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*
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

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

Creating a Data Management Platform

- 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
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
GridBank Architecture

Understand
- Search & Discovery
- Big Data Analytics

Control
- Information Governance
- Multi-Site Replication

Store
- Storage Optimisation
- Object Scalability
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
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.
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

Advanced analytic techniques

Website

ERP

CRM

RFID

Billing

New data sources

Streaming data

Social Media

Network Switches
New Insights with Smarter Analytics

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

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 analyze high volumes of data-in-motion for real-time monitoring of environmental conditions?

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

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

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?
Data-driven insight will help oil and gas companies manage and make more-informed decisions more efficiently, frequently and with greater confidence.

Use Structured Data & Unstructured Data
- Numeric, Text
- Image, Audio, Video
- Online, Offline
- Customers, partners
- Demographic, psychographic, behavioural
- Cross channel

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

- Descriptive Analytics
- Predictive Analytics
- Prescriptive Analytics

Analytics Sophistication
- What happened?
- What could happen? Simulation
- How many, how often, where?
- What if these trends continue? Forecasting
- What exactly is the problem?
- What will happen next if? Predictive Modelling
- What actions are needed?
- How can we achieve the best outcome? Optimization
- How can achieve the best outcome and address variability? Stochastic Optimization

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

Increased Business Value

© 2013 IBM Corporation
Seismic Data Lifecycle

Seismic acquisition and preprocessing

Data transport to shore

Data transport to datacenter

Seismic processing specialists

Processed Seismic data

Interpreted Seismic Version 1

Interpreted Seismic Version 2

Active storage

Tape libraries

HPC Processing center

Processing parameter optimization

Local

Regional
Petroleum Exploration and Production

Processing and Data Lifecycle

Stages
- Tape Input
- 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 significant 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
• Numerical Analysis
• Design Simulation
• Seismic Analysis
• Video Processing (encoding/decoding)
• 3D Rendering

Typical Components & Considerations

<table>
<thead>
<tr>
<th>Component</th>
<th>Considerations &amp; Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPC Clusters</td>
<td>Map Reduce</td>
</tr>
<tr>
<td>Compute Servers</td>
<td>Bare-metal/Physical, Virtual, CPU performance; GPU performance; Large and small memory configurations</td>
</tr>
<tr>
<td>Storage</td>
<td>Disk performance; Scalability; Object Storage</td>
</tr>
<tr>
<td>Network</td>
<td>Fast interconnects; Topology awareness</td>
</tr>
</tbody>
</table>

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

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