recerca amb FUJITSU i en el projecte Edge Twins HPC del tipus Innovation Launchpad. També ha estat aplicat a un cas d'ús del ChEESE CoE.

La nova versió inclou una extensió per millorar la programabilitat de les reduccions definides per l'usuari. S'ha definit una nova directiva `@reduction` per anotar les reduccions definides per l'usuari. El runtime de COMPSs és capaç de descomposar la tasca de reducció en diverses que poden executar-les en paral·lel seguint una forma d'arbre invertit. El runtime executa un algorisme intern que té en compte la localitat de les dades, realitzant primer la reducció de les d'un mateix node i reduint posteriorment les dades entre nodes. Això redueix extremadament el nombre de comunicacions entre nodes en comparació amb una solució que no té en compte la localitat.

El runtime de COMPSs permet el desplegament fent servir contenidors des de la versió 1.4. No obstant això, el suport anterior es limitava a empaquetar tota l'aplicació en un sol contenidor. A partir d'aquesta imatge, el runtime era capaç de desplegar-la en diversos contenidors i iniciar l'execució de l'aplicació sense la interacció de l'usuari mitjançant l'entorn de contenidors. En entorns suportats són Docker, Singularity i Mesos. A la versió 2.8, el suport per a contenidors s'ha ampliat amb una nova directiva `@container` que s'utilitza per anotar tasques incloses en una imatge de contenidor. En aquest sentit, el runtime de COMPSs es converteix en un orquestrador de workflows compostos per tasques natives i tasques incloses en imatges de contenidors, podent interactuar amb el entorn de contenidors sense la intervenció de l'usuari. Les tasques anotades amb `@container` poden ser invocació a binaris externs o codi d'usuari proporcionat per el desenvolupador de l'aplicació. En aquest segon cas, el codi d'usuari s'ha desplegat prèviament a la imatge del contenidor.

La versió 2.8 de COMPSs també inclou: un nou tipus de paràmetre, `IN_DELETE`, que especifica paràmetres d'un sol ús que s'esborren després de l'execució de la tasca; un nou format de dades per a col·leccions en tasques MPI; i la possibilitat de fer servir col·leccions de diccionaris Python com a paràmetres.

A més, la versió 2.8 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>



Nova versió de dislib

El grup també està orgullós d'anunciar la nova versió de [dislib](#) 0.6.0. La Biblioteca de Computació Distribuïda (dislib) proporciona algorismes distribuïts preparats per utilitzar-se com a biblioteca. Fins ara, dislib se centra en algorismes de machine learning i ofereix una interfície inspirada en scikit-learn. L'objectiu principal de dislib és facilitar l'execució d'algorismes d'anàlisi de dades massives en plataformes distribuïdes, com ara clústers, clouds i superordinadors. Dislib s'ha implementat a sobre del model de programació PyCOMPSs, la interfície Python de COMPSs.

Dislib es basa en una estructura de dades distribuïda, ds-array, que permet l'execució paral·lela i distribuïda dels mètodes de machine learning. El codi de la biblioteca de dislib s'implementa com una aplicació PyCOMPSs, on els diferents mètodes s'anoten com a tasques de PyCOMPSs. En el moment d'execució, PyCOMPSs s'encarrega de tots els aspectes de paral·lelització i distribució de dades. No obstant això, el codi final de l'usuari de dislib desconeix els aspectes de paral·lelització i distribució, i s'escriu com scripts Python simples, amb una interfície molt similar a la interfície scikit-learn. Dislib inclou mètodes per fer clustering, classificar, regressió, descomposició, selecció de models i gestió de dades. Un contracte de recerca amb FUJITSU ha finançat parcialment la biblioteca dislib i actualment s'utilitza per avaluar el processador A64FX.

Des de la seva recent creació, dislib s'ha aplicat en casos d'ús d'astrofísica (DBSCAN, GAIA) i en fluxos de treball de dinàmica molecular (Daura i PCA, BioExcel CoE).

La nova versió 0.6.0 inclou una nova regressió lineal multivariant, una implementació paral·lela de la descomposició del valor singular (SVD), una nova versió de l'anàlisi de components principals (PCA) que utilitza aquest SVD paral·lel, l'algorisme ADMM Lasso i l'algorisme de clusterització Daura.

Dos d'aquests codis, PCA amb SVD i Daura clustering, s'han desenvolupat tenint en compte les necessitats de la comunitat BioExcel, que requereix analitzar els resultats de grans trajectòries resultants de simulacions de dinàmiques moleculars. Daura és l'algorisme de clusterització utilitzat principalment per la comunitat de dinàmiques biomoleculars, i s'inclou per exemple en el conjunt d'eines GROMACS. No obstant això, les implementacions anteriors d'aquest algorisme són seqüencials i es limiten a l'ús d'una memòria d'un sol node. Aquesta limitació se supera amb la nova implementació de dislib, que permet agrupar trajectòries de simulació molt grans. De la mateixa manera, una versió anterior del PCA de la dislib utilitzava un mètode seqüencial per trobar els valors propis i els vectors propis d'una matriu de covariança. Aquest mètode seqüencial limitava l'acceleració del PCA i la mida dels problemes a resoldre. El nou PCA ve amb un SVD paral·lel que supera la limitació anterior i pot utilitzar la memòria de diversos nodes.

Dislib 0.6.0 ve amb altres extensions i amb una nova guia de l'usuari. El codi és de codi obert i està disponible per [descarregar-lo](#).

Barcelona Supercomputing Center - Centro Nacional de Supercomputación

Source URL (retrieved on 16 jul 2024 - 09:52): <https://www.bsc.es/ca/noticies/noticies-del-bsc/el-bsc-fa-p%C3%BAblica-la-versi%C3%B3-28-de-compss-i-la-versi%C3%B3-060-de-la-dislib>