## **Psydex Platform**





Psydex Corporation, Copyright 2010 - All Rights Reserved - Contains Confidential and Proprietary Information







Psydex Corporation, Copyright 2010 - All Rights Reserved - Contains Confidential and Proprietary Information





THE R. P.

### Query-Fusion across Petabytes of Streaming and Historical Data



# Semantic Query Service ("SQS") Instance





## **Technology Highlights (Advantages)**

#### Performance and Scalability

•Shared-Nothing MPP (Massively Parallel Processing) Architecture. All In-Memory Indexes •Seamless *Real-time Ad Hoc* Query Across Both Streaming and Historical Data (Terabytes)

#### Adaptive

Semantic Models are Independent of Indexed Data and executed at Run Time
Language Independent. Lexicons can Evolve and Change in *Real Time*Models are Dynamically Applied to Content as Opposed to *"Extracted"*

#### Power and Precision

High Precision Rules for Modeling People, Companies, Sentiment, Events etc...
Powerful Operators (Sets, AND, OR, Proximity, Negation, Lexical Expansion)

Analytics

•Ad Hoc Semantic Statistics ,Time Series, and Data Correlations in *Real Time* 

#### Collection/Ingest

•Standards Based (RSS, NEWSML, NITF, RDF/RFA, JDBC etc.)

•High Speed Wires Direct (AP, BW, DJ, MW, PR, Thomson Reuters)

Packet Capture and Other Streaming Content(TV Caption Recovery)

•Social Media Integration (Twitter, Facebook etc..)

•Other Collectors (Email, AOL IM, SMS, Network Packet Capture)

#### APIs and Feeds

•Full REST API and Custom TCP Feeds



# How Unstructured Data is Handled

- Feed Handlers Collection/Ingest fastest possible live data feeds
   (n) Semantic Query Services ("SQS") receive new data
- Distributed Source Indexing and Querying
  - SQS span (n) processes/nodes
  - Partitioned by source and time period
  - Historical data is indexed when service is initialized
  - Real time data is indexed by SQS on arrival
- Semantic Indexing
  - Tokenization of all characters, words, phrases
  - Tokens are mapped to discrete time slots
  - All meta is indexed as word tokens
  - Only token positions, offsets and times maintained.
- Semantic Topic Models (rules) based on words and phrases
  - Ordinal position, relative and temporal proximity
  - AND, OR, Negation, Sets, Proximity, Expansion, Substitution
  - Models can change independent of content

# PQL ("Psydex Query Language")

() Sets - Phrases separated by commas where the comma means "OR" ex. (Apple Computer, iPhone, ipod, ipod, imac, itablet)

- **\*** "AND" Co-Reference within 1024 word tokens
   ex. (IBM, big blue, International Business Machines) + (acquired, is acquiring,, will acquire)
- {n} Proximity Un-Ordered (Anything in First Set within 5 tokens of Second Set)
   ex. (IBM, big blue, International Business Machines) {5} (acquired, is acquiring, will acquire)
- [n] Proximity Ordered (Anything in First Set within 5 tokens of Second Set) ex. (IBM, big blue, International Business Machines) [5] (acquired, is acquiring)
- Negation
  - ex. (Iraq-nuclear)
- ~ Semantic Expansion

ex. (Economic, Economy) {10} ( ~Worry)

**\$** Nested Topic

ex. (\$IBM) {10} ( \$ACQUISITION)

>?<Extraction

ex. (announced, announcement ,announcing) >< (is acquiring ,acquired, is buying) *Status : Testing....Targeted 3/1/2010* 



## **Partner and Customer Integration**

\*List of Partners Provided Upon request





Psydex Corporation, Copyright 2010 - All Rights Reserved - Contains Confidential and Proprietary Information