Information Coherency with Diverse Data Stores

Cheryl P. Jalbert

JCC Consulting, Inc.

Abstract

A utility company gained major advantages over the years from their Rdb data resource. However, they needed additional applications and were attracted to packages that did not rely on Rdb.

The visual spatialization package (GIS) that was chosen relies on Oracle on Linux for its data store and an intra-net, web-based application using Apache and Java. The web access provided is based on two different prongs. One uses SQL Server running on Windows; the other reads directly from the GIS system. Communication among these is transparent to the users and as close to real time as desired.

The success of the combined architecture has led to improved customer support, greater efficiency, greater safety for the workers, and rave reviews in the press.

This session describes the issues involved in maintaining a coherent approach with different data resources, the use of the JCC LogMiner Loader to replicate the data, as needed, to the Oracle and SQL Server data resources, and the mechanisms developed to provide feedback from those systems to the Rdb database.

CIS (Customer Information System)

- Rdb database designed in 1992.
  - Expanded functionality without need for fundamental change.
  - Rdb V7.2.1.1.1, Open VMS V8.3 with Integrity coming
- Applications
  - DECForms UI replace by Oracle Forms to be replaced by JAVA
  - COBOL, JAVA, BASIC, C, PERL
  - Business logic in stored procedures
- Success level: HIGH!

Applications Covered

- CIS (Customer Information Services)
- Billing
- Cash processing
- Credit
- Load management
- New installations
- Outage reporting
- Service assignments
- Trucks and drivers available

"Companions"

- Accounting
- Bill printing
- Automated telephone answering for crisis times
- Service truck location and messaging
- Mapping
- Internet display of data
- Internet bills payments and information updates

Communication with the "Companions"

- The first few "companion" applications have rather simple interfaces with the Rdb database.
- Accounting
  - COBOL creates flat files from the Rdb database and posts them in specific target locations.
  - Accounting software loads the data from the flat file.
- Bill printing:
  - Data provided to the bill printer in a flat file.
  - Printer sets up the layout, prints and also provides PDF bill images.
  - Application (in testing) to find and display the specific PDF file required to review a specific bill so that the bill shows in the standard UI in the exact same format that the customer is viewing.

Communication Gets Somewhat More Interesting

- Telephone answering service
  - Provides support for overflow calls in large outages.
  - Data on the customers needs to be available to the service.
  - Data from the calls must be available promptly to aid outage analysis.
- Current answer
  - Twice a month, flat file is created from the Rdb data and supplied to the service.
  - Flat files come back. (More on this, in a moment.)
- Future
  - A subset of the Rdb database will be directly accessed and updated by the service.
  - The subset to be created with the LogMiner Loader to provide up to the second accuracy.

Writing Data to the DB from a File

- JCC developed a technology for writing data received in flat file or other format into the Rdb database.
  - Asynchronous
  - Custom
  - Takes advantage of specific procedures for specific functions
  - Uses knowledge of specific db design and specific application to maximize results
  - Generic to some extent
- For the purpose of this discussion, we will call the procedures that implement this technology "services."
  - There is another obvious name, but using it makes this discussion hard to follow.

Outage Reporting Service

- To load the data from the calling service into the database uses an Outage Reporting Service.
- ORS is written in COBOL and JAVA.
- ORS avoids issues of conflicting data updates by utilizing details of the db design.
  - If the outage has not been reported, one is added.
  - If the outage has been reported, columns that are blank can be updated. (The call may provide additional information, such as "There is a tree on the line.")
  - If the outage has been reported and the relevant columns are not blank, a row is written to a table for additional activity relative to the specific outage.
- ORS is somewhat generic in that it can be used for multiple outage reporting services.
  - Telephone service
  - Internet reports
  - Reports from automatic devices

Trucks in the "Field"

- Service trucks work both routine and emergency issues.
- Tracking service trucks and sending messages between the office and the truck is important.
- Communication with the trucks is via satellite and Qualcomm using HTTPS.
- Messages are in XML.
- Data on location and messages is inserted with a JAVA procedure.
- The procedure has many of the characteristics of other services that write data to the db.
  - Asynchronous
  - Custom
  - Embedded with knowledge of the database design
  - Generic, to some extent

Power Outages

Power Outage photo from Wikipedia

Ice Storms → ANGER

5FP011EA12165FP011EA12161a first fri

NEWS, 2A — Man catches baby dropped from third floor

LIFESTYLE/FAMILY, 1D — DON'T LOSE SLEEP OVER SLEEPOVERS

Banish boredom with these events / Upstate Weekend

The Greenville News

FRIDAY, DECEMBER 16, 2005 * FINAL EDITION

greenvilleonline.com

It could take several days to restore power

■ Schools closed again today ■ Ice hits Greenville hardest ■ Commuters facing delays

Residents find no room at local inns

SHARE YOUR STORM PHOTOS:

GET STORM SMARTS:

COPING WITHOUT POWER

Families huddle to fight cold

Icy mess challenges residents

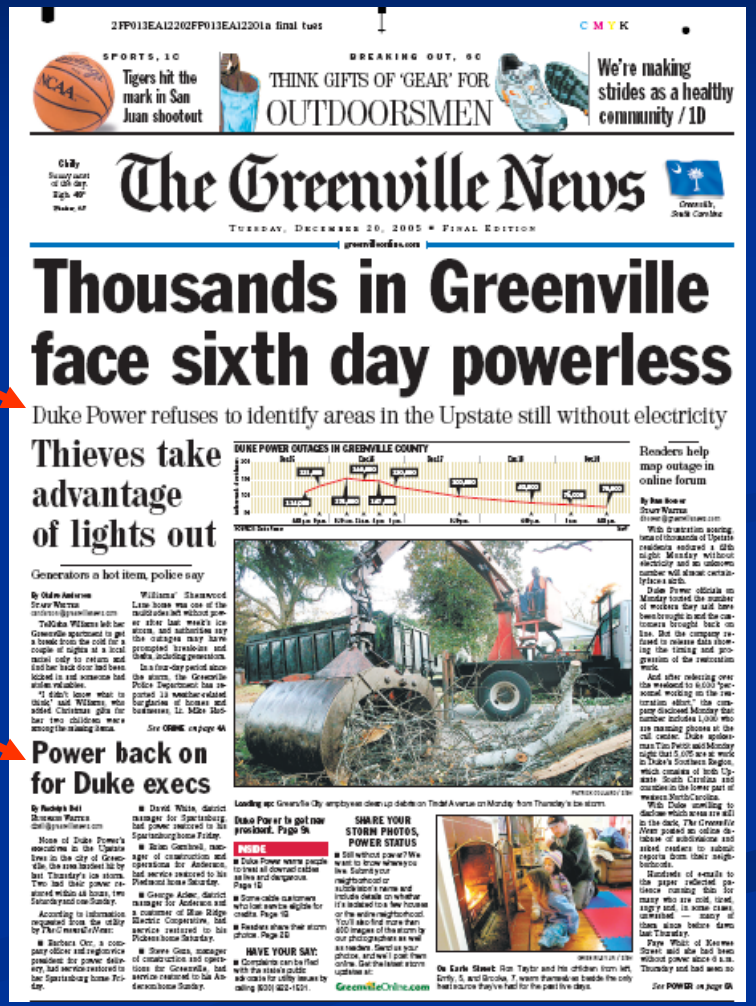
Power outages in the Upstate

INSIDE

NEED SHELTER?

Duke Power
Dec, 2005

When information is not available and power is out, the power company is blamed and anger can rule.



2FP013EA12202FP013EA12201a final tues

SPORTS, 1C — Tigers hit the mark in San Juan shootout

BREAKING OUT, 6C — THINK GIFTS OF 'GEAR' FOR OUTDOORSMEN

We're making strides as a healthy community / 1D

The Greenville News

TUESDAY, DECEMBER 20, 2005 * FINAL EDITION

greenvilleonline.com

Thousands in Greenville face sixth day powerless

Duke Power refuses to identify areas in the Upstate still without electricity

Thieves take advantage of lights out

Generators a hot item, police say

DUKE POWER OUTAGES IN GREENVILLE COUNTY

Readers help map outage in online forum

Power back on for Duke execs

Duke Power is got mar president. Page 9A

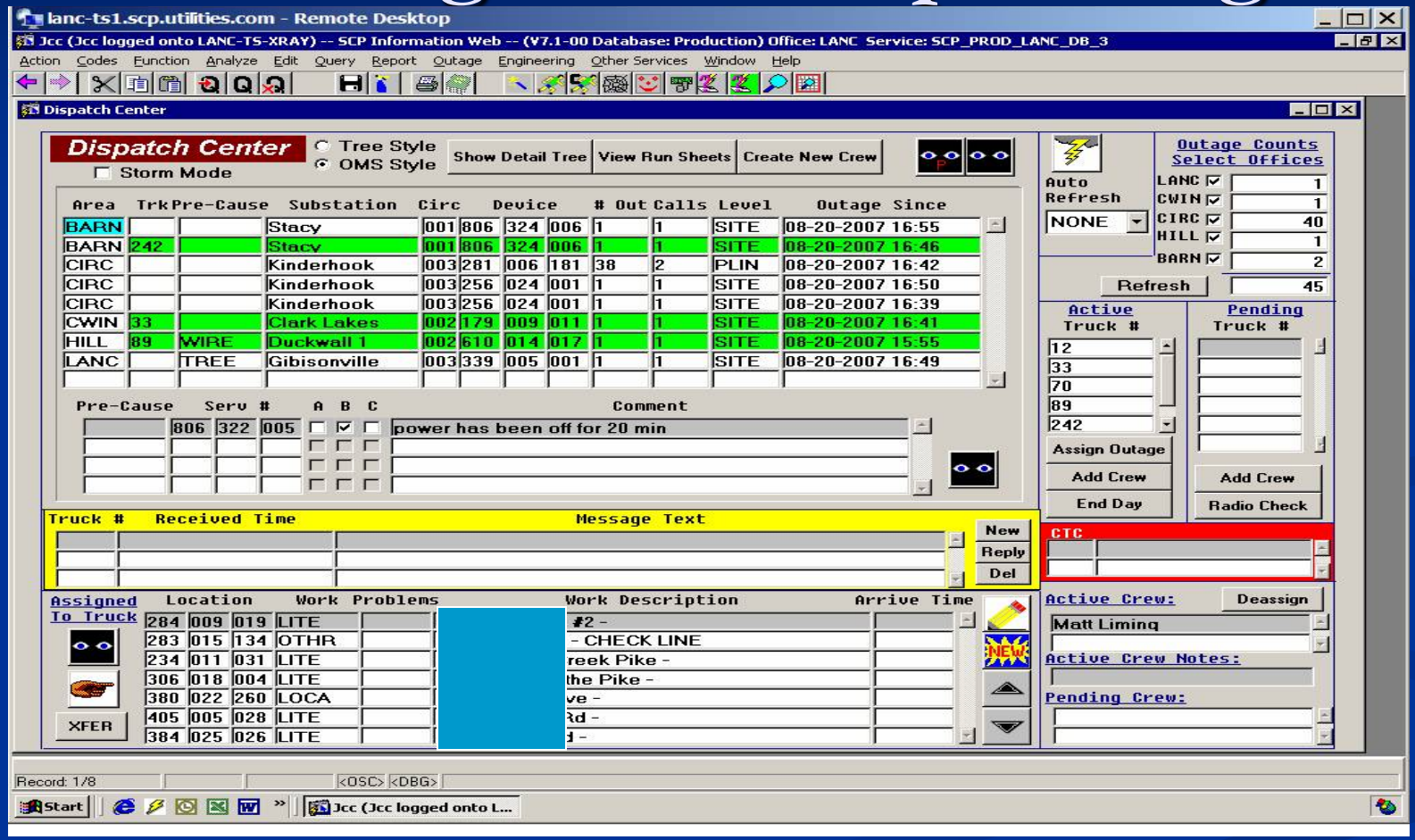
INSIDE

HAVE YOUR SAY:

SHARE YOUR STORM PHOTOS, POWER STATUS

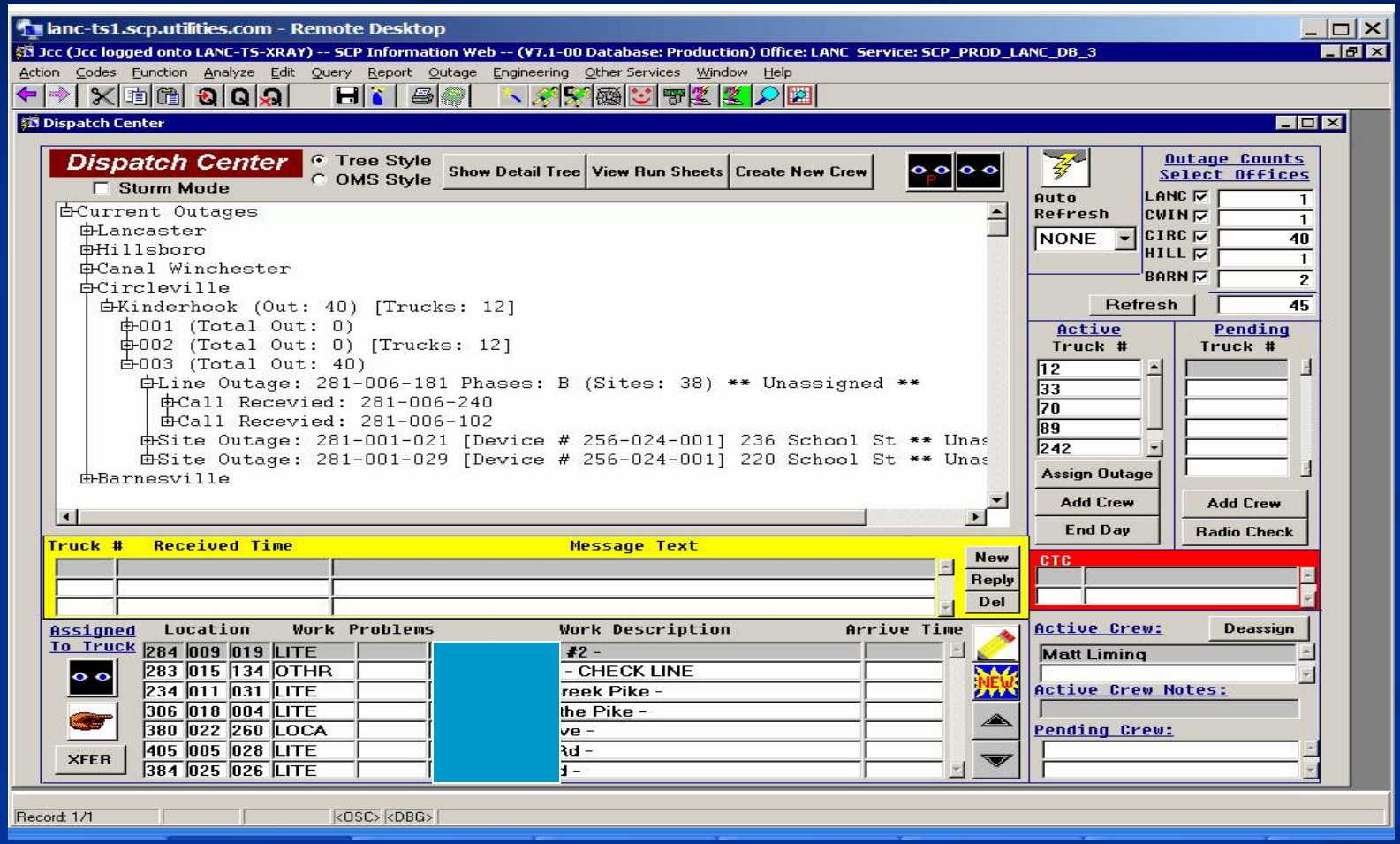
GreenvilleOnline.com

At Our Example Company Outages and Dispatching



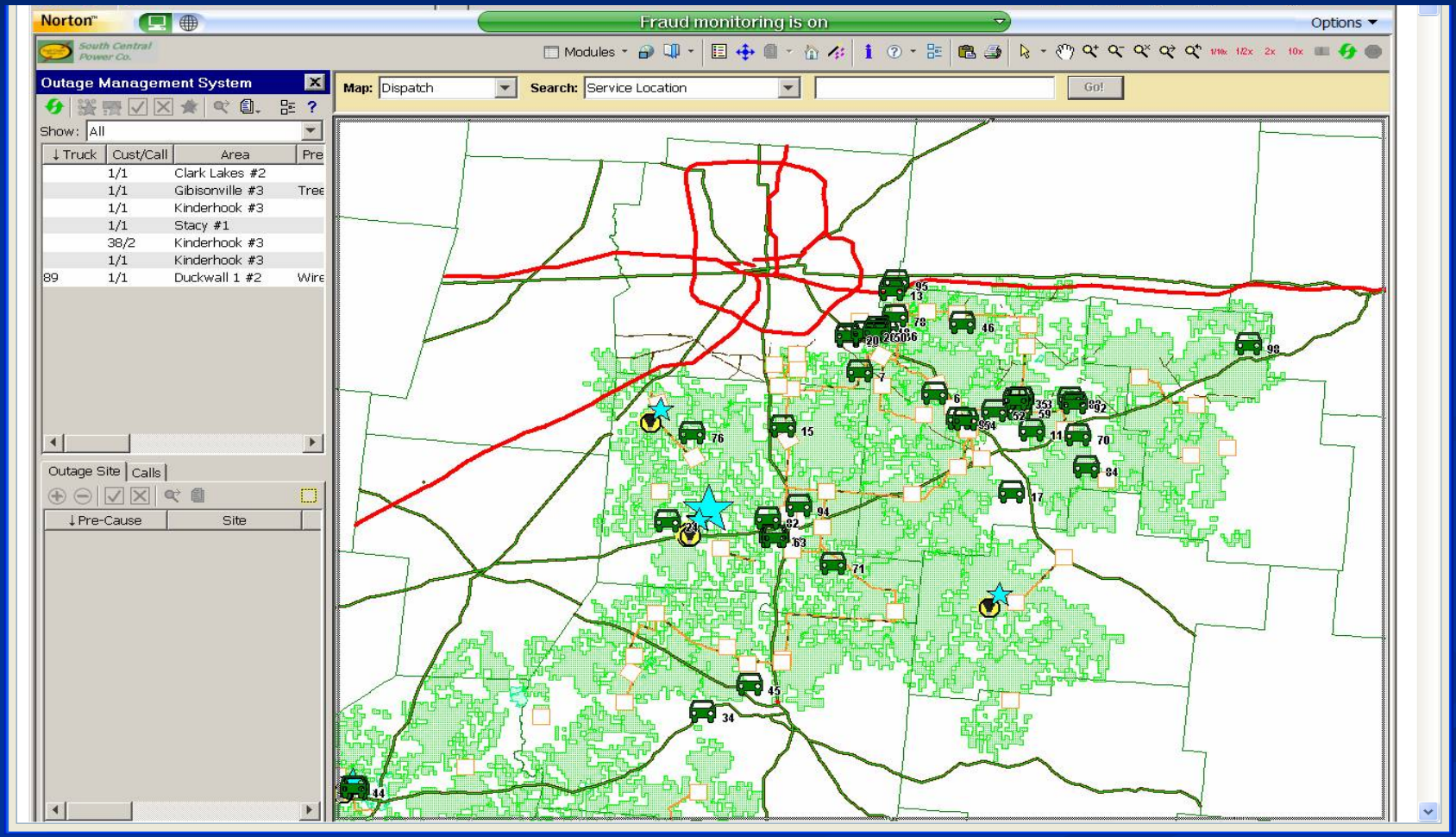
CIS UI lists outage reports, shows the # out for each, and shows some summary.

Outages and Dispatching



Alternate view provides "drill down" capability. However …

Mapping Package Example



Bigger Challenges from "Companion" Applications

- Mapping package
  - Oracle RDMS on Linux
  - Apache and JAVA
  - Intranet presentation
  - Communication of data must be two-way.
- Support for the customers via the internet.
  - Billing and CIS information in SQL Server on Windows
  - Communication of data must be two-way.

Two-Way Data Communication

- Runs the risks of
  - Buried updates
  - Cycles of updates
  - Conflicting information
- Requires
  - Careful definition
  - Careful architecture
  - Leadership
  - Teamwork

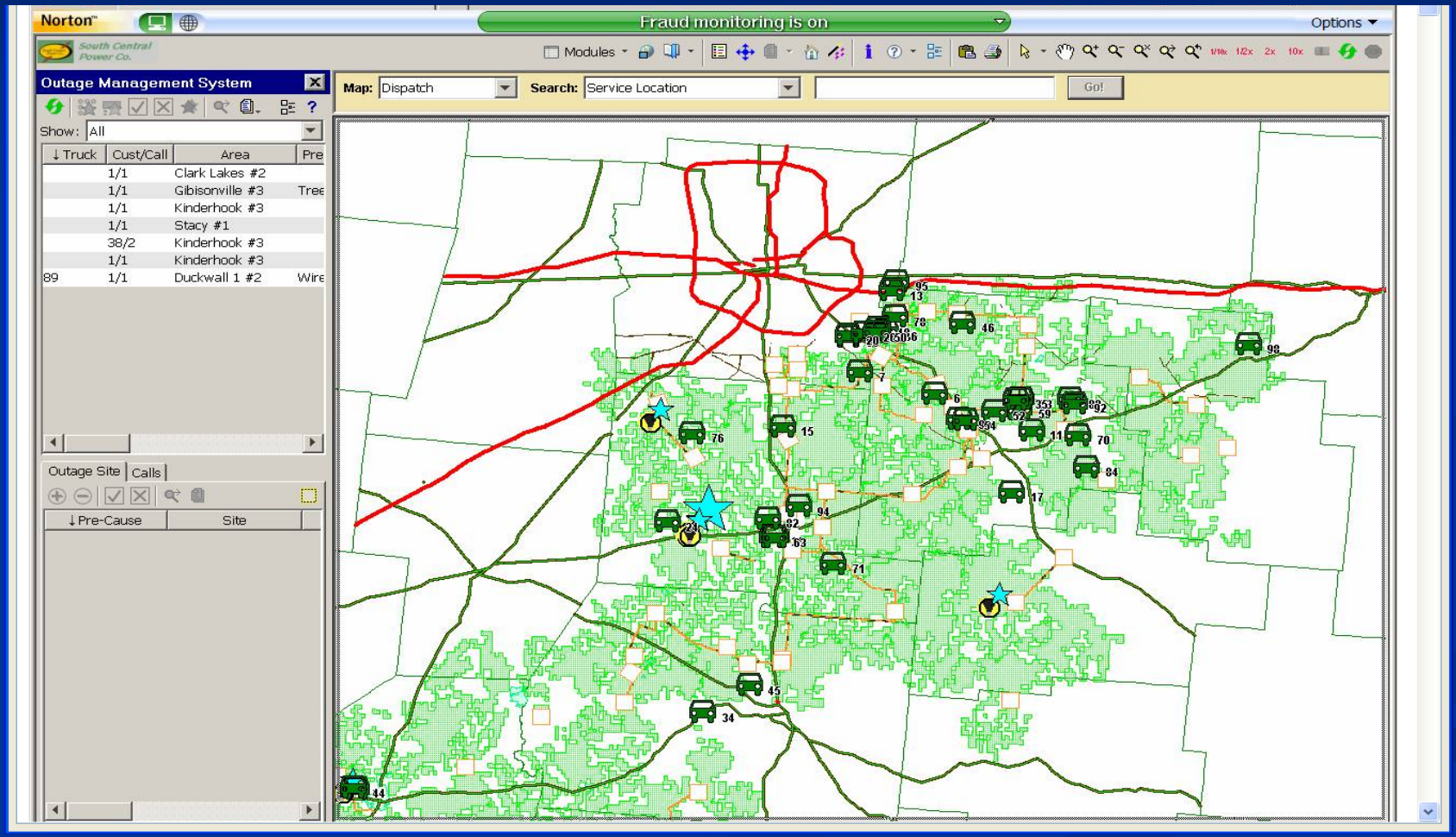
Decisions

- Data consistency suggests a "database of record."
- Definition is required for how the Oracle database gets updates from the Rdb database.
  - Tables of interest.
  - Columns required.
  - Transforms required.
- Definition is required for how the Rdb database gets updates from Oracle.
  - What can be changed via the Oracle interface?
  - What does it map to in the Rdb database?
  - How do we avoid cycles?
- What happens if one of the databases is "down"?

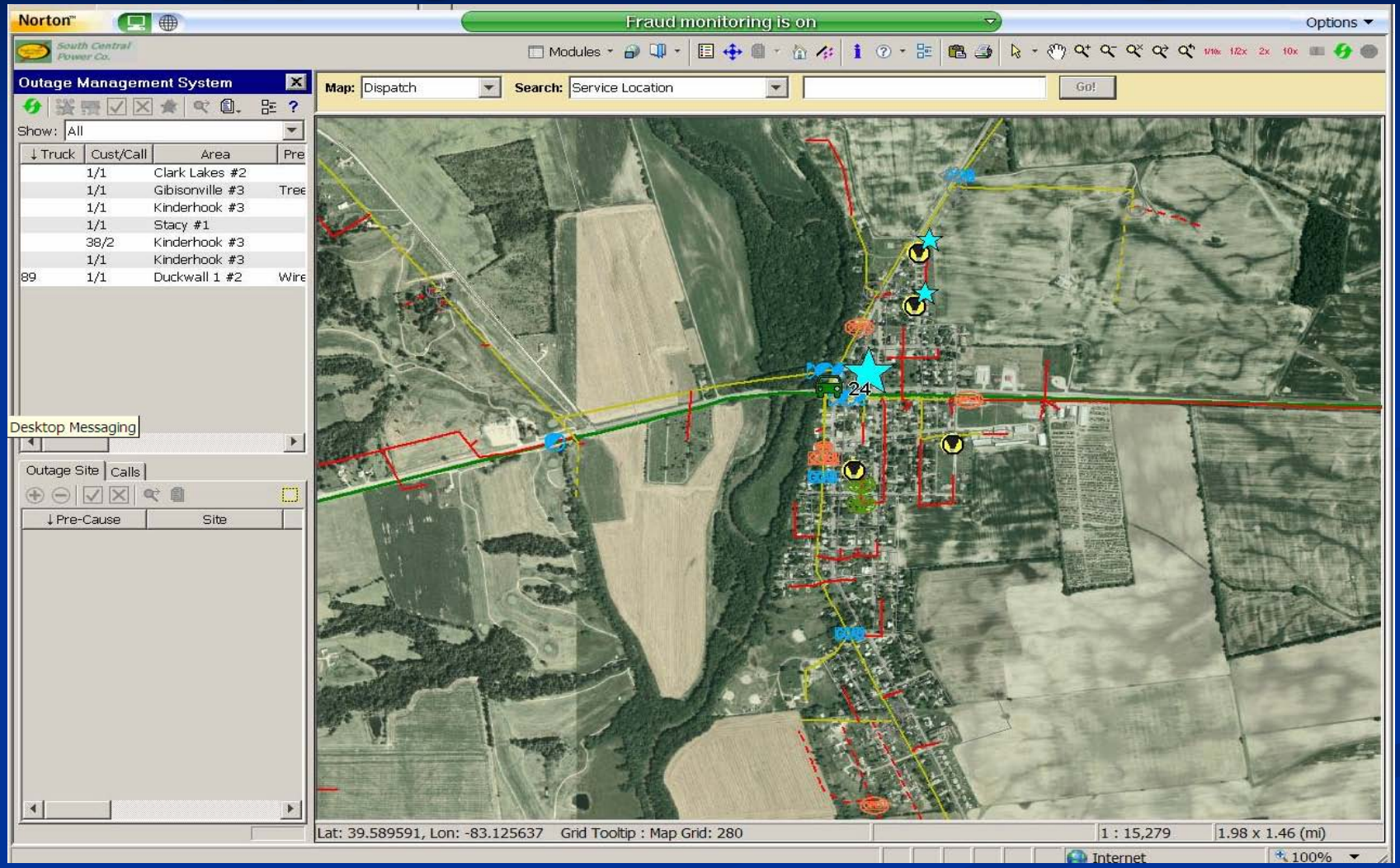
Outages and the Power Grid

- A single outage may be due to something like a tree on a line.
- An ice storm may take out multiple lines, including some trunk lines.
- Difficulty at a substation will affect many residents, but may have a one-stop solution.
- Quick repair benefits from
  - Prompt reporting of information
  - Knowledge of the power grid (substation → circuits → lines → individual sites)
  - Knowing and controlling service truck location

Mapping Package Example

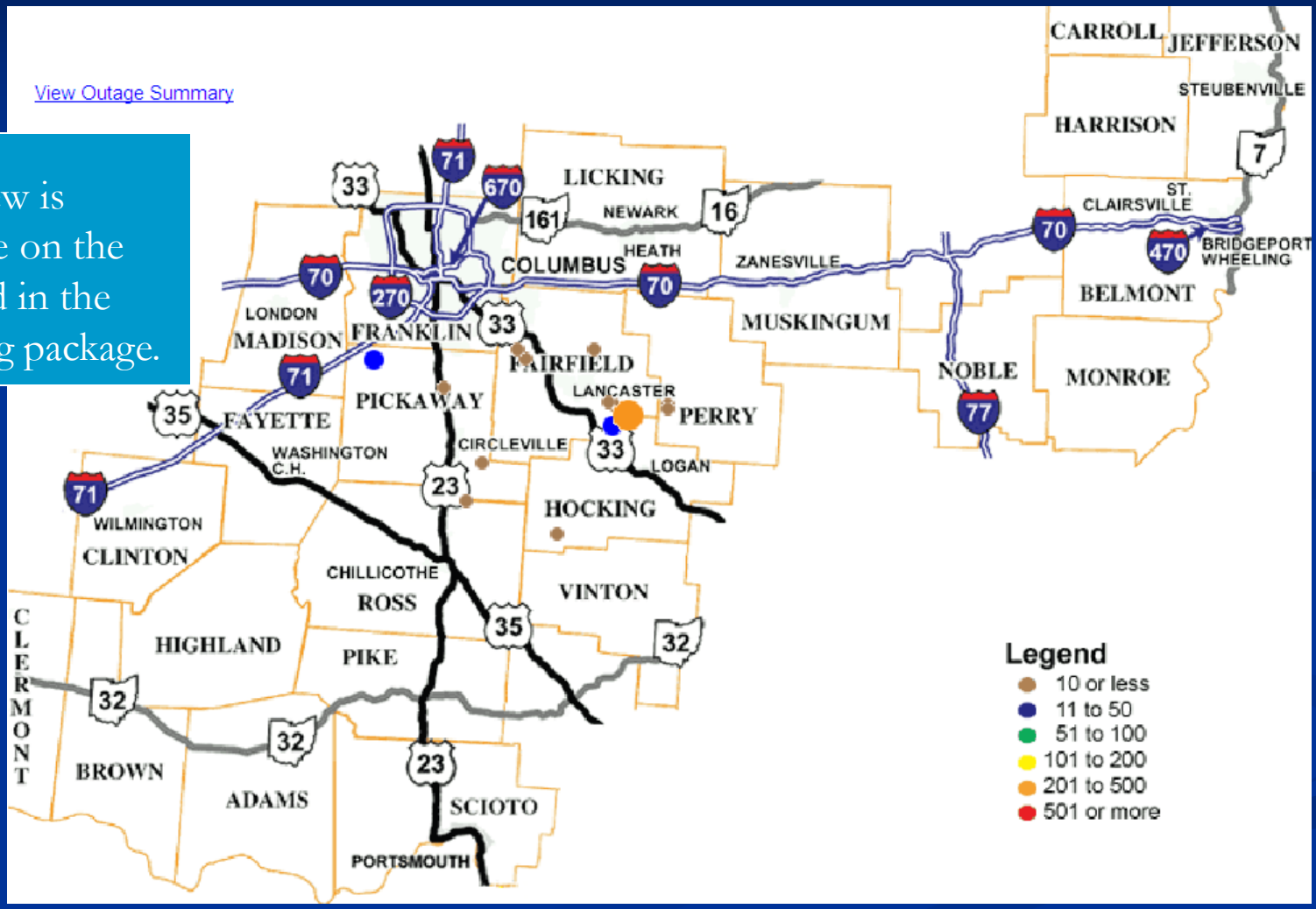


Mapping on Satellite Imagery



Summary View

This view is available on the web and in the mapping package.



Greater Detail

This view is available on the web.

Mapping

- The Mapping package
  - Provides information graphically, providing instant impact.
  - Provides summaries graphically.
  - Includes information on the power hierarchy and connectivity of the grid, not previously available on-line.
- The data synthesis of the map data with the CIS data from the Rdb database and various map sources.
  - Provides the data needed to research a problem and assign a solution → shorter downtime.
  - Avoids duplicate data entry and inconsistent data resources.

Decisions Made

- Rdb is the "database of record".
- Down time is a risk that is within reason.
  - Rdb has proven reliable enough to trust.
  - Disaster recovery systems are available.
  - LogMiner Loader protects against any downtime on the target side.
    - The only requirement is to save the AIJ backup files until they are processed.
    - The Oracle database can be updated when available again.
    - Outages can be logged without the GIS system.

Oracle to Rdb

- GIS application writes some data directly to the Oracle database.
  - Information related to specifics of the line connectivity.
  - Information not duplicated in CIS (Rdb) db.
- Data of general interest, including outages, is written by the GIS application with a DB-link to a virtual Oracle table that is actually in the Rdb database.
- A write to the table in Rdb triggers an AST doorbell.
- The doorbell triggers another "Service."

Service Features

- The services that we are discussing
  - Are middleware.
  - Are asynchronous with other work.
  - Handle some updates and pass other updates to task-specific servers.
- The service for writing updates to Rdb from mapping
  - For outages reported for a specific site: writes the outage
  - For outages at the "higher levels" (substation, circuit, group of lines):
    - Writes the information that will show the outage at the highest level
    - Passes the information to an "outage open/close server" to update information for all the sites and intermediate entities within the hierarchy
  - The performance advantage of this two-fold approach will be obvious in a moment.

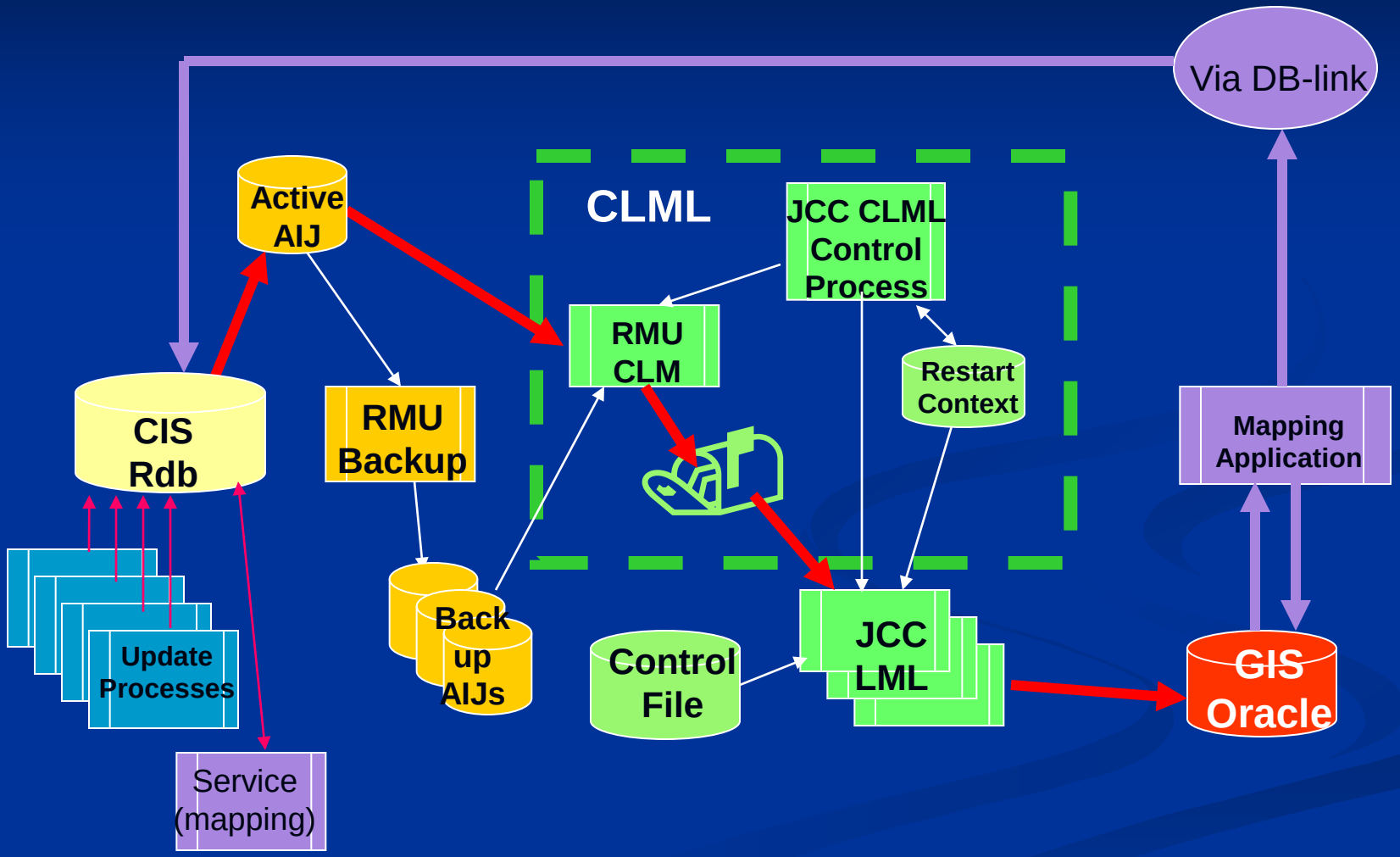
But Wait!

- Remember that the outage data written to the virtual Oracle table is not written directly to the Oracle database.
- As soon as it is written to the Rdb database and committed, it is picked up by LogMiner and the Loader.
- As soon as the Loader writes the data to the Oracle database, the symbol for an outage can be shown on the map.
- This is why it is important to write a substation outage immediately and pass off writing, perhaps, thousands of site outage rows to a server.

The Full Trip

- The full trip sounds lengthy
  - GIS application →
  - DB-link to virtual Oracle table that is in Rdb db →
  - Service →
  - Rdb db outage table →
  - AIJ →
  - LogMiner →
  - LogMiner Loader →
  - Oracle db
- The trip takes on the order of 0.03 seconds.
  - Site outages that are derived (from the site's being connected to the substation that is out) can take somewhat longer without significant impact on planning.

The Full Data Path



Details to Notice

- To the GIS application, the Rdb table accessed via DB-link is just another Oracle table.
- The Service process is just another update process for Rdb.
- The LogMiner Loader Control File is defined to read the data written by the Service process and the servers that it calls, but to exclude the table written via DB-link.
- No cycles of data updates are created.

Time-Out

- JCC's LogMiner Loader works with one source, Rdb, and many targets, including Oracle.
- JCC is often asked if the Loader can add Oracle as a source.
  - JCC has done the preliminary work with excellent support from Oracle Engineering.
  - Completing the work will require development time and funding.
  - There are no immediate plans for completion of this feature.
- However, consider the example.

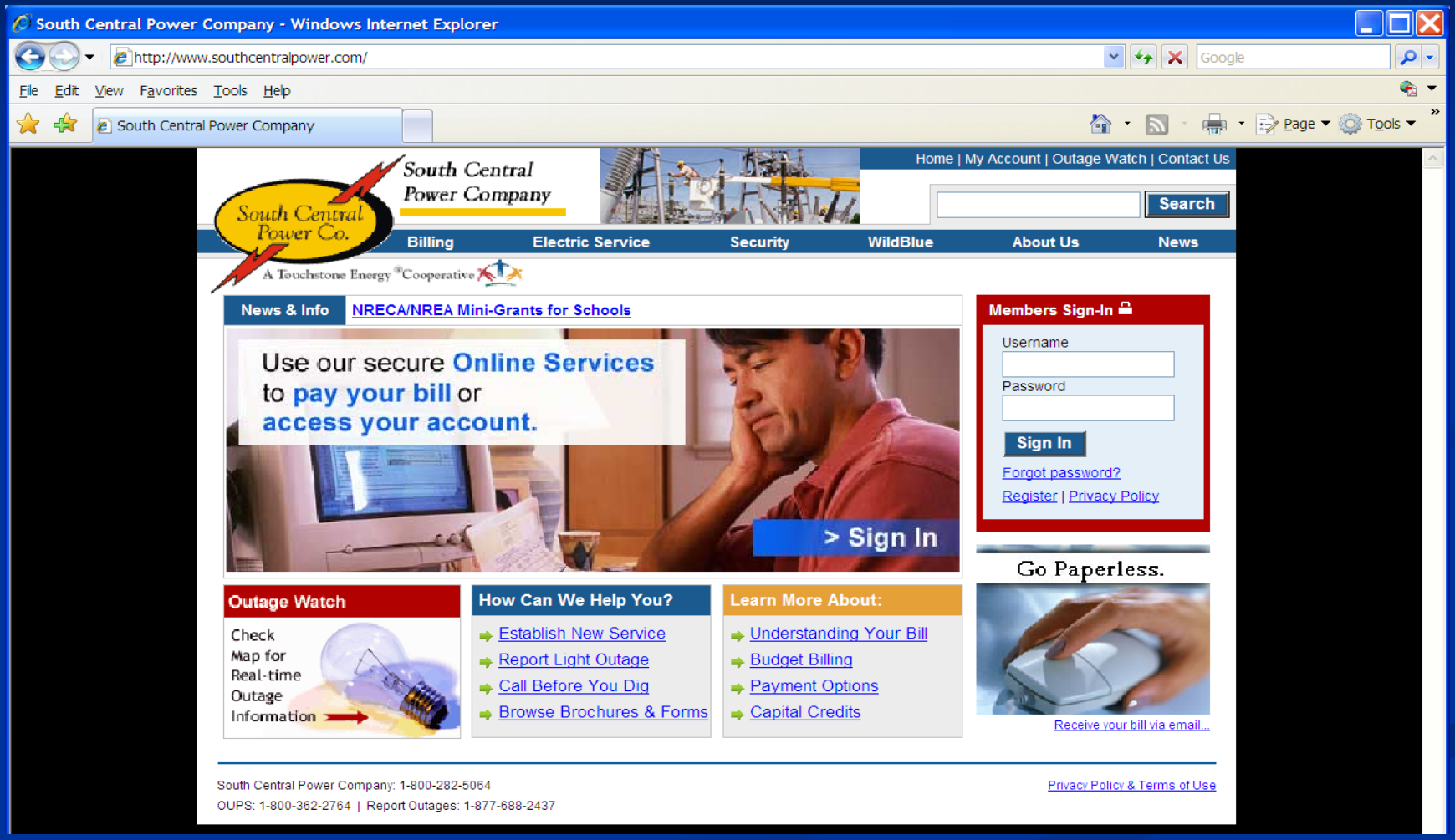
Oracle to Rdb

- The union of the CIS system and GIS system requires writing data that must appear in both an Oracle db and an Rdb db.
  - The GIS application "knows" only Oracle.
- The solution shown is *application specific*.
- The solution takes advantage of *components already in place*.
- Can the architecture be applied elsewhere?

Management

- Sometimes, the biggest impediment to coherency in data resources is a lack of clear leadership.
  - The issues are technical.
  - The path to updating the Oracle DB is not straightforward to express, but it is successful, robust, and timely.
  - The success requires careful definition of which data requires which treatment.
  - The management decision to avoid duplicating data entry was critical to this architecture.
- Applications that rely on varied data resources and are produced by a variety of teams can work coherently.

Web Access



Web Information

- Two web sites
  - Public
    - Outage data publically available
    - Log-in for
      - Outage reporting
      - Bill viewing
      - Data changes to name, address, etc.
      - Bill payment
  - Contractor specific
    - Log-in for
      - Assignments
      - Report completion

Package with SQL Server

- Data
  - In task-specific SQL Server DBs
  - Each a subset of the Rdb DB data
  - Replicated from the Rdb DB by the LogMiner Loader
- Map views read directly from GIS application
- Updating Rdb
  - JAVA application (another "service") connects to the SQL Server DB and inserts into the Rdb DB
  - Triggered by a scheduler job

Service for Web Customer Updates

- Service
  - Reads from SQL Server
  - Writes data changes directly, when no transforms are required.
  - Modifies the format of the SQL Server Payment record into data for the Cash Batch Table.
  - Existing components apply cash batches to the payments table and to the Accounting information.
- Note the similarity of approach to the Mapping Service.

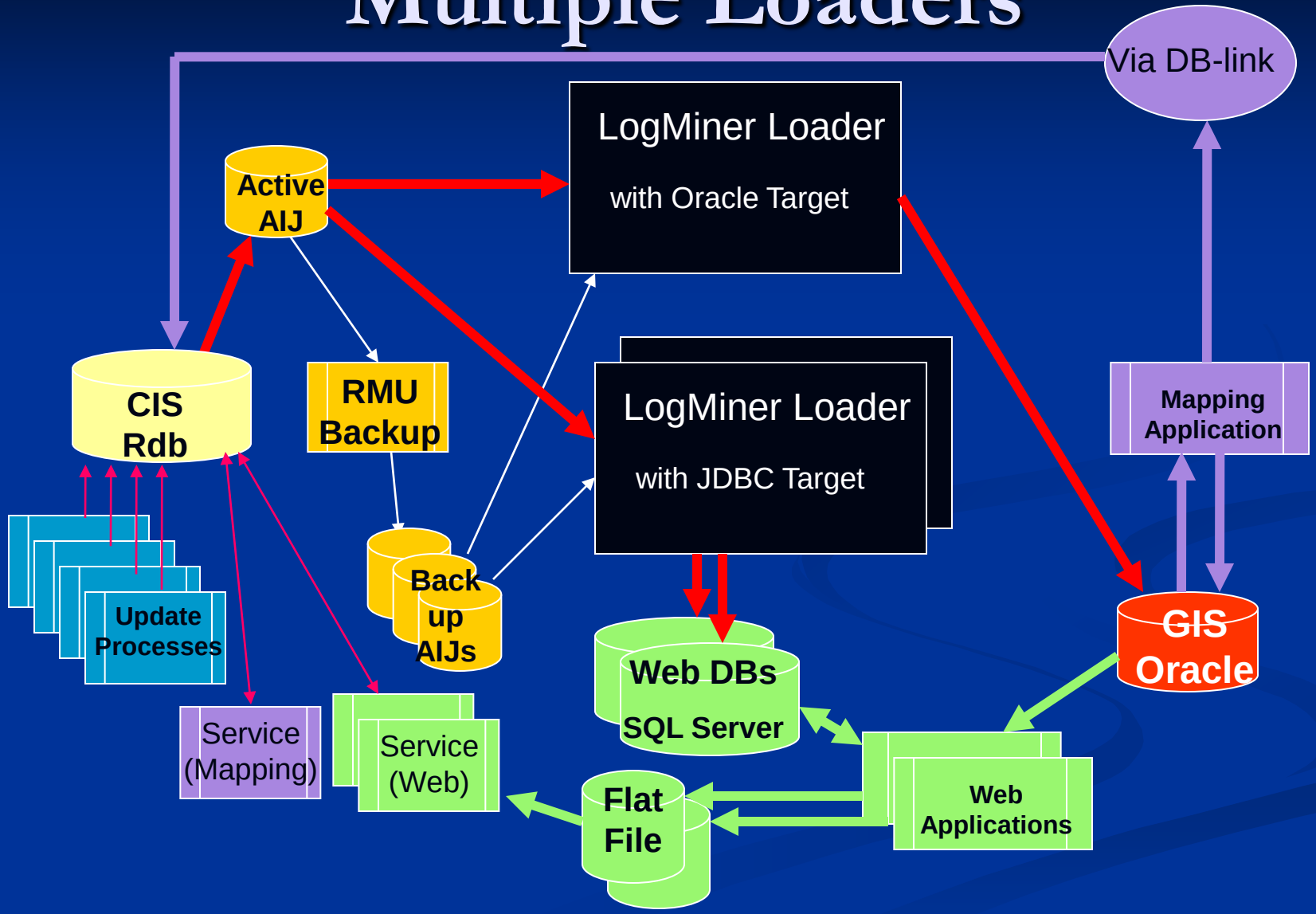
Service for Web Outage Reports

- Web outage reports are fed to the *same* Service that updates the database with reports through the calling service.
  - Custom to the outage reporting activity
  - Generic enough to handle multiple sources

Contractor Web Site

- The web site for contractors supports assignment of work orders.
- Before the web site, orders were generally delayed a day in fulfillment due to inefficient processes.
- Before the Loader the job took seven hours.
- It is now “immediate.”

Multiple Loaders



Coherency

Tools Used

- JCC LogMiner Loader
  - Powerful tool for replication
  - Configurable
  - Supports subsets and some data transforms
  - Provides "near realtime" performance
  - Includes abundant monitoring and testing tools
- Custom "Services" plus existing update servers
  - Custom code
  - Generic concepts

Indications of Success

- Remember the angry press for Duke Power in the December, 2005, ice storm?
- In January, 2007, the Loader user reported here faced a major ice storm.
  - One third of its customers lost power.
  - The company's main office lost power.
  - The company went into its usual crisis mode …with the additional benefit of
    - Mapping management of resource deployment.
    - Outage information being available to its customers (and the press).
- Customers and the press cheered!

Other Details

- Remember that an Integrity computer is arriving soon?
  - Testing in development will run a workload on the alpha and on the Integrity.
  - The Loader will run on each db to capture the results of the work load.
  - The target for each Loader will be XML.
  - Running differences on the two XML outputs will be straightforward.
    - One limit: We don't expect timestamps to match and so will exclude them.

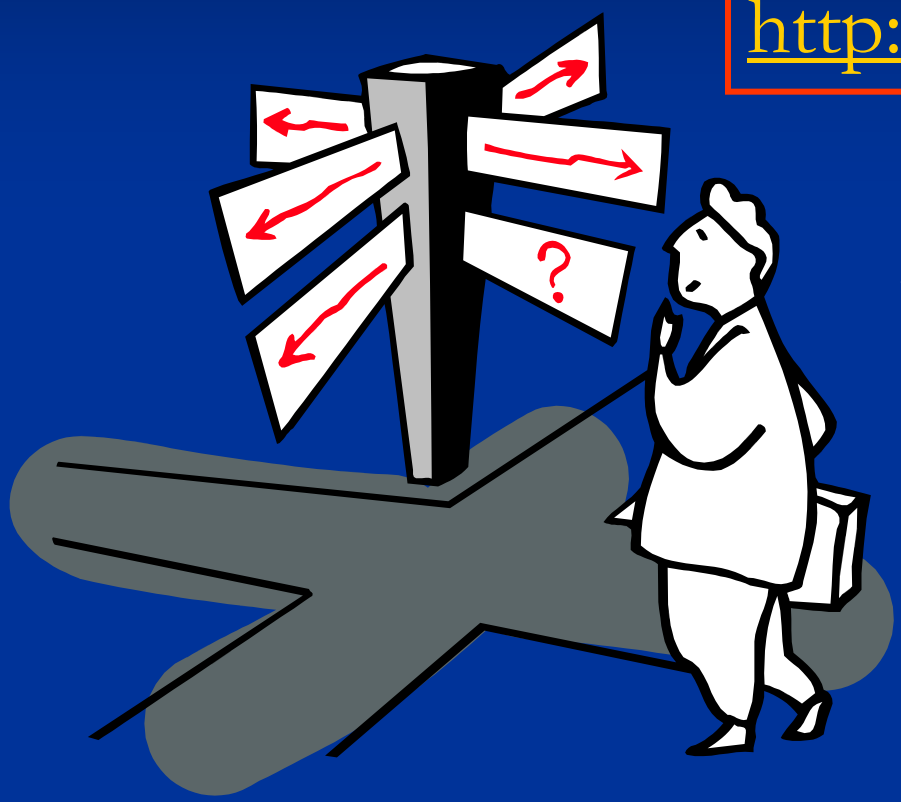
Availability

- The Loader kit is available at FTP.JCC.COM
  - Documentation
  - Kit
- Evaluation license available on request
  - Send mail to info@jcc.com
- Find descriptions of the LogMiner Loader and other information at http://www.jcc.com/LML.htm

Acknowledgements

- Thanks to GateKeeper (Philip A. Naecker [mailto:pan@gatekeeper.com]) for permission to discuss the spatialization (mapping) package.
- Thanks to the "unnamed power company" for permission to share the work and results.
- Thanks to colleagues – Jeff Haidet, Tom Musson, and Jeff Jalbert who developed the server architectures and reviewed my comments.

Questions



http://www.jcc.com/LML.htm

Join the worldwide Rdb community. Send mail to

OracleRdb-request@JCC.com with "SUBSCRIBE" in the body of the message.

Send additional questions to: Info@JCC.com