LexisNexis Active Insights: Evolving the Insurance Industry with the HPCC Systems Open Source Big Data Platform
LexisNexis Active Insights in a Nutshell

• 10,000 data sources updated regularly
• Generate scores, attributes and insights
• Match events to policies in the monitored book of business as data updates
• Proactively notify customers
LexisNexis Active Insights Enables a Highly Proactive Approach to Customer Engagement and Retention

- Increase retention
- Identify opportunities
- Improve loss ratio
- Reduce expenses

When life changes, so do insurance needs:
- New drivers in the home
- First-time homeowner
- New homeowner
- Change in homeownership status
- Marital status changes
- Shopping for a new insurance policy
- Home renovations

*Other events may indicate potential coverage changes too:*
- Foreclosures and vacancies
- New accidents
- Home inspection index score changes
- Moving violations
How do we do it?
The Data Centric Approach

A single source of data is insufficient to overcome inaccuracies in the data. The holes are inaccuracies found in the data.

Our platform is built on the premise of absorbing data from multiple data sources and transforming them to a highly intelligent social network graphs that can be processed to non-obvious relationships.

The holes in the core data have been eliminated.
Data Flow Oriented Big Data Platform

Thor (Data Lake)
- Shared Nothing MPP Architecture
- Commodity Hardware
- Batch ETL and Analytics

ROXIE (Query)
- Shared Nothing MPP Architecture
- Commodity Hardware
- Real-time Indexed Based Query
- Low Latency, Highly Concurrent and Highly Redundant

Raw data from several sources
Batch requests for scoring and analytics

ECL
- Easy to use
- Implicitly Parallel
- Compiles to C++

ESP Middleware Services
Portal
Batch Subscribers
Batch Processed Data
STRIKE Technology Layer View

- **Data Science Portal**
  - Dashboard Creator
  - Workflow Builder
- **Analytics Tools**
  - SALT
    - Cleaning
    - Profiling
    - Normalization
    - MDM
  - KEL
    - Relationship Analysis
    - Attribute Creation
- **Common Programming Language**
  - Predictive Analysis
  - ECL
  - Business Intelligence
- **Data Connect**
  - Interlok
  - Thor
  - ROXIE
Master Data Management with SALT

From disparate data, to clustering, to showing relationships
SALT Enables Content Disambiguation to Increase Productivity

- The acronym stands for “Scalable Automated Linking Technology”
- Entity disambiguation using Inference Techniques
- Templates based ECL code generator
- Provides for automated data profiling, parsing, cleansing, normalization and standardization
- Sophisticated specificity and relatives based linking and clustering

Data Sources

1. Data Preparation Processes (ETL)
   - Profiling
   - Parsing
   - Cleansing
   - Normalization
   - Standardization

2. Record Linkage Process
   - Matching Weights & Threshold Computation
   - Blocking/Searching
   - Weight Assignment & Record Comparison
   - Record Match Decision
   - Linked Data File

42 Lines of SALT
3,980 Lines of ECL
482,410 Lines of C++
SALT’s Superior Linking Technology

SALT eliminates **FALSE NEGATIVES** using probabilistic learning

1. Flavio Villanustre, Atlanta
2. Javio Villanustre, Atlanta

**SALT**

**MATCH** — the system has learnt that “Villanustre” is specific because the frequency of occurrence is small and there is only one present in Atlanta

**ERROR**

**NO MATCH** — because the rules determine that “Flavio” and “Javio” are not the same

SALT eliminates **FALSE POSITIVES** using probabilistic learning

1. John Smith, Atlanta
2. John Smith, Atlanta

**SALT**

**MATCH** — the system has learnt that “John Smith” is not specific because the frequency of occurrence is large and there are many present in Atlanta

**ERROR**

**NO MATCH** — because the rules determine that “John Smith” and the city for both the records match
Relationship Analysis With KEL

KEL — an abstraction for network/graph processing

• Declarative model: describe what things are, rather than how to execute

• High level: vertices and edges are first class citizens

• A single model to describe graphs and queries

• Leverages Thor for heavy lifting and ROXIE for real-time analytics

• Compiles into ECL (and ECL compiles into C++, which compiles into assembler)
LexisNexis Legal & Professional

**THE CHALLENGE**

- Fast insight into case law
- 100+ million documents
- Entity identification and resolution
- Document and topic classification
- Near real-time feedback
THE SOLUTION

- Generation 2 entity recognition employs HPCC PARSE(...) function and pattern rules
  - More entities recognized
  - Faster development
  - Faster operation
- SALT based entity resolution
- Custom resolution for citation entities
  - Case law and statute reference
  - References can be anaphoric (like *infra*)
  - Parallels (same case in more than one book)
- Active learning used to extend classification

THE OUTCOME:

Two enormous benefits:

- Huge lift on entity resolution and document classification because of SALT
- Ahead of customers in terms of performance and maintaining currency of data because of the rapid big data processing capabilities
Smart Hard Hat Ecosystem

THE CHALLENGE

4,000 workers die and millions injured annually while working on the industrial floor

Very high cost for maintaining safety for businesses
Smart Hard Hat Ecosystem

THE SOLUTION

• Equip workers’ hats with smart sensor technology
• Central real-time processing of (high volume) information with real-time alerting capability (HPCC Systems)
• Customizable dashboards, rules framework and data workflow frameworks (HPCC Systems)
• Predictive modeling and analytics (HPCC Systems)

THE OUTCOME: 
Produced an industrial wearable that uses IoT and wireless communications systems to protect and empower industrial workers.
Driver Behavior with Smart Telematics

THE CHALLENGE

- High cost of insurance
- High car accident rates
- Lack of tools to analyze driver behavior
Driver Behavior with Smart Telematics

THE SOLUTION

• Telematics smart phone application
• Central system to collect (very large) data and perform analytics (HPCC Systems)
• Journey based feedback to all drivers to advice and correct behavior (HPCC Systems)
• Insurance enrollment to reduce premiums

THE OUTCOME:

✓ Recommend corrections to driver behavior that would avoid accidents
✓ Reduce overall Insurance costs
✓ Correlate information from drivers data traversing the same path to create an understanding of predictable actions
✓ Examples include periods of traffic congestion, problem areas in the path and hazard detection
Contextual Marketing

Understanding an individual customer’s behavior based on past actions

Technical problem

• Huge volumes of data based on observed cell phone Wi-Fi
• Apply advanced machine learning techniques

THE CHALLENGE
Contextual Marketing

THE SOLUTION

• Central analytics system to collect and analyze data (HPCC Systems)
• Leverage parallel algorithms to perform analytics on large quantities of data (HPCC Systems)

THE OUTCOME:
Created a platform to process any location specific telecom data that can be analyzed rapidly to gauge consumer behavior and in turn help drive context-based marketing
Predict Passenger Volumes in Airports

THE CHALLENGE

- How to interpret 100’s millions of location points while adjusting to flight schedule changes
- Complex clustering algorithm requirements
- Understand passenger behaviors and interaction of local areas of activity
Predict Passenger Volumes in Airports

THE SOLUTION

HPCC used to solve Big Data challenges

• Raw data to refined data
• Clustering analysis
• Forecasting

THE OUTCOME:

Better passenger experience and better airport planning
Resources

- Portal: http://hpccsystems.com
- SALT: https://hpccsystems.com/enterprise-services/purchase-required-modules/SALT
- KEL: https://hpccsystems.com/download/free-modules/kel-lite
- Machine Learning: http://hpccsystems.com/ml
- Online Training: http://learn.lexisnexis.com/hpcc
- HPCC Systems Blog: http://hpccsystems.com/blog
- Our GitHub portal: https://github.com/hpcc-systems
- Community Forums: http://hpccsystems.com/bb
- Case Studies: https://hpccsystems.com/resources/case-studies