

Presentation

on

“ReSearch @ IT Industries”

Presented by
M. B. Chandak
Professor and Head, Department of CSE
www.rknec.edu



@Research Center
Department of Computer Science & Engineering
Shri Ramdeobaba College of Engineering and Management
Nagpur

Research Ongoing



- On going Research in Top IT companies
- Top IT Companies
- Top Trends in 2015

Research Areas: Microsoft



Some interesting research projects

- **Data Mining and Management**

- **Finding pattern and insight in data**

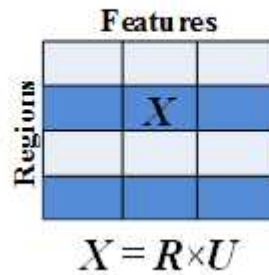
- Solving key problems in database architecture and information management. Our current areas of focus are infrastructure for large-scale database systems; reducing the total cost of ownership of information management; enabling flexible ways to query, browse, and organize rich data sets containing both structured and unstructured data; and the management of database schemas and mappings.

- **Research Project on City Noise:**

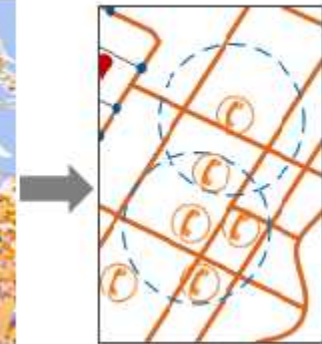
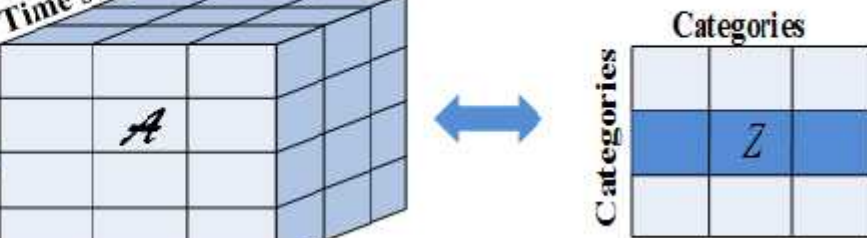
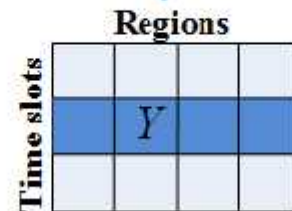
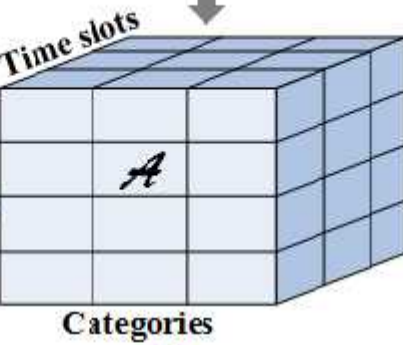
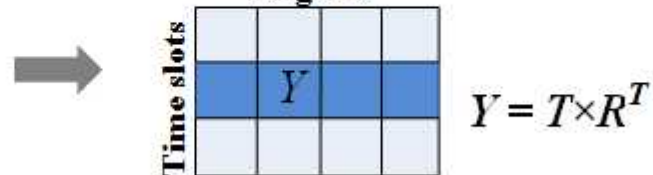
- The project aims to diagnose a city's noise pollution with crowd sensing and ubiquitous data. It reveals the fine-grained noise situation throughout a city and analyzes the composition of noises in a particular location, by using 311 complaint data together with road network data, points of interests, and social media.

Fusion heterogeneous data sources

POIs and Road Networks



Check-ins



Research Areas: Microsoft



Some interesting research projects

•Algorithms and Pattern Recognition

Research Project "Face in Crowd"

The hope that technological innovations will create good, middle-class jobs is part of government and city policies around the world, from New York to New Delhi. At the same time, machine learning and crowdsourcing are being considered as the future; as Kittur et al. (2013) said, "Paid crowd work offers remarkable opportunities for improving productivity, social mobility, and the global economy... But it is also possible that crowd work will fail to achieve its potential, focusing on assembly-line piecework." It is this tension that this project looks to explore.

Research Areas: Microsoft



Some interesting research projects

•Big Data Analytics

Small Summaries of Big Data

Sampling, Sketching to develop summaries

Testing properties of Discrete Distribution

The main focus of this research is the sample complexity of each task as a function of the domain size for the underlying discrete probability distributions. The inference tasks studied include (i) similarity to a fixed distribution (i.e., goodness-of-fit); (ii) similarity between two distributions (i.e., homogeneity); (iii) independence of joint distributions; and (iv) entropy estimation.

Streaming pattern matching

Important is space complexity $O(\log m)$ – Randomized approach

In our online model we first receive the pattern P of size m and preprocess it. After the preprocessing phase, the characters of the text T of size n arrive one at a time in an online fashion. For each index of the text input we indicate whether the pattern matches the text at that location index or not

Research Areas: Microsoft



Some interesting research projects

•Big Data Analytics

Database queries - How can we efficiently resolve database queries on massive amounts of input data? Here the input data may be presented in the form of a distributed data stream.

Machine learning - How can we efficiently solve large-scale machine learning problems? Here the input data may be massive, stored in a distributed cluster of machines.

Distributed computing - How can we efficiently solve large-scale optimization problems in distributed computing environments? For example, how can we efficiently solve large-scale combinatorial problems, e.g. processing of large scale graphs?

Research Areas: Microsoft

Some interesting research projects

- **Machine Learning and AI**

Data Driven Conversation:

This project aims to enable people to converse with their devices. We are trying to teach devices to engage with humans using human language in ways that appear seamless and natural to humans. Our research focuses on statistical methods by which devices can learn from human-human conversational interactions and can situate responses in the verbal context and in physical or virtual environments

Research Areas: IBM



Some interesting research projects

•Security

Protecting your personal data on cloud

One answer is a technology we've been developing for more than a decade called Identity Mixer. Identity Mixer uses a cryptographic algorithm to encrypt the certified identity attributes of a user, such as their age, nationality or address, in such a way that the user is able to reveal only selected pieces to third parties, such as a web service or online retailer.

For example, consider a web-based video streaming service is offering several films which have age restrictions. To stream the 12+ movie, Alice needs to prove that she is at least 12 years of age and that she lives within the appropriate region. The typical way to do this would require Alice to enter her full date of birth and address, but this actually reveals more than is necessary. Identity Mixer can simply confirm that Alice is at least 12 without disclosing the month, date and year of her birth and reveal that she lives in the correct region, i.e. region 1, instead of her full address. This ensures that even if the video streaming service is hacked Alice's personal data remains safe.

Research Areas: IBM



Some interesting research projects

•Cognitive Computing

Artificial Intelligence meet business intelligence

Ability to Sense, Predict, Infer from learning using data in the form of unstructured video, image, symbols, and natural language.

Systems with domain expertise

Unlike **expert systems** of the past which required rules to be hard coded into a system by a human expert, cognitive computers can process natural language and unstructured data and learn by experience, much in the same way humans do. While they'll have deep domain expertise, instead of replacing human experts, cognitive computers will act as a **decision support system** and help them make better decisions based on the best available data, whether in healthcare, finance or customer service

Research Areas: IBM



Some interesting research projects

•Cognitive Computing

Human Machine-Integrated systems

In traditional AI, humans are not part of the equation, yet in cognitive computing, humans and machines work together. To enable a natural interaction between them, cognitive computing systems use image and speech recognition as their eyes and ears to understand the world and interact more seamlessly with humans. It provides a feedback loop for machines and humans to learn from and teach one another. By using visual analytics and data visualization techniques, cognitive computers can display data in a visually compelling way that enlightens humans and helps them make decisions based on data.

Research Areas: Oracle



•Adaptive Optimization

The main idea of this project is that instead of hand-crafting rules that specify actions the system should take, human administrators should instead specify the optimization goal (performance objective), which the system then automatically tries to achieve. This approach greatly increases the ease of system management, makes systems more adaptable to changes in the business environment and increases the productivity of existing resources because the system can use them more optimally.

•FastR

FastR is an implementation of the [R Language](#) in Java atop [Truffle and Graal](#). Truffle is a framework for building self-optimizing AST interpreters. Graal is a dynamic compiler that is used to generate efficient machine code from partially evaluated Truffle ASTs. FastR is an open-source effort of Purdue University, Johannes Kepler University Linz, and Oracle Labs.

Research Areas: Oracle



•Information Retrieval

Information Retrieval

We are interested in **core relevance** in Information Retrieval (IR) systems: determining the set of documents that are most relevant to a given query, using just the query and the content of the documents. This core relevance work can be incorporated into **learning to rank** systems that use Machine Learning to find the best function to use to rank future search results. It can also be incorporated into **results diversity models** that try to show a variety of kinds of relevant documents in response to a query. We are looking at how to use different kinds of signals in learning to rank systems in enterprise and e-commerce search systems.

We are concerned with problems of scale in IR systems: how can systems be built and distributed so that we can search billions of documents in real time. We investigate in how search can be exploited in application specific contexts like email search.

Research Areas: Oracle



•Statistical Natural Language Processing

Statistical Natural Language Processing

We use techniques from the field of Statistical Natural Language Processing (NLP) to do text mining: extracting structured information from unstructured data. We are investigating applications of Statistical NLP like **named entity recognition**, where we extract the names of entities like people, places, organizations, and products from text. Once we have extracted a set of entities we can perform **coreference resolution**, where we try to determine whether possible mentions of an entity are really referring to the same entity, and **entity linking** where we try to link a mention of an entity to a particular entry or set of entries in a structured knowledge base. We can then consider tasks like **relationship extraction**, where we try to find out how the entities in a document are connected, for example, we can try to learn that the person *Larry Ellison* is the CEO of the organization *Oracle Corporation*

Research Areas: Oracle



•Machine Learning

We also have interests in more fundamental aspects of Machine Learning. We're interested in scalable learning and inference techniques for the graphical models that drive our statistical NLP work. We are also concerned with semi- and unsupervised techniques for Machine Learning, since in many situations it is difficult and expensive to obtain the training labels needed for supervised learning techniques. We're also investigating active and cost-sensitive learning techniques, so that we can make the best use of limited resources when acquiring information during the learning process.

Research Areas: Oracle



•Project Q

Project Q's ultimate goal is to enhance the database's applicability by integrating it with multiple DSLs other than SQL and boosting its performance by using just in time (JIT) compilation at runtime. We have several subprojects that are moving us along towards achieving the above stated goal.

DSLs:

Driven by the quest for energy efficiency, computing systems are becoming increasingly parallel and heterogeneous (e.g. consisting of different special-purpose processing elements), and therefore new applications must be capable of exploiting parallelism in order to continue achieving high performance. However targeting these emerging devices often requires using multiple disparate explicit programming models and making decisions that can limit forward scalability. A promising approach to solving this problem is the implicit extraction of parallelism via the use of domain specific languages (DSLs).

DSLs can shield application developers from the complexity of targeting parallel heterogeneous hardware by providing very high level abstractions.

Research Areas: Oracle



•Project Q

Project Q's ultimate goal is to enhance the database's applicability by integrating it with multiple DSLs other than SQL and boosting its performance by using just in time (JIT) compilation at runtime. We have several subprojects that are moving us along towards achieving the above stated goal.

DSLs:

Driven by the quest for energy efficiency, computing systems are becoming increasingly parallel and heterogeneous (e.g. consisting of different special-purpose processing elements), and therefore new applications must be capable of exploiting parallelism in order to continue achieving high performance. However targeting these emerging devices often requires using multiple disparate explicit programming models and making decisions that can limit forward scalability. A promising approach to solving this problem is the implicit extraction of parallelism via the use of domain specific languages (DSLs).

DSLs can shield application developers from the complexity of targeting parallel heterogeneous hardware by providing very high level abstractions.

Top 10 IT companies worldwide



Microsoft ([MSFT](#)): The world leader in software companies, Microsoft continues to maintain its dominance with [total revenues](#) of \$77.85 billion in 2013. Of this, \$65.7 billion, or 84 percent, was from its software stream. Microsoft's software revenues exceed those of its next two competitors combined.

Oracle ([ORCL](#)): Oracle surpassed IBM in 2013 to gain the number two spot in software revenues. Its software revenues were \$29.7 billion in 2013, out of a [total revenue](#) of \$34.74 billion.

International Business Machines ([IBM](#)): The large conglomerate offers a wide range of products and services, including both hardware and software. It consistently derives 25 to 30 percent of revenue from software. Of the [total revenue](#) of \$99.75 billion for 2013, software contributed \$29.1 billion.

SAP ([SAP](#)): The Germany-based multinational software giant generated \$18.9 billion in revenue from its software stream, out of the total revenue of \$22.87 billion. Software represents 83 percent of its business. (SAP's official [report](#) gives revenues in euros. For this article, we converted euros to U.S. dollars using the December 31, 2013 exchange rate of 1.36.)

Top 10 IT companies worldwide



Symantec ([SYMC](#)): Nasdaq-listed Symantec, the global leader in software security solutions, generated \$6.4 billion in software revenue from [total revenues](#) of \$6.9 billion.

EMC ([EMC](#)): NYSE-listed EMC takes the number six spot with \$5.6 billion of software revenues out of [total revenues](#) of \$23.2 billion. (*EMC also owns the software company VMWare which is number eight on this list. The companies are listed as separate entities on the New York Stock Exchange. For that reason, this list treats EMC and VMWare as two separate entities.*)

Hewlett-Packard ([HPQ](#)): The global giant is mainly known for printing products and solutions. From its total revenue of \$112.298 billion, it derived just \$4.9 billion from software. Although this is a small percentage of Hewlett-Packard's total revenues, it still makes the company the seventh largest software company.

VMWare ([VMW](#)): VMWare, owned by EMC, remains separately listed on the New York Stock Exchange. Its software stream revenues for 2013 were \$4.8 billion, out of [total revenues](#) of \$ 5.2 million. If the software revenues of EMC and VMWare considered together, the company would be the fifth largest software company in the world.

Top 10 IT Trends



1. Computing Everywhere

As smart-phone technology advances, smart-phones will be used in new contexts and environments. Along with wearable's, smart-phones will offer connected screens in the workplace and in public. User experience will be key.

2. The Internet of Things (IoT)

The Internet of Things is big and it will continue to grow along with user-oriented computing. Prediction: The Internet of Things will be the focus of digital business products and processes in industrial and operational contexts. Expect technology to be embedded everywhere.

3. 3D Printing

3D printing is about to get cheaper, and its market will grow over the next three years. The expansion will be biggest in industrial, biomedical, and consumer applications helping companies reduce costs.

4. Advanced, Pervasive, Invisible Analytics

Analytics will continue to grow propelled by the Internet of Things, creating large pools of data. Every app will need to be an analytic app. But big data isn't the most important thing: instead we'll need big questions and big answers.

Top 10 IT Trends



5. Context-Rich Systems

Thanks to embedded intelligence and analytics, systems will become alert and responsive to their surroundings. Expect context-aware security as well as other trends.

6. Smart Machines

Analytics and context will pave the way for smart machines that can learn for themselves and act accordingly. These machine helpers will continue to evolve. Prediction: The smart machines era will be the most disruptive in the history of IT.

7. Cloud/Client Architecture

As mobile computing meets cloud computing, centrally coordinated applications that can be delivered to any device will continue to grow. Apps that can use intelligence and storage effectively will see lower bandwidth costs. Expect to be able to use applications simultaneously on multiple devices

Top 10 IT Trends



8. Software-Defined Infrastructure and Applications

Software defined networking, storage, data centers and security are maturing. Cloud service software is configurable thanks to rich APIs. Computing will have to move away from static models to deal with the changing demands of digital business.

9. Web-Scale IT

More and more companies will begin thinking like Amazon, Google and Facebook. As cloud-optimized and software-defined methods become mainstream, we'll see a move towards web-scale IT, starting with DevOps.

10. Risk Based Security and Self - Protection

While 100% security solutions aren't feasible, advanced risk assessment and mitigation will come into play in the next few years. Security will move away from perimeter defense to multi-faceted approaches. Expect security aware application design, dynamic and static application security testing, and runtime application self-protection.

Happy learning!



hodcs@rknec.edu
chandakmb@gmail.com

9822465550
0712-2582844 (477)