ETP 4 HPC

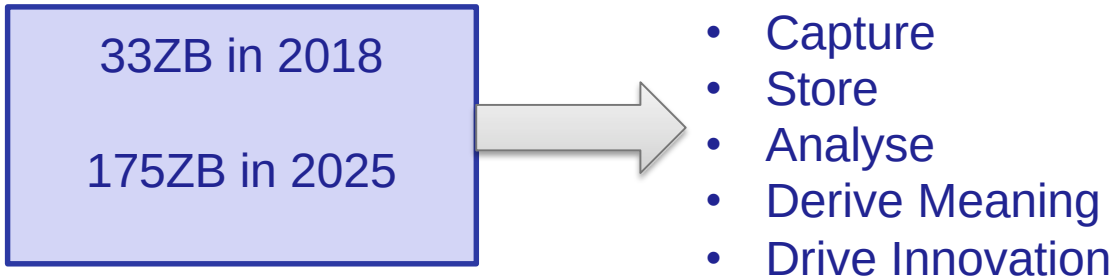
HPC Cloud/Edge Infrastructures: Data Focus areas

Presented at:
European Partnerships for New Digital Age Workshop
Sept'12 2019
, Brussels

Sai B Narasimhamurthy
EU R&D, Seagate
Steering Board Member, ETP4HPC

Importance of Data

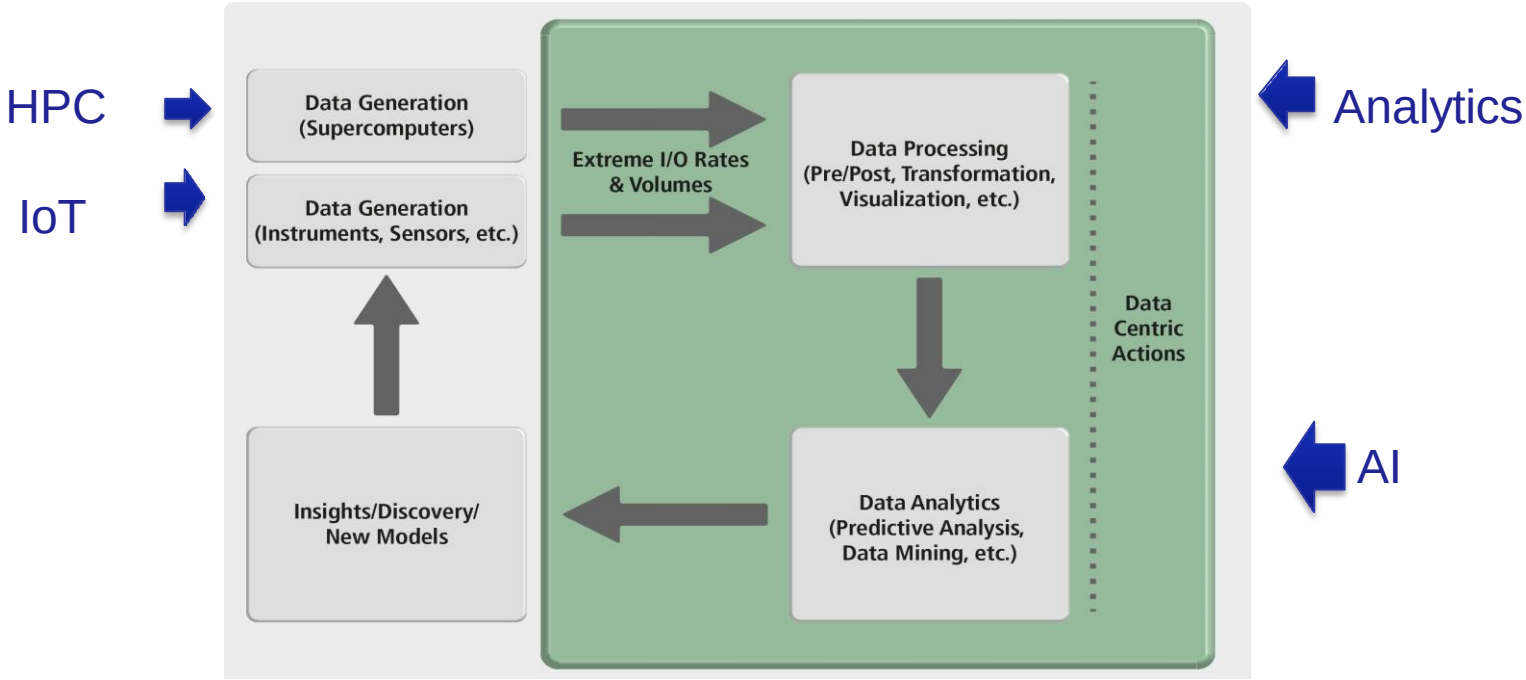
- Evolution of the Methods of Scientific Innovation
  - Observation and Experimentation (1000s of years in the past)
  - Mathamatical Models (Last few centuries)
    - Newtons Laws of Motion, Keplers Laws, etc
  - Simulation Science (Last few decades)
  - Data oriented methods (State of Play!)



Ever Growing Data Sphere!
(Source:IDC)

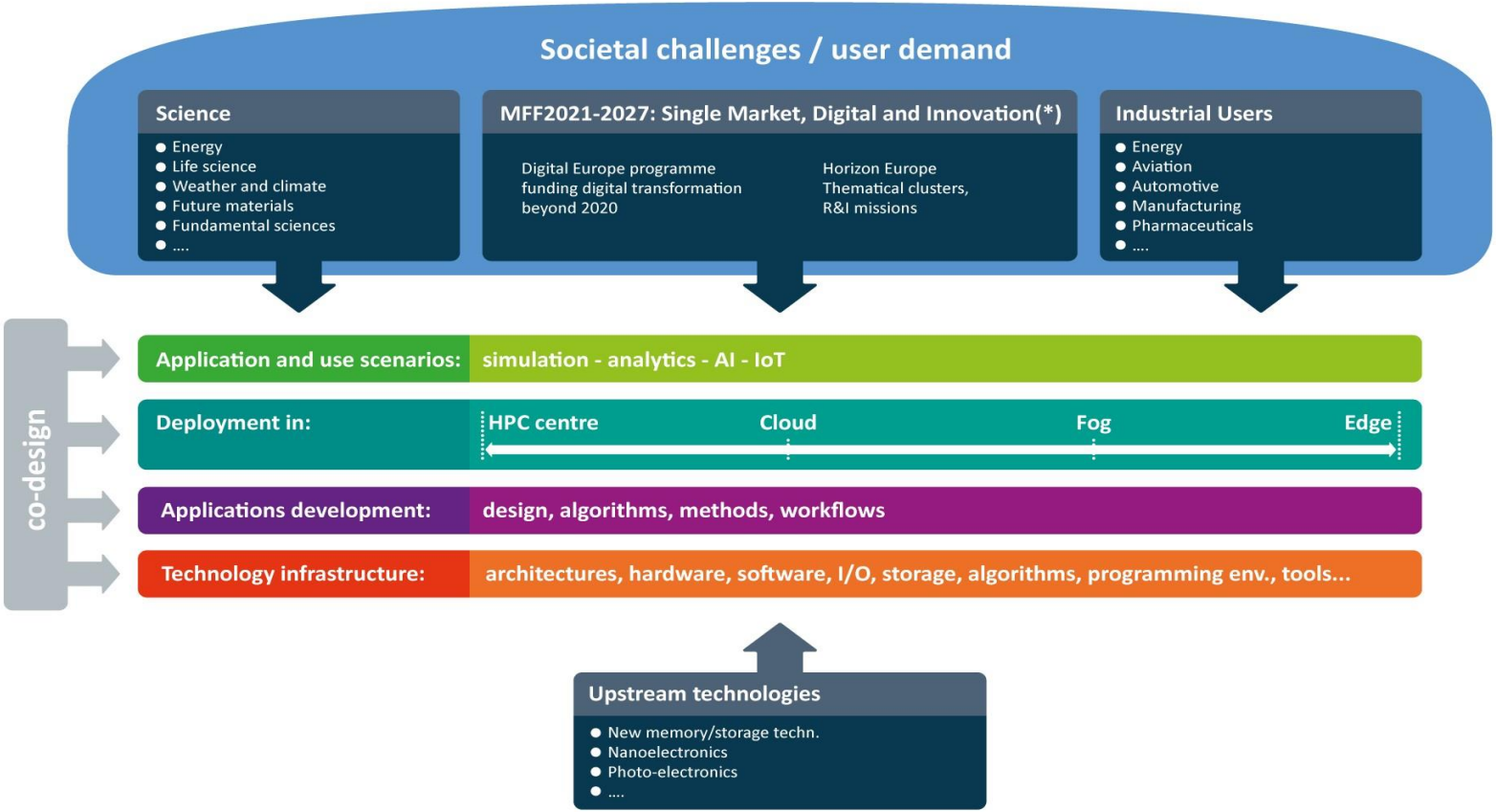
ETP4HPC

A Data Driven Virtous Cycle is evolving in HPC



ETP4HPC

A New HPC Perspective: Increasing Interplay between Simulation, AI, IoT, Analytics



Data & Storage Considerations for the new age

Edge Location

Caching Infrastructure
"The Fog"

Data Core Site
(Eg: SAGE Prototype at Juelich)

Need for better Network Infrastructure (Edge to Core)

Key Data/Storage Considerations for the Edge

- Handling Ephemeral Data Pools
- Time Critical Data Analysis
- Time Critical Learnings from Data
- Low Latency Storage Tech
  Role for SSDs and Non Volatile Memories
- High Density Shippable Drives
  Nothing beats the bandwidth of a truck full of drives (Unfortunate!)

AI/DL workflows are changing long held data and storage tech. assumptions !

Key Data/Storage Considerations for the Core

- New software infrastructure
  (New File Systems, Object Storage Tech)
- Multiple Types of Storage Device Tech
- Hierarchical Storage pools -
  Different performance points
- API's to deal with different data types
- Ability to ship computations to storage
- Big Role for Non Volatile Memories
  (3DXPoint and Beyond)
- Telemetry Analytics on storage Infra.
- Resiliency handling for storage failures
- Quality of Service for Data Access
- Handling Storage Tech. Within Compute Nodes for AI/DL workflows

ETP4HPC

HPC+Edge+IoT - Some Hard Problems

- Tracking Data Provenance - Data Life cycle Management
  - How long to Archive? (Changing storage media)
  - How to deal with changing data formats?
  - How do we deal with exploding metadata?
- How do we deal with data consistency?
- Do we need new paradigms for data federation?
- How do we handle sensitive data across geographies?

ETP4HPC

Opportunities for Collaboration

- Involving IoT community in ETP4HPC's discussions with EuroHPC
  - HPC Communities' Ambitious goal of achieving Exascale in 2023 timeframe
  - Industry User Working Group (IUWG) initiated as part of ETP4HPC
    - We will look for discussions with stakeholders in the IIoT space as well!

- More collaboration of the "data" community through ETP4HPC+BDVA+AIOTI

- FETHPC Projects focussed on storage and I/O
  - SAGE, Sage2, NEXTGenIO..etc
  - May have very good touch points to European Large Scale Pilots Programmes in the IoT Area
  - New round of EuroHPC calls released - Calls due Jan '2020!

ETP 4 HPC

EUROPEAN TECHNOLOGY PLATFORM FOR HIGH PERFORMANCE COMPUTING

Thank you

office@etp4hpc.eu

www.etp4hpc.eu