Managing Seismic Data

Implementing Intelligent Data Management

So How do You Decide which Data to Keep?

Tony Klapcia

Decision Making – how do you decide?

- Which Data is valuable?
- What do you keep?
- Why?
- What is the value
  – Business?
  – Technical?
  – Compliance?
  – Emotional?
- Who decides?
- How long do you keep it?
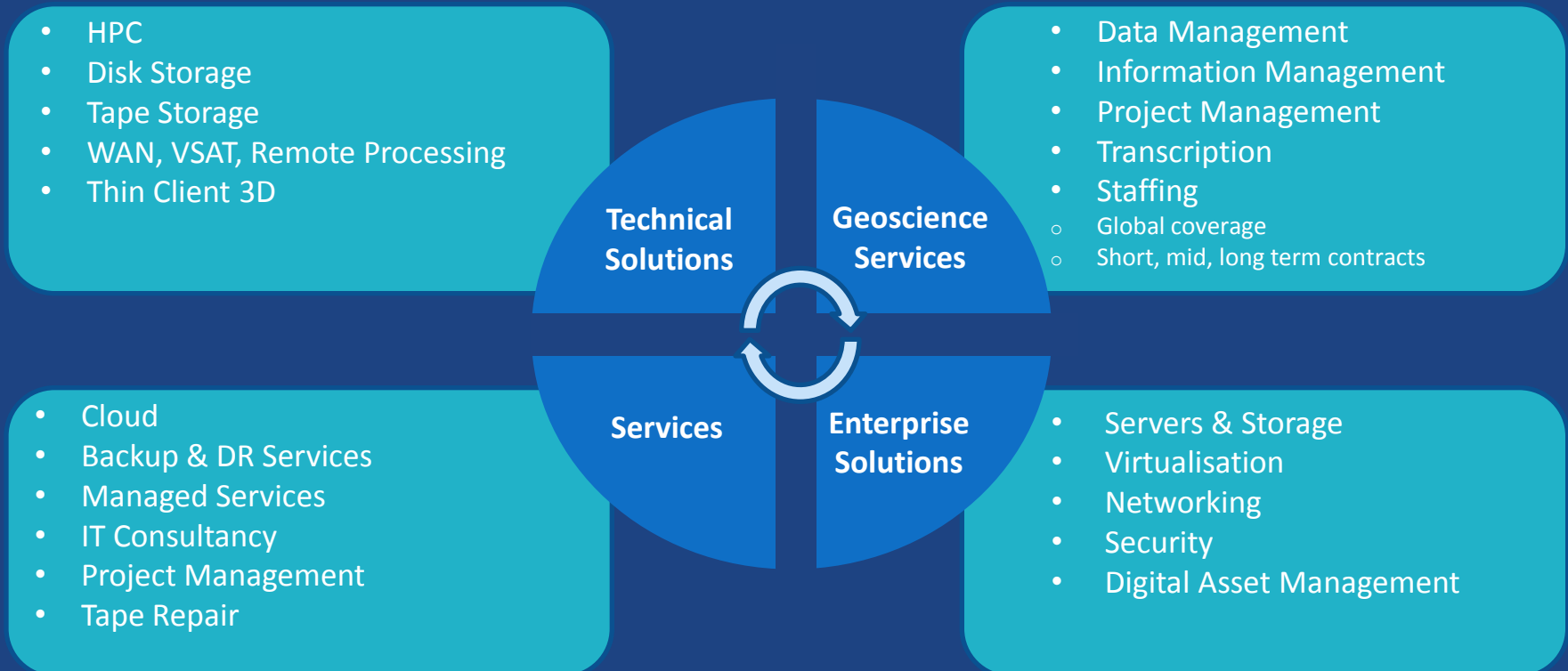- Where?
- How is it protected?

Creates:

- Multiple copies
- Multiple owners
- Geographic Dispersal
- Confused Management
- Lack of Protection
- Poor value to Lines of Business
- Increased Cost

No Best Practice

Eurotech Intro

eurotech computer services ltd

Founded 21 years ago by Upstream Oil & Gas people for Upstream Oil & Gas people
*Understand how to apply **IT** to Data/Info Management*

Technical Solutions

- HPC
- Disk Storage
- Tape Storage
- WAN, VSAT, Remote Processing
- Thin Client 3D

Geoscience Services

- Data Management
- Information Management
- Project Management
- Transcription
- Staffing
  - Global coverage
  - Short, mid, long term contracts

Services

- Cloud
- Backup & DR Services
- Managed Services
- IT Consultancy
- Project Management
- Tape Repair

Enterprise Solutions

- Servers & Storage
- Virtualisation
- Networking
- Security
- Digital Asset Management

IBM Business Partner

TARMIN

www.eurotech-computers.com

Coherent Data Management?

eurotech
computer services ltd

Issues	**Needed**
•Multiple data sources	•Automatic ingestion - Policy driven
•Poor management of data	•Removal of duplicate data
•Stored at multiple locations	•Automatic movement of data – ILM
•No security	•Long term archiving - To Disk, Tape, or Cloud
•Unreliable storage mediums	•Global multi-site coverage
•Different data types	•Compliancy
•No regulatory adherence	•Secure environment
•No internal compliancy	–Protect against loss/corruption
•Visibility limited to select few	–Prevent theft or intrusion

Business Partner

TARMIN

Creating a Data Management Platform

eurotech
computer services ltd

- Policy driven
- Ingestion of semi-structured, unstructured data
- Next generation HSM
- Information Lifecycle Management for G&G
- Archival, - Retention & controlled shredding of data
- Encryption. - E-search, - Audit
- Protected & Secure
- Flexible Performance
- Low TCO

Deliver Active Archive

- Migrate, manage & share data in a secure repository
- Instant access to archived data
- Flexible, automated policies
- Provide:
- Data Integrity, Media Monitoring, Self-healing, Versioning
- Automatic ILM

IBM Business Partner

TARMIN

www.eurotech-computers.com

Flexible Work Flow

- Transparent Data Access
- Centralised access to distributed data
- GIS interaction
- Removal of physical management layers
- Easy internal & external data distribution
- Create Corporate Data Store or Databank
- Pre-stack, post-stack
- Original field data
- Seismic survey & Geometry data
- Well data
- Documents & Images, License data/info
- Production data
- SEGD, SEGY, LIS, LAS, DLIS, etc

Compliancy & Security

- Forbids unauthorised access
- Full audit trails
- Tracks all changes and versions
- Variable retention policies and periods
- No accidental deletion
- Full shredding (*DoD standard*) upon deletion
- Maintain License obligations
- Legal & Industry obligations met
- Adhere to Corporate policy
- Encryption *(standard algorithms: AES, 256, Blowfish, 448)*
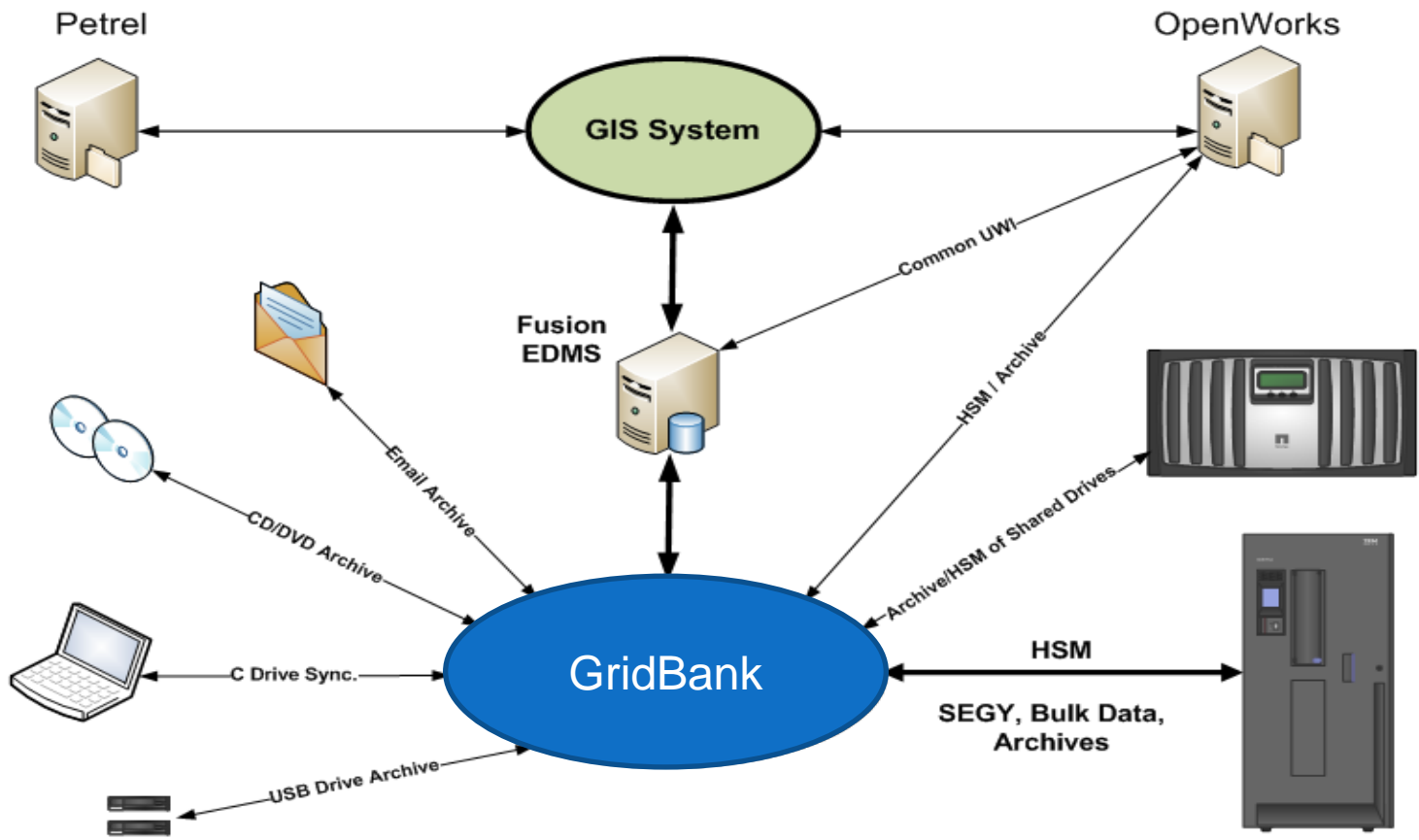- WORM option

Access & Availability

Access

- Anywhere Access
  – LAN, WAN
  – Laptop
  – Mobile
- Secure role based access
- Open interface
  – Protocols include: CIFS, NFS, WebDAV, HTTP
- Google like e-search of archive
- Search on normal strings within metadata
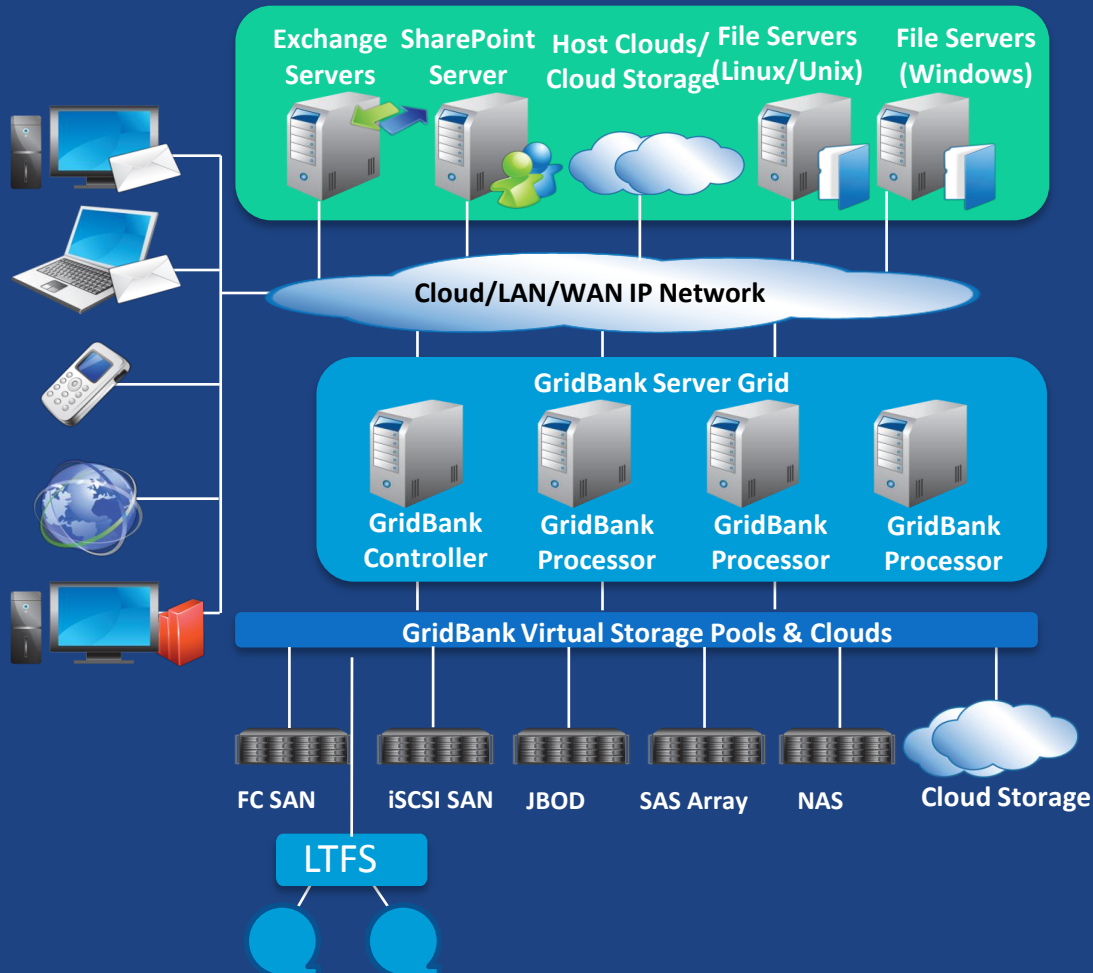- Cloud enabled architecture

Availability

- Simplified data migration
- Application awareness
- Data De-duplication
- Thin provisioning
- Grid architecture for no SPOF
- Built-in real time replication to remote site
  – Disaster recovery
  – Business Continuity

Solution?

GridBank Architecture

eurotech computer services ltd

Exchange Servers

SharePoint Server

Host Clouds/ Cloud Storage

File Servers (Linux/Unix)

File Servers (Windows)

Cloud/LAN/WAN IP Network

GridBank Server Grid

GridBank Controller

GridBank Processor

GridBank Processor

GridBank Processor

GridBank Virtual Storage Pools & Clouds

FC SAN   iSCSI SAN   JBOD   SAS Array   NAS   Cloud Storage

LTFS

Understand
- Search & Discovery
- Big Data Analytics

Control
- Information Governance
- Multi-Site Replication

Store
- Storage Optimisation
- Object Scalability

IBM Business Partner

TARMIN

www.eurotech-computers.com

GridBank

An IM solution not an IT solution

- Information Lifecycle Management that fits G&G
- Manage *all* data – seismic and general unstructured, *and Email*
- Efficient, Flexible Metadata analysis, management and storage
- Reduce storage and operational costs for Secure, long term retention
- But, also – Global Search and eDiscovery capabilities for retained information
- Enables Active Archive environment
- Compliance guaranteed

- Data is Stored, Managed, Protected and Secure – *Automatically*

- *But this then creates an Information platform for Analytics and application of Business Intelligence*
- *Now seismic data can be managed the same as general Corporate Data*

IBM Business Partner

TARMIN

www.eurotech-computers.com

GridBank and IBM

- IBM have adopted GridBank and have tested and validated
- In the Global ISV catalogue
- Working in partnership with Eurotech and Tarmin in Oil & Gas
- Integrated with IBM Servers and Storage to create solutions that -

  – Enables single view of distributed information
  – Create core data platform
  – Apply IBM BI and analytics tools to centrally managed information
  – Utilise IBM Research and Development in Upstream Oil & Gas
  – Exploit many decades of gathered experience and expertise

- Using advanced analytics and integrated solutions to

  – Reduce timescales
  – Increase production
  – Optimize results
  – Lower costs

Large Scale Smarter Analytics for Chemicals and Petroleum Industry

- Douglas McGarrie, CTO IBM Scotland Enterprise Business Unit
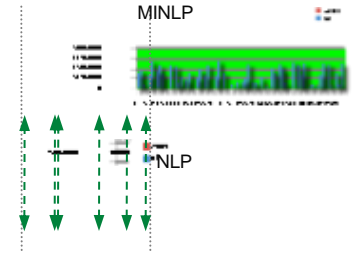
Large Scale for Oil & Gas

The vast array of data sources can be gathered, regardless of their native format, at rest, in motion, and used for Smarter Analytics.
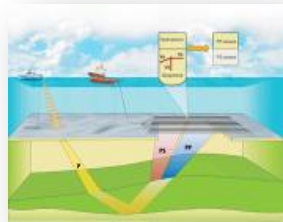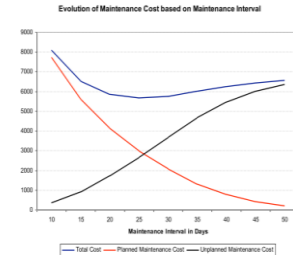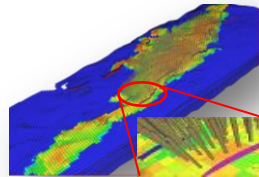
Reservoir Models

Well Files

Lab Results

Seismic Imaging

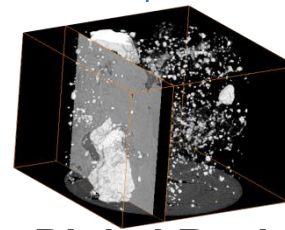**Data Ingestion Sources**

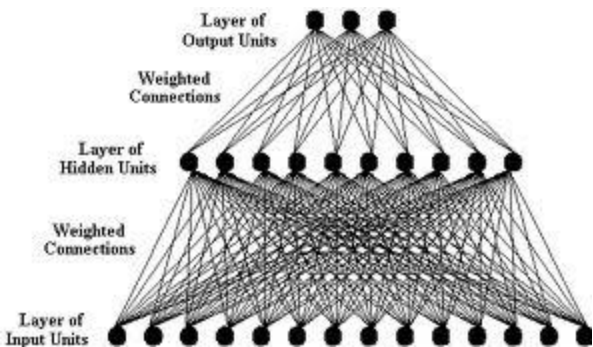**Real-time Well Monitoring**

Well Logs

Historian

Digital Rocks

Daily Drilling Reports

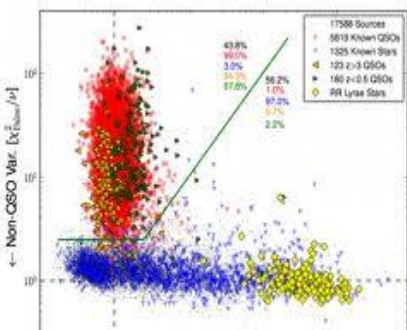# Technology Now Makes it Possible to Analyze ALL Available Data
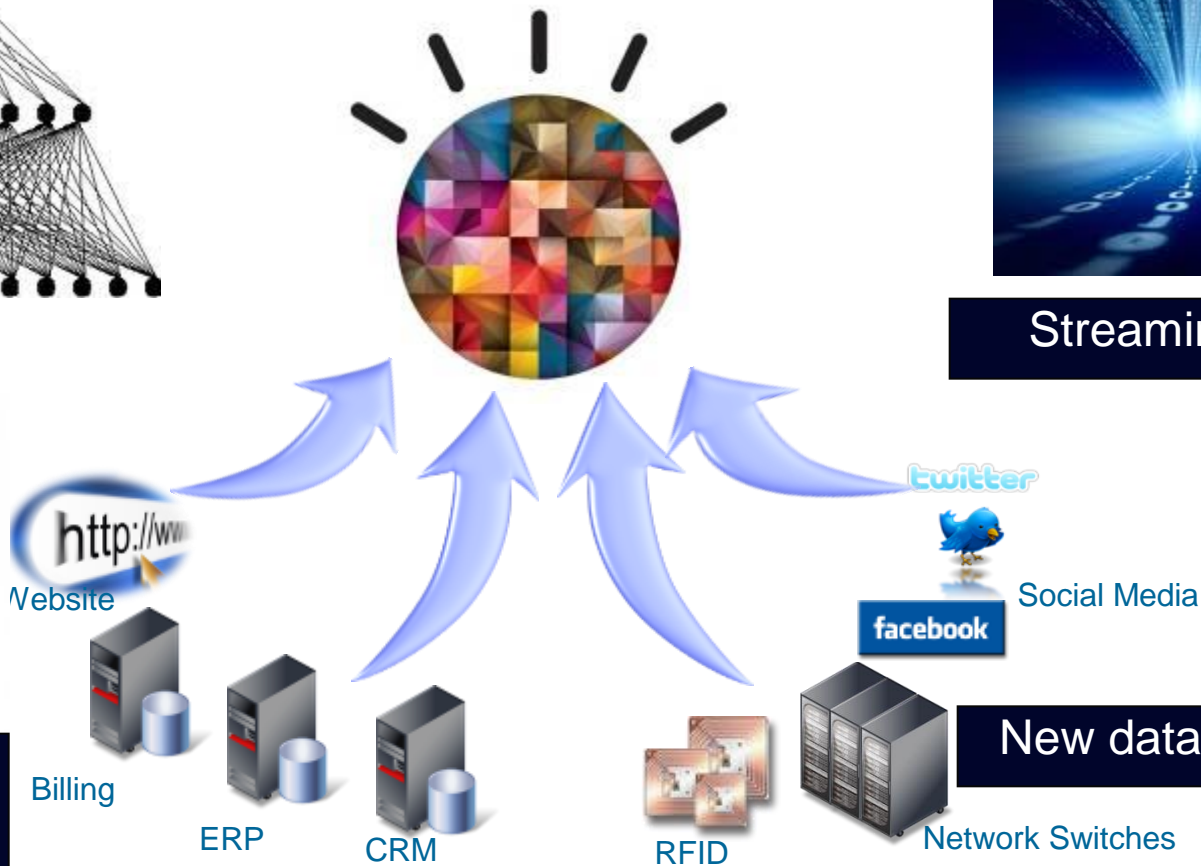
Massively parallel processing

Cost effectively manage and analyze *all available data in its native form*
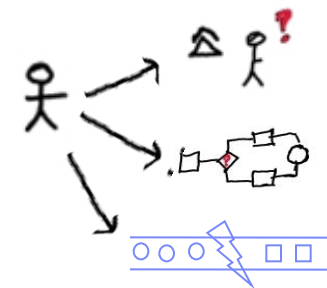
unstructured, structured, streaming

Streaming data

Layer of Output Units
Weighted Connections
Layer of Hidden Units
Weighted Connections
Layer of Input Units

Advanced analytic techniques

Website

Billing

ERP

CRM

RFID

twitter

Social Media

facebook

New data sources

Network Switches

New Insights with Smarter Analytics

What if you could optimize well production yield and lower production cost ?

What if you could discover patterns in operations over time that could point to opportunities for improvements?

What if you could identify the characteristics that tend to increase ownership cost and downtime over the life of a system?

**What if, when an asset is scheduled for maintenance, you could predict what parts are likely to fail in the near future?**

What if you could achieve a more sustained increase in production with a more coordinated effort among monitoring facilities?

What if you could analyze high volumes of data-in-motion for real-time monitoring of environmental conditions?

What if you could quickly mine the thousands of logs that describe the maintenance performed on systems and determine what important observations are being logged by the maintenance team?

IBM
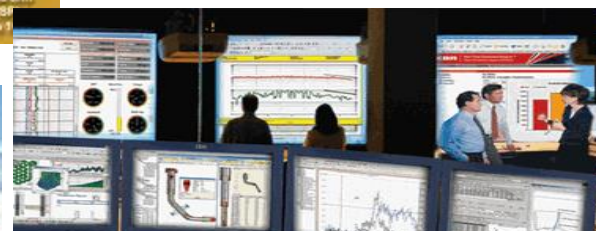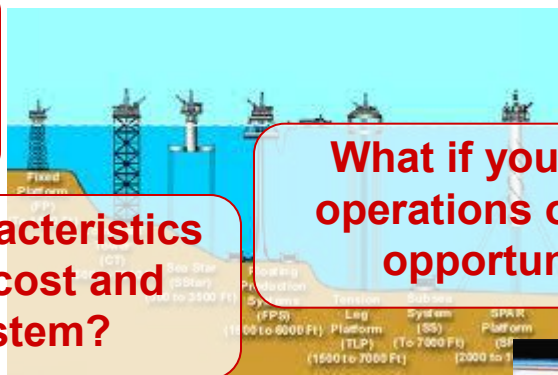
Data-driven insight will help oil and gas companies manage and make more-informed decisions more efficiently, frequently and with greater confidence
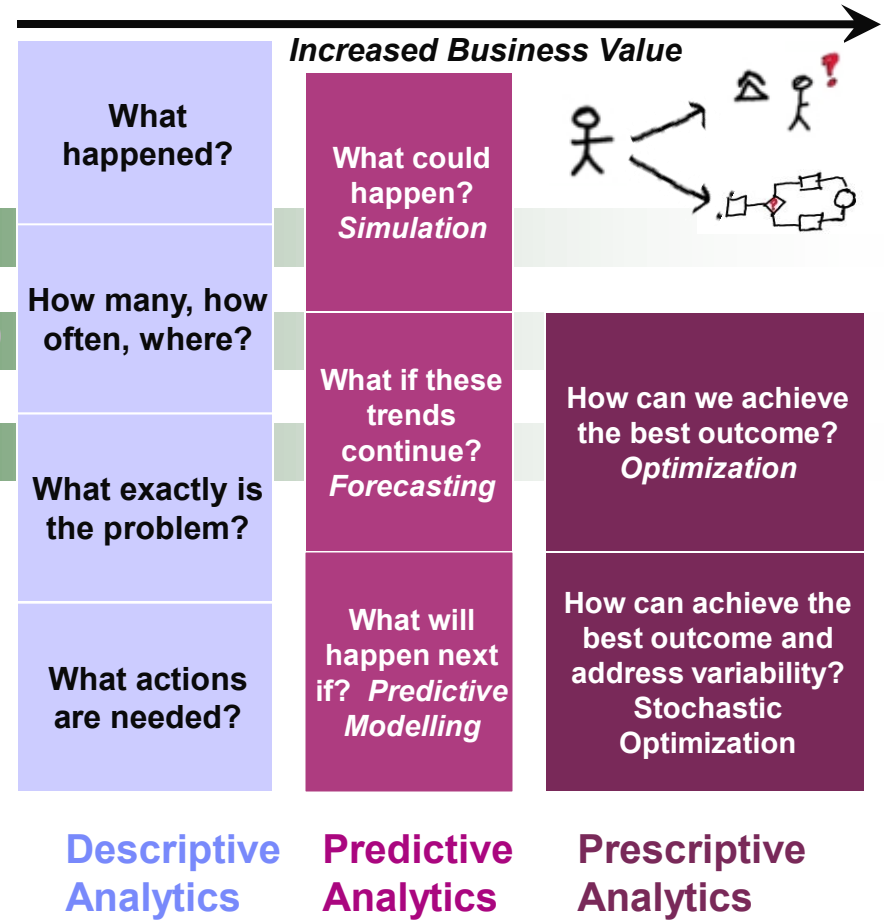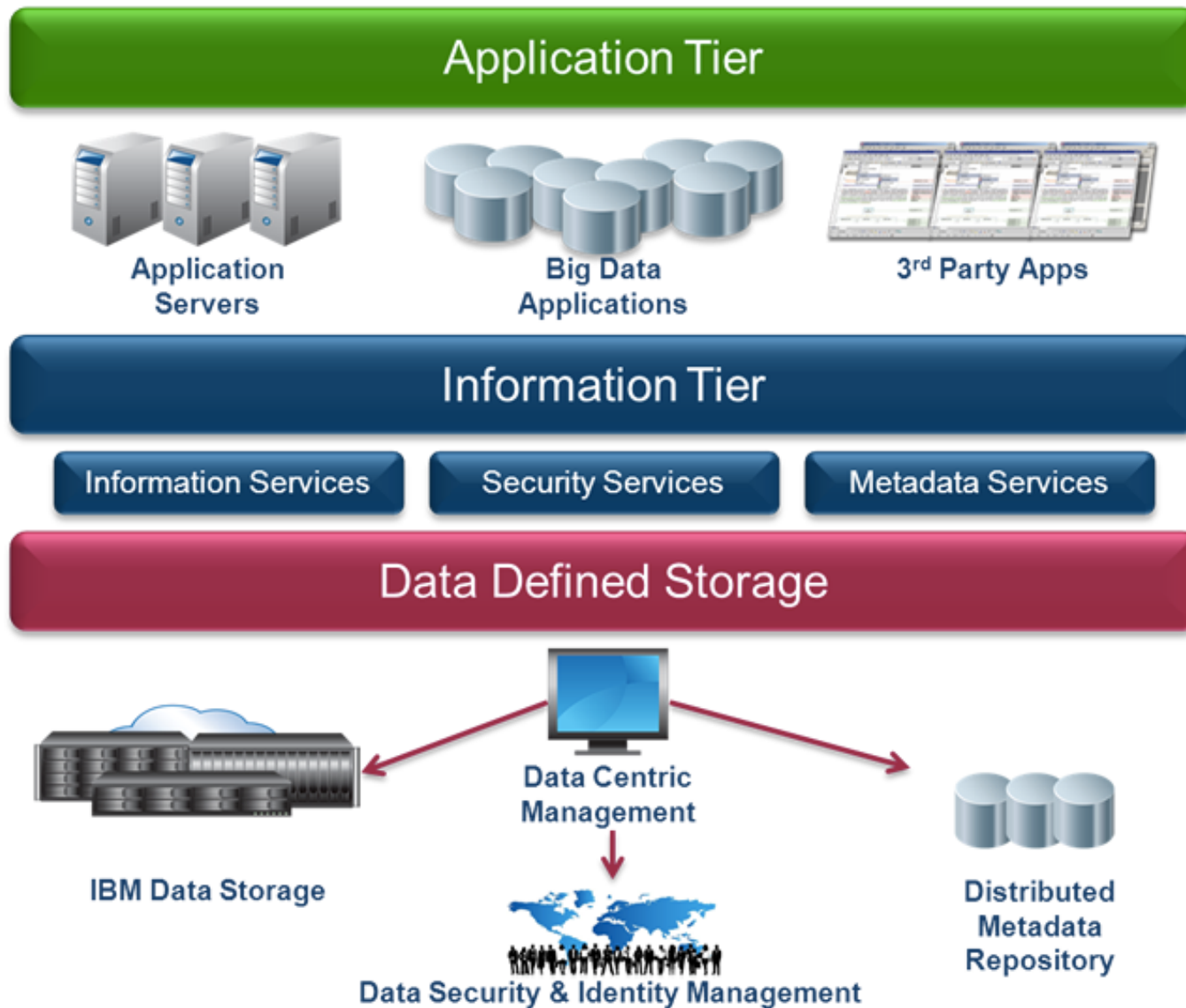
Analytics Sophistication →

Increased Business Value
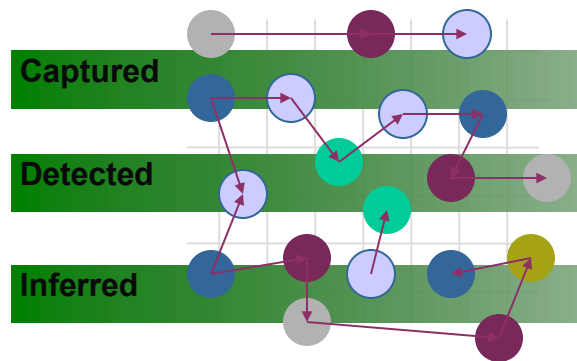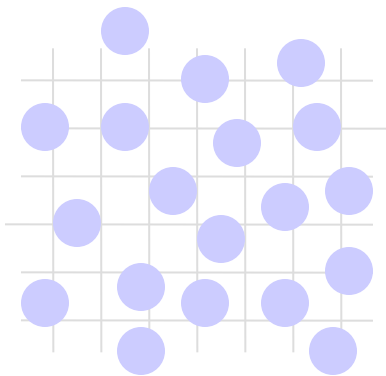
Use Structured Data & Unstructured Data

- **Numeric, Text**
- **Image, Audio, Video**
- **Online, Offline**
- **Customers, partners**
- **Demographic, psychographic, behavioural**
- **Cross channel**

Captured

Detected

Inferred

Made consumable and accessible between digital, subscriber, and partner uses, <u>optimized</u> for their specific purpose, at the point of impact, to deliver <u>better decisions and actions</u> through:

What happened?

How many, how often, where?

What exactly is the problem?

What actions are needed?

What could happen? *Simulation*

What if these trends continue? *Forecasting*

What will happen next if? *Predictive Modelling*

How can we achieve the best outcome? *Optimization*

How can achieve the best outcome and address variability? **Stochastic Optimization**

Descriptive Analytics

Predictive Analytics

Prescriptive Analytics

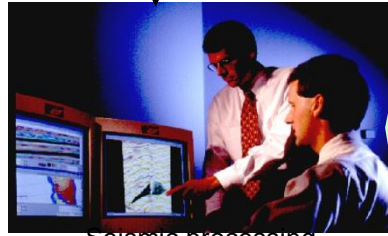# Data Defined Storage: IBM Integration

Seismic Data Lifecycle

Seismic acquisition
and preprocessing

Data transport
to shore

Data transport
to datacenter

Local

Regional

Processing parameter
optimization

Seismic processing
specialists
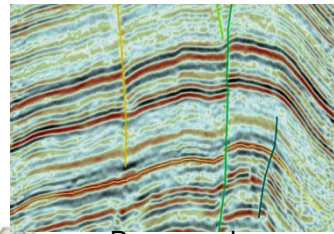
Processed
Seismic data

Interpreted Seismic
Version 1

Interpreted Seismic
Version 2

**Active
storage**

HPC Processing center

Active storage

Tapes

Tape libraries

Petroleum Exploration and Production

Processing and Data Lifecycle
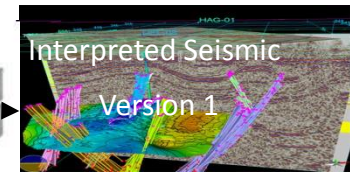

Tape Input

Stages

| Acquisition 3D Seismic Survey | Seismic Processing Pre-processing 2D/3D | 3D Imaging | Seismic Interpretation 3D Visualization | Reservoir Simulation Reservoir Modeling |

Data, Tasks

Seismic Trace Data
Navigation Data

Geological Data

Velocity Data
Other

Sort / filtering
Statics
Velocity analysis

DMO / NMO
Prestack Migrations

Full Wave Migrations
CRAM
RTM

3D Modeling
3D Contours
3D Fault lines
Seismic attributes

Black Oil Compositional

Thermal / Steam Injection
EOR

Tape Archive

Seismic Data Management and Processing Platform

Problem

- Significant volume requirements
  - Petabyte range is common
  - Exponentially high growth
- Many types and formats of seismic data
  - For example: field data, pre-stack/post stack data, intermediate files, raw text, images
- Data needs to be in a state of convenient access and readiness to be interpreted and reinterpreted with HPC computing infrastructure
- Keep Data 'Warm'

Solution

- **New and more versatile platform for storing and processing seismic data**
- An economical, commodity-based storage platform for managing and preparing field, pre-stack, or post-stack data files
- Supports High-Performance Computing (HPC) for processing
- Because the data is organized better, role-based seismic teams can collaborate more easily

- **BigInsights cluster to store data, and support Big Data processing requirements**
- Improve ingest and retrieval through Hadoop indexing
- Combine with MDM for wells, reservoirs, facilities, etc. to correlate seismic data to field data
- Fulfill HPC requirements for complex seismic processing and interpretation calculations, all on the same hardware platform
- Expand seismic-data processing with flexible and expressive tooling for detailed seismic-data analysis

Project: IBM Lossless Seismic Data Compression Package

- There has been a vast increase in the generation of seismic data (e.g. improvements in acquisition modalities)
- Seismic data compression has the following business value
- Reduces archival storage costs for raw, processed data
- Improves seismic workflow performance by the use of on-the-fly compression for static or intermediate data
- Reduces data transport and communication costs

Client Value:

❑ Data compression with no loss of fidelity. Bit-exact data reconstruction.

❑ Significant reduction in storage capacity, hence total cost of ownership of data-center storage.

❑ Can alleviate I/O bottle-necks, thereby reducing computation work-flow time. Reduces compute-center costs.

❑ Light-weight software solution, can be easily integrated into the workflow.

❑ Same solution can be used for compressing raw and intermediate files. Solution works across the wide-range of seismic data file formats and data types which occur in practice.

Results:

❑ Developed for a major ONGC company. Tested extensively on a wide range of seismic data types and formats. The goal was to maximize compression at 10 MBps single-thread throughput. We achieved up to 2.7x compression, significantly out-performing state-of-the-art algorithms. The speed scales with multiple threads.

❑ **There are <u>significant</u> tradeoffs possible in both compression speed and ratio,**

- Software lossless data compression
- No fidelity loss
- Provides state-of-the-art performance
- Specifically targeted to seismic data
- Compresses both raw and intermediate seismic data
- Compresses wide range of seismic data-files and formats
- Significant compression ratio
- Significant throughput, can be scaled up

SoftLayer provides superior technical capabilities critical to a cloud infrastructure, expansive programmable interfaces (APIs) and hundreds of hardware and network configurations

SoftLayer acquisition will help accelerate the delivery of differentiated, high-value platform and application services

- Unified architecture enabled by powerful software
- Customers mix and match bare metal servers, virtual server instances and turnkey private clouds and manage them from a single control pane or API
- All deployed on demand and provisioned automatically in real time
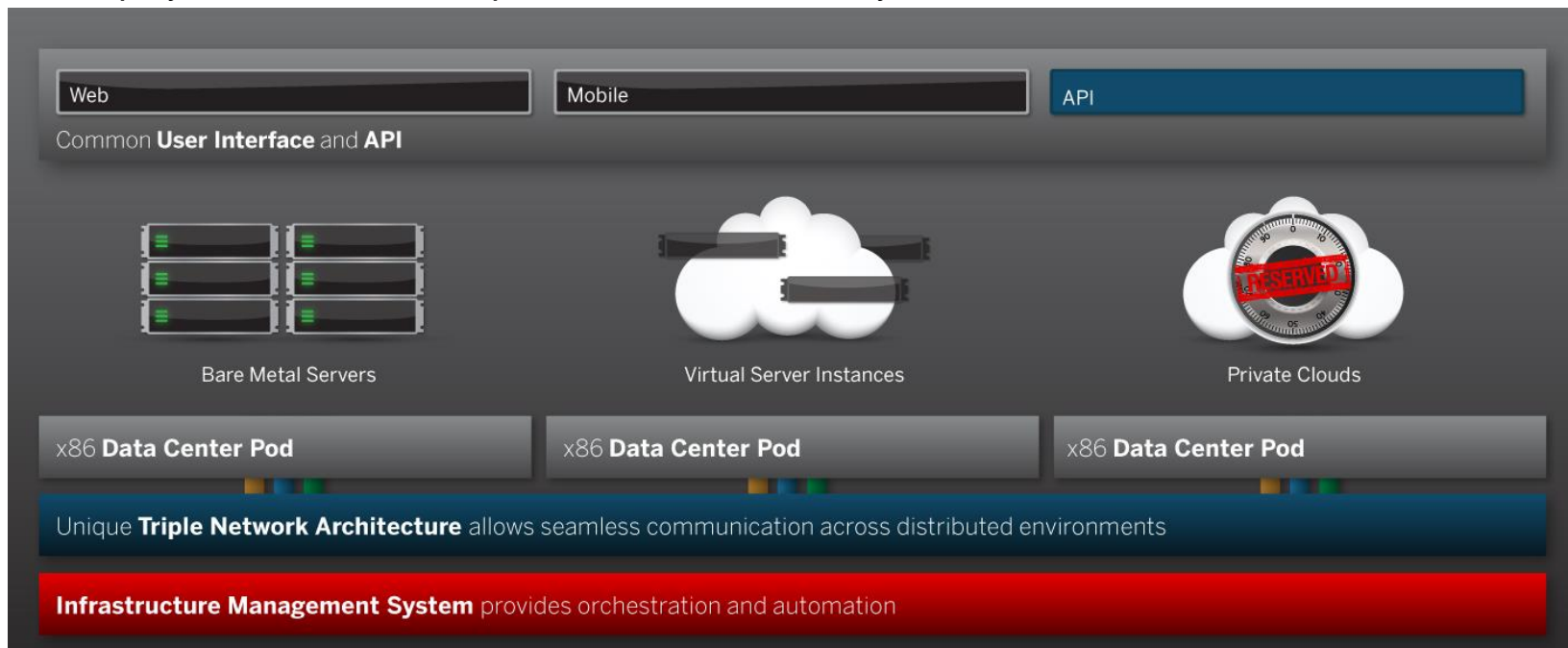
Softlayer - High Performance Computing

The need for High Performance Computing (HPC) capabilities is rapidly crossing all industries and all company sizes.

•SoftLayer offers industry leading HPC cluster technologies required to run HPC applications from entry-level to the most demanding compute intensive workloads.
•The SoftLayer HPC platform delivers performance by offering bare-metal/physical servers, from single to quad processors, and octo-core powerhouses.
•GPU-powered high-performance computing nodes are also available.
•High speed fast SSD storage options for the highest levels of I/O and fast interconnects.

Use Cases	
• Data Mining	• Seismic Analysis
• Numerical Analysis	• Video Processing (encoding/decoding)
• Design Simulation	• 3D Rendering

Typical Components & Considerations	
Component	Considerations & Requirements
HPC Clusters	Map Reduce
Compute Servers	Bare-metal/Physical, Virtual, CPU performance; GPU performance; Large and small memory configurations
Storage	Disk performance; Scalability; Object Storage
Network	Fast interconnects; Topology awareness

Scenario	SoftLayer Solution	Advantage
I need additional power beyond the capacity of today's CPUs.	GPU Acceleration	SoftLayer features NVIDIA Tesla GPUs for accelerated performance.
I need an HPC cluster with multiple server profiles.	HPC clusters provisioned on bare metal and /or virtual servers with specified resources.	HPC clusters provisioned on bare metal, virtual or a combination of servers with a combination resource requirements such as specific memory, storage and architecture requirements.
I need to add additional temporary capacity to my internal HPC cluster.	On-demand Cluster-as-a-service (Bursting) with high speed internet access.	Fully operational rapidly provisioned cluster resources.
I need an environment to run my HPC workloads with predictable performance	Accelerated MapReduce on dedicated storage dense HPC Servers and high performance storage.	High performance, dedicated servers with large storage capacity (up to 36 HDDs).
I need an HPC cluster that can temporarily grow to meet peak demand.	Dynamic and flexible HPC clusters that can grow and shrink.	Cluster footprint grows to meet the needs of an expanding workload, and shrinks as workloads taper to conserve costs.

IBM in the O&G Industry

Sample R&D Projects
Large-scale Basin Modeling
Improved modeling of the complex geological processes

Seismic Ray-tracing
Improved seismic imaging and survey design

Assisted History Matching (AHM)
Nonlinear inversion/optimization in reservoir characterization

Integrated Operations - Turnaround Optimization

Total Field Well Management Optimization
Mixed integer nonlinear optimization

Deep expertise

Dedicated research – focusing on Petroleum Industry challenges

More than 3,000 consultants engaged in Petroleum Industry projects

Innovative solutions: computational geosciences, integrated operations, asset management, logistics, health safety and environment, etc.

Technology Leadership in high performance computing, cloud computing, wireless / RFID, etc.