###############################################Lecture 1#############################################

| ability:36  acceptable:40  accessed:65  accounting:47  accumulate:10  accumulated:10  accumulating:6  accurate:36  achieve:22  acquired:56  acquisition:56-54  action:35  actionable:24-27  actions:35-22  activities:11  actual:40-59  addition:6  additional:10  adheres:16  administered:34  admission:33  adoption:2-37-44-46-47-48-49-51-52  advertisement:63  advertising:25  advisory:48  afternoon:25  agenda:2  agent:32  aggregation:58-54  agriculture:62  aim:33  alerts:65  algorithms:28  alignment:48-49  amounts:7  analysis:4-38-55-56-57-59-60-63  analysts:22  analytics:1-2-21-22-23-24-25-26-27-28-29-30-31-32-33-34-35-36-37-44-46-47-48-49-51-52-53-54-55-56-57-58-59-60-63  analyze:4-22-68  analyzed:66-59-60  analyzing:67-28-29-38  anna:47  answer:32-51  answering:33  answers:59  antarctica:18  application:65-22  applications:64-65-2-66-67-68-69-61-62-63  approach:59  architecture:52  arrive:57-10-58  ascend:46  aspect:39  assess:66  assessed:11  assumption:59  assumptions:59  audio:17  automated:56  automobiles:6  average:41  backing:29  banking:64  based:41-68-29-63  basic:2-20-21-22-23-24-25  ben:47  big:1-2-3-4-5-6-7-8-9-10-11-12-13-14-15-16-17-18-19-21-24-25-29-52-57-62-63-64-68  billions:5  blogs:17  boosting:68  borrower:64  bridges:66  brings:11  broad:28  browsing:63  build:35  building:46  buildings:66  burgers:25  business:2-37-38-39-41-42-12-44-46-47-48-49-51-52-54-55-60  businesses:6  buying:25  called:59-15  calls:32-33  cameras:65  campaigns:25-68-63  capable:29  capture:16  car:65  card:64  carried:32  carrying:59-55  case:54-55  categories:32-33-2-34-35-36-30-31  categorized:32-9  center:17  challenges:11  chances:34  change:51  characteristic:12  characteristics:8-9-10-11-12-13  cities:62  claim:64  classified:59  cleansing:57-28-54 | clear:55  closely:59  cloud:63  collect:63  collected:17-66  collecting:28  collection:63  collections:4  commission:32  commitment:43  common:17-58-19-21  communicate:60  communication:6  company:47  compares:40  complete:28  complex:57  complexity:36  concepts:2-20-21-22-23-24-25  conclusions:22  conditions:67  confirmatory:59  conform:17  conformance:19  connected:65  consistency:19  consolidates:40  content:5-63  context:22-39  continuous:43  convert:52  core:48  corporate:49-51  correct:51-38  correlation:25  corrupt:56  create:25-5  created:10-68  creating:27  credit:64  critical:43  csfs:43  current:67  customer:34-68-6-41-43  customers:48-68  cybersecurity:62  daily:6  dashboard:40  data:1-2-3-4-5-6-7-8-9-10-11-12-13-14-15-16-17-18-19-21-22-23-24-25-26-27-28-29-30-31-32-33-34-35-36-38-43-44-46-47-51-52-53-54-55-56-57-58-59-60-62-63-64-66-67-68  database:16  datacharacteristics:2  datasets:10-4-58-55  day:46  deals:63  decides:25  decision:29  decisions:29-22  dedicated:4-57-58-59-60  deductive:59  deemed:56  deep:36  default:34  define:40  defined:12-16-19-22-55  definitive:59  degree:48  delivers:48  delivery:63  demand:68  demands:10  demonstrates:36  department:41  departments:47  dependent:12  depicted:44  dept:47  depts:47  describing:32  descriptive:32-36-31  detect:65  detected:38  detecting:64  detection:64-65-69  determine:33  determining:67-60  develop:59  developed:29  deviations:67  devices:65-6  diagnosis:67  diagnostic:33-31  differentiate:9  difficult:4  dikw:49-44-46-47  discounts:68  discovers:25  disparate:57  display:40  displayed:40  disprove:59  distinguished:31  distributed:29  divided:54  documents:17  domains:62 | door:65  draw:22  driven:29  drivers:65  earned:32  easily:19  eastern:33  education:43-62  effective:60  efficient:27  effort:38  elevator:46  email:17  embedded:6  employee:47  employees:47  enable:33-29  enables:38  encapsulate:60  encompasses:28  ends:65  energy:62-6  engagement:63  engineer:25  engines:5  enhance:51-38  enhances:52  enormous:10  enriched:51  enterprise:41-12-38  enterprises:11  enters:11  entities:16  environment:27-69-62  environments:10-11-29  equation:60  establishing:57  evaluation:54-55  event:33-34  events:32  examined:51  examples:41-43-46-47  executes:48  executives:48  exhibit:48  exist:60  expedition:18  experience:29  exploratory:59  explored:59  explosive:7  extract:28-22  extracting:57  extraction:57-54  factors:43  factual:29  fail:64  fast:10  fault:67  feet:46  fidelity:11-12  field:4-21  fields:58  figure:40-36-13-44  files:17-19  filtering:56-54  financial:64  floods:69  floors:46  focus:33-47  focused:21  forecasting:34-68  foresight:36-44  forest:69  form:16-32  format:57  formats:11-15  forms:9-19-17  frameworks:28-29  fraud:64  frauds:64  fries:25  fundamentals:1  future:64-34-36  gain:36-38  gained:21  gathered:56  general:31  generally:22  generate:52  generated:38-6-13-51-59  generation:7  geographic:32  gigabytes:7  goal:33-22  goals:48-41-22-55  governing:28  governmental:6  graphically:60  grouped:68  groups:48  growing:5  growth:41-7  handle:47  happen:33-34-35-36 | happened:32-36  happening:51  health:66  healthcare:62  height:46  helps:52-22  high:25-10  higher:12  highly:29  hindsight:36-44  hire:47  hiring:47  history:63  hit:42  holds:12  homes:62  howthe:51  human:47  humans:22  hypothesis:59  identification:54-55  identified:56  identifying:55  illustrates:49  illustrations:13  imagery:6  images:17  impacted:13  implies:21  impose:10  impossible:4  improvement:43  improving:68  include:32-33-34-35-17  includes:34-28  including:17-62  incompatible:57  increase:33-36  increasing:68  increasingly:21  independence:48  indicators:40-41-42-39  inductive:59  industry:67-62  infer:27  influenced:48  info:46  ingested:10  insight:51-36-38  insights:60-24-25-27-28-63  instance:51  institutions:64-6  insurance:64  integrating:58  integration:11  intelligence:38  interactions:17  interesting:22  interests:63  internal:15  internet:65-66-62  interpretation:60  intrusion:65  intrusions:65  intuition:29  invalid:57-11  investigated:59  involves:59  iot:65-66  issues:38  iterative:59  jack:47  jane:47  joined:58  json:19  justification:55  kathy:47  key:40-41-39  knowledge:44-46-47-51-52-21-22-27  kpi:40-41-39  kpis:40-41-42-39  large:4-29-22  laundering:64  layer:48-51  layered:48  layers:48-52  lead:11  learning:23  lecture:1  lens:51-44  level:48-51-19  levels:49-66  leverage:64-63  leveraged:60  lifecycle:2-60-53-54-55-56-57-58-59-28  lifetime:41  limited:62  loan:34  location:32  logistics:62  long:12  longer:12  lydia:1  machine:67-23  making:27-29 | managed:10  management:10-43-28-38  managerial:48-51  margin:41  mark:47  massive:7  mathematical:60  meaning:27-52  meaningful:12-28  measure:41-39  measurements:40  measuring:39  media:17  medicine:34  medicines:35  messages:17  meters:6  methods:29  metric:39  metrics:42-51  middle:48  mining:21-22-23-24-25-59  missed:34  model:16-17-60  modeling:64  models:34-35-60  money:64  monitor:66  monitoring:65-66-69  monthly:32-41-34  months:32-33  motivation:55  multiple:40-35-58-11  music:63  named:9  nature:59-19-60  navigation:65  needed:36  network:66  networks:5-6  noise:11-69  normal:67  number:32-65-47  numbered:7  objectives:48-56-42-27  occupied:48  occur:29  occurred:32-33  occurrence:34  offer:25  offers:68  online:63  operates:48  operating:67  operational:49-51  operations:48-27  order:33-51  organization:48-10-38-15  organizational:52-38  organized:27  organizing:28  orientation:43  originating:33  outcome:34-35  outcomes:35  overview:62  parking:65  parkings:65  patient:33-34  patterns:32-68-22-60-63  pay:64  payment:34  people:25  perform:65  performance:67-38-39-40-41-42-51-63  performing:51  periods:10  personalized:68  petabytes:7  phenomenal:7  phenomenon:33-59  phones:65  pir:65  pis:42  points:39  pollution:69  posts:17  powered:65  powerpoint:17  pre:16  predetermined:59  predict:64-35-36-68  predicting:34-67  prediction:34-35  predictive:34-35-31  preferences:68-63  preparation:10  prescribing:35  prescriptive:35-36-31  presentations:17  presents:55  previous:56  primary:15  process:42-27-4-47  processed:2-10-14-15-16-17-18-19-27-60 | processes:38-10-42-48-28  processing:10-11-4-12  produce:35-60-31  product:34  products:34-68  profit:41  profitability:68  prognosis:67  progress:27  progression:24  project:55  proposed:59  prove:59  provide:59-51-52-62  purchase:34-68  purchased:34  purpose:57  pyramid:49-44-46-47  quality:43-11-69  quantifiable:39  questions:32-33-34-35-51-59  raise:65  range:40-17-62  rate:34  rates:33  ratings:63  raw:25-27  reason:33  received:32  recommendation:63  recommendations:68  recommended:68  recommending:63  record:6  recruitment:47  refer:15  reference:39  refers:65-67-7-10-11  region:33  relational:16-19  relationships:16-60  remote:65  removal:56  remove:11  removing:57  report:51  reporting:43  repository:7  representation:58-54  required:55  requiring:58  resolve:11  resources:47  responses:17  result:7  retail:68-62  retailers:68  retention:41  return:5  reviewing:25  risk:64  river:69  rose:47  rules:57-60  sale:68  sales:32-33-68-41-47-25  sample:32-33-34-35  satellite:6  satisfaction:43-68  scalable:29  schema:16-17  science:21-23  scientific:29  score:64  search:5-63  security:65  seek:33  seeks:48  semi:19-15  send:65  sensing:65  sensors:65-66-6  sequence:24  serve:63  set:60  sets:22  severity:32  short:10  simply:42-51-29  single:7  sites:17  situation:36  size:10  slots:65  smart:65-6  smarter:27  smartphones:6  social:17-5  solution:57  solutions:2-11-14-15-16-17-18-19  sources:56-29-55  span:62  specific:41-42-59-39  speed:10  speeds:10 | spread:58  staff:43  stage:56-57-58-59-60  stages:54  stairs:46  status:42  steer:48-38  stock:35  storage:10-11-4  store:4  stored:16  storing:28  strategic:48-49-51  strategy:48-51  streaming:63  structural:66  structure:19  structured:16-27-19-15  structures:66  study:21  subjected:56  success:41-43-39  summarized:32  sunday:25  supermarket:25  supplier:6  support:32-33-22  supported:11  survey:17  survival:34  systematic:21  systems:64-65-66-68-38-22-27-62-63  tablets:65  tabular:16  tactical:48-49  takes:12  tall:46  targeted:25-68  targets:41-42  task:59  techniques:4-60-22-23  technologies:28-29  terabytes:7  term:28-21-7  terminologies:38-39-40-41-42-43  terms:11-21  text:17  textual:19  threshold:40  ties:52  time:10-35-13  tools:60  toronto:46  tower:46  track:42  tracking:63  trade:35  traditional:4  training:43  transaction:6  transactions:5  transcripts:17  transformation:11  transforming:57  transition:44  triggering:65  turned:12  type:33  types:2-5-41-11-14-15-16-17-18-19-59-63  typically:59  underlying:57  understandable:22  understanding:59-36-55  understandings:60  understood:44  unified:58  unstructured:17-18-19-15  user:63  users:27-60-5-63  utilization:60-54  valid:22  validation:57-54  valuable:25  values:40-34  variety:9-11-4  velocity:9-10-4  vendor:6  veracity:9-11-12-13  vibration:66  video:17-18-63  view:58  visits:63  visualization:60-54  volume:32-9-10-4  volumes:10-68-29  wahid:1  water:69  weather:69  web:5-63  websites:5-63  western:33  whatis:51  whythe:51  wide:17-5-62  wisdom:51-44-46-47  word:17  working:46  works:47  worldwide:5  xml:19 |
| --- | --- | --- | --- | --- | --- | --- |

###############################################Lecture 2#############################################

| actual:11  address:24  addresses:24  adds:27  agenda:2  algorithm:2-8-9-10-11-12-13-14-15-16-17-18-19-20-21-22  analytics:24  appearing:16  apply:10  applying:16-18-11-13  arrive:4  assigned:15  assume:15  average:28-30  based:34-4-36  big:1-4  broken:4  called:11  cluster:5  clusters:4  code:24  collected:24  collection:4  columns:24  combined:4  combiner:21  common:11  composed:5  comprising:33  computation:33-27-36-30  compute:25-27-30-33  computing:24  concurrently:7 | conform:5  conquer:4  consists:5  constituent:11  contact:34-36  copied:16-22  count:25-10-27-33  counts:33  create:22  creates:16  criteria:34-36  customized:16  data:1-4-5-6-7-36-24  dataset:10-11-4-12  datasets:11-4  dec:34-36  defined:18-13  definition:2-3-4-5-6-7  depending:18  deployed:4  directory:11  distribution:15  divide:4  divided:21  divides:11-4-15  emit:18  emits:33-27-36-30  essentially:21  examples:33-2-34-35-36-23-24-25-26-27-28-30-31  executed:5  executes:13  exist:13  explain:10 | explanation:33-27-36-30  extract:21  extracts:36  fair:15  field:27-30  fields:36  figure:6  file:24  files:11  filter:34  filtering:34-36  filters:36  final:18-19  finding:33  finds:30  finished:7  flow:6  format:25-35-28-31  function:33-36-11-13-18-21-22-27-30  generates:13  group:16-27-30  grouped:33-27-30  groups:21-22  guarantees:15  hardware:4  html:34-36  http:36  independently:4-7  input:33-35-36-5-9-10-11-12-13-16-18-20-21-25-27-28-30-31  instances:15  involved:15  items:30  job:33-5 | key:33-36-11-13-15-16-18-21-27-30  large:4  lecture:1  length:24-30  list:33-13-16-22-30  local:36-21  locally:21  log:24  logged:24  logic:18-13  lydia:1  making:18-15  map:36-5-7-9-11-13-14-20-21-22  mapper:33-36-11-13-14-15-21-27-30  mappers:15  mapreduce:1-2-3-4-5-6-7-8-9-10-11-12-13-14-15-16-17-18-19-20-21-22-23-24-25-26-27-28-30-31-33-34-35-36  maps:7  merge:16  merged:16  merging:16  model:5  month:34-36  multiple:4-5-13-15-16  network:16  node:21-22  nodes:16-21  number:25-10  object:18-13  occurrences:10  operations:4  order:22  ordinal:11  output:32-5-9-10-13-15-16-18-19-20-21-22 | pair:16-21-11-13  pairs:33-36-11-13-16-27-30  parallel:4-21-7  parsed:11-36  parses:33-27-36-30  partition:15  partitioner:16-15  partitioners:16  partitioning:21  partitions:15  parts:4  performed:4-21  performs:21  phase:5-9-18-20-21-22  phases:5  position:11  principle:4  problem:10-4  problems:4  process:7  processed:13  processes:4-13  processing:1-4-36  product:20-21-22  products:20  programs:24  quantities:21-22  quantity:20-21  random:15  read:11  receives:33-27-30  record:24-11  records:34-36-11-13-15-21 | reduce:32-5-7-9-16-18-19-20-21-22  reducer:33-15-18-19-21-27-30  reducers:32-15  require:5  required:33-10-30  requires:36  respective:21-7  rest:36  result:4  return:11  row:24  running:16-21  runs:22  sales:21  sends:24  sequentially:5  service:24  shuffle:16-32-22  shuffling:16-17  single:16  smaller:11-4  solved:4  sort:16-32-22  sorted:16  sorting:16-17  sorts:33  spent:28  split:9-11-12-13-21  splits:11-21  stage:32-10-11-12-13-14-15-16-17-18-19-22  stages:5  statement:10  statistics:24 | step:33  subset:34-36  substring:18-13  summarize:18  summation:21  sums:33-22  takes:11  task:16-36-21-22  tasks:21-22-7  technique:4  techniques:1  text:11  time:24-28  times:16  timestamp:24  tracking:24  tuple:33-36  txt:21  unique:22  url:33-36-24-27-30  urls:33  user:18-13  values:16-33-27-30  visit:24-33-30  visited:33-31  visitor:24  visits:24-25-34-33  wahid:1  wait:7  web:24  website:24-25-28-31  word:10  year:36  yyyy:24 |
| --- | --- | --- | --- | --- | --- | --- |

###############################################Lecture 3#############################################

| abstracts:35  accepting:29  access:32-20-21  achieve:41  achieved:50-4  acquiring:38  actions:41-43  active:16  actual:10-38  acyclic:44  adjust:13  advanced:36  agenda:2  aggregate:43-21  aggregated:42  algorithm:48  algorithms:48-51-36-22  allocate:27  allocates:32-26-28  allocating:29-38  allowed:32  amazon:22  analysis:51  analytics:35-20-22-36  analyzing:51  apache:2-7-8-9-10-11-12-15-16-17-19-20-21-22-23-24-34-35-36-37-38-39-40-41-42-43-44-45-46-47-48  api:48-35  apis:36  application:36-37-38-44-23-27-29-30  applicationmaster:27-29  applications:32-33-12-15-51-26-28-29  applicationsmanager:28-29  applied:41  apply:41  architecture:32-38-39-9-10-11-12-28-29-30-31  archive:51  argument:42  array:43  assigned:12-38  assignment:28  automated:23  automatically:16  average:33  avoid:12  balanced:23  balancing:14  based:45-29  basic:2-3-4-5-6  batch:32-48-50-6  batches:6  behalf:45  big:1-35-51-20-22  bigtable:20  blacklist:12  block:13  blocks:37-9-10-13-15-50  blocksize:13  breaking:48  budget:51  build:35  building:51-37-23  built:48-36-47  bundled:44  burdens:35  called:42-19-44-14  capabilities:23  capability:17  capable:48-9-37  cases:51  central:33  centralized:24-23  chain:50  chains:51  changed:13  chosen:12  classification:22  client:12-44  clients:10  cluster:32-33-4-36-37-38-8-9-10-11-12-44-45-15  clustering:22  clusters:8-9-32  code:44-38  collection:40-37  collectively:4  companies:20-22  comparator:43  compatibility:32  compatible:32  complete:45  completed:12 | completing:51  completion:28  components:33-38-9-10-11-46-27-28-29  comprise:4  computation:8-9-41-15-48  computations:9  computed:41-37  computers:8-10-35  computing:17-35-36-47  concept:48-41  concepts:2-3-4-5-6  concurrently:4  conditions:16  configuration:8-24-13-23  configured:17  connection:39  considered:12  consists:9  consume:9  container:29-30  context:14-39  continuous:48  converting:44  converts:44  coordination:24  copy:16  core:47  cores:4-30  cost:14  count:43  cpu:13-30  create:40-41  creates:9-44  creating:19  criteria:33  custom:43  customers:35  cycles:50  dag:27-44  data:1-4-5-6-9-10-11-12-13-15-16-20-21-22-32-35-36-37-38-40-45-48-50-51  database:10-20-39  databases:21  datanode:10-12  dataset:40-42-43-37  datasets:37-8-47-15-19  datastores:47  days:6  dealing:51  debugging:17  decides:12  default:17-14  delivers:48  demand:26  depicts:46  deployed:10-11  designed:8-16-50  desired:14  detection:22  determine:12  developed:19  developers:35  directed:44  directories:10  disk:50-36-30  disruptions:32  distribute:35  distributed:35-4-5-37-8-9-10-42-15-17-19-20-21-23-24  distributes:9  distribution:9  divided:37-15  divides:10  dividing:4  document:22  drastically:48  driven:13  driver:38-40-41-44-45  duplicate:22  dynamic:32  dynamically:27  eager:41  easily:16  easy:35  ecosystem:2-46-47-48-18-19-20-21-22-23-24  efficiencies:48  efficient:32-41  efficiently:50-15  element:42-43  elements:40-42-43  enables:35-9-47-48-24  engine:48-21-47 | engines:32  enormous:9  enterprise:23  entry:38  environments:51  event:16  examples:42-43  exceeds:51  excellent:33  executed:38  executers:38  executing:45-29-30  execution:4-45-44-47  executors:45-38  existing:32-40  exporting:21  extend:32  extended:48  external:40-38-47  extraction:22  extremely:9  facebook:19  facilitates:19  failed:12  failover:23  fails:16-12  failure:16-37-29  failures:14  faster:41-50-36-21  fault:37-40-16-17-50-23  feature:36  features:32-36-15-16-23  file:40-9-10-17  files:10-13  filter:42  finished:30  firm:22  fixed:33  formed:42  forwards:28  framework:8-35-14-30  frameworks:1  fully:17  func:42-43  function:42-43-29-38  functionalities:26-39  functionality:47  functions:21  fundamental:26-37-38  future:45  gateway:39  general:47  generated:6  global:27  goal:4  google:20  graph:48-51-36-44  graphx:48  grouped:14  groups:6  hadoop:1-2-7-8-9-10-11-12-14-15-16-17-18-19-20-21-22-23-24-25-26-27-28-29-30-31-32-33-36-49-50-51  handle:50  handling:15  happening:6  hardware:15  hbase:40-20-21  hdfs:33-40-9-10-15-16-17-47-21  health:30  helps:36  high:48-15  higher:13  highly:16-24-23-15  historical:51  hive:19-21-47  hiveql:19  hours:6  huge:10-15  hundreds:15  idea:26-27  ideally:11  illustration:46  impala:21  implicitly:44  importing:21  improves:48  include:28-29  included:10-11  includes:22  including:21-47  increases:36-14-15 | increasing:14  indexing:23  individual:30  infrastructure:51  input:17-13  inputs:14  instruct:41  instructed:11  instructions:6  int:13-14  interact:8  interaction:6  interactive:32-6  interface:47  intermediate:50-36  internal:10  internet:23  intersection:42  introduced:26  involved:15  involves:4  iterable:14  iterative:48-51  java:8-17-36  job:38-12-14-26-27-29-30  jobs:38-11-12-51-26-27  jobtracker:33-26-11-12  join:21  json:47  key:42-14  language:19  languages:19-36  large:8-9-19-35  larger:4  largest:23  latin:19  launches:45  lazy:41  leads:32  learning:48-36-22  lecture:1  legal:14  level:13-38  library:48-22  light:13  limited:51  list:14  load:14-23  local:8-17  locates:12  location:12  logical:44-37  logically:44  lost:16-50  lowers:14  lucidworks:22  lydia:1  machine:4-36-10-11-16-17-48-22  machines:8-17-4  magnitude:21  mahout:22  main:36-38-9-28-29  maintaining:24  maintains:16-10  making:32  manage:32-15  managed:10-11  management:33-26-27-28  manager:32-38-45-28-29-30  manages:11  managing:33-19  manipulated:48  manner:32  map:32-33-36-9-42-11-13  mapreduce:33-9-41-11-12  maps:11-13  mark:12  master:38-10-11-26-28-30  measured:6  mechanism:10  memory:36-47-48-50-51  metadata:10  method:14  microbatches:48  microseconds:6  microsoft:21  mining:22  mllib:48  mode:17  model:51 | modeled:20  modern:21  modes:17  monitor:45  monitored:12  monitoring:26-30  monitors:26  mrjobconfig:13  multi:32  multiple:32-35-4-36-37-9-15-17  mutual:21  mysql:21  namenode:10-12  naming:24  natural:43  navigation:23  negotiates:45  negotiating:29-30  negotiator:26  network:15  networked:4  node:38-11-13-45-16-50-28-30  nodes:32-37-38-8-9-10-11-12-45-15  note:10  notify:12  num:13  number:36-43-13-14-50  numpartitions:42  objects:37  offering:8  offers:17  open:8-20-23  operated:40  operation:17  operations:41-51-36  optimization:33  optimizations:44  oracle:21  order:43-21  ordering:43  original:50  otherdataset:42  output:17  overcome:16  overhead:14  pair:14  pairs:42  parallel:4-5-40-9-51  parallelism:13  parallelizing:40  partitioned:40-37-38  partitions:37  pass:41  passing:42  perform:8-41-35-38  performance:48-50-21  performs:9-26-44-29  person:6  personalized:20  phase:14-22  physical:44-30  physically:4  pig:19  place:15  placement:45  plan:44  platform:35-47-16-19-20-23  point:45-38  poll:12  popular:22  portability:16  portable:16  powerful:23-15  powers:23  primary:19  prime:35  problems:16  process:48-17-30  processed:10  processes:17  processing:32-1-35-4-5-6-36-8-9-38-48-50-51-26-27-28  processors:4  production:17  program:40  programming:35  programs:19  progress:30  provide:10-6  providing:24-23  pseudo:17 | python:36  quality:48  queries:19-36-47  query:21  querying:23  ram:30  random:20  rapid:9  rdd:37-40-41-42-43  rdds:40-41-37-38  read:36-40-10-50-20  reading:19-47  reads:50  real:32-35-6-50-51-20-21  realtime:20  rebuilding:37  receives:28  recommender:20  recomputed:50  record:12  records:40  recovery:23  reduce:32-33-4-36-9-42-11-43-14  reducebykey:42  reducers:14  reduces:50-15  reducing:36  reduction:14  referencing:40-47  relational:20-21  reliability:17  reliable:24-9-23  replicas:9  replicating:50  replication:16-23  reporting:30  request:28  requested:15  requests:10  required:6  requirements:29  residing:19  resilient:37  resource:32-33-26-27-28-29-30  resourcemanager:27-28  resources:32-38-11-45-26-27-28-29-30  responsible:38-10-11-27-28-29-30  restarting:29  resubmit:12  result:41-21  return:42-43  returns:42  robustness:16  running:11-36-29  runs:17-45-38  scala:36  scalability:32-17-15  scalable:20-23-22-15  scale:8-48  schedule:27-11  scheduled:12  scheduler:32-28-29  schedules:45  scheduling:26-27-29  search:23  select:21  selecting:42  send:45  sends:30  sensitive:51  separate:17-4  series:48  server:8-21  service:24-29-38  serving:10  set:10-45-13-14  setnumreducetasks:14  sets:51-10-35  shared:40  shoulders:35  signals:12  signifies:16  simultaneous:4  single:4-8-15-17-26-27-30  sites:23  size:51-13  slave:38  slavearchitecture:10-11  slaves:10-11  slots:33-12 | slowing:50  smaller:4  software:19  solr:23  source:8-42-20-23  sources:40  spark:1-2-34-35-36-37-38-39-40-41-42-43-44-45-46-47-48-49-50-51  specialized:21  specific:8-11-12-29  speed:50-36  split:26  sql:19-36-21-47  sqoop:21  stage:44  stages:44  standalone:17  standard:47  stands:26-37  stating:30  status:12-30  step:44-45  storage:8-9-10-40-15-47-19-20  store:10-20  stored:10-21-15  stores:50-36  storing:50  stream:32-48-51-6  streaming:48-36  structure:37  stumble:20  submissions:29  submit:12  submits:12-44  submitted:44-30  suitable:33  support:47  supports:32-36  synchronization:24  systems:47  takeordered:43  takes:35-15  talks:12-45  task:4-8-12-45-15  tasks:35-4-38-9-11-12-13-14-44-45  tasktracker:11-12  tenancy:32  thousands:8-32  throughput:15  time:32-35-4-6-50-51-20-21  times:6  tolerance:16-17-50  tolerant:16-40-37-23  tool:21  tracking:30  traffic:15  transformations:41-42-50  translation:38  tree:10  true:42  twitter:19-20-22  types:41  typically:4  underlying:47  unfavorable:16  units:44  unreliable:12  updates:12  usage:50-30  user:38-44-6-14  users:19-47  utilization:32  utilized:17  utilizes:17  values:42-37-14  version:33-26  view:45  volume:10  wahid:1  ways:40  web:23  work:35-12  worker:45-38  workers:38  workflow:44-45  working:30  works:21  writable:14  writablecomparable:14  write:10-20-50-36  writes:50  writing:19-47  written:8  yahoo:19-20-22  yarn:32-33-2-35-25-26-27-28-29-30-31  zookeeper:24 |
| --- | --- | --- | --- | --- | --- | --- |

###############################################Lecture 4#############################################

| absence:26  accept:27  accuracy:41  address:21  agenda:2  aggregates:37  alarm:42  algorithm:8-9-27-21  allowing:36  amounts:22  analytics:1-4  appears:30  application:36  applied:22  applies:31  applying:2-35-36-37-20-21-22-23-24  approach:22  arithmetic:45  array:9  ascending:9  assign:9-18-5  assigns:34-29  assume:33  assumes:26  assumption:30  attribute:32-33-36-37-30  attributes:28-29  attributevalue:36-37  average:48-19-45  averaged:19  based:8-9-17-22-26  basis:9  bayes:32-33-2-34-35-36-6-37-25-26-27-28-29-30-31  big:1-2-35-36-37-20-21-22-23-24  building:32-19  calculate:48-9  calculated:32-45-47  calculating:17  calculation:17  calculations:30  called:42-43  cases:8-22  categorial:18-12  categorical:27  category:8-18  character:5  characters:5  choose:9 | class:32-33-34-36-5-37-39-9-26-27-28-30  classes:48-45-46-47  classification:5-6-22-24-26-31  classified:40-41-29-5  classifier:32-33-2-34-35-5-36-37-39-40-42-25-26-30-31  classifiers:2-38-39-40-41-42-43-44-45-46-47-48  classifies:8  classify:5-22  cluster:22  cms:17  collect:32  complexity:21  computation:36-21  compute:24  computed:33-42-43-19-23  computes:36  computing:32  conditional:32-26-28-30  conditionally:30  confusion:47-39  consumption:21  continuous:27  convert:27  correctly:40-41-43  corresponds:29  cross:19  customer:13  data:1-2-35-4-36-37-8-9-17-19-20-21-22-23-24  dataset:8-9-22  datasets:21  deal:21  decision:6  decreases:36  defined:41  defining:41  definitive:24  denominator:30  denoted:28  determine:24-23  dimensions:34  discretization:27  distance:8-9-11-12-14-17-18-23  distances:21  distribute:22  divided:41-36-23  dividing:19  drawbacks:22  earlier:31 | easily:21  email:5  emitted:23  equal:19-12  equation:41-28  error:42-19  estimates:19  euclidean:9-11-14  evaluation:2-38-39-40-41-42-43-44-45-46-47-48  event:28  events:26  exact:24  examples:5-6-22  exceed:21  exercise:48-46-47  explained:9  extended:30  false:40-42-45  feature:26-34  features:34-11-12-18-21-26  figure:39  final:24  fns:45  fold:19  folds:19  form:29  fpr:42  fps:45  frequency:36-37  frequent:9  function:37  functions:8-11-12  general:29  generally:27  groups:19  hamming:9-18-12  handling:18  handwritten:5  harmonic:44  height:17-13  holding:19  identified:40-43  ignore:30  impact:30  implementation:24  implementing:9  include:5  includes:27 | independence:30  independent:17-30  individual:45  influence:17  input:27-36  instance:5  instances:40-41-43-12-21  integer:9  introduction:2-3-4-5-6  involves:19  key:36-37  kgs:17  knn:2-6-8-9-16-20-21-22-23-24  knowing:32  label:32-34-36-5-37-27-28-29-30  labels:32-33-34  lacks:21  large:21-22  largest:29  law:26  layout:39  learning:4-5  learns:5  lecture:1  length:12  limitations:16  list:24  load:9  lydia:1  machine:4-5  machines:6  macro:48-45  main:21  manhattan:9-11  map:24-36-23  mappers:37  mapreduce:36-22  maps:24  marked:42-43  matrix:47-39  maximizes:34  measure:8-12  measured:17  memory:21  mentioned:31  method:9-26  methods:17  metric:41 | micro:48-45  mining:4  model:41-45-19-21-24  motivate:22  multi:48-45-46-47  multiple:36-29-23  nearest:2-6-7-8-9-10-11-12-13-14-15-16-17-18-19-22-23-24  negative:40  negatives:40-42-45  neighbor:2-6-7-8-9-10-11-12-13-14-15-16-17-18-19-22  neighbors:8-24-23  networks:6  neural:6  nodes:22  normalization:17  number:34-37-40-41-11-12-18-21  object:29  observed:28  occurred:28  occurring:28  optimal:19  order:9  original:24  output:27  outputs:36-37  pair:36-37  parallel:24  percent:42  percentage:43  performance:2-38-39-40-41-42-43-44-45-46-47-48  phase:36-37-8-23-24  point:9  points:9  positive:40-42-43  positives:40-45  precision:48-43-44-45  prediction:5  predictive:1-4  presence:26  probabilistic:26  probabilities:32-33-26  probability:27-28-30  problems:21-36-5  procedure:19  process:27-19  processes:36  processing:22  product:31 | promising:21  proportional:31  purposes:5  ram:21  randomly:19  rapid:21  rate:41-42-43-19-24  reality:40  recall:48-43-44-45  record:34  records:41  reduce:24-37-23  referred:27  relationship:26  relative:30  remaining:19  removing:30  repeated:19  repeats:19  require:21  required:32  requirement:32  requires:32  result:31-23  row:9  rows:9  runtime:21  scalability:21  scale:17-21  scaling:17  score:48-27-44-45  scores:45-30  selected:19  set:32-34-19-5  shirt:13-14  similarity:8  simplification:30  simplifications:30-31  simplify:30  simultaneously:22  single:36  size:19-13-14  smallest:19  sort:9  sorted:9  spam:5  specific:39  split:36 | splits:36-23  standardization:17  statistics:32-36  step:9  store:21  stores:8  strings:12  subset:19  success:41  suitable:12  sum:41-45  supervised:4-5  support:6  svm:6  symbols:12  table:39  techniques:4-6-22  test:9-34-22  tested:34  testing:34  theorem:26-28-29-30  time:19  times:19-36-31  total:41  tpr:43  tps:45  traditional:21  train:19  training:32-36-8-9-17-19-21-22-23  trees:6  true:40-43-45  type:42  typically:27  units:17  unrelated:26  unseen:5-22  validate:19  validation:19  values:32-33-19-23-30  variable:17  variables:17-27  variations:27  variety:21  vector:6  visualization:39  wahid:1  ways:27  weight:17-13  weighted:48-45  wide:21  working:9  yields:19 |
| --- | --- | --- | --- | --- | --- | --- |

###############################################Lecture 5#############################################

| account:12  actual:20  adding:17  advantages:18  agenda:2  agglomerative:9  aim:20  algorithm:14-15-19-20-22-32-33-34-35-36-37-38-39-40-41-42-43-44-45  algorithms:32-54-22  analysis:54-6-30  analytical:13  analytics:1-4  answer:27-15  appearing:47  application:12-54  applications:6  apply:44  applying:2-49-50-51-52-21-22-23-24-25-53  approach:32-9  approaches:9-10  apriori:32-33-34-35-36-37-38-39-40-41-42-43-44-45  aprioriproperty:34  areas:10  arguments:52  arithmetic:13  assign:14  assigned:15  assignment:12  assigns:20  association:2-26-27-28-29-30-31-32-33-34-35-36-37-38-39-40-41-42-43-44-45-46-47-48-49-50-51-52-53-54  assume:41  assumes:12  assuming:35  attributes:13  automatically:12  average:24-13  based:6-7-10-11-12-13-16-51-54  basis:34  basket:30  behavior:28  belong:8  belongs:10  benefit:17  bhattacharya:54  big:1-2-49-50-51-52-21-22-23-24-25-53  blocks:50  bodoia:54  bought:28  bread:33-34-35-36  broadcast:23  broadly:6  business:28  buy:27  buyya:54  calculate:41-42-43  calculated:11-15  calculates:24-25  called:20-30  candidate:32-37-46-47-48-50-52  card:6  case:17-20  cell:11  cells:11  center:20-13-15  centroid:17-19-14-15  centroids:14-15-17-18-19  cereal:29  change:15  characteristics:6  characterize:6  check:37  checks:38  choose:14  choosing:19  chosen:16-19-11-13 | classifying:6  clicked:31  close:9-17  closest:17-20-14-22  closure:34  cluster:6-8-9-10-11-13-15-17-20-54-23-24-25  clustering:2-5-6-7-8-9-10-11-12-13-14-15-16-17-18-19-20-21-22-23-24-25-54  clusters:9-10-12-13-14-15-16-17-20  coefficients:10  coincidental:48  coincidentally:46  collection:28-13-31  collects:51-52  combiner:24  common:37  commonly:12  communication:12  compared:47  component:33  computationally:18  compute:23-37-14-15  computed:39-43-15-20-25  computes:22  condition:9-37  confidence:48-45-46-39  consequent:48  considered:33-10-34-44  consists:28  constraint:12-7  constraints:12  constructs:8  cont:15  continuous:9  controlling:32  converge:18-19  convergence:25-18-15  converges:15  correspond:22  count:41-42-43-52  create:8-42-20  creates:9  credit:6  criteria:41  criterion:33-35-38-12-25  current:23  customer:6-30  customers:27  data:1-2-4-6-8-9-10-12-14-16-17-19-20-21-22-23-24-25-49-50-51-52-53  database:8-50  dataset:40-47-17-20-27  datasets:18  day:31  decided:11  decomposition:9  defined:17-10-33-15  defines:12-14  degree:10  densities:11  density:10-11-7  descriptive:1-27-4  desired:12  detection:6  determine:16-35-12  determined:13  determining:16-17  difference:47  dimensional:13  discarded:52  discards:35  disclosed:27  discover:27-28  discovery:6  disease:54  dissimilarities:20  distance:19-14-23 | distances:17  distributed:23  distribution:6  divided:11-22  divisive:9  documents:6  downward:34  drawbacks:18  earliest:32  easy:18  eds:54  eggs:35-36  elements:37  eliminate:37  eliminated:37  eliminates:32  emitted:51-52-23  emitting:50  engineering:54  ensure:48-19  entities:31  enumerated:32  equation:17-14-15  estimating:12  evaluated:36  evaluation:48-46-47  examined:16  exchanged:51  exclude:45  expand:43  expectation:12  expected:46-47  expensive:18  explaining:16  exponential:32  exponentially:32  expressed:14  extracts:50  fact:30  fast:18  faster:19  features:14-15  file:52  filter:48  finite:11  form:11-30  formally:33  forming:9  formulate:22  fraud:6  frequency:50  frequent:33-34-35-36-37-39-40-42-43-44-27  frequently:28  function:6-23  fundamental:32  fuzzy:10-7  gain:6  generally:44-31  generate:35-36-37-39-40  generated:12  generating:32-44  generation:37  global:18  goal:28  good:16  govardhan:54  greater:11-46  greatly:17  grid:11-7  group:9  grouped:51  groups:9-13-6  grows:38  growth:32  guarantee:18 | guess:16  guesses:14  hadoop:54  happening:46  hard:10  helps:6  heuristic:16  hidden:27  hierarchical:9-7  high:48  higher:10-39  holds:9  huge:24  hyperlinks:31  identified:48-42-43-36  identifies:35-13  identify:35-36-37-6-48  image:6  implement:18  improve:8  include:43  included:50  including:7  incorporation:12  independent:46-47  individual:31  infrequent:34  initial:8-19-14  initialization:18-19  innovations:54  input:50-51  insight:6  instances:24  interactive:12  interesting:48-27-28  intermediate:51  introduction:2-3-4  item:41-10-37-30  items:33-35-41-10-43-28-30-31  itemset:33-34-38-39-41-42-43-44-48-50-51-52-30-31  itemsets:32-35-36-37-38-40-42-43-44-27  iterating:25  iteration:38-9-15-22-23  iterations:44-38  iterative:8  ith:17  jam:34  job:50  join:37  joining:37  key:50-51-52-23  khatua:54  kind:31  large:32-18-27-28  larger:46-47  lead:18  learning:27-4  lecture:1-54  length:38  leverage:48-46-47  lift:48-46  likelihood:12  list:51-28-52  lydia:1  machine:4-23  major:33  map:50-51-22-23-24  mapper:50  mapreduce:54-50-22  maps:51  mark:11  market:30  markets:6  mass:15  maximum:12 | meaning:30  meaningful:28  meansis:13  measurable:13  measure:46  measures:48-46-47  medoid:20  medoids:20  member:10  membership:10  mentioned:25  merge:11  merging:9  met:25-38  method:8-9-10-11-12-13-27  methods:7-8-9-10-11-12  metric:16  milk:35-36-29  minimize:24-20  minimizes:23  minimum:33-35-39-41-42-43-45-18-52  mining:54-4-6-30  misleading:48  model:12-7  mondal:54  moving:8  multivariate:6  nearest:19  neighboring:11  networks:54  newly:15  noise:10-12  note:43-23  notes:54  notion:47  number:32-8-11-12-14-16-17-24  object:8-9-11-20  objects:6-8-9-10-11-12-13-17-20  observation:10  observations:20  observe:6  observed:30  occur:46  occurs:28  operates:23  optimal:16  order:20-23  oriented:12  original:12  oscillate:15  outlier:12-6  outliers:18-20  output:24-51-52  outputs:24  overview:28-29-30-31  pair:50-51-37-23  paired:36  pairs:42-51-52  pam:20  parameters:12  parses:50  partition:8-51  partitioner:50-51-52  partitioning:8-20-13-7  partitions:8-10  passing:52  patterns:6  percent:39  percentage:33  perform:37  performed:12  person:27  perspective:28  phase:24-25-23  phases:22 | pin:14-15  pioneered:32  point:14-15-17-19-23  points:10-15-17-19-20-22  popularity:30  predefined:16-38  probability:19-47  problem:22  procedure:25-38  process:12-38  processing:50-6  produce:10  product:27  products:27  properties:12  property:34-43-37  proportional:19  provide:12  proximity:13  prune:41-42-43-37  pruned:41  prunes:35  pruning:32  purchase:27  purchased:27-28-29-31  purchasing:28-6  questions:27  random:28  randomly:19-11  ratio:47  reach:38  reached:44-15  reasonable:16  received:50  recover:12  reduce:37-17-52-22-25  reducer:51  reducers:52  references:54  referred:34-30  refers:12-31  reformulate:22  relationship:28-47-31  relationships:27-28  relocation:8  remainder:10  remaining:37  repeat:15  repeated:25-38  repeating:38  represent:8  representative:20  represented:27  required:10  requirement:16  respective:17-52  rest:19  result:52  retail:31  retains:35  rule:48-54  rules:2-26-27-28-29-30-31-32-33-34-35-36-37-38-39-40-41-42-43-44-45-46-47-48-49-50-51-52-53  runs:38  saini:54  satisfy:41  sayal:54  science:54  searches:20  selected:45  sends:50  sensitive:18-19-20  sensitivity:18  separate:9-10  serves:6  session:31 | set:33-35-37-9-10-20-23-31  sets:37-6-22  setups:18  shared:25  shopping:30  shuffled:51  single:51-31  size:24-51-37  small:17-18  smaller:9  smarter:19  soft:10  solving:22  space:11  sparse:10  specifies:51  split:9-23  splits:50  squared:23  squares:16-17  standard:10-19-12-22  stanford:54  start:9  statistically:46-47  statistics:12  step:35-36-37-41-42-43-44-45-15  steps:35-36-37-38-39-14-15  strength:46  strong:18  stronger:47  structure:16-11  subset:34  subsets:43-37  suggests:46  sum:16-17-20-52  sums:52  superset:43  supersets:34  support:32-33-34-35-37-38-41-42-43-46-48-52  supportcount:52  supports:43  suppose:8-40  systems:54  task:50-51-52  tasks:31  technique:8-13-22  techniques:4  tend:27  term:17-33-31  termination:9  threshold:35-11-45-39  time:50-29  times:46  tool:6  transaction:50-28-30-31  transactions:33-34-35-39-40-28-30  traversed:11  triplets:37  trustworthy:48  types:7  uncovered:30  university:54  unsupervised:27-4  updated:25  user:12-31  values:25-19-52-51  vector:13  video:6  view:27  viewed:30  wahid:1  web:6  worse:16  written:52  wss:16-17 |
| --- | --- | --- | --- | --- | --- | --- |

###############################################Lecture 6#############################################

| ability:10  abstraction:10  access:35-7-8-9-10-22-24-25  accessed:33-14  accessibility:30  accessible:8-24-23  accommodate:39  action:34  add:10  addition:41  additional:43-11  adjust:44  administrative:16-18  advertisement:23  agenda:2  allocation:43-31  allowed:25  alternative:7  application:35-37  applications:10-37-39  apps:38  architecture:1-2-33-36-37-38-39-40-41-42-43-44-45-46-47-27-29-30  artifact:6  assess:39  automated:43  automatic:44  avenues:23  balancer:41  balancing:40  bandwidth:11  bare:16  based:33-6-7-8-9-43-14-24-27  basic:33-6-7  basis:10  benefits:10  bottlenecks:39  boundary:8-5-7  browser:33-34  burden:16  business:39  called:31  capabilities:9-4  capable:41  capacity:40-44-15  center:37-29-30  central:15  centralizing:24  centric:14  challenges:11  changing:7  characteristics:29  choose:26  client:33-34  close:30  cloud:1-2-3-4-5-6-7-8-9-10-11-12-13-14-15-16-17-18-19-20-21-22-23-24-25-26-27-28-29-30-31-32-33-34-35-36-37-38-39-40-41-42-43-44-45-46-47  clouds:8-9-23-24-27-29-30  collectively:29-31  combination:13  commercialized:23  common:13-22  commonly:44  communicates:37 | communication:37  communications:33  community:25-22  components:33-37-29-38  comprise:33  comprised:34-14-16-26-27  compute:39  computing:1-2-3-4-10-47-24-29  concepts:1-47-6-7  configuration:14  configured:16-43  connected:27-30  connectivity:37-14  considered:24-7  constraints:11  consumer:6-11-43-16-18-24-26-27  consumers:6-10-14-23-24-25  contained:27-14  contemporary:29  content:9-34-33  contract:43  control:25-18-14  conventional:7  converting:31  cost:23  create:14-31  created:35  creating:31  creation:23  custom:6  customers:4  data:33-37-11-26-29-30  day:29  dedicated:9-27  defined:16  defining:25  definitions:4-5  delete:34  delivered:4  delivery:2-12-13-14-15-16-17-18-19-20-23  demand:10-43  denote:5  departments:24  deploy:26  deployed:16  deployment:2-21-22-23-24-25-26-27  design:35-39  designed:6-39  determines:43  device:6  devices:10  dictate:43  dictated:43  dispersed:30  distances:11  distinct:6  distinguished:22  distributed:4-37  distributing:41  distribution:40-41-42  duplication:44  dynamic:40-43-44-45-46 | dynamically:43  easily:10  efficiency:30  efficiently:43  elastic:40-4  elements:33  enable:24-35-29  enabled:4  enables:43-37  enabling:32-33-2-34-35-28-29-30-31  enterprise:7  environment:5-6-37-8-39-14-15-16-22-23-26-27  environments:9  erl:47  established:13  evenly:41  evolving:25  examples:40  exchange:33  execute:34  existing:39  explore:39  expressing:33  external:4  features:29  fine:10  finite:8  fluctuating:11-44  fluctuations:43  form:4  formalized:13  foundation:30  fundamental:33-40-41-42-43-44-45-46  gartner:4  general:14  generally:33-18-23-25-31  generic:18  geographic:11  geographically:30  grained:10  granted:25-18  grouping:30  guarantee:25  hall:47  hand:9  handle:44  hardware:37-6-14-15-30  high:14  higher:30  hops:11  horizontal:44  horizontally:41  host:34-44-31  hosted:27-7  hosts:24  html:33-34  http:33-34  hybrid:26-22  hypervisor:31  iaas:16-13-14-15  identical:41  identified:34 | identifiers:33  identify:34  images:47-31  implementation:33-18  improved:30  include:14  includes:34  increased:11  individual:8  infrastructure:37-38-10-13-14-15-16-31  initiate:44  installation:31  instances:44  institute:4  intend:14  inter:27  interact:7  interaction:43  interface:33-37-38  interfaces:14  interlinked:33  internet:33-34-4-8-9-29-30  introduce:11  introduces:4  jointly:25  key:29  kinds:37  languages:33  latency:11  leased:15  lecture:1  level:10-14  limited:25-18-11-31  listener:43-44  load:40-41  loads:39  local:15  locate:34  located:7  locates:34  locations:24-10  locked:10  logic:41-35-43  long:24  longer:10-11  lydia:1  maintaining:16  maintenance:23  manage:31  managed:27-14  management:33-31  manual:43  markup:33  measured:4-6  mechanism:43  medium:33  member:25  members:25  membership:25  memory:39-15  metadata:33  metered:8 | middleware:37  model:13-14-16-22-27  models:2-4-12-13-14-15-16-17-18-19-20-21-22-23-24-25-26-27  modern:29  monitors:44  moved:10-7  multiple:35  multitenant:35-29  national:4  needed:10  network:37-6-11-14-30  networking:30-39  nist:4  offered:13-23  offering:18  offers:8  open:8  operating:14-31  operation:34  organization:24  organizational:7  owned:8-24-25-23  ownership:22  paas:16-17-13  packaged:13  parties:25  parts:24  party:25-6-23  pay:10  peak:39  performance:39  performs:34  physical:6-31  physically:24  platform:16-17-13  point:33  portability:11  positioned:18  potential:11  power:30  pre:16-13  premise:7  premises:24-7  prentice:47  primary:29-15  principles:39  private:24-26-27-22  privately:8  proceed:34  process:31  processing:9-26-43-44  processor:15  product:18  profile:18  program:18-6  protocol:33  protocols:9  provide:33-14  provider:6-11-13-18-23-24-25-27  providers:11-6  provision:8  provisioned:18-23 | provisioning:43-4-6-30  provisions:25  proximity:30  public:22-23-25-26-27  publicly:23  published:9  purpose:6-14  qualify:7  questions:39  read:34  ready:16  reclaimed:43  reduces:41  reference:47  refers:6  release:10  relocated:44  relocation:44  remote:9-30  remotely:8-24-4-6  remove:10  replication:44  represent:7  represents:13-14-16-18-22  request:34  requested:34  requests:44  require:11-14  requirements:15  requiring:43  reside:24  resource:34-6-7-40-41-44-15-30-31  resources:33-4-6-8-9-10-41-43-13-14-16-23-24-25-30-31  response:34  responsibility:25-11-14  responsible:23  risks:11  running:39  runs:37-31  runtime:41  saas:18-13  scalability:40-43-44-45-46  scalable:4-6  scale:39  scaled:41-44  scaling:43-44  security:11-38  sensitive:26  server:33-34-6-15-31  servers:30-15  service:4-40-13-14-15-16-17-18  services:33-26-37-38  set:29  setting:16  share:25  shared:18-11-30  sharing:30  short:10  sides:37  signals:44  simultaneously:35  single:24-44 | size:22  software:37-6-13-18-31  source:47  space:15  spared:16  specialized:4  specific:8-25-13-22  standards:4  statements:34  step:31  storage:38-30-15  style:4  supplying:9  support:39  supported:29-31  symbol:5  systems:14  tables:47  team:39  technically:24  technologies:4  technology:32-1-2-33-4-34-35-47-24-28-29-30-31  tenant:35  tenants:35  term:10-7  terminology:6-7  terms:43  thomas:47  thresholds:43  tools:14-31  type:22  types:44  typical:18-15  typically:8-16-25  unlimited:10  unnecessary:43  update:34  url:33-34  usage:43-30  user:9-37-38  users:35  utility:18  utilization:4-39-41-43-14  variable:43  versa:7  vertical:44  vice:7  view:35  vim:31  virtual:27-31-6-15  virtualization:32-37-29-31  virtualized:31  visible:37  vulnerabilities:11  wahid:1  web:33-34-8-9-29  wide:9-33  working:16  workload:40-41-42-43  workloads:44-30-39  write:34  xml:33-34 |
| --- | --- | --- | --- | --- | --- | --- |

###############################################Lecture 7#############################################

| access:4,7,9,23,27,30  advanced:9,20  aks:13,38  amazon:32,33,2,24,25,26,27,28,29,30,31  analytics:25,4,6,22  analyze:27,6,31  analyzing:37,22  apache:27,37,6  apis:25,23  appendix:2,36,37,6,7,38,39,40,11,13,21,27  application:8,25  applications:32,5,38,8,9,23,26,29  apps:9,37,7  architecture:32,29,30,31  assets:16,11,14,15  assignment:2,35,34  automated:33,47,17,19,20  automating:17,38  automl:2,40,41,47,17,18,19,20,21  aws:32,1,2,33,24,25,26,27,28,29,30,31  azure:1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,33,35,38,41,42,43,46,47  based:37,6,9,17,20,22,23,27  big:27,37,6,7  build:8,9,23  called:26,28,37  capabilities:29,5 | categories:25,35,4  category:4,5,6,7  check:6,7,11,13,21,27  choose:19,29  class:2,35,34  classification:2,35,40,21,22  click:50,53,47  cliv:35,19  cloud:1,4,6,7,8,9,10,38,13,14,20,22,25,26,27  cluster:13,52,53,27,28,29,30,31  clusters:38,6,7  code:8,16,19,12  cognitive:33,22,23  component:16,11,28  components:16,28  compute:4,37,7,11,13,52,26  computers:26,37,39  computing:1,4,37,25  configure:41,51,20,52,53,54,55,26  containerized:8,38  cpu:37,13  create:5,7,41,42,43,12,44,47,48,49,50,52,53,23,26,30  created:11,30,47  custom:22,15  data:37,6,39,9,11,12,14,16,18,20,26,27,28,29,30,31  dataset:48,41,50,49 | datastore:11,12  default:31,15  definitions:2,36,37,38,39  demand:25,7  demo:41,21  dependencies:8,15  deploy:8,13,38,7  deployment:25,38  design:19,20,6  details:51,44  develop:8,5  developer:25,4  development:8,17,13  devops:8,4  distributed:37,7  elastic:26,27  emr:32,33,27,28,29,30,31  enter:51,44  enterprise:9,37  environment:11,13,15  experiments:19,20  extraction:5,22  file:14,30  frameworks:27,29,31  gpu:37,13  hadoop:6,27,29,30,31 | hdfs:28,30  hive:32,29  hpc:37,7  hybrid:9,4  identity:9,4  include:25,11  includes:26,11,4  infrastructure:8,9,4,38  instance:26,28,13,30  integration:9,4  intended:38,15  internet:25,4,39  ipsec:27,39  job:41,11,47,51,20,52,53,56  jobs:12,31  kubernetes:13,38  language:33,35,5,22,23  latest:33,5  layers:32,29,30,31  learning:4,5,6,11,12,13,14,15,16,17,18,19,20,21,25,33,35,37,38,41,43,46,47  lecture:1,35  libraries:8,32,23  lifecycle:8,30  load:48,41,50,49  local:20,14,30  machine:4,5,6,11,12,13,14,15,16,17,18,19,20,21,25,26,33,35,37,38,41,43,46,47 | manage:8,9,4,7  management:4,38,25,29,31  manages:26,28  managing:38,31  mapreduce:32,27,31  metadata:26,14  metrics:12,20  microsoft:1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,33,35,37,41  migration:10,4  mixed:10,4  mobile:25,4  model:38,11,14,16,17,18,20,21  models:5,41,13,14,20  multicloud:9,4  natural:35,22  network:27,39  networking:25,26,4  nlp:35,22  node:28,30  nodes:28,13,38  note:30,23  number:18,37,14  objects:10,26  operate:8,7  optimize:18,7  output:16,12 | parameters:16,18,20  pay:25,7  platform:27,4,37  platforms:1,25  portal:42,23  private:27,39  problem:35,19  process:6,38,8,17,27,31  processing:35,37,16,22,28,29,31  programs:32,29  provide:25,39  provision:6,7  python:37,15  reality:10,4  referred:17,28  resource:42,12,23,29,31  resources:4,7,10,11,12,13,31  responsible:31,15  running:56,27,28  scale:37,7  scoring:16,11,38,15  scripts:12,15  sdkv:35,19  security:4,39  select:50,47  server:37,6 | service:32,33,38,13,18,22,23,26,29,30,31  services:32,33,2,4,5,6,7,8,9,10,24,25,26,27,28,29,30,31  settings:20,15  software:8,28,38,15  source:20,38  spark:32,6,27,29,31  steps:8,19,20  storage:4,12,14,20,25,26,29,30  store:28,6,30  storm:37,6  streaming:32,6  studio:41,46,47,19,21,23  supports:32,13,15  task:41,21,54,55,28  tasks:37,11,16,17,28  text:5,22  time:35,6  tools:25,10,4  training:11,13,15,16,18,19,20  types:30,13,14,15  understanding:5,22  users:26,38  virtual:4,39,7,26,27  vision:35,5  visualize:8,6  web:32,2,4,19,23,24,25,26,27,28,29,30,31  work:8,18,12,14  workspace:41,42,11,12,13,43,44,47 |
| --- | --- | --- | --- | --- | --- | --- |

###############################################Lecture 8#############################################

| analysis:9,5,6,7  analytical:11,6,7  application:9,15  applications:9,4  architecture:2,13,14,15,16,17,18  bit:33,34  bitmap:33,34  built:8,11  called:33,20  collection:8,11,4,15  complex:20,7  consists:25,28,23  constellation:24,28  cube:21,22 | data:1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,32,33,34,35  database:35,4,5,20,24  databases:9,11,4  decision:8,4,7  defined:8,11,4  defining:2,3,4,5,6,7,8,9,10,11,12  delete:9,6  dimension:22,23,25,28,29  dimensional:21,22,25,27,29  dimensions:20,23  fact:23,24,25,28,29  form:25,6,15  historical:5,6,8,9,29  includes:4,15 | indexed:33,34  insert:9,6  insertions:16,4  lakes:11,2,10,12  larger:6,22  load:16,14,15  making:8,7  management:8,16,5,6  manager:18,14,15  marts:11,2,10,12  massive:18,12  mining:35,7  model:20,23  modeling:32,33,2,34,19,20,21,22,23,24,25,26,27,28,29 | models:24,20,22  moved:16,5  multidimensional:20,22,23  multiple:8,28,22  normalized:9,27,6,25  olap:9,35,6,7  oltp:9,4,5,6  online:4,6,7  organization:8,11  organizes:18,21  oriented:8,9  performance:33,20,29,6  point:8,29  process:8,16 | processing:4,6,7  purposes:5,7  quarter:21,22  queries:28,20,5,6  query:18,20,14  raw:16,12  records:29,5  region:21,22  relational:24,20  required:18,15  retrieval:4,6  row:34,29  sales:29,21,22  schema:35,24,25,26,27,28 | set:16,28  single:25,11  smaller:6,22  snowflake:24,27  specific:29,21  star:24,25,26,27  storage:33,12  stored:18,22  stores:12,29  structured:16,12  subject:8,9  support:33,4,6,7,8  systems:8,35,7  table:25,29,23 | tables:6,23,25,27,28  techniques:33,35  term:5,22,7  time:8,29  traditional:9,4  transaction:9,4  transactional:4,29  types:24,29,23  typically:6,22  university:35,5  update:9,6  usage:9,18  user:18,15  users:18,11  vector:33,34  vectors:33,34  warehouse:1,2,3,4,5,6,7,8,9,11,12,13,14,15,16,17,18,22,24,28,33,35  warehouses:2,4,6,7,11,19,20,21,22,23,24,25,26,27,28,29,32,33,34  warehousing:18,35 |
| --- | --- | --- | --- | --- | --- | --- |

###############################################Lecture 9#############################################

| achieve:29,6  achieved:10,20  acid:40,37  attaining:21,39  availability:33,35,37,40,10,20,25  base:32,2,37,38,39,40  basically:37,38  cap:32,33,2,35,34,36,37  chosen:36,14  clients:40,39  collectively:4,29  combined:25,26  communication:33,35,36,15  concern:40,15 | concurrency:37,20,21  consistency:33,35,36,37,38,39,40,15,20,21,26  consistent:39,13,15,16,21,22  copied:16,18,22,15  data:1,33,35,4,38,7,8,10,11,16,20,22,28  database:33,35,36,37,38,39,20,21,22  dataset:10,4,29,6  distributed:33,4,37,36  ensure:35,20,15  ensures:40,10  eventual:37,38,39  eventually:16,20,21,22  fails:10,14  failure:8,33,29,38 | fault:36,10,25,26,29  figure:28,31  form:33,38  helps:8,29  horizontal:13,6  illustrates:28,31  improved:36,6  inconsistency:16,20,15  inconsistent:35,38,40,16,21,22  increased:25,29  involved:29,21  loads:13,6  locking:40,20,21  managed:26,13 | master:2,10,11,12,13,14,15,16,17,18,21,26,28,29  multiple:33,4,6,8,40,10,11,20,26,29  node:33,4,6,8,10,11,13,14,15,18,28,31  nodes:33,35,4,36,6,39,10,11,13,14,18,26  occur:11,20  occurs:26,22  operations:26,29  optimistic:37,20,21  partition:33,35,36  partitioning:8,4  peer:2,10,18,19,20,21,22,23,29,31  peers:18,20,21,29,31  performance:8,13  previous:28,31 | process:4,39  propagated:20,39  properties:33,35,37  provide:33,26  query:8,7  read:33,6,38,11,13,15,16,22,26,29  reads:38,7,39,14,18,21,28,31  record:40,20,15  remain:35,21  remains:20,21  replicas:10,26,31  replicated:10,11,29,31  replication:2,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,28,29,31  request:33,38 | responsible:4,6,7,29,31  scalability:36,6,10,25,29  scaling:13,6  scenario:16,22  served:28,31  shard:4,7,8,26,28,29,31  sharding:2,3,4,5,6,7,8,24,25,26,28,29,31  shards:4,7,8,26,28,31  simultaneously:18,35  single:8,26,29  slave:2,10,11,12,13,14,15,16,17,18,21,26,28  slaves:26,15  soft:40,37,38,39  stored:8,4 | subset:29,7  success:33,38  supported:36,29,14  systems:40,33,36,37  theorem:32,33,2,35,34,36,37  time:20,21  tolerance:33,35,36,8,10,25,26,29  unavailable:35,20  update:11,20,15  updated:16,38,20,22  updates:16,20,21,22  user:16,38,22  write:33,6,39,11,13,18,20,26,29  writes:7,13,14,18,28,31  written:11,38 |
| --- | --- | --- | --- | --- | --- | --- |

##############################################Lecture 10#############################################

| access:39,9,10,46,49,20,56,31  accessed:10,11,31  acid:32,35,36,13,51,20,31  aggregate:29,22,30  amounts:43,13  analytics:37,5,40,42,51  apis:56,20,46  application:46,14,15,22,56  applications:51,28,30  approach:49,21  approaches:4,5  attributes:32,33,28  availability:9,20,21,49  base:35,20  based:32,33,5,41,10,49,19,20,24,56  batch:43,5  big:1,4,5,40,10,42,23  capability:17,10,20,23  capable:9,14  capacity:40,41,39  change:56,30,31  characteristics:19,20,21,22,23  client:48,53,55  cluster:40,19,5,14  column:24,31  compared:41,49  compliance:32,20  compliant:35,36,13,31  comprising:10,30 | consistency:32,35,49,19,28  consistent:19,20,30  cont:20,21,22  continuous:42,55  cost:41,39,7  costly:33,13  creates:17,10  customer:48,53  data:1,2,4,5,9,10,11,13,14,15,17,19,20,21,22,23,24,27,28,30,31,33,36,39,40,41,42,43,44,45,46,47,48,49,51  database:2,4,7,44,13,14,17,50,51,20,52,22,53,54,55,56  databases:2,12,13,14,15,17,18,19,20,21,22,23,24,25,27,28,29,30,31,32,33,34,35,36,37  datasets:9,10,11,43,23  delete:27,30,31  describes:48,53,55  develop:19,39  developing:51,42,37,21  device:32,33,39,40,9,41,11,42,43,17,21,27,28,30,31  devices:32,35,36,41,44,19,20,21,22,23,24,51,29  differentiate:19,20,21,22  disk:2,5,6,7,8,9,10,11,12,13,14,15,17,18,34,40,41,49,51,56  distributed:2,7,8,9,10,11  document:24,29,30  drives:39,5,7  employ:51,13  enterprise:43,4  entities:32,33  entity:27,22  event:42,20,55  eventual:35,19,49 | evolution:31,30,23  examples:49,51,37  exist:19,28  expensive:41,5  failure:20,14  family:24,31  fast:35,39,9,10,42,13,17  faster:40,49  fault:35,19,36,49  field:29,31  fields:30,31  file:2,7,8,9,10,11  files:10,11  form:48,19,46  fully:27,22  generally:32,4,36,7,39,11,13,17,20  good:21,13  graph:24,33,32,22  grid:44,46,47,48,49  grids:2,45  hard:40,39,7  high:37,9,17,49,20,21,27  highly:35,39,17,19,23  horizontal:13,21,5  ideal:17,10  imdb:44,51,52,53,54,55,56  imdbs:49,51  imdg:44,46,47,48,49  imdgs:49,46 | implementations:20,37  implemented:9,44,7  inappropriate:33,43,28,30,31  include:49,51,37  increased:40,11  insert:17,27,30,31  introduction:2,3,4,5  involve:30,31  involving:27,13  issues:51,55  joins:31,22,30  key:46,48,24,27,28,29  large:33,35,9,10,11,43,13  latency:40,42,17,49,51,23  leverage:51,37  list:19,20,21,22  logic:14,15  long:43,7  main:40,29,39  manipulation:48,28,36  manner:10,4  medium:44,39  memory:2,5,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56  model:35,19,36  multiple:9,11,14,46,19,28,30  nature:9,51  nested:33,22  newsql:2,35,34,37,36  node:32,33,43,13,20 | nodes:33,5,43,46,49,19,20  nosql:2,18,19,20,21,22,23,24,25,27,28,29,30,31,32,33,35,36,49,51,56  number:33,10  object:46,48,22  objects:49,22  operation:33,28,29  operations:35,36,27,30,31  oracle:49,14  order:43,14  pace:42,13  pairs:48,46  partial:29,30  patterns:27,30,31  performance:27,51,11,5  performed:42,30,31  previous:48,53,55  principal:19,20,21,22  processing:4,37,5,40,42,11,43,14,49  properties:36,13  provide:32,35,39,41,11,49,19,20,21,23  queries:33,20,31  query:42,14,20,55,27,30,31  ram:51,43,39  random:11,13  raw:19,11  rdbms:2,36,12,13,14,15,17,56  rdbmss:35,13,19,20,21,22  read:35,39,40,10,13,15,27  reads:19,11,15 | real:37,40,42,14,51  record:53,15  relational:2,4,36,9,13,14,17,49,19,51,53,22,56,31  replication:9,21  requests:48,53  require:17,43,28,30  required:32,33,42,11,14,56,27,28,30,31  requires:42,20  requiring:51,20,21  result:17,35  returned:48,53,15  scalability:36,5  scalable:35,19,4  scale:35,20,13,23  scaling:40,21,13  scenario:48,55,53,15  schema:9,46,17,19,23,28,29,30,31  searches:30,31  searching:33,28  select:55,27,29,30,31  semi:4,9,46,17,21,23,30  serialized:48,46  shard:14,15  simple:9,27  single:11,13,14,28,30  small:10,13  solid:5,39  solutions:42,39  sql:36,19,20,53,56,31 | storage:1,2,4,5,6,7,8,9,10,11,12,13,14,15,17,18,19,20,21,22,23,24,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56  store:33,4,46,49,22,23,24  stored:32,11,44,48,53,22,28,29,30  storing:9,33,17,30  structure:27,29,30,31  structured:4,9,17,51,21,23,30  support:35,4,36,9,41,49,51,21,23,56,31  supported:17,20,29  systems:35,37,8,9,10,11,13,21  technologies:1,19,44,21  technology:51,4  time:4,37,40,42,43,17,51,21  tolerance:49,36  tolerant:35,19  traditional:19,20,21,22  transaction:35,37,51,28,30  transactional:35,13,31  traversal:32,33  types:24,25,21  unstructured:4,9,46,17,51,21,23,27  update:17,20,28,30,31  updated:28,29,30,55  updates:33,29  utilizes:48,39,7  values:27,28,46,31  velocity:17,10,23  volumes:37,4,13  write:35,39,40,10,13,17,19,23  writes:27,11,15 |
| --- | --- | --- | --- | --- | --- | --- |