


RECOMMENDATION FOR 123COMPANY  
(A REAL ESTATE CORP) TO UTILIZE  
ADVANCED ANALYTICS

**Bao Nguyen**

November 2017

*MY THOUGHTS ON ENGINEERING AND DATA MONETIZATION FOR A  
REAL ESTATE COMPANY TO HELP MODERNIZE THEIR ANALYTICS  
TECH STACK TO MAXIMIZE REVENUE AND PROFITS*

# CONTENT


- ▶ Objective Statement
  - ▶ Execution Plan Summary
  - ▶ Execution Plan Overview
  - ▶ Real Estate Advanced Analytics Opportunities
  - ▶ Real Estate Analytics Implementation Concepts
  - ▶ Conclusion
- 

# OBJECTIVE STATEMENT

Integration of end-to-end real estate processes with financial insights and advanced industry leading tools to improve financial returns for consumers, agents, brokers, and I23COMPANY.

# EXECUTION PLAN SUMMARY

Execution Strategies to help fulfill objective statement:

- ▶ Build high-performance and efficient teams that can execute
  - ▶ Implement modern and efficient development processes and practices
  - ▶ Automate everything!
  - ▶ Foster collaborate via Project Management “art” & collaboration tools
  - ▶ Ensure on solid existing products, architecture and roadmap
  - ▶ Predictive and advanced analytics-based features for differentiation and high growth
  - ▶ Use distributed and API based architecture for rapid development
- 

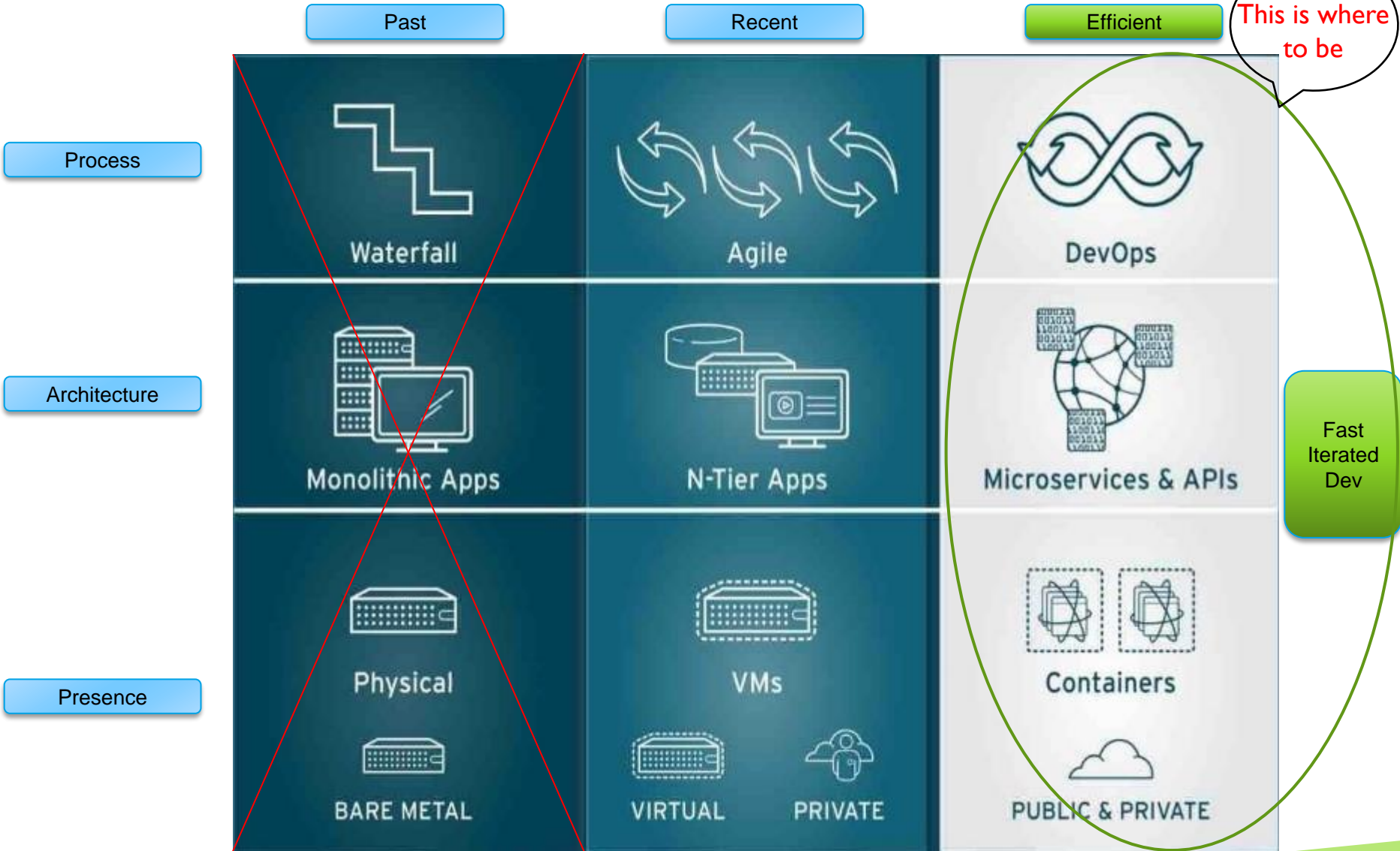
# EXECUTION PLAN OVERVIEW

# BUILDING HIGH PERFORMANCE TEAMS

Build teams that are collaborative, focus, and can execute

- ▶ Mentor and partner with key performers
- ▶ Appraise and reprimand teams in real-time to ensure lessons learned
- ▶ Understand and leverage team strengths and mentor their weaknesses
- ▶ Know and manage individual / team goals via sound project management
  - ▶ It is more efficient with clear and understood expectations
  - ▶ Track and measure goals for continuous improvement and communication
- ▶ Team scaling with:
  - ▶ Networks
  - ▶ Meetup groups – great venue to identify and possibly recruit talents
  - ▶ University and college recruiting events
  - ▶ LinkedIn contacts
- ▶ Team collaboration has to be the culture
- ▶ Turn over is normal part of business – plan and manage it!

# Modern Development Practices





# Modern Development Practices

Improving efficiencies, development speed, products expandability and integration with other technologies and cross functional department

- ▶ DevOp – Collaborate with Engineering to deploy, operational development, improve and operate systems / platforms
- ▶ API and distributed based architecture
  - ▶ Faster technology updates to adapt to fast changing markets
  - ▶ Quicker integration with newly acquired companies and technologies
- ▶ Hybrid cloud and private networks
  - ▶ Provides mobility, network scale, technology scale, faster development
  - ▶ Provides data security needed available on private networks
- ▶ Automate everything!
  - ▶ Avoid performing repeated processes and mundane tasks; Time is \$\$\$
  - ▶ Automate software deployment, configuration and management
  - ▶ Leverage machine learning for complex automation and process scaling
  - ▶ Reduce human errors and time can be used on more value-add tasks

# Collaboration and Efficiency Tools

For efficiencies, collaboration tools are need to share info, goals and improving communication among stakeholders

- ▶ **Project Management**
  - ▶ This is a must have tool as tasks need to be clearly understood and tracked
  - ▶ Project Management is a goal setting, measuring, and control tool
  - ▶ Focus on “the art” and discipline of project management
- ▶ Use info sharing and documentation tools, example, Confluence, Jira
- ▶ Use cloud collaboration tools such as Slack and video tools that spans desktop and mobile for real-time collaboration

# Popular Development Efficiencies Approach

- ▶ Componentized and APIs
- ▶ Hybrid cloud and hardware
- ▶ Continuous Integration and Continuous Delivery
- ▶ Agile and DevOps approach
- ▶ Source Control = Git / Stash
- ▶ Integration and Build = Jenkins, Java, Maven, rpmbuild
- ▶ Packaging = Jar, RPM
- ▶ Repository = Nexus (jar), CADS (rpm)
- ▶ Testing = TestNG, Junit, Jbehave
- ▶ Issues Tracking = Jira
- ▶ Deploy = Chef
- ▶ Provisioning = ORB (VM and Hardware)

# Solid Product Roadmap and Future Approach

Ensure existing product features and architecture are sound and leveraged as much as possible for short to mid-term growth

- ▶ Deep dive in both features and architecture are required to develop more details
- ▶ Implement product interfaces so that newly acquired products and services can be integrated and leveraged quickly
- ▶ Don't ignore basic existing product improvements, example
  - ▶ Are there more streamlining opportunities: Automation efficiencies?
  - ▶ Do existing products have enough automation for operational efficiency to reduce costs thus increasing financial performance for stakeholders?
  - ▶ Do existing products have rich reporting capabilities such as demographic trends, transaction financial performance, leads performance, etc? The data should already exist for these purposes.
  - ▶ Can existing products integrate and expand with newly acquired products quickly?

# PREDICTIVE ANALYTICS OPPORTUNITIES IN REAL ESTATE

# PREDICTIVE ANALYTICS AND REAL ESTATE

There are many opportunities when mixing big data, predictive analytics and real-estate information

Use of predictive analytics, capabilities once only available to large brokers, are now possible to I23COMPANY clients – leveling the playing field

# PREDICTIVE ANALYTICS REAL ESTATE OPPORTUNITIES

With big data, analytics, and machine learning, here are some possibilities

- ▶ Real-time and more accurate appraisal allows
  - ▶ Realtors to price properties automatically, quicker and more accurately
  - ▶ Buyers and sellers can make more efficient offers and pricing
  - ▶ Lenders can reduce investment risks
- ▶ Local and relevant information collection automatically for agents
  - ▶ Saves time on info collecting and researching
  - ▶ Behavioral analytics can predict and match buyers/sellers to more relevant properties to improve probably of an actual transaction
- ▶ Future value prediction on properties
  - ▶ With proper data sources and machine learning modeling, this is a differentiator
  - ▶ This feature helps manage and reduce risks for all stakeholders
- ▶ Digital advertising
  - ▶ With user's behavioral analytics and real-estate analytics, accurate and targeted advertising can be lucrative and act as another competitive differentiator

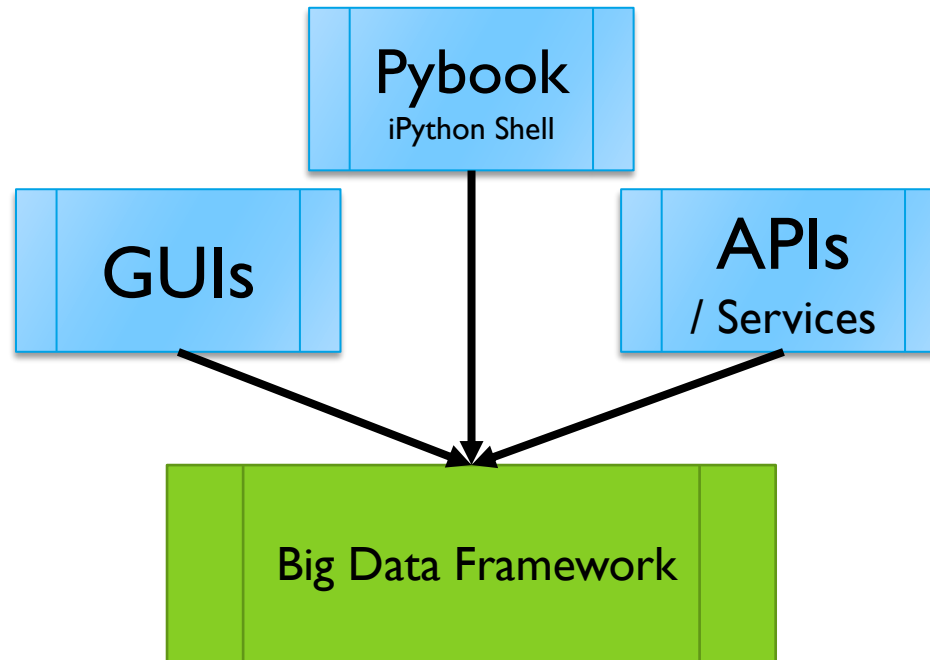
# PREDICTIVE ANALYTICS IMPLEMENTATION OVERVIEW

- ▶ Data acquisition
  - ▶ Mine internal existing data
  - ▶ Purchase data from 3<sup>rd</sup> party such as Online Residential; behavioral data is plentiful for purchase from many online advertisers
  - ▶ Generate differentiating data via mobile apps SDKs and marketing websites
- ▶ Data management aggregation distribution platform
  - ▶ Fairly straight forward to implement or lease
  - ▶ Use cloud services such as AWS to not reinvent the wheels
  - ▶ Solid data management upfront will save a lot of costs and time later on
- ▶ Domain expertise and data scientists
  - ▶ To evaluate market and features such as in previous slide
  - ▶ To unlock insights in data to seed data for machine learning
- ▶ Establish a rapid analytics framework with interfaces for quick experiment and development – see next slide example



# RAPID ANALYTICS DATA INSIGHTS APPROACH

Framework for rapid experiment and unlocking data insights



**ANALYTICS STARTS WITH DATA**

# FOCUS AREAS

- ▶ On having the right data, not necessary a lot of data
- ▶ On right features selection
- ▶ On solid training data
- ▶ On data model & model validation
- ▶ On hybrid cloud
- ▶ Do you need streaming analytics?
  - ▶ Customer interaction data in real-time?
  - ▶ Web logs?
  - ▶ Click events?
  - ▶ Mobile events data?

# DATA QUALITY UPFRONT IS KEY

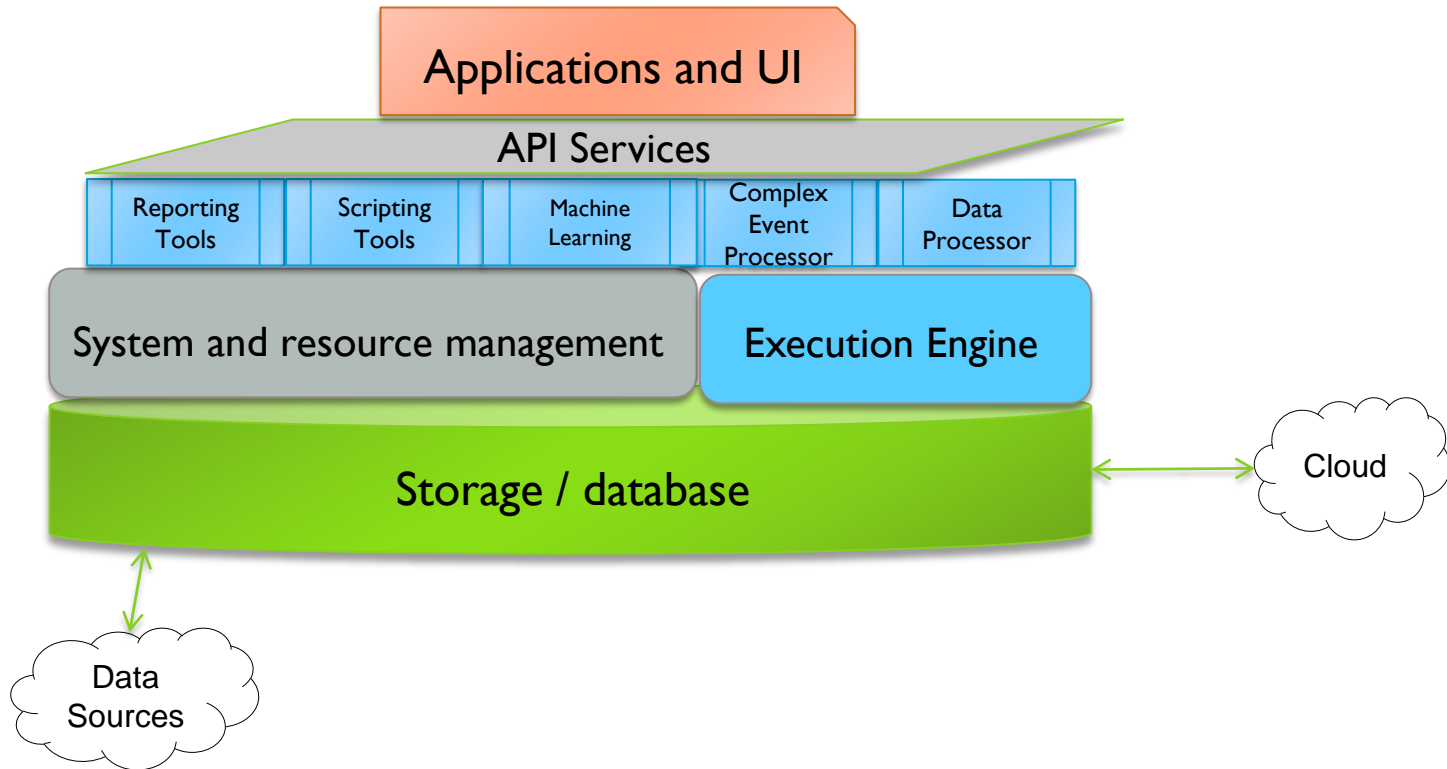
- ▶ Garbage-in garbage-out
- ▶ Invest in collecting quality & focus on having stable data sources
- ▶ Operationally **monitor** & manage data sources – this a common and costly overlook
- ▶ Do not use data without knowing the data!!
- ▶ Know the data and clean the data
- ▶ More data doesn't mean better; **relevant** data is key
- ▶ Does data latency matter? Need to understand that!
- ▶ **Automate** everything!
- ▶ **Huge costs and time later if the above are not done**

# DATA MANAGEMENT AND ANALYTICS PLATFORM (DMAP)

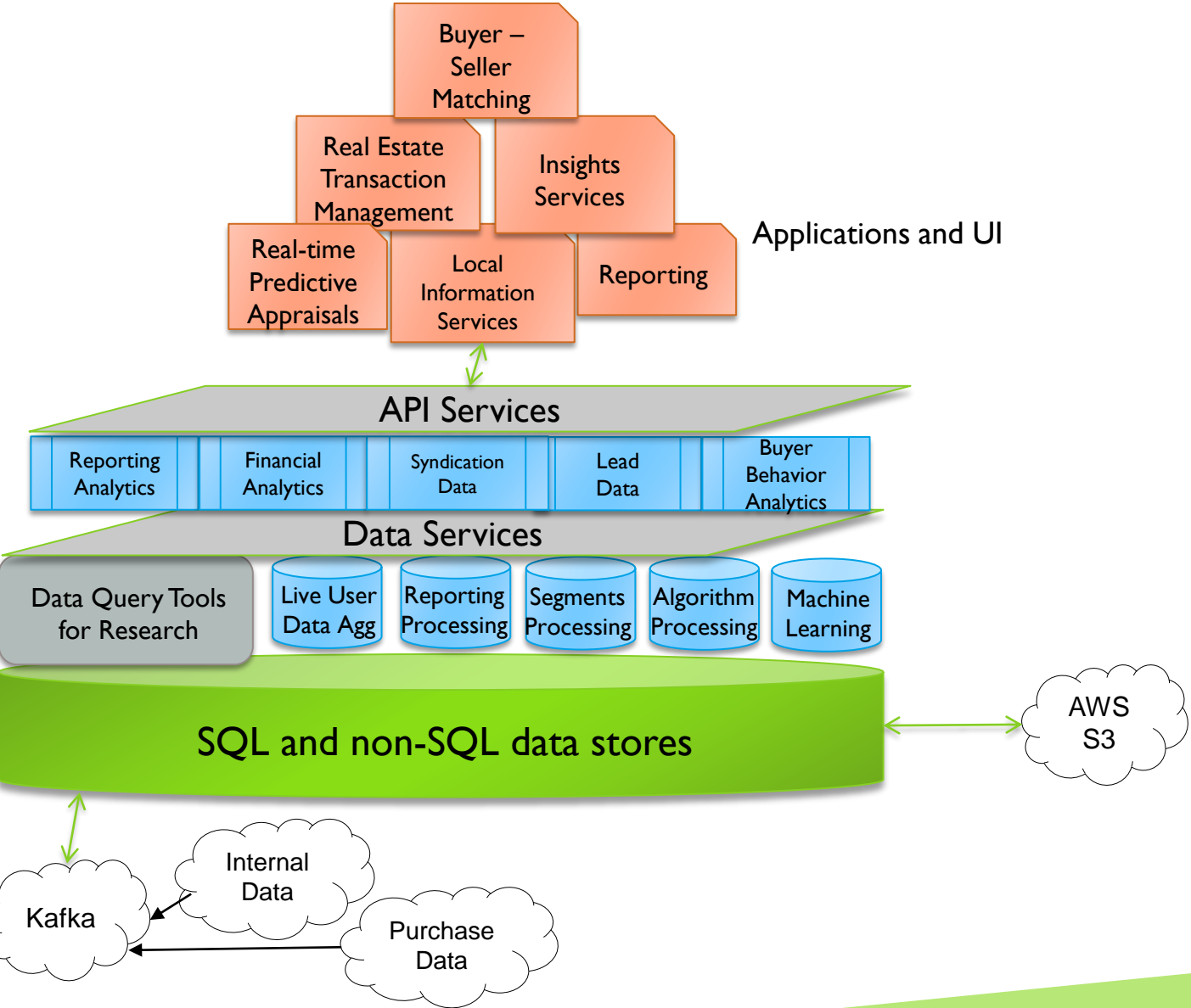
# THE DATA MANAGEMENT & ANALYTICS PLATFORM

- ▶ Two Functions: Data Platform and Data Analytics
  - ▶ Data Management
    - ▶ Handling of large amount of data from many data sources
    - ▶ Data ETL, normalization and distribution framework
    - ▶ A data exploration and analysis framework for Data Engineers
    - ▶ A data quality monitoring system and data system of records
    - ▶ A data modeling & enrichment framework – Data should always be improved
    - ▶ Data layer and API architecture to share data to internal and external consumers
    - ▶ A clean and controlled data hand-off point
    - ▶ Componentized and likely cloud based
  - ▶ Data Analytics and machine learning components
    - ▶ Getting insights from data sources
    - ▶ Use of algorithms and machine learning – stay simple
- ▶ The other framework: Data ingestion and data distribution

# DATA MANAGEMENT & ANALYTICS STACK



# REAL ESTATE ANALYTICS FRAMEWORK





# ANALYTICS TECHNOLOGY STACK CONSIDERATION

Function	Technologies
Platform	Apache Hadoop, .NET, Java
Data ETL	Kafka
Data Storage & Management	MapReduce, Hbase, Zookeeper, Cassandra, Couchbase
Real-time processing	Spark ML, Redis
Research	Impala, Spark ML, Octave, R, Python
Algorithms, Machine Learning, Real-time analytics	MapReduce, Mahout, Bayes, Random Forest, Spark ML
Data Warehousing, Query	AWS, Hive
Metadata, table management	MySQL
CI / CD Development Methodology	Jenkins, Docker, Ansible, Rundeck
Cluster Operational Management	Yarn
Data Serialization	Avro
Reporting	Vertica? Existing database

# REAL-TIME PREDICTIVE ANALYTICS

- ▶ Application consumes data from the Advanced Analytics Service Platform to generate offline analytics
  - ▶ Random Forest ensemble learning (multiple algos of both classifier as well as mod linear regression analysis), Naïve Bayes classifiers
- ▶ Real-time streaming analytics component was added to process online streaming live events
- ▶ Identifies and updates user's behavior characteristics in real-time for advertising campaigns to target – Spark MLlib
- ▶ System profiles more than 750M users
- ▶ In addition to Predictive Analytics revenue, additional monetization was driven by providing to 3<sup>rd</sup> party real-time analytics data. Clients often add private data, generating even more revenues
- ▶ Sounds complicated but not actually

# POTENTIAL ALGORITHMS USAGE

# TYPES OF ALGORITHMS TO CONSIDER

Use algorithms when scale is needed or process is not deterministic – Start simple & experiment, perhaps with Bayes

- ▶ Classification – Organize records by group types
  - ▶ Binary classification: simple and effective for 0 / 1 class
- ▶ Clustering – Organize records based on similarity
- ▶ Association learning – Learn what often appears with what
  - ▶ Logistics Regression and Linear Regression
- ▶ Prediction – Advanced analysis to gain insights in scale
  - ▶ Regression analysis with data scientist's crafted seed data
  - ▶ Linear regression algorithm

# OPEN SOURCE MACHINE LEARNING OPTIONS

- ▶ Amazon – Deep Scalable Sparse Tensor (DSSTNE)
- ▶ Google – TensorFlow: ML library
- ▶ IBM – SystemML: ML platform
- ▶ MicroSoft – Distributed Machine Learning Toolkit (DMTK)
- ▶ Facebook – Torch: ML deep learning algorithms
- ▶ Yahoo – CaffeOnSpark: ML library

# CONCLUSION

- ▶ There are many opportunities for I23COMPANY to integrate with the FNF acquired companies with proper data integration and APIs. This should significantly increase I23COMPANY market leadership and revenue. Streamlining real estate processes, automation, and providing reporting should also improve I23COMPANY products
- ▶ Additionally, there are opportunities in real estate advance analytics to add products and features such as real-time appraisal, automatic of matching of buyers and sellers, and digital advertising that are not only differentiators but also opportunities to grow I23COMPANY significantly