Collaborative Analytics Environment (CARE):
An End-User Programming Environment For Thinkers

Doug Talbott
Bedarra Research Labs
doug.talbott@bedarra.com

About Bedarra Research Labs

Private industrial research lab

Focus on next-generation computing and communication technologies

Launched in 2003 by Dave Thomas and Brian Barry

Previously founded Object Technology International –1988-1996

- ENVY/Developer, ENVY/Smalltalk, and IBM VisuallAge

Sold to IBM in 1996

- VisualAge for Java
- J9 Embedded JVM (Foundation for IBM Websphere)
- Eclipse and IBM Java Development Tools
- The Eclipse Foundation

CARE Design Goals

Analyst	Product
Domain experts in a range of disciplines	Support a range of thinking or organizational models (e.g. spreadsheets, tables, flowcharts, lists, etc)
Wide range of computer expertise	Support novice to expert programming (e.g. scripting, visual programming, DSL & wide spectrum programming)
Often work in teams when solving problems	Support local or remote collaborative shared spaces Support fine grained asset version control (i.e. cell level) Support workflow version control (i.e. replays) Support degrees of workflow and data visibility (i.e. security)
Often use big data sets from clouds to embedded sensors	Support large data sets Support small footprint for deployment Support scalability and concurrency
Use iterative approach as an exploratory tool with strong emphasis on hypothesis formulation and visualization	Support a “Think-Compute-See” problem solving workflow (e.g. multiple programming models & visualization techniques)
High performance computation is an essential part of their discovery and design process	Needs to be very fast and interactive
Analysts will use any combination of tools to get the job done.	Support interoperability with other languages and applications. (e.g. Fortran, C++, Java, Python, R, Matlab, Office)

CARE Functional Design

- 50+:1 of ratio of analysts to expert developers
- Experts surface new functionality to analysts as DSL library extensions

Analyst Application Programming Model

- Wide Spectrum DSLs [Spreadsheet, Tables, SQL, Boxes and Arrows, Visual Query...]
- Narrow Domain Specific DSL [intrusion detection, finance, geographic, etc...]
- Visualization [Scatter, Table, Bar, Histogram, Binning, Directed Graph, Map...]
- Leverage existing standards and user models
- Interoperable with R, MatLab, MS Office

Expert Programming Model

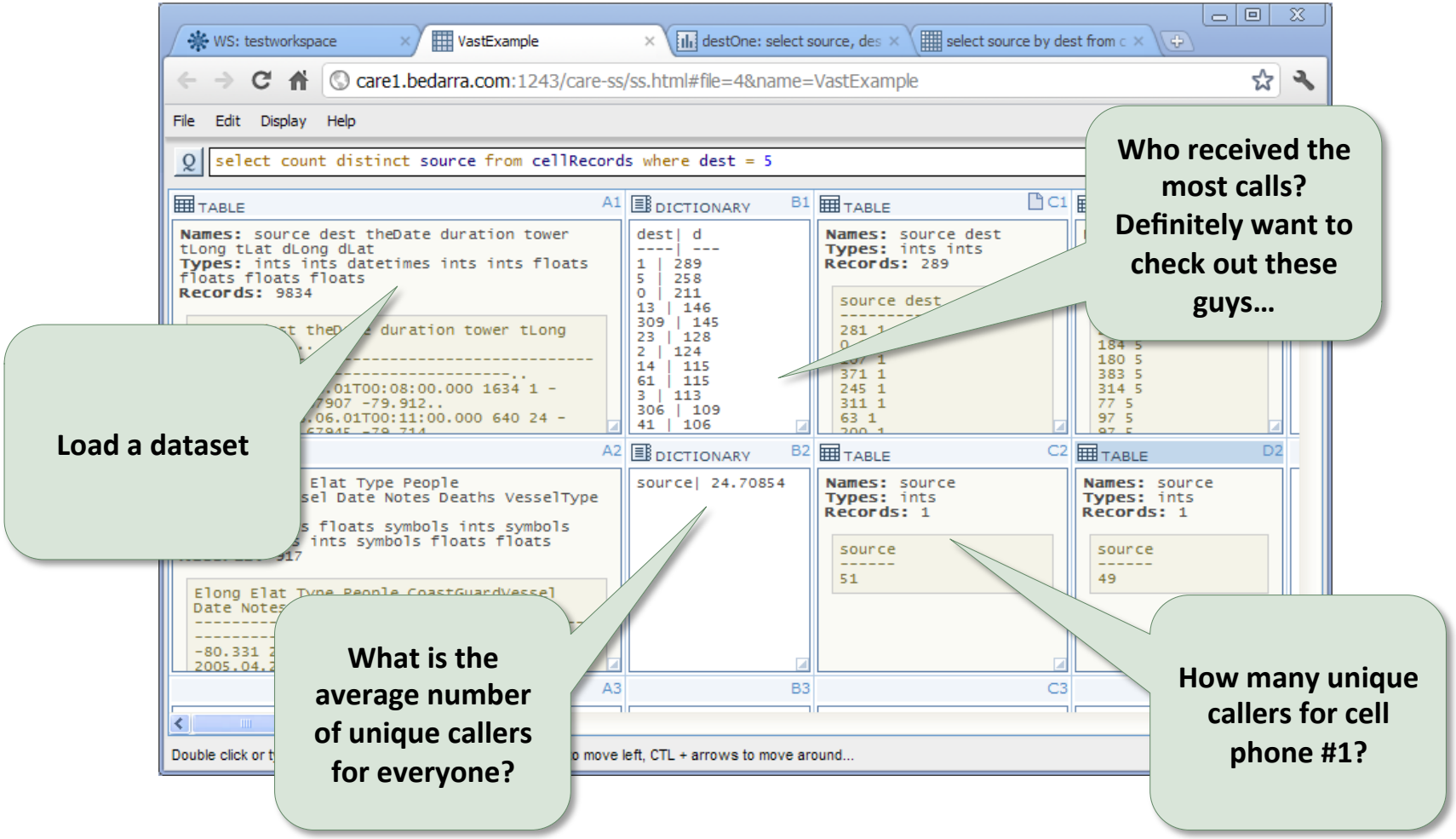
- Wide Spectrum Functional Vector Language
- Full Interoperable with current technologies such as ODBC, Java, C#, C++, Web

CARE Core Platform

- Column store, Core Libraries, Platform interoperability
- Virtual Execution Environment (VEE) – High Performance Vector Functional Runtime

CARE Think, Compute, See Example

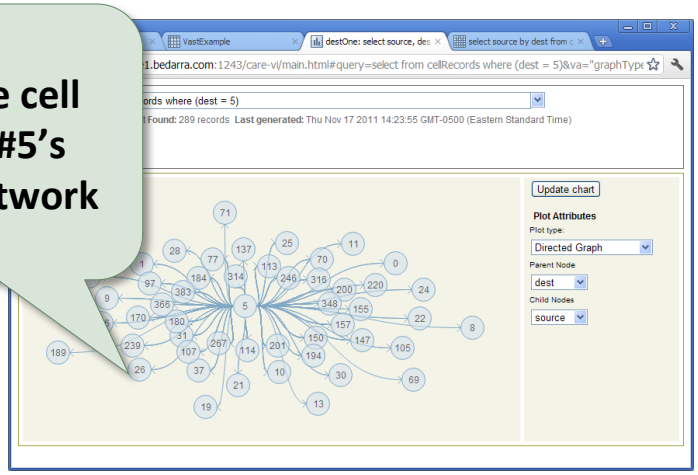
VAST2010 cell phone challenge...

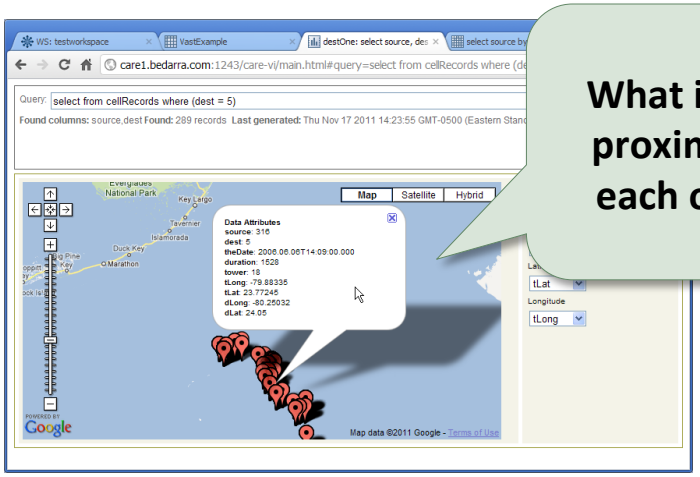


CARE Think, Compute, See Example

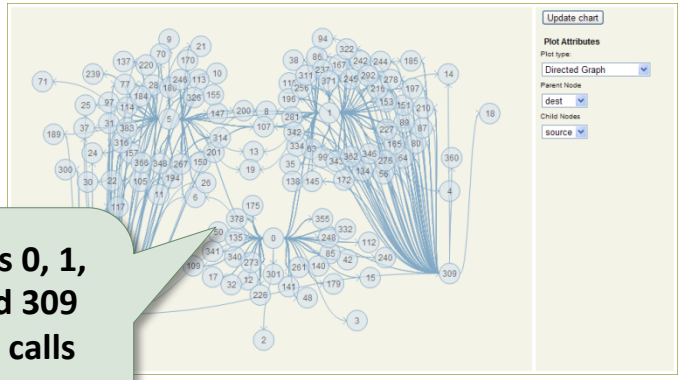
Spot a relationship between 0, 1, 5, 309, and 306 within minutes

Let's see cell phone #5's social network

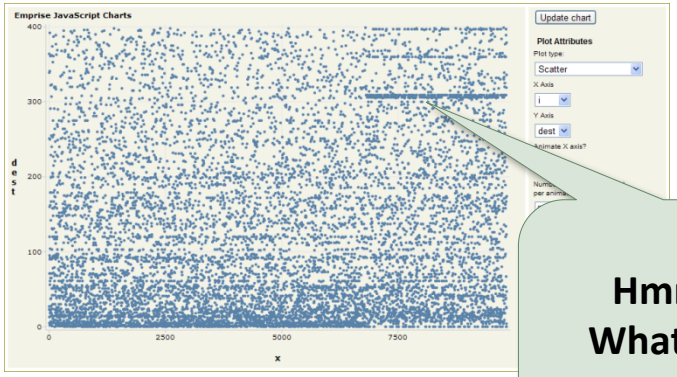




What is their proximity to each other?



Cell phones 0, 1, 5, 306, and 309 are taking calls from the same social network



Hmmm… What is this pattern?

CARE Summary

- A collaborative computing environment
  - Order of magnitude improvement in productivity
  - High-performance analytics over large volumes of data
  - Domain-oriented programming by end-user analysts
  - Rich and efficient library for scalar and collection programming
  - Minimal impedance between persistent and live data
  - Multi-paradigm programming for software experts
  - Fine grain versioning of assets and workflow

- Ongoing research activity
  - Currently developing v3.0