

DAFTAR GAMBAR DAN ILUSTRASI

Gambar I-1 Grafik penyebab data tidak akurat di Amerika Serikat (Thomas, 2013)	1
Gambar I-2 Perkiraan kenaikan volume data tahun 2016 (Lehmann, Roy, & Winter, 2016).....	2
Gambar I-3 Penyebab lemahnya kualitas data tahun 2016 (Lehmann, Roy, & Winter, 2016).....	2
Gambar I-4 Runtunan melakukan data quality management.....	3
Gambar II-1 Proses dari analisis kolom (Olson, 2013).....	8
Gambar II-2 Proses dari analisis struktur (Olson, 2013).....	10
Gambar II-3 Klasifikasi tugas data profiling(Abedjan, Golab, & Naumann, 2016)	11
Gambar II-4 Gambar proses data profiling (Olson, 2013).....	12
Gambar II-5 Alur data profiling berdasarkan HCL technology (R, Srinivasan, & S, 2008).	13
Gambar II-6 Arsitektur Metanome (Papenbrock, Bergmann, Finke, Zwiener, & Naumann, 2015)	14
Gambar II-7 Pipeline untuk menemukan kemiripan entitas pada LOD (Spahiu, Xiey, Rula, Maurino, & Caiy, 2016)	15
Gambar II-8 Hasil profiling dengan OpenRefine (Kusumasari & Fitria, 2016) ...	17
Gambar II-9 Matrik perbandingan Talend Open Studio dan	18
Gambar II-10 Arsitektur three-tier (www.ibm.com).....	23
Gambar III-1 Model konseptual penelitian	25
Gambar III-2 Sistematika penelitian	26
Gambar IV-1 Data quality magic quadrant.....	29
Gambar IV-2 Arsitektur OpenRefine (Github OpenRefine/OpenRefine : Architecture, 2013)	34
Gambar IV-3 Arsitektur aplikasi SQL Power Architech (SQL Power Group Inc., 2009)	36
Gambar IV-4 Arsitektur aplikasi SQL Power Architect (SQL Power Group Inc., 2009)	36

Gambar IV-5 Arsitektur aplikasi Talend Open Studio (Talend, Inc, 2017).....	37
Gambar IV-6 Arsitektur aplikasi Data Cleaner (Human Inference B.V, 2008- 2015).....	39
Gambar IV-7 Arsitektur aplikasi Pentaho (web wiki pentaho).....	40
Gambar IV-8 Arsitektur aplikasi usulan	42
Gambar IV-9 Arsitektur Pentaho Data Integration	42
Gambar IV-10 Usecase diagram	43
Gambar IV-11 Activity diagram login.....	47
Gambar IV-12 Activity diagram admin-Clustering	47
Gambar IV-13 Activity diagram Admin-monitoring.....	48
Gambar IV-14 Activity diagram Managerial-monitoring.....	48
Gambar IV-15 Class diagram.....	49
Gambar IV-16 Sequence diagram login.....	50
Gambar IV-17 Sequence diagram view list DataSource.....	50
Gambar IV-18 Sequence diagram view list Clustering.....	51
Gambar IV-19 Sequence diagram Edit nama.....	51
Gambar IV-20 Sequence diagram Monitoring.....	52
Gambar IV-21 Sequence diagram view List null/blank.....	52
Gambar IV-22 Monitoring manager	53
Gambar IV-23 Halaman login.....	53
Gambar IV-24 Halaman dashboard	54
Gambar IV-25 Halaman tabel detail	54
Gambar IV-26 Halaman tabel fingerprint	55
Gambar IV-27 Halaman editing.....	55
Gambar IV-28 Halaman monitoring admin	56
Gambar IV-29 Halaman monitoring manajerial	57
Gambar V-1 Struktur tabel master tb_pabrik pada Database Lama	59
Gambar V-2 Struktur tabel master tb_trader pada Database Lama	60
Gambar V-3 Pembuatan koneksi database lama.....	60
Gambar V-4 Struktur tabel pada Database Baru.....	60
Gambar V-5 Pembuatan koneksi database baru.....	61
Gambar V-6 Logika clustering dengan metode fingerprint.....	61

Gambar V-7 Konfigurasi tabel input.....	62
Gambar V-8 Hasil konfigurasi tabel input	62
Gambar V-9 Konfigurasi calculator	63
Gambar V-10 Hasil konfigurasi calculator	63
Gambar V-11 Konfigurasi Replace String (punctuation)	63
Gambar V-12 Hasil konfigurasi punctuation	64
Gambar V-13 Pengaturan komponen string operations	64
Gambar V-14 Hasil konfigurasi string operations	64
Gambar V-15 Konfigurasi replace string (mengubah menjadi satu spasi seluruhnya)	65
Gambar V-16 Sampel data sebelum spasi diubah.....	65
Gambar V-17 Sampel data setelah dilakukan replace spasi.....	65
Gambar V-18 Hasil dari <i>sorted</i>	65
Gambar V-19 Hasil dari <i>grouped</i>	66
Gambar V-20 Konfigurasi tabel output	66
Gambar V-21 Penyimpanan hasil clustering pada tabel data_pentaho di Database baru	67
Gambar V-22 Pengaturan komponen Group by	67
Gambar V-23 Proses stream lookup data.....	68
Gambar V-24 Pengaturan komponen Stream Lookup.....	68
Gambar V-25 Pengaturan penyimpanan tabel output data_nama.....	69
Gambar V-26 Hasil dari tabel output data_nama di Databse baru.....	69
Gambar V-27 Logika pengecekan null atau blank.....	70
Gambar V-28 Konfogurasi tabel input.....	70
Gambar V-29 Hasil dari konfigurasi tabel input.....	70
Gambar V-30 Konfigurasi filter rows	71
Gambar V-31 Hasil eksekusi dari filter rows.....	71
Gambar V-32 Hasil dari tabel output data_null di Databse baru	72
Gambar V-33 Implementasi job di Pentaho.....	72
Gambar V-34 konfigurasi interval jalannya job.....	72
Gambar V-35 Konfigurasi transformation clustering	73
Gambar V-36 Konfigurasi transformation show null	73

Gambar V-37 Tampilan halaman login.....	74
Gambar V-38 Tampilan halaman dashboard admin	74
Gambar V-39 Isi dari tabel data_nama	75
Gambar V-40 Tampilan halaman tabel detail.....	75
Gambar V-41 Tampilan halaman tabel clustering	76
Gambar V-42 Tampilan halaman edit clustering	76
Gambar V-43 Tampilan halaman setelah data terupdate	77
Gambar V-44 Perubahan data setelah di update pada website	77
Gambar V-45 Gambar V 44 Perubahan data setelah di update pada Database baru	78
Gambar V-46 Tampilan halaman monitoring admin	78
Gambar V-47 Tampilan tabel data_monitor di Database baru	79
Gambar V-48 Tampilan halaman setelah data monitoring tersimpan	79
Gambar V-49 Tampilan halaman monitoring	80
Gambar V-50 Tampilan halaman list null.....	80
Gambar V-51 Tampilan halaman dashboard manager.....	81
Gambar V-52 Tampilan halaman monitoring manager	81
Gambar V-53 Data pabrik.....	86
Gambar V-54 Melakukan clustering pada nama pabrik	86
Gambar V-55 Hasil clustering dengan OpenRefine.....	86
Gambar V-56 Pemilihan nama baru.....	87
Gambar V-57 Hasil dari pemilihan nama baru	87
Gambar V-58 Data trader.....	88
Gambar V-59 Melakukan facet blank pada kolom TELP.....	88
Gambar V-60 Hasil shownull dengan OpenRefine.....	88
Gambar V-61 Pembuatan koneksi database MySQL pada Talend	89
Gambar V-62 Sampel hasil profiling dengan Talend Open Studio	89
Gambar V-63 Pemilihan kolom untuk di cluster	89