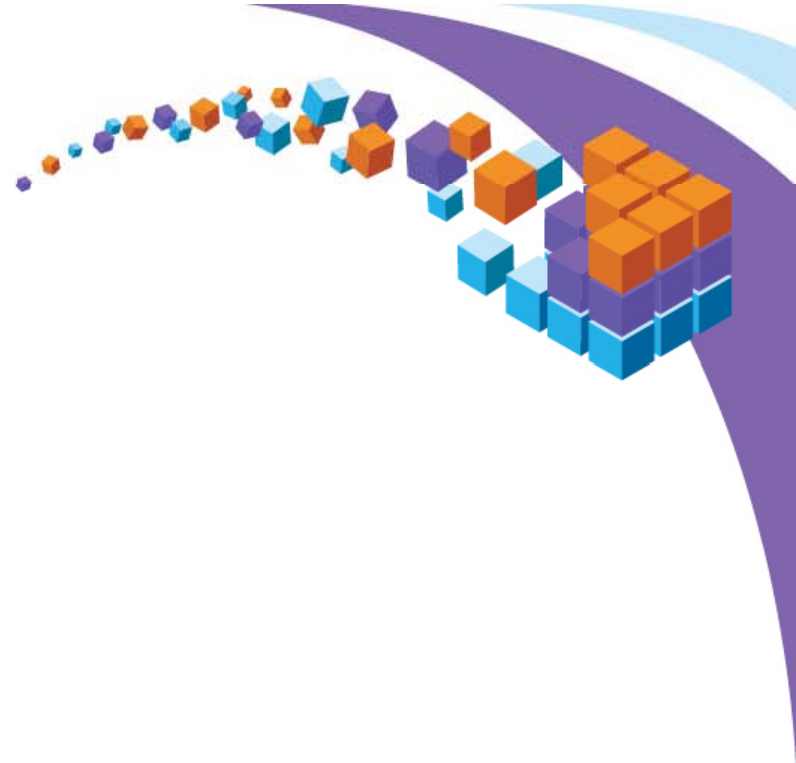# IBM INFORMATION INTEGRATION & GOVERNANCE SYMPOSIUM 2012

Delivering Trusted Information for Smarter Business Decisions

Smarter Decision Making:
New Insight from Big Data

Timothy Young, AP Marketing Executive

Topics

- Big Data is here and now
  - But it's more than just a data scale challenge
- Big Data is an opportunity, not a problem
  - Companies are gaining competitive edge TODAY
- No cookie cutter approach
  - Different and situational approaches are required
- IBM is at the forefront of this trend
  - Technology, innovation, people, partners, clients

"It's life, Jim, but not life as we know it."

twitter: *Follow @ANZ_IM or mention #IIGS*

Citigroup and IBM Watson

Smarter Planet

The world is becoming more instrumented, interconnected and intelligent....

Today, organisations are now at an inflection point, as the importance of analytics shifts.

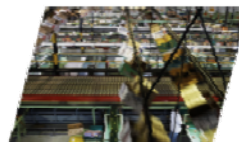Neonatal Care

Public Safety

Telecom

Fraud Prevention

Resource Management

Traffic Control

Supply Chain Visibility

Trading

All the Data, all the Time

Real-time Traffic Flow Optimisation

Fraud & risk detection

Act on customer sentiment

Accurate and timely threat detection

Act on intent to purchase

Prevent churn

twitter: *Follow @ANZ_IM or mention #IIGS*

Sources

12 TB of tweets being created each day.

Every day, NYSE captures 1 TB of trade information.

5 Billion mobile phones in 2010. Only 12% were smartphones.

One second of HD video generates 2,000x as many bytes as required to store a single page of text.

More than 30M networked sensor, growing at a rate >30% per year.

What is your business doing with it?

twitter: *Follow @ANZ_IM or mention #IIGS*

Competing on Analytics

Studies show that organizations using analytics
are outperforming those that are not

220%

Organizations competing on analytics substantially
outperform their peers

twitter: *Follow @ANZ_IM or mention #IIGS*

Little Data Gets Big Quickly

All information
- Social media, emails, chats •
- Transactions • Data warehouses •
- Documents • Sensors • Video •
- Location…

All people
- All departments •
- Experts and non-experts •
- Executives and employees •
- Partners and customers •

At the point of impact

All perspective
- Past – historical, aggregated •
- Present – real-time •
- Future – predictive •

All decisions
- Major and minor •
- Strategic and tactical •
- Routine and exceptions •
- Manual and automated •

Big Data Challenge

Extracting insight from an immense volume, variety and velocity of data, in a timely and cost-effective manner.

Volume: Scale from terabytes to zettabytes

Variety: Manage the complexity of multiple relational and non-relational data types and schemas

Velocity: Streaming data and large volume data movement

twitter: *Follow @ANZ_IM or mention #IIGS*

New Analytics

Analyse a Variety of Information
Novel analytics on a broad set of mixed information that could not be analysed before

Analyse Information in Motion
Streaming data analysis
Large volume data bursts & ad-hoc analysis

Analyse Extreme Volumes of Information
Cost-efficiently process and analyse petabytes of information
Manage & analyze high volumes of structured, relational data

Discover & Experiment
Ad-hoc analytics, data discovery & experimentation
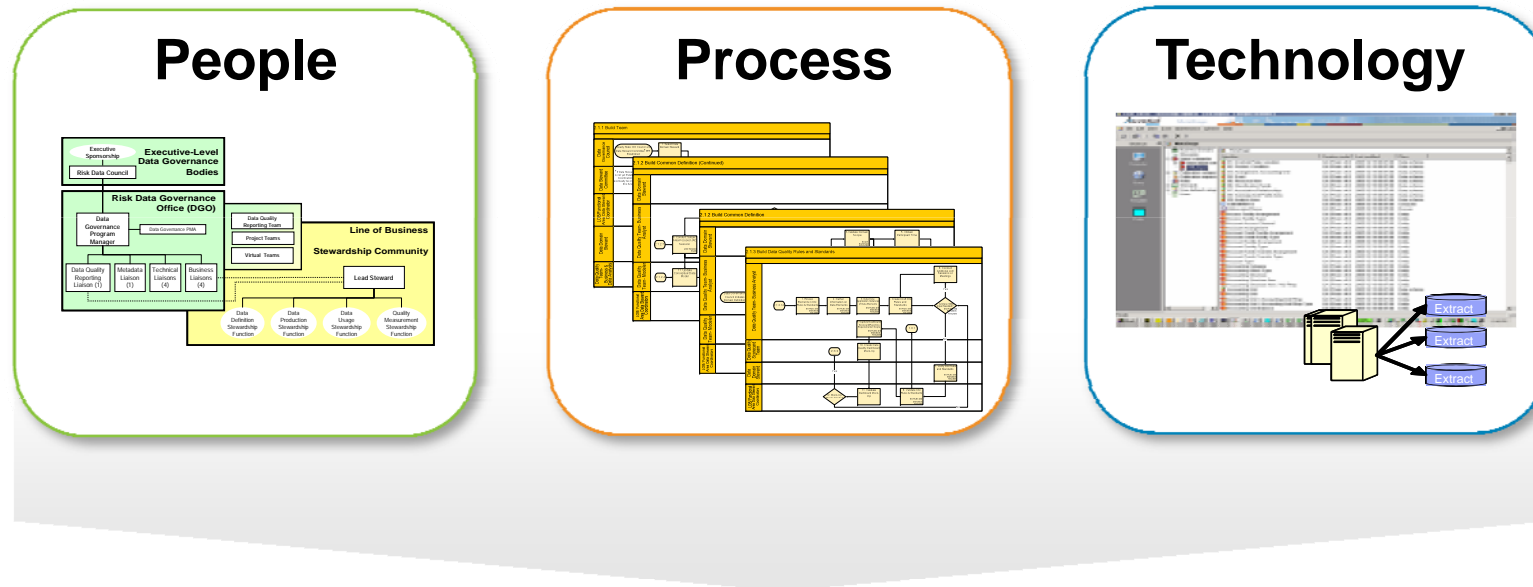
Manage & Plan
Enforce data structure, integrity and control to ensure consistency for repeatable queries

Big Data Enterprise Platform for V³

twitter: Follow @ANZ_IM or mention #IIGS

Information Governance

People

Process

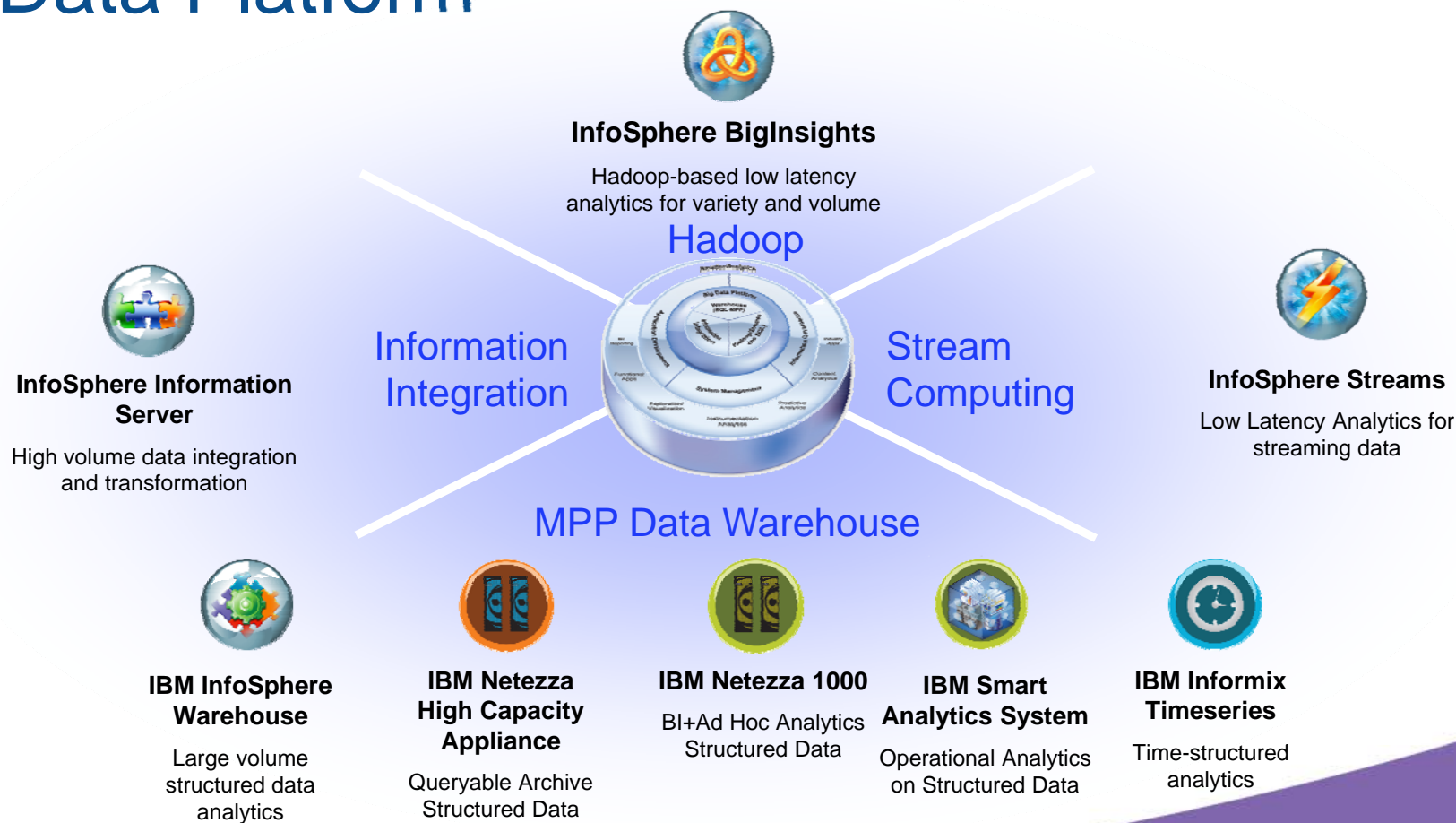**Technology**

Ensure information is understood and consistently defined.

Increase the use and trust of information as an enterprise asset.

Protect information, reduce risk and comply.

Big Data Platform

InfoSphere BigInsights

Hadoop-based low latency
analytics for variety and volume

Hadoop

Information
Integration

Stream
Computing

**InfoSphere Information
Server**

High volume data integration
and transformation

InfoSphere Streams

Low Latency Analytics for
streaming data

MPP Data Warehouse

**IBM InfoSphere
Warehouse**

Large volume
structured data
analytics

**IBM Netezza
High Capacity
Appliance**

Queryable Archive
Structured Data

IBM Netezza 1000

BI+Ad Hoc Analytics
Structured Data

**IBM Smart
Analytics System**

Operational Analytics
on Structured Data

**IBM Informix
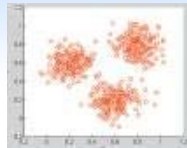Timeseries**

Time-structured
analytics

Analytic Accelerators Designed for Variety
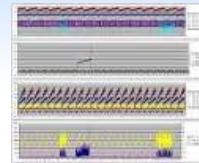
Text
(listen, verb),
(radio, noun)
Simple & Advanced Text
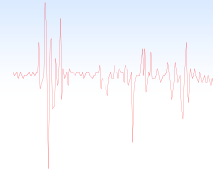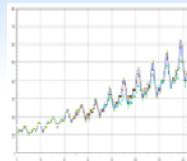
Acoustic

Mining in Microseconds

Advanced Mathematical Models

Predictive

$$\sum_{population} R(s_t, a_t)$$ **Statistics**

GeoSpatial

Image & Video

twitter: *Follow @ANZ_IM or mention #IIGS*

InfoSphere BigInsights

- **Scalable**
  - New nodes can be added on the fly.

- **Affordable**
  - Massively parallel computing on commodity servers

- **Flexible**
  - Hadoop is schema-less, and can absorb any type of data.

- **Fault Tolerant**
  - Through MapReduce software framework

IBM Innovation

- **Performance & reliability**
  - Adaptive MapReduce, Compression, BigIndex, Flexible Scheduler

- **Analytic Accelerators**

- **Productivity Accelerators**
  - Web-based UIs
  - Tools to leverage existing skills
  - End-user visualization

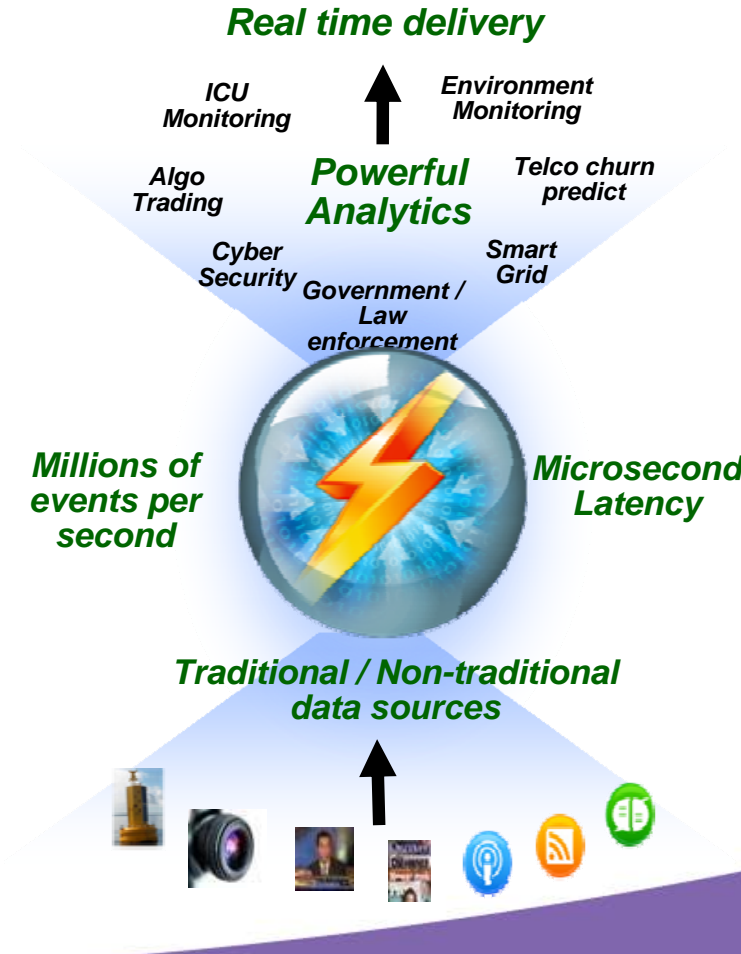- **Enterprise Integration**
  - Extend & enrich information supply chain

- **Security**
  - Role Based / LDAP

InfoSphere Streams

A platform to
run in-motion
analytics
on BIG Data

Real time delivery

ICU
Monitoring

Environment
Monitoring

Algo
Trading

**Powerful
Analytics**

Telco churn
predict

Cyber
Security

Government /
Law
enforcement

Smart
Grid

**Millions of
events per
second**

**Microsecond
Latency**

**Traditional / Non-traditional
data sources**

twitter: *Follow @ANZ_IM or mention #IIGS*

Data in Motion

120 children monitored
120K messages/sec
(ie billions/day)

17
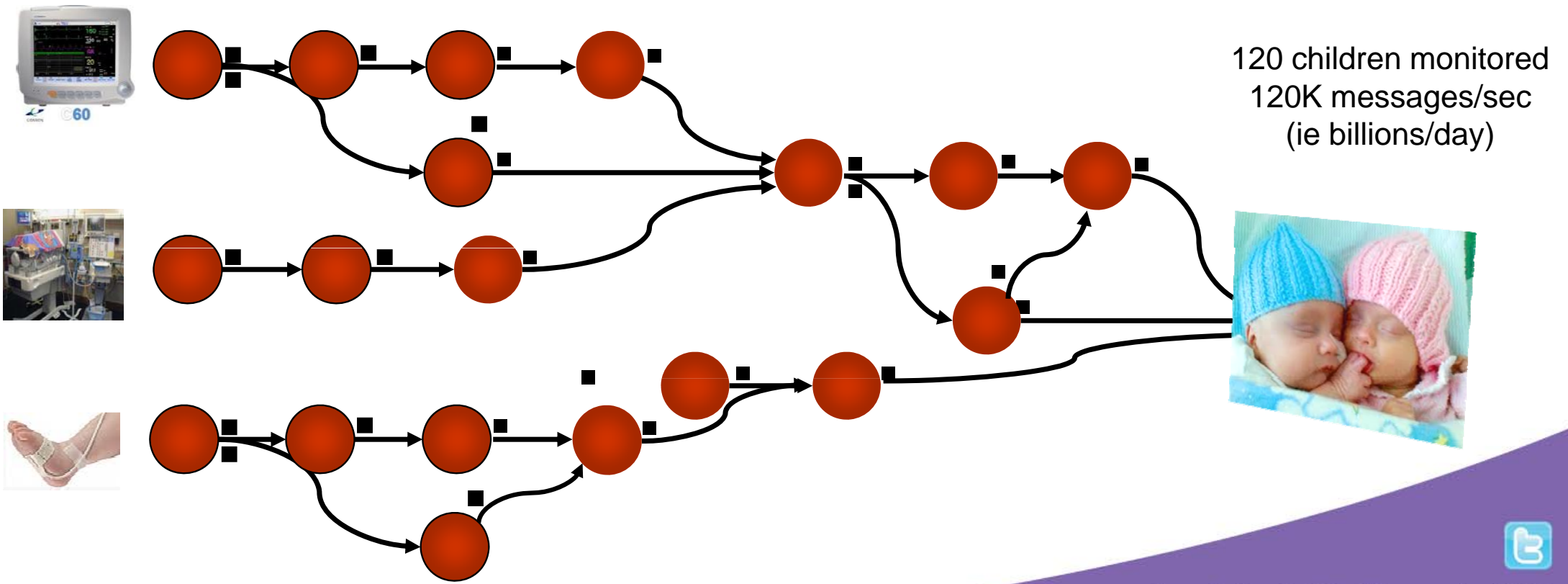
Use Case - Telco

- Reduce billing costs and improve customer satisfaction
- Capabilities
  - Stream computing
  - Analytic Accelerators
- Real-time mediation and analysis of 6B CDRs / day
- Data processing time reduced from 12 hours to 1 second
- Hardware costs reduced to 1/8th
- Proactively address issues
  - Eg dropped calls impacting customer satisfaction (leading to churn)

Use Streams Where

- It would be too expensive to store before analysing

- Data fusion across multiple, disparate streams brings advantage

- True real-time data analysis can provide better business outcomes

- Ability to run multiple analytic models or applications against the same data

Big Data Platform – Data Warehousing

- Workload optimised systems

  – Deep analytics appliance

  – Configurable operational analytics appliance

  – Data warehousing software

- Capabilities

  • Massive parallel processing engine

  • High performance OLAP

  • Mixed operational and analytic workloads

twitter: *Follow @ANZ_IM or mention #IIGS*

Data Warehouse Appliance

- Purpose-built analytics engine

- Integrated database, server & storage

- Standard interfaces

- Low total cost of ownership

- **Speed**: 10-100x faster than traditional systems

- **Simplicity**: Minimal administration and tuning

- **Scalability**: Peta-scale user data capacity

- **Smart**: High-performance advanced analytics

NYSE Euronext

- Peak Data Volumes & Availability
  - 4 billion rows / day
  - 1 TB / day
  - Growing at 100% - 200% / year
  - 7 Year data retention requirement
- Technology Drivers
  - Performance / scalability to meet volume growth
  - Cost reduction
  - Administrative simplification
  - Commercial Innovation
- IBM Netezza 1000 appliance:
  - Market Surveillance, to help identify and investigate trading anomalies
  - Research
  - BI & Adhoc Reporting

twitter: *Follow @ANZ_IM or mention #IIGS*

Use Case – Farming Social Media

Social Media

Online flow: Data-in-motion analysis

Timely Decisions

Stream Computing and Analytics

Data Ingest and Prep → Text Analytics: Timely Insights → Entity Analytics: Profile Resolution → Predictive Analytics: Action Determination

Dashboard

Data Warehouse / Hadoop and Analytics

Social Media Data → Text Analytics → Entity Analytics and Integration → Comprehensive Social Media Customer Profiles → Predictive Analytics → Customer Models

Reports

Offline flow: Data-at-rest analysis

twitter: *Follow @ANZ_IM or mention #IIGS*

23

IBM and Big Data

1. ***Big Data in Real-Time***

- Ingest, analyse and act on massive volumes of streaming data
- Faster AND more cost-effective for specific use cases
- 10x volume of data on the same hardware

2. ***Fit for purpose analytics***

- Analyses a variety of data types, in their native format – video, audio, text, time series, geospatial & more

3. ***Enterprise Class***

- Open source enhanced for the enterprise – reliability, performance and security
- High performance warehouse software and appliances
- Ease of use with end users, admin and development UIs

4. ***Integration***

- Integration into your information management architecture
- Pre-integrated analytic applications

twitter: *Follow @ANZ_IM or mention #IIGS*

Summary

- Big Data is upon us
- Seize the day – get competitive edge
- Be opportunity driven
- IBM has always been at the forefront of this trend
  - Technology, innovation, people, partners, use-cases

twitter: *Follow @ANZ_IM or mention #IIGS*