

A blurred photograph of a crowd of people walking across a street with white zebra crossings. The motion blur suggests a busy, fast-paced environment.

Percipient StorAGe for Exascale Data Centric Computing



per·cip·i·ent (pr-sp-nt)

adj.

Having the power of perceiving, especially perceiving keenly and readily.

n.

One that perceives.



Introducing: Seagate

Cloud Systems and
Electronic Solutions



You Probably Think of Seagate as a Hard Drive Manufacturer...

- We are this AND
also a company that

- serves many types of customers and businesses
- has deep expertise and unique IP in storage and data management
- runs supply chain operations recognized as world class

Now we are expanding our mission and focus...

Seagate Drives



HDD

- 55,000+ personnel worldwide
- 40% of the world's digital information stored on Seagate products
- Supplier to 8 of the 10 largest global cloud service providers
- Technology leader with 14,000+ patents
- Over 2 Billion drives shipped

+ Flash and Controllers via LSI Acquisition



HDD	FLASH
SILICON	

- Adds world-class capabilities in Flash and Controller technologies
- #2 PCIe Flash supplier
- Ability to work with all NAND suppliers for innovation and cost benefits

+ Systems via Xyratex Acquisition



+

xyratex

SYSTEMS

HDD

FLASH

SILICON

- Over 17,000 Exabytes shipped
- Strong and growing IP portfolio
- Leader in High Performance Computing (over 100 engineers)
- Leader in OEM Storage Systems (over 400 engineers)
- Design wins in Energy, Genomics, HPC

+ Data Protection & Cloud Services via EVault Acquisition



HYBRID SOLUTIONS

SYSTEMS

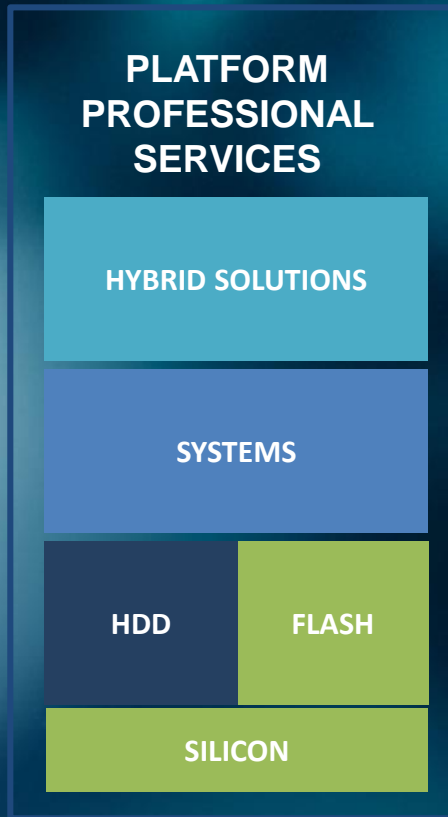
HDD

FLASH

SILICON

- Deep cloud storage and operations expertise
- Serving 43,000+ cloud storage customers
- Key Partnerships with IBM, Iron Mountain, Microsoft and others

What It Means to You



- Broad solutions portfolio that scale as customers need
- Ability to combine silicon, component, system and software innovation in new ways
- Backed by world-class operations and supply chain and long-term financial strength and operations expertise



Seagate Springtown Manufacturing Plant

- Manufacturing in Northern Ireland since 1993
- Manufacturing ~60% of all Seagate read/write heads
- The only major disk drive manufacturing plant in Europe
- 1,400 Personnel
- £722M investment



**Seagate Systems
Havant
R&D and Manufacturing
Plant**

- Building storage systems since 1994
- Over 17 Exabytes of storage manufactured to date
- Systems IP registered in Europe
- Leading supplier of enterprise-class storage systems

Strategic European Development Partnerships

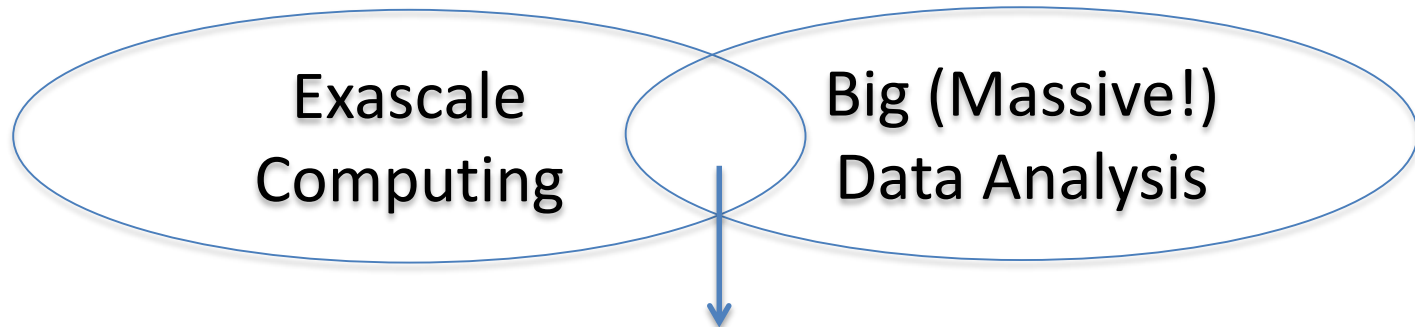




Percipient Storage



Future



Exascale Data Centric Computing
(**B**ig **D**ata **E**xtr^em^e **C**omputing, or *BDEC*)

We cannot just build storage for Exaflop scaled processing
we must accommodate the Ingest, Store, Process and
Management of extreme amounts of data

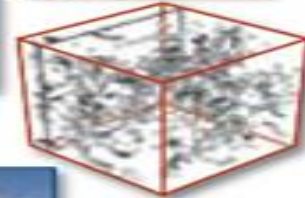
Data Centric Computing

A Perspective

Science Paradigms

- Thousand years ago:
science was **empirical**
describing natural phenomena
- Last few hundred years:
theoretical branch
using models, generalizations
- Last few decades:
a **computational** branch
simulating complex phenomena
- Today: **data exploration** (eScience)
unify theory, experiment, and simulation
 - Data captured by instruments or generated by simulator
 - Processed by software
 - Information/knowledge stored in computer
 - Scientist analyzes database/files using data management and statistics

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - K\frac{c^2}{a^2}$$



SEAGATE

Source: "The 4th Paradigm" - report from Microsoft Research

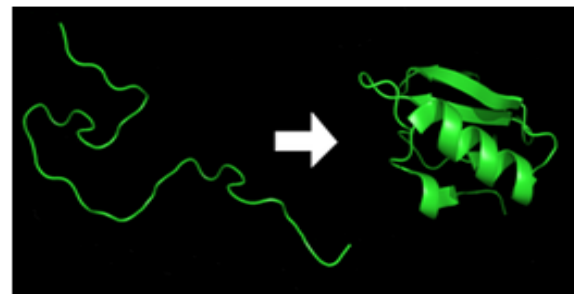
Data Centric Computing : Big Science



High Energy Physics
1 PB of raw data/ sec



Biology : 8PB for 100M Neurons



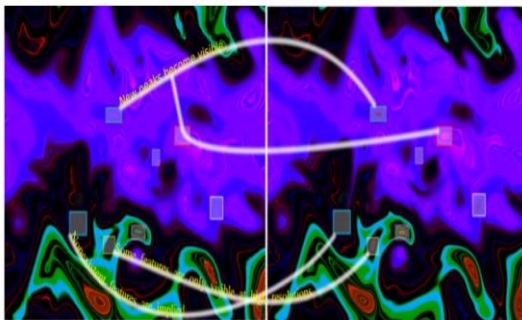
Medicine : Exascale to tackle mysteries of Protein folding



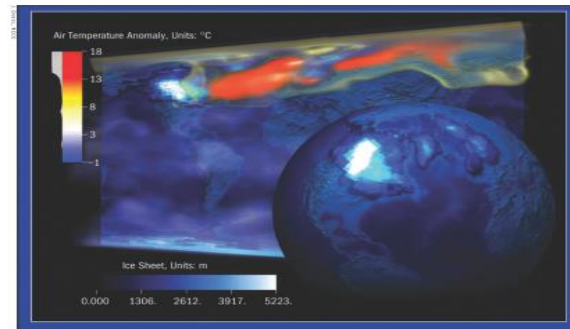
Astronomy
SKA:100PB - 3EB/yr



S E A G A T E



Industrial (eg:Oil and gas)
Visualizations and Volume rendering



Climate change effects

Data Centric Computing

HPDA(High Performance Data Analytics)

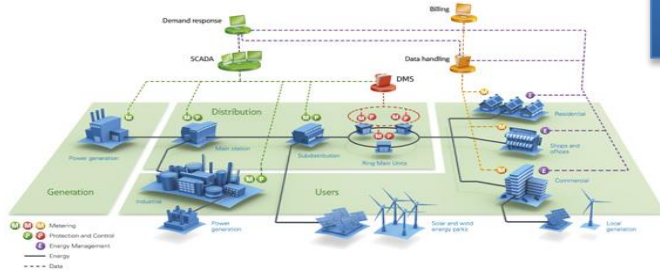


Automated Fraud Detection

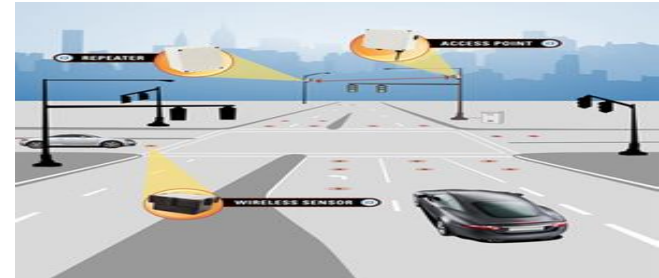


Social Network Data Analysis

HPC for Big Data Analysis



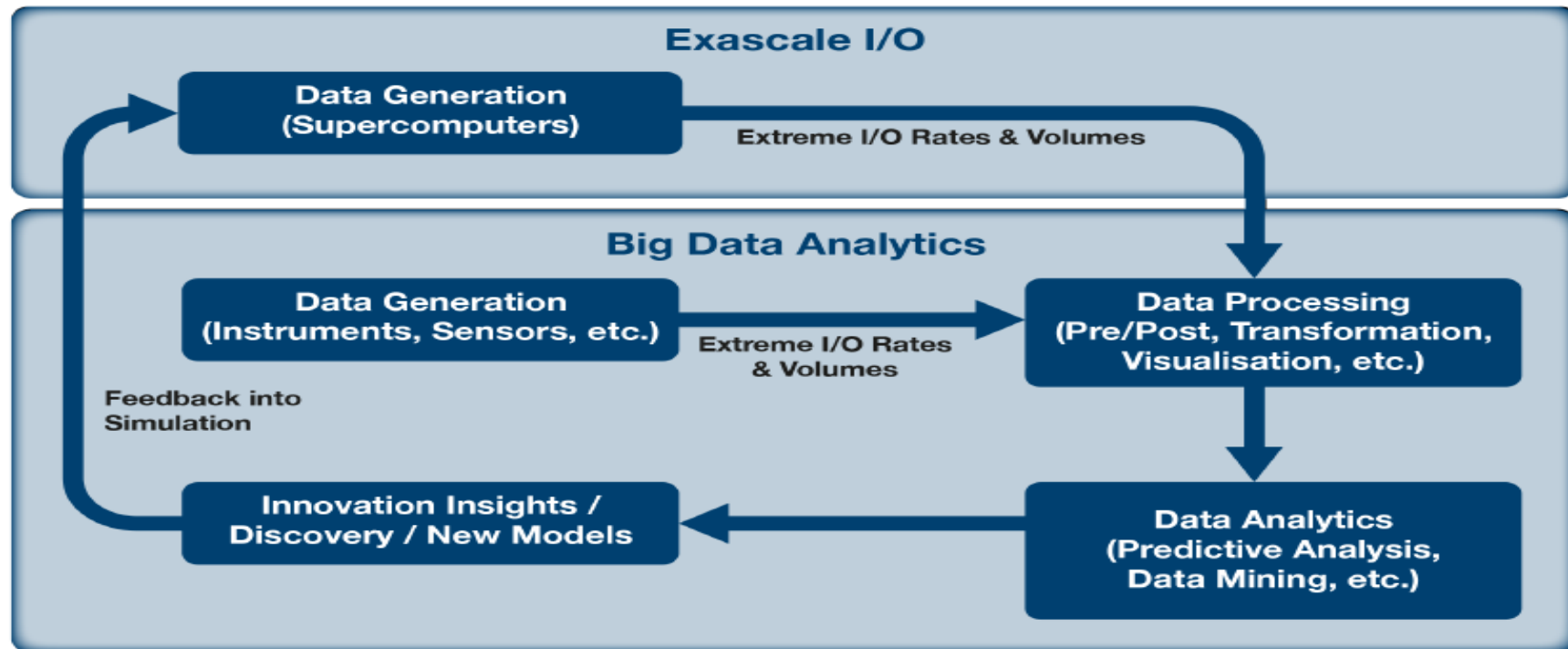
Smart Grid Data



Sensor Network Data

Knowledge Discovery in Data Centric Computing

The Links between Exascale I/O and Big Data Analytics



Key Solution Requirements

Dimensions of Needed Innovation*

Data Storage
Innovation

Data & System
Management
Innovation

Data Processing
and Analytics
Innovation

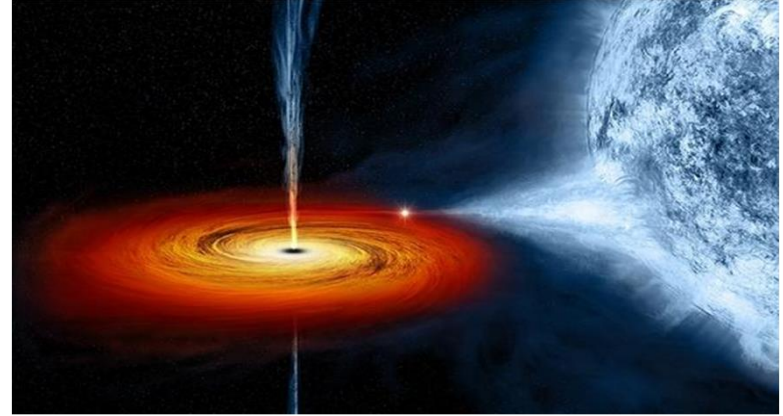
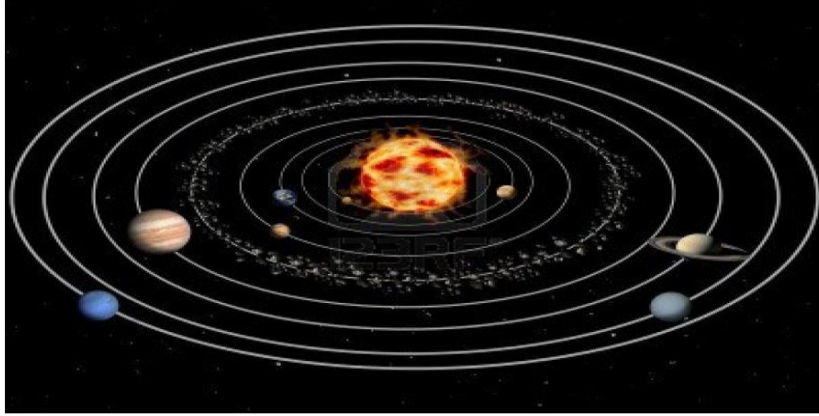
A “Percipient” Storage & I/O Oriented Solution

* *Based on Research Priorities for Storage & I/O identified by the Strategic Research Agenda(SRA) for Europe being continuously developed by the European Technology Platform for HPC (ETP4HPC)*



S E A G A T E

The Quality of “Percipience”



↗
The Old Paradigm
'Isolated' Storage &
Computing

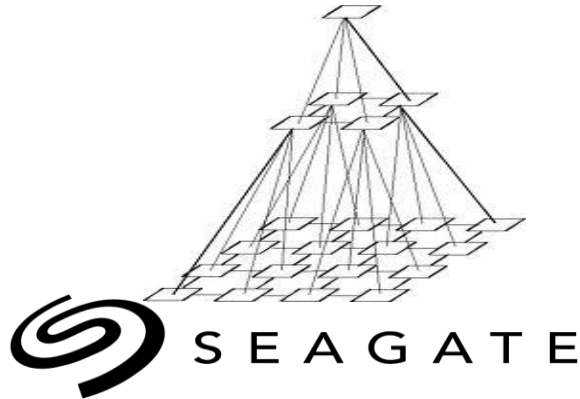
“PERCIPIENCE”

↖
The New Paradigm

**Having a homogenous view of
data anywhere/everywhere**

Data Storage

Data Storage Innovation



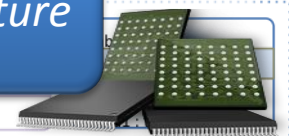
Balanced I/O: Graceful scaling with compute side innovations, Objects

Resiliency/Reliability: Hierarchical check pointing & High Availability Schemes

Deep I/O tiers: Incorporating future NVRAM Technologies

Autonomics: Dynamically Provide appropriate "Containers" to apps, etc.

Standardised APIs:
For Apps & for data storage infrastructure
(Eg. Object Storage)



Data Processing and Analytics

Data Processing and Analytics Innovation

Extreme Data Processing: *incl. Big Science Experiments, need advanced Big Data Analytics Methodologies*

Active Storage: *Storage Systems with compute capabilities*

On the Fly Data Manipulation: *In-situ data processing in the I/O stack/infrastructure (eg: Network)*

Energy Reduction Strategies: *Computation takes less energy than data movements*

System & Data Management

System & Data
Management
Innovation

S
Y
S
T
E
M

Infrastructure Telemetry, Analytics

Infrastructure Simulation

HSM/ILM: Storage Management with
many Layers

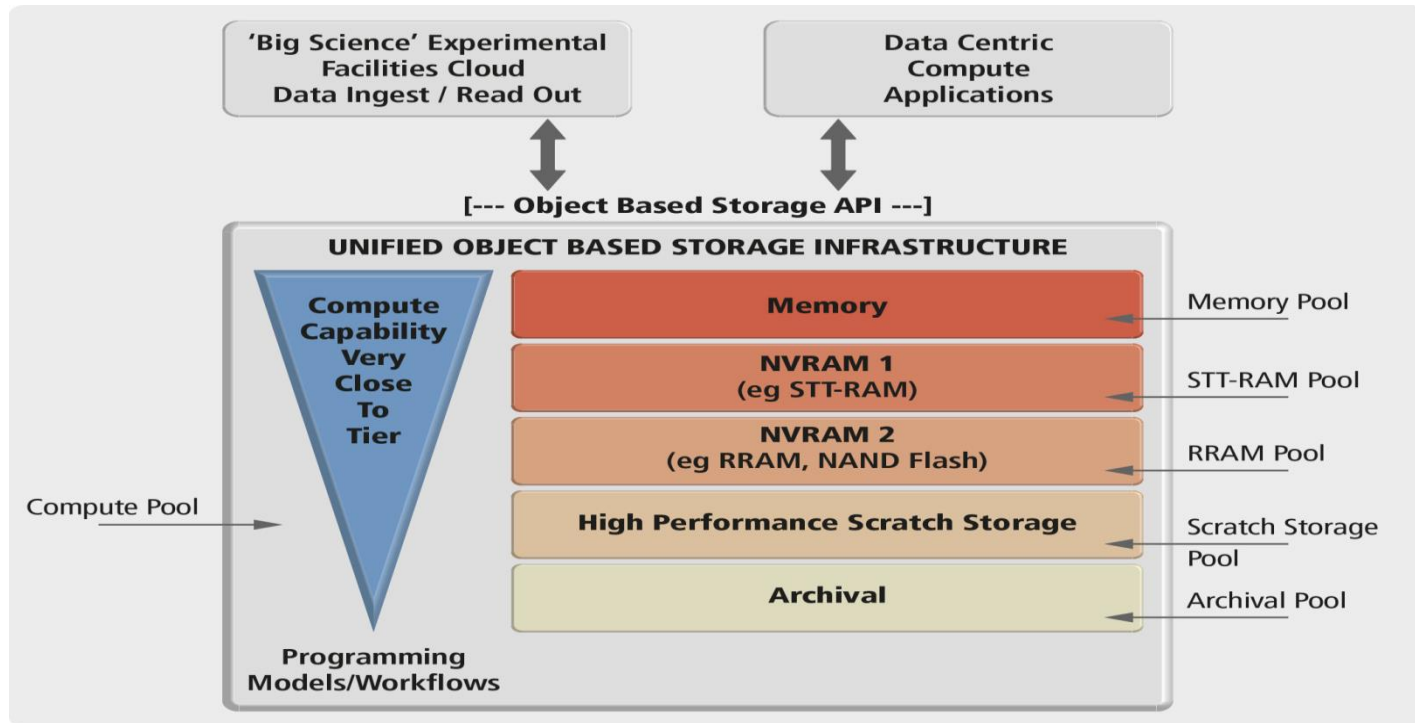
Data Layouts/Transformation:
*Different "Views" to the same data
(eg: POSIX, HDF5, NetCDF)*

Data Indexing/Provenance Tracking/ILM

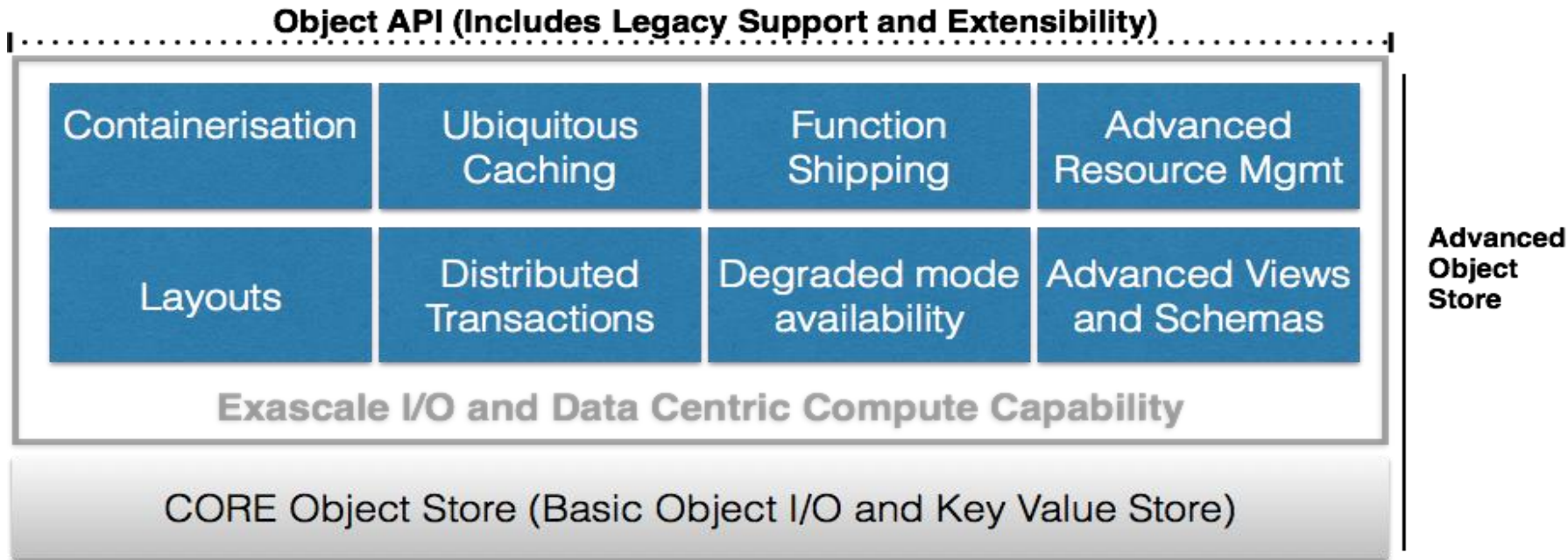
Plug-in Frameworks for adding data
management "apps"



A Percipient Storage Architecture



Advanced Object Store in Percipient Storage





Malcolm Muggeridge