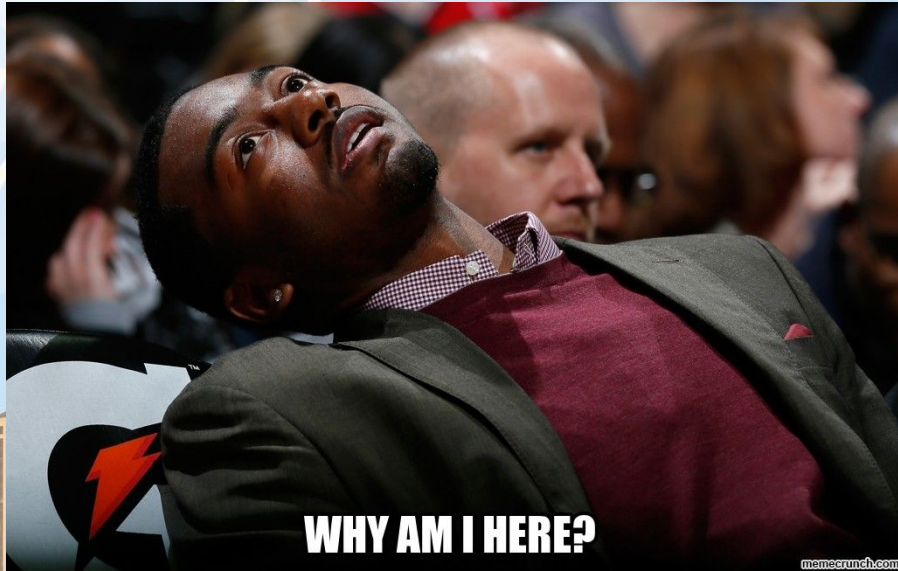


Open Source Data Mapping based on RESO Standards



2017  **RESO™ TECHNOLOGY SUMMIT**
ORCHESTRATING COLLABORATION & INNOVATION

WHY AM I STILL HERE?



Agenda

- How realtor.com maps your data today
- Modernizing our Listing Processing Platform
 - Logic & Workflow for our new system
- How we leveraged RESO's Data Dictionary to reduce implementation times
- Discuss & Demo an open source ETL Tool that anyone can use

Speaker

Rick Herrera

Director, Data Services

realtor.com

Rick.Herrera@realtor.com

Speaker

Alex Makhanko

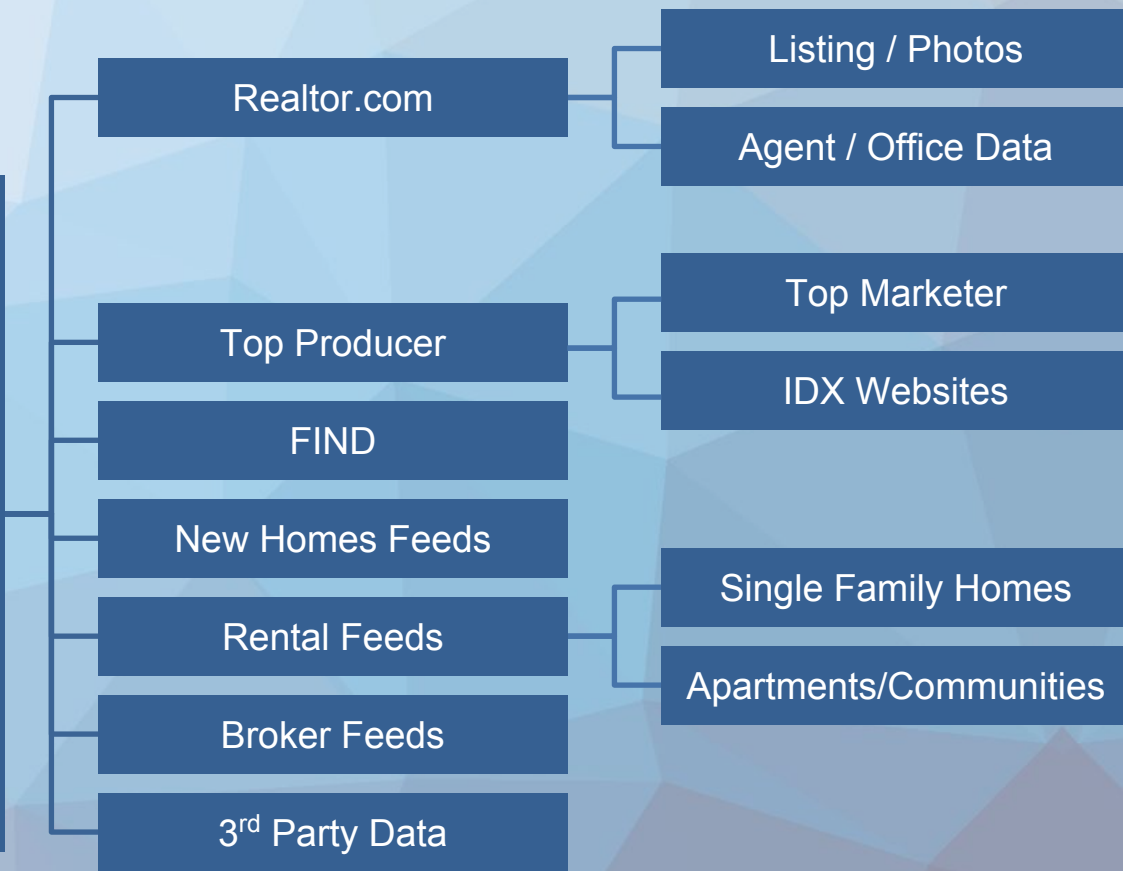
MLS Engineering Mgr

[realtor.com](https://www.realtor.com)

Alex.Makhanko@realtor.com

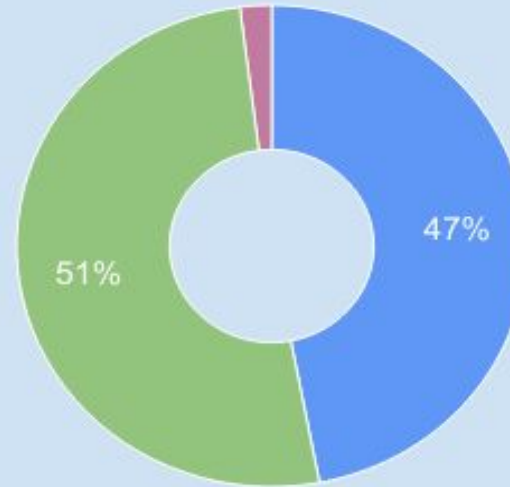
Our Responsibilities

Data Serv. Team



Update Frequencies

Update Frequencies



● 1 Minute Updates ● 15 Minute Updates ● Other

Current State

- Property Type based mapping

MLSs	Data Maps	Total Maps
750	4-6	3,750

Current State

- Property Type based mapping
- Maintenance is rough...

MLSs	Data Maps	Total Maps
750	4-6	3,750
		5,000

Current State

- Property Type based mapping
- Maintenance is rough
- Setup takes too long
- Doesn't leverage RESO Data Dictionary well

MLSs	Data Maps	Total Maps	Unique Fields	Bus Rules
750	4-6	3,750	33,000	24,000
		5,000		

The Time Is Now!

“The best time to plant a tree was 20 years ago.
The second best time is now.”

-Chinese Proverb

Objectives:

- Turn Days into Hours(!)
- Consolidate rules / simplify
- Cloud-based (AWS)
- Deeper understanding of data
 - Empower Search

WHAT WE'VE ACHIEVED SO FAR

- Reduced New implementation Time-to-Market by 70%!
- Leveraged RESO DD to make that happen
- Consolidated 24,000+ mapping rules down to ~ 200 Global rules
- All on an Open Source Cloud-based platform that anyone can duplicate

The Right Tool for the Job



EMC²



Nasdaq



CAT®



MOODY'S



NetApp™



4SIGHT
BUSINESS INTELLIGENCE INC.



DELTA DENTAL®



TERADATA



LOMA LINDA UNIVERSITY
School of Medicine

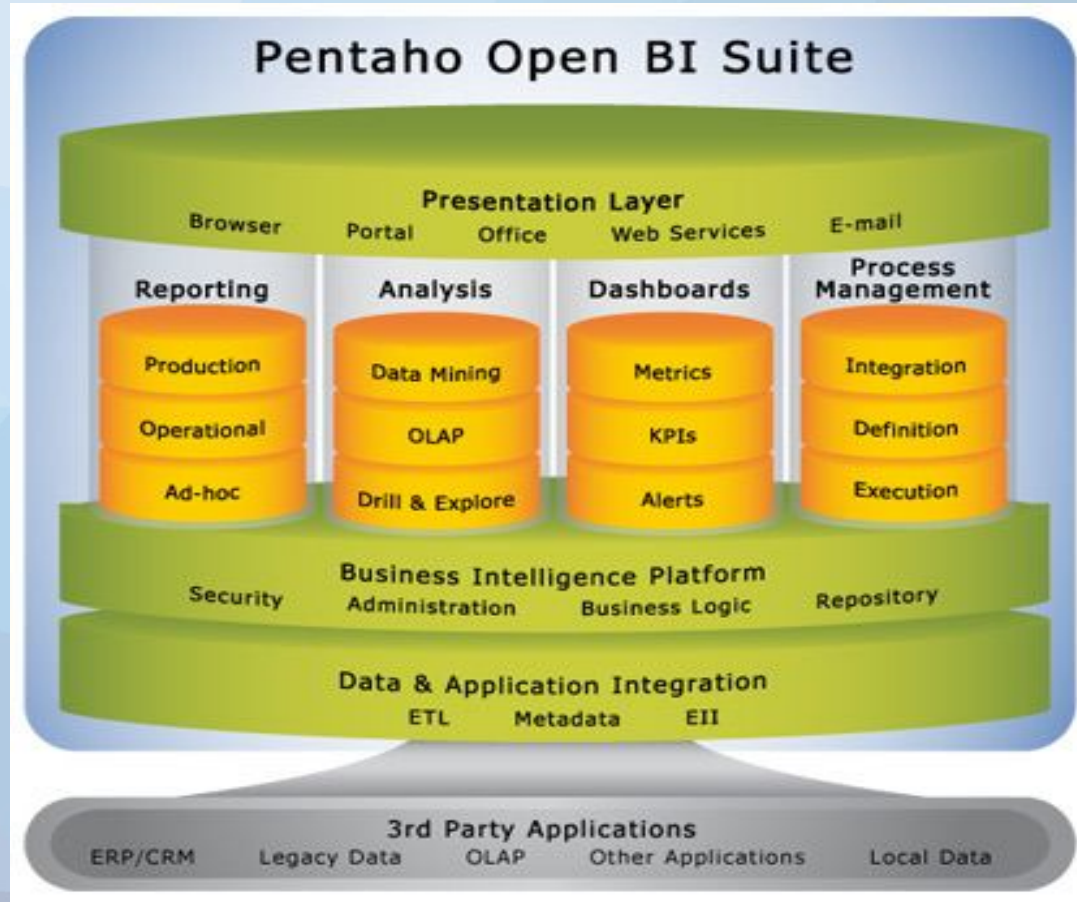


The University of Oklahoma
TULSA
SCHUSTERMAN CENTER



The New York Times

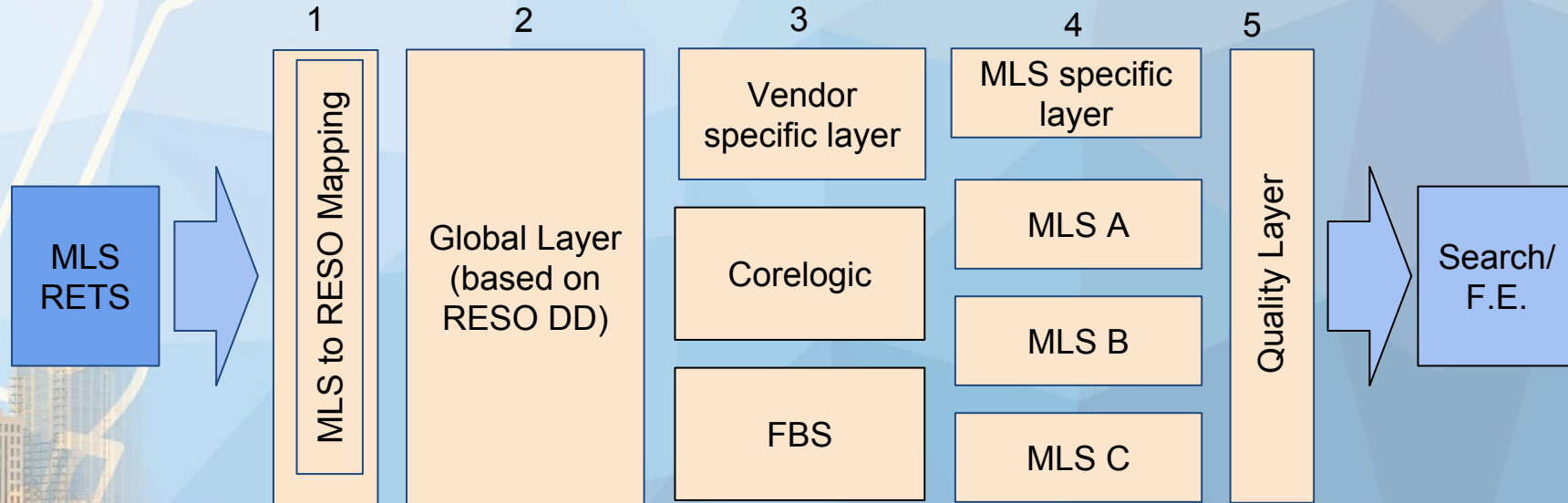
What is Pentaho



What is Pentaho

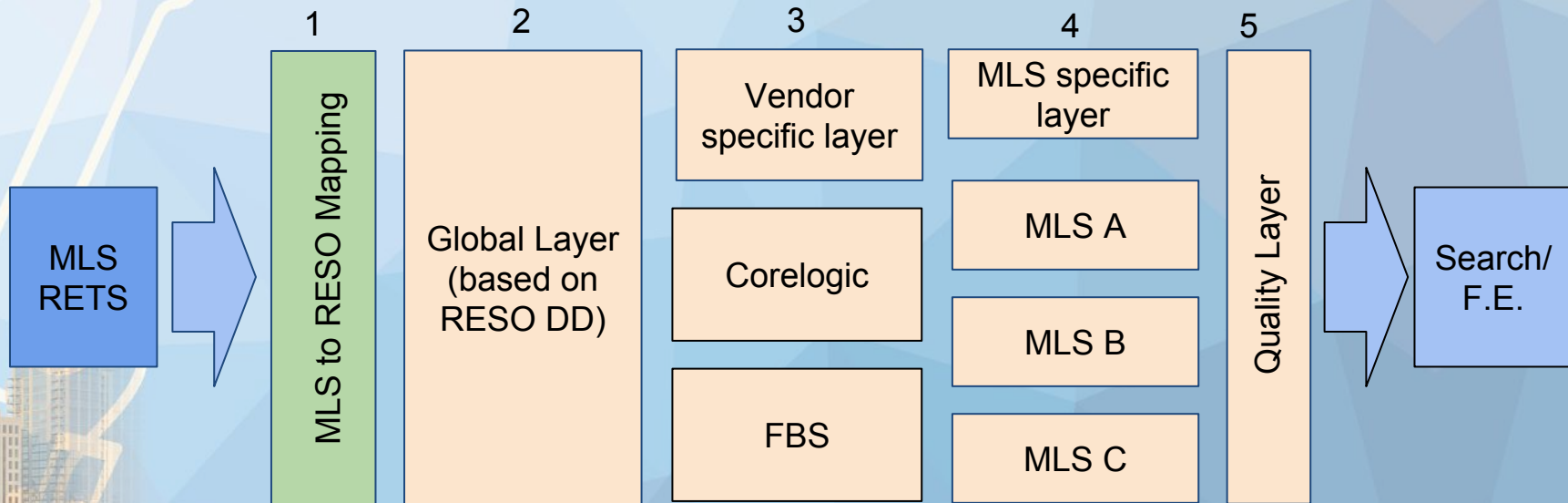
- Open Source ETL Tool
- Connectors to support many data sources (extensible)
- Incredible set of B.I. tools
- Everyone's an Analyst - Data Sparks Ideas
- Support for Cloud and big data technologies
 - AWS, ElasticSearch, Docker, Hadoop

Basic Workflow



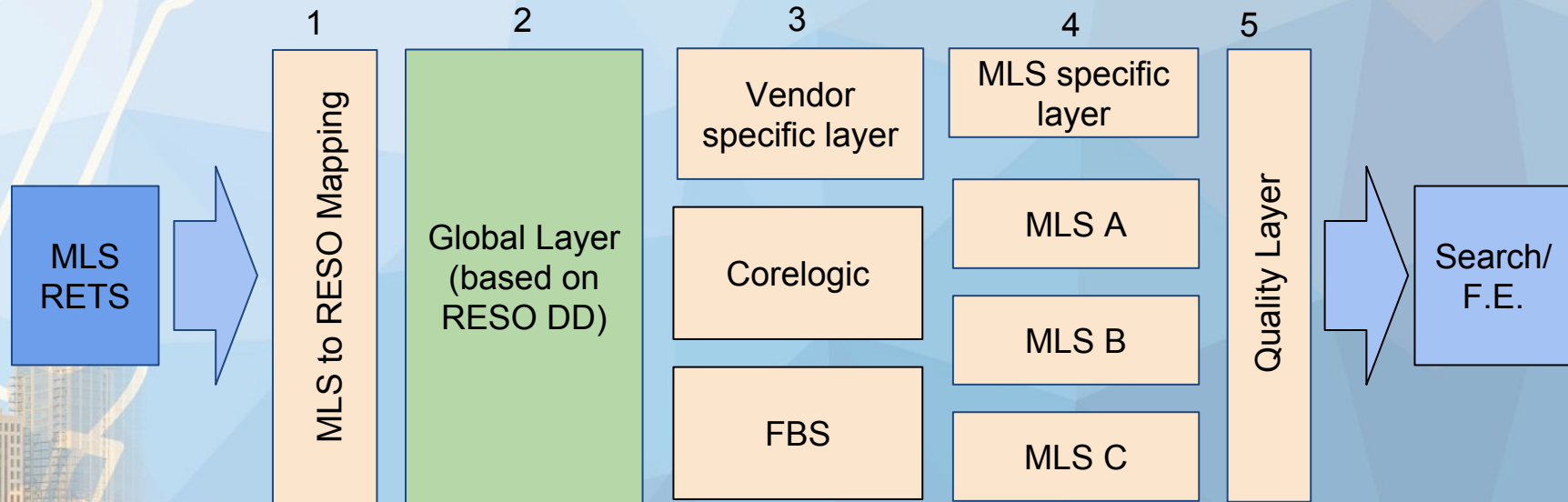
- Translate MLS data field names to Data Dictionary
- RESO compliant MLSs greatly reduce reconvert effort

Step One: Match RESO DD!



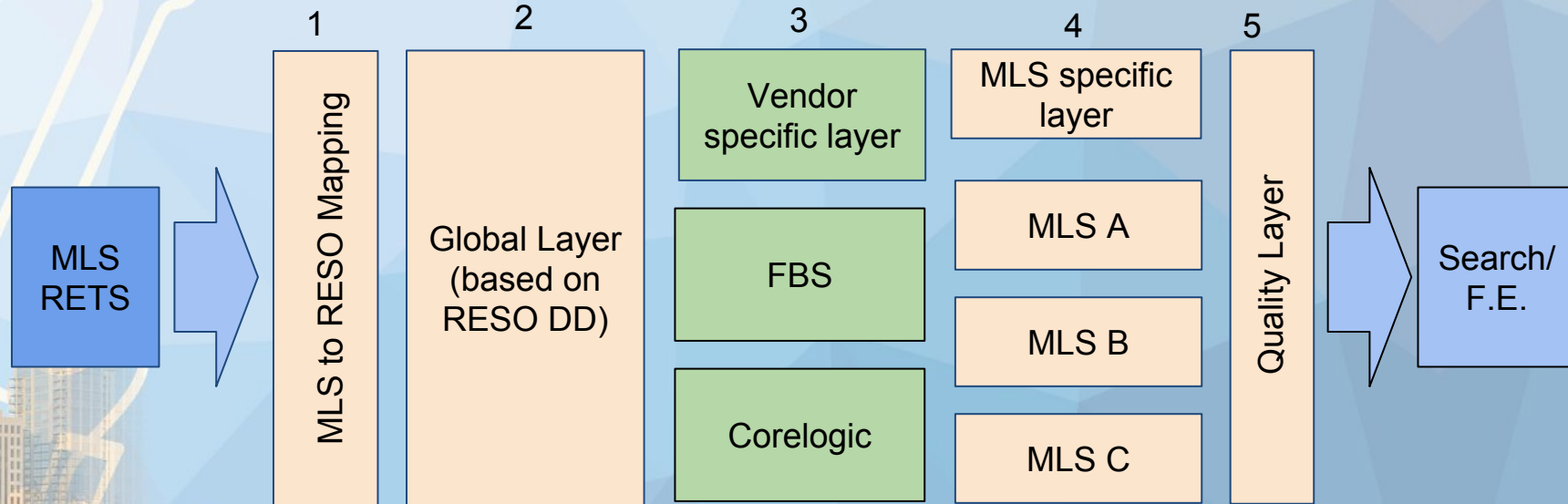
- Translate MLS data field names to Data Dictionary
- RESO compliant MLSs greatly reduce reconvert effort

Step Two: Global Rules



- Data typing, formatting, ending zeros, cleansing...
- RESO compliant MLSs greatly reduce reconvert effort
- Reduced rules from 36,000 to <200!!

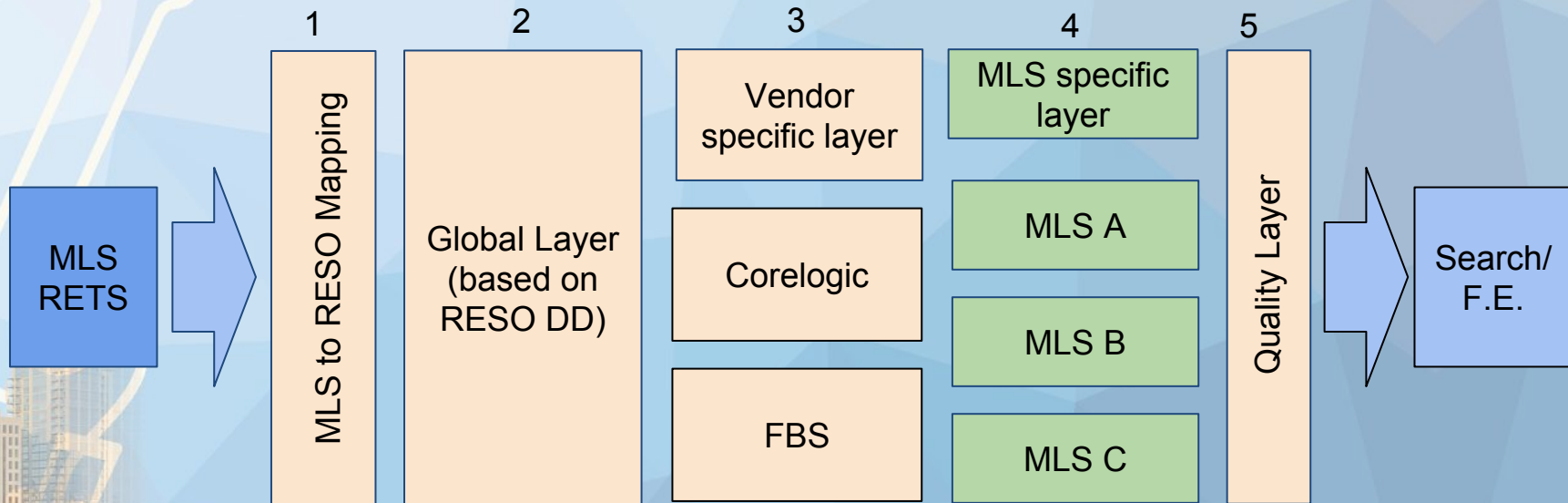
Vendor Specific Rules



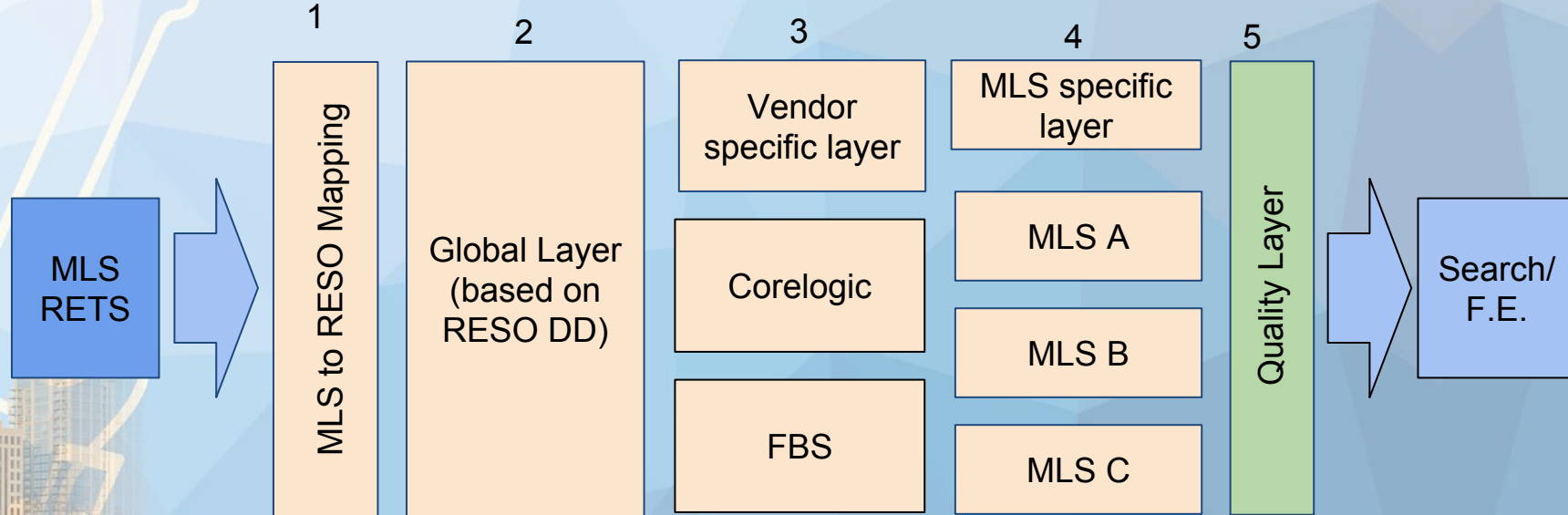
Examples:

- FBS - extended remarks in a separate class
- Corelogic/Matrix - room and unit information in a separate class
- Easy to maintain when vendors make changes

MLS Specific Transformation



Data Quality Layer

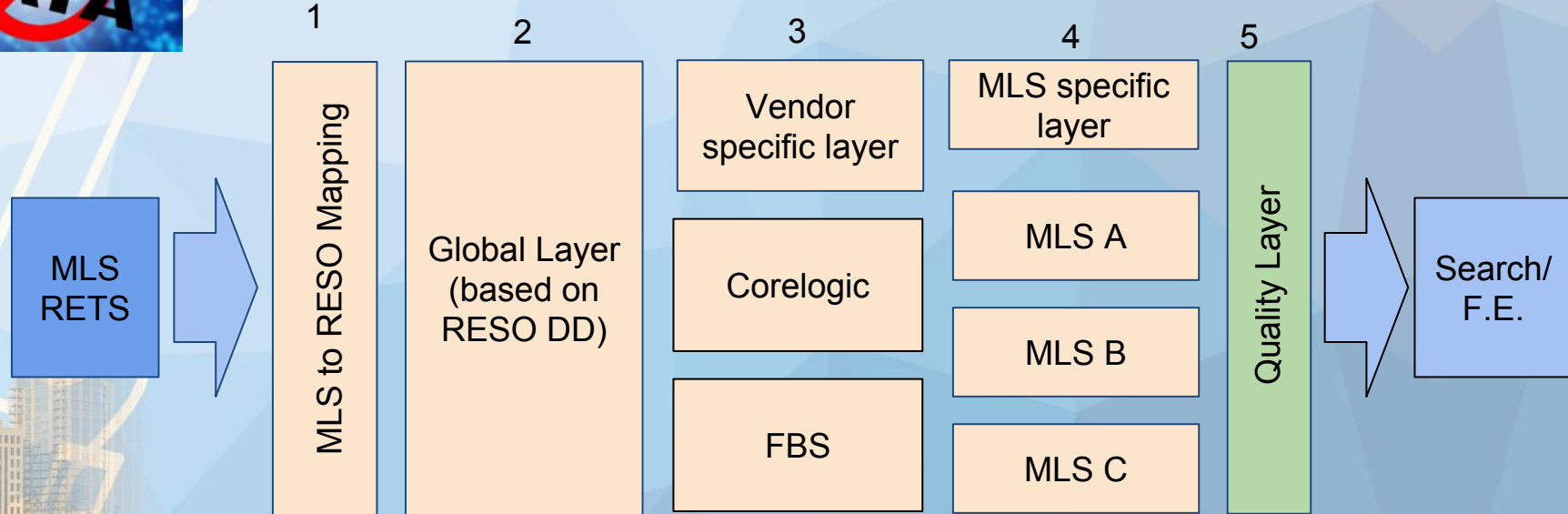


Data Quality Layer





Data Quality Layer



- Series of algorithms that seek anomalies and trends
- Capture questionable records
- Prevent bad data from infecting downstream systems

DEMO: From Days to Hours



<https://meet.google.com/rzz-vsby-zty>

DEMO: From Days to Hours

- Demonstrate transformations and data flow for a RESO platinum certified MLS using Pentaho Open Source ETL
- Input:
 - MLS data sample data from CRMLS
 - RETS client and Datasource have been pre-setup
- Process
 - Local instance of Pentaho ETL tool called Kettle
 - Application of RESO standard mapping
 - Application of all rules
- Output:
 - Processed data
 - Simulated feature breakdown
- <https://meet.google.com/rzz-vsbi-zty>



**Don't forget to take the session survey in the
conference app!**