EDA 3.0

Leon Stok, VP EDA, IBM
My.. IWLS history..

- IWLS 2003: May 28 - 30, 2003, Laguna Beach, California
- IWLS 2002: June 4 - 7, 2002, New Orleans, Louisiana
- IWLS 2001: June 12 - 15, 2001, Lake Tahoe, California

- IWLS 2000: May 31 - June 2, 2000, Dana Point, California, General Chair.
- IWLS 1999, Vice Chair., Lake Tahoe
- IWLS 1998, Program Chair, Lake Tahoe
- .....
- My First IWLS.. 1993, Lake Tahoe
Outline

- Introduction
- Little history lesson: EDA 1.0 and EDA 2.0
- Why EDA 3.0?
- What will EDA 3.0 look like?
- Who is going to build it?
- Conclusions and Discussion
Predictions

- The facts expressed here belong to everybody, the opinions to me. The distinction is yours to draw...

- Howard H. Stevenson writes that prediction in business "... is at least two things: Important and hard."[1]

- Past performance is no guarantee of future results.
Question

- Why does it take about 2 seconds to look up the shortest path from my home to JFK on my phone, and does it take a couple of hours to find the critical path in a large chip in an advanced technology on a sizeable server?
EDA 1.0

- Small Circuits
  - Circuit simulation
- Boards
  - Placement
  - Partitioning
  - Routing
- Schematics
  - Some Logic synthesis
  - Some equivalence checking
- Relatively small, higher complexity algorithms.
Moore’s Law

- Device density doubles every 12-18 months

Integrated Circuits
- Put EDA tools on the Moore’s law curve.
- Increasingly large design sizes
EDA 1.0 implementations

- Higher level languages
  - Lots more logic written
  - Logic synthesis became serious business.
- $O(n \log n)$ or $O(n \log^2 n)$
  - Spectral methods
  - Network flow
  - Approximations
- Data structures:
  - Quad trees, KD trees, BDD’s, ATPG engines
Discrete Design Flow

[Hathaway, Integrated, Incremental, and Modular EDA Tools, 1996]
Gate and Interconnect Scaling

- Gate Delay
- Sum of Delays, Al & SiO2
- Sum of Delays, Cu & Low $\kappa$
- Interconnect Delay, Al & SiO2
- Interconnect Delay, Cu & Low $\kappa$

Materials:
- Al: $3.0 \mu\Omega/cm$
- Cu: $1.7 \mu\Omega/cm$
- SiO2
- Low $\kappa$
- Al & Cu Line: $0.8 \mu$ Thick, $43 \mu$ Long

Graph:
- Delay (ps) vs. Generation (nm)
- Gate with Al & SiO2
- Gate with Cu & Low $\kappa$
- Gate
The Demise of Prediction

- **Prediction is becoming very difficult:**
  - The impact of later designs steps is increasing
  - New cost functions are becoming more important
  - Aggressive designs allow less guard-bandng
- **Design decisions interact**
  - The iterations not only became cumbersome and slow, but often did not even converge.

[Hathaway, Integrated, Incremental, and Modular EDA Tools, 1996]
EDA 2.0: The Integration Era

Increasing Automation/Integration

[Darringer, IBM]
Design Infrastructure in EDA 2.0

Skill, Python
1-2M lines

IP, model, lib: ~1M lines of tcl code
IO vs Computation in a real design flow

- Reports/Post Timing
- CPPR
- Timing
- Build Model
- Read Parasitics
- Read Assertions
- Load Design

Design loading speedups
Assertion loading speedups
Parallel reports and forked post-processing
Skill, Python
1-2M lines

Tcl

Control, 1-2M lines of tcl code
IP,model,lib: ~1M lines of tcl code
Design Environment

- **Designer**, I want to
  - get to my DATA.. from anywhere any place
  - the DATA be there without me waiting for it
  - analyze the DATA with whatever tools I can lay my hands on
  - know how to improve my design [data]
  - know how to get from A to B through my design process
  - be like……..

- Brayton / Cong 2009 NSF workshop, EDA Past, Present, Future
  - Intuitive, simplified and standardized design environments.
  - Scalable design methodologies
  - Predictable design flows
11 fascinating facts about Google Maps

- **1. How much data has Google Maps accumulated**
  - Combining satellite, aerial and street level imagery, Google Maps has over 20 petabytes of data, which is equal to approximately 21 million gigabytes, or around 20,500 terabytes

- **2. How often are the images updated?**
  - Depending on data availability, aerial and satellite images are updated every two weeks. Street View images are updated as quickly as possible, though Google wasn't able to offer specific schedules, due to its dependence on factors such as weather, driving conditions, etc.

- **8. In the history of Google Maps, how many Street View images have been taken?**
  - The Street View team has taken tens of millions of images since the Street View project began in 2007, and they've driven more than 5 million unique miles of road.

[http://mashable.com/2012/08/22/google-maps-facts/ ]
Design Data Volume

- 3Gb Oasis
- 1 Tb product engineering
- 5Tb test and diagnostics data
- 5km wire

Tweets/day
Big Data spans four dimensions: Volume, Velocity, Variety, and Veracity.

- **Volume**: Enterprises are awash with ever-growing data of all types, easily amassing terabytes—even petabytes—of information.
  - Turn 12 terabytes of Tweets created each day into improved product sentiment analysis
  - Convert 350 billion annual meter readings to better predict power consumption

- **Variety**: Big data is any type of data - structured and unstructured data such as text, sensor data, audio, video, click streams, log files and more. New insights are found when analyzing these data types together.
  - Monitor 100’s of live video feeds from surveillance cameras to target points of interest
  - Exploit the 80% data growth in images, video and documents to improve customer satisfaction

- **Veracity (aka Variability)**: 1 in 3 business leaders don’t trust the information they use to make decisions. How can you act upon information if you don’t trust it? Establishing trust in big data presents a huge challenge as the variety and number of sources grows.

- **Velocity**: Sometimes 2 minutes is too late. For time-sensitive processes such as catching fraud, big data must be used as it streams into your enterprise in order to maximize its value.
  - Scrutinize 5 million trade events created each day to identify potential fraud
  - Analyze 500 million daily call detail records in real-time to predict customer churn faster
EDA Evolution

- EDA 1.0
  - Point tools on individual workstations
- EDA 2.0
  - Integration of design tools on distributed servers
- EDA 3.0
  - Integrated Design Flow on very large Clusters

What will EDA 3.0 look like?
- Start to the improve the Data Analytics
- Learn from Big Data: other data and graph parallel systems
- Capitalize on the Changing nature of IT so we will be able to run this effectively.
Data Explosion

Capacity

Time

Design Data

Processed Information

Capacity of the Human Brain

Analytics and Visualization
YARN

- YARN: Yet Another Resource Negotiator
- Resource management and job tracking/scheduling got split up.

[tomsitpro.com]
Scaling Apache Giraph to a trillion edges

Four minutes
One Iteration of PageRank

https://www.facebook.com/notes/facebook-engineering/scaling-apache-giraph-to-a-trillion-edges/10151617006153920
YARN

BATCH, INTERACTIVE & REAL-TIME DATA ACCESS

YARN: Data Operating System
(Cluster Resource Management)

HDFS
(Hadoop Distributed File System)

[hortonworks.com]
Spark

- Transformations and Actions
- Lazy evaluation of transformations
  - Lineage graph
Spark Advantages

- Tight integration gives the ability
  - to build applications that seamlessly combine different processing models.
    - in Spark you can write one application that uses machine learning to classify data in real time as it is ingested from streaming sources.
    - Simultaneously, analysts can query the resulting data, also in real time, via SQL (e.g., to join the data with unstructured logfiles).
    - More sophisticated data engineers and data scientists can access the same data via the Python shell for ad hoc analysis.
    - Others might access the data in standalone batch applications.
- Second, the costs associated with running the stack are minimized, because instead of running 5–10 independent software systems, an organization needs to run only one. These costs include deployment, maintenance, testing, support, and others.
- Each time a new component is added to the Spark stack, every organization that uses Spark will immediately be able to try this new component.
- When Spark’s core engine adds an optimization, SQL and machine learning libraries automatically speed up as well.
- This changes the cost of trying out a new type of data analysis from downloading, deploying, and learning a new software project to upgrading Spark.
- All the while, the IT team has to maintain only one system.

[Learning Spark, by Holden Karau, Andy Konwinski, Patrick Wendell, and Matei Zaharia]
Spark and Graphx

The GraphX Stack

- PageRank (5)
- Connected Comp. (10)
- Shortest Path (10)
- SVD (40)
- ALS (40)
- K-core (51)
- Triangle Count (45)
- LDA (120)

Pregel API (28)

GraphX (3575)

Spark

Quadratic Programming Solver for Non-negative Matrix Factorization with Spark


©LS 2015
Where do we run this? Software Defined Network

- Software Defined Networking, Storage and Compute
WHO IS GOING TO BUILD THIS?
Sources of Innovation in EDA 2006

- **Corporate CAD**
  - Great source of contributions in the EDA1.0
  - Spawned many startups in EDA2.0
  - Overall slump in Corporate research labs.

- **University**
  - Lots of fresh, innovative ideas
  - Worked really well in EDA1.0
  - Created several of the mini-platforms e.g. Spice, MIS, CUBDD, Magic, Timberwolf
  - Started to erode in EDA2.0
    - Lack of access to leading edge integration infrastructure, leading edge designs, leading edge design problems.

- **Startups**
  - Usually based on “existing” invention.
    - Creates micro environment to rapidly turn this into innovation
    - Works really well in the EDA 1.0
  - Starts to erode in EDA 2.0
    - Have access to and focus on leading edge designs and problems
    - But spend their time:
      - 10% innovation
      - 90% creation of integrated infrastructure

- **BIG CAD**
  - Acquire significant portions of innovation from startups and 90% of their time re-creating the acquired technology with their infrastructure.

- **IP**
  - Usually based on “existing” invention.
  - Creates micro environment to rapidly turn this into innovation
  - Works really well in the EDA 1.0
  - Starts to erode in EDA 2.0
    - Have access to and focus on leading edge designs and problems
    - But spend their time:
      - 10% innovation
      - 90% creation of integrated infrastructure

[Walden Rhines, Sociology of Design and EDA, DATE2006]
Changing landscape

Corporate CAD

Research Funding

Academia

Start-Ups < $40 Million 300 + Co.

Small Public $40-150M 6-10 Co.

“Big-3” Rev. > $500 M

VC Funding $155M in 2005

Graduates & Ideas

Technology & People

Cash & Stock

Innovators Serial Startups

Graduates & Ideas

Research Funding

[Walden Rhines, Sociology of Design and EDA, DATE2006]
EDA Eco-system 2015

[Edward Walden Rhines, Sociology of Design and EDA, DATE2006]
EDA Eco-system 2015

Corporate CAD

Research Funding

Graduates & Ideas

Start ups
Web

Graduates
Ideas

Academia

Research Funding

Big 3

> 1.2B
Graph Databases

[Graph showing the popularity changes of various database types from Jan 2013 to Jan 2015. The X-axis represents the years from Jan 2013 to Jan 2015, and the Y-axis represents the popularity changes. The graph includes lines for Graph DBMS, Wide column stores, Document stores, RDF stores, Key-value stores, Search engines, Native XML DBMS, Object oriented DBMS, Multivalue DBMS, and Relational DBMS. Each line shows a trend of increasing popularity over time.]
Will it ever become EDaaS?

- Pre-dominantly a business model discussion.
  - Currently paying for the tools we do NOT use vs paying for tools by usage.
- Not as long as ‘one’ pays per the license.
- But… could change if one pays for the ‘data’
Why API?

*9148 total published APIs as of May/19/2013 - ProgrammableWeb

Lower integration costs -
• By 2017, over two thirds of new integration flows will extend outside the enterprise firewall
• By 2018 50% of the cost of implementing new large systems will be integration

APIs also
• Drive customer engagement
• Support mobile app development
• Extend business models, offering new revenue opportunities for existing company assets

By 2014, 75% of the Fortune 1000 will offer public APIs

Source: Gartner Research
Who is going to pay for it?

What will EDA tools licensing model look like? [50th DAC, 2013]

Value is moving from the algorithms to the DATA

Gary Smith was right (again..)

“Give away the tools.. Charge for the models” (DATA)
EDA 3.0 Advantages

- Phenomenal design flow overall TAT improvements and efficiencies.
- Fantastic new insights in our design and design flows through advanced analytics
- Significant Standardization throughout the software stack
  - Spawn significant new wave of innovation
- Supporting the tools becomes a lot easier.