



Data Quality On Demand Must Accommodate Imprecision

What's the Solution?

Clint Bidlack, President and CTO, ActivePrime

ActivePrime Inc

ORACLE

Author: Clint Bidlack, President and CTO of ActivePrime.



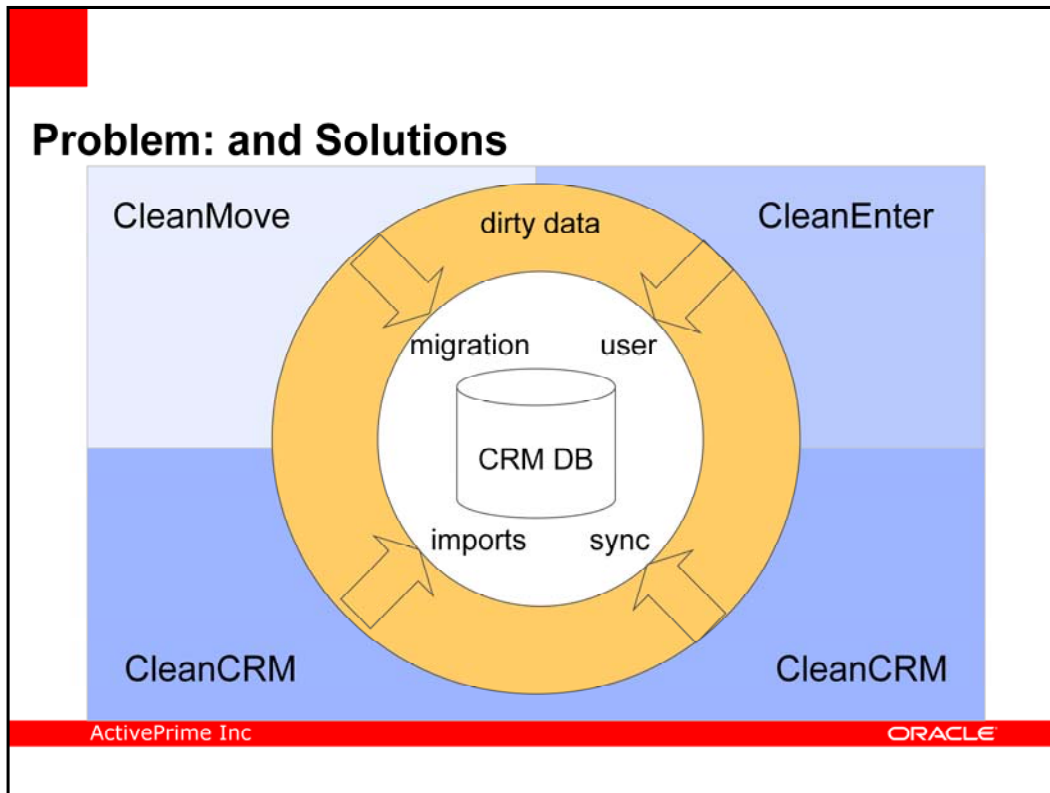
Overview: Think Big

- Dirty data everywhere
- **Big** corporate headache
- **Think** about context
- ActivePrime has solutions

ActivePrime Inc

ORACLE

Think differently about DQ



Problem: There are many ways for dirty data to enter a databaes. These include initial data migrations (data loads), users entering data, ongoing imports, and syncs with external applications (like Outlook, ERP systems, etc).

Solution: ActivePrime has solutions for all these sources of dirty data. For instance, with CRM data,

- CleanMove to support clean data upon initial migrations.
- CleanEnter to support clean data entry by all users of the system.
- CleanCRM to support ongoing cleansing of data to detect bad data sync'd into the system as well as ongoing importing of data.

Analogy: Dirty Water



ActivePrime Inc

ORACLE

Modern enterprise software systems with dirty data are like...

Clean plumbing with beautiful faucets. But that's deceptive... no one will drink the water.

The customer is disappointed!

Google: Missing Context



The screenshot shows a Google search interface. At the top left is a red square. Below it, the Google logo is followed by a search input field containing the text "willy clintkton". To the right of the input field is a "Search" button and a link for "Advanced Search". Below the search bar, there is a "Web" tab and a link for "Show options...". The main content area displays the message: "Your search - **willy clintkton** - did not match any documents." Below this, under the heading "Suggestions:", there is a bulleted list of advice: "Make sure all words are spelled correctly.", "Try different keywords.", "Try more general keywords.", and "Try fewer keywords.". At the bottom of the page, there is a footer with links for "Google Home", "Advertising Programs", "Business Solutions", "Privacy", and "About Google". A red horizontal bar at the very bottom contains the text "ActivePrime Inc" on the left and "ORACLE" on the right.

Most people recognize what “willy clintkton” should return. Without context (don’t know that it’s a person), this is difficult!

ActivePrime: Context Aware

- Fuzzy search
- Detect duplicates
- Real-time

The screenshot displays the ActivePrime Search interface. On the left, a smaller window shows the search input 'willy clintkton' and a 'Search' button. An arrow points to the main search interface on the right. The main interface has a header 'ActivePrime Search' with links for 'Learn' and 'About'. Below the header is a section titled 'Search for a Contact' with three input fields: 'First Name' containing 'willy', 'Last Name' containing 'clintkton', and 'Account' which is empty. There are 'Search' and 'Clear' buttons below these fields. A second section titled 'Use an existing Contact' contains a table with the following data:

	Account	First Name	Last Name
Use	MVP Health Care Inc	William	Clinton
Use	MVP Health Care Inc	William	Clinton
Use	MVP Health Care	Bill	Clinton
Use	MVP Health Care	Billy	Clinton

The footer of the interface includes 'ActivePrime Inc' on the left and 'ORACLE' on the right.

Because context is known and it is leveraged, an ActivePrime search for “willy clintkton” returns records as expected.

Real time fuzzy search engine applied to data quality.

Fuzzy search is a novel, exceedingly valuable approach to data quality

<http://activeprime.com/cleanenter/sod/>

Who: A Few ActivePrime Customers



ActivePrime has 6000 users, in 40 countries, representing 15 languages.

For instance:

- Administaff had an Oracle search application for their CRM On Demand system, but the search wasn't able to identify duplicates. The search wasn't fuzzy search.
- Administaff bought ActivePrime to enable their users to quickly search and find records, even with misspellings, nicknames, and various alternative spellings of global 2000 company names.
- Today Administaff benefits by rapidly finding all the relevant data in their system and can leverage ActivePrime products to then keep their CRM On Demand system clean, usable and their staff very efficient. No more time spent trying to find data in the system ActivePrime does that for them automatically!



Customers: Value

- More efficient sales reps
- Better (accurate) analytics
- Effective integration
- Less waste
- Perception of competence



Example: Dirty Data

- MA
 - Mass
 - Massachusetts misspelled many ways
 - Massachusetts
-
- What were my sales in Q2 in MA?

This is very important (handling imprecise data) because... say you want to know what were my sales in Q2 in the state of Massachusetts. What would your query to the system look like? A mess!



Insight: Dirt Everywhere

- CRM
- ERP
- SCM
- On Premise and On Demand
- Etc...



Performance: Essential!

- Millions records
- Response time in seconds
 - ✓ Search engine grade speed
- No impact on application performance
- User adoption



Engineering: Artificial Intelligence (AI)

- Imprecision (Prune Search Space)
 - Context (Domain Knowledge)
 - Rules (Expert Systems)
-
- Think different – Think AI

Detailed discussion if time permits

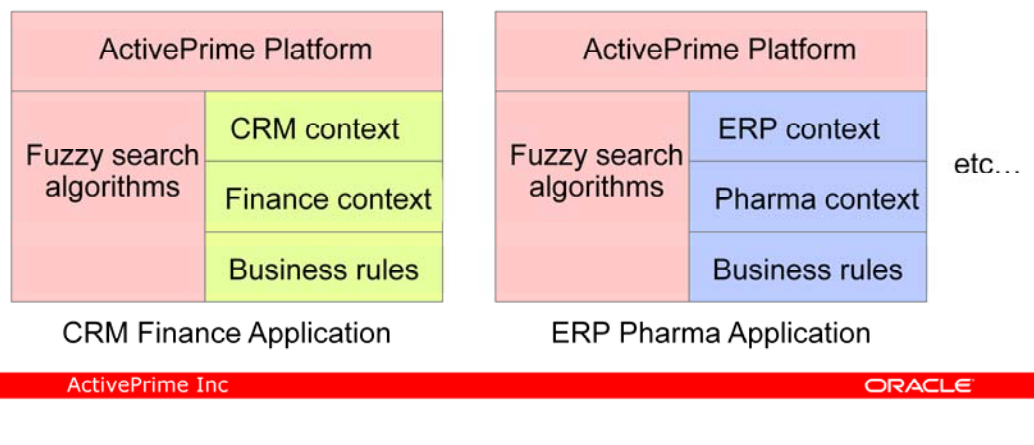
ActivePrime Inc

ORACLE

See last four slides for details.

Platform: Future Enabled

- Leverage investments
- Across applications and industries



Building systems from the AI perspective allows one to leverage investments in a platform that can work across applications and across industries.

The ActivePrime platform stays the same across applications and industries.

The application level context, industry context, and business rules adjust accordingly.



Future: Growth

- Immediate
 - ✓ Grow user base via current/new CRM On Demand
- New Products
 - ✓ Expand products to other Oracle CRM systems
 - ✓ Develop new apps for ERP, SCM and others
 - ✓ OEM the ActivePrime platform

ActivePrime Inc

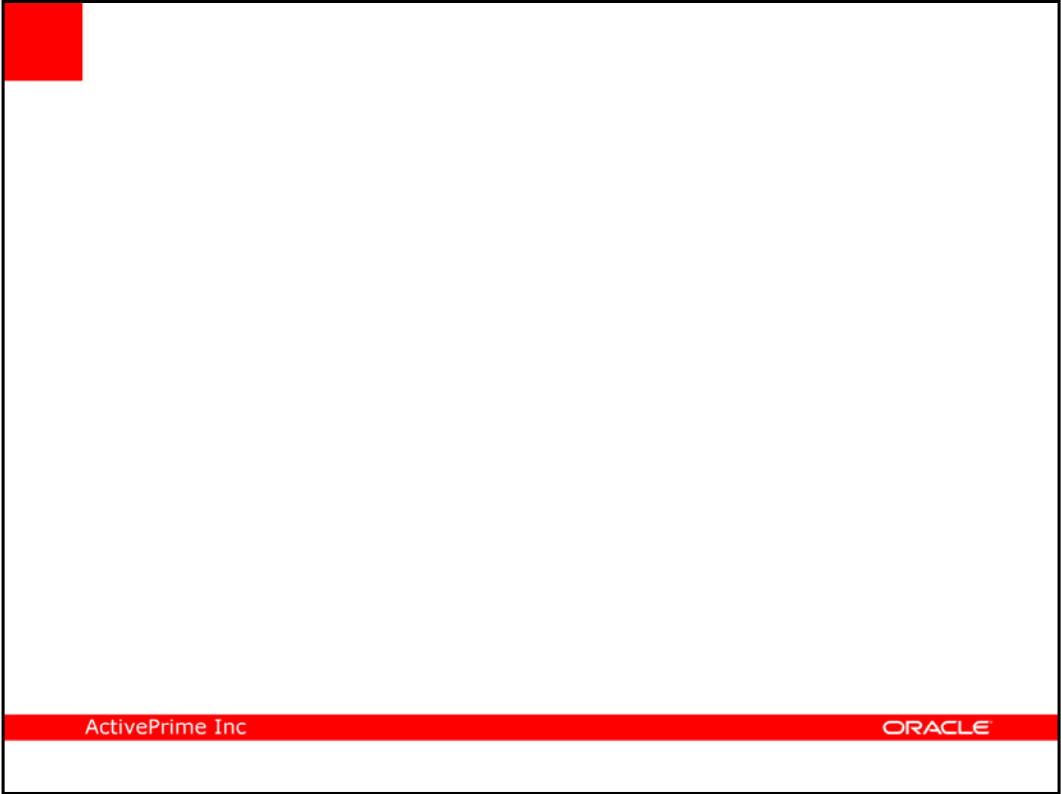
ORACLE

Immediate revenue growth

- Grow into existing and new user base of Oracle CRM On Demand by focusing on product and process innovations for ActivePrime Search, CleanEnter, CleanCRM and CleanMove.

New product derived growth

- Expand existing ActivePrime products into other Oracle CRM systems.
- Develop new applications for other Oracle products including ERP, SCM, and others.
- OEM the ActivePrime platform via Web Services





Case Study: Oracle

- Josh Saffren
- Oracle Application Sales Manager
- Oracle CRM On Demand
- User of CleanEnter and CleanCRM



Problem: Walking in Mud

- Working with dirty data is like ...
- ... selling while walking in mud
- Old and out of date data
- Inaccurate contacts
- Many duplicates



Goal: Increase Revenue

- Scale better
- Cover more territory
- Meaningful calls and emails



Ideal: Simplicity

- Easy to use
- Seamless integration
- Looks like CRM On Demand
- In the cloud

ActivePrime Inc

ORACLE



Solution: ActivePrime

- Great results
- Integrated and easy-to-use
- CleanCRM for initial cleanup
- CleanCRM for ongoing uploads
- CleanEnter for manual entry

ActivePrime Inc

ORACLE



Results: Increased Revenue

- Cover more customers & prospects
- ... with fewer people
- Increased team revenue!



Solutions: ActivePrime Products

- ActivePrime Search
 - ✓ Fuzzy enterprise search engine
- CleanEnter
 - ✓ Real-time dedupe and lead conv.
- CleanCRM
 - ✓ Batch cleansing
- CleanMove
 - ✓ Migration services



Thank You!

ActivePrime Inc

ORACLE



How: Three Dimensions

- Imprecision
- Context
- Rules

ActivePrime Inc

ORACLE

Here is the new, novel way to think about data quality.

How does one deliver search engine centric data quality solutions that have very high performance?

Three dimensions go hand-in-hand when building such solutions: imprecision, context, and heuristics.

Can't have an effective solution without all three.

Can't have broadly applicable solution without all three.



How: Imprecision

- Massachusetts and Nasachusets
- How close are they?
- Large volume (indexing)

ActivePrime Inc

ORACLE

Lots of small problems (for example, Massachusetts and Nasachusets being close, but not the same) result in a huge problems for companies. The problem grows exponentially with volume. The issue IS the volume!



How: Context

- An example (USA address context)
 - Mass means Massachusetts
- Mass has 20+ meanings (wikipedia)
- Multiple levels
 - Application
 - Industry
 - Corporate

ActivePrime Inc

ORACLE

Context narrows the meaning down to the appropriate one (of many).

There are multiple levels of context.

1. Application level. For instance, working with a CRM system we know that the data is “people centric”.
2. Industry level. For instance, being in the finance industry provides context. For example, the meaning of SEC.
3. Corporate level. For instance, your company may have specific abbreviations for your product names.



How: Rules

- Apply appropriate processing
- Example
 - If State, apply address context
 - If Account, apply business context
 - Etc...

ActivePrime Inc

ORACLE

Business rules require context specific processing. How to perform processing of imprecise data, given the context.