Atmosphere
iPlant Collaborative’s Cloud Computing
1. Introduction

1.1 Cloud Computing
1.2 Cloud Types
1.3 Cloud Service
1.1 Cloud Computing

Cloud Computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and related with minimal management effort or service provider interaction (The National Institute of Standards and Technology)
1.2 Cloud Types

- Public Cloud
- Private Cloud
1.3 Cloud Service

- **Infrastructure as a Service (IaaS)**
  A platform virtualization environment/service. Rather than purchasing servers, software, data-center space or network equipment, clients instead buy those resource as a fully outsourced service.

- **Platform as a Service (PaaS)**
  A computing platform and/or solution stack as a service. It facilitated deployment of applications without the cost and complexity of buying and managing the underlying hardware and software layers.

- **Software as a Service (SaaS)**
  Software as a service over the Internet, eliminating the need to install and run the application on the customer’s own computers and simplifying maintenance and support.
2. Cloud in Industry

2.1 List of Cloud offers by Industry

2.2 Is industry offers enough to run your research?
2.1 List of Cloud Services

- Amazon Web Services
- Rack Space
- Windows Azure
- IBM Smart Cloud
- Google (App Engine, etc)
2.2 Is industry offers enough to run your research?
3. iPlant Cloud / Atmosphere
What we are offering

3.1 What Atmosphere is
3.2 Atmosphere’s Service Layers
3.3 Case Study I – Workshops
3.4 Case Study II – Researchers uses
3.5 Case Study III – Developers uses
3.6 What makes people like Atmosphere?
3.7 Architecture of Atmosphere
3.8 Features
3.9 APIs
3.1 What Atmosphere is

- iPlant Collaborative’s infrastructure as a Cloud Service based on on-demand server/storage virtualization
- Easy-to-use web based interface that can be run by researchers, IT staffs, and developers
- Providing various image packages for agile research project phase
3.2 Atmosphere’s Service layers
3.3 Case Study I – Workshops

3.3.1 Atmosphere uses for workshops

3.3.1 Next-Generation Sequence Workshops - CBSU at Cornell University

3.3.2 Cold Spring Harbor Laboratories, CSHL, Workshops
3.3.1 Atmosphere uses for workshops

- Atmosphere’s on-demand cloud resources are suitable for science workshops that requires isolations between resources and pre-deployed application stacks.

- Isolated work space (Application stacks on virtualized machine) provides stable/independent work environment

- Pre-deployed application stacks in VM provides agile research phase.
3.3.2 Next-Generation Sequence Workshops - CBSU at Cornell University

Workshops on the topic of next-generation sequencing, organized jointly by the Computational Biology Service Unit (CBSU), the Center for Vertebrate Genomics (CVG), and the Cornell Center for Comparative and Population Genomics (3CPG), and open to members of CVG and 3CPG laboratories.
3.3.2 Next-Generation Sequence Workshops - CBSU at Cornell University

Workshop machine images/applications
- Burrows-Wheeler Aligner (BWA)
- Samtools
- Picard
- Genome Analysis Toolkits (GATK)
- Annovar: Functional annotation of genetic variants from high-throughput sequencing data
- Integrative Genomics Viewer (IGV)

30 Workshop attendees launching their own computing resources with 16G Ram, 4 CPUs
3.3.3 Cold Spring Harbor Laboratories Workshops
3.4 Case Study II – Researchers use

3.4.1 Atmosphere uses for researchers
3.4.2 COGE – University of California at Berkley
3.4.3 Douglas-fir climate change transcriptome observatory by Corvallis Forestry Science Lab, USDA
3.4.1 Atmosphere uses for researchers

- For plant science researchers, Atmosphere is the space that they can obtain on-demand computational resources.
- Atmosphere provides not only computational resources but also stacks of applications and ready to run environment.
- Researchers can get benefits from running existing application on Atmosphere and/or running Atmosphere based design application.
3.4.2 COGE

- COGE is a comparative genomics platform for all genomes across all domains of life in any state of assembly.
- COGE-X and COGE-API were developed for cloud / distributed style’s COGE use.
- COGE Users access Atmosphere and launch COGE instance/application and access COGE Dataware house.
- http://www.genomeevolution.org/CoGe/
• COGE / Atmosphere architecture figure goes here
3.4.3 Douglas-fir climate change transcriptome observatory

Brian J. Knaus (Corvallis Forestry Science Lab, USDA Forest Service)

“We are looking for cyberinfrastructure to support the analysis of our funded Douglas-fir climate change transcriptome observatory. This project seeks to characterize differential expression in Douglas-fir (Pseudotsuga menzeisii) throughout a complete growing season. Tissue collection is being performed bi-monthly for RNA-Seq transcriptome profiling. Concurrently, biotic and abiotic measurements are being made in order to help correlate gene expression patterns with environmental conditions and the phenotypic response to these conditions.

This experiment is projected to generate over 85 billion lines of genotypic data (Illumina/ Solexa sequencing output). Organizing and analyzing 26 temporal sample points throughout a single growing season requires random access to a large and growing database of sequence data which will be compared to databases of historic data (from previous research using these accessions) and contemporary phenotypic and abiotic data.

Computational requirements are evolving but currently include:

- RNA-Seq analysis pipelines using flat file and R/MySQL database systems
- Housing of phenotypic and abiotic data in MySQL databases
- GMOD services

Our collaborators are distributed throughout the United States. Use of the Atmosphere cyberinfrastructure would allow researchers of differing backgrounds to share their entire analytic pipeline in a reproducible, modifiable and persistent fashion. It would also allow us to share these results via GMOD to the greater plant genetics community.”
3.4.3 Douglas-fir climate change transcriptome observatory

Requested and provided application stack for Douglas-fir climate change transcriptome observatory by Atmosphere

- b2g4pipe (blast2go.org)
- bioperl (www.bioperl.org)
- blast (blast.ncbi.nlm.nih.gov)
- blat (genome.ucsc.edu)
- bowtie (bowtie-bio.sourceforge.net)
- bwa (bio-bwa.sourceforge.net)
- cap3 (seq.cs.iastate.edu)
- cashx (jcclab.science.oregonstate.edu)
- cufflinks (cufflinks.cbcb.umd.edu)
- lastz (www.bx.psu.edu/~rsharris/lastz)
- mafft (mafft.cbrc.jp/alignment/software)
- muscle(www.drive5.com/muscle)
- Perl (www.perl.org)
- R (www.r-project.org)
- samtools (samtools.sourceforge.net)
- tophat (tophat.cbcb.umd.edu)
- Velvet (www.ebi.ac.uk/~zerbino/velvet/)

- R Packages (mirrored and installed within R)
  - Ape
  - phangor
  - Seqinr
  - RMySQL (required by tileQC)

- Bioconductor (R) packages
  - edgeR
  - Qvalue
  - Biostrings

- Perl Modules
  - Inline::C (http://search.cpan.org/~sisyphus/Inline-0.47/C/C.pod)
3.5 Case Study III – Developer’s use

- iPlant Collaborative’s Discovery Environment
- iPlant’s Tool Editor/Development application is based on virtual environment that Atmosphere provides
- Developing prototype application on extremely flexible and safe environment
- Developers can get root access of VM and do various programming experience and programming
3.6 What makes people like Atmosphere?

- 3.6.1 Easy to use, Really
- 3.6.2 Agility
3.7 Architecture of Atmosphere
3.8 Features
3.9 APIs

- 3.8.1 APIs
- 3.8.2 Integration with existing project
3.9.1 APIs

- Atmosphere provides HTTP protocol based RCP call style API
- Atmosphere APIs are used by Atmosphere Web Interface as well as
  - iPlant Discovery Environment’s Tool Development environment
  - Android Application for Atmosphere
- Simple to easy but powerful APIs that provide full access of Atmosphere’s feature
3.9.2 Integration with existing project

- With provided Atmosphere’s API / Web Hook / Image APIs, third party applications can easily integrate Atmosphere features into their existing applications.

- For example, launching/controlling/terminating Atmosphere’s application/instance within existing application or web site can be done with simple web/http calls.
4. Demo