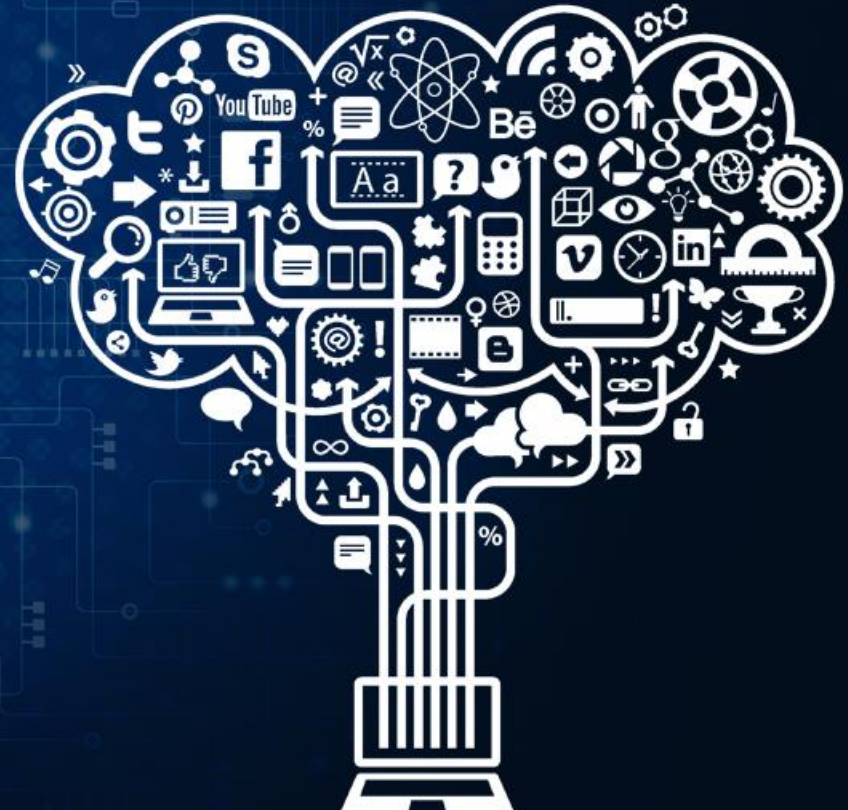


20 September 2016

# The Rise of the AI in **BIG DATA**



## **Lifecycle Analitik Data**

Imam Cholissodin | [imam.cholissodin@gmail.com](mailto:imam.cholissodin@gmail.com)

# **Pokok Bahasan**

- 1. Pengantar**
- 2. Key Roles Kunci Sukses Proyek Analitik**
- 3. Lifecycle Analitik Data: Discovery, Data Preparation, Model Planning, Model Building, Communicate Results, Operationalize**
- 4. Introduction to R Software**
- 5. Tugas**

# Pengantar

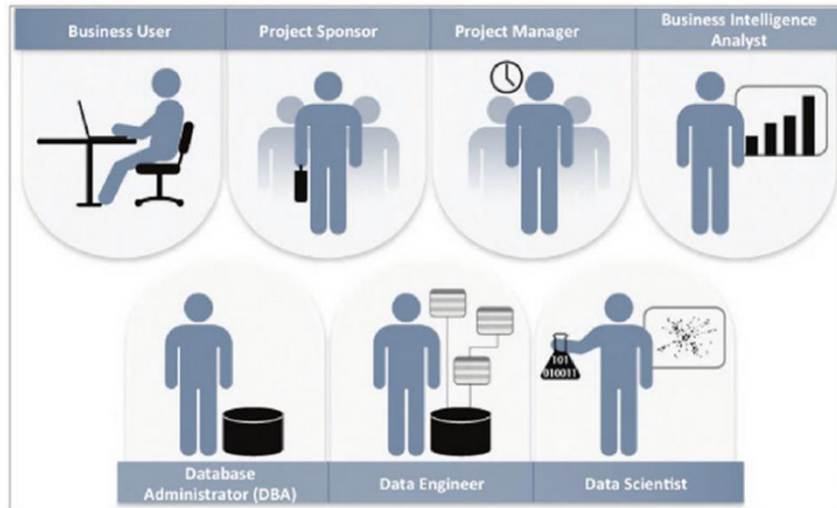
- Fakta-fakta terkait dengan kondisi *existing* Perusahaan:



- Informasi apa saja yang bisa digali dari Big Data pada perusahaan di atas?
- Strategi apa saja yang bisa dilakukan dari masing-masing perusahaan di atas terkait Analitik Data.
  - Key Roles Kunci Sukses Proyek Analitik
  - Lifecycle Analitik Data

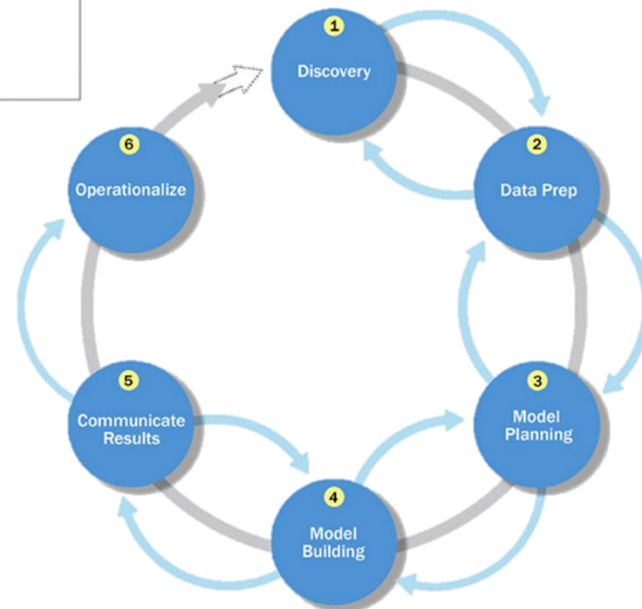
# Pengantar

## ■ Key Roles Kunci Sukses Proyek Analitik



## ■ Lifecycle Analitik Data:

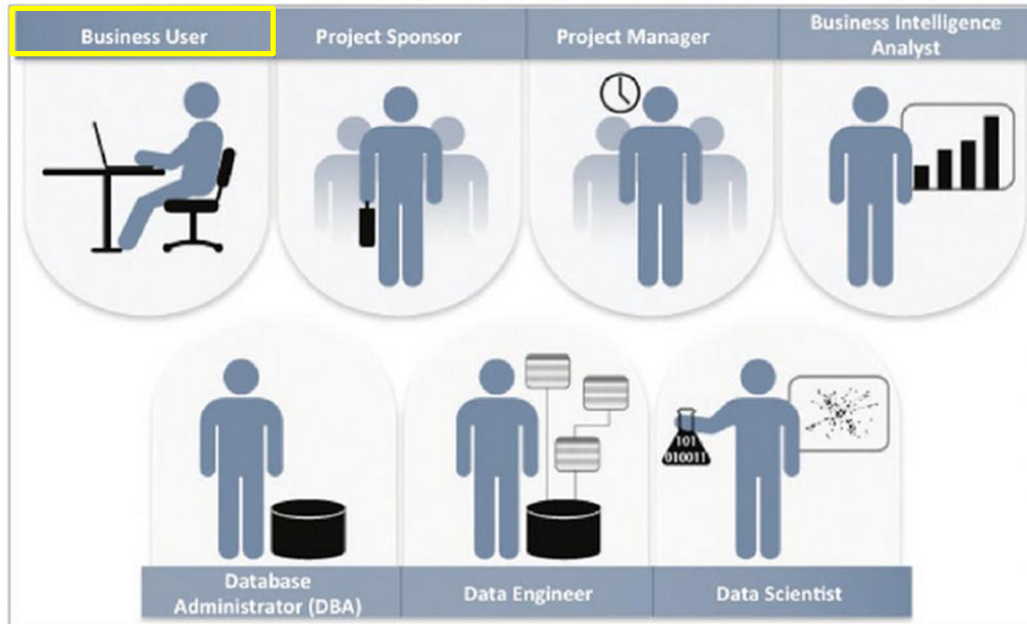
1. Discovery:
2. Data Preparation
3. Model Planning
4. Model Building
5. Communicate Results
6. Operationalize





# Key Roles Proyek Analitik

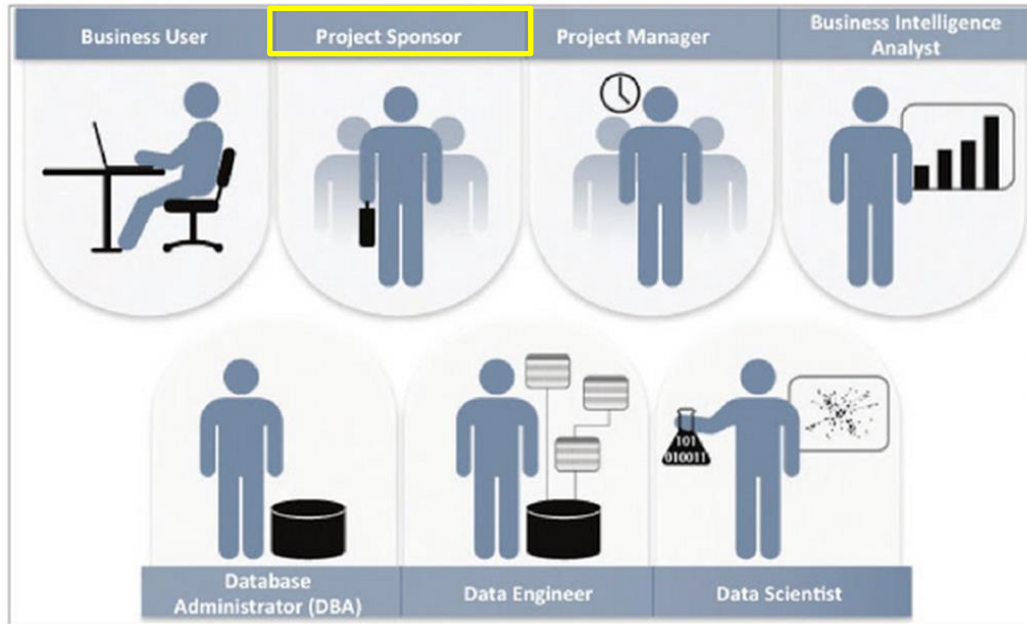
- Key Roles Kesuksesan Proyek Analytics



- **Business User:** Seseorang yang **memahami wilayah domain (kondisi *existing*)** dan dapat mengambil manfaat besar dari hasil proyek analitik, dengan cara konsultasi dan menyarankan tim proyek pada *scope* proyek, hasil, dan operasional *output* (terkait dengan cara mengukur suatu variabel). Biasanya yang memenuhi peran ini adalah **analisis bisnis**, manajer lini, atau ahli dalam hal pokok yang mendalam.

# Key Roles Proyek Analitik

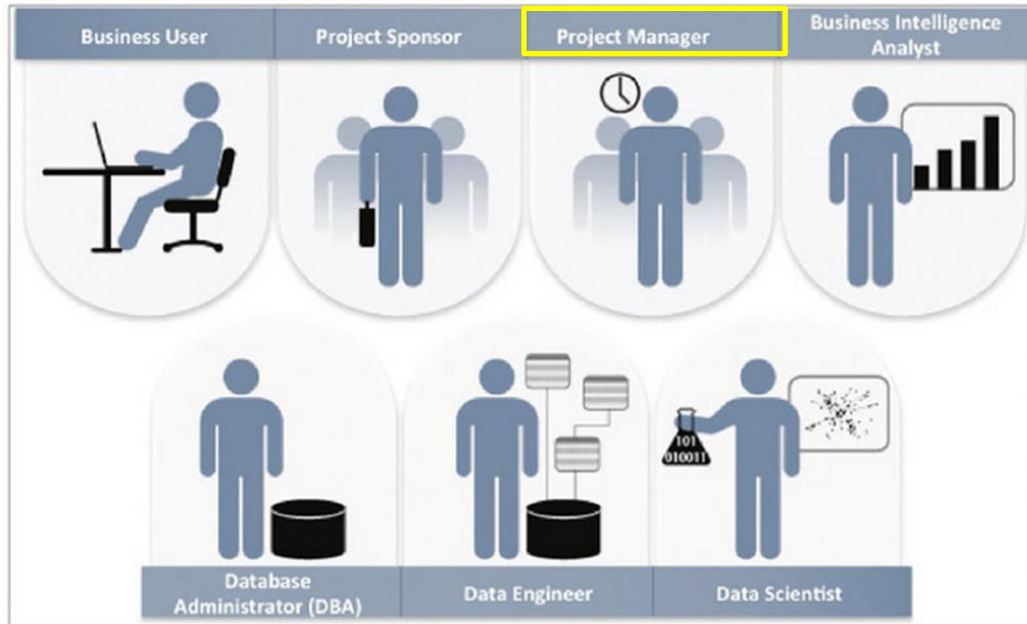
- Key Roles Kesuksesan Proyek Analytics



- **Project Sponsor:** Bertanggung jawab terkait asal proyek. Memberi dorongan, persyaratan proyek dan mendefinisikan masalah core bisnis. Umumnya **menyediakan dana** dan konsep pengukur tingkat nilai *output* akhir dari tim kerja. Menetapkan prioritas proyek dan menjelaskan *ouput* yang diinginkan.

# Key Roles Proyek Analitik

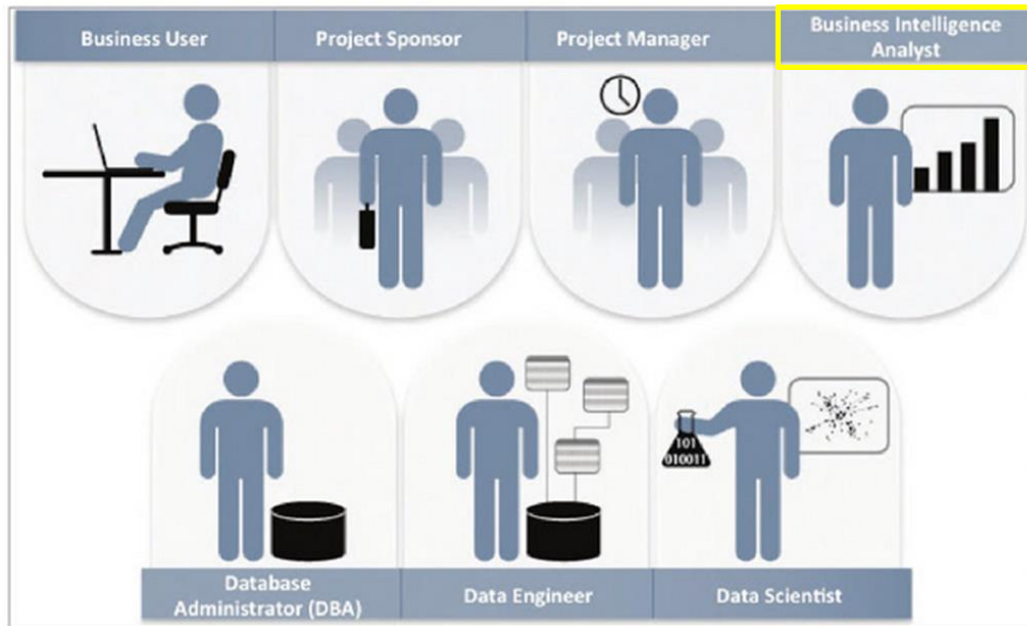
- Key Roles Kesuksesan Proyek Analytics



- **Project Manager:** Memastikan bahwa pencapaian utama proyek dan **tujuan terpenuhi tepat waktu** dan sesuai dengan **kualitas** yang diharapkan.

# Key Roles Proyek Analitik

- Key Roles Kesuksesan Proyek Analytics

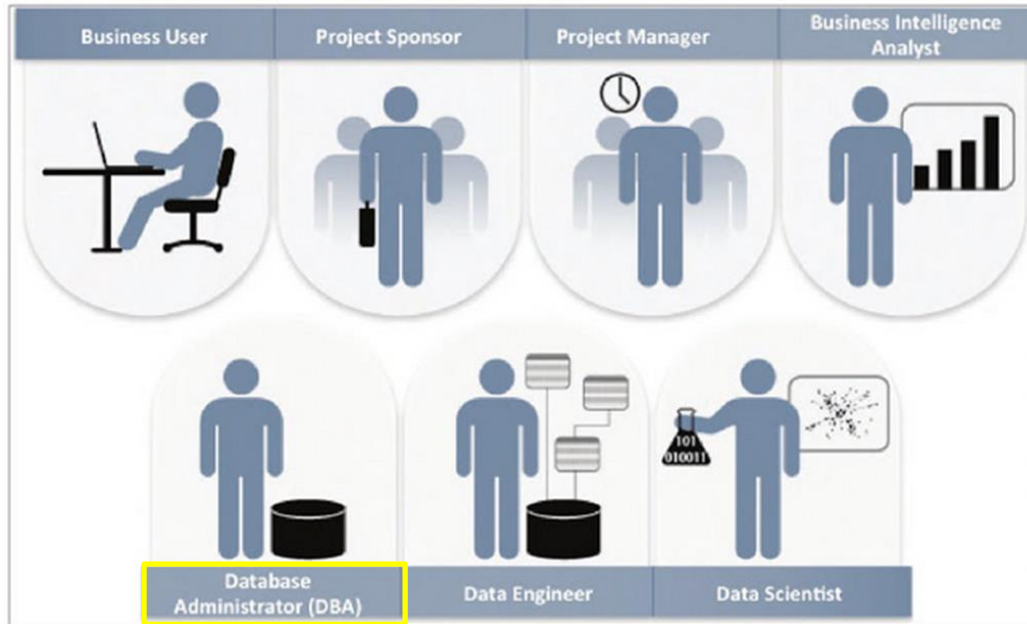


- **Business Intelligence Analyst:** Menyediakan keahlian dalam **domain bisnis** berdasarkan pemahaman yang mendalam mengenai data, indikator kinerja utama (KPI), metrik kunci, dan intelijen bisnis dari **perspektif pelaporan**. Analisis Business Intelligence umumnya membuat dashboard (panel kontrol) dan laporan dan memiliki pengetahuan tentang sumber data dan mekanismenya.



# Key Roles Proyek Analitik

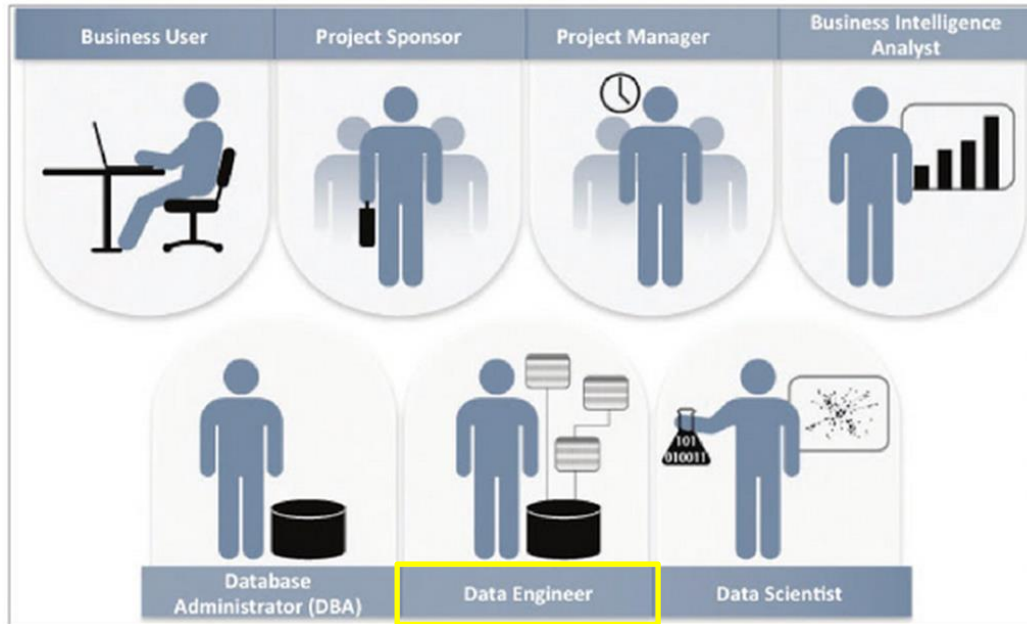
- Key Roles Kesuksesan Proyek Analytics



- **Database Administrator (DBA):** Set up dan mengkonfigurasi database untuk mendukung kebutuhan analitik. Tanggung jawab ini mungkin termasuk menyediakan akses ke database keys atau tabel dan memastikan tingkat keamanan yang sesuai berada di tempat yang berkaitan dengan penyimpanan data.

# Key Roles Proyek Analitik

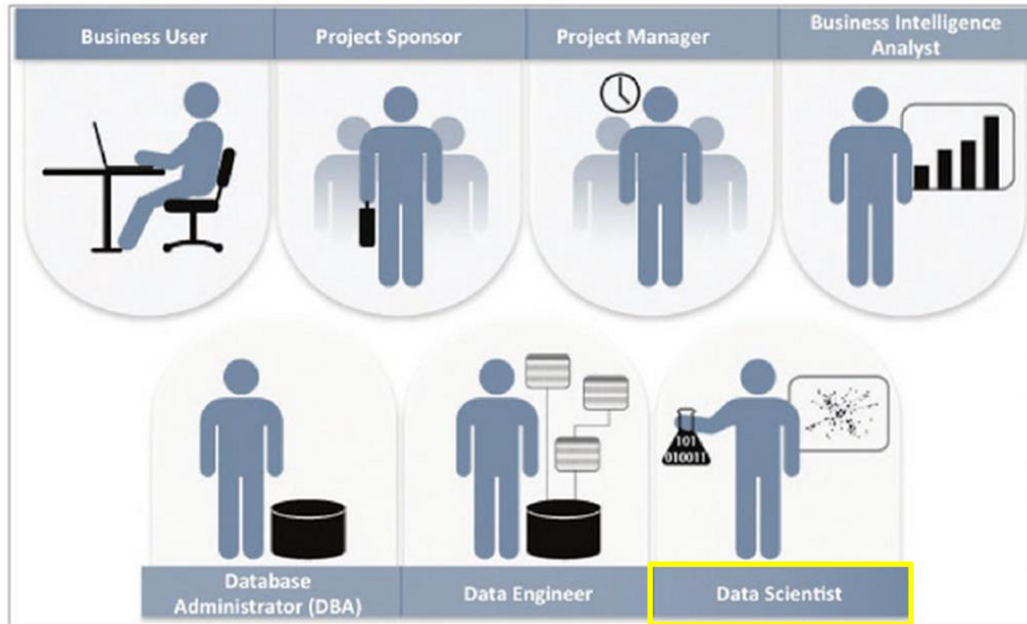
- Key Roles Kesuksesan Proyek Analytics



- **Data Engineer:** Memiliki keterampilan teknis yang mendalam untuk membantu **penyetelan query SQL** untuk **pengelolaan data** dan **ekstraksi data**, dan mendukung untuk konsumsi data ke dalam sandbox analitik. Data Engineer melakukan **ekstraksi data aktual** dan **melakukan manipulasi data** yang cukup besar untuk memfasilitasi kebutuhan proyek analitik. Insinyur data (**Data Engineer**) bekerja sama dengan ilmuwan data (**Data Scientist**) untuk membantu membentuk data yang sesuai dengan cara yang tepat untuk analisis.

# Key Roles Proyek Analitik

- Key Roles Kesuksesan Proyek Analytics



- **Data Scientist (Ilmuan Data):** Menyediakan **keahlian** untuk **teknik analitis**, **pemodelan data**, dan **menerapkan teknik analitis** yang **valid** untuk masalah yang diberikan. Memastikan melalui keseluruhan analitik tujuannya dapat terpenuhi. **Merancang** dan **mengeksekusi metode analitis** dan melakukan **pendekatan lainnya** dengan data yang tersedia untuk proyek tersebut.

# Lifecycle Analitik Data

- Gambaran Umum dari Lifecycle Analitik Data:





# 1. Discovery

- Fase 1 adalah **Discovery**: Pada tahap ini, tim Ilmuwan data (Data Scientist) harus **belajar, mencari dan menyelidiki fakta-fakta, masalah** (identifikasi *problem base*), mengembangkan konteks dan pemahaman, dan belajar tentang sumber data yang dibutuhkan dan yang telah tersedia untuk kesuksesan proyek analitik.
- Selain itu, tim merumuskan hipotesis awal yang nantinya dapat diuji dengan data.



# 1. Discovery

- Fase 1 adalah **Discovery**: Tim belajar domain bisnis, termasuk **kriteria dari data history yang relevan**, seperti, apakah organisasi atau unit bisnis telah mencoba proyek serupa di masa lalu (apa saja yang sudah mereka pelajari dari data). Tim menilai sumber daya yang tersedia untuk mendukung proyek tersebut dari segi SDM, teknologi, waktu, dan data.
- Kegiatan penting dalam fase ini meliputi membingkai masalah bisnis sebagai tantangan analitik yang dapat dibahas dalam fase berikutnya dan merumuskan hipotesis awal (IHs) untuk menguji dan mulai mempelajari data.

## 2. Data Preparation

- Fase 2 adalah **Data Preparation**: Tahap ini membutuhkan adanya sandbox analitik, di mana tim dapat bekerja dengan data dan melakukan analitik selama proyek tersebut. tim perlu melaksanakan proses ekstrak, load dan transformasi (ELT) atau ekstrak, transform dan load (ETL) untuk menyiapkan data ke sandbox.
- ETLT adalah proses integrasi data untuk mentransfer data mentah dari server sumber ke sebuah gudang data pada server target dan kemudian menyiapkan informasi untuk keperluan hasil akhir.

## 2. Data Preparation

- Data Sandbox, dalam konteks Big Data, adalah **platform terukur dan berkembang** yang digunakan untuk **mengeksplorasi informasi besar** suatu **perusahaan**. Hal ini memungkinkan perusahaan untuk mewujudkan nilai investasi yang sebenarnya dalam Big Data.
- Sebuah sandbox data, utamanya **dieksplorasi** oleh tim **Data Scientist** yang menggunakan platform sandbox stand-alone, misal **untuk analitik data marts**, logical partitions pada suatu media penyimpanan di perusahaan. **platform Data sandbox menyediakan komputasi** yang diperlukan **bagi** para **ilmuwan Data** (Data Scientist) untuk mengatasi beban kerja analitik yang biasanya kompleks.



# 3. Model Planning

- Tahap 3 adalah **Model Planning**, di mana tim menentukan metode, teknik, dan alur kerja.
- Mereka berniat untuk mengikuti tahap pembentukan model berikutnya. Tim mengeksplorasi data untuk belajar tentang **hubungan antara variabel** dan kemudian **memilih variabel kunci** dan **model** yang **paling cocok** untuk digunakan.

# 4. Model Building

- Tahap 4 adalah **Model Building**, tim mengembangkan dataset untuk **pengujian (*testing*)**, **pelatihan (*training*)**, dan tujuan produksi (menghasilkan data baru dari data yang ada).
- Selain itu, dalam fase ini tim membangun dan mengeksekusi model yang didasarkan pada kerja yang dilakukan di dalam fase **Model Planning**.
- Tim juga mempertimbangkan apakah ini alat yang ada akan cukup untuk menjalankan model, atau jika itu akan membutuhkan lingkungan yang lebih robust untuk mengeksekusi model dan alur kerja (misalnya, **hardware yang cepat**, teknik **dekomposisi data** dan **pemrosesan paralel**, jika dapat diterapkan).

# 5. Communicate Results

- Pada fase 5 adalah **Communicate Results**, tim bekerja sama dengan pemangku kepentingan (*stakeholders*) utama, **menentukan apakah hasil proyek** tersebut **sukses** atau mengalami **kegagalan** berdasarkan kriteria yang dikembangkan di Fase 1.
- Tim harus mengidentifikasi temuan kunci, mengukur nilai bisnis, dan mengembangkan narasi untuk meringkas dan menyampaikan temuan kepada para pemangku kepentingan.

## 6. Operationalize

- Pada fase 6 adalah **Operationalize**, tim memberikan **laporan akhir, pengarahan, kode, dan dokumen teknis**. Selain itu, tim dapat menjalankan **pilot project** untuk menerapkan model dalam lingkungan produksi.
- Pilot Project adalah sebuah studi percontohan, proyek percontohan atau **studi pendahuluan skala kecil** yang dilakukan untuk mengevaluasi kelayakan, waktu, biaya, efek samping, dan efek ukuran dalam upaya **untuk memprediksi ukuran sampel** yang tepat dan **memperbaiki design penelitian** sebelum kepada proyek penelitian skala penuh.



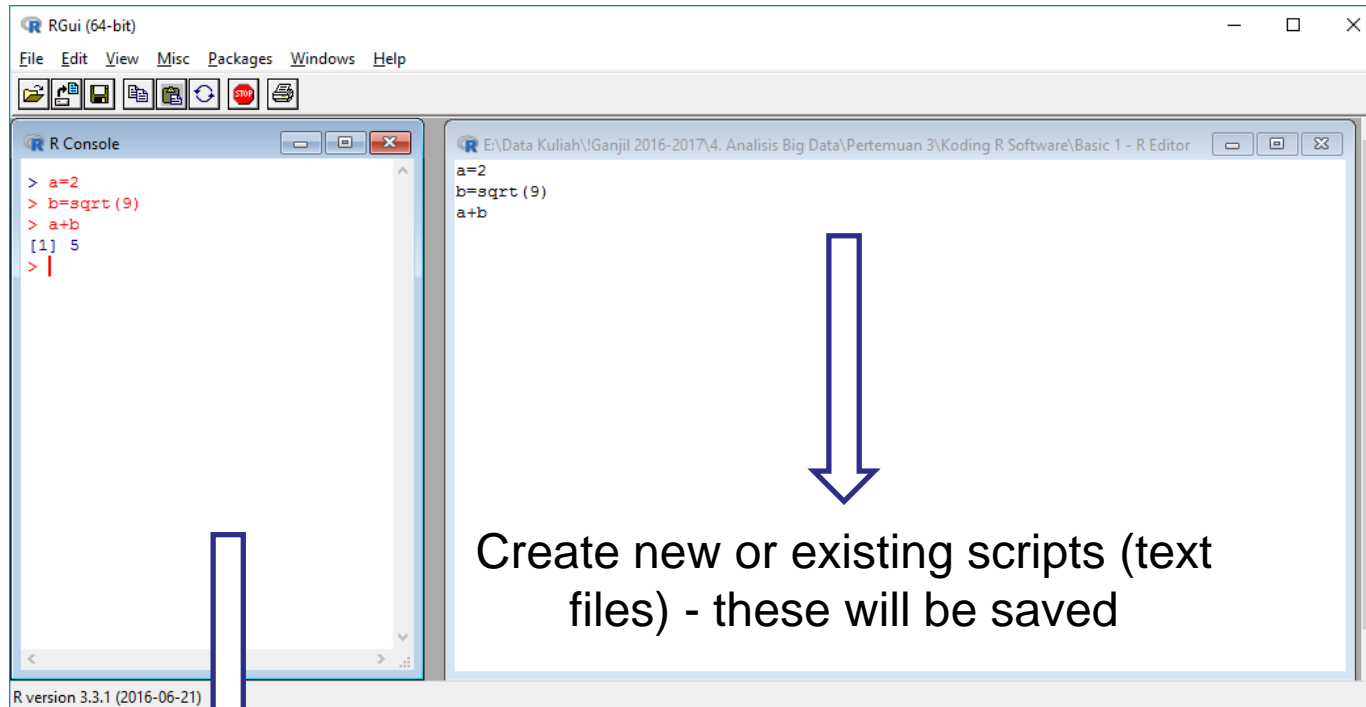
# Introduction to R

- Tentang Software R:
  - Free
  - Didukung dan dimaintain oleh para ahli dibidangnya
  - tersedia di semua platform
  - perbaikannya terus-menerus (continuous)
  - Tersedia di [www.r-project.org](http://www.r-project.org)
- Software R berisi:
  - Basic operations (assignment operation: `a <- 2+sqrt(5)`)
  - Data creation + I/O (`read.csv("c:/data/yearly_sales.csv")`)
  - Component extraction (`d[r,]`: rth row of object d)
  - Plots
  - Basic statistics (mean, stdev)
  - Regression analysis
  - Etc.



# Introduction to R

- Tentang R Console dan R Editor:



Create new or existing scripts (text files) - these will be saved

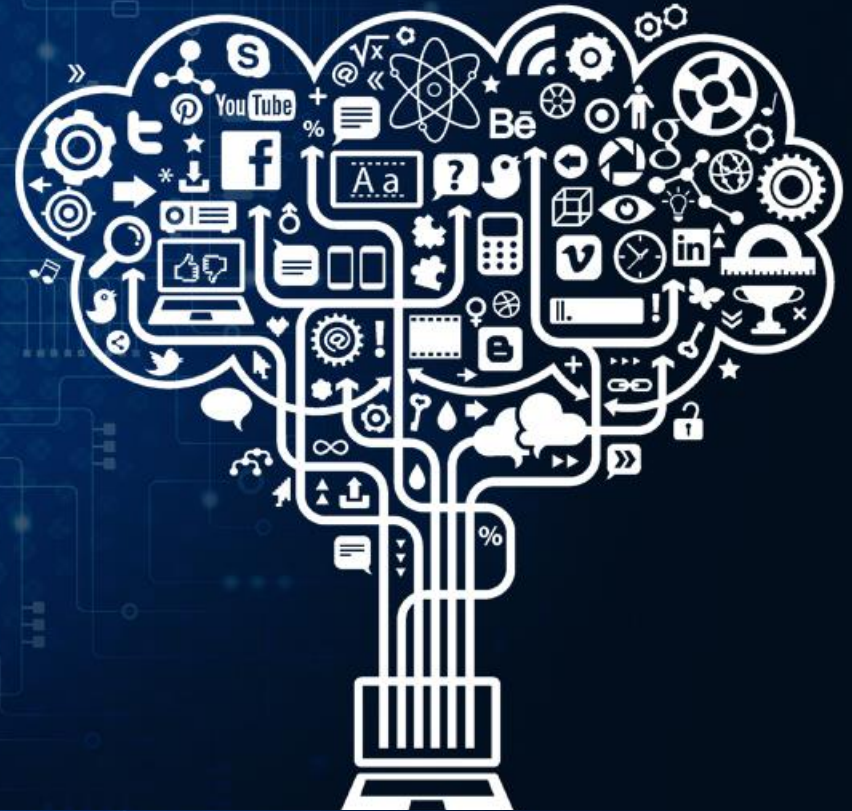
Terminal – output & temporary input - usually unsaved

# Tugas Kelompok

1. Jelaskan apa yang dimaksud dengan Data Sandbox pada konteks Big Data!
2. Jelaskan perbedaan antara Data Science dengan Business Intelligence!
3. Jelaskan perbedaan antara Data Science dengan Data Engineer!
4. Dari beberapa macam “Key Roles Kunci Sukses Proyek Analitik”, manakah 2 pekerjaan paling banyak dibutuhkan pada saat ini, terutama diperusahaan besar?
5. Dari “Gambaran Umum dari Lifecycle Analitik Data”, Buatlah studi kasus dengan mengambil salah satu perusahaan besar yang ada di Indonesia atau perusahaan Asing di dunia untuk melakukan fase ke-1, yaitu **Discovery**. Berikan penjelasan detail terkait hasil penyelidikan anda dari:
  - a. Fakta-fakta (Analisis kondisi *existing* yang ada disana)
  - b. Permasalahan yang ditemukan (identifikasi *problem base*)
  - c. Dari hasil penjabaran permasalahan pada point (b), manakah permasalahan yang menurut anda paling kompleks? Jelaskan!

20 September 2016

# The Rise of the AI in **BIG DATA**



# Terimakasih

Imam Cholissodin | [imam.cholissodin@gmail.com](mailto:imam.cholissodin@gmail.com)