



Contribution ID: 115

Type: **Plenary**

THE POTENTIAL OF GENOMIC MEDICINE—THE MANY ASPECTS OF COMPUTING ARE DEMANDING AND THE COMPLEXITY MUST BE MANAGED.

Monday, 1 September 2014 11:50 (35 minutes)

Janis will speak about Building an architected genomics pipeline platform that will extend to support the analytics, data management, data provenance, long-term retention, and especially the issues as we genomics becomes part of the clinical information for patients.

High-performance best practices in computing and storage solutions are required to process the data produced by Next Generation Sequencing that is doubling every five months. Of the four phases in a sequencing project: a. Experimental design and sample collection, b. Sequencing c. Data management and d. Downstream analysis, IBM has optimized the data management aspect. It takes a highly optimized HPC platform to keep pace with the genomic data analysis. The algorithms are typically I/O intensive. Additionally, the genomics pipeline workflow must be optimized in order to fully utilize the available resources. The data must typically be archived so that it stored cost effectively and, in the cases of clinical genomics, for many years. IBM, working with our life cycle management and hierarchical storage management will address the requirement for long-term data archive.

The primary goal of a sequencing project is to use the data for extensive downstream analytics with corresponding phenotypic information, image analysis, published scientific discovery, and the use of other internal and external data sources, allowing researchers to obtain insights. IBM will address the work being done to integrate genomics data into translational platforms, with the goal of personalized medicine.

Primary author: LANDRY-LANE, Janis (IBM)

Presenter: LANDRY-LANE, Janis (IBM)

Session Classification: Plenary

Track Classification: Computing Technology for Physics Research