Adaptive Analysis of Sensor Data

- - Using Stream Computing

John Palfreyman
Contents

- Stream Computing – Why? What?
- InfoSphere Streams - Overview
- Case Study – Marine Life Detection in Galway Bay
- Other Applications
Big (too much) Data?

- **Instrumented Battle Space**
  - Ubiquitous sensors
  - Ground, sea, air, (cyber) space
  - Interconnection

- **Instrumented Supply Chain**
  - RFID tagging
  - Consignment tracking

- **Proactive Data Collection**
  - HUMINT
  - IMINT
  - COMINT
  - ELINT
  - OSINT
Traditional Approaches

(1) Data Mining
- Normalize data (different formats, sources)
- Store in database
- Query Database

(2) Digital Signal Processing
- Fast, dedicated purpose hardware / firmware
- Intercept data stream
- Look for patterns, trends, characteristics
Stream Computing

- Sensor Array
- Filter / Sample
- Transform
- Annotate
- Correlate
- Classify
Contents

- Stream Computing – Why? What?
- InfoSphere Streams - Overview
- Case Study – Marine Life Detection in Galway Bay
- Other Applications
A brief history . . .

2002 2003 2004 2005 2006 2007 2008 2009 2010

Work Started with Federal Government

First System Running

IBM Research begins applying Streams to other areas

 Streams made available to academic community

“Data Baby” and Stockholm Traffic Management

Commercial Work Begins

Full Commercial Release

"The US Government has been working with IBM Research since 2003 on a radical new approach to data analysis that enables high speed, scalable and complex analytics of heterogeneous data streams in motion. The project has been so successful that US Government will deploy additional installations to enable other agencies to achieve greater success in various future projects”

US Government
Components

Development Environment

Runtime

Tools, Adapters & Samples
Runtime

Adapts to changes in resources, workload, data rates

Capable of using generic or specialized hardware
Contents

- Stream Computing – Why? What?
- InfoSphere Streams - Overview
- Case Study – Marine Life Detection in Galway Bay
- Other Applications
Galway Bay Marine Mammal Project

Project Goals

- Identify marine mammals
  - Species
  - Count
  - Distance
  - Individual returning mamals
    (extended goal)

Method

- Analysis of hydrophone data
  - High frequency (500 kHz)
  - Medium resolution (16bit mono)
  - Contain environmental (natural and artificial) noise
Species Identification

- “Click Detection” and “Click Profiling”
- Three stages process
Pre-click Detection

About 0.5s of WAV data

High Pass Filter → Pre-click detector → Fast Fourier Transform (FFT) → Mean Frequency

Porpoise $f=137-144kHz$

Dolphin $f=115-120kHz$
Sound pressure level (signal strength) determined by:

- Distance
- Salinity
- Temperature

Apply filter based on:

- Species “hint” (frequency)
- Sound pressure level

Calculate Sound Pressure Level

Porpoise

f=137-144kHz

Band Pass Filter (175 dB)
Band Pass Filter (161 dB)
Band Pass Filter (151 dB)

Dolphin

f=115-120kHz

Band Pass Filter (230 dB)
Band Pass Filter (216 dB)
Band Pass Filter (210 dB)
Click Profiling & Detection

- Pre-click detection
- Dynamic filtering
- Click profiling & detection

Fast Fourier Transform → Mean Frequency → Band Energy → Peak Position & Width → Click Length → Click Counter

Spectral frequency in click

© 2011 IBM Corporation
Contents

- Stream Computing – Why? What?
- InfoSphere Streams - Overview
- Case Study – Marine Life Detection in Galway Bay
- Other Applications
Cross Industry Examples

- Neonatal Care
- Trading Advantage
- Environment
- Law Enforcement
- Radio Astronomy
- Telecom
- Manufacturing
- Traffic Control
- Fraud Prevention
Cyber Security Analytics

Real-time Results (Tickets, Monitoring) → 1 → Collect Results + Evidence → 2 → Trends, History → 3 → Adapted Analytics Models → 4

Real Time  Offline

DNS  NetFlow

Operating System  Transport System S Data Fabric

Dashboard

Real-time Results

Evidence

Trends, History

Adapted Analytics Models

Analytics
HyperSpectral Camera

- Airborne Platform Mount
- Near Real Time Image Processing
- Improved Resilience / Autonomy

Method

- Break image cube (X, Y, Freq.) into slices
- Parallel processing over machine cluster
- “At least once” transactions manager

Demonstration

- Real time processing of image cube data proven
- Recovery from software & processing node failure shown
- Next Step: sensor system application discussions
Traditional Analysis & Classification

- Hydrophone Array
- Beam Forming
- Bearing / Time
- Detection
- Classification
- Tracking

Digital Signal Processing
- Fast, dedicated purpose hardware / firmware
- Intercept data stream
- Look for patterns, trends, characteristics
Adaptive Analysis & Classification

Stream Computing
- As fast, low latency
- Signal Processing Functions
- Adaptive

Big Insights
- Offline Analysis
- Build Models & Patterns
- Condition Real Time Processing

Hydrophone Array → Beam Forming → Bearing / Time → Detection → Classification → Tracking

History
Spectrum of Defence Applications

Anywhere where Digital Signal Processing is used . . .

- Passive Sonar
  - Analysis
  - Classification

- SIGINT
  - COMINT
  - ELINT

- Radar Processing

- Image Processing

- Track Fusion / Sensor Tasking
Summary

Why?
- Vast Data Volumes, Types, Sources
- Analysis of Data in Motion
- Alternative to Custom Digital Signal Processing

What?
- Development Workbench
- Runtime
- Connectors, Adaptors, Samples

Proof?
- Galway Bay – Marine Life Detection (Passive Sonar)
- Cyber Security – Attack Detection
- Government – Sensitive Projects
John Palfreyman

2dsegma@uk.ibm.com