On the Centrality of HPC for taking NGS to the next frontier: Clinical application at scale

Monica Martinez-Canales, PhD
Principal Engineer
Big Data Solutions Group
monica.martinez-canales@intel.com

Stephen R. Wheat, PhD
Director, HPC Pursuits
HPC and Big Data Business Group
stephen.r.wheat@hpe.com

www.intel.com/healthcare/bigdata

HPC Alliance Accelerates Customer Value

HP and Intel partner to deliver innovation across the entire HPC Solution Framework

Customer Value

HPC Center of Excellence – Customer Experience Innovation

- Code Modernization
- Performance Optimization
- Engineering

- Benchmarking / POC
- Service Engagement
- Resource Utilization

- Business Outcomes
- Rapid Deployment
- Expert Community

HPC Solutions Framework – HP Innovation

Industry Expertise

Geosciences (Oil & Gas) 

Life Sciences 

Financial Services 

Ecosystem Optimization

ANSYS | BIOVIA | Gaussian | Halliburton | Abaqus | Paradigm | Schlumberger | SIMULIA | Synopsys

Solutions Design

- Purpose built and optimized for HPC workloads
- End-to-end Integration to simplify deployment and management
- Uniquely designed for Density, Energy Efficiency, and Performance

HPC System Framework – Intel Innovation

Intel® Xeon Phi™ coprocessors 

Intel® Omni-Path Architecture

Intel® Lustre* Software

Intel® Xeon® processors 

Value

- Scalability & Resiliency
- Power Efficiency
- Price/performance

Customer Value


Optimal Performance


Exceptional Efficiency


Rapid Deployment

HP Apollo Systems Family

- Purpose-Build HPC Solutions

Apollo 8000

Supercomputing

Purpose Built liquid-cooled supercomputing
Delivering the highest level of performance and density for the most demanding workloads

Liquid Cooled

Apollo 6000

Rack Scale HPC

Purpose built rack scale HPC solutions
Delivering shared infrastructure efficiency optimized for specialized workloads

Rack Scale

SL4500/Apollo 4000

Server Solutions Purpose Built for Big Data

Purpose Built Server Storage Solutions
Delivering Industry leading density, efficiency, and price performance at Hyperscale for Big Data and Object Storage Solutions

Server Based Storage

Apollo 2000

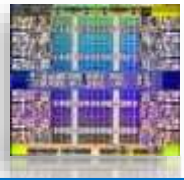
Enterprise Bridge to Scale-Out Compute

Purpose Built Density Optimized solutions

Delivering Hyperscale efficiency and performance for traditional enterprise and SME datacenters

Traditional Datacenters

What is the optimal platform for code modernization?



	The world is going parallel – stick with sequential code and you will fall behind.	*Intel® Xeon® Processor E5-2600 v3 Product Family formerly codenamed* **Haswell**	*Intel® Xeon Phi™ x100 Product Family formerly codenamed* **Knights Corner**	***Future Xeon Phi*** **Knights Landing**	...	***Future Xeon***
Cores		18	61	>60		TBA
Threads/Core		2	4	4		TBA
Vector Width		256-bit	512-bit	512-bit (x2)		512-bit
Peak Memory Bandwidth		68 GB/s	352 GB/s	>400 GB/s		TBA

Haswell coupled with Knights Corner provides the optimal scale to get performance now and get ready for Knights Landing and future Xeon

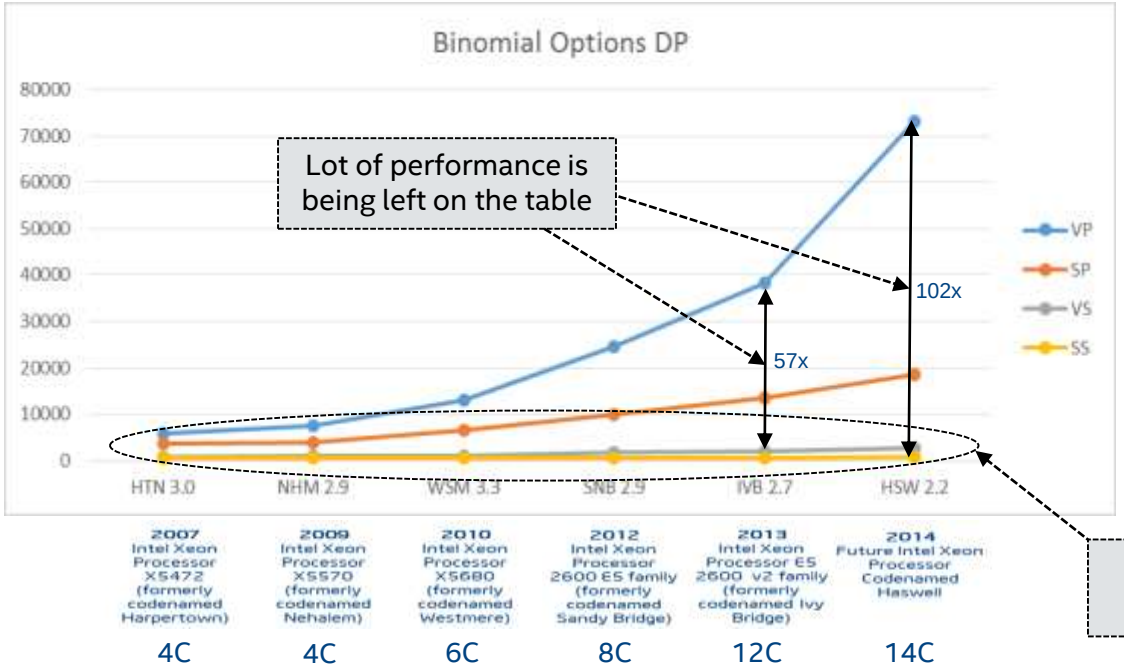
† Xeon = Intel® Xeon® processor
† Xeon Phi = Intel® Xeon Phi™ coprocessor

Intel is Making an Enormous Pipeline Investment

Scaling R&D and touching next-gen via Intel® Parallel Computing Centers (IPCC)

*Other names and brands may be claimed as the property of others.

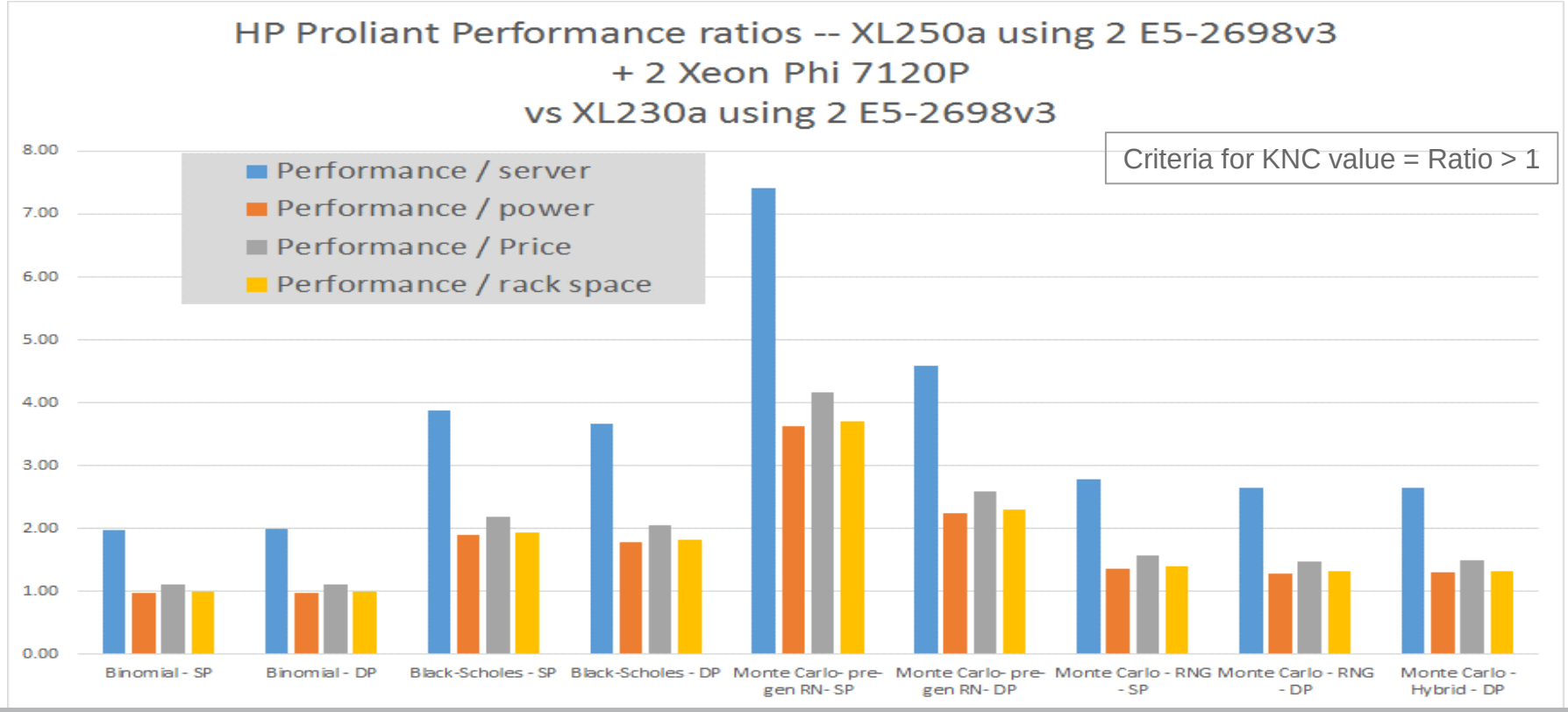
How can I get higher performance for my apps?



VP = Vectorized & Parallelized
SP = Scalar & Parallelized
VS = Vectorized & Single-Threaded
SS = Scalar & Single-Threaded

Modernization (i.e. parallelization and vectorization) of your code is the solution

Key TCO metrics using HP Apollo 6000 with 16C Xeon† & Xeon Phi†



Code modernized using Xeon Phi delivers compelling TCO metrics for FSI on Apollo 6000 servers

† Xeon = Intel® Xeon® processor
† Xeon Phi = Intel® Xeon Phi™ coprocessor

The Oklahoma Health Prediction Center

Applying NGS-enabled solutions to a community-wide effort to combat cancer.

Tulsa's Collaborative Health Network Platform has already demonstrated:

- Better Primary Care
- Better Coordination of Care
- Enhanced clinical services to communities

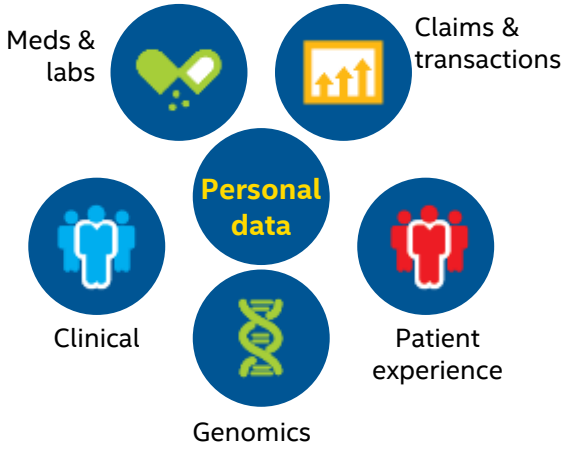
with support, trust, and advocacy driven by:

- Information Security Expertise
- Bio-Informatics Expertise
- Health Law Expertise
- Molecular Biology Capacity
- And, a community-first focus

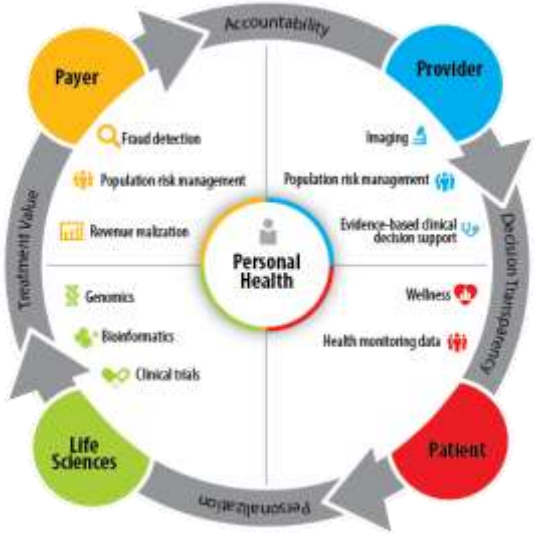
Creating a "National Model" for Community-Supported City-wide Health Collaboration and Partnerships.

Intel's Vision for Precision Medicine

Today: Many disparate data types, streams…

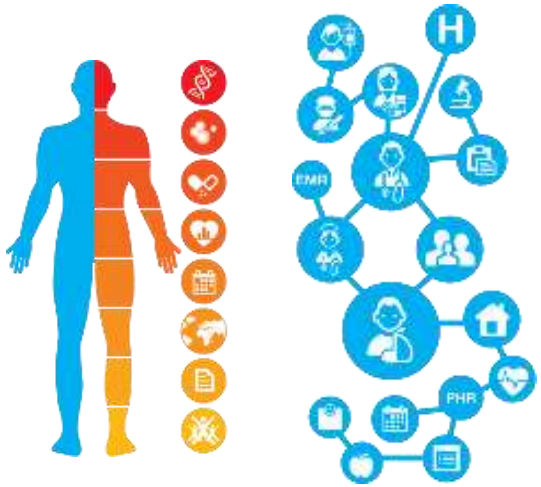


Future: connected clinically-actionable information



- Value-based Care
- Higher Quality at Reduced costs

Leading to better decisions



- Patient-centric approach
- Healthier population outcomes

Computing Innovation for Billions of People Worldwide

Health IT

- Performance‡
- Security
- Manageability

Powering the Health Workforce

- Mobility
- Datacenter
- Cloud
- Health information exchange
- Sensors and Wearables

Healthcare Solutions

- Intel®-powered medical imaging & devices
- Security software (McAfee, Wind River)
- Analytics (Intel® Math Kernel Library, Intel® Analytics Toolkit, Lustre, Cloudera)

Life Sciences

- Extreme computing for big data
- Open, interoperable clouds
- Appliances
- Code optimization

Ecosystem Support

- People-centered R&D
- Policy
- Standards

Intel in Health & Life Sciences

Big Data and High Performance Computing


Sequencer


Analytics Appliance

HPC Pipeline Cluster


Storage and Networking


Research Computing


Precision Medicine

Intel® Xeon® / Intel® Xeon Phi™ • Rack Scale Architecture • Integrated Fabric • Software Defined Network • Solid-State Drives • Workload Optimization

Hadoop* • Lustre* • Cloudbursting • Virtualization • Encryption • Cloud-based Analytics

Intel® Architecture is present from sequencers to appliances to high-performance computing cloud.

At the Intersection of Transformative Forces

10^{18} Enabling extreme-scale computing on massive data sets

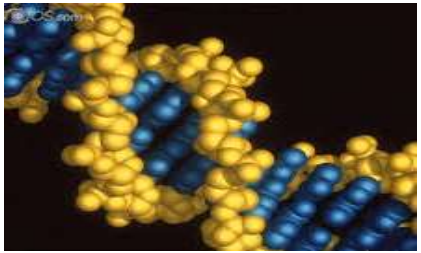

Helping enterprises build open, interoperable clouds


Contributing code and fostering the ecosystem

Trends & Challenges in Life Sciences

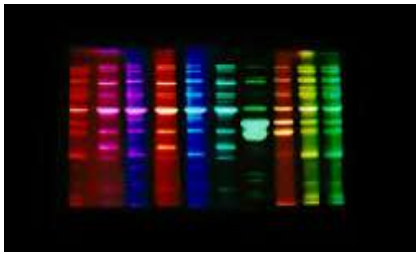
Big Data in Life Sciences

- Sequencer advances – 4x data in 50% less time .5TB/device/day
- 4D molecular imaging produces 2TB/device/day



Burdens of Data Management

- Store, manage, share, ingest and move PBs of research & clinical data
- Need to reliably 'snapshot' pipelines with archive to tiered storage

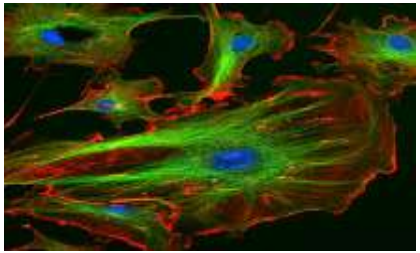


Innovation Drives Change

- Rapid iteration of algorithms far outpace IT, requiring flexibility, agility
- Most applications do not fully leverage available infrastructure

Converged Infrastructure

- Workloads converging between local and cloud-based HPC/Big Data
- Advanced orchestration required to maximize throughput & efficiency



Optimizing Top Applications and Pipelines

Intel working with industry experts worldwide

- Genomics, Molecular Dynamics and Molecular Imaging applications targeting both Intel® Xeon® processors and Xeon® Phi™ coprocessors
- Fine- and coarse-grained optimization at the node and cluster level
- Work with code authors to release optimizations, disseminate best practices

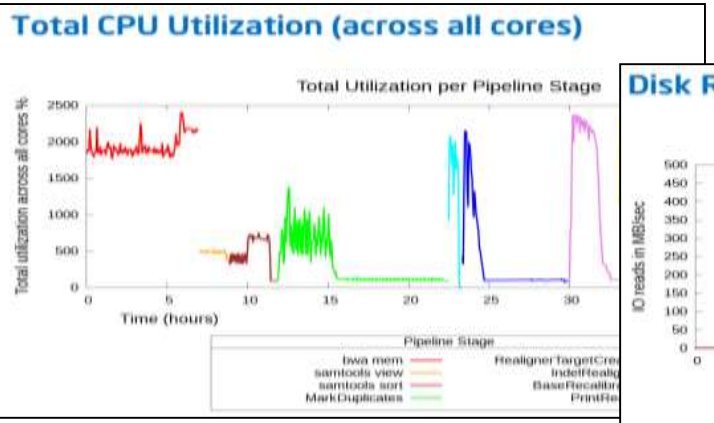
GENOMICS
ABySS*
BLAST*
Bowtie* TopHat* Cufflinks*
BWA*
GATK* Picard* SAMtools*
MPI-HMMER*
Velvet*

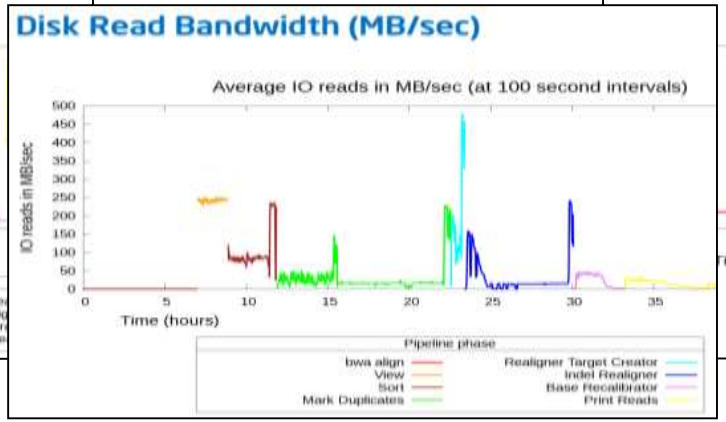
MOLECULAR DYNAMICS
AMBER*
CAS-Soft Sphere* CAS-IPE*
CP2K*
CPMD*
DLPOLY*
GAMESS*
Gaussian*
GROMACS*
LAMMPS*
NAMD*
NWChem*
Quantum Espresso*
VASP*

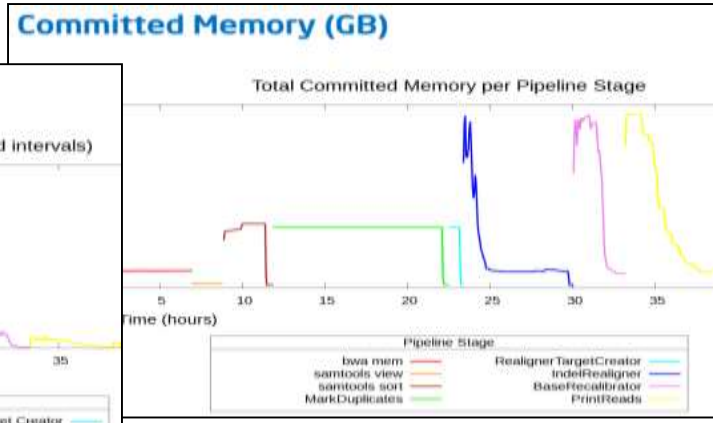
www.intel.com/healthcare/optimizecode

*Some names and brands may be claimed as the property of others.

Intel Workflow Profiler
coarse-grained profiling of long-running workflows







Open Source Distribution:
https://01.org/workflow-profiler

- Automates data collection & charts from standard Linux OS tools
- Quickly identify CPU, Memory & I/O Constraints
- Pareto analysis of hotspots in user-defined steps of workflow
- Target areas which will benefit from newer algorithms and technologies

The future of precision medicine

1 patient visit

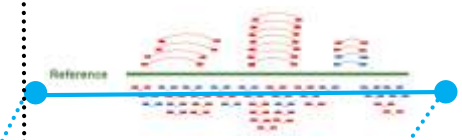
2 causal genes identified & molecular pathways determined

3 targeted therapeutics & companion diagnostics; treatment begins in earnest

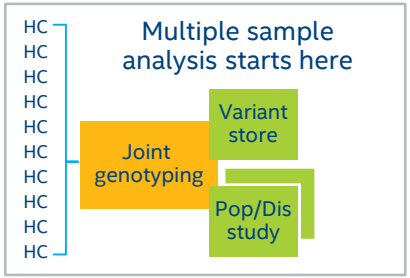
PRIMARY ANALYSIS

Individual

SECONDARY ANALYSIS, DNA/RNA PIPELINE + MORE



Multiple individuals



PRECISION MEDICINE

1 to 4 days | ~~weeks~~ **hours** | ~~months~~ **days**

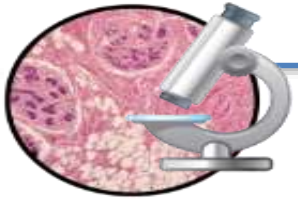
Precision medicine is the standard of care, integral to wellness by 2020

Applying Precision Treatments for Pediatric Cancer

Neuroblastoma:
From 15 days to less than 4 hours for personalized treatment

Patient / Physician diagnosis, treatment, ongoing management

Treatments with a more reasonable chance of a cure

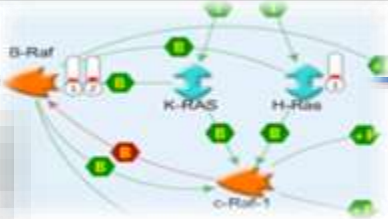


Tumor Sample

Minimizing trial and errors

Complete molecular characterization of the diseased tumor

Understanding the individual disease



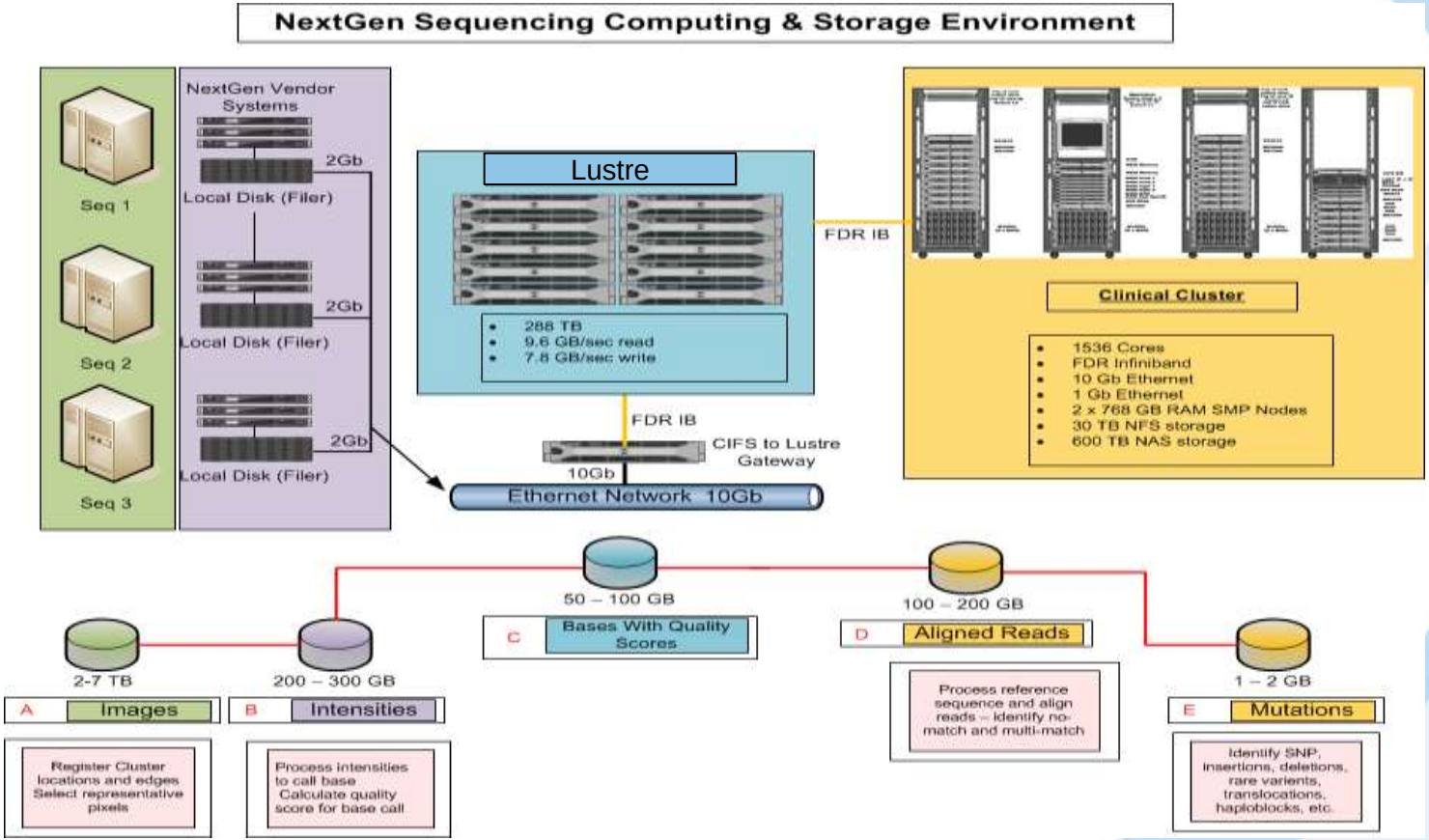
Analytical tool for mapping patient data against database for recommended treatment

Accelerating targeted treatment options

Integration of scientific & clinical evidence for future research

Creating platform to scale to 100k+ patients

Genomics Data Processing Pipeline



*Some names and brands may be claimed as the property of others.

An Idea for Clinical Decision Support Applications Combining Clinical, Genetic/Genome, and Family Health History Data

Goal:

Promote widespread use of clinical decision support that will help clinicians/counselors in assessing risk and assist genetic counselors in ordering genetic tests.

Build a scalable CDS that leverages standardized data that includes:

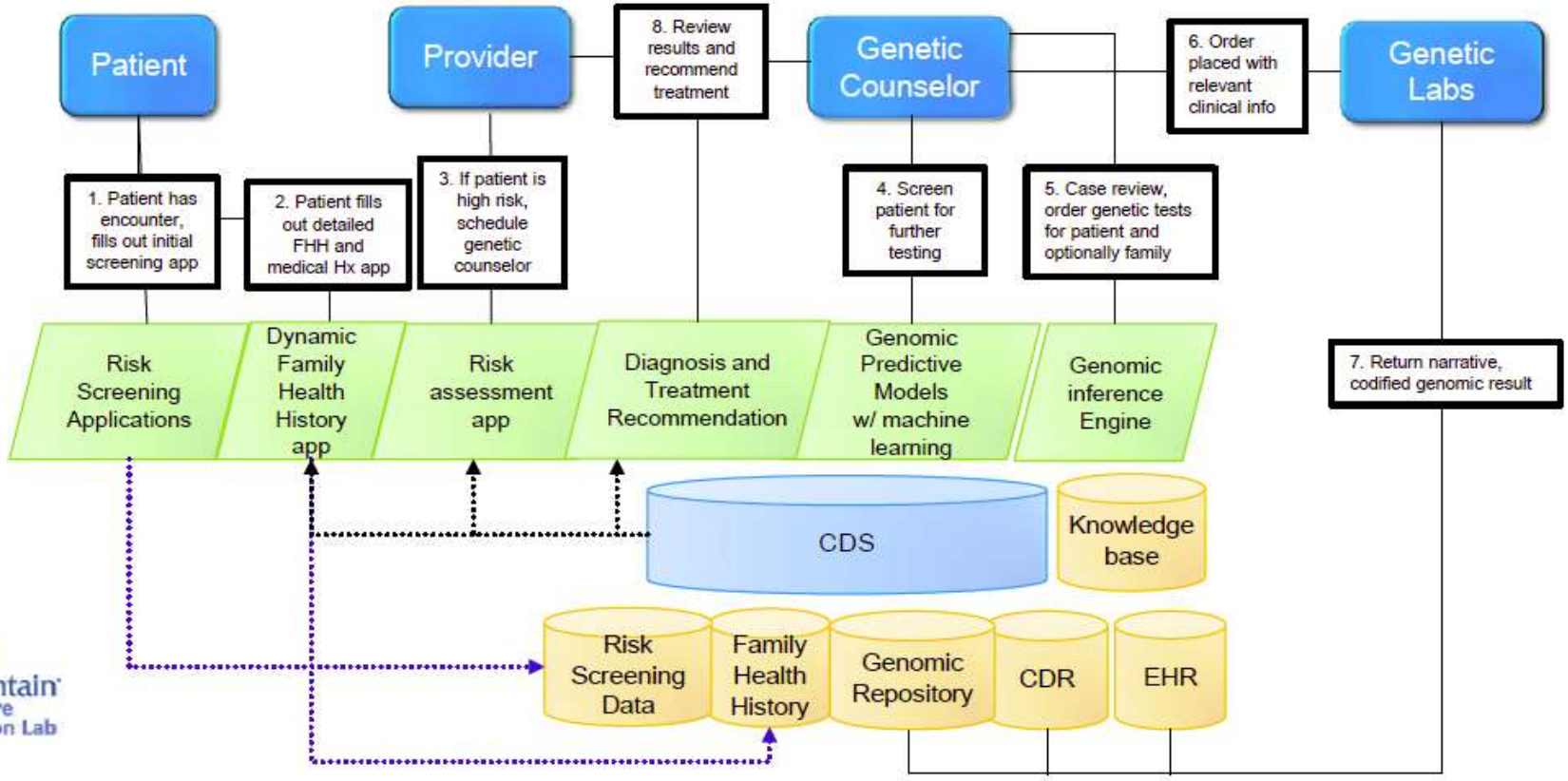
- Family Health History
- Clinical data and Screening
- Genomic Data

Solution Considerations:

- Be agnostic to data collection tools.
- Be scalable to different clinical domains (grow beyond Breast Cancer) and other healthcare institutions.
- Be standards based where they exist
- Work across all EHRs, but start with Cerner
- Leverage Intel technologies (infrastructure, Intel Data Platform, etc.)
- Be flexible to incorporate other data sources (e.g. imaging data, personal device data)

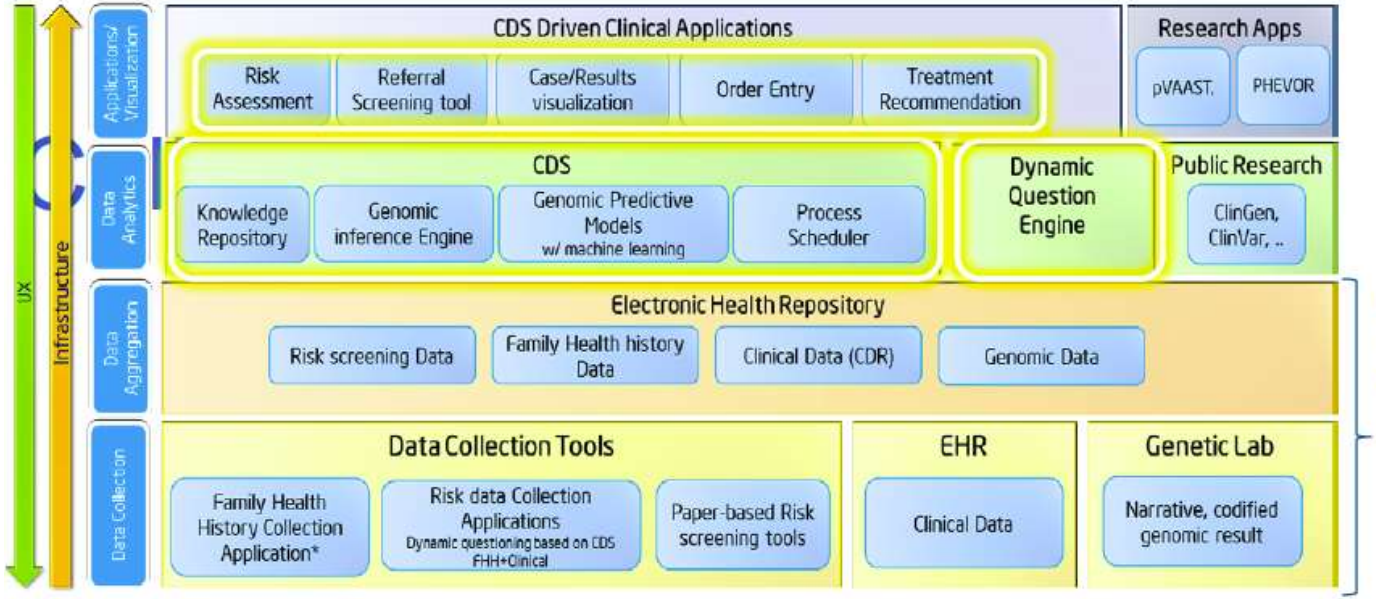
K. Paranjape, General Manager, Health & Life Sciences, Intel, “Personalized Medicine and IT”, IOM, 2014

Sample Clinical Workflow with Clinical Decision Support



Intermountain Healthcare Transformation Lab

K. Paranjape, General Manager, Health & Life Sciences, Intel, "Personalized Medicine and IT", IOM, 2014



Baseline infrastructure

- Risk Screening Applications
- Structured and coded Family health History data
- Data Mart that combines FHH, Clinical, and Genomic data

Intermountain Healthcare Transformation Lab

Solution Considerations

Utilize the following where appropriate:

1. Health Services Platform (HSPC), HealtheDecision, Open CDS
2. Intel Data Platform for Machine Learning, Graph Analytics, Mining
3. HL7 standards, FHIR + SMART Apps for clinician facing applications

K. Paranjape, General Manager, Health & Life Sciences, Intel, “Personalized Medicine and IT”, IOM, 2014

Sample Analytics

Areas	Benefits, solutions
Utilization and Treatment Analysis	Combine trends with individual treatment analysis
Treatment Effectiveness	Build large-scale treatment effectiveness monitoring
Diagnosis and Treatment correlation	Discover diagnosis/treatment connections
Managed Care Optimization	Optimize resources for managed care
Diagnosis Treatment and Trends and Predictions	Determine overall trends, but put them at the disposal of individual diagnosticians
Drug Utilization and Expense Prediction	Build dynamic precise drug utilization prediction models
Treatment and Outcomes Analysis and Optimization	Predict treatment prognosis, optimize based on individual's complete picture
Demand Forecasting	Better demand preparedness
Price Analysis and Determination	Optimize quality and revenue through price monitoring
Epidemiology Research	Discover trends by analyzing data from disperse sources
Provider Ratings and Benchmarking	Use all source of data to benchmark and monitor providers
Patient History and Digital Records Archiving and Analysis	Combine patient history records from disparate sources, greatly improve the quality of patient care
Contract Optimization	Optimize contract resource utilization

K. Paranjape, General Manager, Health & Life Sciences, Intel, “Personalized Medicine and IT”, IOM, 2014

Imagine what is possible

We are working with industry experts like you –

- to overcome systemic challenges,
- maximize the use of available infrastructure, and
- drive innovation through open standards and platforms.

These technologies facilitate more accurate science, enabling scientists & clinicians to imagine what is possible, instead of framing their research in terms of constraints.

Accelerate Science. Translate Results. Deliver Today.

www.intel.com/healthcare/bigdata
www.intel.com/healthcare/optimizecode

Intel Across Healthcare: Learn More

Mobile Healthcare

Care Coordination

Secure Cloud

Devices and Imaging

Performance

Big Data

Policy and Standards

Aging

www.intel.com/healthcare
www.intel.com/healthcare/optimizecode

Health & Life Sciences at Intel
Where information and care meet


intel®
Look Inside.™

Backup

Life Sciences Key Industry Challenges and Solutions

- Many (most) applications are single-threaded, single address space

  Intel is delivering optimizations working with open source community, developing NGS+HPC curriculum

- Some algorithms scale quadratically with the size of the problem. Large data sets exceed available memory and storage

  Innovations in acceleration, compute, storage, networking, security, and *-as-a-service.

- International collaboration is an imperative, bioinformatics expertise is scarce
- ***Intel is working closely with the ecosystem to address enterprise to cloud transmission of terabyte payloads***
- Databases are distributed, data is siloed and will likely stay that way

  Tools like Hadoop, Lustre, Graphlab, In-Memory Analytics, etc.

Health & Life Sciences at Intel

Where information and care meet