Data and Analytics Architecture
Are you Keeping Up?

Marina Kerbel, Principal at UpOnData

November 9, 2017

Introduction

- **Why Do Architects Need New Skills?**
  - Unprecedented pace of changes in the industry creates confusion and can make architecture assumptions outdated in a matter of months.
  - Technical savvy business areas leverage emerging capabilities and develop their own applications that need to be integrated with other systems.
  - Agile teams require 'hands-on' architects or they make architects irrelevant.

- **New Learning Approaches are Needed**
  - It is very challenging to keep up with changes, remain credible and avoid 'accidental architecture'.
  - New learning approaches are needed.
  - This presentation shows how the requirements for data skills are changing and how a new online learning ecosystem can help you gain the right skills at the right time.

UpOnData

Data and Analytics Market is Exploding

Example - Big Data Landscape 2017

The term "Big Data" continues to gradually fade away, but the Big Data space itself is booming.
http://mattturck.com/wp-content/uploads/2017/05/Matt-Turck-FirstMark-2017-Big-Data-Landscape.png

Frequent Changes Make Past Knowledge Obsolete

Examples - Common Myths about Big Data based on Outdated Information

1. Big Data = HDFS + MapReduce

2. Big Data = Unstructured Data

3. Big Data = No SQL

4. Big Data is Batch Only

5. Big Data is Cheap

UpOnData

Requirements for Data Skills are Changing

Snapshot of Twin Cities Job Postings on October 2017
Big Data Architects Requirements

UpOnData

Requirements for Data Skills are Changing

Snapshot of Twin Cities Job Postings on October 2017
Big Data Architect Requirements - Observations

SQL is "a must", thumps NoSQL

UpOnData

Snapshot of Twin Cities Job Postings on October 2017
Big Data Architect Requirements - Observations

Programming is "a must": Scala, Python, Java

Requirements for Data Skills are Changing

Snapshot of Twin Cities Job Postings on October 2017
Big Data Architect Requirements - Observations

HDFS and Map Reduce are not as prominent as Hadoop, Spark and other requirements

UpOnData

Requirements for Data Skills are Changing

Snapshot of Twin Cities Job Postings on October 2017
Big Data Architect Requirements - Observations

Deep Learning and Machine Learning start showing as requirements
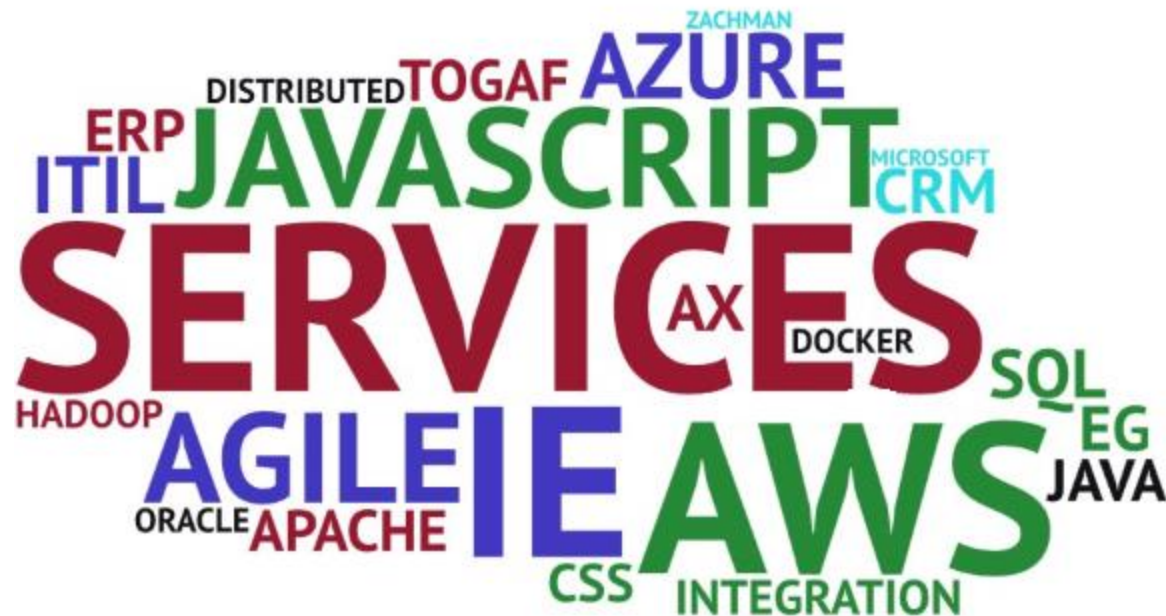
UpOnData

Snapshot of Twin Cities Job Postings on October 2017
Big Data Architect Requirements - Observations
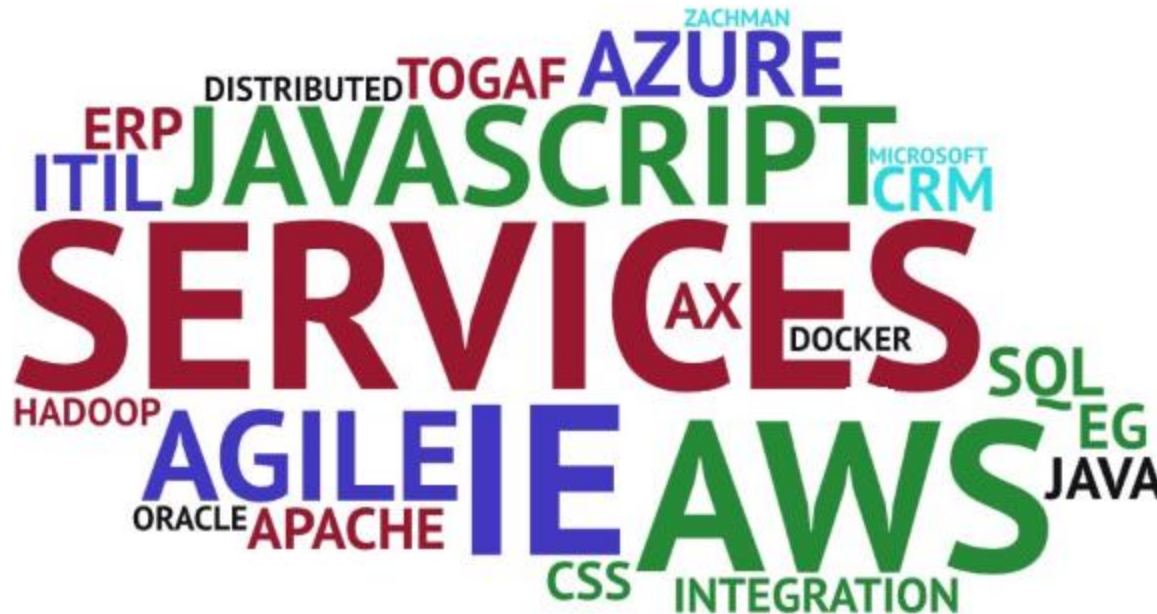

Cloud requirements are rare for Data Architects (yet!)

UpOnData

Data Skills are Relevant to Other Domains

Snapshot of Twin Cities Job Postings on October 2017
Solution Architects Requirements

Snapshot of Twin Cities Job Postings on October 2017
**Solutions Architect Requirements - Observations**

Data skills such as Apache, Hadoop, SQL appears in about 20% of jobs postings.

UpOnData

Many Data Skills are Transferable, Some are New

Data Skills	Traditional => Modern (green – *almost* same, yellow – extendable, red – new)
1. **Data Analysis**	Data is still the KING – understanding and preparing data was taking 70-80% of all data work 10+ years ago. The number stays the same today despite advances in discovery, profiling, visualization.
2. **SQL**	Almost all new data tools use SQL in some form. Demand for SQL skills is only increasing.
3. **Data Modeling**	Modeling rules are relaxed with NoSQL and 'Schema on Read', but there is no silver bullet. Data still needs associated structure to be understood. *Requires more collaboration with business & developers.*
4. **Programming**	Programming is similar to what it was years ago. Coding is a more recent requirements for data architects, particularly for *object-oriented languages such as Java or Python.*
5. **Data Warehouse / Data Lake Architecture**	Traditional DW & Data Lake have similar pipelines: from data sources to analytics & reporting. They get more and more alike as architecture patterns and tools for both are expanding (e.g. streaming, noSQL for DW or Metadata & SQL for Data Lake). *Requires understanding of newer patterns and tools.*
6. **Data Integration & Reporting**	All traditional integration methods (ETL, EII, EAI /Messages, and Services) are used for big data. More focus on messaging, streaming, self-service, visualization. *Requires understanding of new tools.*
7. **Data Governance, Metadata, Security**	Similar needs as before: understand, clean, share and protect data. More difficult to implement for Big Data solutions as they are often more dynamic, lack predefined data structure and have course-grain data processing. *Requires understanding of new approaches and new security and metadata tools.*
8. **Administration**	Similar capabilities as before: configuration, logging, monitoring, deploying, etc. Increased focus on distributed data stores; starting Cloud. *Requires familiarity with new administration and DevOps tools.*
9. **A.I., Machine Learning, Deep Learning**	Previously developed independently by data scientists with advanced degrees, the capabilities become democratized and analytical products need to be integrated with core systems. *Requires closer partnership with data scientists and understanding data, infrastructure, tools, security, and integration needs for the entire lifecycle of model development, deployment and management.*

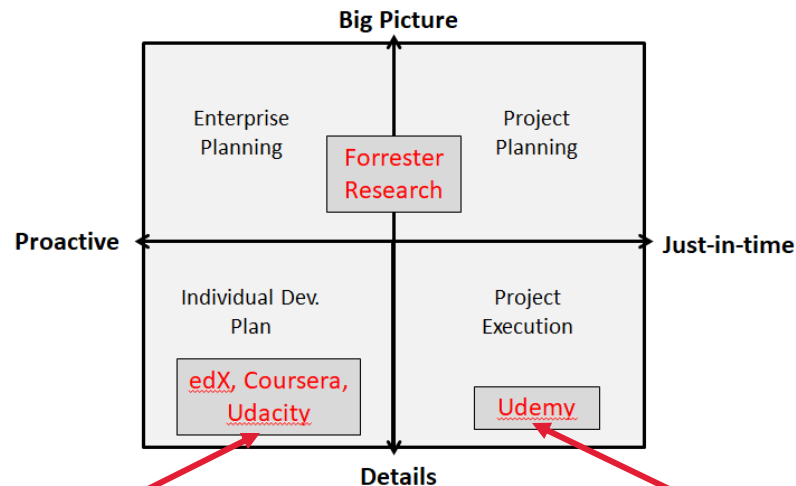UpOnData

Examples of Approaches for Online Learning
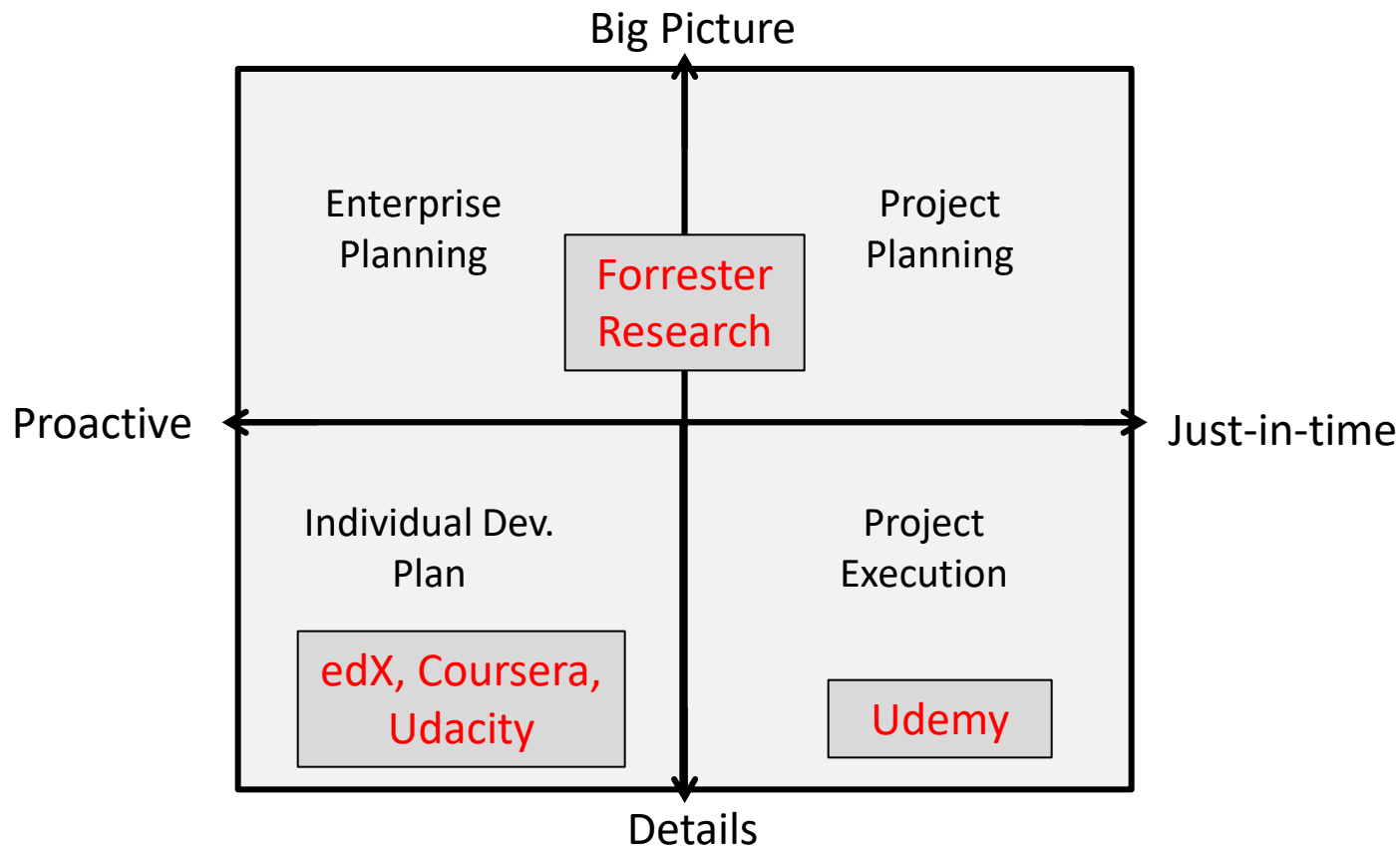
MOOC (Massive Open Online Course) is a relatively new education channel open to everyone via internet. EdX, Coursera, Udacity, and Udemy are examples of MOOC platforms.

New Learning Ecosystem Provides Many Options

edX, Coursera, Udacity

- Associated with top universities
- Structured, comprehensive, academic
- Self-paced or timed programs (up to 12 months)
- May require waiting until the program is available
- Udacity offers nano degree, Coursera offer specialization degree
- Cost – from free to $200/month or several thousand dollars per program

Examples of MOOC Platforms

Big Picture

Enterprise Planning	Project Planning

Forrester Research

Proactive ← → **Just-in-time**

Individual Dev. Plan	Project Execution

edX, Coursera, Udacity

Udemy

Details

Disclaimer – this information represent a point of view at this time. All MOOC platforms periodically change their approaches and pricing.

Udemy

- Anyone can be an instructor
- Courses are validated to ensure acceptable quality
- Practical, can be the first place to cover a new capability or tool
- Self-paced with lifetime access to class
- Instant access
- Cost – one time from free to $200 per course
- Often offers deep discounts (as low as $10 per class for past customers)

UpOnData

New Learning Ecosystem Provides Many Options

Suppose, you want to learn Deep Learning for a customer-centric transformational initiative. Here are examples of how you can use different learning options.

Forrester example

- Report on Deep Learning for clients

For Customer Insights Professionals

REPORT: Deep Learning: The Start Of An AI Revolution For Customer Insights Professionals

Continuous Improvement: The Customer Analytics Playbook

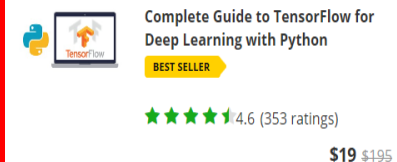May 12, 2017 | Brandon Purcell, Mike Gualtieri, Diego Lo Giudice

Coursera example

- Deep Learning 5-course Specialization (a few weeks for each course)
- Starts on Nov 7
- Cost $49 a month

Deep Learning
5-course Specialization
deeplearning.ai
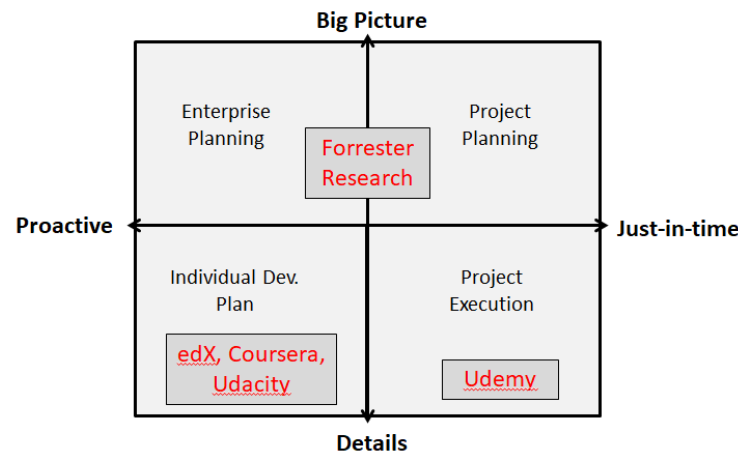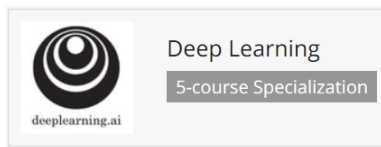
Udemy example

- Course - Complete Guide to TensorFlow
- 12 hours of videos + hands-on practice
- $195 full price ($19 promotional price)

Complete Guide to TensorFlow for Deep Learning with Python
BEST SELLER
★★★★½ 4.6 (353 ratings)
$19 $195

Big Picture

Enterprise Planning

Project Planning

Forrester Research

Proactive ←→ **Just-in-time**

Individual Dev. Plan

Project Execution

edX, Coursera, Udacity

Udemy

Details

UpOnData

Conclusion

- This is a good time to be a data expert.

- A lot of new and exciting changes are happening in the data field.

- The new learning ecosystem provides many different options to help us keep up with these changes.

UpOnData

Questions

Contact

- http://upondata.com/

- marina@upondata.com

UpOnData