SPARK - A Supercomputing Pipeline for The Ark

Thilina Ranaweera, **Enes Makalic, John Hopper, Adrian Bickerstaffe**
Centre for Epidemiology and Biostatistics, *The University of Melbourne*

Introduction

Advances in genotyping technologies have paved the way for cost-effective and large scale human genetic research. Genome-wide Association Studies (**GWAS**) detect genetic associations with complex human diseases by analysis of very large data sets. The first approaches have been conservative and left the vast majority of familial associations unexplained.

Research groups across the globe are developing novel GWAS analysis techniques to try to resolve more of the 'missing heritability'. These require large amounts of computational power and are limited to supercomputers or the like, with limited access. **SPARK** addresses this problem by developing a supercomputer pipeline to The Ark, a web-based data management system.
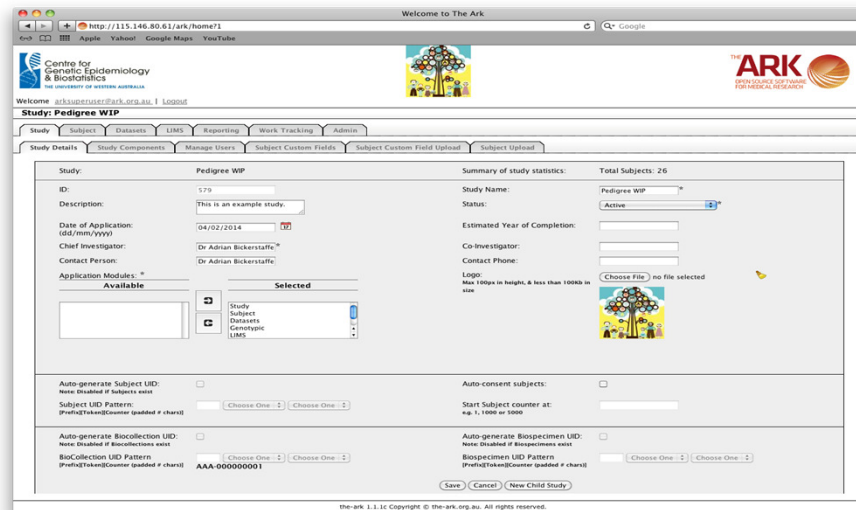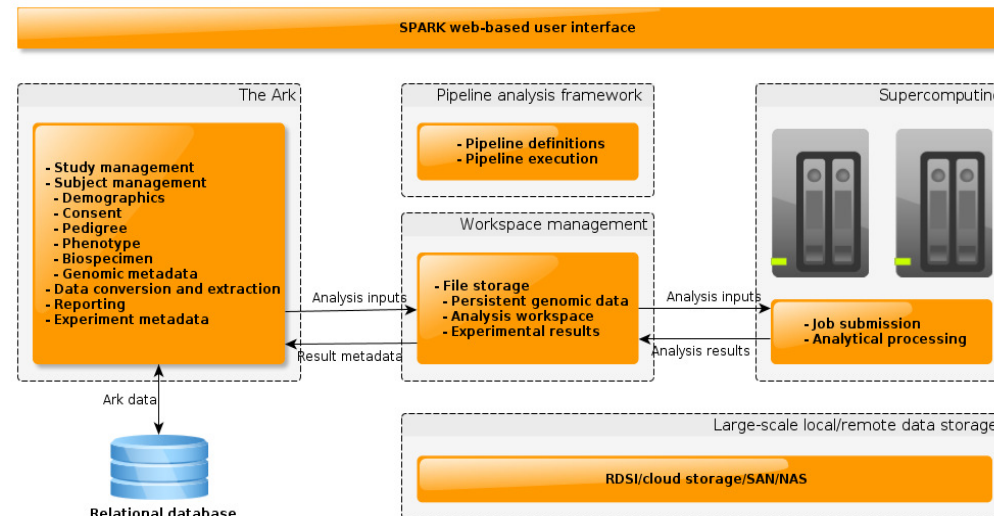
The Ark

The Ark is an open source web-based biomedical data management system with role based high security.

Modules

❑ Study management
❑ Subject (participant) management
❑ Pedigree management
❑ Phenotypic data management
❑ Laboratory data management
❑ Work tracking and invoicing
❑ Reporting
❑ Data extraction

SPARK Architecture

SPARK

The proposed **SPARK** platform will extend The Ark to handle large genomic datasets and provide a user-friendly web interface to massively parallel high-performance computing resources. This will greatly enhance GWAS research by providing crucial data management, processing and modelling facilities.

SPARK Objectives

❑ Bring state-of-the-art complex genomic data management to the international GWAS research community.
❑ Provide the health and medical research community with user-friendly access to massively parallel computing resources.
❑ Foster collaboration among the global GWAS community.

References

P. M. Visscher, M. A. Brown, M. I. McCarthy, and J. Yang. Five years of GWAS discovery. American Journal of Human Genetics, 90(1):7–24, 2012.

S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. De Bakker, M. J. Daly, and P. C. Sham. PLINK: A tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics, 81(3):559–575, 2007.

Collaborators

Centre for Genetic Origins of Health and Disease
The University of Western Australia

"To create a leading open source knowledge discovery platform that leverages massively parallel computational power for heterogeneous genomic datasets"