| 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 |
| --- | --- | --- | --- | --- | --- | --- |