# **IAB206 MODERN DATA MANAGEMENT - ASSIGNMENT 3**

Ву

Aaron Gietzel - n9261419

Ryan Indrananda – n10852565

Hoang Minh Nguyen - n10375694

Pawel Hepnar - n10906479

# **Table of Contents**

Task 1	3
Task 2	4
Task 3	5
Task 5	8
Task 6	10
Task 7	12
Task 8	13
Task 9	14
Task 10	15
Task 11	16
Task 12	17
Task 13	18
Task 14	19
Task 15	20
Task 15 - Name: Pawel Hepnar	21
Task 15 - Name: Ryan Indrananda	23
Task 15 - Name: Aaron Gietzel	25
Task 15 - Name: Hoang Minh Nguyen	29
DECLARATION FORM	31
APPENDIX – STUDENT EVALUATIONS	32

Task 1: Consider the dataset provided as your organisation's dataset and your team to be the data analytics team. Familiarise yourself with the dataset. Have a look at the fields, understand what they mean, and the data they contain. Based on your evaluation of the dataset, provide a one-page report that is expected to be presented to the management of your organisation so that they are familiar with the dataset when you present the analysis to them (see Task 15).

The team at Ace Analytics Ltd have been commissioned by ACME Software, a new Australian start up enterprise who have developed and successfully launched a new cloud-based business resource planning software application. In discussions with the Chief Operating Officer at ACME it has been indicated that the company is interested in expanding operations by taking their new software to prospective customers who may likely be interested in their new product offering.

# Data and Service offering of Ace Analytics.

Ace Analytics have at their disposal a data repository which contains an extensive amount of corporate data which includes information pertaining to 18,800 companies. Information available upon each company is quite detailed and there are approximately 38 fields available which may be queried upon. Some fields contain extended arrays which may contain very large amounts of list data. Employees (past and present) and product lists of organisations in the database are in some cases very large. One key field available in the database is called 'category code' which contains the business sector in which each respective company operates. One such sector in the database is nominally called 'software' which may be especially interesting to the management of ACME software as this field may be used to perform queries and gain insights specific to the software sector in which they operate. The database also includes many other attributes one might expect of a database dedicated to containing company information. Attributes such as company addresses, websites, social media channels, company foundation dates, contact emails and phone numbers are readily available. Perhaps some of the more interesting fields where a significant amount of information include Products Offered, total employees, lists of employees (past and present), capital raising information with detailed tranche data, organisational acquisitions & investments, office addresses with GPS locations, organisational milestones, external websites, news links and business partners. Some companies in the database contain information relevant to Initial Public Offerings, stock exchange codes, date and currency codes. There are also additional fields which may also be used for analytic purposes which contains embedded images and videos too.

The 'companies' dataset contains a significant amount of data relevant to numerous companies that can be used to analyse embedded information. Ace Analytics may with a high degree of confidence be able to derive and deduce from the 'companies' database insightful information such as the size of a company, number of employees, acquisitions, valuations, investment opportunities and ability to raise capital in the software sector on behalf of ACME Software.

Task 2: Describe a scenario and write a query to summarise (summarise would mean finding the mean, median, and other value of interest) a field of your choice from the entire dataset.

**Scenario**: Our company wants to know an average number of employees working in software industry from 2012.

**Business value**: With over 21,000 employees that produce innovative technologies to make the world a better place, we are constantly in on a lookout for new talents.

# **Query screenshot**:

```
MongoDB Enterprise atlas-laagab-shard-0:PRIMARY> db.companies.aggregate([
... {$match: {founded_year: {$gte: 2012}, category_code: "software"}},
... {$group: {_id: "$item",avg_number_of_employees_from_2012_in_software: { $avg: "$number_of_employees"}}}
... ]).pretty()
{
        "_id": null,
        "avg_number_of_employees_from_2012_in_software": 31.6666666666668
}
MongoDB Enterprise atlas-laagab-shard-0:PRIMARY> _
```

Task 3: Describe a scenario and write a query that uses \$match operator to match documents against two conditions (e.g. age greater than 20 and year less than 2010) and then uses \$project operator to project any three fields.

**Scenario**: We are interested in a list of the companies that have more than 20 employees and was founded before 2010.

**Business value**: With increasingly raising demands for innovation, we are on the lookout for experienced companies that could help us widen our field of expertise.

# **Query screenshot:**

```
db.companies.aggregate([

{$match: {number_of_employees: {$gt: 20}, founded_year: {$lt: 2010}, category_code: "software"}},

{$project: {_id:0, name:1, total_money_raised:1, founded_year:1}}
```

# **Output screenshot:**

]).pretty()

```
MongoDB Enterprise atlas-laagab-shard-0:PRIMARY> db.companies.aggregate([
  .. \ \{shatch: \{number\_of\_employees: \{\$gt: 20\}, founded\_year: \{\$lt: 2010\}, category\_code: \ "software of the content of the c
  ... {project: {_id:0, name:1, total_money_raised:1, founded_year:1}}
 ...]).pretty()
{ "name" : "Zoho", "founded_year" : 2005, "total_money_raised" : "$0" }
{ "name" : "MotionDSP", "founded_year" : 2005, "total_money_raised" : "$0" }
                        "name" : "GoingOn",
                        "founded_year" : 2008,
"total_money_raised" : "$8.5M"
                         "name" : "Adobe Systems",
                        "founded_year" : 1982,
"total_money_raised" : "$0"
                        "name" : "PayScale",
                        "founded_year" : 1999,
"total_money_raised" : "$31.1M"
                        "name" : "Bazaarvoice",
                        "founded_year" : 2005,
"total_money_raised" : "$131M"
                        "name" : "spigit",
                        "founded_year" : 2006,
"total_money_raised" : "$55.6M"
                         "name" : "Jive Software",
                        "founded_year" : 2001,
"total_money_raised" : "$69.4M"
      "name" : "Vidoop", "founded_year" : 2006, "total_money_raised" : "$0" }
                        "name" : "INgage Networks",
                        "founded_year" : 1999,
"total_money_raised" : "$0"
                         "name" : "Dropbox",
                        "founded_year" : 2007,
"total_money_raised" : "$257M"
                         "name" : "RatePoint",
                        "founded_year" : 2006,
"total_money_raised" : "$24.5M"
                         "name" : "Kaltura",
                        "founded_year" : 2006,
"total_money_raised" : "$69.1M"
                         "name" : "Digitalsmiths",
                        "founded_year" : 2006,
"total_money_raised" : "$31.5M"
                         "name" : "Trutap",
                        "founded_year" : 1989,
"total_money_raised" : "$14.5M"
      "name" : "xkoto", "founded_year" : 2005, "total_money_raised" : "$11M" }
```

Task 4: Describe a scenario and write a query that has the following elements:

- Matches the documents against a given condition (e.g. number of employees greater than 500)
- Groups the documents using a logical \_id field
   Provides aggregated information for each group (you may use \$max, \$min, \$avg, \$sum)

**Scenario**: Our business has an interest in investing into a software company, they want to know the number of software companies that have been established later than the 2000s.

**Business value**: As the software is thriving, on the purpose of investment into building software companies, the number of founded software companies each year is valuable, business can analyse the competition and market potential in the industry.

# **Query screenshot:**

```
db.companies.aggregate([
{$match: {founded_year: {$gte: 2000}, category_code: "software"}},
{$group: {_id: {founded_year: "$founded_year"},
total_companies: {$sum: 1}}},
{$sort: {"_id": 1}}
]).pretty()
```

Task 5: Describe a scenario and write a query that skips through some documents in the third stage of the aggregation pipeline.

**Scenario**: Our company wants to seek out the names of former software managers of large companies which have a number of employees greater than 300, skip the first name then sorting as number of employees because the company should not be too small.

**Business value**: With the information about the former software managers of large companies, the business can recruit or analyse them to take advantage of talented employees who can contribute significantly to the company.

# **Query screenshot:**

```
MongoDB Enterprise atlas-u2uss9-shard-0:PRIMARY> db.companies.aggregate([
... {$unwind: {path:"$relationships"}},
... {$match:
... {number_of_employees: {$gte: 300},
... category_code: "software",
... "relationships.is_past": true,
... "relationships.title": {$regex:/software.*manager/i}
11
         }},
{$skip:1},
{$project: {_id: 0, name: 1, "relationships.title": 1, "relationships.person": 1}},
{$sort: {"number_of_employees": -1}}
]).pretty()
                   "name": "Adobe Systems",

"relationships" : {
    "title" : "Senior Software Engineer and Manager",
    "person" : {
        "first_name" : "Mark",
        "last_name" : "Hagerty",
        "permalink" : "mark-hagerty"
                   "name" : "Microsoft",

"relationships" : {

    "title" : "Software Engineer/Development Manager / Program Manager",
    "person" : {

        "first_name" : "Constantine",
        "last_name" : "Korovkin",

        "permalink" : "constantine-korovkin"
                    "name" : "Microsoft",

"relationships" : {

    "title" : "Software Architect,,Principal Development Manager",
    "person" : {

        "first_name" : "Hans-Martin",
        "last_name" : "Will",
        "permalink" : "hans-martin-will"
                   "name" : "Microsoft",

"relationships" : {

    "title" : "Software Development Manager",
    "person" : {

        "first_name" : "Guy",
        "last_name" : "Taylor",
        "permalink" : "guy-taylor"
                    "name" : "Microsoft",
"relationships" : {
"title" : "Software Design Engineer, Program Manager",
                                        "person" : {

"first_name" : "Daniel",

"last_name" : "Shimshoni",

"permalink" : "daniel-shimshoni"
MongoDB Enterprise atlas-u2uss9-shard-0:PRIMARY>
```

Task 6: Describe a scenario and write a query that sorts documents in the third stage of the aggregation pipeline.

**Scenario**: Our company is interested in investing in rapid-growth companies, and one of the most important criteria is the large company founded later than 2005 with most offices.

**Business value**: As companies established in recent years scaling up, the scaling branches culture could be the element of successful companies. Being knowledgeable of these companies, our business can have better insights in selection and analysis in investment.

# **Query screenshot**:

```
MongoDB Enterprise atlas-laagab-shard-0:PRIMARY> db.companies.aggregate([
 ... {$match: {founded_year: {$gte: 2005}, category_code: "software"}},
...]).pretty()
            "name" : "Veeam Software",
            "number_of_employees" : 1500,
            "num_offices" : 18
  "name" : "Xero", "number_of_employees" : 507, "num_offices" : 7 }
            "name" : "Basho Technologies",
            "number_of_employees" : 110,
"num_offices" : 5
  "name" : "AtelMedia", "number_of_employees" : 13, "num_offices" : 4 }
"name" : "Integrate", "number_of_employees" : 120, "num_offices" : 4 }
"name" : "Integrate", "number_of_employees" : 120, "num_offices" : 4 }
"name" : "ZeroTurnaround", "number_of_employees" : 70, "num_offices" : 4 }
"name" : "Synthesio", "number_of_employees" : 75, "num_offices" : 4 }
            "name" : "Maas Global Solutions Corporation",
            "number_of_employees" : 43,
            "num_offices" : 4
  "name" : "Veeva", "number_of_employees" : 510, "num_offices" : 4 }
            "name" : "Tiempo Development",
            "number_of_employees" : 230,
"num_offices" : 4
            "name" : "Varonis Systems",
"number_of_employees" : 120,
            "num_offices" : 4
   "name" : "FireScope", "number_of_employees" : 65, "num_offices" : 4 }
"name" : "Cogniance", "number_of_employees" : 300, "num_offices" : 4 }
            "name" : "Luma International",
            "number of employees" : 40,
            "num_offices" : 4
  "name" : "spigit", "number_of_employees" : 120, "num_offices" : 3 }
"name" : "Zmags", "number_of_employees" : 95, "num_offices" : 3 }
"name" : "Friend2Friend", "number_of_employees" : 20, "num_offices" : 3 }
"name" : "WSO2", "number_of_employees" : 228, "num_offices" : 3 }
"name" : "CohesiveFT", "number_of_employees" : 20, "num_offices" : 3 }
 ype "it" for more
```

Task 7: Describe a scenario and write a query that limits results to a certain number of documents in the third stage of the aggregation pipeline.

**Scenario**: Our company wants to know the top ten companies having the most competitors and what industry they are in.

**Business value**: Being knowledgeable of the most competitive industry of surrounding businesses will give our company an insight to shape current strategy and come up with new formulas to enter a market or analyse financial aspects.

# Query screenshot:

```
{ "name" : "Yelp", "category_code" : "search", "num_competitions" : 47 }
{ "name" : "Google", "category_code" : "search", "num_competitions" : 42 }

    "name" : "LibraryThing",
        "category_code" : "web",
        "num_competitions" : 39
}

{    "name" : "Dropbox",
    "category_code" : "software",
    "num_competitions" : 37
}

{    "name" : "YouTube",
    "category_code" : "games_video",
    "num_competitions" : 35
}
{    "name" : "Groupon", "category_code" : "web", "num_competitions" : 35 }
{    "name" : "37signals",
    "category_code" : "public_relations",
    "num_competitions" : 29
}
{    "name" : "Evite", "category_code" : "web", "num_competitions" : 28 }
{    "name" : "Digg", "category_code" : "news", "num_competitions" : 28 }
{    "name" : "Skype", "category_code" : "software", "num_competitions" : 26 }
```

Task 8: Describe a scenario and write a query to reshape a document in your dataset such that the names of two fields is displayed as a fieldname of your choice in the output.

**Scenario**: Where can I find a complete set of contact details to discuss the possibility of entering a joint venture with a fellow business?

**Business value**: With the prospect and possibility of investing and merging with fellow companies, it is important to be knowledgable of all the available methods in which one company may contact the other to discuss such possibilities. From an analytical standpoint, this provides a clear, healthy communication between companies, and competitors, alike.

# Query screenshot:

```
db.companies.aggregate( [ {$match: { category_code: "software" }},
{ $project: { contact_details: { $concat: [ "$email_address", " - ", "$phone_number" ] }, name:
1, _id: 0 } } ] )
```

Task 9: Describe a scenario and write a query that uses any two of these functions: \$concat, \$substr, \$toLower, \$toUpper

**Scenario**: When was the last time a company's website was updated? For readability purposes, I want the result to display the company name in all uppercase.

**Business value**: With growing societal demands, businesses have been required to continually update their business activities (such as in the marketing aspect) to consistently appeal to said demands. In essence, those businesses more consistent may be viewed as devoted to catering the public and remaining relevant within their market.

# **Query screenshot:**

```
db.companies.aggregate([ {$match: { category_code: "software" }}, { $project: { name: { $toUpper: "$name" }, _id : 0, dateTimeSubstring: { $substr: [ "$updated_at", 0, 23 ] }, yearSubstring: { $substr: [ "$updated_at", 24, -1 ] } } ])
```

```
{ "name" : "ZOHO", "dateTimeSubstring" : "Wed Oct 30 00:07:05 UTC", "yearSubstring" : "2013" }
{ "name" : "SKYPE", "dateTimeSubstring" : "Wed Dec 11 13:03:57 UTC", "yearSubstring" : "2013" }
{ "name" : "REGONLINE", "dateTimeSubstring" : "Mon Jun 17 16:44:54 UTC", "yearSubstring" : "2013" }
{ "name" : "MOTIONDSP", "dateTimeSubstring" : "Thu Jul 11 09:37:17 UTC", "yearSubstring" : "2013" }
{ "name" : "GOINGON", "dateTimeSubstring" : "Thu Apr 04 09:03:17 UTC", "yearSubstring" : "2013" }
{ "name" : "SPIGIT", "dateTimeSubstring" : "Fri Nov 01 10:28:32 UTC", "yearSubstring" : "2013" }
{ "name" : "BAZAARVOICE", "dateTimeSubstring" : "Fri Oct 18 13:47:49 UTC", "yearSubstring" : "2013" }
{ "name" : "GOLIGHTLY", "dateTimeSubstring" : "Wed Dec 11 05:28:29 UTC", "yearSubstring" : "2013" }
{ "name" : "JUMPBOX", "dateTimeSubstring" : "Thu Jul 29 02:51:22 UTC", "yearSubstring" : "2010" }
{ "name" : "BOONEX", "dateTimeSubstring" : "Fri Mar 22 17:58:07 UTC", "yearSubstring" : "2013" }
{ "name" : "PAYSCALE", "dateTimeSubstring" : "Mon Dec 30 07:38:06 UTC", "yearSubstring" : "2013" }
{ "name" : "ALSTRASOFT", "dateTimeSubstring" : "Fri Mar 29 09:25:50 UTC", "yearSubstring" : "2013" }
{ "name" : "INGAGE NETWORKS", "dateTimeSubstring" : "Fri Mar 29 09:25:50 UTC", "yearSubstring" : "2013" }
{ "name" : "DROPBOX", "dateTimeSubstring" : "Wed Jan 01 22:51:34 UTC", "yearSubstring" : "2014" }
{ "name" : "DIGITALSMITHS", "dateTimeSubstring" : "Wed Jan 01 22:51:34 UTC", "yearSubstring" : "2013" }
{ "name" : "BIGITALSMITHS", "dateTimeSubstring" : "The Mar 26 23:07:38 UTC", "yearSubstring" : "2013" }
{ "name" : "SNAPTELL", "dateTimeSubstring" : "The Mar 26 03:07:38 UTC", "yearSubstring" : "2013" }
{ "name" : "ATTRIBUTOR", "dateTimeSubstring" : "The Mar 26 03:07:38 UTC", "yearSubstring" : "2013" }
{ "name" : "ATTRIBUTOR", "dateTimeSubstring" : "Non Mar 25 03:16:23 UTC", "yearSubstring" : "2013" }
{ "name" : "ATTRIBUTOR", "dateTimeSubstring" : "Sat Oct 12 05:22:39 UTC", "yearSubstring" : "2013" }
{ "name" : "THUTAP", "dateTimeSubstring" : "Sat O
```

Task 10: Describe a scenario and write a query that uses any two of these functions: \$add, \$divide, \$mod, \$multiply, \$subtract

**Scenario**: My company is interested in acquiring another company and we want an analytical perspective on how costly this new venture might be. For example, how much revenue must we generate yearly to break-even with past acquisitions?

**Business value**: Breaking even on any activity signifies the moment in which total cost and total revenue are equal, and thus net profit is zero. Acquisitions are generally costly ventures for a business to take, so it is important to gain a relevant perspective on its financial aspects. Being knowledgeable of these perspectives will give businesses the upper hand in the decision-making process of undertaking new ventures, such as acquiring a new business.

#### **Query screenshot:**

```
db.companies.aggregate( [ { $match: { acquisitions: { "$exists": true, "$not": { $size: 0 } }, category_code: "software" } }, { $project: { name: 1, _id : 0, acquisitionCostPerYear: { $divide: [ { $arrayElemAt: [ "$acquisitions.price_amount", 0 ] }, { $subtract: [ 2021, { $arrayElemAt: [ "$acquisitions.acquired_year", 0 ] } ] } } } ])
```

```
"acquisitionCostPerYear" : 15000000 }
              "Skype".
   'name"
              "spigit", "acquisitionCostPerYear" : null }
             "Bazaarvoice", "acquisitionCostPerYear" : 16777777.77777776 }
  "name"
             "Adobe Systems", "acquisitionCostPerYear" : 212500000 }
  "name"
             "Dropbox", "acquisitionCostPerYear" : null }
  "name"
  "name"
             "Digitalsmiths", "acquisitionCostPerYear" : null }
             "Jive Software", "acquisitionCostPerYear" : null }
  "name"
             "LinkShare", "acquisitionCostPerYear" : null }
  "name"
            "DocuSign",
                            "acquisitionCostPerYear" : null }
  "name"
             "uTest", "acquisitionCostPerYear" : null }
"Vringo", "acquisitionCostPerYear" : null }
"Splunk", "acquisitionCostPerYear" : null }
"xkoto", "acquisitionCostPerYear" : null }
  "name"
  "name"
  "name"
  "name"
             "HubSpot", "acquisitionCostPerYear" : null }
             "Microsoft", "acquisitionCostPerYear" : 3571428.5714285714 }
"Workface", "acquisitionCostPerYear" : null }
"NetApp", "acquisitionCostPerYear" : 9230769.23076923 }
  "name"
  "name"
  "name"
             "Evernote", "acquisitionCostPerYear" : null }
  "name"
             "Opera Software", "acquisitionCostPerYear" : 727272.72727273 }
  "name"
  "name" : "Intuit", "acquisitionCostPerYear" : 13076923.076923076 }
Type "it" for more
```

# Task 11: Describe a scenario and write a query that uses \$redact, \$\$descend and \$\$prune command

**Scenario**: With my company finally deciding to undertake an acquisition, we believe it might be easier to discuss with businesses in our area (Sydney). What companies in our area are not yet acquired?

**Business value**: Being knowledgeable of surrounding businesses in your area promotes healthy business-to-business relationships and competition. These existing relationships may aid in providing more fruitful discussions on various ventures in comparison to those businesses undertaking the same venture with an altogether new business. Surrounding businesses may already be knowledgeable of your reputation, and thus be more willing to collaborate.

# **Query screenshot:**

```
db.companies.aggregate([ { $match: { acquisition: null, category_code : "software" } }, {
$redact: { $cond: { if: { $eq: [ { $arrayElemAt: [ "$offices.city", 0 ] }, "Sydney" ] }, then:
"$$DESCEND", else: "$$PRUNE" } } }, { $project: { name: 1, _id: 0 } } ])
```

```
{ "name" : "Boonex" }
{ "name" : "TechAnts" }
{ "name" : "CodeSense" }
{ "name" : "Shane Longman" }
{ "name" : "Campaign Monitor" }
{ "name" : "Novera" }
```

Task 12: Describe a scenario and write a query that uses the \$graphlookup operator. Limit your search to 20 documents.

**Scenario**: Acquisitions of organisations occurs quite frequently and sometimes it may be useful to know when an organisation was acquired and by who. Holding companies can influence strategic direction, governance and a range of other factors including value and this may be particularly so in the software industry. Therefore, a list of software organisations accompanied by acquisiton companies may be of value to business stakeholders and potential investors too. Please note: It is acknowledged by targeting of software companies only, the depth of a graph search where a non-software company participates in a takeover may be diminished. This is a targeted trade-off to ensure the returned result data set is on point.

**Business value**: A repository listing of software organisaations which were acquired by others could represent valuable information for organisational managers, stakeholders and investors too. Such information may provide business ownership and takeover insights not otherwise readily available.

# **Query screenshot**:

```
db.companies.aggregate([ { $match : { $and: [ {category_code : "software"},{acquisition: { "$ne": null}}]}},
```

{ \$graphLookup: { from: "companies", startWith: "\$acquisition.acquiring\_company.name", connectFromField: "acquisition.acquiring\_company.name",

```
connectToField: "name", as: "Acquiring Company" }}, { $limit: 20 },{$project: {_id:0, name:1, acquisition:{acquired_year:1, acquired_month:1}, "Acquiring Company": {name:1, homepage_url:1} }}]).pretty()
```

```
"name" : "Skype
"acquisition"
        "acquired_year" : 2005,
        "acquired month" : 9
),
"Acquiring Company" : [
                 "name" : "eBay"
                 "homepage_url" : "http://ebay.com"
        }
"name" : "SnapTell",
"acquisition'
         "acquired_year" : 2009,
        "acquired_month" : 6
},
"Acquiring Company" : [
        {
                 "name" : "Amazon'
                                  "http://amazon.com"
                 "homepage_url" :
        }
```

Task 13: Think from the perspective of a data analyst. You want to present some interesting information from the dataset, which may be useful for the organisation. Describe a scenario and write a query using the aggregation framework of MongoDB that outputs this information. The query should not be similar to any queries you have written in Tasks 2-12 but may include any operator used in the prior tasks.

**Scenario**: It is common knowledge that Silicon Valley is home to a large number of established software companies as well as many new start-ups too. With so many software companies operating in the same location, this may bring advantages for all companies alike when they may seek employee resources, partners or perhaps wish to keep abreast of competitor activities too. Unlike Silicon Valley, for software businesses who operate in Australia it may be very useful to know which other software organisations operate locally in the region.

**Business value**: A repository listing of businesses who operate within the same sector of a single country such as Australia could be very valuable information for businesses who operate in the same domain for reasons identified in the corresponding scenario.

# Query screenshot:

Task 14: Describe a scenario and write a query to create a simple mapreduce function for the dataset provided to you.

**Scenario**: The software industry has certainly changed over the last few decades with the explosion of Information Technologies. With that said, are software companies able to reliably source finance and raise capital in today's markets?

**Business value**: The ability to successfully raise capital for strategic initiatives and operational activities is an essential requirement for all businesses. A year-on-year comparison of software industry wide capital raisings will help entrepreneurs and business financiers understand trends which may give insight as to how easily a software company may secure required capital for their company and/or initiatives.

#### **Query screenshot:**

```
var mf1 = function() { for (var idx = 0; idx < this.funding_rounds.length; idx++) {
/* The IF statement ensures only funding tranches in USD are captured */
if(this.funding_rounds[idx].raised_currency_code == "USD"){ var key =
this.funding_rounds[idx].funded_year;
var value = { qty: this.funding_rounds[idx].raised_amount }; emit(key, value); } };
var rf1 = function(keyITEM, countObjVals) { reducedVal = { qty: 0 };
for (var idx = 0; idx < countObjVals.length; idx++) { reducedVal.qty += countObjVals[idx].qty; }
return reducedVal; };
db.companies.mapReduce( mf1,rf1, { out: { merge: "software_industry_raisings_usd" }, query:{"category_code":"software"} } )
db.software_industry_raisings_usd.aggregate([ { $match : {"_id" : {$gt : 1900, $lt : 2014 }}}, { "$project": { "_id" : 0, "Year": "$_id", "Raised_USD" : "$value.qty" } } ])</pre>
```

```
Year" : 1960,
                 "Raised USD" : 1800000 }
 ear" : 1998,
                "Raised USD" : 700000 }
  ear" : 1999,
                "Raised_USD" : 250000 }
       : 2000,
                "Raised_USD" : 152800000 }
        : 2001,
                "Raised_USD" : 40500000
                "Raised_USD" : 56950000 }
"Raised_USD" : 81155000 }
"Raised_USD" : 169383000 }
"Raised_USD" : 1111774000
 Year"
       : 2002,
rear" : 2003,
"Year" : 2004
         2004,
'Year":
          2005,
                "Raised_USD" :
Year":
          2006,
                                   1365851000
Year":
          2007, "Raised_USD" :
                                   2081684650
Year":
                "Raised_USD" : 2441469500
         2008,
Year" :
                "Raised_USD" : 1648441107
         2009,
Year" :
                "Raised USD" : 1603010425
         2010,
                "Raised USD" : 1842100284
         2011,
'Year" : 2012,
                "Raised_USD" : 1516591372
Year" : 2013, "Raised_USD" : 2431373675
```

Task 15\*: Write a short report summarizing your findings of Tasks 2-14. Based on the summary, provide at least 5 recommendations for improvement of business operations. Your findings and recommendations should be understandable by the top management.

# Task 15 - Name: Pawel Hepnar

Student Number: n10906479

#### Introduction

Improving business processes would be a great way to add value to your business. As your organisation develops and transforms, it becomes increasingly complex. More people are involved in day-to-day business, data is exchanged between different systems, and customer expectations and expectations rise. As a young software start-up, you are looking into business improvements through various suggestions.

This report will suggest several business operation improvements based on the dataset analysis provided.

#### **Human Resources**

In this era of technical developments and the continuous effective growth of technologies, the human workforce is regarded as one of the most significant organisational assets. Although technology and automation are replacing many labour-intensive activities, we believe that human resources are not only the foundation of progress but also responsible for technology and its economic implications. Although most HR departments prioritise the well-being of their employees, the approach they adopt may differ depending on the size of the company. In other words, a small business may view HR and its job in an entirely different light than a large one. It's critical to comprehend some distinctions between HR in small firms and huge corporations.

Most of the workforce has experienced job growth over the previous year due to the expansion of the information technology industry. Numerous small and medium-sized companies are growing, and larger companies are expanding their markets and finding ways to become more competitive. This implies that they must focus on the future viability of the business.

When the business changes, the workplace culture also focuses on organisational development. This is a reasonable examination of the couple of years that demonstrates how the software industry was fluctuating between 1960 and 2013 including the financial crisis in 2008.

# **Acquisitions**

We are during another major wave of corporate acquisitions after the feverish merger activity of the late 1960s. The common reason for the recent merger frenzy is that the market is undervaluing many good small and medium companies, making them far less expensive to buy than to construct.

When we combine this belief with the fact that many corporations have relatively strong cash positions, as well as the widely held belief that increased government regulation and economic uncertainty make internal growth strategies unattractive. It's easy to see why mergers and acquisitions have become an increasingly important part of corporate growth strategy.

# Marketing

It's easy to waste money on ineffective marketing. Social media is a great low-cost, low-risk way to promote our business. Twitter is a great tool for creating a social existence and drawing attention to our business. Global marketing can increase brand awareness, credibility, and desirability, generating synergies and economies of scale by having multiple offices across multiple countries, for example.

Shared material, videos, and photographs on social media can help by increasing relevancy in search results on social networking sites like Facebook, Twitter, LinkedIn, YouTube, and Instagram, as well as search engines like Google.

#### Investment

Software development in Australia is one of the world's fastest expanding businesses, with a significant amount of offshore and nearshore software development outsourcing to companies in other countries. Each year, each software development firm in Australia contributes significantly to the domestic economy and propels the country to new heights. In the Asia Pacific area, Australia is becoming one of the fastest expanding education and technology hubs. It attracts numerous engineering students as well as professional and certified engineers and developers to pursue their life ambitions in this country. As a result, software engineers in Australia are in high demand.

For a variety of reasons, Australia is an excellent site for software development companies to prosper. A few extremely important factors are the fast-growing software development business, the ease with which an Australian software firm can be established, and the availability of a large demand for services not only in the local market but also in software exports.

#### Collaboration

We could strengthen our relationships with a range of people and businesses to assist us to build our corporate messaging. There are several methods for accomplishing this, but they all serve the same goal: to get people talking about our brand. A strategic partnership can strengthen both organisations' defences against outsiders while decreasing the status of one partner relative to the other. Making a detailed collaboration agreement will help both parties define their time, resources, and money commitments. This outlines each company's responsibilities and aids in the project's planning, launch, and execution from start to finish.

# Task 15 - Name: Ryan Indrananda

#### Student Number: n10852565

To reiterate, ACME Software are a company in the 'software' sector looking to expand operations. As with any new ventures and prospective expansions, it is imperative to conduct our due research which will smoothen this transition and guide our decision making in both early and later stages of the business.

There is a multiplicity of key takeaways provided within the findings. The dataset utilized provides us with a constant supply of perspectives into talents within the existing sector. This is also further benefitted from our research into experienced former managers that have worked and managed in well-regarded companies in the past. In the event of expansion, the workforce must be well-equipped to tackle new difficulties and experiences that may come with the expansion, and recruiting said individuals and talents will strengthen our place in the software industry and market. Furthermore, there is an expansive array of experience denoted by the years a company has been active in the market, for example. While competitions may be produced, having a knowledge of these experienced companies aid in widening our field of expertise and market potential in the software industry, an imperative aspect to note for ACME. Further inquiries into these companies will aid in analysis into the operations and financial aspects that may come with expanding operations in the industry and offering our cloud-based business resource planning software application.

When it comes to becoming familiar with the market and the industry, creating healthy connections and relationships with fellow businesses is important for ACME to establish. It is quite possible that as our experience and reputation within the industry increases, opportunities such as joint ventures and other collaborative business activities may present itself. Being knowledgeable of where and how to contact competitors is imperative for the success when ACME may inevitably take these opportunities, and our research has underlined this and made it simple to access relevant contact details. Of course, associating with the right companies goes a long way in our success in the industry. For example, knowing which businesses are consistently appealing to new demands in the market establishes their reputation as devoted businesses and should such increase our interest in conducting collaborative activities with them.

Gaining insights into the financial side of our activities is likely the most imperative aspect of understanding the software industry. In our research, we looked at perspectives of how long and how much revenue must be achieved for external companies within the dataset to break-even with their past acquisitions. Understanding and finding patterns and trends of these costs will prepare us for the likelihood of ourselves venturing into the possibility of acquiring companies. Our research has also thus provided an initial list of companies of which has not been acquired and operate within the software industry in Australia, as well as a report on organisations and their acquisition status and who they have acquired or have been acquired by, as potential companies we may acquire or conduct other activities with in the future.

In lieu of recommendations, Ace Analytics advises ACME to utilize this report accordingly to bolster their preparations and readiness pre-expansion. Analysis of item 2 recognizes the average number of employees of a given year. ACME must use this as a perspective of how extensive their workforce must be to accommodate their activities within the software sector (an average of around 32 employees). With this analysis limited to businesses from 2012, it provides a relevancy for ACME due to the modernity of its scope. Furthermore, analysis of item 5 provides a list of experienced managers of which ACME may hire to manage and supervise this expansion. It is not enough from a business perspective to

have an extensive workforce without the hiring of someone to organize said workforce. ACME may cross-reference this list in their decision to head-hunt this position.

As mentioned, perspectives are a great aid for new businesses looking to expand within an industry or new market. Analysis of item 6 provides ACME with such a perspective of relatively new businesses within the industry that has experienced rapid growth. Regarding this, we may further analyse these specific businesses and what they have done to achieve relatively quick success within the industry. This can be used in succession with analysis of item 8, which provides ACME with the contact details of all businesses within the software industry. Research is imperative in understanding a market, and primary research is of utmost importance in obtaining the correct and relevant information. Connecting with said businesses provides not only a gateway for ACME to gain said information and perspectives, but forming healthy business-to-business relationships (and therefore competitions) will open doors for them to conduct collaborative business activities in the future.

In looking to the future, analysis of item 10 provides perspectives into the financial costs of conducting acquisition activities in the software industry. Understanding this again prepares ACME for when they may look into acquiring businesses in the future, such that they may consult with their financial team and provide this data to ensure correct decisions are made and hurdles are passed before they do indeed look into these ventures. Furthermore, analysis of items 11 and 13 provides an initial list of businesses conducting activities in the software industry within Sydney and Australia, some yet to be acquired. Connecting with these businesses early in the expansion will bolster ACME's reputation within the local software industry and increases their chances of success in lieu of acquiring local businesses in the future.

Therefore, ACME must use these findings, summarizations, and recommendations to prepare for their upcoming expansion in offering their new cloud-based business resource planning software application to prospective customers.

Task 15 - Name: Aaron Gietzel

Student Number: n9261419

#### **Summary of Findings**

This contents of this report is the product of an engagement requested by ACME Software for Ace Analytics to perform investigative data analysis to determine insights which may assist them with their plans to roll out a new software offering and expand the business. Upon completion of investigative analysis works, Ace Analytics were able to determine a number of insights and findings on behalf of ACME Software. The team at Ace Analytics considered key factors in planning processes to formulate a targeted analysis approach to deliver insights that may be both specific and relevant to ACME Software by keeping in mind ACME is a start-up software company operating in Australia.

# **Financing Of Software Companies**

Upon inspection of year on year data returned per [Task 14] it was found that total capital raisings across the software industry worldwide started to occur from the late 1990s and from that point in time capital raising activity rapidly expanded. From 2005 companies of the software industry secured capital raising of more than a trillion dollars USD worldwide and only 2 years later that number had more than doubled to exceed 2 trillion dollars. Whilst a minor decline occurred around the time of the Global Finance Crisis it is clear that the software industry has since recovered and ongoing capital raising through to 2013 continued to trend upwards in what is now a large and well established industry and is continuing to grow.

#### **Investment Data**

An interesting insight found in this report was returned via [Task 4] where the number of software companies founded in the year 2000 was 130 and this number increased fairly steadily to reach 263 by 2008. However, following the GFC formation of new software companies dramatically diminished to single figures only in subsequent years. Perhaps this data returned is influenced by a lag in new companies being added to the companies database but none-the-less this insight remains significant as data lags could not reasonably account for the extent of the drop for new software companies being created. Indeed the report returned via [Task 10] indicated many software companies have very large spending budgets to acquire new companies. Adobe for example spends more than 200 million dollars annually on acquiring new investment and strategic acquisition of target companies.

The companies database included the respective acquisition company if an acquisition had occurred for a company and such information may be very useful. Ownership structures can sometimes be opaque and in some cases ownership information may not be easily found. Holding companies may well likely influence subordinate organisations decisions and operations which could have implications for ACME Software if it's management for example were to embark upon an investment, collaboration or partnership with the nominal target organisation [Task 12].

Upon consideration of the companies data from a different viewpoint, ACE Analytics maintain that In some cases it could be useful to know if a company has never been acquired previously [Task 11]. If the management of ACME wished to pursue new investment opportunities they may duly consider if they would like to acquire a new start-up organisation for its' cultural and entrepreneurial leadership dynamics which may likely have been retained if the company has never been acquired previously. Alternately ACME could consider to acquire a company that was previously purchased by a larger organisation which may with higher likelihood have more established organisation governance systems and protocols implemented.

Of course, any investment initiative may be subject to due diligent functions. Information relevant to companies which have a large amount of competitors may be quite useful information and could influence investment decisions made by the ACME Software executive leadership team [Task 7]. Companies such as Yelp, Google and Dropbox are all large and very successful organisations, and any attempts to buy into a specific channel in which they already operate should be considered with a high degree of caution. Whilst the number of competitors of a company is a useful metric, other information may be viewed from different perspectives also which could influence investment decisions. For example, poor and infrequent maintenance of an organisations' website can indicate that the respective company may have poor governance practices and may not maintain social media channels well. Data available on such observations could sway an investment decision to be declined. Such information could be used not only for appraisal of investment decisions but could also be used to assess potential partners & collaborators also [Task 9].

#### **Operations in Australia**

Surprisingly, there were a number of software companies which have office/s within Australia [Task 13]. Also, it was also confirmed that there are several software companies operating in Sydney which have not previously been acquired [Task 11]. Such factors considered together indicate that Sydney may be a viable location in Australia in which to operate a software company and there may well be enough other local companies operating to deduce that an industry of some collective significance exists. Moreover, this observation may be of strong significance for ACME Software as future sourcing of potential partners, collaborators and developer resources may likely be possible as there is evidence of an existing significant software *industry* that is based in Australia.

#### **Software Companies On Average Are Small**

The average number of employees for software companies in 2012 was about 32 employees [Task 2]. Considering there are very large software companies in existence such as Google and Microsoft which may influence and skew average numbers, this indicates that there must be a significant number of software companies which employee a very small number of employees. In many ways this may represent encouraging news for a company like ACME Software in that it is common for software companies to start off small and have potential to grow very rapidly and can attract investment interest from larger companies. Evidence from the companies database demonstrates that it is possible for software companies to grow very rapidly [Task 6].

# **Resourcing Considerations**

ACME Software will be very aware that development of software requires very specialist skillsets. Good developers who have acquired several years of experience can be invaluable to positively influence successful software projects. If a software industry exists in a regional location or country as was discussed earlier under the heading "Operations in Australia" then it may be reasonable to deduce that developer resources may be reasonably accessible which is important with regards to future operations of ACME Software.

The companies database includes extensive data on employees past and present [Task 5]. The employee query performed with this engagement was limited only to software companies but such a query could easily be modified or further restricted to a particular county or target region for which ACME management may be consider to be of value.

# **Recommendations to ACME Software**

# **Financing**

The software industry is very large and continues to grow. The industry is demonstrated to be well supported by financiers and therefore ACME Software should reasonably expect to be able to reliably secure finance for their future business operations.

# **Investment and Expansion Considerations**

The companies database can be used to support investment and to some extent operational decisions. The software industry has grown significantly over the last 10 years or so and it may likely continue to grow in such a manner for the foreseeable future. However, prudent investment strategy and utilisation of insights available from the companies database should be considered with any investment decisions by the management of ACME Software.

#### **Operations in Australia**

The software industry in Australia appears to be of sufficient size and scale to be of a positive benefit for ACME Software. Partners, collaborators and human resources should be sufficiently available to assist ACME Software to execute its' business plans.

# Start Small, Think Big

Software companies are typically small, but they can grow very quicky and are regularly the target of major software companies as takeover targets. It is recommended the management of ACME Software consider the industry in which they intend to operate in. Hopefully their developed software application is well received by the market.

# **Resourcing Requirements**

With the existence of a relatively significant software industry in Australia, ACME Software should expect to be able to reliably secure the necessary human resources to execute their business plan. ACME should be able to begin planning of recruitment initiatives to secure human resources with a reasonable amount of confidence.

# Task 15 - Name: Hoang Minh Nguyen

#### Student number: n10375694

ACME Software is an enterprise planning to expand operations and they regconises that data analysis process provides with more insights into customers and the software industry. As behalf of Ace Analytics, the report presents the findings after giving order, structure and meaning from raw given data.

With the findings provided in this report, ACME will gain insights of acquiring new companies or entering new field more successfully after applying more deep analysis with the results of this report. Understanding the importance of evaluation in existing companies, we provide the information of local software companies in item 13, which ACME can take advantage of it to analyse and increase the chance to enter the local market efficiently. The item 9 keeps our information up-to-date when providing the time of updated information, which makes the data more reliable. Furthermore, the relationships and reference are more important than ever if a company wants to go far, so the contacts of emails and phone numbers in item 8 also give them a huge opportunity. Being knowledgable of the industry is a must, so the number of year-to-year founded software companies in task 4 is a resource for the analysis of industry growth rate. For example, the financial crisis in 2008 makes the number of founded companies dropped dramatically after a long time consistantly hitting the peak. The information of industry also relies on the competition of each sector the companies are in, so the iteam 7 extracts top ten companies having most competitors and what industry they are in. ACME should analyse this information and know more about their competitors. The rapid growth companies are also listed in item 6 by measurement of the number of offices when they expand business.

The report not only gives the information of the market, it also prepares the beneficial information of workforce in Task 2 and 5. To illustrate, the talents having experience in large companies in the past within the existing sector are the valuable resource for ACME to recruit. These talents benefit in human resources and products quality due to their experienced knowledge. Also, to ensure being the successful the new entrants of this industry, they can take average number of employees (32 employees) as a reference of building a new software company. The information shows that software companies are not too large. However, the outlier data of large companies with multiplication in the number indicate that this industry can grow really fast and have considerable competition. Thus, the preparation step is hugely important, ACME should take small steps and set a proper target for the business growth.

The financial aspects and funding rounds could also be considerably analysed. The yearly total capital raisings in the industry shown in task 14 indicate the amount of raised capital increasing overtime until the financial crisis, it then goes to a signal of uptrends after 2013. The total money raised of lastest companies also is also shown in Task 3 to see more details of raised money. Some companies have raised huge amount of money (over 200 million dollars), which shows that this market really absorbs investment. As the support to the interest of ACME's acquisition, the list of acquired software companies provided in task 12. After analysing this acquisition information, the actual local companies are not acquired are shown in Task 11. If ACME stakeholders take this investing into consideration, this could be the huge opportunity for them to take advantage of. However, the investment should be considered carefully because it could be a risk in the future.

These findings could aid in the ACME's further analysis and decision-making process. The recommendations provided in the analysis are also considerably taken as a reference. In

the future forward-looking, these considerations of investment or recruiting should take some small steps and analysis them thoroughly.

# **DECLARATION FORM**

By submitting this assignment, I am/We are aware of the University rule that a student must not act in a manner which constitutes academic dishonesty as stated and explained in the QUT Manual of Policies and Procedures. I/We confirm that this work represents my individual/our team's effort, I/we have viewed the final version and does not contain plagiarised material.

Full Name	Student Number	Signature
Hoang Minh Nguyen	n10375694	HAM.
Aaron Gietzel	n9261419	ahl
Ryan Indrananda	n10852565	4.
Pawel <u>Hepnar</u>	n10906479	Pawel <u>Hepnar</u>

# APPENDIX - STUDENT EVALUATIONS

# Peer Evaluation Form - Aaron Gietzel

Write the name and student number of each of your group members in a separate column. For each person, indicate the extent to which your team agrees with the statement on the left (SA - Strongly disagree; D = disagree; A=agree; SA=strongly agree).

Evaluation Criteria	Student Name: Aaron Gietzel Student ID: n9261419	Student Name: Ryan <u>Indrananda</u> Student ID: n10852565	Hoang Minh	Student Name: Pawel <u>Hepnar</u> Student ID: n10906479
Attends group meetings regularly and contributes meaningfully to group discussions.	SA	SA	SA	SA
Completes assigned tasks on time.	SA	SA	SA	SA
Prepares work in a quality manner.	SA	SA	SA	SA
Demonstrates a cooperative and supportive attitude.	SA	SA	SA	SA
Contributes significantly to the success of the project.	SA	SA	SA	SA
Based on these considerations, state a peer mark that each team member should receive out of 10.	10 /10	10 /10	10 /10	10 /10

# Peer Evaluation Form - Ryan Indrananda n10852565

Write the name and student number of each of your group members in a separate column. For each person, indicate the extent to which your team agrees with the statement on the left (SA - Strongly disagree; D =disagree; A=agree; SA=strongly agree).

Evaluation Criteria	Aaron Gietzel	Student Name: Ryan <u>Indrananda</u> Student ID: n10852565	Student Name: Hoang Minh Nguyen Student ID: n10375694	Student Name: Pawel <u>Hepnar</u> Student ID: n10906479
Attends group meetings regularly and contributes meaningfully to group discussions.	SA	SA	SA	SA
Completes assigned tasks on time.	SA	SA	SA	SA
Prepares work in a quality manner.	SA	SA	SA	SA
Demonstrates a cooperative and supportive attitude.	SA	SA	SA	SA
Contributes significantly to the success of the project.	SA	SA	SA	SA
Based on these considerations, state a peer mark that each team member should receive out of 10.	10 /10	10 /10	10 /10	10 /10

#### Peer Evaluation Form for Group Work

Your name \_\_\_\_\_ Pawel Hepnar - n10906479

Write the name of each of your group members in a separate column. For each person, indicate the extent to which you agree with the statement on the left, using a scale of 1-4 (1=strongly disagree; 2=disagree; 3=agree; 4=strongly agree). Total the numbers in each column.

Evaluation Criteria	Group member: Aaron Gietzel - n9261419	Group member: Ryan <u>Indrananda</u> – n10852565	Group member: MINH - n10375694	Group member: Pawel Hepnar - n10906479
Attends group meetings regularly and arrives on time.	SA	SA	SA	SA
Contributes meaningfully to group discussions.	SA	SA	SA	SA
Completes group assignments on time.	SA	SA	SA	SA
Prepares work in a quality manner.	SA	SA	SA	SA
Demonstrates a cooperative and supportive attitude.	SA	SA	SA	SA
Contributes significantly to the success of the project.	SA	SA	SA	SA
TOTALS	10/10	10/10	10/10	10/10

# Peer Evaluation Form - Hoang Minh Nguyen - n10375694

Write the name and student number of each of your group members in a separate column. For each person, indicate the extent to which your team agrees with the statement on the left (SA - Strongly disagree; D =disagree; A=agree; SA=strongly agree).

Evaluation Criteria	Student Name: Aaron Gietzel Student ID: n9261419	Student Name: Ryan <u>Indrananda</u> Student ID: n10852565	Student Name: Hoang Minh Nguyen Student ID: n10375694	Student Name: Pawel <u>Hepnar</u> Student ID: n10906479
Attends group meetings regularly and contributes meaningfully to group discussions.	SA	SA	SA	SA
Completes assigned tasks on time.	SA	SA	SA	SA
Prepares work in a quality manner.	SA	SA	SA	SA
Demonstrates a cooperative and supportive attitude.	SA	SA	SA	SA
Contributes significantly to the success of the project.	SA	SA	SA	SA
Based on these considerations, state a peer mark that each team member should receive out of 10.	10 /10	10 /10	10 /10	10 /10