Emerging HPC Use Cases and Deployment Models

Spiros Liolis
HPC & AI

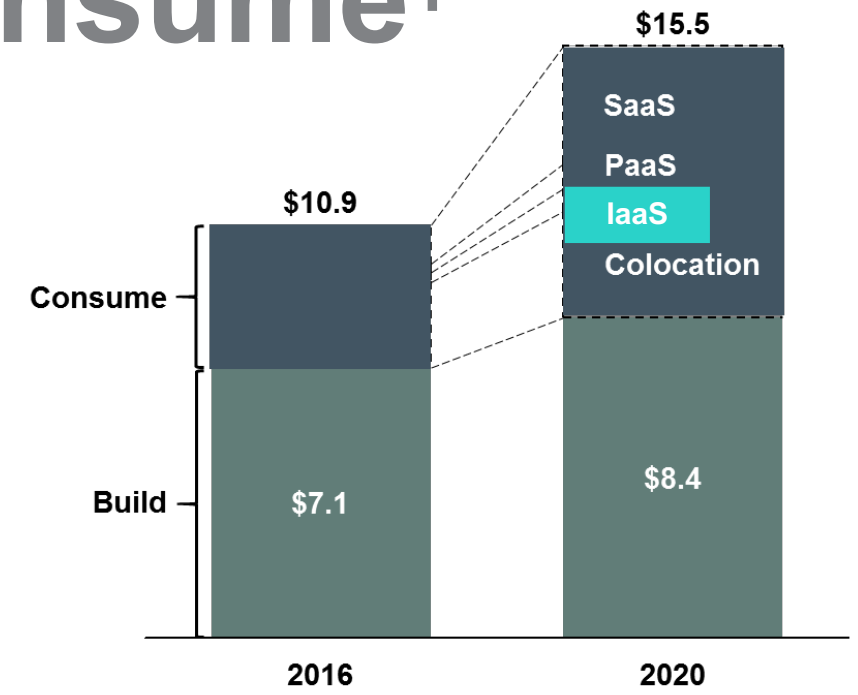
September 27th, 2017

Market Opportunity

Hewlett Packard
Enterprise

“By 2020, 45% of the HPC market will have shifted from built to consume[1] models”

[1] IDC studies



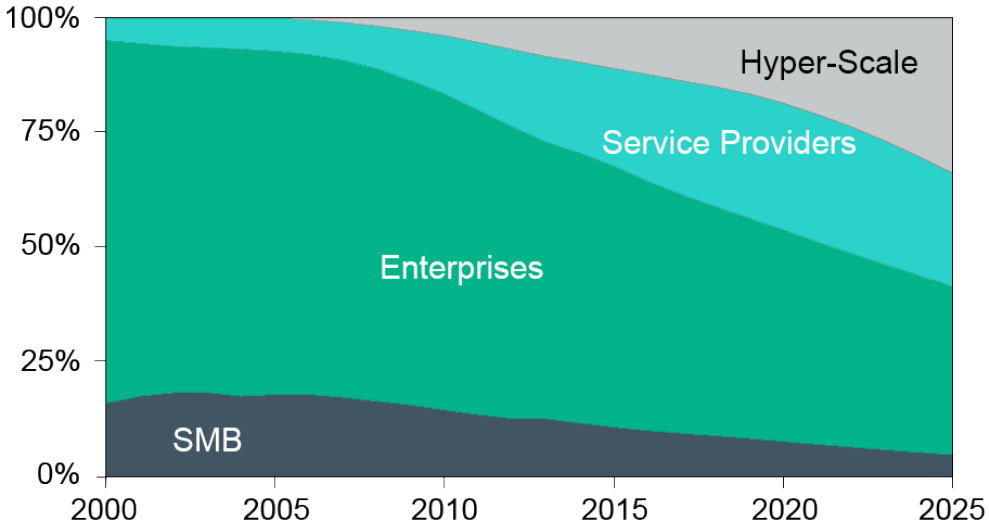
HPC Market Size by Deployment ($B)

Hewlett Packard Enterprise

Customer buying patterns are shifting

Key Trends

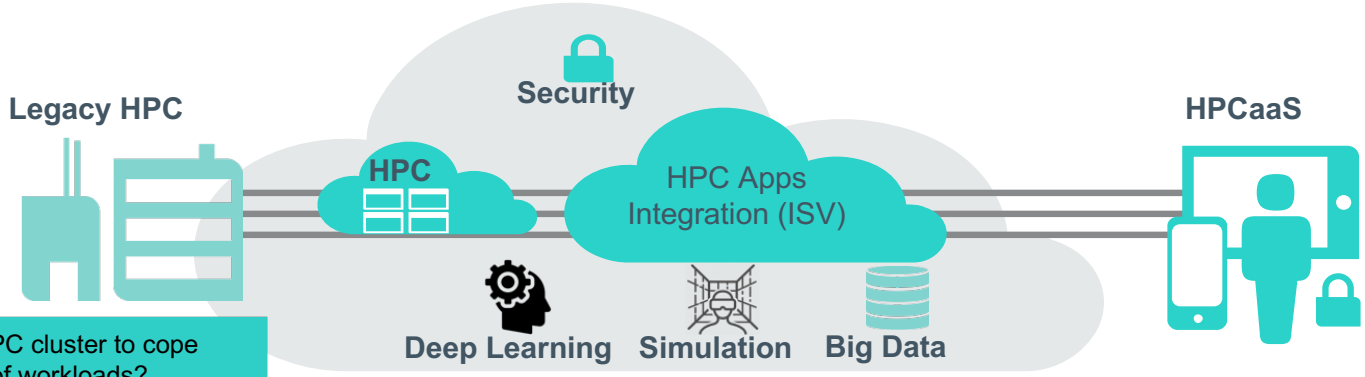
- Organizations are adopting **hybrid IT** with multiple service delivery models to meet their unique business and technical needs
- "Cloud" IT spending is increasingly shifting towards consumption-based **services from 3rd party providers** vs building their own
- **Speed of deployment, business agility, financial flexibility** and **cost effectiveness** are key drivers for consumption-based model



Gartner Research:
Service Provider and Hyperscale Data Centers will contain 60% of world's compute power by 2025

HPCaaS solves the challenges of a modernized HPC infrastructure

Bridging the legacy world with Cloud



Challenges

- How to repurpose an HPC cluster to cope with a growing number of workloads?
- How to interconnect HPC with the other Information Systems?
- DC space? What deployment models? on premise, off premise, public cloud?
- How to develop cloud native applications?
- How to onboard an increasing number of stakeholders and non-skilled end users?

Introducing HPC attributes into Cloud

- Service catalogue
- Jobs driven
- Performance built

Results

- Driven by business objectives
- Easily acquired and integrated with business apps
- Stable, scalable and secure

Hewlett Packard Enterprise

Typical reasons and ways to use cloud for HPC workloads

Main reasons to move HPC to the “Cloud”

- Testing new architectures (AI / Deep learning, GPGPU)
- Collaborating with partners (H2020 IMEC)
- Moving to an OPEX mode
- Burst computation requirements (Running out of on premise resources, New industrial model, end of year, …)
- Faster time to get resources for user
- Shortage of Data Center space (high density)

How to consume Resources for HPC

- Application (SaaS)
- Cluster (IaaS)

HPE in HPCaaS

POCs

- **Complete**: CERN, Jaguar Land Rover, TETRAPAK, Valeo Lighting, NEVS, MAN, TET, Simulia – Living Heart, Total phase 1
- **Ongoing** : Total (Advania) phase 2
- **Planned** : ANSYS Virtual Benchmark, DTU, PSA, IMEC – 8 big pharmas

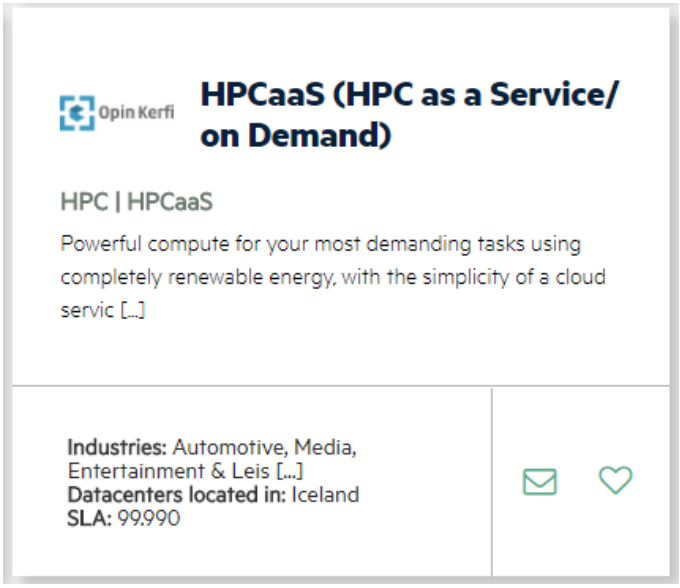
Use Case

- ADVANIA HPC with HPE – ADVANIA provides the necessary processing power to TOTAL, for their seismic simulations and analysis
- OPIN KERFI HPC with HPE – Opin Kerfi provides the necessary processing power to clients in Automotive, Media and Entertainment clients


Hewlett Packard Enterprise

Seismic Processing jobs run through a remote processing centre

HPE run Total's application using existing software built on hybrid compute nodes.


Opin Kerfi

HPCaaS (HPC as a Service/ on Demand)

HPC | HPCaaS

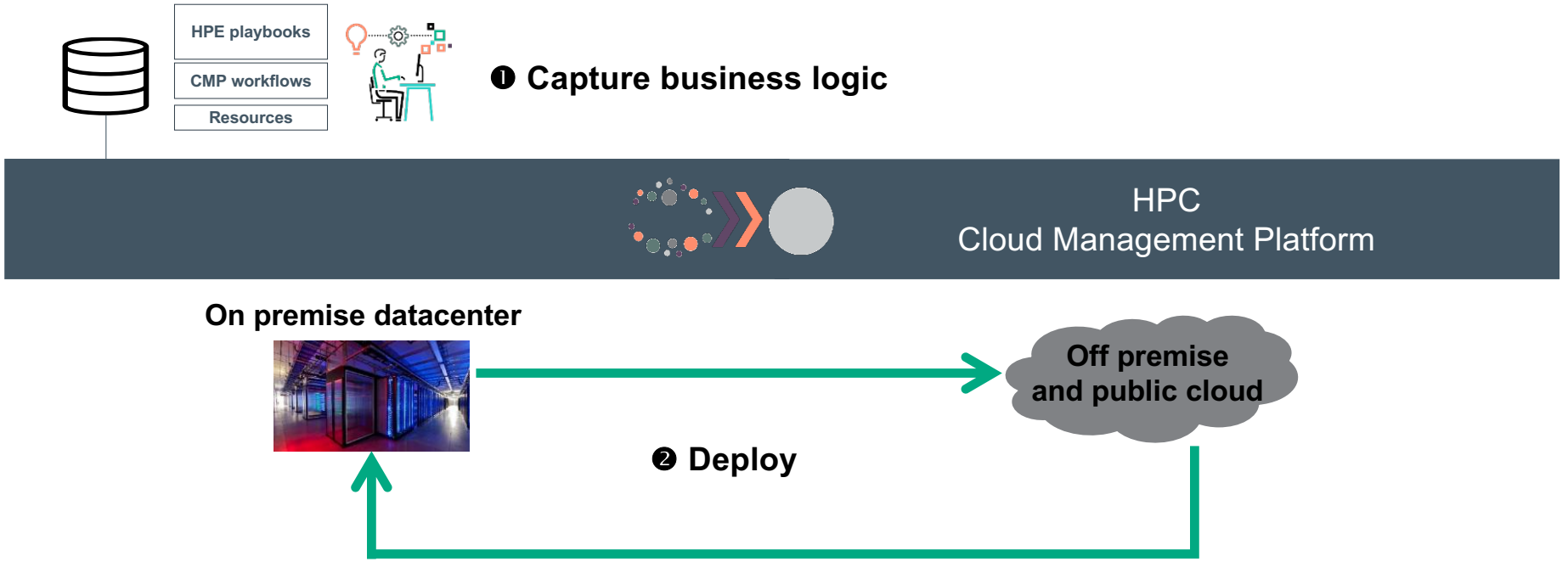
Powerful compute for your most demanding tasks using completely renewable energy, with the simplicity of a cloud servic [...]

Industries: Automotive, Media, Entertainment & Leis [...]
Datacenters located in: Iceland
SLA: 99.990

Hewlett Packard Enterprise

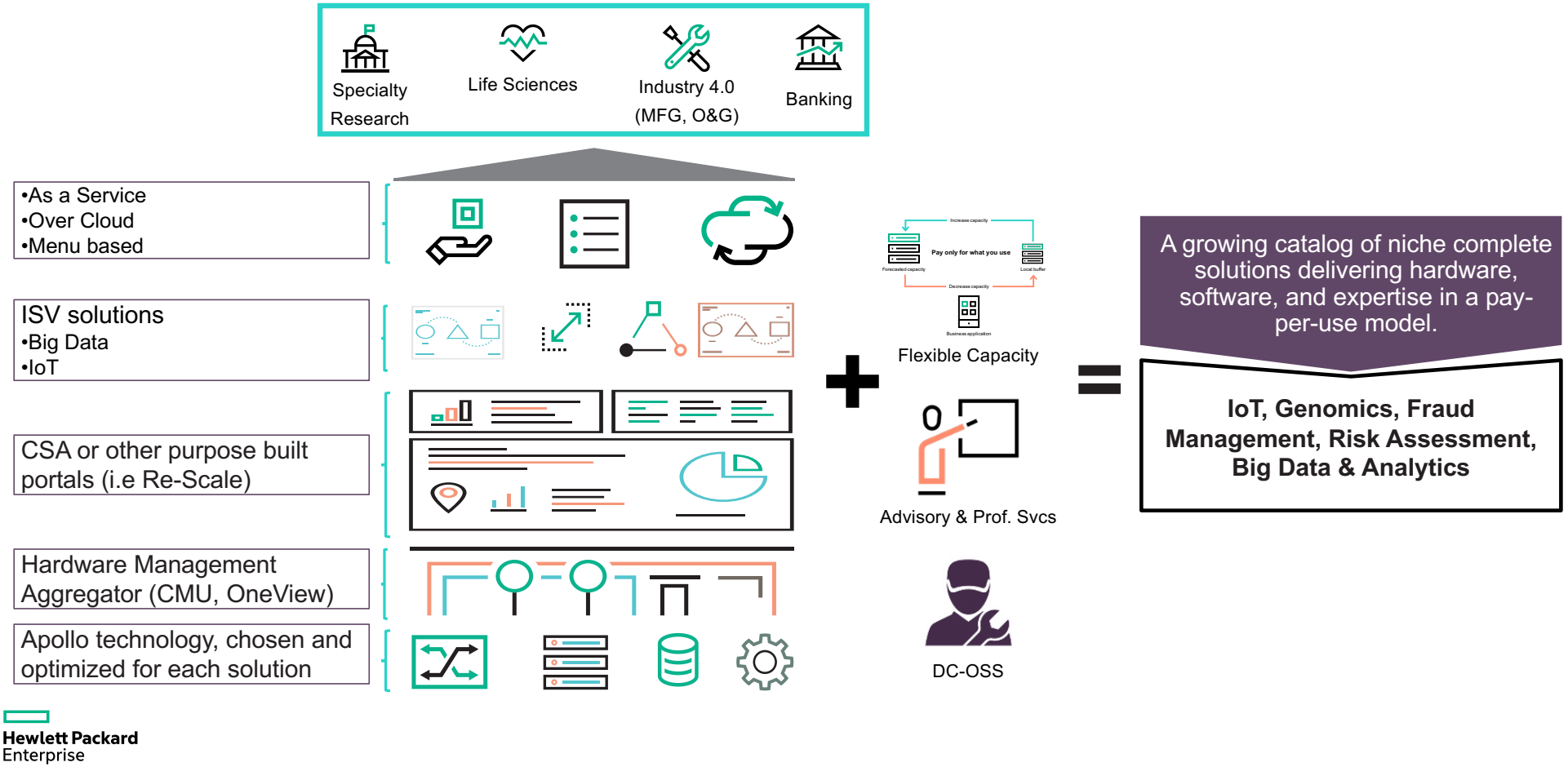
Overview of HPCaaS Capabilities & Value Proposition

HPCaaS in 3 steps



Hewlett Packard
Enterprise

HPC in the Cloud – Reference Architecture



Hewlett Packard Enterprise

HPE enables HPC from on premise to off premise & public cloud

Relevant Simulation, Big Data Analytics and AI capabilities

Current App environment

Cloud Management Platform

Infrastructure

Service Providers and Advisory Service

Expertise	HPE POINTNEXT			
Software		MICRO FOCUS, SUSE, CLOUD CRUISER	**DMF – Tiered Data Management**	
Recent HPE innovations		**HPE HPC IaaS**	**Apollo, MDC, Machine . . .**	**ClearPass, UIoT platform, EdgeLine**
Expanded Partnerships	Uber Cloud, ANSYS, SIMULIA, GE Digital, ptc	rescale	panasas, intel, BeeGFS developed by Fraunhofer	PCPC Direct, DXC.technology, EQUINIX, SCALEMATRIX, Cloud28+, X-ISS, advania, R SYSTEMS, Opin Kerfi, CPU 24|7

HPCaaS Centers of Excellence

Extensive technology and industry expertise to deliver comprehensive solutions

Hewlett Packard Enterprise + intel

Advice, planning, design, benchmarking

- **HPCaaS capabilities in worldwide COE facilities**
- **Locations include Bangalore, Grenoble, Houston and Reykjavik*****
- **Target workloads are in HPC, Virtualization and Deep learning**
- **Support new Apollo and Intel products, customer deals, POC and benchmarks**
- **Targeted engagement through Sales Rep**

*** *Sites : HPE Bangalore, HPE Grenoble, HPE Houston, UTA Arlington, Mesosphere and ADVANIA COE*

HPE Pointnext Services

Helping accelerate your transformation

Bringing technical expertise and a comprehensive portfolio of services to advise, transform, and manage each unique journey, with an end-to-end lifecycle approach, to help address IT and business needs

Advise

- Focusing on business outcomes and goals
- Design your transformation and build a road map tuned to your unique requirements

Transform

- Help build new IT or evolve existing IT, applications, and business processes to customer's future desired state
- Collaborate with your IT team from technical design to implementation, build to migration, and into production

Manage

- New ways of delivering IT
- Managing and optimizing workloads, resources, and capacity, on-prem and in the cloud,
- Simplify the experience and offer choice in where to land workloads and what to self-manage or out-task

A services partner built for your business today and tomorrow

HPCaaS – HPE's Value Proposition

Elements of differentiation

Apollo technology, chosen and optimized for each solution

Software specific to the solution

Flexible Capacity model for pay per use based on unique business metric; manage the evolution of the solution **(Optional)**

Designed and integrated with Pointnext Advisory and Professional services

Infrastructure stack operated and supported with included DC-OSS and support services **(Optional)**

A catalog of niche complete solutions delivering hardware, software, and expertise in a pay-per-use model.

Growing Catalog of solutions for SP, FSI, MFG, Healthcare

IoT, Genomics, Fraud Management, Risk Assessment, Big Data & Analytics

Hewlett Packard Enterprise

HPCaaS – Use and Study Cases

Our journey towards a robust HPE Industry Solutions Portfolio

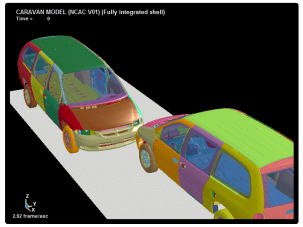
Why Industry Solutions?

- Expand market reach
  - Go beyond infrastructure to **higher layers of the solution stack**
- Deliver higher value
  - Focusing on key workloads and solving business challenges
- Create clear differentiation in market
  - **Unique and differentiated** solution features
- Stronger partner ecosystem
  - Team up with market leading **ISV partners**

What is a solution?

- Focus on **key growth vectors** in the HPC enterprise space
- **Reference Architecture** Model, allowing customer customization
- **Solution Packaging:** H/W + S/W + Services
- Full Collateral Suite:
  - WP, Customer Presentation, Sales Guide, etc.
- Industry leading application benchmarking SME's to provide optimization and benchmarking guidance

Priority Industries

FSI	LSI	MFG
		
TAM: $0.9B, +9.5% CAGR	TAM: $1.7B, +20.6% CAGR	TAM: $1B, +4% CAGR

Hewlett Packard Enterprise

Any Industry: Massive density for Hadoop and Big Data Analytics

Hortonworks cloudera hadoop

Analytics

Hadoop optimized
3 servers in 4U chassis ideal for Hadoop-based analytics with 3-copy data replication

At scale

Efficient analytics scaling
Up to 30 servers per 42U rack with 15 HDDs/SSDs per server

Versatile performance

For Big Data variety
Customize for Hadoop workload variety and NoSQL analytics with disk, CPU, I/O and interconnect options

Unleash the full value of Big Data with Hadoop

Any Industry: Phishing and Malware Detection

Applying data mining principles to large evidentiary data sources

Requirements

- unique intelligence when combined with powerful analytics
- identifying cybercriminals, prioritizing cybercrime investigations and protecting consumers, corporations, and governments
- data retrieved and processed in timely manner regardless the size of the service dataset
- scalable service regardless the size of data and the number of users and present near real-time data analytics

Use Case / Architecture

- product with friendly User Experience, real-time data analysis using cutting edge technologies for big data, full text indexing and aggregation analytics generation.
- A Single Page Application (SPA) with a responsive UI, including multiple criteria search forms, configurable results views, dynamic dashboard analytics and graphs
- A scalable streamlined data processing and indexing layer based on Spring Integration, Spring Batch and Elastic (previously known as Elasticsearch)

Why Apollo?

- Lower upfront investment
  - Low $/GB
  - Purchase in smaller, more affordable increments
- Space efficient 4U form factor
- Range of high performance fabrics

Any Industry: Sentiment Analysis

Uses social media data to understand customer opinions and incorporate feedback into design

Requirements

- House large amounts of unstructured data in minimal amount of space – 4PB+
- High speed inter-connect to accelerate throughput
- Balance of price/performance

Use Case / Architecture

- Gather millions of social media blogs, videos, tweets, posts, etc. to store on HDFS repository
- Perform analytics on SoMe data to better understand customer sentiments about car design
- 12 racks of Apollo 4530

Why Apollo?

- Lower upfront investment
  - Low $/GB
  - Purchase in smaller, more affordable increments
- Space efficient 4U form factor
- Range of high performance fabrics

Hewlett Packard
Enterprise

Life Sciences

HPE Next Generation Sequencing Solution

Scalable, Reliable, Flexible Genome Analytics Solution for Small to Large Enterprises

Life Sciences

Reliably predictable genomics cluster solution

Turn-Key Solution — Rapid, simple integration into end users genome analytics environment

Wide Market Appeal — From traditional R&D to emerging clinical-based patient treatment

HPE Apollo 2000

NGS performance & workload capacity

HPE Apollo 4520

Scalable HPC storage for NGS workloads

HPE ProLiant DL560

Ideal for genome assembly workloads

BIGstack Broad-Intel Genomics Stack
- Turnkey GATK Best Practices Workflows
- Reducing setup time

intel BROAD INSTITUTE — NEW

Data Management Framework (DMF)
- Data Management and Data Migration
- Job specific data staging

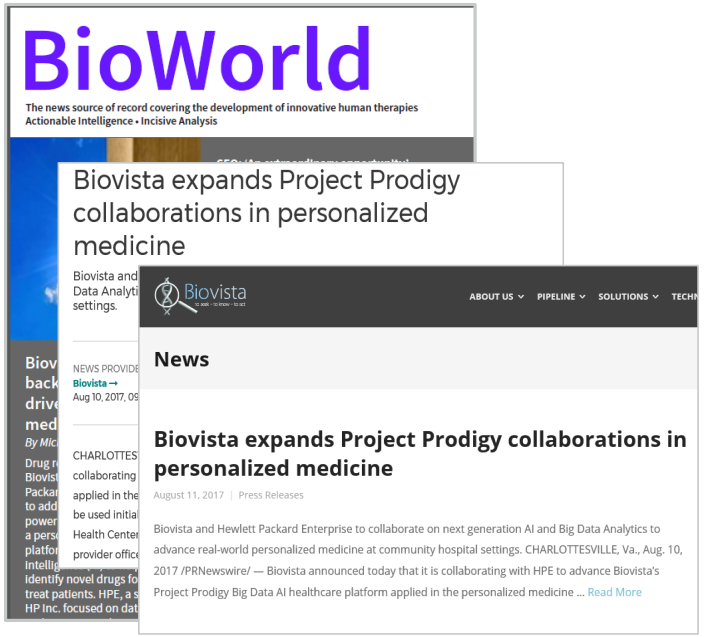
NEW

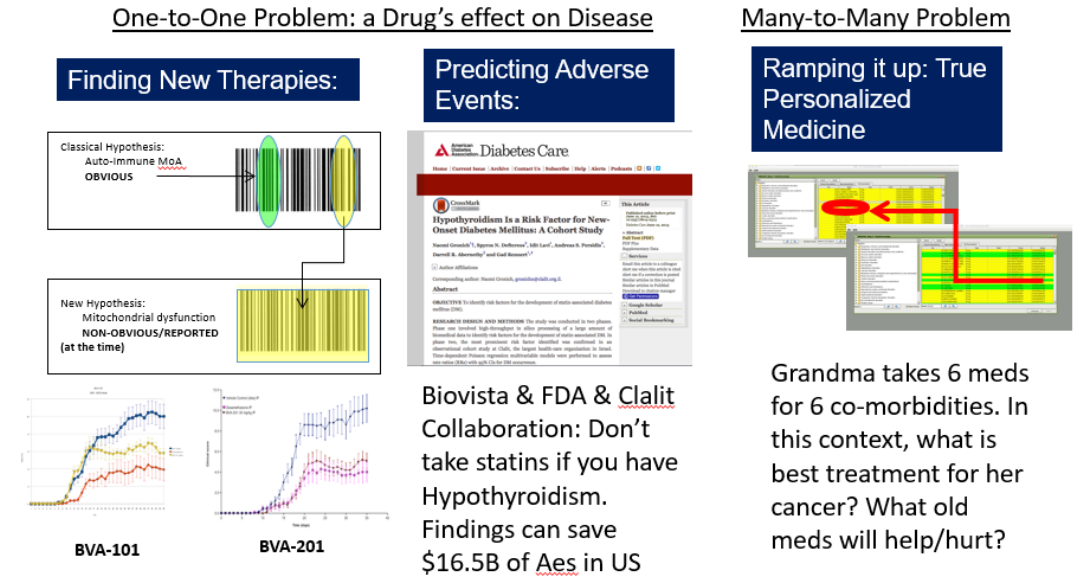
HPE POINTNEXT
- Customized design
- Pipeline tuning and config optimization
- HSM strategy implementation

NEW

Hewlett Packard Enterprise

HPE & Biovista collaboration: Bringing AI and HPC to Medicine





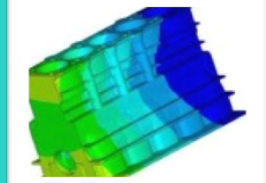
HPC & AI Helps solve One-to-one Problems. HPC & AI is Essential for Many-to-Many Problems.

Is it possible to combine patient and general biomedical data to transform medicine?

Hewlett Packard
Enterprise

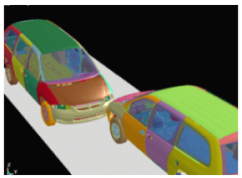
Manufacturing

CAE Disciplines in Virtual Product Design



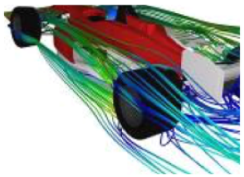
Computational Structural Mechanics (CSM) For Implicit FEA

Simulate the strength and vibration characteristics of product



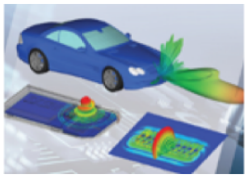
Computational Structural Mechanics (CSM) For Explicit FEA

Simulate the shock impact of products over short duration



Computational Fluid Dynamics (CFD)

Simulate Aerodynamics; cooling; mixing of fluids such as air, water and chemicals



Computational Electromagnetics (CEM)

Simulate radar signature/scattering to assess/prevent detection and identification; Antenna performance, ASIC package simulations

“Air as a service” – Creating new Business Models

Reduced maintenance costs and spare parts inventories, improve customer satisfaction

Requirements

- **Enablement of classic industrial equipment** with next generation IoT technology
- **Real-time data feeds** from **globally distributed** customer equipment
- **Predictive analytics** in real-time to identify potential failure before it occurs
- **New Product offering** for large environments to protect Market positioning

Solution Value Prop

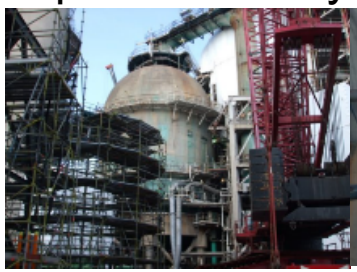
- Real-time analytics on sensor data from distributed equipment
- Large memory and processing capability to meet workload performance needs
- **Business Model development** based on the new capabilities
- SAP HANA implementation with Predictive Maintenance components

Customer Value

- Est. **$10M annual inventory savings** via predictive analytics
- 60% reduction in downtime
- **New Product introduction (as a service)**
- **Competitive advantage** via increased customer care and uptime
- Future product improvement
- Platform to develop addtl. services on other product segments

Fire and Safety

Improve security and efficiency





Requirements

- Improve security and efficiency
- Increase value proposition

Solution Value Prop

- Real-time analytics on sensor data from distributed equipment
- Large memory and processing capability to meet workload performance needs
- HPE (mobile) Wifi solution based on mesh technology
- Active Wi-Fi tags, track personnel and assets indoors and out
- Wi-Fi personnel tags with man-down alert

Customer Value

- Helps direct first responders by showing who is mustering in place and who is not
- Improves operational efficiency: optimizes labor dispatch and cost reduction

Dubai Smart City Platform

Dubai smart city platform – Integration of infrastructure and apps development

“Our goal is for the entire city’s services and facilities to be available on smartphones. We want to provide a better quality of life for all”

His Highness Shaikh Mohammad Mohammed Bin Rashid Al Maktoum,
Vice-President and Prime Minister of the UAE and Ruler of Dubai

Solution

- Platform to transform Dubai into a smart city for the coming 5 years
- HPE will build the platform, integrate with data sources, design dashboards and maintain and operate the setup
- Spearheaded by Consulting leveraged HPE servers and storage, Helion OpenStack and 3rd party vendors (Hortonworks/Hadoop)

Value to business

- Bring Dubai to the top tier of Smartest Cities on the planet
- Ensure a collaborative approach fueling entrepreneurship & Global competitiveness
- 6 dimensions: Smart economy, smart governance, smart people, smart living, smart mobility and smart environment
- 100 initiatives and 1000 smart services

Value to Dubai gov. and citizens

- Dubai Smart City Platform will empower government, public and private organizations and citizens/residents to harness 100% of the data sources through optimal information systems and analytics
- Contribute to the Happiness Index of Dubai
- A platform for additional digital services

No external reference yet

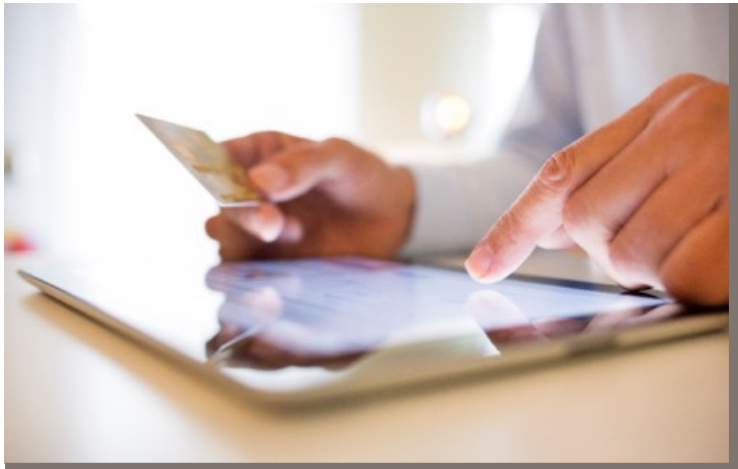
Hewlett Packard Enterprise

Hewlett Packard
Enterprise

Financial Services

HPE's Fraud Detection and Prevention solution is the answer

Real-time, AI-enhanced (deep-learning) fraud analysis



Fraud Detection and Prevention solution details

Artificial Intelligence Engine

ProLiant DL380, **Apollo** 2000 and **Apollo** 6500 with **Nvidia**
Mellanox InfiniBand Fabric

Hewlett Packard Enterprise · Mellanox Technologies · NVIDIA · kinetica · intel

Deep-Learning Framework
accelerated by **NVIDIA GPU's**

SW-based Applications
for Fraud Detection

Detect fraudulent transactions easier, more accurately, in real-time across various platforms and data sets

Solution benefits

- **Optimal Performance:** Real-time fraud detection with in-memory databases accelerated by GPU's - ideal for machine-learning & data-science workloads. Cross-functional, virtual team integration and monitoring
- **Cost Effective**: Right-sized configurations for any size FSI account; more cost effective than upgrading legacy technology. Providing underlying infrastructure as a service for data collection, integration, and analytics software (e.g., Flex Capacity)
- **Confidence:** Identify fraud patterns to prevent future fraudulent transactions
- **Tested**: Vetted with industry analysts, Pre-tested & validated on HPE platforms

Hewlett Packard
Enterprise

Artificial Intelligence and Deep Learning

Deep Learning - Rapidly growing customer use

Artificial Intelligence (AI) based Pattern Matching

Video | Text | Images | Audio

Neural networks—networks of hardware and software that approximate the web of neurons in the human brain.

Massively Parallel Processing

Training Module

- Minimize Training Time (scales **exponentially** with number of parameters)
- Data Fusion from multiple data sets for inference
- Ultra-scale real time data analytics

Inference Engine

Infrastructure requirements

- GPU capacity for Parallel Processing
- Parallel Performance (e.g. I/O, Networking, Data movement)
- TCO
- Simplicity
- Scalability

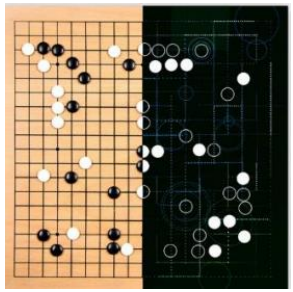
Game of GO Example

- 19 x 19 board
- 250 moves per turn
- More possible positions than atoms in the universe

Google AlphaGo System

- 170 GPU cards
- 1,200 standard CPUs

System isn't just learning from data provided by humans. **It's learning by playing itself, by generating its own data.**



Possible use cases

- Robotics
- Scientific research
- Siri-like mobile digital assistants
- Financial investments

★ **Huge Breakthrough for Artificial Intelligence**
Google's AI Beats a Top Player at the Game of GO

Deep Learning requires extreme massively parallel processing performance for Training models

Scenario 1

Manufacturers developed AI solution for long-haul trucks to drive and deliver autonomously on divided highways

- **Proposal**: **Retailers and Distributors** would like to implement this technology but need to apply **AI Models** to determine routes and driver schedules to optimize this new capability. **Airlines** are willing to contribute to the project a portion of their Autopilot data and scheduling which has been in production for many years. The DoT also contributes road data. HPE contributes resources to pilot the Autonomous trucking solution.
- **Solution**:
  - Develop new models for routing and scheduling based on the data of the members
  - Models may be commercialized and take them to other companies around the world
- **Benefits**:
  - Cost and Energy from the new Semi-Autonomous Trucking solution.
  - Quick implementation of scheduling and routing changes at all members

Scenario 2

Outbreak of a dangerous communicable disease occurs in San Francisco, NY or London. Companies need to quickly find out the status of their employees and control exposure to risk

- **Proposal**: The **CDC** and the **American Red Cross** sponsor the membership to develop new models for determining affected individuals based on recent travel patterns and define a mechanism to regularly contact exposed individuals to identify potential symptoms.
- **Solution**:
  - A task force of **corporate, government and non-profits** develop models for understanding traffic flows with the areas in question.
  - An **AI bot** and database is created to contact people who have travelled to the area
  - By applying this model potentially affected individuals are quickly located
  - More sophisticated models may be developed for ongoing tracking a predicting the effect of the change
- **Benefits**:
  - Quick response to unexpected issue through understanding Data and models of the member companies.
  - Ability to generalize the solution and results for a broader set of enterprises.

Scenario 3 - The Samantha Morton Example

Pre-cognitive abilities to predict crimes before they happened by "predictive policing" techniques

- **Proposal**: Large US cities, London, Dubai, Shanghai, already have begun to deploy AI technologies for public safety and security. These include cameras for surveillance that can detect anomalies pointing to a possible crime, drones, and predictive policing applications. Machine learning and AI is already used to combat white collar crime such as fraud. It is also used to automatically scan social media to highlight people of risk of being radicalized by ISIS.
- **Solution**:
  - A team from **HPE, Universities, Law Enforcement and Criminologists**, develop models on crowd simulations
  - Detect plans for disruptive events from social media
  - Monitor activity at large gatherings of people to analyze security
- **Benefits**:
  - Reduction of Crime.

Summary

Leader in HPC

Pioneering with Industry Solutions

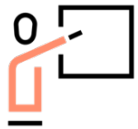
Cloud28+

Largest partner ecosystem including Cloud28+

HPE is leader in HPC and HPCaaS

…and leads with Services and Service Providers

Largest Service Providers Ecosystem



Advisory & Transformation

Professional Services

Operational Services

Hewlett Packard
Enterprise

Thank you

Contact information