Object and Graph Databases

Leon Guzenda - Objectivity, Inc.
OBJECT DATABASE
INDUSTRY
The ODBMS Players

The Object-Oriented Database System Manifesto


Matisse

VBase (Ontologic) → Ontos

GemStone (Servio Logic → Servio → GemStone)

Objectivity/DB

ObjectStore (Object Design → Excelon → Progress)

GBase (Graphael)

Versant Object Database (Versant)

O2 (→ Ardent → Ascential)

UniSQL

Poet → FastObjects

Note that many of these companies were founded earlier, e.g. Objectivity, Inc. was founded in June 1988.

Berkeley DB (Sleepycat → Oracle)

db4objects

InfiniteGraph

© Objectivity, Inc. 2010

Leon Guzenda at ICOODB 2010
Objectivity, Inc.

• A privately held, profitable company founded in 1988 to provide object database software for engineering, scientific and defense applications.

• Defence and Intelligence Community customers include: US DoD/DoE/DHS, BaE, Boeing, General Dynamics, L-3, Lockheed Martin, NGC, Raytheon, Titan...

• Equipment manufacturing customers include: Alcatel, Ciena, Emerson, Ericsson, Iridium, NEC, Qualcomm, Siemens...

• “Big Science” customers include Berkeley NL, Brookhaven NL, CERN, Lawrence Livermore NL, FermiLab, JHU Sloan Digital Sky Survey, Los Alamos NL, Max Planck...
Technology Partnerships Are Key

• High Performance Communications
  – Brocade, Arista Networks, Bivio Networks...

• High Performance Processing
  – Appro, Bull, Cray, Convey Computer, HP, IBM, Inspur, Tilera...

• High Capacity Memory Systems
  – Violin Memory, ioFusion, TMS...

• Massively Scaleable File and Storage Systems
  – DDN, EMC, Hyperion, Kaminario, Panasas, whamcloud, Xyratex...

• Visualization
  – Activu, Tom Sawyer, SL Graphics,
TYPICAL DEPLOYMENTS
Typical Deployments...

Objectivity/DB reduces the time it takes to accurately locate foes by more than 90%.

SIGINT

Uses network-centric techniques to horizontally integrate multiple ISR assets, providing machine-to-machine (M2M) interaction of multi-INT sensors, to create actionable information on time-sensitive targets.

Multi-INT with Imagery, GMTI, and SIGINT
Objectivity/DB is used to aggregate knowledge and disseminate it back into the field.

HUMINT

Who They Are
Where They Are
How Well They're Doing

Legacy applications

What They're Doing

When They Operate
...Typical Deployments...

Astronomy
...Typical Deployments

Zeus Anesthesia  Base Station Controller  Iridium LEO Network

Process Control  High Energy Physics
**ODBMS Evolution**

- **1980s**
  - “Performance, Performance, Performance!”
  - Primarily scientific and engineering applications

- **1990s**
  - Reliability and Scalability
  - New languages and Operating Systems
  - Large deployments in the scientific domain

- **2000s**
  - Ease of use and instrumentation
  - Query languages
  - Performance and scalability
  - Grids and Clouds
  - Embedded systems, government and more...

---

**DATA MANIPULATION**

Applications tended to generate data and relationships

**RELATIONSHIP ANALYTICS**

Applications ingest and correlate data and relationships
CURRENT CHALLENGES
Current Challenges

• Faster, more robust ISR* networks deliver more data than ever before.
• There is an ever increasing range of sensor types and platforms.
• Fusing the data for immediate and analytic use and dissemination is becoming a huge challenge.
• Conventional hardware and software approaches are being overwhelmed.

* ISR = Intelligence, Surveillance and Reconnaissance
High Volume Data Ingest...

Sources → Ingest → Correlation & Analytics → Visual Analytics → Actions

- Sources: Various data sources such as drones, remote sensing equipment, and satellite imagery.
- Ingest: The process of importing and organizing data from sources.
- Correlation & Analytics: Analysis of data for pattern recognition and decision-making.
- Visual Analytics: Graphical representation of data for easier interpretation.
- Actions: Application of insights to take physical actions.

---

© Objectivity, Inc. 2011
...High Volume Data Ingest

• 2000
  – System handles 1 billion objects per day
  – Correlates them with existing knowledge
  – Raises alerts when new information is found
  – Services about 100 analysts at one agency

• 2010
  – Multi-petabyte system handles 10s of billions of objects per day
  – Services 1000+ analysts at five agencies

• 2012
  – Each sensor platform will collect 200 Terabytes per day...
Fusion & Metadata Indexing

• Disparate streams of data must be correlated
  – Geospatial and temporal properties
  – Add metadata to enhance situational awareness

• Information and derived knowledge must be:
  – Timely
  – Accurate
  – Actionable

• Get the right knowledge to the right people at the right time
Objects

- Accessible via a unique identifier (OID)
- Can be linked by named relationships
- Can be indexed in multiple ways
- Can be included in multiple collections
- Can be named in multiple scopes
- Can be dynamically clustered
- Can be versioned
- Can be transient or persistent
- Interoperable across multiple languages (including SQL++) on multiple platforms
STORAGE ARCHITECTURE

64 bit Object Identifier (OID)

3456 - 7890 - 1234 - 56

Logical Database #
Logical Container #
Logical Page #
Logical Slot #

Federated Database → Partition → Database → Container → Object → VArray

Associations

Single Logical View of Millions of Exabytes
Single Logical View of Millions of Exabytes
Faster Navigation

**PROBLEM:**
Find all of the Suspects linked to a chosen Incident

Relational solution:
\( N \times 2 \) B-Tree lookups
\( N \times 2 \) logical reads

Objectivity/DB solution:
1 B-Tree lookup
1 + N logical reads
Lower Query Latency

Relational

O/R Mapping → Send Request → Receive & Interpret Request → Optimize → Access Indices → Qualify → Read Data → Create View → Return Result → Loop

Objectivity / DB

Initialize Iterator → Start Loop → Access Indices → Qualify → Open Object → Loop

Qualified objects are returned as soon as they are found.
Distributed Architecture

Enhances scalability and availability
Parallel Query Engine

The Task Splitter aims queries at specific databases and containers.

Filters can run complex qualification methods. Gateways can access other databases or search engines.

Replaceable components for smarter optimization.
Objectivity/DB Advantages

- Fully Distributed with Client-Side Smart Caching and Flexible Clustering of objects
- Highly efficient storage and navigation of relationships
- Scalable Collections
- Customizable Parallel Query Engine
- Quorum-based Replication and High Availability
- Flexible, Multi-mode Indexing
- Fully Interoperable Across Platforms and Languages
- Significantly reduces development time and effort.
NoSQL?
NoSQL?

• All of these features could have been obtained from “Commercial Off The Shelf” ODBMSs:

• Unique object/document IDs
• Sharding
• Shared-Nothing
• Fully distributed
• No lock and novel transaction modes
• Iterators and fast, predictable traversals
• Fast scans
• Optimization for random access
• In Memory Database configurability

• Flexible object clustering
• Effectivity (data in a relationship)
• Geospatial and multi-dimensional indexing
• Hash table (key-value) lookups
• Hyperspace = single logical view of a federation
• Multi-way replication
• High Availability
• Text searching
Objectivity/DB & Hadoop

- **Problem:** HDFS works best when transferring 64+ MB blocks of data. Objectivity/DB works best with 16-64KB blocks.

- **Solution:** Implement a memory cache for the HDFS blocks.
RELATIONSHIP ANALYTICS
Relationships Are Complex And Dynamic
Relationship Analytics...

• It’s all about the relationships

• Relational databases aren’t good at handling complex relationships
  – Join tables create scalability and performance problems
  – SQL isn’t designed to find relationships between data in different tables, linked by data in multiple other tables

• Object databases represent and handle relationships much more efficiently.
...Relationship Analytics...

Relational Database

Object Database
The Link Hunter

Searches 100 million records for relationships out to 5 degrees of separation in minutes rather than days
Graph Databases

- Vertex objects represent nodes
- Edge objects represent relationships
  - Edges may be weighted
- Both are regular object classes
  - Properties (data fields)
  - Methods (algorithms)
  - Inheritance (reuse).
InfiniteGraph...

- Dedicated Graph API
- Easy to use and deploy
- Available on Amazon EC2 and GoGrid clouds
- Placement Manager controls distribution of databases and containers
- Java now, Gremlin soon...
- Multiple indexing options
- Advanced parallel ingest
- Built on Objectivity/DB
- Will be fully interoperable
A Typical InfiniteGraph Installation
Graph DBMS Challenges

• Lack of standards
  – SPARQL, or an adaptation of it, may become the query language
  – GraphML many be used for transferring data
  – Gremlin is a promising API, but it’s inefficient

• Distributed, parallel queries
  – InfiniteGraph will leverage the Objectivity/DB Parallel Query Engine

• Visualization
  – Current tools don’t scale well
  – InfiniteGraph has partnered with Tom Sawyer
Questions?

• White papers, downloads etc.
  – objectivity.com
  – infinitegraph.com

• Presenter: Leon Guzenda - leon@objectivity.com