

STARTER KIT FOR DATA ANALYTICS JOURNEY WITH **BIG DATA** OR **SMALL DATA**



Embarking on a data analytics journey for any enterprise is quite often a challenge considering complexities involved in making it a success. This involves everyone rowing in the same direction and associated technology and tools should make their life easy.

3 Main Challenges to work upon

- 1. Defining why you need it**
- 2. Right Solution Stack**
- 3. Implement & Sustain**

1. Defining why you need it

Enterprise has a set of goals defined for a period and necessary activities are performed towards meeting the same. There are lot of processes, tools involved to keep track of the data related to these activities. If you have issues in Interconnecting them and using that data to make your decision then you need of a proper data analytics solution

Volume

You have more data from any of the following history, regular transactions, data from Weblogs, sensors, Social data

Variety

Your data is structured, Semi-structured, unstructured Documents, images, in files, RDBMS, XML, Excel etc.,

Velocity

On a given frequency your Systems, transactions generate Lot of new data Eg., Weblogs, Sensor data, retail Transaction, mobile / webapps Activities etc.,

If any or all of these problems / challenges exist in your enterprise environment related to data, then data analytics is a need to stay competitive by making effective data based decisions rather than intuitions.

Scope Workshops

When you understand there is a definite need, then choosing the Right Use cases which will add more value and setting the right priority is the key to start seeing the benefits at the earliest, to do that conduct Internal Scope Workshops

Objective:

- Identify Challenges and Pains
- Understand Key Business Drivers
- KPAs & KPI's
- List data sources and formats (Volume, Variety, Velocity)

Involve

- Executive sponsor
- Business Users, IT stakeholders, and SME's

Collect

- Business Requirements
- KPI Catalog

Execution Time:

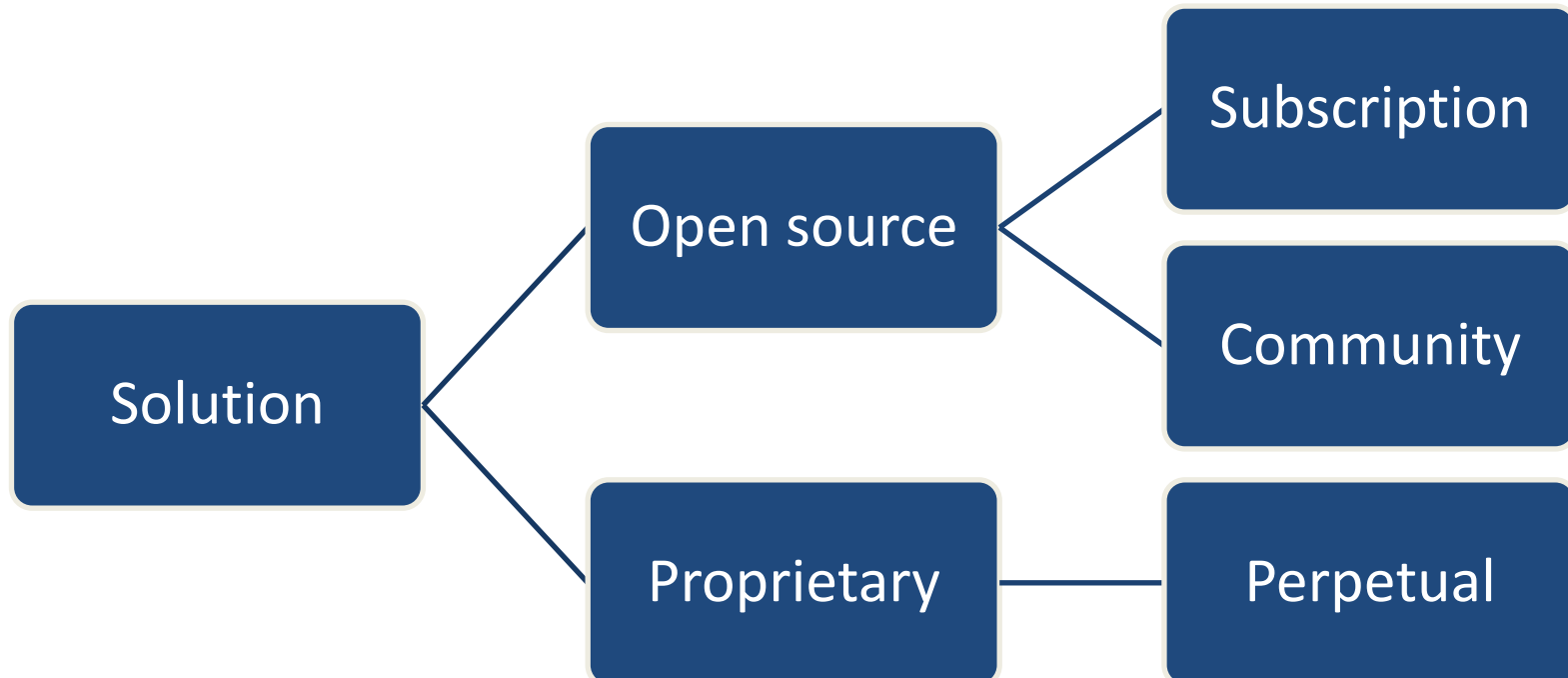
- Complete within 5 days, via online or onsite 5 sessions with a consulting team

Outcome:

- Inception Report

2. Right Solution Stack

Data Analytics is provided by various tools both on cloud and on premise Solution exists, in addition there are proprietary and open source solutions and wherever required nice blend of these solutions put together will work effectively will work for your needs, choosing the right stack is the key challenge as this architecture should help you with the following



Solution Highlights

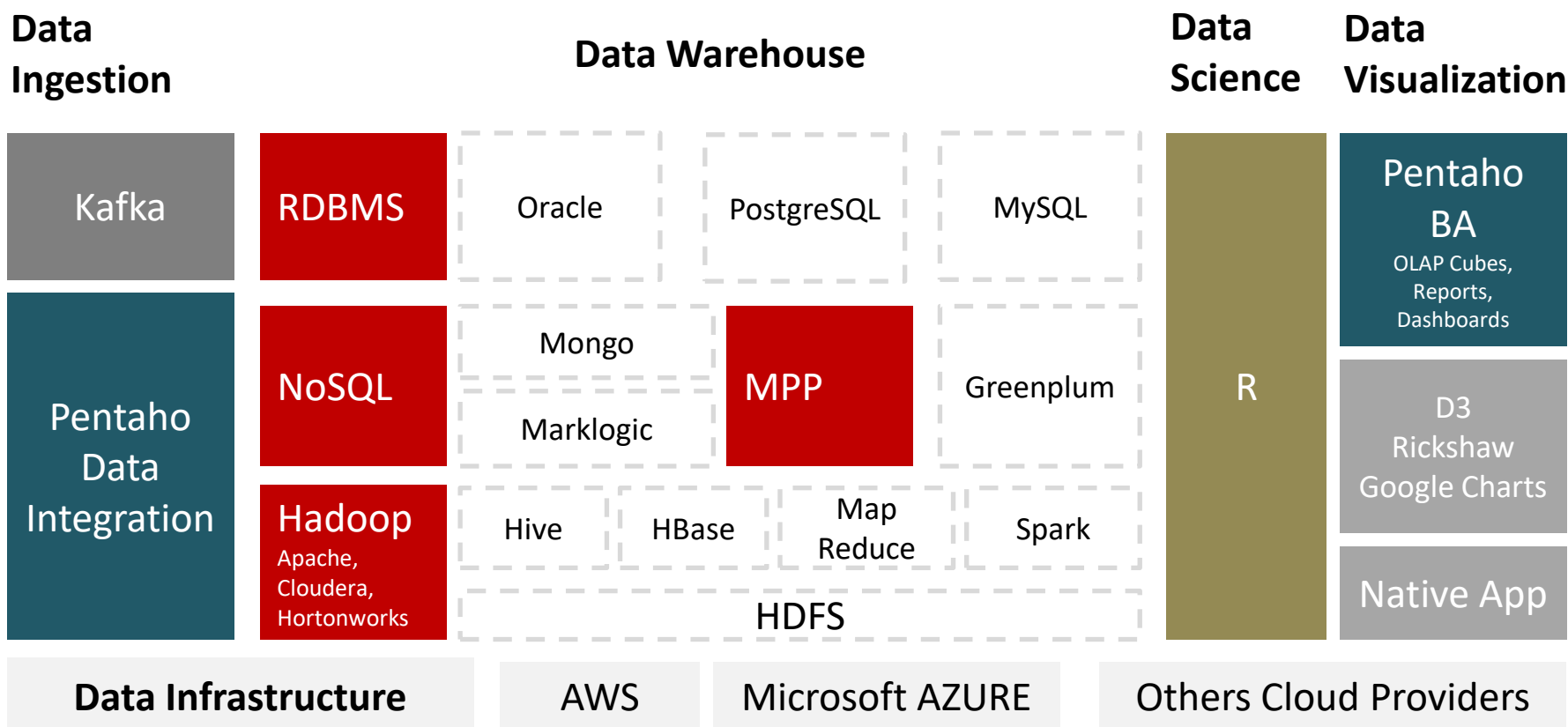
The solution architecture to provide the following major benefits:

- Eases the pain of Data On boarding from existing environment to a distributed analytics database
- Reduced development and maintenance efforts, flexible to sense and accommodate source structure changes
- Self service capabilities for end users
- Extract dynamic datasets and auto generate models and visualize on demand
- Scheduled data extraction in defined formats and sharing it with end consumers
- Secured access control at granular level
- Scalable and Distributed
- Supports Realtime
- Open System for integration
- Embeddable
- Platform agnostic
- Extended big data eco system support

BIG DATA STACK



Below blueprint is one possible option, components can be Altered depending on your enterprise needs



Solution Features

Check for the following features

- Data Integration with visual workflow designer
- Designer tools for reports, dashboards, cubes
- Interactive analyzer and report builder to reduce IT dependency
- Smart tools to suggest aggregations
- Easy to monitor administration tool
- Mobile Support
- Ability to run reports offline and burst
- Handle high volume of data
- Support for extensive graphs
- Ability to export and share data via web services
- Integration with your Single Sign on systems
- Integration with data science tools for descriptive, predictive modelling

Proof Of Concept

Evaluate the solution stack with your data with defined use cases, use Case to cover one sample set of data from your various sources and it should be really usable and the first set of users of that use case should have the required pain to adapt to the analytics solution

Environment

- Choose on cloud, on premise or hybrid model

Infrastructure

- Identify Concurrent Usage Needs
- Define data volume and velocity based on current data and consider growth for next few years

Blueprint

- Identify the right components required for your needs from the overall ecosystem considering all the factors listed above

FIT/GAP

- Evaluate your needs against the solution stack picking few use cases

Duration

- 3 -4 Weeks

Outcome:

- Live POC Demo

3. 1 Implement

Implementation approach should be iterative and should be incremental
User adoption is the key to successful implementation and ensuring
Solution caters their needs in a iterative manner will add more confidence
To users. Data accuracy is the prime concern of any implementation
making users to participate in data verification and acceptance will
guarantee usage. Implementation will most likely have these following
high level activities

- 1. Data Infrastructure Setup**
- 2. Data Warehouse & Integration**
- 3. Data Visualization**
- 4. Data Science**

3.1 Implement

Data Infrastructure set up for the customer landscape. Installation and configuration of stack components based on the solution architecture blueprint to meet the business need

Data Warehouse Integrating all the relevant data sources for your business - including both transactional data in relational databases and unstructured data from social media, Internet of Things or device/sensor data, and more in an analytics-ready fashion. Build an Enterprise Data Warehouse for Reporting and Analytics

Data Visualization Build custom analytics dashboards, interactive Reports, Cubes. Gain in-depth and shareable insights into your business. Provide appropriately designed graphically represented information in making decision

Data Science is a field driven by data. A process in which by implementing various scientific and statistical methods, knowledge/insights is obtained from the data. Build statistical models, predictive and prescriptive models using various algorithm and machine learning.

3. 2 Sustain

Upon go live the key is to ensure analytics solution is stable and available and data accuracy is ensured. Making sure all users know how to use the system and having an internal Tier 1 support to assist, run some jobs, export data is available in house with required training, following activities are important to sustain the system and make sure users get adapted and addicted to the analytical solution

Support to support and Address all bugs/errors/gaps in the requirements as agreed in SRS, Data Validations between source and target systems, Verify all developed visualizations, Monitor Job execution Performance tuning, Monitor manage Infrastructure load

Train Online and public classroom courses provide a quick and convenient way to get your team up-and-running. Provide a quick and convenient way to get the end user team effectively use the Pentaho BI Application

SERVICES we offer



CONSULTING & TRAINING

- Architecture Workshop
- Installation & Configuration
- Evaluate & Rationalize
- Training

IMPLEMENTATION & SUPPORT

- Data warehouse modeling
- Data Integration
- Analysis (OLAP)
- Visualization
- Data Science
- Support

EMBEDDED ANALYTICS

- White Label
- Theming
- Infuse

Pentaho technology
innovation award winners

www.tenthplanet.in/pentaho
info@tenthplanet.in

91-44-42961000