Making Sense of Big Data in Insurance

Amir Halfon, CTO, Financial Services, MarkLogic Corporation
“BIG” DATA?..
The Evolution of Data Management

Mainframe Era

- Application- and hardware-specific

Relational Era

- Normalized, tabular model
- Application-independent query
- User control

Modern Era

- Schema-agnostic
- Massive scale
- Query and search
- Analytics
- Application services
- Faster time-to-results

For your application data!
- Application- and hardware-specific

“For your structured data!”
- Normalized, tabular model
- Application-independent query
- User control

“For all your data!”
- Schema-agnostic
- Massive scale
- Query and search
- Analytics
- Application services
- Faster time-to-results
Fraud Detection and Prevention

Beyond algorithmic transaction analysis

- From claim-centric to person-centric - beneficiary behavior across claims
- Data sources beyond the firewall
- Networks of people rather than individuals

Information across all parties involved in a claim

- Counter-parties as well as partners (e.g. auto repair shops)
- Notes; images; sound files…

Integrated event processing

- Alerting; case management
Customer Insight

Beyond policy administration
- Regulatory drivers
- Increase customer retention and satisfaction
- Personalized insurance

Single customer view
- Across multiple channels and LoBs
- On-boarding docs; customer care logs; clickstreams; telemetry; geospatial…
Claims Management

Beyond structured data
- Supporting notations, images and video
- Hierarchical ACORD standard formats
- Mainframe modernization

Integrated workflow and information
- Event driven architecture
- Faster claim processing
- Innovative interaction models
Underwriting – Commercial and Reinsurance

Beyond re-entering data

- Supporting information at quotation (loss histories, property schedules, etc.)
  - Diverse file formats, rarely standard
  - Typically assessed manually by the underwriter
  - ..then re-examined if a claim is subsequently made
  - Time consuming, inefficient
  - Information remains locked in the documents, with little or no analytical use.

Data needs to be searchable and discoverable
THE TECHNICAL ANGLE...
Last Generation

- OLTP
- Warehouse
- Archives
- Data Marts
- Reference Data

“Unstructured”

- Documents, Messages
- Video
- Audio
- Signals, Logs, Streams
- Social
- Search

Metadata

{ }
Hadoop
Hadoop’s Limitations

- Hadoop was designed for batch processing
  - Does not support real-time applications on its own
  - Limited support for SQL-based analytics
- Requires expertise to configure, deploy and manage
- Until recently has not supported mainstream security authentication technologies
- You still need a database
ACID IS INDISPENSABLE
ENTERPRISE HARDENED

RDBMS: THE GRAND DADDY OF DATABASES

30 YEARS OF SUCCESS

TODAY'S DATA DOESN'T FIT

DATA MODELING IS SLOW

SCALING IS SLOW & EXPENSIVE
Hadoop + Relational Database

- Split 
  \[ [k1, v1] \]
- Sort by \( k1 \)
- **Merge**
  \[ [k1, [v1, v2, v3 ...]] \]

Output Data
NOSQL: THE YOUNG UPSTART

- **Schemaless**
- **Fast & Agile**
- **Runs on Scale-Out Clusters**

- **Data is more than keys & values**
- **Business requires ACID transactions**
- **IT requires HA/DR/Security**
ENTERPRISE NOSQL: THE BEST OF BOTH

12 YEARS OF SUCCESS

ACID IS INDISPENSABLE

ENTERPRISE HARDENED

SCHEMA AGNOSTIC

FAST, AGILE & POWERFUL

RUNS ON SCALE-OUT CLUSTERS
Enterprise NoSQL

- Schema-agnostic design
- Real-time indexing for search and query
- Reverse queries for event processing and alerting
- Scale-out shared-nothing cluster topology
- Analytics and Visualization
- Transactions
- Hadoop Integration
Database + Search

- Interactive access to unstructured text
- Best if integrated at the DB layer:
  - Quicker time to information
  - Greater than the sum of its parts
  - No separate indexes to maintain
  - One less friction point
- Search as a text processing engine
  - Turn text into numbers, create new dimensions
  - Infer new information from text search and enrich
New Generation Data Management

- Load and index data “as is” from any sources
- Schema-agnostic, poly-structure data, Full text search
- Deliver data to each consumer in the right format
- Security
- Compliance
- Fraud Detection
- Risk Management
- etc.

Transactions

Interactions

Connections

etc.

Hadoop File System integration
Example: Fraud and Financial Crime Prevention

Profile Configuration Tools

Profile Data Extracted From multiple sources

Profile includes social graphs

Analytics

Decision Support

Example: Fraud and Financial Crime Prevention
Final Thoughts

- One way to do more with less is to do less..
  - ETL, schema design, database maintenance
  - .. to spend less
  - on complex infrastructure and manual processes
- … and focus more
  - on harnessing new generation technology to maximize agility
THANK YOU