MOVING BIODIVERSITY TO THE CLOUD

Change on the presentation due to interest

Javier de la Torre
Andrew W Hill
Tim Robertson

Friday, November 13, 2009
What is cloud computing?
Scalable web applications

Data Processing
Detecting novel influenza
http://biodivertido.blogspot.com/
Intersecting GBIF network data with the World Database on Protected Areas
160M points intersect 150K polygons
We need data, tools and methodologies
TDWG Cloud Computing beerSession

Friday, November 13, 2009
1) Data Sets on the Cloud

GBIF species occurrence
BHL text data
BHL aging data
IUCN range maps
IUCN species status
EOL content
BOLD

ALA
TreeBASE
WDPA
MorphBank
PESI
AntBase
Public Data Sets on AWS

Public Data Sets on AWS provides a centralized repository of public data sets that can be seamlessly integrated into AWS cloud-based applications. AWS is hosting the public data sets at no charge for the community, and like all AWS services, users pay only for the compute and storage they use for their own applications.

Previously, large data sets such as the mapping of the Human Genome and the US Census data required hours or days to locate, download, customize, and analyze. Now, anyone can access these data sets from their Amazon Elastic Compute Cloud (Amazon EC2) instances and start computing on the data within minutes. Users can also leverage the entire AWS ecosystem and easily collaborate with other AWS users. For example, users can produce or use prebuilt server images with tools and applications to analyze the data sets. Users can also discuss the services and solutions in the dedicated forums.

By hosting this important and useful data with cost-efficient services such as Amazon EC2, AWS hopes to provide researchers across a variety of disciplines and industries with tools to enable more innovation, more quickly.
OpenModeller
Maxent
Hadoop
Disco
HBase/HTable
Voldemort
BlastReduce
Clustal
MAFFT
PyCogent
GDAL
Clustr
HIT Harvester
AMQP
MrBayes
Paup*
Muscle
MAFFT
ImageMagick
GraphViz
SAM
TaxonFinder
ECAT Name Parser
Tesarac
GeoCR
ABCD Parse
This code site relates to the Biodivertido blog and serves as a repository for biodiversity informatics code snippets in several programming languages that eventually will be integrated into a common, reusable library for each language.

Languages of focus initially are Java, PHP, Ruby, Python and .Net

The biodivertido blog was created to compile all different "experiments" the authors do related to Biodiversity Informatics. Main topics have to do with UI, GIS, Grid technology, Computing in the cloud, Scalable architectures, RIAs, Web services, public web APIs, standards.

Project Home | Downloads | Wiki | Issues | Source | Administer

Summary | Updates | People

Code license: Apache License 2.0
Labels: biodiversity, tdwg, dwc

Blogs: Biodivertido
Feeds: Project feeds

Project owners: wixner, jatorre, timrobertson100

Project committers: movesyside, dhobem, peter.desmet.cubc, a.kehlebecker@bgbm.org, ben.richardson@dec.wa.gov.au, rycroft, mikegiddens@silverbiology.com