

Application
Transformations

Users define analyses
via transformations

Variant calling & analysis,
RNA-seq analysis, etc.

Presentation
Enriched Models

Enriched models provide convenient
methods on common models

Enriched Read/Variant

Evidence Access
MapReduce/DBMS

The evidence access layer
efficiently executes transformations

Spark, Spark-SQL,
Hadoop

Schema
Data Models

Schemas define the logical
structure of basic genomic objects

Avro Schema for reads,
variants, and genotypes

Materialized Data
Columnar Storage

Common interfaces map logical
schema to bytes on disk

Load data from Parquet and
legacy formats

Data Distribution
Parallel FS

Parallel file system layer
coordinates distribution of data

HDFS, Tachyon, HPC file
systems, S3

Physical Storage
Attached Storage

Decoupling storage enables
performance/cost tradeoff

Disk, SDD, block
store, memory cache