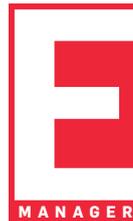


# 15-MINUTE



TRAINING
EDUCATION
WORKPLACE
COMPENSATION
WORKFORCE TRENDS
SKILLS DEVELOPMENT
PERSONAL DEVELOPMENT

HIIT- NEW  
WAY TO BE  
HEALTHY  
PAGE 36



**Inputs of a Big Data Architecture:** PAGE 37

**Healthy Tips: HIIT--The New Mantra** PAGE 36

**Product Review- Nokia X Phone is simple** PAGE 38

**The CIO's Role is Well-Defined** PAGE 49

BY SUNIL S RANKA

With big data and analytics playing an influential role helping organisations achieve a competitive advantage, IT managers are advised not to deploy big data in silos but instead to take a holistic approach toward it and define a base reference architecture even before contemplating positioning the necessary tools. This is a mechanism to map bgi data and use analytics to make the most of them.

### Myths that surround big data

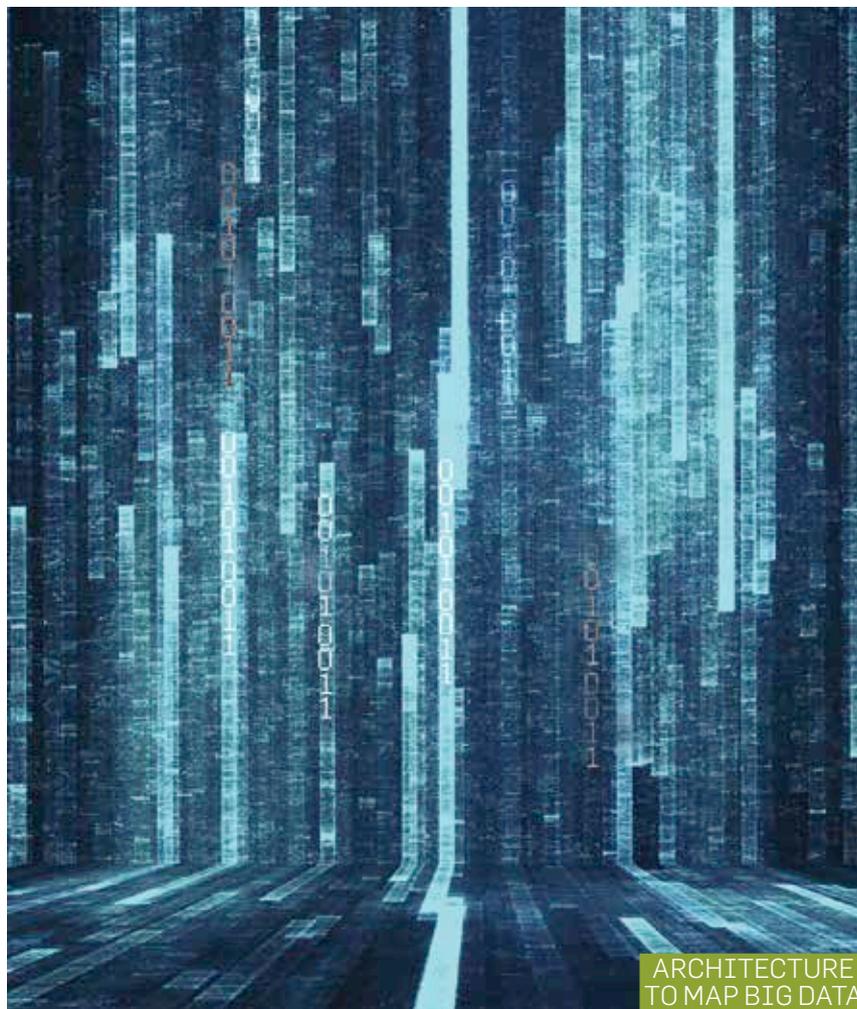
The industry is witnessing a paradigm shift in the way that big data is being used to transform and analyze information in assisting informed decisions.

Everchanging technology is the driving force of these myths about big data and they are creating considerable confusion among its users.

This makes it even more important to understand the reference architecture before deploying big data. But before looking at the architecture and its uses, it is critical for IT managers to recognize some of these myths -

- Hadoop is big data
- By using big data, information processing can be done in a matter of seconds
- Big data acts as a magic wand for all data processing issues

IMAGE: THINK STOCK PHOTOS.IN



ARCHITECTURE  
TO MAP BIG DATA

## STYLE GUIDE TO MAP BIG DATA

IT managers are advised to take a holistic approach to big data and define a reference architecture.

## 15-MINUTE MANAGER

- Big data automatically discovers all 'Unknown Unknowns'
- Big data is inexpensive
- Big data can replace any BI and EDW system

### What is Reference Architecture and why should one opt for it?

Like any other base architectures associated with any technology, the 'Reference Architecture' in case of big data will describe the ins and the outs of the platform. In a nutshell, a reference architecture will provide:

- A common lexicon and taxonomy
- A common (architectural) vision for the big data project
- Modularization and the complementary context of the project

With user needs and technological dynamics changing constantly, it is essential to standardize the key components of a reference architecture for scalability purposes. According to Sunil S Ranka, a BI professional, "Our experience on working with various clients compelled us to have a reference architecture in place which would define the very objective of the big data project and evolve a mechanism to support a unified vision for deploying information management and analytics."

Reference architecture is described in terms of technological components that achieve the capabilities and drive the vision of the project. Big data technologies are mapped to the architecture in order to illustrate how the architecture can be implemented and deployed. Organisations can use this reference architecture as a preparatory point for outlining their own distinctive and custom-tailored architecture.

"With heavy investment in current BI systems, customers want to enhance the current capability of their analytics by bringing in big data solutions to their existing enterprise system's landscape, but the million dollar question is on how this is to be done.

### How can one fit big data in the existing landscape of the enterprise?

According to a study conducted at the Embedded Systems Institute and Ste-



## HEALTHY TIPS

# HIIT—THE NEW MANTRA FOR FITNESS

High-intensity interval training, a type of workout that consists of very brief bouts of very strenuous exercise, has become enormously popular in recent years.

The science of intensive interval training has, though, been lagging behind the workout's popularity. Past studies of HIIT, as the practice is commonly known, had established that as measured by changes in cellular markers, standard short-burst HIIT training may improve aerobic fitness up to 10 times as much as moderate endurance training, reported Gretchen Reynolds in *The New York Times*.

But scientists had not determined whether a single sustained interval likewise improves fitness, or the ideal number of HIIT sessions

per week. So to clarify those issues, researchers at two of the laboratories most noted for HIIT science set out to learn more about the best way to do interval training. First, Ten of them were asked to exercise on two separate days. Separately, the remaining seven volunteers did the continuous four-minute workout three times a week for six weeks.

The researchers collected blood and muscle samples, and monitored changes. When collated and compared, the data showed that the physiological differences among the two groups of riders were notable and, in some ways, strange. When the researchers checked blood and muscle tissue in the second group of riders after they had completed six weeks of single-interval training, some of the pending improvements seemed to have evaporated, while six weeks of standard short-burst HIIT exercise resulted in significant, sustained gains in these markers.



**“Our experience on working with various clients compelled us to have a reference architecture in place which defined the objective”**

— **Sunil S Ranka**, BI Professional

## INPUTS OF A BIG DATA REFERENCE ARCHITECTURE

- Existing Architectures
- Customer needs and business needs
- Essence architecture patterns
- Product portfolio future requirements
- Proven concepts and known problems
- Exploration and analysis
- Data Mining and vision
- Triggers new changes
- Guides evolution

vens Institute of Technology, there are processes to create and maintain a reference architecture.

This suggests that a reference architecture captures a previous experience by mining, or by generalizing existing architectures. To be of value for future architectures, a reference architecture is based on proven concepts. The validation of the concepts in a reference architecture is often derived from their predecessors. This is especially true in cases where disruptive technologies or innovative applications are introduced and it is challenging to have sufficient proof to support a reference architecture. In these cases reference implementations, prototyping and an incremental approach might be an appropriate alternative for validation and proof.

More often than not, preceding architectures are mined for these proven concepts. Architecture renovation, innovation validation and proof can be based on reference implementations and prototyping. According to an architecture, the big data solution should comprise traditional BI reports, charts, information access application, data warehouses for information integration, and information delivery and decision making, which drives unstructured data transformation through ETL systems. The enterprise systems and content stores absorb con-

tent management systems, data bases, unstructured data, files systems, ERP, CRM, SCM, SOA, ESB, Web service and so on. These are integrated with the solutions to analyze them.

Using this process, big data components are designed to refer to four focus areas in terms of how the data is analyzed. To further elaborate, a reference architecture elaborates mission, vision and strategy to provide guidance to multiple organisations. What does that mean? A reference architecture is strongly linked to company mission, vision and strategy. The strategy determines what multi-dimensions have to be addressed what the scope of the reference architecture is, what it means, such as synergy, are available to realize mission and vision. There are 2 principles that the deployers need to follow.

Principle 1: A Reference Architecture facilitates a shared understanding across multiple products, organisation, and disciplines about the current architecture and the vision on the future direction. Principle 2: A Reference architecture is based on proven concepts. A reference architecture often consists of a list of functions and some indication of their interfaces (or APIs) and interactions with each other and with functions located outside of the scope of the reference architecture.

### The Big data Analysis

The architecture lays down the four step approach to analyze the data -

- Data ingestion
- Data storage
- Data processing
- Data visualization

**The value of a reference architecture is evident in environments with high multiplicity factors that are creating social, organizational, business, application and technical complexities.**

PRODUCT REVIEW

# NOKIA X REVIEW



## Build and design

The Nokia X is an overgrown Asha device. The back panel is all plastic and comes in a variety of colors including bright green, bright red, cyan, yellow, black, white. The material feels a bit like the polycarbonate used on the Lumia devices, with a matte finish. The front is dominated by a protective glass covering the 4-inch display. The handset has a very simple and clean design approach. The size and dimensions seem perfect and the handset fits into your palms perfectly (unless you have huge hands).

## UI and performance

This is where the Nokia made a very cunning move. At the time of announcement, the company mentioned that the handset runs on a custom built Android thanks to the AOSP (Android open source project). The UI is just like the Asha Platform with a mix of Windows Phones tiles, so you only have two screens, one where all your apps and widgets can be placed and the second is your Fastlane screen where all your recent apps and notifications are placed.

The UI is quite confusing in the beginning and feels a bit empty if you compare it to a proper Android running smartphones. All installed apps are placed automatically in colored blocks on the first homescreen and one can also make folders and add apps into them. The fastlane section, just like the Asha platform, gives you the recently opened apps, your system as well as social network notifications, alarms, birthdays, etc.

The OS doesn't support true multitasking so when you exit an app, you essentially close the app. You can resume working on the app by going to Fastlane, but switching between multiple apps is not possible.

The major issue with OS is that there is no Google framework. This means that you cannot have access to all the apps that require a Google account such as Gmail, Google Maps, and so on.

The Nokia X is essentially an affordable handset which offers smartphone-like features and supports Android apps. We didn't expect much in the performance department but we couldn't resist carrying out some benchmark tests. As the handset has the capability to sideload apps, we tried Smartbench, Antutu and Quadrant to see how it scores.

The battery test proved to be a little off-scale as we found a loss of about 32% after running a full HD video for one hour.

## Data ingestion:

The system must be able to acquire data despite high volumes, velocity and variety. It may not be necessary to persist and maintain all the data that has been received. Some may be ignored or discarded while others are kept for varying amounts of time. The following components can be used for data ingestion:

Apache Flume is a distributed, reliable service for efficiently collecting, aggregating and moving large amounts of log data from many different sources to a centralized data store. It has a simple and flexible architecture based on streaming data flows. It is robust and fault-tolerant with tunable reliability mechanisms, and many failover and recovery mechanisms.

Apache Sqoop(TM) is a tool that was designed for efficiently transferring bulk data between Apache Hadoop and structured data stores such as relational databases.

Java JDBC, using simple java class data from any of the sources can be ingested in HDFS.

## Data Storage:

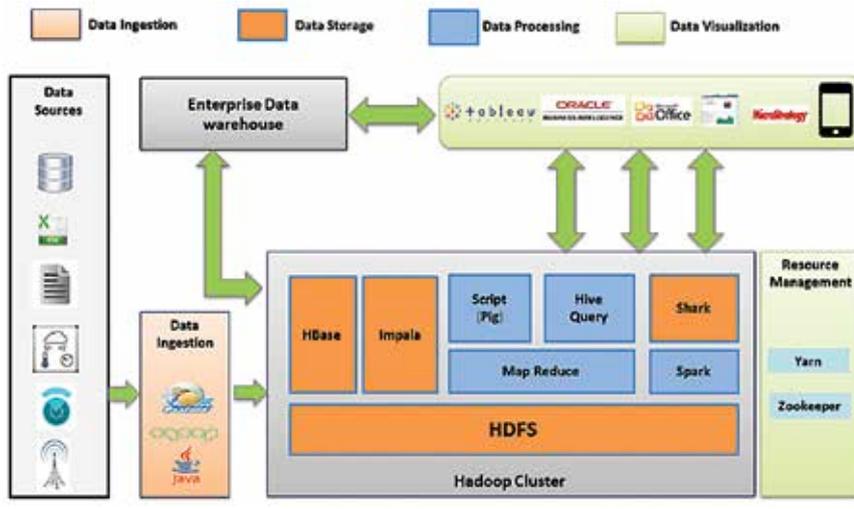
To implement the data storage layer, Hadoop uses a feature known as HDFS or the Hadoop Distributed File System. HDFS is not a file system in the traditional sense. Along with HDFS, data can also be stored in Hive (metadata), Impala, HBase and Shark.

HDFS is the primary storage system used by Hadoop applications. HDFS creates multiple replicas of data blocks and distributes them on computer nodes throughout a cluster to enable reliability and extremely rapid computations.

Apache HBase is an open-source, distributed, versioned, non-relational database modeled after Google's 'Bigtable: A Distributed Storage System for Structured Data by Chang et al.' Just as Bigtable leverages the distributed data storage provided by the Google File System, Apache HBase provides Bigtable-like capabilities on top of Hadoop and HDFS.

The Apache Hive data warehouse software facilitates querying and

## 4 step approach to big data.



Courtesy: Embedded Systems Institute and Steven Institute

managing large datasets that reside in distributed storage. Hive provides a mechanism to project structure and to query this data using a SQL-like language called HiveQL. This language also allows traditional map/reduce programmers to plug in their custom mappers and reducers when it is inconvenient or inefficient to express this logic in HiveQL.

The Apache Shark is an open source distributed SQL query engine for Hadoop data. It brings state-of-the-art performance and advanced analytics to Hive users.

Impala raises the bar for query performance while retaining a familiar user experience. With Impala, you can query data, whether it is stored in HDFS or Apache HBase – including SELECT, JOIN and aggregate functions – in real time. Furthermore, it uses the same metadata, SQL syntax (Hive SQL), ODBC driver and user interface (Hue Beeswax) as Apache Hive and so, provides a familiar and unified platform for batch-oriented or real-time queries (enabling Hive users to utilize Impala with little overhead set up).

### Data Processing:

MapReduce is a framework that parallel-processes large data sets across a

large number of nodes. Computational processing can occur on data that is stored either in a file system (unstructured) or in a database (structured). MapReduce can take advantage of the locality of data, processing data on or near the storage assets to decrease transmission of data.

Apache Pig is a platform for analyzing large data sets that consists of high-level language for expressing data analysis programs. It also has the infrastructure to evaluate these programs. The salient property of Pig programs is that their structure is amenable to substantial parallelization which is what enables them to handle very large data sets.

Apache Spark is a fast and general-purpose cluster computing system. It provides high-level APIs in Scala, Java and Python that make parallel jobs easy to write, and an optimized engine that supports general computation graphs. It also supports a rich set of higher-level tools including Shark (Hive on Spark), MLlib for machine learning, GraphX for graph processing and Spark Streaming.

### Data Visualization:

For deriving value out of a mammoth dataset, visualization tools are very important for a data scientist and an

analyst. As big data is a platform for parallel processing with fault tolerant behavior, it has no part for open source visualization that is included in the technology stack. With the availability of connectors, most of the BI tools provide connectivity to a Hadoop ecosystem. Some of the popular ones are Tableau, Micro strategy, R and MS Excel.

### How can a reference architecture help?

According to the study, reference architecture captures knowledge from existing architecture. Then, based on an elaboration of the mission, vision and strategy, and on future customer needs the reference architecture is transformed into architecture that provides guidance to multiple organisations that evolve or create new architectures.

Reference architectures should address the technical, business architectures and the context.

One of the main challenges here is to make this inherently abstract concept of a reference architecture more concrete and understandable by providing sufficient specific information and guidelines.

The value of a reference architecture is evident in environments with high multiplicity factors that are creating social, organizational, business, application and technical complexities. This is a young, relatively under-developed area which has more questions than answers. These questions range from the proven value to the lifecycle of reference architectures.

### Key Imperatives

Reference architectures should address technical and business architectures and the context. One of the main challenges as per the study from Embedded Systems Institute and Steven Institute is to make this inherently abstract Reference Architecture concrete and understandable by providing sufficient specific information and guidelines. **ITNEXT**

# Riding IT with Style

**Santhosh Kempaiah, Senior Manager–Information Security, Volvo IT, a deep believer in IT, a biker at heart, self-inspired senior IT Manager calls his own shots.**

BY SUBHANKAR KUNDU

Santhosh Kempaiah derives inspiration from his team members. And that's the motivating factor that propels him to go to work every day--the team gives him the flexibility and space to share his ideas. His IT journey started right from his graduation days when the IT boom was in its nascent stages in India; and even before he decided to leap into a career in Commerce, Bangalore was being branded the 'Silicon City.' He went against the conventions of society, as his family wanted him to pursue the civil service exams.

He says, "My gut feel was that no other field could give me the exposure that IT could. I saw a future full of potential and unlimited creativity in this sector." His gut feel paid dividends.

His stint at Thomson Reuters gave him vast exposure. He gained experience from end-user support to process transitioning. He considers

it to be the turning point in his career.

At Volvo, he plays a critical role in the Global IT Security team. He manages risk management, compliance, and awareness & training, and leads a team of security co-ordinators across the globe.

He also plays an important role in managing end to end security requirements which include implementing and managing group compliance, directives, and processes for entire APAC spanning India, China, Japan, South Korea, Thailand and Indonesia.

Kempaiah says, "We have a very robust IT architecture team with strong processes and structure, which currently runs like a well-oiled engine, servicing new demands for services in the organisation."

From the innovation perspective, his focus is on PPT (People, Process and Technology).

Volvo used to follow traditional ways of provid-



Be your own  
inspiration