Big Data Science Challenging The Oil Industry
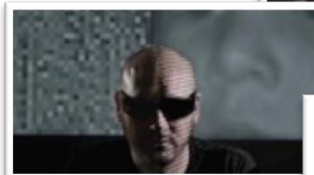
Mahesh Seshadri
EMC Distinguished Engineer,
CTO Global Services
EMC Corporation

EMC²

ATEA

big•data \ datasets so large they break traditional IT infrastructures.

EMC²

ATEA

Transforming Business, And Our Planet!

Over 200 Stories About How Big Data
Is Changing Our Planet

Big Data Characterized by Volume, Variety and Velocity

Structured Information In Relational Databases

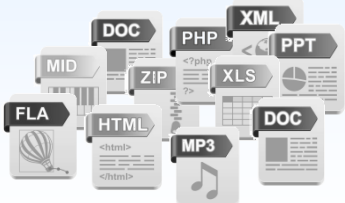## Managed & Unmanaged Unstructured Information

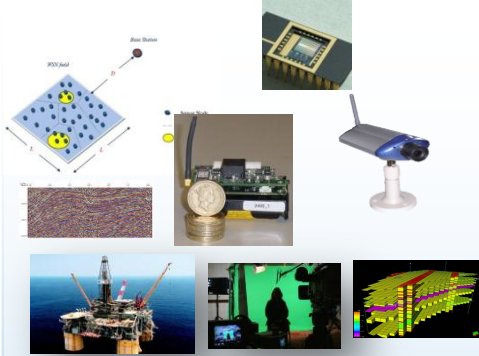## Internet Of Things

Telemetry, Location-Based Information etc.

Non-Enterprise Information

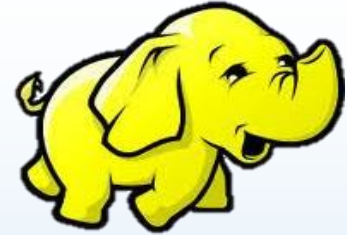**The Digital Universe Is Growing By 7,600 PB / Day**

Big Data Technology Enablers

More Data Sources, More Samples, More Sensors More Variety

Lower Cost To Store and Process Data

Availability of Powerful Algorithms & Tools

Goal: Gain A Competitive Edge!

Platform changes…

Existing Architecture

Emerging Architecture

Apps

HW

Network

Software Defined Network

New Analytic Platforms Are Different

Structured & Unstructured Data

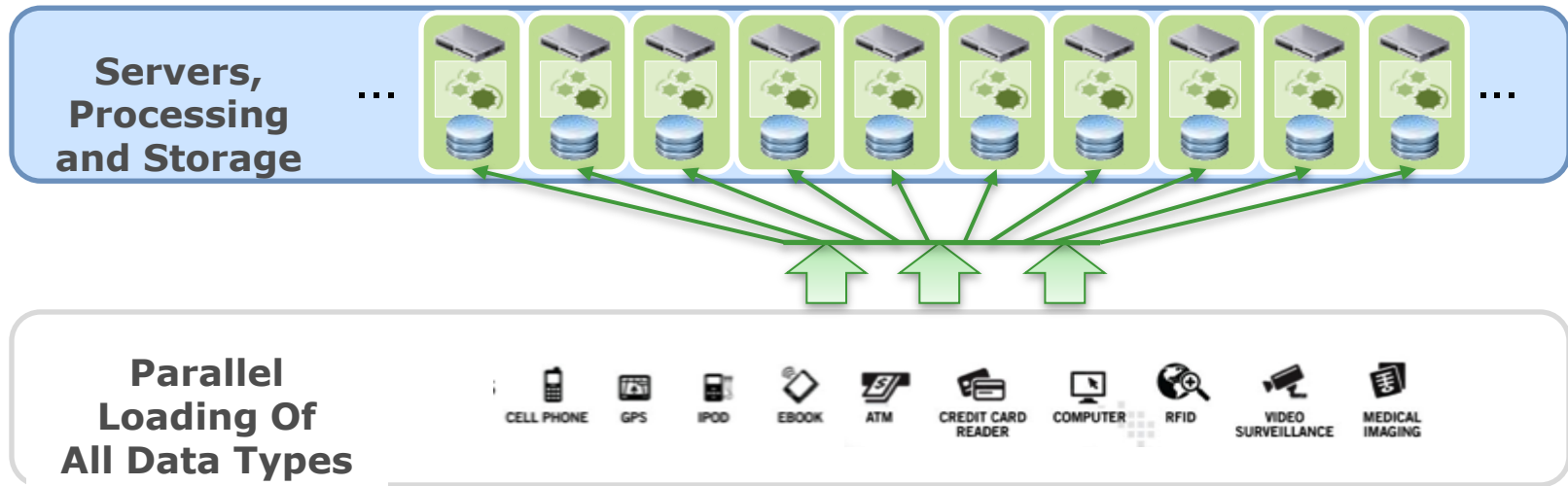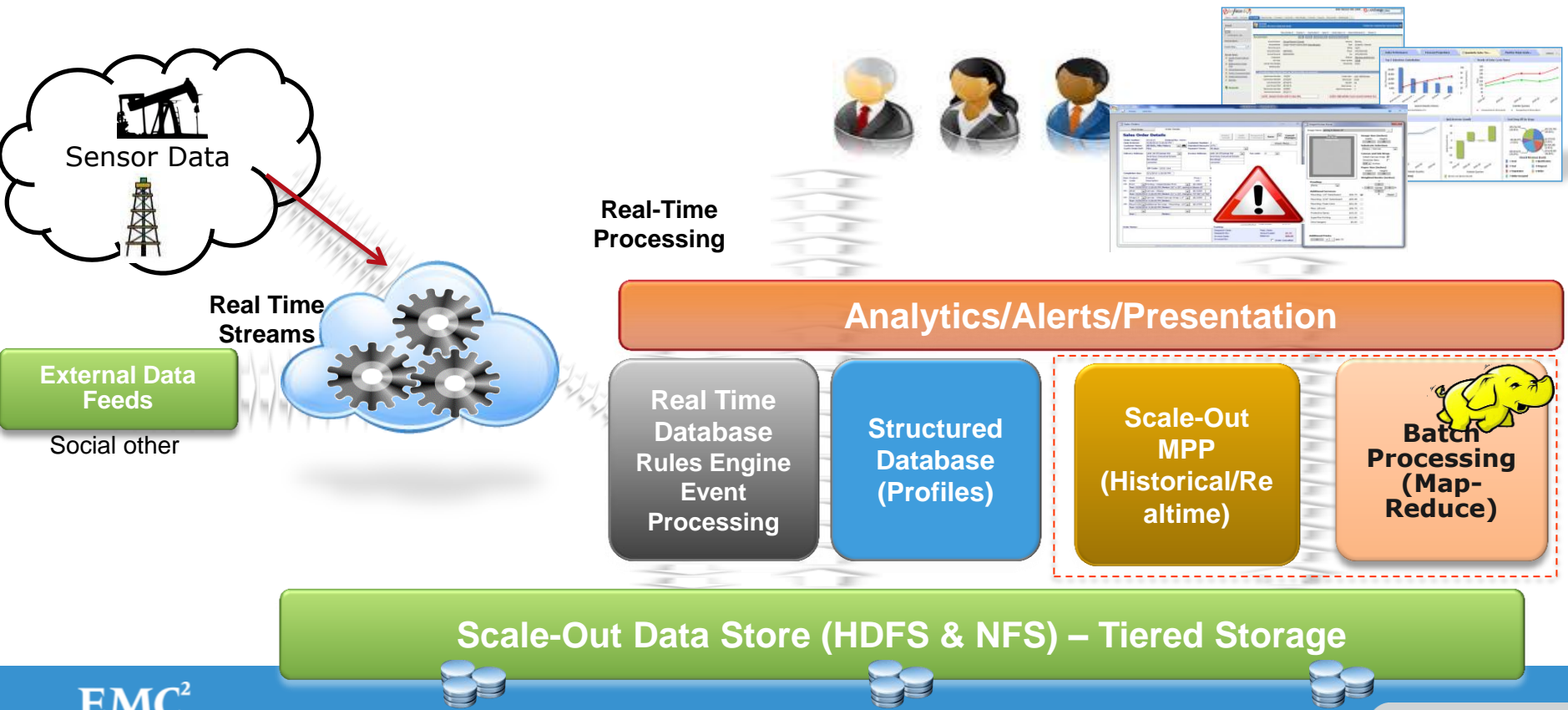Analyze Petabytes Of Current Data

Virtual, Scale Out Architecture

Scale-Out Infrastructure MPP and NoSQL/New SQL Architectures enabling Big Data Analytics

Provides extreme performance through parallelism

Servers, Processing and Storage

Parallel Loading Of All Data Types

CELL PHONE · GPS · IPOD · EBOOK · ATM · CREDIT CARD READER · COMPUTER · RFID · VIDEO SURVEILLANCE · MEDICAL IMAGING

EMC²

New Open Scalable "Big Data" Systems:

Sensor Data

External Data Feeds

Social other

Real Time Streams

Real-Time Processing

Analytics/Alerts/Presentation

Real Time Database Rules Engine Event Processing

Structured Database (Profiles)

Scale-Out MPP (Historical/Realtime)

Batch Processing (Map-Reduce)

Scale-Out Data Store (HDFS & NFS) – Tiered Storage

EMC²

ATEA

BI focuses on managing and reporting on **existing data** to **monitor** and **manage** concerns within the enterprise
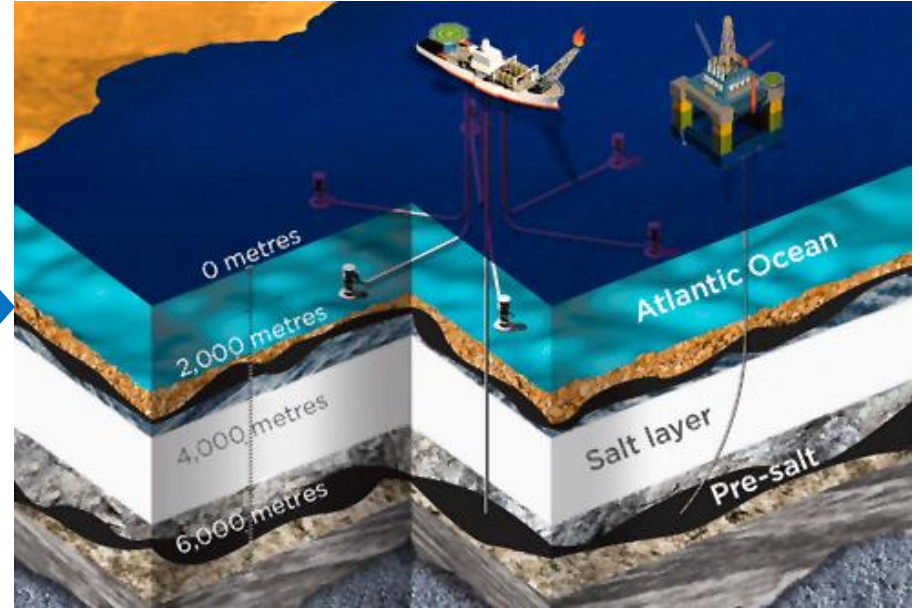
Data Science applies advanced **analytical** tools and algorithms to generate **predictive insights** and **new** product **innovations** that are a direct result of the data
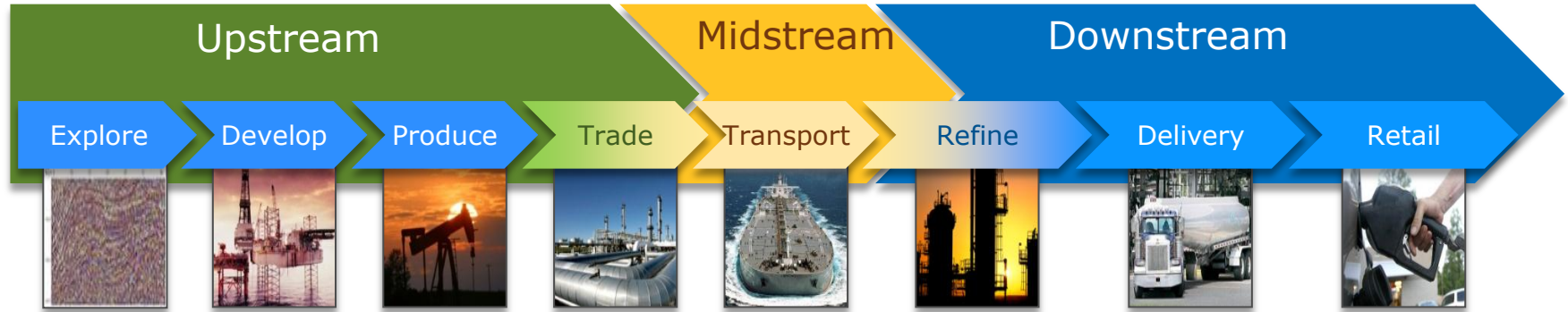
Harvard Business Review

Data Scientist:
The Sexiest Job of the 21st Century

Who Is The Data Scientist?

EMC²

ATEA

The End of the Era of Easy Oil Drives More Complexity into Big Data Science

The Ultimate Goal of "Big Data Science"

| Upstream				Midstream	Downstream		
Explore	Develop	Produce	Trade	Transport	Refine	Delivery	Retail

- Find more Hydrocarbons ? – Speed & Accuracy
- How can we extract, develop and produce safely, efficiently and with minimal environmental impact ? – Speed, Safety, Security
- How can we refine and distribute products in the most efficient and cost effective manner ? –Speed, Efficiency (scheduling optimization)
- How to we manage customers ? – Identify Opportunities, markets

Big Data in Upstream

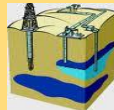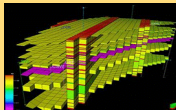## "Volume", "Variety" and "Velocity" of data in upstream processes

Exploration   **Development**   **Engineering and Production**
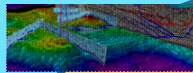
- Seismic acquisition
- Seismic processing

- Seismic interpretation
- Geological interpretation

- Reservoir Modeling
- Reservoir Simulation

- Facilities & Reservoir Engineering
- Drilling program
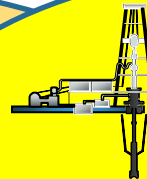
- Drilling & Test

- Production Development and Optimization
- SCADA/Sensors

- Production

• Seismic
  – SEGD, Pre-Stack, Post-Stack (PB)
  – Wide azimuth offshore data acquisition

• Navigation

• Recently Acquired and Historical
  – Log curves   – Tops
  – Production data   – Lithology
  – Drilling and test   – Cultural
  – Micro seismic   – Cores

• Real Time streaming data from well-heads and drilling equipment
  – LWD   – Flow
  – MWD   – Pressure
  – Mud logging – Rate of Penetration

Volume   **Variety**   **Velocity**

EMC²   ATEA

Big Data Challenges for Upstream

Bigger data at Higher quantities analyzed More Analytics moving at Wider distances needing to be Better managed accessible over Longer lifecycles of time

Innovation Trends

- Higher precision models
- More measures, more often, more places
- More iterations
- Longer production periods

- More scarcity of even more specialized skills
- Greater collaboration and automation
- More analytical processes over longer time frames
- Access and Back-Haul Bandwidth constraints

EMC²

ATEA

Innovation Through Applied Research

EMC's Brazilian Research Program
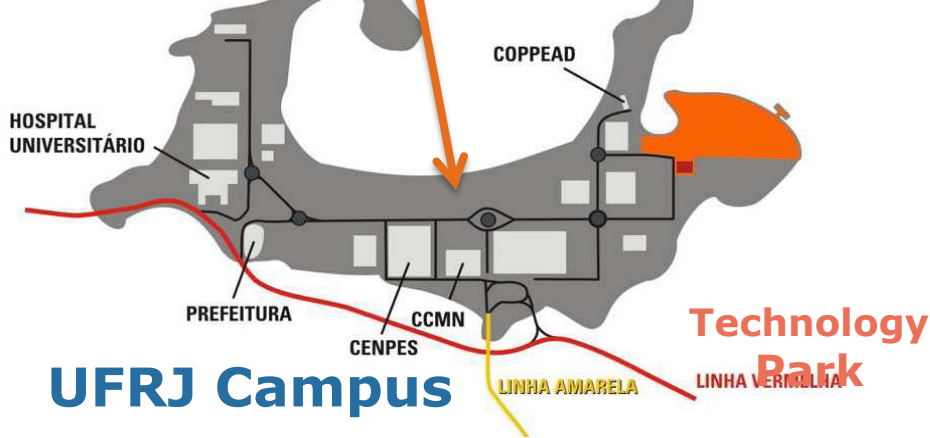
- **R&D Center** in the Federal University of Rio de Janeiro Tech Park
  - On the UFRJ Campus with Petrobras R&D, Schlumberger, Halliburton, GE & others

- **30+ Big Data Scientists**
  - Collaborating with 50 others on Campus
  - Partnering with other IT suppliers

- **Research** in **Oil & Gas**
  - Acquisition, Analysis, Collaboration & Visualization of Seismic Data

Research Focused on Oil & Gas Key Requirements

- Improve performance of seismic data processing time by orders of magnitude

- Reduce oil production cost by reducing time-to-oil & improving computational accuracy

- Minimize human presence at platforms through automaton & remote monitoring (smart fields)

- Improve information management to deal with the explosion of information throughout the Upstream process

- Speed up all analytics performed during Exploration & Production

- Improve quality & time to visualize information by improving accuracy & enabling sharing

ATєA

Big Data Use Cases - Upstream

Upstream			Midstream		Downstream		
Explore	Develop	Produce	Trade	Transport	Refine	Distribute	Retail

- ➤ **Exploration** — Using Hadoop and Greenplum Database for storage, quick visualization and comprehensive processing and imaging of seismic data
- ➤ **Production and Operations** — Making the drilling platforms and pipeline infrastructure smart to anticipating issues and acting to prevent failures and increase productivity
  - ➤ **Drilling Safety[1]** – Prevent undesired events. Must act in real time on drilling data and formation evaluation data. Use this data for predictive modeling.
  - ➤ **Asset / Reservoir Management –**Simulating fluid flow adds a layer of complexity to a model. Determine drilling coordinates through oil shale to optimize the # of wellheads needed for efficient extraction of oil
- ➤ **R&D Optimization –** Root cause analysis of design flaws w/in Engineer projects. Understand incident reports to develop strategy to address critical design issues

EMC²

ATEA

Reduce Exploration Risk

- Parallel ingest and processing of seismic data with minimal data movement
- Every minute saved results in thousands of dollars in savings
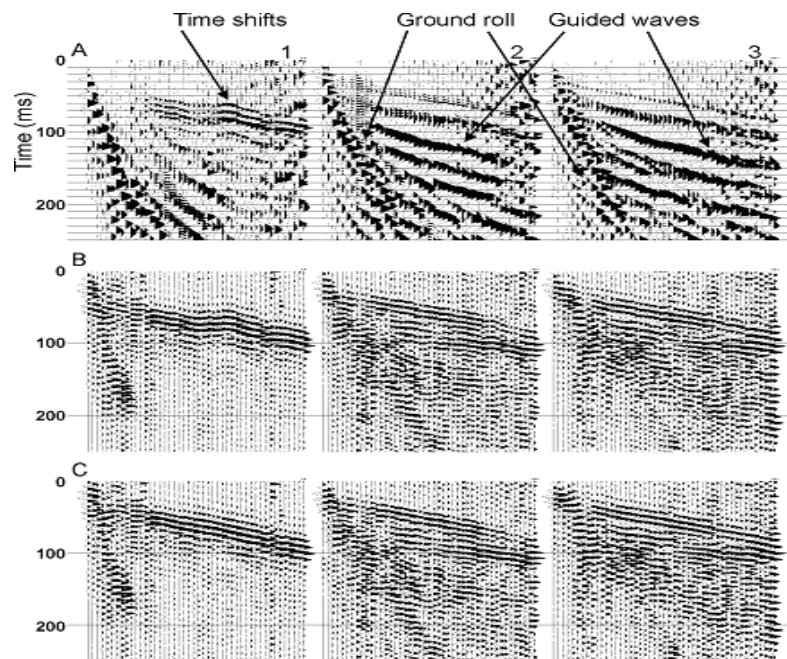
Traditional Seismic Processing

Geoscientist
"I don't care if the storage is full. I need my interpretation projects to stay online…..forever."

Infrastructure Management

"My storage is full, again. My budget has been cut. Do I really need to buy more? If so, how much?"

Primary Storage Expense

$

$

Primary Tier 1 Storage

Primary Tier 1 Storage

Primary Tier 1 Storage

Tape

Tape

Tape
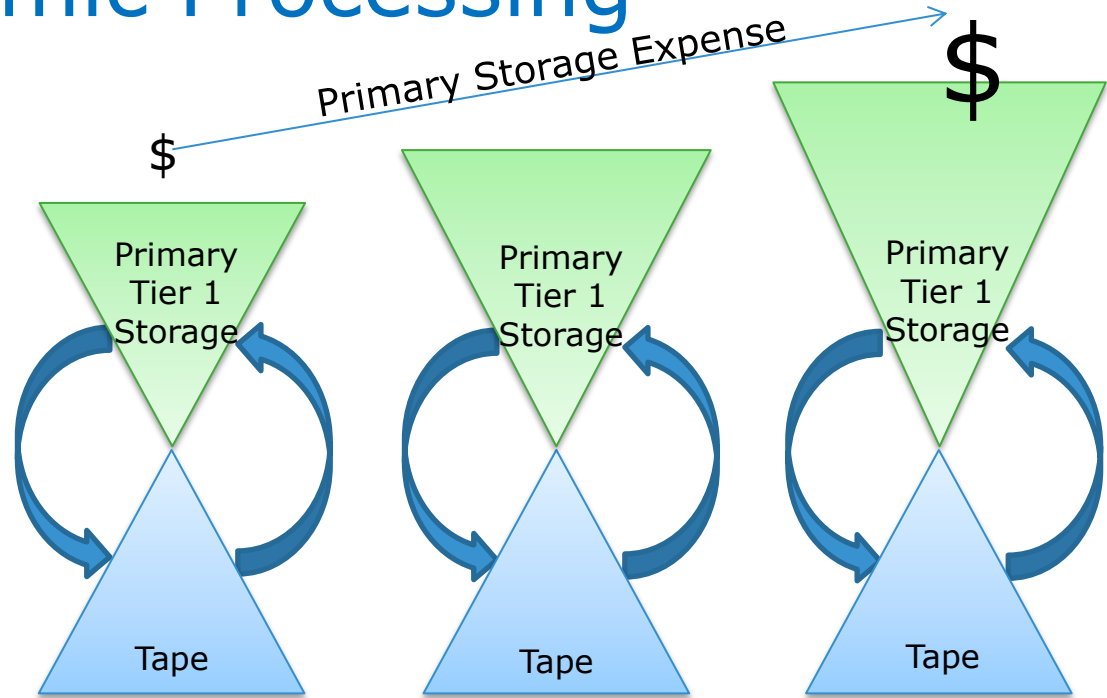
" Two Tiers….online and gone"

Increasing Online Data Requirements

ATEA

Intelligent Storage Model

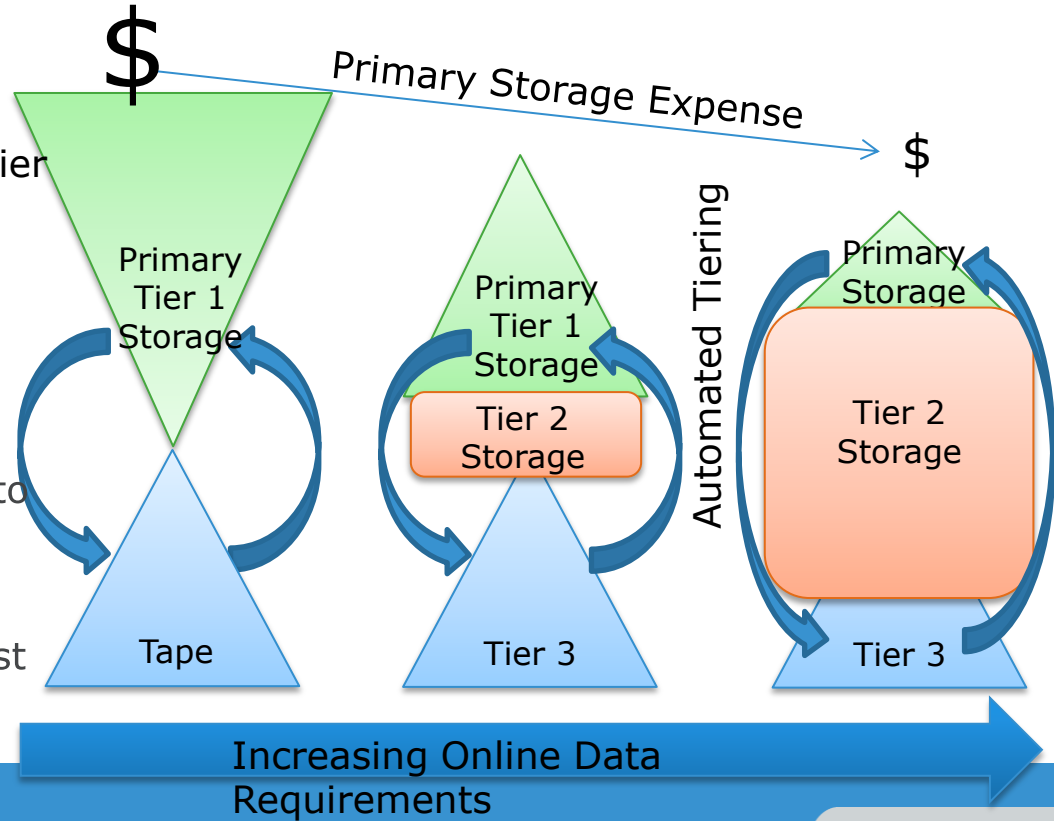- The right data, in the right tier. Automatically move data up and down Tier

- Onsite or offsite repository

- Get more data online (e.g.Pre-Stack SEGY)

- Stage projects for processing or interpretation

- Archived projects stay online and easy to access

- Reduce primary storage by 70%

- Scale Out NAS Reduce management cost

$

Primary Storage Expense

$

Automated Tiering

Primary Tier 1 Storage

Tape

Primary Tier 1 Storage

Tier 2 Storage

Tier 3

Primary Storage

Tier 2 Storage

Tier 3

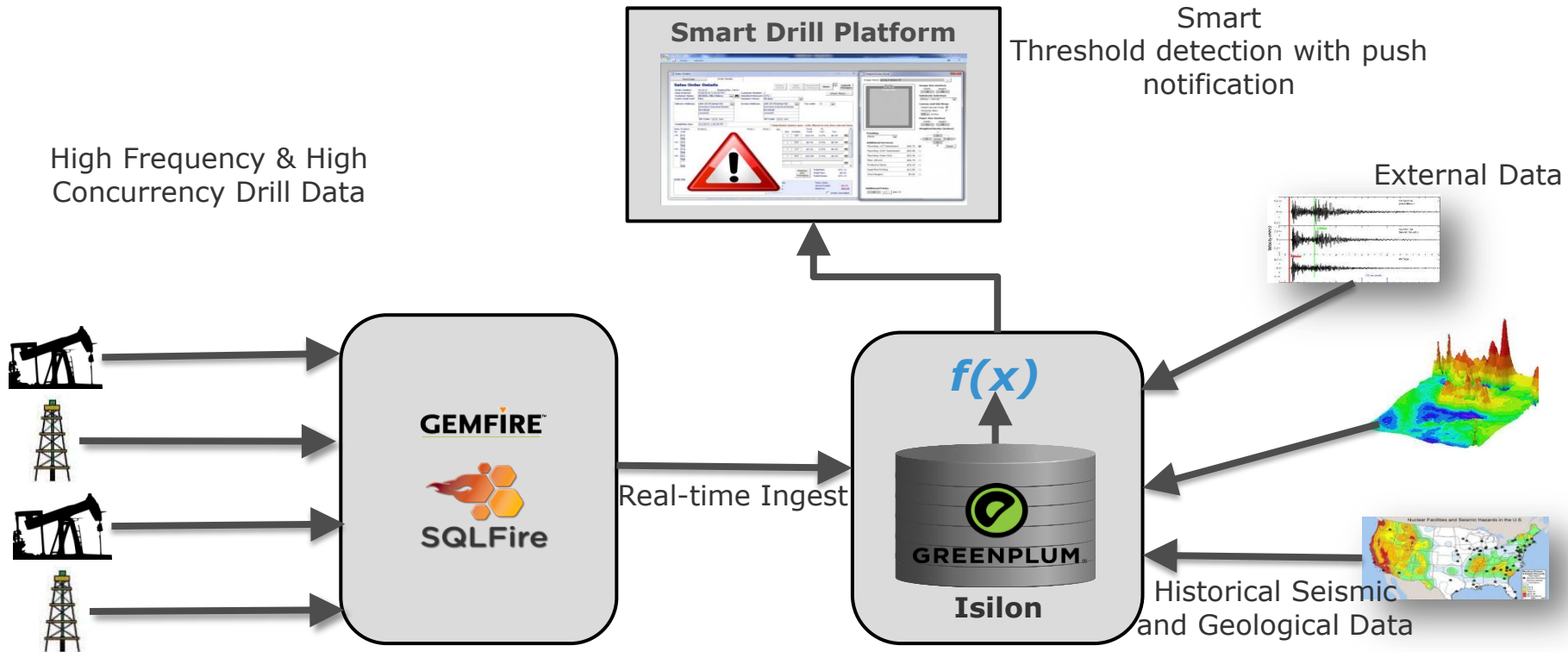Increasing Online Data Requirements

EMC²

ATEA

Optimized I/O and Compute Platform for Seismic Processing

- Study seismic processing workflows and algorithms (such as Reverse Time Migration and Full Wave Inversion) to create an architecture and API to optimize data access performance throughout the HPC stack.

- Time consuming tasks: jump between compute intensive and I/O intensive phases

- Parallelization opportunities and alternatives

- Numerous algorithms from O&G companies and Software vendors

- Explore mechanisms to move compute closer to the storage

- Context-aware storage intelligence based on data structures and algorithms

Monitoring & Prediction within Real-Time O&G Sensor Networks

- Create an integrated platform, based on "real-time" <u>data analytics</u> and <u>predictive modeling,</u> to better control and secure critical real-time data streams, engineering data, processes and process control systems for the O& G industry, improving <u>risk management</u> and mitigation from the seismic model to engineering, plant, and facilities operations.

- Key points
  - Integration of various different data sources
  - Prediction from historical data+simulations+domain knowledge
  - Boundary conditions based on current and possible future states
  - Continuous monitoring and evaluation of boundary conditions
  - Support for decision making when boundary conditions are not satisfied
  - Automatic update of prediction models when necessary

Predictive Analytics with Real-time Ingest

Smart Drill Platform

Smart Threshold detection with push notification

High Frequency & High Concurrency Drill Data

External Data

GEMFIRE™

SQLFire

Real-time Ingest

f(x)

GREENPLUM

Isilon

Historical Seismic and Geological Data
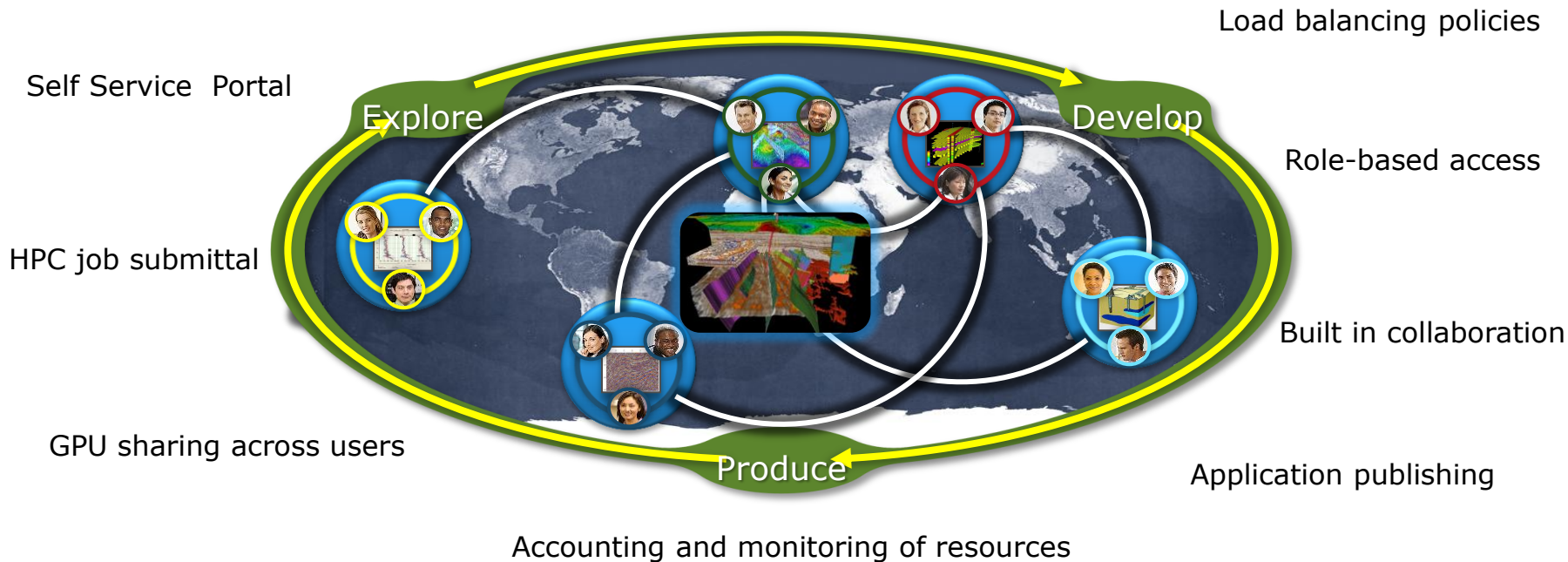
EMC²

ATEA

Improve Drill Accuracy and Safety

- Build and train drilling models based on all existing well data
- Refresh models based on incoming sensor data from drill rig
- Predict drill maintenance/downtime, optimize drill parameters, and prevent blowout accidents

Next Generation Upstream- Enabling Global Collaboration Anytime, Anywhere

Remote 3D visualization for OpenGL apps

Load balancing policies

Self Service Portal

Explore

Develop

Role-based access

HPC job submittal

Built in collaboration

GPU sharing across users

Produce

Application publishing

Accounting and monitoring of resources

EMC²

ATEA

Big Data Uses Cases: Midstream

Upstream				Midstream		Downstream	
Explore	Develop	Produce	Trade	Transport	Refine	Distribute	Retail

- **Environmental Monitoring –** Real-time ingestion of sensor data and storage to conform to gov't policy and company protocol. Help predict maintenance based on level of pollution emissions
- **Crude Assay –** Provide unique service to refining customer by predict the oil going to the refinery. Reduce set up time by understanding the quality of the raw crude being processed.
- **Predictive Maintenance –** Identify problem areas in advance and reduce non-productive time. Store maintenance logs and inspection data for future liability risk. New sensor equipment is becoming more sensitive and capable of collecting more data, must be prepared to ingest and analyze in short time period

Big Data Uses Cases : R&M

Upstream | Midstream | Downstream

Explore | Develop | Produce | Trade | Transport | Refine | Distribute | Retail

- **Refining:**
  - **Preventative Maintenance Optimization:** Turnarounds[1] are usually planned 1 – 2 years in advance and last 1-4 weeks. Develop real-time equipment sensor monitoring system that networks all machines in order to analyze and optimize maintenance based on cost and needs
  - **Simulating Refining Operations:** Distillation / FCCU[2] optimization determines inputs & parameters for optimal control of the refining process. Conducting a crude oil assay[3] to help streamline process is not currently done.
- **Marketing / Retail:**
  - **Gas Station Automation Optimization[3]:** Connecting the many systems and developing new systems like RFID identification to that run the daily activity of a gas station can improve overall services for customers who buy primarily of price. Build holistic view of life time value of each location and optimize distribution network based on location
  - **Loyalty Programs/Marketing:** Give customers a reason to come back w/ targeted incentives and services. Expand fleet services as well and commercial vehicle offerings to build enterprise business relationships

EMC²

ATEA

Summary

1 Big Data Is Transforming How We Work and Live

2 The Approach To Big Data is Different

3 Embrace Big Data Now For Business Success

ATEA

Strong partnership EMC & Atea

- Leading Nordic and Baltic IT infrastructure provider
  - Largest in Nordic, 2nd. largest in Europe
  - No. 1 in all markets
  - 20+ Billion. NOK revenue

- Most certified
  - More than 7500 technologycertifications

- Strong local presence
  - 82 locations in 7 countries
  - 6 000+ employees

- Unique combination of local presence and European size and knowledge

EMC²

ATGA
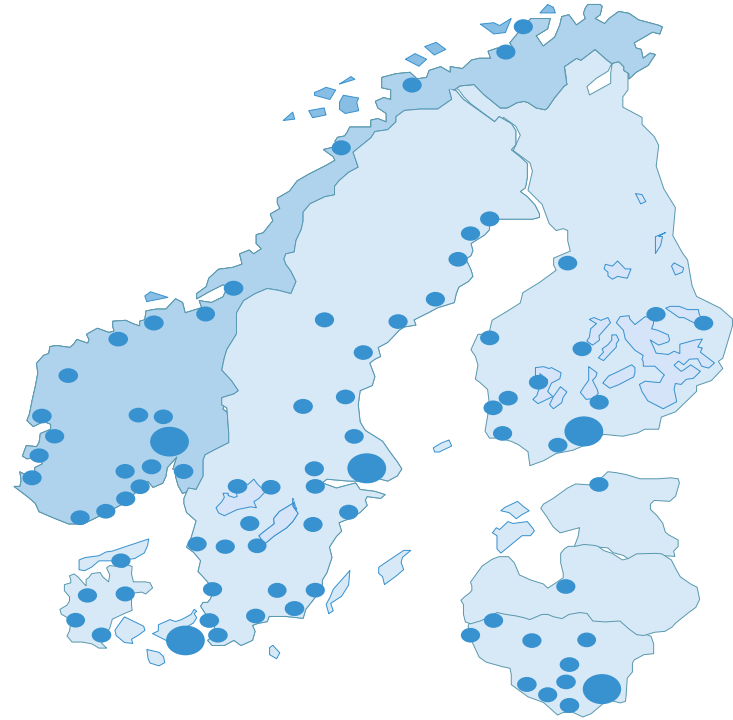
Upstream - Business Drivers

- Better identification of reservoirs and reserves (improved geoscience) (50% success rate at a rate of $23M/ well development)
  - Reduce waste of drilling exploration wells
  - Improved understanding of future strategy based on available oil (bbl's)
  - Predict the chances of success of turning reservoir into a production well
  - Optimization of drilling resources by not over drilling well site (wellhead to hydrocarbon volume ratio)

- Safety, Speed, Accuracy
  - These help reduce risk and cost to an E&P operation

We Have Entered The Information Era
Creating Value From Information *Is* The New Economy

THE ERA OF

BIG DATA

IS HERE

EMC²

ATEA