

BICube™

BICube™ 是针对客户的需求把大数据商业模式以One-Stop的形式实现的大数据机器学习平台。

大数据解决方案的限度

关于大数据，绝大部分的解决方案只关注单一的技能。因此，要实现商业模式只能选择，**Architecturer** 多样的解决方案并加以组合。

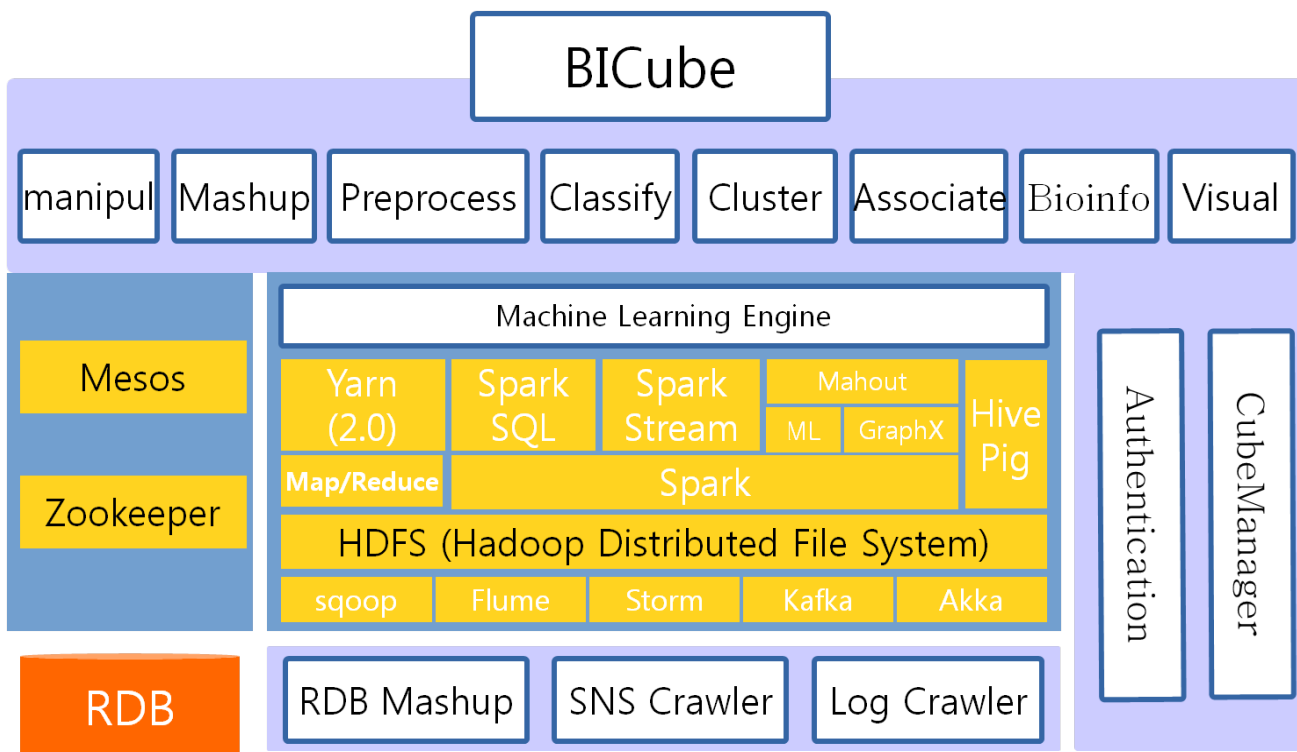
如此以来，产生的各种问题，例如工具(Tool)之间呼唤问题，额外购买解决方案产生的费用，怎样稳定运行新增加解决方案，技术援助 等，都是实现客户所需的大数据商业模型的绊脚石。

BICube™ – 单一平台

BICube™ 是把各个按技能分类的Layer合并后实现了单一的大数据平台。

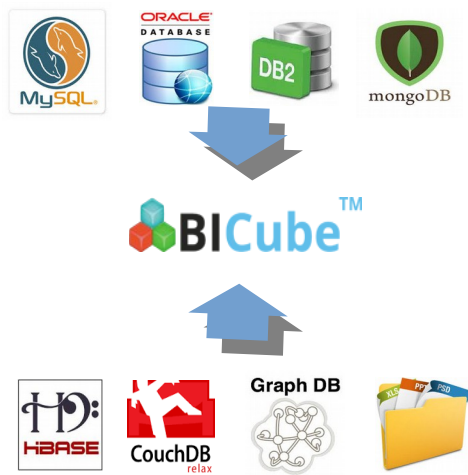
- Layer 1 – 数据提取，变换以及装载
- Layer 2 – 数据存储
- Layer 3 – 机器学习引擎
- Layer 4 – 应用层(实现商业模型的层)

Vertical 结构的大数据平台是世界唯一的平台，而且在单一的平台内实现了商业模式。



多样化数据的连接器

在大数据分析平台，数据链接是最基本的而且是最重要的。以HDFS做为基础数据存储库的 BICube™，为了从HDFS提取数据，提供各种不同的数据连接器。以此为基础，BICube™ 可以聚集多样化来源的数据于一身。



SNS Crawler / LOG Crawler

在大数据分析领域，不能忽视的部分是社交网络数据分析和系统登录数据分析。BICube™ 具备的 SNS Crawler 用于社交网络数据分析和 LOG Crawler 用于登录数据分析。



实时流式传输

为实时分析流式传输 BICube™ 多采用了开放源代码，而且为提高性能，自制研发并持有流式传输处理模块。



数据变换模块

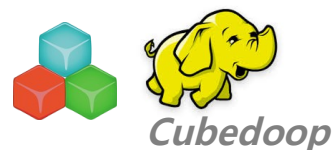
平台内包含的多样化数据变化模块是用于预处理定型数据，并且可以简便的执行基础数据的变化。
- 数据探测器, Table 关联运算, Row/Column 运算, Summary 运算, 关系链接等。



Cubedoop

BICube™ 使用的基础存储库是 Hadoop。要是直接使用从Apache提供的Hadoop为Enterprise用途，运行上会产生很多麻烦的问题。

为了圆满的监视和管理Hadoop，专门为 BICube™ 制造了特化的Hadoop配布版。



为什么需要机器学习?

在业界，虽然兴起大数据流行的年份已有4-5年，但是很少报道关于大数据成功的事迹。其主要原因之一是错误的业务处理方式。

很多人对大数据业务，错误的理解为在传统的数据分析模板上只是增加了要分析的数据的量。

能帮助摆脱这种局面的唯一解决方案就是机器学习。作为分析对象的庞大数据遇上机器学习这门技术时，必会创造出我们意想不到的新价值。

已有很多利用机器学习的公司，正在进行新兴产业。对此，也有媒体报道着相应的应用事迹。

多样的 Algorithm

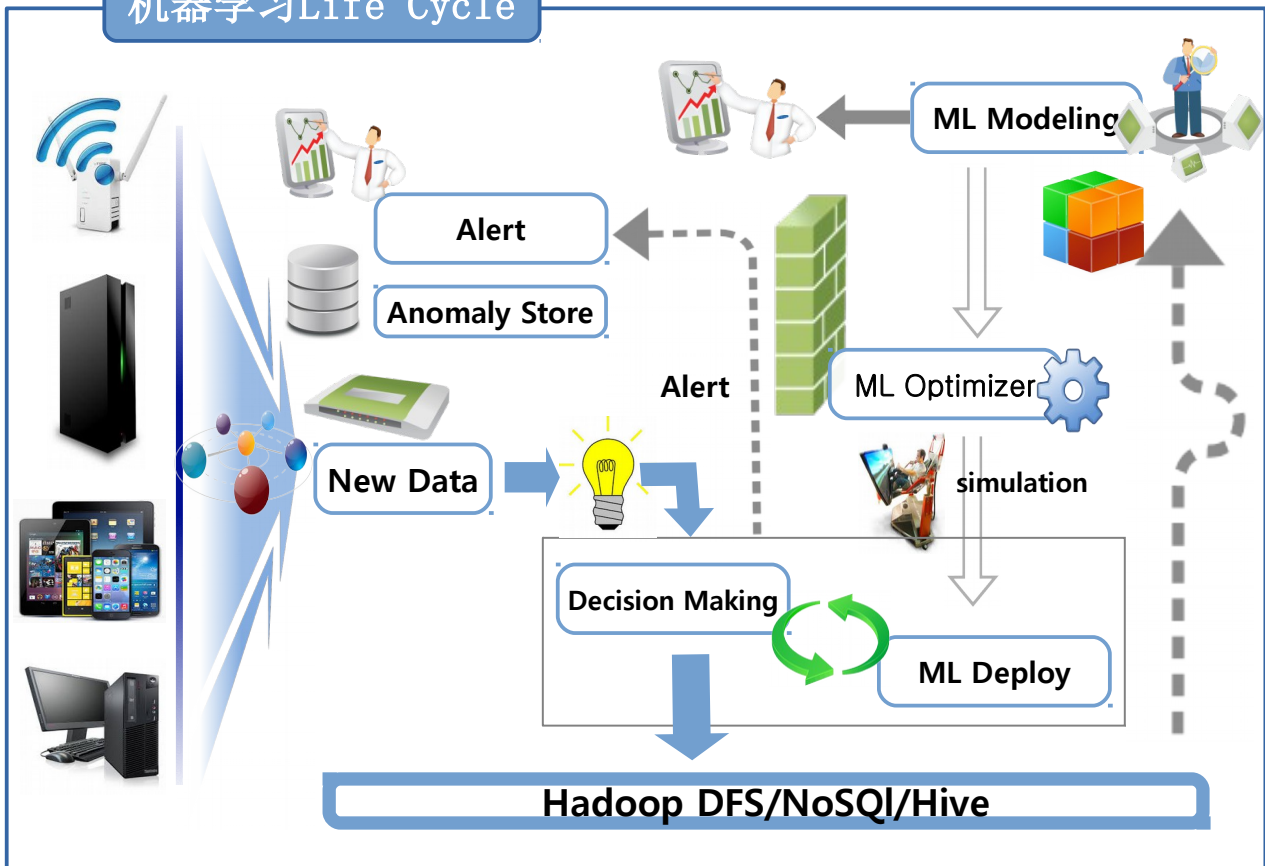
敝司认知到机器学习的必要性，在 Framework 内包含了机器学习引擎。机器学习引擎内体现了各种不同的 Algorithm。

[体内的机器学习 Algorithm 清单]

- MultiLinearRegression, KoreanMorpheme
- Logistic Regression, SVM, Adaboost, Neural Network
- RandomForest, NaiveBayssian, Timeseries Analysis
- Kmeans, LDA, HMM, PCA, SVD, CRF, TDA
- Covariance, Correlation, ROC, LSA, EigenFace Image
- CF Recommender, ConfusionMatrix, AUC
- Bagging, Wagging, ArcX4, LogitBoost, ModestAdaboost
- SAMME, BinaryClibration, IsonicCalibration...
- Discriminant Adaptive Nearist Neighbor
- Hierachical agglomerative clustering
- Kernel Recursive Least Square, Radial Base Network
- Local scaled density base clustering, EMGaussianMixture
- Sequential minimum optimization, Dual coordinate
- T-Test, F-Test, G-Test, Chi-square-Test Anova etc...

机器学习Life Cycle

BICube™ 是对机器学习Life Cycle的管理提供管理环境。



业务逻辑的体现

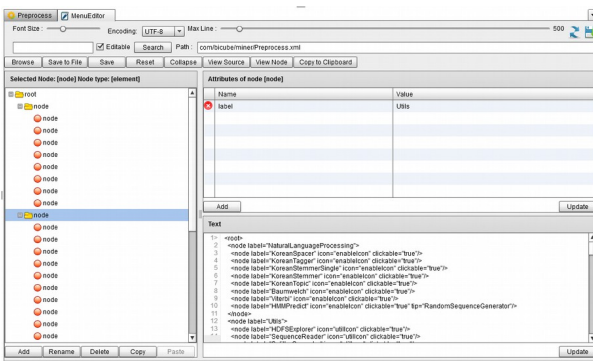
BICube™ 的应用层作为框架中最高层，是体现业务逻辑的层面。共分为8个组织，每个组织的作用如下。

- ◇ Manipulation : 各种控制远程系统的组件
(Java, Spark, Scalar, Python, R, Hadoop, Graph)
- ◇ Mashup : 数据连接器和Crawler, Open API组件
- ◇ Preprocess : 自然语处理, 各种分析程序, Incorder组件
- ◇ Classify : 体现各种类型 Algorithm的组件
- ◇ Cluster : 体现各种集团 Algorithm的组件
- ◇ Associate : 体现关于关联分析 Algorithm的组件
- ◇ Bioinfomatic : 为分析生物的组件
- ◇ Visual : 各种视角化组件

→ 利用各种组件, 体现客户所需的商业模型。

DIY 用户对界面的管理

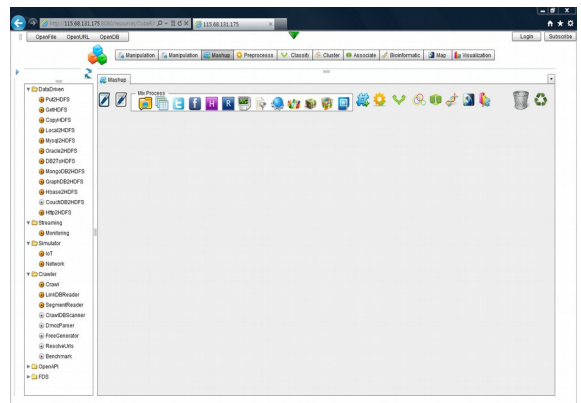
BICube™ 可以根据客户所需的方式组成菜单。因此, 使作业环境的组成成为了可能, 而大大提高了作业效率。



已合并的 GUI作业空间

为了体现商业逻辑, 系统开发者可登录熟悉的网络基础画面。根据需要体现的技能, 点击最上端的菜单, 左侧会显示相应的 **Catergorized Tree Menu**. 选择需要的组件后, 拖放到右边的作业空间, 进行详细的设定。

此外, 每个组件可以与 **Link** 相连接, 像这样的 **UML** 方式的逻辑体现方式, 它的逻辑体现很简便, 而且因可续性的优越, 也很容易维护系统的运行。



For more Infomation
<http://www.bicube.co.kr>
 daengky@bicube.co.kr