Cloud computing and NGS:
massively parallel computing for massively parallel sequencing

Pareja-Tobes E. Manrique M, Pareja-Tobes P, Pareja E, Tobes R

Era7 Information Technologies SLU, Granada. Spain

eduardopt@era7.com

More info at: **www.era7bioinformatics.com**

Research using new massively parallel sequencing technologies is continuously generating huge amount of data that are especially suited to be managed by means of new massively parallel computing and storage models.

Next generation sequencing and other recent technologies have improved the accessibility of sequencing but in many cases the bottle-neck has now moved to the bioinformatics analysis, tightly related to the availability of high-throughput computation infrastructure. Cloud computing fits perfectly with these new bioinformatics needs and, improving the accessibility of high-throughput computation, specifically with "infrastructure as a service" cloud services.

Paradigms such as Map/Reduce, XML-based queryable data storage, SOAP/REST SOA architectures, messaging frameworks (AWS SQS, cloudMQ, ...), concurrent programming models such as Actors (Scala, Erlang) are especially adapted both to cloud computing and bioinformatics needs.

Key factors of cloud computing are automatic provisioning of infrastructure, on-demand scalability, capacity to adapt programmatically your computing infrastructure in real-time by means of API calls, security and reliability.

NGS data management using cloud computing allows you to maintain the focus on knowledge-driven analysis not on infrastructure provisioning:
manage your data, not your datacenters.

http://www.era7bioinformatics.com